Towards a formal notion of impact metric for cyber-physical attacks*

Ruggero Lanotte[1 [0000−0002−3335−234X]], Massimo Merro[2 [0000−0002−1712−7492]], and Simone Tini[1 [0000−0002−3991−5123]]

[1] Dipartimento di Scienza e Alta Tecnologia, Università dell'Insubria, Como, Italy
{ruggero.lanotte,simone.tini}@uninsubria.it
[2] Dipartimento di Informatica, Università degli Studi di Verona, Verona, Italy
massimo.merro@univr.it

Abstract. Industrial facilities and critical infrastructures are transforming into "smart" environments that dynamically adapt to external events. The result is an ecosystem of heterogeneous physical and cyber components integrated in cyber-physical systems which are more and more exposed to *cyber-physical attacks*, *i.e.*, security breaches in cyberspace that adversely affect the physical processes at the core of the systems. We provide a formal *compositional metric* to estimate the *impact* of cyber-physical attacks targeting sensor devices of *IoT systems* formalised in a simple extension of Hennessy and Regan's *Timed Process Language*. Our *impact metric* relies on a discrete-time generalisation of Desharnais et al.'s *weak bisimulation metric* for concurrent systems. We show the adequacy of our definition on two different attacks on a simple surveillance system.

Keywords: IoT system · Cyber-physical attack · Impact metric · Probabilistic metric semantics

1 Introduction

The *Internet of Things* (IoT) is heavily affecting our daily lives in many domains, ranging from tiny wearable devices to large industrial systems with thousands of heterogeneous cyber and physical components that interact with each other.

Cyber-Physical Systems (CPSs) are integrations of networking and distributed computing systems with physical processes, where feedback loops allow the latter to affect the computations of the former and vice versa. Historically, CPSs relied on proprietary technologies and were implemented as stand-alone networks in physically protected locations. However, the growing connectivity and integration of these systems has triggered a dramatic increase in the number of *cyber-physical attacks* [14], *i.e.*, security breaches in cyberspace that adversely affect the physical

* Partially supported by the project "Dipartimenti di Eccellenza 2018-2022", funded by the Italian Ministry of Education, Universities and Research (MIUR), and by the Joint Project 2017 "Security Static Analysis for Android Things", jointly funded by the University of Verona and JuliaSoft Srl.

processes, *e.g.*, manipulating *sensor readings* and, in general, influencing physical processes to bring the system into a state desired by the attacker.

Cyber-physical attacks are complex and challenging as they usually cross the boundary between cyberspace and the physical world, possibly more than once [11]. Some notorious examples are: (i) the *STUXnet* worm, which reprogrammed PLCs of nuclear centrifuges in Iran [6], (ii) the attack on a sewage treatment facility in Queensland, Australia, which manipulated the SCADA system to release raw sewage into local rivers [32], or the (iii) the recent *BlackEnergy* cyber-attack on the Ukrainian power grid, again compromising the SCADA system [15].

The points in common of these systems is that they are all safety critical and failures may cause catastrophic consequences. Thus, the concern for consequences at the physical level puts *CPS security* apart from standard *IT security*.

Timing is particularly relevant in CPS security because the physical state of a system changes continuously over time and, as the system evolves in time, some states might be more vulnerable to attacks than others. For example, an attack launched when the target state variable reaches a local maximum (or minimum) may have a great impact on the whole system behaviour [17]. Also the *duration of the attack* is an important parameter to be taken into consideration in order to achieve a successful attack. For example, it may take minutes for a chemical reactor to rupture, hours to heat a tank of water or burn out a motor, and days to destroy centrifuges.

The estimation of the *impact* of cyber-physical attacks on physical components of the target system is a crucial task when protecting CPSs [10]. For instance, in industrial CPSs, before taking any countermeasure against an attack, engineers first try to estimate the impact of the attack on the system functioning (*e.g.*, performance and security) and weight it against the cost of stopping the plant. If this cost is higher than the damage caused by the attack (as is sometimes the case), then engineers might actually decide to let the system continue its activities even under attack. Thus, once an attack is detected, *impact metrics* are necessary to quantify the perturbation introduced in the physical behaviour of the system under attack.

The *goal* of this paper is to lay theoretical foundations to provide formal instruments to precisely define the notion of impact of cyber-physical attack targeting physical devices, such as *sensor devices* of IoT systems. For that we rely on a timed generalisation of *weak bisimulation metrics* [5] to compare the behaviour of two systems up to a given tolerance, for time-bounded executions.

Weak bisimulation metric allows us to compare two systems M and N, writing $M \simeq_p N$, if the weak bisimilarity holds with a *distance* or *tolerance* $p \in [0, 1]$, *i.e.*, if M and N exhibit a different behaviour with probability p, and the same behaviour with probability $1 - p$. A useful generalisation is the *n-bisimulation metric* [3] that takes into account bounded computations. Intuitively, the distance p is ensured only for the first n computational steps, for some $n \in \mathbb{N}$. However, in timed systems it is desirable to focus on the passage of time rather than the number of computational steps. This would allow us to deal with situations where

it is not necessary (or it simply does not make sense) to compare two systems "ad infinitum" but only for a limited amount of time.

Contribution. In this paper, we first introduce a general notion of *timed bisimulation metric* for concurrent probabilistic systems equipped with a discrete notion of time. Intuitively, this kind of metric allows us to derive a *timed weak bisimulation with tolerance*, denoted with $\approx_p^k$, for $k \in \mathbb{N}^+ \cup \{\infty\}$ and $p \in [0,1]$, to express that the tolerance p between two timed systems is ensured only for the first k time instants (tick-actions). Then, we use our timed bisimulation metric to set up a formal *compositional* theory to study and measure the *impact* of cyber-physical attacks on IoT systems specified in a simple probabilistic timed process calculus which extends Hennessy and Regan's *Timed Process Language* (TPL) [12]. IoT systems in our calculus are modelled by specifying: (i) a *physical environment*, containing informations on the physical state variables and the sensor measurements, and (ii) a *logics* that governs both accesses to sensors and channel-based communications with other cyber components.

We focus on *attacks on sensors* that may eavesdrop and possibly modify the sensor measurements provided to the controllers of sensors, affecting both the *integrity* and the *availability* of the system under attack.

In order to make security assessments of our IoT systems, we adapt a well-know approach called *Generalized Non Deducibility on Composition* (GNDC) [7] to compare the behaviour of an IoT system M with the behaviour of the same system under attack, written $M \parallel A$, for some arbitrary cyber-physical attack A. This comparison makes use of our timed bisimulation metric to evaluate not only the *tolerance* and the *vulnerability* of a system M with respect to a certain attack A, but also the *impact* of a successful attack in terms of the deviation introduced in the behaviour of the target system. In particular, we say that a system M *tolerates an attack* A if $M \parallel A \approx_0^\infty M$, *i.e.*, the presence of A does not affect the behaviour of M; whereas M is said to be *vulnerable* to A in the time interval $m..n$ with impact p if $m..n$ is the smallest interval such that $M \parallel A \approx_0^{m-1} M$ and $M \parallel A \approx_p^k M$, for any $k \geq n$, *i.e.*, if the perturbation introduced by the attack A becomes observable in the m-th time slot and yields the maximum *impact* p in the n-th time slot. In the concluding discussion we will show that the *temporal vulnerability window* $m..n$ provides several informations about the corresponding attack, such as *stealthiness* capability, duration of the *physical effects* of the attack, and consequent room for possible run-time *countermeasures*.

As a case study, we use our timed bisimulation metric to measure the impact of two different attacks injecting *false positives* and *false negative*, respectively, into a simple surveillance system expressed in our process calculus.

Outline. Section 2 formalises our timed bisimulation metrics in a general setting. Section 3 provides a simple calculus of IoT systems. Section 4 defines cyber-physical attacks together with the notions of tolerance and vulnerability *w.r.t.* an attack. In Section 5 we use our metrics to evaluate the impact of two attacks on a simple surveillance system. Section 6 draws conclusions and discusses related

and future work. In this extended abstract proofs are omitted, full details of the proofs can be found in the technical report [23].

2 Timed Bisimulation Metrics

In this section, we introduce *timed bisimulation metrics* as a general instrument to derive a notion of timed and approximate weak bisimulation between probabilistic systems equipped with a discrete notion of time. In Section 2.1, we recall the semantic model of *nondeterministic probabilistic labelled transition systems*; in Section 2.2, we present our metric semantics.

2.1 Nondeterministic Probabilistic Labelled Transition Systems

Nondeterministic probabilistic labelled transition systems (pLTS) [30] combine classic LTSs [16] and discrete-time Markov chains [34] to model, at the same time, reactive behaviour, nondeterminism and probability. We first provide the mathematical machinery required to define a pLTS.

