

Trace-Checking CPS Properties: Bridging the Cyber-Physical Gap

Claudio Menghi
University of Luxembourg
Luxembourg, Luxembourg
claudio.menghi@uni.lu

Enrico Viganò *
University of Luxembourg
Luxembourg, Luxembourg
enrico.vigano@uni.lu

Domenico Bianculli
University of Luxembourg
Luxembourg, Luxembourg
domenico.bianculli@uni.lu

Lionel C. Briand
University of Luxembourg
Luxembourg, Luxembourg
University of Ottawa
Ottawa, Canada
lionel.briand@uni.lu

Abstract—Cyber-physical systems combine software and physical components. Specification-driven trace-checking tools for CPS usually provide users with a specification language to express the requirements of interest, and an automatic procedure to check whether these requirements hold on the execution traces of a CPS. Although there exist several specification languages for CPS, they are often not sufficiently expressive to allow the specification of complex CPS properties related to the software and the physical components and their interactions.

In this paper, we propose (i) the Hybrid Logic of Signals (HLS), a logic-based language that allows the specification of complex CPS requirements, and (ii) ThEodorE, an efficient SMT-based trace-checking procedure. This procedure reduces the problem of checking a CPS requirement over an execution trace, to checking the satisfiability of an SMT formula.

We evaluated our contributions by using a representative industrial case study in the satellite domain. We assessed the expressiveness of HLS by considering 212 requirements of our case study. HLS could express all the 212 requirements. We also assessed the applicability of ThEodorE by running the trace-checking procedure for 747 trace-requirement combinations. ThEodorE was able to produce a verdict in 74.5% of the cases. Finally, we compared HLS and ThEodorE with other specification languages and trace-checking tools from the literature. Our results show that, from a practical standpoint, our approach offers a better trade-off between expressiveness and performance.

Index Terms—Monitors, Languages, Specification, Validation, Formal methods, Semantics

I. INTRODUCTION

Cyber-physical systems (CPSs) combine cyber and physical capabilities [1]. Cyber capabilities are typically provided by software components that sense and act on the physical environment, while physical capabilities are provided by the environment in which the software is deployed. Therefore, CPSs combine software and physical dynamics. Physical dynamics are typically modeled through formalisms that capture the continuous evolution—according to physical laws—of the environment over time (e.g., differential equations); the corresponding behaviors are typically represented as continuous signals. Software (i.e., cyber) dynamics are typically modeled with discrete event systems (e.g., finite state machines), whose

behavior is typically represented by a sequence of events. Cyber-physical systems exhibit *hybrid* dynamics since they combine both physical and software capabilities.

Engineers collect traces (i.e., logs) describing the behavior of a CPS both when the CPS is simulated and, by means of instrumentation and logging mechanisms, during the actual execution of the CPS. A trace is a sequence of records that contain some information about the execution (or the simulation) of the cyber-physical components (e.g., the state of the system variables). Trace records are usually labeled with time-stamps representing the time instants at which the recorded information was obtained.

Engineers analyze these traces to check whether they conform to the system’s requirements specifications; this activity can be automated by means of trace-checking tools. Specification-driven trace-checking tools usually take as input a trace to be analyzed and a requirement specification; they yield a Boolean verdict indicating whether the trace satisfies the specification. The algorithms implemented by trace-checking tools are typically language-specific. In the context of trace checking, there exist two main categories of languages used for specifying CPS requirements: *time-based* and *sequence-based* languages.

Time-based languages (e.g., STL [2], STL* [3], RFOL [4], SFO [5], TPTL [6]. and SB-TemPsy-DSL [7]) interpret the records of the cyber and physical components as signals over a *time domain*. Specifications written in a time-based language express time relations over the occurrence of events. Such languages are suitable to express CPS requirements related to physical quantities; an example of such requirement is P1: “between 2 s and 10 s (measured starting from the origin of the trace) the speed of the satellite is lower than 10 m/s”. However, time-based languages are not easily amenable to specifying requirements related to software components. As an example, let us consider the requirement P2: “*whenever the satellite changes its mode from safe to normal, the speed of the satellite decreases*”. To express the first part of this requirement (marked in italics), one should specify that 1) in the trace there are two *consecutive* records; 2) the first record captures that the satellite is in “safe mode”; and 3) the second record captures that the satellite is in “normal mode”. This requirement cannot be easily expressed in time-

*This work was carried out while the author was a student at Politecnico di Milano (Italy), during an internship at the SnT Centre of the University of Luxembourg.

based languages since they generally do not provide constructs specifically designed to specify the first condition, i.e., that a record immediately follows another one in the trace. Indeed, expressing such a condition requires the specification language to provide access to the indices (i.e., positions in the trace) of the different records.

On the other hand, in sequence-based languages—such as LTL [8] (and domain-specific languages based on one of its extensions, like the one in the SpeAR tool [9]), FRETISH [10], and CoCoSpec [11]—traces are sequences of consecutive records, whose temporal model is represented by the sequence of *discrete indices* of the records. This class of languages interprets the records of the CPS software and physical components as discrete-time signals. Specifications in these languages constrain the indices in which events can occur; such specifications are used to express properties that mostly refer to the CPS software components, such as the first part of the aforementioned P2 property. However, these languages cannot express time relations over the occurrence of events, such as the one in property P1.