The state space in a pLTS is given by a set $\mathcal{T}$, whose elements are called *processes*, or *terms*. We use $t, t', ..$ to range over $\mathcal{T}$. A (discrete) *probability sub-distribution* over $\mathcal{T}$ is a mapping $\Delta\colon \mathcal{T} \to [0,1]$, with $\sum_{t\in\mathcal{T}} \Delta(t) \in (0,1]$. We denote $\sum_{t\in\mathcal{T}} \Delta(t)$ by $|\Delta|$, and we say that Δ is a *probability distribution* if $|\Delta| = 1$. The *support* of Δ is given by $\lceil\Delta\rceil = \{t \in \mathcal{T} : \Delta(t) > 0\}$. The set of all sub-distributions (resp. distributions) over $\mathcal{T}$ with finite support will be denoted with $\mathcal{D}_{\text{sub}}(\mathcal{T})$ (resp. $\mathcal{D}(\mathcal{T})$). We use Δ, Θ, Φ to range over $\mathcal{D}_{\text{sub}}(\mathcal{T})$ and $\mathcal{D}(\mathcal{T})$.

Definition 1 (pLTS [30]). *A* pLTS *is a triple* $(\mathcal{T}, \mathbf{A}, \rightarrow)$*, where: (i)* $\mathcal{T}$ *is a countable set of* terms*, (ii)* $\mathbf{A}$ *is a countable set of* actions*, and (iii)* $\rightarrow \subseteq \mathcal{T} \times \mathbf{A} \times \mathcal{D}(\mathcal{T})$ *is a* transition relation*.*

In Definition 1, we assume the presence of a special deadlocked term $\mathsf{Dead} \in \mathcal{T}$. Furthermore, we assume that the set of actions $\mathbf{A}$ contains at least two actions: τ and tick. The former to model internal computations that cannot be externally observed, while the latter denotes the passage of one time unit in a setting with a discrete notion of time [12]. In particular, tick is the only *timed action* in $\mathbf{A}$.

We write $t \xrightarrow{\alpha} \Delta$ for $(t, \alpha, \Delta) \in \rightarrow$, $t \xrightarrow{\alpha}$ if there is a distribution $\Delta \in \mathcal{D}(\mathcal{T})$ with $t \xrightarrow{\alpha} \Delta$, and $t \not\xrightarrow{\alpha}$ otherwise. Let $der(t,\alpha) = \{\Delta \in \mathcal{D}(\mathcal{T}) \mid t \xrightarrow{\alpha} \Delta\}$ denote the set of the derivatives (i.e. distributions) reachable from term t through action α. We say that a pLTS is *image-finite* if $der(t,\alpha)$ is finite for all $t \in \mathcal{T}$ and $\alpha \in \mathbf{A}$. In this paper, we will always work with image-finite pLTSs.

Weak transitions. As we are interested in developing a *weak* bisimulation metric, we need a definition of weak transition which abstracts away from τ-actions. In a probabilistic setting, the definition of weak transition is somewhat complicated by the fact that (strong) transitions take terms to distributions; consequently if we are to use weak transitions then we need to generalise transitions, so that they take (sub-)distributions to (sub-)distributions.

To this end, we need some extra notation on distributions. For a term $t \in \mathcal{T}$, the *point (Dirac) distribution at* t, denoted $\overline{t}$, is defined by $\overline{t}(t) = 1$ and $\overline{t}(t') = 0$ for all $t' \neq t$. Then, the convex combination $\sum_{i\in I} p_i \cdot \Delta_i$ of a family $\{\Delta_i\}_{i\in I}$ of (sub-)distributions, with I a finite set of indexes, $p_i \in (0,1]$ and $\sum_{i\in I} p_i \leq 1$, is the (sub-)distribution defined by $(\sum_{i\in I} p_i \cdot \Delta_i)(t) \stackrel{def}{=} \sum_{i\in I} p_i \cdot \Delta_i(t)$ for all $t \in \mathcal{T}$. We write $\sum_{i\in I} p_i \cdot \Delta_i$ as $p_1 \cdot \Delta_1 + \ldots + p_n \cdot \Delta_n$ when $I = \{1, \ldots, n\}$.

Thus, we write $t \xrightarrow{\hat{\tau}} \Delta$, for some term t and some distribution Δ, if either $t \xrightarrow{\tau} \Delta$ or $\Delta = \overline{t}$. Then, for $\alpha \neq \tau$, we write $t \xrightarrow{\hat{\alpha}} \Delta$ if $t \xrightarrow{\alpha} \Delta$. Relation $\xrightarrow{\hat{\alpha}}$ is extended to model transitions from sub-distributions to sub-distributions. For a sub-distribution $\Delta = \sum_{i\in I} p_i \cdot \overline{t_i}$, we write $\Delta \xrightarrow{\hat{\alpha}} \Theta$ if there is a non-empty set of indexes $J \subseteq I$ such that: (i) $t_j \xrightarrow{\hat{\alpha}} \Theta_j$ for all $j \in J$, (ii) $t_i \not\xrightarrow{\hat{\alpha}}$, for all $i \in I \setminus J$, and (iii) $\Theta = \sum_{j\in J} p_j \cdot \Theta_j$. Note that if $\alpha \neq \tau$ then this definition admits that only some terms in the support of Δ make the $\xrightarrow{\hat{\alpha}}$ transition. Then, we define the *weak transition relation* $\stackrel{\hat{\tau}}{\Longrightarrow}$ as the transitive and reflexive closure of $\xrightarrow{\hat{\tau}}$, *i.e.*, $\stackrel{\hat{\tau}}{\Longrightarrow} = (\xrightarrow{\hat{\tau}})^*$, while for $\alpha \neq \tau$ we let $\stackrel{\hat{\alpha}}{\Longrightarrow}$ denote $\stackrel{\hat{\tau}}{\Longrightarrow}\xrightarrow{\hat{\alpha}}\stackrel{\hat{\tau}}{\Longrightarrow}$.

2.2 Timed Weak Bisimulation with Tolerance

In this section, we define a family of relations $\approx_p^k$ over $\mathcal{T}$, with $p \in [0,1]$ and $k \in \mathbb{N}^+ \cup \{\infty\}$, where, intuitively, $t \approx_p^k t'$ means that *t and t' can weakly bisimulate each other with a tolerance p accumulated in k timed steps*. This is done by introducing a family of *pseudometrics* $\mathbf{m}^k \colon \mathcal{T} \times \mathcal{T} \to [0,1]$ and defining $t \approx_p^k t'$ iff $\mathbf{m}^k(t,t') = p$. The pseudometrics $\mathbf{m}^k$ will have the following properties for any $t, t' \in \mathcal{T}$: (i) $\mathbf{m}^{k_1}(t,t') \leq \mathbf{m}^{k_2}(t,t')$ whenever $k_1 < k_2$ (tolerance monotonicity); (ii) $\mathbf{m}^{\infty}(t,t') = p$ iff p is the distance between t and t' as given by the weak bisimilarity metric in [5] in an untimed setting; (iii) $\mathbf{m}^{\infty}(t,t') = 0$ iff t and t' are related by the standard weak probabilistic bisimilarity [27].

Let us recall the standard definition of pseudometric.

Definition 2 (Pseudometric). *A function* $d \colon \mathcal{T} \times \mathcal{T} \to [0,1]$ *is a* 1-bounded pseudometric *over* $\mathcal{T}$ *if*

- $d(t,t) = 0$ *for all* $t \in \mathcal{T}$,
- $d(t,t') = d(t',t)$ *for all* $t, t' \in \mathcal{T}$ *(symmetry),*
- $d(t,t') \leq d(t,t'') + d(t'',t')$ *for all* $t, t', t'' \in \mathcal{T}$ *(triangle inequality).*

In order to define the family of functions $\mathbf{m}^k$, we define an auxiliary family of functions $\mathbf{m}^{k,h} \colon \mathcal{T} \times \mathcal{T} \to [0,1]$, with $k, h \in \mathbb{N}$, quantifying the tolerance of the weak bisimulation after a sequence of computation steps such that: (i) the sequence contains exactly k tick-actions, (ii) the sequence terminates with a tick-action, (iii) any term performs exactly h untimed actions before the first tick-action, (iv) between any i-th and $(i{+}1)$-th tick-action, with $1 \leq i < k$, there are an arbitrary number of untimed actions.

The definition of $\mathbf{m}^{k,h}$ relies on a *timed and quantitative* version of the classic bisimulation game: The tolerance between t and t' as given by $\mathbf{m}^{k,h}(t,t')$ can be

below a threshold $\epsilon \in [0,1]$ only if each transition $t \xrightarrow{\alpha} \Delta$ is mimicked by a weak transition $t' \stackrel{\hat{\alpha}}{\Longrightarrow} \Theta$ such that the bisimulation tolerance between Δ and Θ is, in turn, below ϵ. This requires to lift pseudometrics over $\mathcal{T}$ to pseudometrics over (sub-)distributions in $\mathcal{D}_{\text{sub}}(\mathcal{T})$. To this end, we adopt the notions of *matching* [37] (also called coupling) and *Kantorovich lifting* [4].

Definition 3 (Matching). *A* matching *for a pair of distributions $(\Delta,\Theta) \in \mathcal{D}(\mathcal{T}) \times \mathcal{D}(\mathcal{T})$ is a distribution ω in the state product space $\mathcal{D}(\mathcal{T} \times \mathcal{T})$ such that:*

- $\sum_{t' \in \mathcal{T}} \omega(t,t') = \Delta(t)$*, for all* $t \in \mathcal{T}$*, and*
- $\sum_{t \in \mathcal{T}} \omega(t,t') = \Theta(t')$*, for all* $t' \in \mathcal{T}$*.*

We write $\Omega(\Delta,\Theta)$ to denote the set of all matchings for (Δ,Θ).

A matching for (Δ,Θ) may be understood as a transportation schedule for the shipment of probability mass from Δ to Θ [37].

Definition 4 (Kantorovich lifting). *Assume a pseudometric $d\colon \mathcal{T} \times \mathcal{T} \to [0,1]$. The* Kantorovich lifting *of d is the function $\mathbf{K}(d)\colon \mathcal{D}(\mathcal{T}) \times \mathcal{D}(\mathcal{T}) \to [0,1]$ defined for distributions Δ and Θ as:*

$$\mathbf{K}(d)(\Delta,\Theta) \stackrel{def}{=} \min_{\omega \in \Omega(\Delta,\Theta)} \sum_{s,t \in \mathcal{T}} \omega(s,t) \cdot d(s,t).$$

Note that since we are considering only distributions with finite support, the minimum over the set of matchings $\Omega(\Delta,\Theta)$ used in Definition 4 is well defined.

Pseudometrics $\mathbf{m}^{k,h}$ are inductively defined on k and h by means of suitable *functionals* over the complete lattice $([0,1]^{\mathcal{T}\times\mathcal{T}}, \sqsubseteq)$ of functions of type $\mathcal{T} \times \mathcal{T} \to [0,1]$, ordered by $d_1 \sqsubseteq d_2$ iff $d_1(t,t') \leq d_2(t,t')$ for all $t,t' \in \mathcal{T}$. Notice that in this lattice, for each set $D \subseteq [0,1]^{\mathcal{T}\times\mathcal{T}}$, the supremum and infimum are defined as $\sup(D)(t,t') = \sup_{d\in D} d(t,t')$ and $\inf(D)(t,t') = \inf_{d\in D} d(t,t')$, for all $t,t' \in \mathcal{T}$. The infimum of the lattice is the constant function zero, denoted by $\mathbf{0}$, and the supremum is the constant function one, denoted by $\mathbf{1}$.

Definition 5 (Functionals for $\mathbf{m}^{k,h}$). *The functionals $\mathbf{B}, \mathbf{B}_{\mathsf{tick}}\colon [0,1]^{\mathcal{T}\times\mathcal{T}} \to [0,1]^{\mathcal{T}\times\mathcal{T}}$ are defined for any function $d \in [0,1]^{\mathcal{T}\times\mathcal{T}}$ and terms $t,t' \in \mathcal{T}$ as:*

$$\begin{aligned}
\mathbf{B}(d)(t,t') \quad &= \max\{\, d(t,t'),\\
&\qquad \sup_{\alpha\in\mathbf{A}\setminus\{\mathsf{tick}\}} \max_{t\xrightarrow{\alpha}\Delta} \inf_{t'\stackrel{\hat{\alpha}}{\Longrightarrow}\Theta} \mathbf{K}(d)\big(\Delta, \Theta + (1- |\Theta|)\overline{\mathsf{Dead}}\big),\\
&\qquad \sup_{\alpha\in\mathbf{A}\setminus\{\mathsf{tick}\}} \max_{t'\xrightarrow{\alpha}\Theta} \inf_{t\stackrel{\hat{\alpha}}{\Longrightarrow}\Delta} \mathbf{K}(d)\big(\Delta + (1- |\Delta|)\overline{\mathsf{Dead}}, \Theta\big)\ \}\\
\mathbf{B}_{\mathsf{tick}}(d)(t,t') &= \max\{\, d(t,t'),\\
&\qquad \max_{t\xrightarrow{\mathsf{tick}}\Delta} \inf_{t'\stackrel{\widehat{\mathsf{tick}}}{\Longrightarrow}\Theta} \mathbf{K}(d)\big(\Delta, \Theta + (1- |\Theta|)\overline{\mathsf{Dead}}\big),\\
&\qquad \max_{t'\xrightarrow{\mathsf{tick}}\Theta} \inf_{t\stackrel{\widehat{\mathsf{tick}}}{\Longrightarrow}\Delta} \mathbf{K}(d)\big(\Delta + (1- |\Delta|)\overline{\mathsf{Dead}}, \Theta\big)\ \}
\end{aligned}$$

where $\inf \emptyset = 1$ *and* $\max \emptyset = 0$.

Notice that all max in Definition 5 are well defined since the pLTS is image-finite. Notice also that any strong transitions from t to a distribution Δ is mimicked by a weak transition from t', which, in general, takes to a sub-distribution Θ. Thus, process t' may not simulate t with probability $1- |\Theta|$.

Definition 6 (Timed weak bisimilarity metrics). *The family of the* timed weak bisimilarity metrics $\mathbf{m}^k \colon (\mathcal{T} \times \mathcal{T}) \to [0,1]$ *is defined for all* $k \in \mathbb{N}$ *by*

$$\mathbf{m}^k = \begin{cases} \mathbf{0} & \text{if } k = 0 \\ \sup_{h \in \mathbb{N}} \mathbf{m}^{k,h} & \text{if } k > 0 \end{cases}$$

while the functions $\mathbf{m}^{k,h} \colon (\mathcal{T} \times \mathcal{T}) \to [0,1]$ *are defined for all* $k \in \mathbb{N}^+$ *and* $h \in \mathbb{N}$ *by*

$$\mathbf{m}^{k,h} = \begin{cases} \mathbf{B}_{\mathsf{tick}}(\mathbf{m}^{k-1}) & \text{if } h = 0 \\ \mathbf{B}(\mathbf{m}^{k,h-1}) & \text{if } h > 0. \end{cases}$$

Then, we define $\mathbf{m}^\infty \colon (\mathcal{T} \times \mathcal{T}) \to [0,1]$ *as* $\mathbf{m}^\infty = \sup_{k \in \mathbb{N}} \mathbf{m}^k$.

Note that any $\mathbf{m}^{k,h}$ is obtained from $\mathbf{m}^{k-1}$ by one application of the functional $\mathbf{B}_{\mathsf{tick}}$, in order to take into account the distance between terms introduced by the k-th tick-action, and h applications of the functional $\mathbf{B}$, in order to lift such a distance to terms that take h untimed actions to be able to perform a tick-action. By taking $\sup_{h \in \mathbb{N}} \mathbf{m}^{k,h}$ we consider an arbitrary number of untimed steps.

The pseudometric property of $\mathbf{m}^k$ is necessary to conclude that the tolerance between terms as given by $\mathbf{m}^k$ is a reasonable notion of behavioural distance.

Theorem 1. *For any* $k \geq 1$, $\mathbf{m}^k$ *is a 1-bounded pseudometric.*

Finally, everything is in place to define our timed weak bisimilarity $\approx_p^k$ with tolerance $p \in [0,1]$ accumulated after k time units, for $k \in \mathbb{N} \cup \{\infty\}$.

Definition 7 (Timed weak bisimilarity with tolerance). *Let* $t, t' \in \mathcal{T}$, $k \in \mathbb{N}$ *and* $p \in [0,1]$. *We say that* t and t' are weakly bisimilar with a tolerance p, which accumulates in k timed actions, *written* $t \approx_p^k t'$, *if and only if* $\mathbf{m}^k(t,t') = p$. *Then, we write* $t \approx_p^\infty t'$ *if and only if* $\mathbf{m}^\infty(t,t') = p$.

Since the Kantorovich lifting $\mathbf{K}$ is monotone [26], it follows that both functionals $\mathbf{B}$ and $\mathbf{B}_{\mathsf{tick}}$ are monotone. This implies that, for any $k \geq 1$, $(\mathbf{m}^{k,h})_{h \geq 0}$ is a non-decreasing chain and, analogously, also $(\mathbf{m}^k)_{k \geq 0}$ is a non-decreasing chain, thus giving the following expected result saying that the distance between terms grows when we consider a higher number of tick computation steps.

Proposition 1 (Tolerance monotonicity). *For all terms* $t, t' \in \mathcal{T}$ *and* $k_1, k_2 \in \mathbb{N}^+$ *with* $k_1 < k_2$, $t \approx_{p_1}^{k_1} t'$ *and* $t \approx_{p_2}^{k_2} t'$ *entail* $p_1 \leq p_2$.