A third class of specification languages is the one of *hybrid* languages (e.g., STL-MX [12], HyLTL [13], HRELTL [14], Differential Dynamic Logic [15], HTL [16]), which support the specification of both continuous and discrete behaviors. However, these languages typically extend existing languages (e.g., LTL) to support the specification of hybrid behaviors in specific contexts (e.g., using signal derivatives). Therefore, they provide ad-hoc solutions that inherit some of the intrinsic limitations of the base language, thus hindering the expressiveness of the resulting hybrid language. For example, a hybrid language based on LTL cannot support metric operators to constrain the time distance between events.

The goal of this paper is to tackle the challenge of specifying hybrid behaviors of CPSs, in a way amenable to practical and efficient trace-checking. To reach this goal we propose:

(i) the *Hybrid Logic of Signals (HLS)*, a new specification language tailored to specifying CPS requirements. HLS allows engineers to express CPS requirements as properties (i.e., specifications) that refer both to the time-stamps and to the indices of the records of CPS traces. In this way, HLS specifications can easily express the behavior of both cyber and physical components, as well as their interactions.

(ii) *ThEodorE*, an efficient trace-checking approach for properties expressed in HLS. ThEodorE reduces the problem of checking an HLS property on a trace to a satisfiability problem, which can be solved using off-the-shelf Satisfiability Modulo Theories (SMT) solvers. The latter have efficient decision procedures for several background theories, thus making it possible to check whether a formula expressed in a first-order logic is satisfiable.

We evaluated our contribution using an *industrial* case study in the satellite domain, in collaboration with the engineers who developed the satellite’s on-board system.

(i) We assessed the expressiveness of HLS by checking whether it could express the 212 requirements of our case study. Our results show that HLS could express all the

requirements of our case study. We also compared HLS with SB-TemPsy-DSL [7] and STL [2], two specification languages proposed in the literature and for which trace-checking tools are available. The results show that HLS is significantly more expressive than SB-TemPsy-DSL and STL, which could only express 145 and 102 requirements, respectively.

(ii) We evaluated the trace-checking support provided by ThEodorE by assessing its applicability on 20 large traces provided by our industrial partner and obtained by simulating the behavior of the satellite across representative, different scenarios. We ran the ThEodorE trace-checker on 747 trace-requirement combinations. ThEodorE completed the verification in 74.5% of the cases within one hour, a reasonable time-out considering typical CPS development contexts. ThEodorE yielded a verdict for 67.9% of the 337 trace-requirement combinations containing a requirement that can not be verified by any of the other trace-checkers. We compared the applicability of ThEodorE with SB-TemPsy-Check [7] and Breach [17], for the trace-requirement combinations containing requirements expressible in SB-TemPsy-DSL and STL. For these combinations, SB-TemPsy-Check and Breach were 21.9% and 4.9% more often applicable than ThEodorE, respectively. SB-TemPsy-Check and Breach were also more efficient, but not to a point where it had practical implications.

Our results show that ThEodorE is broadly applicable as it allows engineers to specify a large variety of requirements while providing an efficient trace-checking procedure. Since in practical applications it is generally difficult to know in advance which requirement types engineers will need to specify, our findings suggest that ThEodorE is good default choice. However, if ThEodorE is not able to produce a verdict, and the requirement are expressible in SB-TemPsy-DSL or STL, engineers should then use SB-TemPsy-Check or Breach.

The paper is organized as follows. Section II describes our case study. Section III illustrates the syntax and semantics of HLS. Section IV presents ThEodorE. Section V evaluates our contribution based on an industrial case study. Section VI discussed related work. Section VII concludes the paper.

II. CASE STUDY AND MOTIVATIONS

Our industrial partner LuxSpace [18] developed, in collaboration with ESA [19] and ExactEarth [20], a maritime satellite to collect tracking information from vessels operating on Earth and to relay those data to the ground. This is a representative CPS made of complex software component interacting with many actuators and sensors and the physical environment where the satellite is to be deployed. This system should satisfy many varied requirements regarding the behavior of the software system itself but also its interactions with hardware and the satellite physical dynamics in space. Its development relies on technologies and practices typically seen in CPS contexts, e.g., Model-in-the-loop development with Simulink®.

Software engineers check the compliance of the satellite behavior to its requirements [21] both while the software is being developed and at run time. This is done by (i) collecting

ang-rate	20.1	22.2	23.3	20.4	21.1	3.2	1.1
mode	0	1	0	0	3	3	3
timestamp	0	0.2	0.9	1.8	3.0	4.9	5.7
index	0	1	2	3	4	5	6

Record r_3

Figure 1: A fragment of an execution trace of our case study.

execution traces of the system, and (ii) checking whether those traces satisfy the system requirements.

Figure 1 shows a fragment of an execution trace, which we will use to motivate this work. A trace is a sequence of records that contain some information about the execution of the system. In this example, the records include data about the *angular rate* (`ang-rate`) and the (*satellite*) *mode* (`mode`). The angular rate is a physical quantity represented by a real value measured by sensors. The mode is an enumeration of values that represent the state of the satellite software. There are four different modes: “Idle Mode”, “Safe Spin Mode”, “Normal Mode Coarse”, and “Normal Mode Fine”, which are represented in the trace by the values 0, 1, 2, and 3, respectively. In addition, each record is associated with a timestamp, representing the time instant at which the recorded information was obtained, and a progressive index value.