We conclude this section by comparing our behavioural distance with the behavioural relations known in the literature.

We recall that in [5] a family of relations $\simeq_p$ for *untimed* process calculi are defined such that $t \simeq_p t'$ if and only if t and t' weakly bisimulate each other with tolerance p. Of course, one can apply these relations also to timed process calculi, the effect being that timed actions are treated in exactly the same manner as untimed actions. The following result compares the behavioural metrics proposed in the present paper with those of [5], and with the classical notions of probabilistic weak bisimilarity [27] denoted $\approx$.

Proposition 2. *Let* $t, t' \in \mathcal{T}$ *and* $p \in [0,1]$. *Then,*

- $t \approx_p^\infty t'$ *iff* $t \simeq_p t'$
- $t \approx_0^\infty t'$ *iff* $t \approx t'$.

3 A Simple Probabilistic Timed Calculus for IoT Systems

In this section, we propose a simple extension of Hennessy and Regan's *timed process algebra* TPL [12] to express *IoT systems* and *cyber-physical attacks*. The goal is to show that timed weak bisimilarity with tolerance is a suitable notion to estimate the impact of cyber-physical attacks on IoT systems.

Let us start with some preliminary notations.

Notation 1 *We use x, x_k for* state variables*, c, c_k, for* communication channels*, z, z_k for* communication variables*, s, s_k for* sensors devices*, while o ranges over both channels and sensors.* Values*, ranged over by v, v', belong to a* finite *set of admissible values $\mathcal{V}$. We use u, u_k for both values and communication variables. Given a generic set of names $\mathcal{N}$, we write $\mathcal{V}^{\mathcal{N}}$ to denote the set of functions $\mathcal{N} \to \mathcal{V}$ assigning a value to each name in $\mathcal{N}$. For $m \in \mathbb{N}$ and $n \in \mathbb{N} \cup \{\infty\}$, we write $m..n$ to denote an* integer interval*. As we will adopt a discrete notion of time, we will use integer intervals to denote* time intervals*.*

State variables are associated to physical properties like *temperature*, *pressure*, etc. *Sensor names* are metavariables for sensor devices, such as *thermometers* and *barometers*. Please, notice that in cyber-physical systems, state variables cannot be directly accessed but they can only be tested via one or more sensors.

Definition 8 (IoT system). *Let $\mathcal{X}$ be a set of state variables and $\mathcal{S}$ be a set of sensors. Let $range : \mathcal{X} \to 2^{\mathcal{V}}$ be a total function returning the range of admissible values for any state variable $x \in \mathcal{X}$. An* IoT system *consists of two components:*

- *a* physical environment $\xi = \langle \xi_{\mathrm{x}}, \xi_{\mathrm{m}} \rangle$ *where:*
  - $\xi_{\mathrm{x}} \in \mathcal{V}^{\mathcal{X}}$ *is the* physical state *of the system that associates a value to each state variable in $\mathcal{X}$, such that $\xi_{\mathrm{x}}(x) \in range(x)$ for any $x \in \mathcal{X}$,*
  - $\xi_{\mathrm{m}} : \mathcal{V}^{\mathcal{X}} \to \mathcal{S} \to \mathcal{D}(\mathcal{V})$ *is the* measurement map *that given a physical state returns a function that associates to any sensor in $\mathcal{S}$ a discrete probability distribution over the set of possible sensed values;*
- *a* logical (or cyber) component P *that interacts with the sensors defined in ξ, and can communicate, via channels, with other cyber components.*

We write $\xi \bowtie P$ to denote the resulting IoT system, and use M and N to range over IoT systems.

Let us now formalise the *cyber component* of an IoT system. Basically, we adapt Hennessy and Regan's *timed process algebra TPL* [12].

Definition 9 (Logics). Logical components *of IoT systems are defined by the following grammar:*

$$P, Q \ ::= \ \mathsf{nil} \mid \mathsf{tick}.P \mid P \parallel Q \mid \lfloor pfx.P \rfloor Q \mid H\langle \tilde{u} \rangle \mid \mathsf{if}\ (b)\ \{P\}\ \mathsf{else}\ \{Q\} \mid P \backslash c$$
$$pfx \ ::= \ o!v \mid o?(z)$$

The process $\mathsf{tick}.P$ sleeps for one time unit and then continues as P. We write $P \parallel Q$ to denote the *parallel composition* of concurrent processes P and Q. The

process $\lfloor pfx.P \rfloor Q$ denotes *prefixing with timeout*. We recall that o ranges over both channel and sensor names. Thus, for instance, $\lfloor c!v.P \rfloor Q$ sends the value v on channel c and, after that, it continues as P; otherwise, if no communication partner is available within one time unit, it evolves into Q. The process $\lfloor c?(z).P \rfloor Q$ is the obvious counterpart for channel reception. On the other hand, the process $\lfloor s?(z).P \rfloor Q$ reads the sensor s, according to the measurement map of the systems, and, after that, it continues as P. The process $\lfloor s!v.P \rfloor Q$ writes to the sensor s and, after that, it continues as P; here, we wish to point out that this a *malicious activity*, as controllers may only access sensors for reading sensed data. Thus, the construct $\lfloor s!v.P \rfloor Q$ serves to implement an *integrity attack* that attempts at synchronising with the controller of sensor s to provide a fake value v. In the following, we say that a process is *honest* if it never writes on sensors. The definition of honesty naturally lifts to IoT systems. In processes of the form $\mathsf{tick}.Q$ and $\lfloor pfx.P \rfloor Q$, the occurrence of Q is said to be *time-guarded*. *Recursive processes* $H\langle \tilde{u} \rangle$ are defined via equations $H(z_1, \ldots, z_k) = P$, where (i) the tuple $z_1, \ldots, z_k$ contains all the variables that appear free in P, and (ii) P contains *only time-guarded occurrences* of the process identifiers, such as H itself (to avoid *zeno behaviours*). The two remaining constructs are standard; they model conditionals and channel restriction, respectively.

Finally, we define how to compose IoT systems. For simplicity, we compose two systems only if they have the same physical environment.

Definition 10 (System composition). *Let $M_1 = \xi \bowtie P_1$ and $M_2 = \xi \bowtie P_2$ be two IoT systems, and Q be a process whose sensors are defined in the physical environment ξ. We write:*

- *$M_1 \parallel M_2$ to denote $\xi \bowtie (P_1 \parallel P_2)$;*
- *$M_1 \parallel Q$ to denote $\xi \bowtie (P_1 \parallel Q)$;*
- *$M_1 \backslash c$ as an abbreviation for $\xi \bowtie (P_1 \backslash c)$.*

We conclude this section with the following abbreviations that will be used in the rest of the paper.

Notation 2 *We write $P \backslash \{c_1, c_2, \ldots, c_n\}$, or $P \backslash \tilde{c}$, to mean $P \backslash c_1 \backslash c_2 \cdots \backslash c_n$. For simplicity, we sometimes abbreviate both $H(i)$ and $H\langle i \rangle$ with H_i. We write $pfx.P$ as an abbreviation for the process defined via the equation $H = \lfloor pfx.P \rfloor H$, where the process name H does not occur in P. We write $\mathsf{tick}^k.P$ as a shorthand for $\mathsf{tick}.\mathsf{tick}.\ldots.\mathsf{tick}.P$, where the prefix tick appears $k \geq 0$ consecutive times. We write Dead to denote a deadlocked IoT system that cannot perform any action.*

3.1 Probabilistic Labelled Transition Semantics

As said before, sensors serve to observe the evolution of the physical state of an IoT system. However, sensors are usually affected by an *error/noise* that we represent in our measurement maps by means of discrete probability distributions. For this reason, we equip our calculus with a probabilistic labelled transition system. In the following, the symbol ϵ ranges over distributions on physical environments,

$$\text{(Write)}\ \frac{-}{\lfloor o!v.P \rfloor Q \xrightarrow{o!v} P} \qquad \text{(Read)}\ \frac{-}{\lfloor o?(z).P \rfloor Q \xrightarrow{o?(z)} P}$$

$$\text{(Sync)}\ \frac{P \xrightarrow{o!v} P' \quad Q \xrightarrow{o?(z)} Q'}{P \parallel Q \xrightarrow{\tau} P' \parallel Q'\{^{v}/_{z}\}} \qquad \text{(Par)}\ \frac{P \xrightarrow{\lambda} P' \quad \lambda \neq \mathsf{tick}}{P \parallel Q \xrightarrow{\lambda} P' \parallel Q}$$

$$\text{(Res)}\ \frac{P \xrightarrow{\lambda} P' \quad \lambda \notin \{o!v, o?(z)\}}{P \backslash o \xrightarrow{\lambda} P' \backslash o} \qquad \text{(Rec)}\ \frac{P\{^{\tilde{v}}/_{\tilde{z}}\} \xrightarrow{\lambda} Q \quad H(\tilde{z}) = P}{H\langle \tilde{v} \rangle \xrightarrow{\lambda} Q}$$

$$\text{(Then)}\ \frac{[\![b]\!] = \mathsf{true} \quad P \xrightarrow{\lambda} P'}{\mathsf{if}\ (b)\ \{P\}\ \mathsf{else}\ \{Q\} \xrightarrow{\lambda} P'} \qquad \text{(Else)}\ \frac{[\![b]\!] = \mathsf{false} \quad Q \xrightarrow{\lambda} Q'}{\mathsf{if}\ (b)\ \{P\}\ \mathsf{else}\ \{Q\} \xrightarrow{\lambda} Q'}$$

$$\text{(TimeNil)}\ \frac{-}{\mathsf{nil} \xrightarrow{\mathsf{tick}} \mathsf{nil}} \qquad \text{(Delay)}\ \frac{-}{\mathsf{tick}.P \xrightarrow{\mathsf{tick}} P}$$

$$\text{(Timeout)}\ \frac{-}{\lfloor pfx.P \rfloor Q \xrightarrow{\mathsf{tick}} Q} \qquad \text{(TimePar)}\ \frac{P \xrightarrow{\mathsf{tick}} P' \quad Q \xrightarrow{\mathsf{tick}} Q'}{P \parallel Q \xrightarrow{\mathsf{tick}} P' \parallel Q'}$$

Table 1. Labelled transition system for processes

whereas π ranges over distributions on (logical) processes. Thus, $\epsilon \bowtie \pi$ denotes the distribution over IoT systems defined by $(\epsilon \bowtie \pi)(\xi \bowtie P) = \epsilon(\xi) \cdot \pi(P)$. The symbol γ ranges over distributions on IoT systems.

In Table 1, we give a standard labelled transition system for logical components (timed processes), whereas in Table 2 we rely on the LTS of Table 1 to define a simple pLTS for IoT systems by lifting transition rules from processes to systems.

In Table 1, the meta-variable λ ranges over labels in the set $\{\tau, \mathsf{tick}, o!v, o?(z)\}$. Rule (Sync) serve to model synchronisation and value passing, on some name (for channel or sensor) o: if o is a channel then we have standard point-to-point communication, whereas if o is a sensor then this rule models an *integrity attack* on sensor s, as the controller is provided with a fake value v. The remaining rules are standard. The symmetric counterparts of rules (Sync) and (Par) are omitted.

According to Table 2, IoT systems may fire four possible actions ranged over by α. These actions represent: internal activities (τ), the passage of time (tick), channel transmission ($c!v$) and channel reception ($c?v$).

Rules (Snd) and (Rcv) model transmission and reception on a channel c with an external system, respectively. Rule (SensRead) models the reading of the value detected at a *sensor* s according to the current physical environment $\xi = \langle \xi_{\mathrm{x}}, \xi_{\mathrm{m}} \rangle$. In particular, this rule says that if a process P in a system $\xi \bowtie P$ reads a sensor s defined in ξ then it will get a value that may vary according to the probability distribution resulting by providing the state function ξ_{x} and the sensor s to the measurement map ξ_{m}.

Rule (Tau) lifts internal actions from processes to systems. This includes communications on channels and malicious accesses to sensors' controllers. According

$$\text{(Snd)}\ \frac{P \xrightarrow{c!v} P'}{\xi \bowtie P \xrightarrow{c!v} \overline{\xi} \bowtie \overline{P'}} \qquad \text{(Rcv)}\ \frac{P \xrightarrow{c?(z)} P'}{\xi \bowtie P \xrightarrow{c?v} \overline{\xi} \bowtie \overline{P'\{^{v}/_{z}\}}}$$

$$\text{(SensRead)}\ \frac{P \xrightarrow{s?(z)} P' \quad \xi_{\mathrm{m}}(\xi_{\mathrm{x}})(s) = \sum_{i \in I} p_i \cdot \overline{v_i}}{\xi \bowtie P \xrightarrow{\tau} \overline{\xi} \bowtie \sum_{i \in I} p_i \cdot \overline{P'\{^{v_i}/_{z}\}}}$$

$$\text{(Tau)}\ \frac{P \xrightarrow{\tau} P'}{\xi \bowtie P \xrightarrow{\tau} \overline{\xi} \bowtie \overline{P'}} \qquad \text{(Time)}\ \frac{P \xrightarrow{\mathsf{tick}} P' \quad \xi \bowtie P \xrightarrow{\tau} \not\quad \xi' \in \mathit{next}(\xi)}{\xi \bowtie P \xrightarrow{\mathsf{tick}} \overline{\xi'} \bowtie \overline{P'}}$$

Table 2. Probabilistic LTS for a IoT system $\xi \bowtie P$ with $\xi = \langle \xi_{\mathrm{x}}, \xi_{\mathrm{m}} \rangle$

to Definition 10, rule (Tau) models also channel communication between two parallel IoT systems sharing the same physical environment.

A second lifting occurs in rule (Time) for timed actions tick. Here, ξ' denotes an admissible physical environment for the next time slot, nondeterministically chosen from the *finite* set $\mathit{next}(\langle \xi_{\mathrm{x}}, \xi_{\mathrm{m}} \rangle)$. This set is defined as $\{\langle \xi'_{\mathrm{x}}, \xi_{\mathrm{m}} \rangle : \xi'_{\mathrm{x}}(x) \in \mathit{range}(x)$ for any $x \in \mathcal{X}\}$.[3] As a consequence, the rules in Table 2 define an *image-finite* pLTS.

For simplicity, we abstract from the *physical process* behind our IoT systems.

4 Cyber-physical Attacks on Sensor Devices

In this section, we consider attacks tampering with sensors by eavesdropping and possibly modifying the sensor measurements provided to the corresponding controllers. These attacks may affect both the *integrity* and the *availability* of the system under attack. We do not represent (well-known) attacks on communication channels as our focus is on attacks to physical devices and the consequent impact on the physical state. However, our technique can be easily generalised to deal with attacks on channels as well.

Definition 11 (Cyber-physical attack). *A (pure) cyber-physical attack A is a process derivable from the grammar of Definition 9 such that:*

- *A writes on at least one sensor;*
- *A never uses communication channels.*

In order to make security assessments on our IoT systems, we adapt a well-known approach called *Generalized Non Deducibility on Composition (GNDC)* [7]. Intuitively, an attack A affects an honest IoT system M if the execution of the composed system $M \parallel A$ differs from that of the original system M in an observable manner. Basically, a cyber-physical attack can influence the system under attack in at least two different ways:

[3] The finiteness follows from the finiteness of $\mathcal{V}$, and hence of $\mathit{range}(x)$, for any $x \in \mathcal{X}$.

- The system $M \parallel A$ might have non-genuine execution traces containing observables that cannot be reproduced by M; here the attack affects the *integrity* of the system behaviour (*integrity attack*).
- The system M might have execution traces containing observables that cannot be reproduced by the system under attack $M \parallel A$ (because they are prevented by the attack); this is an attack against the *availability* of the system (*DoS attack*).

Now, everything is in place to provide a formal definition of *system tolerance* and *system vulnerability* with respect to a given attack. Intuitively, a system M tolerates an attack A if the presence of the attack does not affect the behaviour of M; on the other hand M is vulnerable to A in a certain time interval if the attack has an *impact* on the behaviour of M in that time interval.

Definition 12 (Attack tolerance). *Let M be a honest IoT system. We say that M* tolerates an attack *A if $M \parallel A \approx_0^\infty M$.*

Definition 13 (Attack vulnerability and impact). *Let M be a honest IoT system. We say that M is* vulnerable to an attack *A* in the time interval *$m..n$* with *impact $p \in [0,1]$, for $m \in \mathbb{N}^+$ and $n \in \mathbb{N}^+ \cup \{\infty\}$, if $m..n$ is the smallest time interval such that: (i) $M \parallel A \approx_0^{m-1} M$, (ii) $M \parallel A \approx_p^n M$, (iii) $M \parallel A \approx_p^\infty M$.*[4]

Basically, the definition above says that if a system is vulnerable to an attack in the time interval $m..n$ then the perturbation introduced by the attack starts in the m-th time slot and reaches the maximum impact in the n-th time slot.

The following result says that both notions of tolerance and vulnerability are suitable for *compositional reasonings*. More precisely, we prove that they are both preserved by parallel composition and channel restriction. Actually, channel restriction may obviously make a system less vulnerable by hiding channels.