The requirements to be checked on the system traces refer both to the software and to the physical dynamics of the satellite. For example, let us consider requirement $\mathcal{R}1$: *Whenever the satellite mode switches from “Idle Mode” to “Normal Mode Fine”, the satellite angular rate shall reach a value lower than $1.5^\circ/\text{s}$ within 10 s. Moreover, the angular rate shall stabilize around an arbitrary value c lower than or equal to $1.5^\circ/\text{s}$.* $\mathcal{R}1$ specifies a constraint on a physical quantity, i.e., the angular rate of the satellite, which shall be ensured as a reaction to a software change, i.e., the satellite switching its mode from “Idle” to “Normal Mode Fine”.

One way to express that *the mode of the satellite switches from “Idle Mode” to “Normal Mode Fine”*, is to specify that the trace contains: 1) two records with *consecutive* indices; 2) the first record captures that the satellite is in “Idle Mode”; 3) the second record captures that the satellite is in “Normal Mode Fine”. This requirement cannot be easily expressed in the vast majority of time-based languages since they do not provide access to the indices of the different records. To compensate for this limitation when using time-based languages, engineers can apply ad-hoc solutions, such as adding a new Boolean flag to the trace records. In our example, such a flag would be true whenever the mode of the satellite switches from “Idle Mode” to “Normal Mode Fine”. In this way, the aforementioned requirement fragment would be rephrased as *the flag switch-from-IDLE-to-NORMAL-MODE-FINE is true*. However, this is impractical in real scenarios because: (i) the number of flags to add in the trace records can quickly grow and become unmanageable. For example, given

the four possible values for the satellite mode in our case study, to consider all possible combinations for switching satellite mode, engineers would need to add 16 values in each record (one for each mode switching combination). (ii) The requirement is reformulated and its connection to the actual software component behavior is lost.

Furthermore, requirement $\mathcal{R}1$ cannot be expressed using sequence-based languages because they do not support time relations over the occurrence of events. More specifically, expressing that “the [...] angular rate shall reach [...] **within 10 s**” requires to access the timestamps associated with the trace records (and compute a distance). This feature is not provided by sequence-based languages.

Moreover, to the best of our knowledge, among the time-based and sequence-based languages mentioned in the previous section, SFO [5] is the only language that allows users to use quantified variables in specifications, (as in “(there exist) **an arbitrary value** c lower than or equal to $1.5^\circ/\text{s}$ around which [...] shall stabilize”). This type of requirements is quite common in practical CPS applications, since engineers often want to check that the system stabilizes around a given value (e.g., the steady-state value). Although engineers know some properties of the steady-state value c (i.e., c shall be lower than or equal to $1.5^\circ/\text{s}$), they generally do not know its exact value, which has to be indicated as a generic variable in the requirement specification.

This example, extracted from our case study, shows the need for an expressive language for specifying hybrid behaviors of CPSs. In the next section, we will introduce a new specification language for CPSs, which overcomes the limitations—in terms of expressiveness—of state-of-the-art languages and is supported by an effective trace-checking procedure.

III. HYBRID LOGIC OF SIGNALS

In this section, we illustrate *HLS (Hybrid Logic of Signals)*, our new specification language for CPSs. We first discuss the design goals of the language (section III-A). Then, we define the mathematical model of the traces considered in this work (section III-B). Finally, we present the syntax (section III-C) and the semantics (section III-D) of the language.

A. Design goals

We designed HLS to provide a language for specifying CPS properties that seamlessly combine the features of sequence-based and time-based languages. Therefore, HLS extends existing time-based languages (e.g., STL [2], RFOL [4], and SFO [5]) and sequence-based languages (e.g., LTL [8], FRETISH [10], and CoCoSpec [11]) to allow engineers *to refer both to trace indices and to timestamps* in the logical specifications, *to arbitrarily combine them* to define properties describing the expected behavior of a CPS, and to express properties by quantifying over the values of the variables. More specifically, HLS allows engineers to use first-order *existential* and *universal* quantifiers with:

- *timestamp variables*, to express properties that refer to specific time instants and to the distance among them,

such as “*there exists a time instant t within 10s from the current time instant [...]*”;

- (trace) index variables, to express properties that refer to the indices of trace records, such as “*for every trace index i , such that the corresponding record captures that the satellite is in “Idle Mode”, and the immediately following record (at trace index $i + 1$) captures that the satellite is in “Normal Mode Fine” [...]*”;
- real-valued variables, to express properties that refer to arbitrary signal values, such as “*there exists a value c lower than or equal to $1.5^\circ/\text{s}$ around which the signal ang-rate shall stabilize*”.

Additionally, HLS supports specifications that use:

- the value of a signal *at a certain timestamp* or associated with a record *at a certain index*;
- the timestamp associated with the record *at a certain index*;
- the index of the record *with a certain timestamp*;
- expressions combining time variables, trace indices, and real-valued variables, using arithmetic and relational operators.

B. Traces

Let $\mathbb{J} = \{0, 1, 2, \dots, j, \dots, m\}$, with elements $j \in \mathbb{N}$, be a set of indices. Let \mathbb{T} be an interval of \mathbb{R} ; we call \mathbb{T} a time domain. Let $S = \{s_1, s_2, \dots, s_i, \dots, s_n\}$ be a set of variables (hereafter called “signals”) of the systems being monitored, with $s_i \in \mathbb{R}$. A trace π is a finite sequence of records $r_0, r_1, \dots, r_j, \dots, r_m$, with $j \in \mathbb{J}$.