Theorem 2 (Compositionality). *Let $M_1 = \xi \bowtie P_1$ and $M_2 = \xi \bowtie P_2$ be two honest IoT systems with the same physical environment ξ, A an arbitrary attack, and $\tilde{c}$ a set of channels.*

- *If both M_1 and M_2 tolerate A then $(M_1 \parallel M_2)\backslash\tilde{c}$ tolerates A.*
- *If M_1 is vulnerable to A in the time interval $m_1..n_1$ with impact p_1, and M_2 is vulnerable to A in the time interval $m_2..n_2$ with impact p_2, then $M_1 \parallel M_2$ is vulnerable to A in a the time interval $\min(m_1, m_2)..\max(n_1, n_2)$ with an impact $p' \leq (p_1 + p_2 - p_1 p_2)$.*
- *If M_1 is vulnerable to A in the interval $m_1..n_1$ with impact p_1 then $M_1\backslash\tilde{c}$ is vulnerable to A in a time interval $m'..n' \subseteq m_1..n_1$ with an impact $p' \leq p_1$.*

Note that if an attack A is tolerated by a system M and can interact with a honest process P then the compound system $M \parallel P$ may be vulnerable to A. However, if A does not write on the sensors of P then it is tolerated by $M \parallel P$ as well. The bound $p' \leq (p_1 + p_2 - p_1 p_2)$ can be explained as follows. The likelihood that

[4] By Proposition 1, at all time instants greater than n the impact remains p.

the attack does not impact on M_i is $(1-p_i)$, for $i \in \{1,2\}$. Thus, the likelihood that the attack impacts neither on M_1 nor on M_2 is at least $(1-p_1)(1-p_2)$. Summarising, the likelihood that the attack impacts on at least one of the two systems M_1 and M_2 is at most $1-(1-p_1)(1-p_2) = p_1+p_2-p_1p_2$.

An easy corollary of Theorem 2 allows us to lift the notions of tolerance and vulnerability from a honest system M to the compound systems $M \parallel P$, for a honest process P.

Corollary 1. *Let M be a honest system, A an attack, $\tilde{c}$ a set of channels, and P a honest process that reads sensors defined in M but not those written by A.*

- *If M tolerates A then $(M \parallel P)\backslash\tilde{c}$ tolerates A.*
- *If M is vulnerable to A in the interval $m..n$ with impact p, then $(M \parallel P)\backslash\tilde{c}$ is vulnerable to A in a time interval $m'..n' \subseteq m..n$, with an impact $p' \leq p$.*

5 Attacking a Smart Surveillance System: A Case Study

Consider an alarmed ambient consisting of three rooms, r_i for $i \in \{1,2,3\}$, each of which equipped with a sensor s_i to detect unauthorised accesses. The alarm goes off if at least one of the three sensors detects an intrusion.

The logics of the system can be easily specified in our language as follows:

$$
\begin{aligned}
\mathit{Sys} &= (\mathit{Mng} \parallel \mathit{Ctrl}_1 \parallel \mathit{Ctrl}_2 \parallel \mathit{Ctrl}_3)\backslash\{c_1,c_2,c_3\}\\
\mathit{Mng} &= c_1?(z_1).c_2?(z_2).c_3?(z_3).\mathsf{if}\ (\textstyle\bigvee_{i=1}^{3} z_i{=}\mathsf{on})\ \{\mathit{alarm}!\mathsf{on}.\mathsf{tick}.\mathit{Check}_k\}\ \mathsf{else}\ \{\mathsf{tick}.\mathit{Mng}\}\\
\mathit{Check}_0 &= \mathit{Mng}\\
\mathit{Check}_j &= \mathit{alarm}!\mathsf{on}.c_1?(z_1).c_2?(z_2).c_3?(z_3).\mathsf{if}\ (\textstyle\bigvee_{i=1}^{3} z_i = \mathsf{on})\ \{\mathsf{tick}.\mathit{Check}_k\}\\
&\quad\ \mathsf{else}\ \{\mathsf{tick}.\mathit{Check}_{j-1}\} \quad \text{for } j>0\\
\mathit{Ctrl}_i &= s_i?(z_i).\mathsf{if}\ (z_i{=}\mathsf{presence})\ \{c_i!\mathsf{on}.\mathsf{tick}.\mathit{Ctrl}_i\}\ \mathsf{else}\ \{c_i!\mathsf{off}.\mathsf{tick}.\mathit{Ctrl}_i\}\ \text{for } i{\in}\{1,2,3\}.
\end{aligned}
$$

Intuitively, the process *Sys* is composed by three controllers, Ctrl_i, one for each sensor s_i, and a manager *Mng* that interacts with the controllers via private channels c_i. The process *Mng* fires an alarm if at least one of the controllers signals an intrusion. As usual in this kind of surveillance systems, the alarm will keep going off for k instants of time after the last detected intrusion.

As regards the physical environment, the physical state $\xi_{\mathrm{x}} : \{r_1,r_2,r_3\} \to \{\mathsf{presence},\mathsf{absence}\}$ is set to $\xi_{\mathrm{x}}(r_i) = \mathsf{absence}$, for any $i \in \{1,2,3\}$. Furthermore, let p_i^+ and p_i^- be the probabilities of having *false positives* (erroneously detected intrusion) and *false negatives* (erroneously missed intrusion) at sensor s_i[5], respectively, for $i \in \{1,2,3\}$, the measurement function ξ_{m} is defined as follows: $\xi_{\mathrm{m}}(\xi_{\mathrm{x}})(s_i) = (1{-}p_i^-)\,\overline{\mathsf{presence}} + p_i^-\,\overline{\mathsf{absence}}$, if $\xi_{\mathrm{x}}(r_i) = \mathsf{presence}$; $\xi_{\mathrm{m}}(\xi_{\mathrm{x}})(s_i) = (1{-}p_i^+)\,\mathsf{absence} + p_i^+\,\overline{\mathsf{presence}}$, otherwise.

Thus, the whole IoT system has the form $\xi \bowtie \mathit{Sys}$, with $\xi = \langle \xi_{\mathrm{x}}, \xi_{\mathrm{m}} \rangle$.

We start our analysis studying the impact of a simple cyber-physical attack that provides fake *false positives* to the controller of one of the sensors s_i. This attack affects the *integrity* of the system behaviour as the system under attack will fire alarms without any physical intrusion.

[5] These probabilities are usually very small; we assume them smaller than $\frac{1}{2}$.

Example 1 (Introducing false positives). In this example, we provide an attack that tries to increase the number of false positives detected by the controller of some sensor s_i during a specific time interval $m..n$, with $m, n \in \mathbb{N}$, $n \geq m > 0$. Intuitively, the attack waits for $m-1$ time slots, then, during the time interval $m..n$, it provides the controller of sensor s_i with a fake intrusion signal. Formally,

$$A_{\mathsf{fp}}(i,m,n) = \mathsf{tick}^{m-1}.B\langle i, n-m+1\rangle$$
$$B(i,j) = \mathsf{if}\ (j=0)\ \{\mathsf{nil}\}\ \mathsf{else}\ \{\lfloor s_i!\mathsf{presence.tick}.B\langle i,j-1\rangle \rfloor B\langle i,j-1\rangle\}\,.$$

In the following proposition, we use our metric to measure the perturbation introduced by the attack to the controller of a sensor s_i by varying the time of observation of the system under attack.

Proposition 3. *Let ξ be an arbitrary physical state for the systems $M_i = \xi \bowtie Ctrl_i$, for $i \in \{1,2,3\}$. Then,*

- $M_i \parallel A_{\mathsf{fp}}\langle i,m,n\rangle \approx_0^j M_i$, *for* $j \in 1..m{-}1$;
- $M_i \parallel A_{\mathsf{fp}}\langle i,m,n\rangle \approx_h^j M_i$, *with* $h = 1-(p_i^+)^{j-m+1}$, *for* $j \in m..n$;
- $M_i \parallel A_{\mathsf{fp}}\langle i,m,n\rangle \approx_r^j M_i$, *with* $r = 1-(p_i^+)^{n-m+1}$, *for* $j > n$ *or* $j = \infty$.

By an application of Definition 13 we can measure the impact of the attack A_{fp} to the (sub)systems $\xi \bowtie Ctrl_i$.

Corollary 2. *The IoT systems $\xi \bowtie Ctrl_i$ are vulnerable to the attack $A_{\mathsf{fp}}\langle i,m,n\rangle$ in the time interval $m..n$ with impact $1-(p_i^+)^{n-m+1}$.*

Note that the vulnerability window $m..n$ coincides with the activity period of the attack A_{fp}. This means that the system under attack recovers its normal behaviour immediately after the termination of the attack. However, in general, an attack may impact the behaviour of the target system long after its termination.

Note also that the attack $A_{\mathsf{fp}}\langle i,m,n\rangle$ has an impact not only on the controller $Ctrl_i$ but also on the whole system $\xi \bowtie Sys$. This because the process *Mng* will surely fire the alarm as it will receive at least one intrusion detection from $Ctrl_i$. However, by an application of Corollary 1 we can prove that the impact on the whole system will not get amplified.

Proposition 4 (Impact of the attack A_{fp}). *The system $\xi \bowtie Sys$ is vulnerable to the attack $A_{\mathsf{fp}}\langle i,m,n\rangle$ in a time interval $m'..n' \subseteq m..n$ with impact $p' \leq 1-(p_i^+)^{n-m+1}$.*

Now, the reader may wonder what happens if we consider a complementary attack that provides fake *false negatives* to the controller of one of the sensors s_i. In this case, the attack affects the *availability* of the system behaviour as the system will no fire the alarm in the presence of a real intrusion. This because a real intrusion will be somehow "hidden" by the attack.

Example 2 (Introducing false negatives). The goal of the following attack is to increase the number of false negatives during the time interval $m..n$, with $n \geq m > 0$. Formally, the attack is defined as follows:

$$A_{\mathsf{fn}}(i,m,n) = \mathsf{tick}^{m-1}.C\langle i, n-m+1\rangle$$
$$C(i,j) = \mathsf{if}\ (j=0)\ \{\mathsf{nil}\}\ \mathsf{else}\ \{\lfloor s_i!\mathsf{absence.tick}.C\langle i,j-1\rangle \rfloor C\langle i,j-1\rangle\}\,.$$

In the following proposition, we use our metric to measure the deviation introduced by the attack A_{fn} to the controller of a sensor s_i. With no surprise we get a result that is the symmetric version of Proposition 3.

Proposition 5. *Let ξ be an arbitrary physical state for the system $M_i = \xi \bowtie Ctrl_i$, for $i \in \{1,2,3\}$. Then,*

- $M_i \parallel A_{\mathsf{fn}}\langle i,m,n\rangle \approx_0^j M_i$, *for* $j \in 1..m{-}1$;
- $M_i \parallel A_{\mathsf{fn}}\langle i,m,n\rangle \approx_h^j M_i$, *with* $h = 1-(p_i^-)^{j-m+1}$, *for* $j \in m..n$;
- $M_i \parallel A_{\mathsf{fn}}\langle i,m,n\rangle \approx_r^j M_i$, *with* $r = 1-(p_i^-)^{n-m+1}$, *for* $j > n$ *or* $j = \infty$.

Again, by an application of Definition 13 we can measure the impact of the attack A_{fn} to the (sub)systems $\xi \bowtie Ctrl_i$.

Corollary 3. *The IoT systems $\xi \bowtie Ctrl_i$ are vulnerable to the attack $A_{\mathsf{fn}}\langle i,m,n\rangle$ in the time interval $m..n$ with impact $1-(p_i^-)^{n-m+1}$.*

As our timed metric is compositional, by an application of Corollary 1 we can estimate the impact of the attack A_{fn} to the whole system $\xi \bowtie Sys$.

Proposition 6 (Impact of the attack A_{fn}). *The system $\xi \bowtie Sys$ is vulnerable to the attack $A_{\mathsf{fn}}\langle i,m,n\rangle$ in a time interval $m'..n' \subseteq m..n$ with impact $p' \leq 1-(p_i^-)^{n-m+1}$.*

6 Conclusions, Related and Future Work

We have proposed a timed generalisation of the n-bisimulation metric [3], called *timed bisimulation metric*, obtained by defining two functionals over the complete lattice of the functions assigning a distance in $[0,1]$ to each pair of systems: the former deals with the distance accumulated when executing untimed steps, the latter with the distance introduced by timed actions.

We have used our timed bisimulation metrics to provide a formal and *compositional* notion of *impact metric* for *cyber-physical attacks* on *IoT systems* specified in a simple timed process calculus. In particular, we have focussed on cyber-physical attacks targeting sensor devices (attack on sensors are by far the most studied cyber-physical attacks [38]). We have used our timed weak bisimulation with tolerance to formalise the notions of *attack tolerance* and *attack vulnerability with a given impact p*. In particular, a system M is said to be vulnerable to an attack A in the time interval $m..n$ with impact p if the perturbation introduced by A becomes observable in the m-th time slot and yields the maximum impact p in the n-th time slot. Here, we wish to stress that the *vulnerability window* $m..n$ is quite informative. In practise, this interval says when an attack will produce observable effects on the system under attack. Thus, if n is finite we have an attack with *temporary effects*, otherwise we have an attack with *permanent effects*. Furthermore, if the attack is quick enough, and terminates well before the time instant m, then we have a *stealthy attack* that affects the system late enough to allow *attack camouflages* [11]. On the other

hand, if at time m the attack is far from termination, then the IoT system under attack has good chances of undertaking countermeasures to stop the attack.

As a case study, we have estimated the impact of two cyber-physical attacks on sensors that introduce *false positives* and *false negatives*, respectively, into a simple surveillance system, affecting the *integrity* and the *availability* of the IoT system. Although our attacks are quite simple, the specification language and the corresponding metric semantics presented in the paper allow us to deal with smarter attacks, such as *periodic attacks* with constant or variable period of attack. Moreover, we can easily extend our threat model to recover (well-known) attacks on communication channels.

Related Work. A number of papers have recently proposed different methodologies for assessing the direct and indirect impact of attacks on CPSs.

Bilis et al. [1] proposed a systematic approach that uses five metrics derived from complex network theory to assess the impacts of cyber attacks on electric power systems. The metrics were used to rank nodes in a graph-based representation of an electric grid. Sgouras et al. [31] evaluated the impact of cyber attacks on a simulated smart metering infrastructure; the denial-of-service attacks against smart meters and utility servers caused severe communications interruptions. Sridhar and Govindarasu [33] evaluated the impacts of attacks on wide-area frequency control applications in power systems; their research showed that cyber attacks can significantly impact system stability by causing severe drops in system frequency. Genge et al. [10] introduced a methodology, inspired by research in system dynamics and sensitivity analysis, to compute the covariances of the observed variables before and after the execution of a specific intervention involving the control variables. Metrics are proposed for quantifying the significance of control variables and measuring the impact propagation of cyber attacks. Orojloo and Azgomi [25] investigated how an attack against system parameters can affect the values of other parameters. The system parameters are divided into two classes of cause and effect parameters, which can be same as or different from each other. They proposed metrics to prioritise the sensor readings and control signals based on their sensitivity to conducted attacks. Urbina et al. [35] defined an evaluation metric for attack-detection algorithms that quantifies the negative impact of stealthy attacks and the inherent trade-off with false alarms. The authors showed that the impact of such attacks can be mitigated in several cases by the proper combination and configuration of detection schemes. Huang et al. [13] proposed a risk assessment method that uses a Bayesian network to model the attack propagation process and infers the probabilities of sensors and actuators to be compromised. These probabilities are fed into a stochastic hybrid system (SHS) model to predict the evolution of the physical process being controlled. Then, the security risk is quantified by evaluating the system availability with the SHS model.

Notice that only this last paper adopts formal methodologies. More generally, we are aware of a number of works using formal methods for CPS security,

although they apply methods, and most of the time have goals, that are quite different from ours.

Vigo et al. [36] proposed an untimed calculus of broadcasting processes equipped with notions of failed and unwanted communication. They focus on DoS attacks without taking into consideration timing aspects or attack impact. Bodei et al. [2] proposed a different untimed process calculus, IoT-LySa, supporting a control flow analysis that safely approximates the abstract behaviour of IoT systems. Essentially, they track how data spread from sensors to the logics of the network, and how physical data are manipulated. Rocchetto and Tippenhaur [29] introduced a taxonomy of the diverse attacker models proposed for CPS security and outline requirements for generalised attacker models; in [28], the same authors proposed an extended Dolev-Yao attacker model suitable for CPSs. Nigam et al. [24] worked around the notion of timed Dolev-Yao intruder models for cyber-physical security protocols by bounding the number of intruders required for the automated verification of such protocols. Following a tradition in security protocol analysis, they provided an answer to the question: How many intruders are enough for verification and where should they be placed? Lanotte el al. [19] did a static security analysis, based on model-checking, for a non-trivial engineering case study to statically detect a variety of attacks targeting sensors and/or actuators of the system under investigation. Finally, Lanotte et al. [20] defined a hybrid process calculus to model both CPSs and cyber-physical attacks; they defined a threat model for cyber-physical attacks to physical devices and provided a proof methods to assess attack tolerance/vulnerability with respect to a timed trace semantics (no tolerance allowed). They also advocated a timed formalisation of the impact of an attack in terms of the deviation introduced in the runtime behaviour of the system under attack.

Future Work. Recent works [18,8,21,22,9] have shown that bisimulation metrics are suitable for compositional reasoning, as the distance between two complex systems can be often derived in terms of the distance between their components. In this respect, Theorem 2 and Corollary 1 allows us compositional reasonings when computing the impact of attacks on a target system, in terms of the impact on its sub-systems. We believe that this result can be generalised to estimate the impact of parallel attacks of the form $A = A_1 \parallel \ldots \parallel A_k$ in terms of the impacts of each malicious module A_i. As future work, we also intend to adopt our impact metric in more involved languages for *cyber-physical systems and attacks*, such as the language developed in [20], with an explicit representation of physical processes via differential equations or their discrete counterpart, difference equations.