Each record r_j is a tuple $\langle j, t, v_1, v_2, \dots, v_n \rangle$, where $j \in \mathbb{J}$ is the index associated with the record, $t \in \mathbb{T}$ is the timestamp at which the recorded information was obtained, and $v_1, v_2, \dots, v_n \in \mathbb{R}$ are the values associated with signals s_1, s_2, \dots, s_n in the record. For a trace π we use the array notation “[j]” to denote the j -th record of π , and we use the dot notation to denote an element of a record; we also introduce the notation t_j , short for $\pi[j].t$ for a given trace π . For example, let π^e be the fragment of the trace depicted in Figure 1; it contains seven records. Record r_3 is denoted by $\pi^e[3]$; it is represented by the tuple $\langle 3, 1.8, 0, 20.4 \rangle$, where $\pi^e[3].t = t_3 = 1.8$ is the value of the timestamp, $\pi^e[3].\text{mode} = 0$ is the value of signal mode, and $\pi^e[3].\text{ang-rate} = 20.4$ is the value of signal ang-rate.

We assume that the values associated with the timestamps are monotonically increasing, i.e., $t_j < t_{j+1}$, since records refer to consecutive timestamps. We say that a trace has a *fixed sample rate* sr if, for every $j, 0 \leq j < m$, $t_{j+1} - t_j = sr$, where sr is a constant value; otherwise, we say that the trace has a *variable sample rate*. For example, trace π^e in Figure 1 has a variable sample rate.

Additionally, we introduce a function $\iota_\pi: \mathbb{T} \rightarrow \mathbb{J}$: given a timestamp value t , $\iota_\pi(t)$ is the value of the index j of the record in π with the highest timestamp t_j such that $t_j \leq t$; we will omit the trace subscript when it is clear from the

context. For example, for trace π^e in Figure 1, $\iota_{\pi^e}(2.5) = 3$. In this work, we consider two definitions of ι :

$$\begin{aligned} \iota^V(t) &::= [t_0 \leq t] \cdot [t < t_1] \cdot 0 + [t_1 \leq t] \cdot [t < t_2] \cdot 1 + \\ &\quad \dots + [t_{m-1} \leq t] \cdot [t < t_m] \cdot (m-1) + [t_m = t] \cdot m \\ \iota^F(t) &::= \left\lfloor \frac{t}{sr} \right\rfloor \end{aligned}$$

Definition $\iota^V(t)$ assumes that the trace has a variable sample rate. Notice that the notation $[P]$, where P is a logical predicate, is the Iverson bracket; it evaluates to 1 if P is true, and to 0 otherwise. The resulting arithmetic formula checks where the timestamp t provided in input is situated w.r.t. the timestamps of the trace (i.e., t_0, t_1, \dots, t_m), and returns the value of the index of the record that has the highest timestamp that is smaller than or equal to t . For example, if the parameter t is greater than timestamp t_2 and lower than timestamp t_3 , the only expression in $\iota^V(t)$ that does not evaluate to 0 is $[t_2 \leq t] \cdot [t < t_3] \cdot 2$; therefore the index returned will be 2.

Definition $\iota^F(t)$ assumes that the trace has a fixed sample rate. In such a case, the index associated with a timestamp can be simply retrieved by computing the floor of the ratio of the timestamp t over the sample rate sr .

In this work, we assume that all the variables are sampled at each timestamp. This is a necessary requirement to enable the evaluation of the satisfaction of the system requirements at each timestamp. For systems that do not sample all the variables at each timestamp, engineers can use a pre-processing step to generate values to be assigned to variables for which the value is missing at certain timestamps. In this work, we consider two complementary pre-processing strategies:

$\mathcal{A}1$: In each record, an interpolation function (e.g., piecewise constant, linear, cubic) specific to each signal, is used to generate values for unassigned variables. Notice that this approach does not alter the original sample rate of the trace, since it keeps the same records as the original trace and only generates (in each record) values for the *unassigned* variables.

$\mathcal{A}2$: If the trace has a variable sample rate, it is converted into a trace with a fixed sample rate. This is done by generating a fresh set of records with a fixed sample rate equal to the smallest sample rate (i.e., the minimum time distance between two records) of the original trace, and by using the interpolation functions (as in the case of strategy $\mathcal{A}1$) to generate the values of *all* variables.

As we will discuss in Section V, the strategy used to generate the values of unassigned variables determines the trace accuracy. The latter influences the trace checking verdict and may impact on the correctness of the trace-checking procedure.

C. Syntax

An HLS formula is defined according to the grammar presented in Figure 2, whose start symbol is \wp . In the grammar, we use the symbol f to represent a generic (binary) arithmetic function; the symbol $|$ separates alternatives. In the following, we illustrate the various language constructs; in the explanations, we will refer to the set $TV = \{\tau_0, \tau_1, \dots\}$ of timestamp variables over \mathbb{T} , the set $IV = \{\sigma_0, \sigma_1, \dots\}$ of index variables

over \mathbb{J} , and the set $RV = \{\rho_0, \rho_1, \dots\}$ of real-valued variables over \mathbb{R} .

A *term* (non-terminal tm) can be either a *time term*, an *index term*, or a *value term*.

A *time term* (non-terminal tt) allows engineers to refer to timestamps in the specifications. A time term can be a timestamp variable $\tau \in TV$, a literal denoting a value $t \in \mathbb{T}$, the value returned by the operator $\mathbf{i2t}$, or an arithmetic expression over these entities. The operator $\mathbf{i2t}(\text{it})$ takes an index term as argument and returns the timestamp associated with the record at the (trace) index it . An example of time term is the expression $\tau_0 + 5.5 + \mathbf{i2t}(2)$.

An *index term* (non-terminal it) allows engineers to refer to trace indices in the specifications. An index term can be an index variable $\sigma \in IV$, a literal denoting a value $j \in \mathbb{J}$, the value returned by the operator $\mathbf{t2i}$, or an arithmetic expression over these entities. The operator $\mathbf{t2i}(\text{tt})$ takes a time term as argument and returns the index j of the trace record with timestamp t_j , where t_j is the highest timestamp value for which $t_j \leq \text{tt}$. An example of index term is the expression $\sigma_0 + 2 + \mathbf{t2i}(3.3)$.

A *value term* (non-terminal vt) allows engineers to refer to real values (e.g., signal values) in the specifications. A value term can be a real-valued variable $\rho \in RV$, a literal denoting a value $x \in \mathbb{R}$, the value of a signal returned by the operators $\mathbf{@i}$ (“at index”) and $\mathbf{@t}$ (“at timestamp”), or an arithmetic expression over these entities. The $\mathbf{@i}$ operator is an infix operator that takes two arguments: a signal s and an index term it ; it returns the value of signal s associated with the record at the (trace) index it . Similarly, the $\mathbf{@t}$ operator is an infix operator that takes two arguments: a signal s and a time term tt ; it returns the value of signal s associated with a record at timestamp t_j , where t_j is the highest timestamp value in the trace for which $t_j \leq \text{tt}$. An example of value term is the expression $(s_1 \mathbf{@i} 2) + (s_2 \mathbf{@t} 3.3) + \rho_0 + 5.2$, where s_1 and s_2 are signals, 2 is an index term, 3.3 is a time term, ρ_0 is a real-valued variable, and 5.2 is a numeric literal.

A *formula* (non-terminal p) is a relational expression over terms, a logical expression over other formulae defined using Boolean connectives, or an existentially quantified formula. As anticipated in section III-A, HLS supports three types of quantification:

- (i) over timestamp variables, as in “**exists** τ **in** I_T [...]”, where I_T is a time range with bounds in \mathbb{T} ;
- (ii) over index variables, as in “**exists** σ **in** I_J [...]”, where I_J is a range of index values with bounds in \mathbb{J} ;
- (iii) over real-valued variables, as in “**exists** ρ [...]”.

For example, the formula **exists** σ_0 **in** $[3, 5]$ **such that** $(s_1 \mathbf{@i} \sigma_0) < 2.5$ specifies that there exists a record with index greater than or equal to 3 and lower than or equal to 5, in which the value of signal s_1 is less than 2.5.

The language is further extended with additional relational operators, additional logical connectives (e.g., implication (**implies**), conjunction (**and**)), and universal quantifiers (**forall**) on timestamp variables, index variables, and real-valued variables, using the standard logical conventions.

<i>Term</i>	$\text{tm} ::= \text{tt} \mid \text{vt} \mid \text{it}$
<i>Time Term</i>	$\text{tt} ::= \tau \mid t \mid \mathbf{i2t}(\text{it}) \mid f(\text{tt}_1, \text{tt}_2)$
<i>Index Term</i>	$\text{it} ::= \sigma \mid j \mid \mathbf{t2i}(\text{tt}) \mid f(\text{it}_1, \text{it}_2)$
<i>Value Term</i>	$\text{vt} ::= \rho \mid x \mid (s \mathbf{@i} \text{it}) \mid (s \mathbf{@t} \text{tt}) \mid f(\text{vt}_1, \text{vt}_2)$
<i>Formula</i>	$\text{p} ::= \text{tm}_1 < \text{tm}_2 \mid \mathbf{not} \text{p} \mid \text{p}_1 \mathbf{or} \text{p}_2$ $\mid \mathbf{exists} \tau \mathbf{in} I_T \mathbf{such\ that} \text{p}$ $\mid \mathbf{exists} \sigma \mathbf{in} I_J \mathbf{such\ that} \text{p}$ $\mid \mathbf{exists} \rho \mathbf{such\ that} \text{p}$

$$t \in \mathbb{T}, j \in \mathbb{J}, x \in \mathbb{R}, \tau \in TV, \sigma \in SV, \rho \in RV, s \in S$$

Figure 2: Syntax of the Hybrid Logic of Signals.