Acknowledgements. We thank the anonymous reviewers for valuable comments.

References

1. Bilis, E.I., Kröger, W., Cen, N.: Performance of Electric Power Systems Under Physical Malicious Attacks. IEEE Systems Journal **7**(4), 854–865 (2013)

2. Bodei, C., Degano, P., Ferrari, G., Galletta, L.: Tracing where IoT data are collected and aggregated. Logical Methods in Computer Science **13(3)**, 1–38 (2017). https://doi.org/10.23638/LMCS-13(3:5)2017
3. van Breugel, F.: On behavioural pseudometrics and closure ordinals. Inf. Process. Lett. **112**(19), 715–718 (2012)
4. Deng, Y., Du, W.: The Kantorovich Metric in Computer Science: A Brief Survey. ENTCS **253**(3), 73–82 (2009)
5. Desharnais, J., Jagadeesan, R., Gupta, V., Panangaden, P.: The metric analogue of weak bisimulation for probabilistic processes. In: LICS 2002. pp. 413–422. IEEE Computer Society (2002). https://doi.org/10.1109/LICS.2002.1029849
6. Falliere, N., Murchu, L., Chien, E.: W32.STUXnet Dossier (2011)
7. Focardi, R., Martinelli, F.: A Uniform Approach for the Definition of Security Properties. In: Wing, J.M., Woodcockm, J., Davies, J. (eds.) FM 1999. LNCS, vol. 1708, pp. 794–813. Springer (1999)
8. Gebler, D., Larsen, K.G., Tini, S.: Compositional Bisimulation Metric Reasoning with Probabilistic Process Calculi. Logical Meth. Comput. Sci. **12**(4), 1–38 (2016)
9. Gebler, D., Tini, S.: Sos specifications for uniformly continuous operators. Journal of Computer and System Sciences **92**, 113–151 (2018)
10. Genge, B., Kiss, I., Haller, P.: A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures. IJCIP **10**, 3–17 (2015)
11. Gollmann, D., Gurikov, P., Isakov, A., Krotofil, M., Larsen, J., Winnicki, A.: Cyber-Physical Systems Security: Experimental Analysis of a Vinyl Acetate Monomer Plant. In: Zhou, J., Jones, D. (eds.) ACM CCPS 2015. pp. 1–12. ACM (2015). https://doi.org/10.1145/2732198.2732208
12. Hennessy, M., Regan, T.: A process algebra for timed systems. Information and Computation **117**(2), 221–239 (1995)
13. Huang, K., Zhou, C., Tian, Y., Yang, S., Qin, Y.: Assessing the Physical Impact of Cyberattacks on Industrial Cyber-Physical Systems. IEEE Trans. Industrial Electronics **65**(10), 8153–8162 (2018)
14. Huang, Y., Cárdenas, A.A., Amin, S., Lin, Z., Tsai, H., Sastry, S.: Understanding the physical and economic consequences of attacks on control systems. IJCIP **2**(3), 73–83 (2009)
15. ICS-CERT: Cyber-Attack Against Ukrainian Critical Infrastructure, https://ics-cert.us-cert.gov/alerts/IR-ALERT-H-16-056-01
16. Keller, R.M.: Formal verification of parallel programs. Communications of the ACM **19**, 371–384 (1976)
17. Krotofil, M., Cárdenas, A.A., Larsen, J., Gollmann, D.: Vulnerabilities of cyber-physical systems to stale data - Determining the optimal time to launch attacks. IJCIP **7**(4), 213–232 (2014)
18. Lanotte, R., Merro, M.: Semantic analysis of gossip protocols for wireless sensor networks. In: Katoen, J.P., König, B. (eds.) CONCUR 2011. LNCS, vol. 6901, pp. 156–170. Springer (2011)
19. Lanotte, R., Merro, M., Munteanu, A.: A Modest Security Analysis of Cyber-Physical Systems: A Case Study. In: Baier, C., Caires, L. (eds.) FORTE 2018. LNCS, vol. 10854, pp. 58–78. Springer (2018). https://doi.org/10.1007/978-3-319-92612-4
20. Lanotte, R., Merro, M., Muradore, R., Viganò, L.: A formal approach to cyber-physical attacks. In: CSF 2017. pp. 436–450. IEEE Computer Society (2017). https://doi.org/10.1109/CSF.2017.12

21. Lanotte, R., Merro, M., Tini, S.: Compositional weak metrics for group key update. In: Larsen, K.G., Bodlaender, H.L., Raskin, J.F. (eds.) MFCS 2017. LIPIcs, vol. 42, pp. 72:1–72:16. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2017). https://doi.org/10.4230/LIPIcs.MFCS.2017.72
22. Lanotte, R., Merro, M., Tini, S.: A Probabilistic Calculus of Cyber-Physical Systems. CoRR **abs/1707.02279** (2017)
23. Lanotte, R., Merro, M., Tini, S.: Towards a formal notion of impact metric for cyber-physical attacks (full version). CoRR **abs/1806.10463** (2018)
24. Nigam, V., Talcott, C., Urquiza, A.A.: Towards the Automated Verification of Cyber-Physical Security Protocols: Bounding the Number of Timed Intruders. In: Askoxylakis, I.G., Ioannidis, S., Katsikas, S.K., Meadows, C.A. (eds.) ESORICS 2016. LNCS, vol. 9879, pp. 450–470. Springer (2016)
25. Orojloo, H., Azgomi, M.: A method for evaluating the consequence propagation of security attacks in cyber-physical systems. Future Generation Comp. Syst. **67**, 57–71 (2017)
26. Panangaden, P.: Labelled Markov Processes. Imperial College Press (2009)
27. Philippou, A., Lee, I., Sokolsky, O.: Weak Bisimulation for Probabilistic Systems. In: Palamidessi, C. (ed.) CONCUR 2000. LNCS, vol. 1877, pp. 334–349 (2000)
28. Rocchetto, M., Tippenhauer, N.O.: CPDY: Extending the Dolev-Yao Attacker with Physical-Layer Interactions. In: Ogata, K., Lawford, M., Liu, S. (eds.) ICFEM 2016. LNCS, vol. 10009, pp. 175–192 (2016). https://doi.org/10.1007/978-3-319-47846-3
29. Rocchetto, M., Tippenhauer, N.O.: On Attacker Models and Profiles for Cyber-Physical Systems. In: Askoxylakis, I.G., Ioannidis, S., Katsikas, S.K., Meadows, C.A. (eds.) ESORICS 2016. LNCS, vol. 9879, pp. 427–449. Springer (2016)
30. Segala, R.: Modeling and Verification of Randomized Distributed Real-Time Systems. Ph.D. thesis, MIT (1995)
31. Sgouras, K.I., Birda, A.I., Labridis, D.L.: Cyber attack impact on critical Smart Grid infrastructures. In: IEEE PES ISGT 2014. pp. 1–5. IEEE (2014). https://doi.org/10.1109/ISGT.2014.6816504
32. Slay, J., Miller, M.: Lessons Learned from the Maroochy Water Breach. In: Goetz, E., Shenoi, S. (eds.) Critical Infrastructure Protection. IFIP, vol. 253, pp. 73–82. Springer (2007)
33. Sridhar, S., Govindarasu, M.: Model-Based Attack Detection and Mitigation for Automatic Generation Control. IEEE Trans. Smart Grid **5**(2), 580–591 (2014)
34. Stewart, W.J.: Introduction to the Numerical Solution of Markov Chains. Princeton University Press (1994)
35. Urbina, D.I., Giraldo, J.A., Cárdenas, A.A., Tippenhauer, N.O., Valente, J., Faisal, M.A., Ruths, J., Candell, R., Sandberg, H.: Limiting the Impact of Stealthy Attacks on Industrial Control Systems. In: Weippl, E., Katzenbeisser, S., Kruegel, C., Myers, A.C., Halevi, S. (eds.) ACM CCS 2016. pp. 1092–1105. ACM (2016). https://doi.org/10.1145/2976749.2978388
36. Vigo, R., Nielson, F., Nielson, H.R.: Broadcast, Denial-of-Service, and Secure Communication. In: Johnsen, E.B., Petre, L. (eds.) iFM 2013. LNCS, vol. 7940, pp. 412–427. Springer (2013)
37. Villani, C.: Optimal transport, old and new. Springer (2008)
38. Zacchia Lun, Y., D'Innocenzo, A., Malavolta, I., Di Benedetto, M.D.: Cyber-Physical Systems Security: a Systematic Mapping Study. CoRR **abs/1605.09641** (2016)