We now present an application of HLS for the specification of one of the requirements in our case study. Let us consider a fragment of requirement $\mathcal{R}1$: *Whenever the satellite mode switches from “Idle Mode” to “Normal Mode Fine”, the satellite angular rate shall reach a value lower than $1.5^\circ/\text{s}$ within 10s.* We recall that the satellite mode is represented by the signal `mode`, for which value 0 corresponds to “Idle Mode” and value 3 corresponds to “Normal Mode Fine”; also, the angular rate is represented by the signal `ang-rate`. This fragment can be specified in HLS as:

```
forall  $\sigma_0$  in  $[0, 5]$  such that
((mode @i  $\sigma_0$ ) = 0 and (mode @i ( $\sigma_0 + 1$ )) = 3)
implies exists  $\tau_0$  in  $[0\text{s}, 10\text{s}]$  such that
(ang-rate @t ( $\tau_0 + \mathbf{i2t}(\sigma_0)$ ) < 1.5)
```

The sub-formula $((\text{mode } \mathbf{@i} \sigma_0) = 0 \mathbf{and} (\text{mode } \mathbf{@i} (\sigma_0 + 1)) = 3)$ detects when the satellite switches from “Idle Mode” to “Normal Mode Fine” over two consecutive records (notice the use of the “at index” operator to refer to the consecutive indices σ_0 and $\sigma_0 + 1$). This expression is within the scope of the outer universal quantifier, which iterates over a range of values for the index variable σ_0 . This range depends on the length of the trace and on the use of σ_0 in the formula. In this case, since the requirement says “*whenever* [the satellite mode switches. . .]”, in the specification we want to cover the full length of the trace fragment π^ϵ in Figure 1, where record index values span from 0 to 6. We achieve this by setting the lower bound to zero and the upper bound to five; in this way, the term $\text{mode } \mathbf{@i} (\sigma_0 + 1)$ always refers to a record index of the example trace.

The inner quantification over the timestamp variable τ_0 checks whether the angular rate of the satellite reaches a value lower than $1.5^\circ/\text{s}$ within 10s. More specifically, the expression $(\text{ang-rate } \mathbf{@t} (\tau_0 + \mathbf{i2t}(\sigma_0)) < 1.5)$ represents the value of signal `ang-rate` at timestamp $\tau_0 + \mathbf{i2t}(\sigma_0)$, where τ_0 is in the interval $[0\text{s}, 10\text{s}]$ (corresponding to the distance of 10s) and $\mathbf{i2t}(\sigma_0)$ is the timestamp at which the satellite switches from “Idle Mode” to “Normal Mode Fine”, i.e., the timestamp associated with the record at index σ_0 .

D. Semantics

To evaluate whether an HLS formula is true or false over a trace π , we must first define how time, index, and value terms are interpreted and evaluated.

Let $\mu^{TV}, \mu^{IV}, \mu^{RV}$ be variable assignments, respectively, for timestamp, index, and real-valued variables; for example, μ^{TV} is a mapping from a timestamp variable in TV to a value in \mathbb{T} . Let μ denote, collectively, the family of variable assignment functions $\mu^{TV}, \mu^{IV}, \mu^{RV}$. We evaluate a generic term tm on a trace π , using the variable assignment functions in μ , by means of an interpretation function $\llbracket \text{tm} \rrbracket_{\pi, \mu}$.

The interpretation of HLS terms is defined inductively at the top of figure 3. For all three term types, the interpretation of a literal is the value denoted by the literal itself; a variable is interpreted using the variable assignment function for the corresponding type; an arithmetic expression defined using a function f is interpreted by applying the interpretation of the function symbol f to the interpretation of the corresponding arguments. The operators **i2t**, **t2i**, **@i**, and **@t** are interpreted according to the informal semantics provided in the previous section.

The semantics of an HLS formula ϕ is defined over a trace π and a variable assignment μ ; we use the notation $(\pi, \mu) \models \phi$ to indicate that trace π satisfies formula ϕ under variable assignment μ . The satisfiability relation of HLS formulae is defined inductively at the bottom of figure 3. The formula $\text{tm}_1 < \text{tm}_2$ is satisfied if and only if (iff) the interpretation of term tm_1 is lower than the interpretation of term tm_2 . The semantics of the Boolean connectives **or** and **not** is the standard one. A formula with an existential quantifier over a timestamp variable, of the form **exists** τ **in** I_T **such that** p , is satisfied iff there exists a timestamp $t_j \in I_T$, such that when substituting timestamp t_j for τ in the formula p (denoted by $p[\tau \leftarrow t_j]$), the resulting formula is satisfied. The semantics of the other two types of formulae with an existential quantifier is defined in a similar way.

IV. TRACE CHECKING HLS FORMULAE

In this section, we present ThEodorE, our trace checker for HLS. ThEodorE reduces the problem of checking an HLS property on a trace to a satisfiability problem, which can be solved using off-the-shelf SMT solvers.

ThEodorE takes as input a property ϕ expressed in HLS and a trace π . The first step of ThEodorE is to automatically translating property ϕ and trace π formulae expressed using a target logic \mathcal{L} . This translation relies on two translation functions **h** (for HLS formulae, see Section IV-B) and **t** (for traces, see Section IV-A) and guarantees, that $(\pi, \mu) \models \phi$ iff $\mathbf{h}(\neg\phi) \wedge \mathbf{t}(\pi)$ is not satisfiable, where μ is a model for $\mathbf{h}(\neg\phi) \wedge \mathbf{t}(\pi)$, i.e., μ is a variable assignment leading to the property violation, consistent with the values of the variables of the trace records.

The second step of ThEodorE is checking the satisfiability of formula $\psi \equiv \mathbf{h}(\neg\phi) \wedge \mathbf{t}(\pi)$, expressed in the target logic \mathcal{L} using an SMT solver. Based on the condition stated above, when ψ is satisfiable, it means that ϕ does not hold on the

Time Term Interpretation	
$\llbracket \tau \rrbracket_{\pi, \mu} = \mu^{TV}(\tau)$, for all $\tau \in TV$;	
$\llbracket t \rrbracket_{\pi, \mu} = t$, for all $t \in \mathbb{T}$;	
$\llbracket \mathbf{i2t}(it) \rrbracket_{\pi, \mu} = \pi[\llbracket it \rrbracket_{\pi, \mu}].t$;	
$\llbracket f(t_1, t_2) \rrbracket_{\pi, \mu} = \llbracket f \rrbracket_{\pi, \mu}(\llbracket t_1 \rrbracket_{\pi, \mu}, \llbracket t_2 \rrbracket_{\pi, \mu})$;	
Index Term Interpretation	
$\llbracket \sigma \rrbracket_{\pi, \mu} = \mu^{IV}(\sigma)$, for all $\sigma \in IV$;	
$\llbracket j \rrbracket_{\pi, \mu} = j$, for all $j \in \mathbb{J}$;	
$\llbracket \mathbf{t2i}(tt) \rrbracket_{\pi, \mu} = \iota_{\pi}(\llbracket tt \rrbracket_{\pi, \mu})$;	
$\llbracket f(it_1, it_2) \rrbracket_{\pi, \mu} = \llbracket f \rrbracket_{\pi, \mu}(\llbracket it_1 \rrbracket_{\pi, \mu}, \llbracket it_2 \rrbracket_{\pi, \mu})$;	
Value Term Interpretation	
$\llbracket \rho \rrbracket_{\pi, \mu} = \mu^{RV}(\rho)$, for all $\rho \in RV$;	
$\llbracket x \rrbracket_{\pi, \mu} = x$, for all $x \in \mathbb{R}$;	
$\llbracket (s \ @i \ it) \rrbracket_{\pi, \mu} = \pi[\llbracket it \rrbracket_{\pi, \mu}].s$;	
$\llbracket (s \ @t \ tt) \rrbracket_{\pi, \mu} = \pi[\iota_{\pi}(\llbracket tt \rrbracket_{\pi, \mu})].s$	
$\llbracket f(vt_1, vt_2) \rrbracket_{\pi, \mu} = \llbracket f \rrbracket_{\pi, \mu}(\llbracket vt_1 \rrbracket_{\pi, \mu}, \llbracket vt_2 \rrbracket_{\pi, \mu})$;	
Formula Satisfaction	
$(\pi, \mu) \models \text{tm}_1 < \text{tm}_2$	iff $\llbracket \text{tm}_1 \rrbracket_{\pi, \mu} < \llbracket \text{tm}_2 \rrbracket_{\pi, \mu}$
$(\pi, \mu) \models \mathbf{not} \ p$	iff $(\pi, \mu) \not\models p$
$(\pi, \mu) \models p_1 \ \mathbf{or} \ p_2$	iff $(\pi, \mu) \models p_1$ or $(\pi, \mu) \models p_2$
$(\pi, \mu) \models \mathbf{exists} \ \tau \ \mathbf{in} \ I_T$	iff $(\pi, \mu) \models p[\tau \leftarrow t_j]$
$\quad \mathbf{such \ that} \ p$	for some $t_j \in I_T$
$(\pi, \mu) \models \mathbf{exists} \ \sigma \ \mathbf{in} \ I_J$	iff $(\pi, \mu) \models p[\sigma \leftarrow j]$
$\quad \mathbf{such \ that} \ p$	for some $j \in I_J$
$(\pi, \mu) \models \mathbf{exists} \ \rho$	iff $(\pi, \mu) \models p[\rho \leftarrow v]$
$\quad \mathbf{such \ that} \ p$	for some $v \in \mathbb{R}$

Figure 3: Semantics of the Hybrid Logic of Signals.

trace π . Vice-versa, when ψ is not satisfiable, it means that ϕ holds on the trace π .

The final verdict yielded by ThEodorE can be “satisfied”, “violated” or “unknown”; it is based on the answer of the solver. ThEodorE yields the *definitive verdicts* “satisfied” or “violated” when the solver returns “UNSAT” or “SAT”, indicating, respectively, that ψ is unsatisfiable or satisfiable. However, the solver may return an “UNKNOWN” answer, since the satisfiability of the underlying target logic \mathcal{L} is generally undecidable. In our case, this indicates that no conclusion is drawn on the satisfiability of formula ψ , resulting in an “unknown” verdict returned by ThEodorE. Assessing whether this is a frequent case in practical applications is part of our evaluation (Section V).

The target logic \mathcal{L} to be selected for trace checking of HLS properties in ThEodorE shall fulfill two goals:

G1: be sufficiently expressive to encode the logic-based representation of an input trace π and the (semantics of an) HLS formula ϕ . This means that it should include linear real arithmetic (to support real-valued and timestamp terms), quantifiers (since HLS is a first-order logic), and arrays (since a trace can be seen as an array of records).

G2: be supported by an efficient solver, so that the trace checking procedure for HLS formulae can be completed within practical time limits.

We have identified the *AUFLIRA* (Closed linear formulae with free sort and function symbols over one- and two-dimensional arrays of integer indices and real values) fragment of the SMT-LIB (Satisfiability Modulo Theories LIBrary)

logic [22] as a suitable target logic for ThEodorE. The theories used by AUFLIRA are identifiable through its name: **A**: arrays; **UF**: extension allowing free sort and function symbols; **LIRA**: linear integer and real arithmetics. Furthermore, AUFLIRA does not restrict the formulae to be quantifier-free. Based on the list of supported theories, AUFLIRA satisfies G1. It also satisfies G2, since it is included in the SMT-LIB logic, whose satisfiability can be verified using highly efficient and optimized solvers, as shown in the annual SMT competition [23].

In the following subsections we will describe functions \mathbf{t} and \mathbf{h} . For simplicity, we will present the translation using the syntax of the Z3 Python API [24].

A. Translating a Trace into the Target Logic

Function \mathbf{t} translates a trace π into a logic formula expressed using the target logic \mathcal{L} .

To represent the sequence of timestamps in π , the translation creates an array variable \mathbf{t} ; the type of the array indices (i.e., the domain of \mathbf{t}) is \mathbb{Z} , whereas the type of the array values (i.e., the range of \mathbf{t}) is \mathbb{R} . Then, the translation defines a series of constraints on the values in \mathbf{t} : the value of array \mathbf{t} at position i (denoted by $\mathbf{t}[i]$) is constrained to be equal to the value of the timestamp contained in the record at index i of trace π .

In addition, the translation creates an array variable for each signal whose values are recorded in the trace; the variable name is the string obtained by concatenating $\mathbf{v_}$ with the name of the signal. For each of these array variables representing signals, the translation defines a series of constraints on the values of the array: the value of the array in position i is constrained to be equal to the value of the corresponding signal in the record at index i of trace π .

B. Translating an HLS Formula into the Target Logic

Function \mathbf{h} translates an HLS formula into a logic formula expressed using the target logic \mathcal{L} .

First, the translation declares a new variable for each timestamp, index, and real-valued variable used in the HLS formula; the name of the new variable is the string obtained by concatenating $\mathbf{v_}$ with the name of the original variable. The type of the new variables is `Real` for timestamp and real-valued variables, and `Int` for index variables.

Afterwards, the translation recursively evaluates each node in the parse tree of the input formula, starting from the root node; each node is translated using the rules shown in Figure 4.

The translation of time, index, and values term nodes is defined as follows. Nodes referring to HLS variables are translated into the corresponding variables in the target logic formula. Literal nodes are mapped into literals in the target logic formula. Arithmetic expressions using a function f are translated by converting the function symbol into the equivalent in the target language, and then by applying it to the translation of its arguments. A time term node of the form $\mathbf{i2t}(it)$ is translated into an expression that accesses the element of the array \mathbf{t} in position $\mathbf{h}(it)$. An index term node of the form $\mathbf{t2i}(tt)$ is translated into the application of the translation of function \mathbf{t} to $\mathbf{h}(tt)$. A value term of the

Time Term
$\mathbf{h}(\tau) = \mathbf{v_}\tau$, for all $\tau \in TV$;
$\mathbf{h}(t) = t$, for all $t \in \mathbb{T}$;
$\mathbf{h}(f(tt_1, tt_2)) = \mathbf{h}(f)(\mathbf{h}(tt_1), \mathbf{h}(tt_2))$;
$\mathbf{h}(\mathbf{i2t}(it)) = \mathbf{t}[\mathbf{h}(it)]$;
Index Term
$\mathbf{h}(\sigma) = \mathbf{v_}\sigma$, for all $\sigma \in IV$;
$\mathbf{h}(j) = j$, for all $j \in \mathbb{J}$;
$\mathbf{h}(f(it_1, it_2)) = \mathbf{h}(f)(\mathbf{h}(it_1), \mathbf{h}(it_2))$;
$\mathbf{h}(\mathbf{t2i}(tt)) = \mathbf{h}(t)(\mathbf{h}(tt))$;
Value Term
$\mathbf{h}(\sigma) = \mathbf{v_}\sigma$, for all $\sigma \in RV$;
$\mathbf{h}(x) = x$, for all $x \in \mathbb{R}$;
$\mathbf{h}(f(vt_1, vt_2)) = \mathbf{h}(f)(\mathbf{h}(vt_1), \mathbf{h}(vt_2))$
$\mathbf{h}((s \ @i \ it)) = \mathbf{v_s}[\mathbf{h}(it)]$;
$\mathbf{h}((s \ @t \ tt)) = \mathbf{v_s}[\mathbf{h}(t)(\mathbf{h}(tt))]$;
Formula (with $I_T = [t_a, t_b]$ and $I_J = [a, b]$)
$\mathbf{h}(tm_1 < tm_2) = \mathbf{h}(tm_1) < \mathbf{h}(tm_2)$;
$\mathbf{h}(p_1 \ \mathbf{or} \ p_2) = \mathbf{Or}(\mathbf{h}(p_1), \mathbf{h}(p_2))$;
$\mathbf{h}(\mathbf{not} \ p) = \mathbf{Not}(\mathbf{h}(p))$;
$\mathbf{h}(\mathbf{exists} \ \tau \ \mathbf{in} \ I_T \ \mathbf{such} \ \mathbf{that} \ p) =$ $\mathbf{Exists}(\mathbf{v_}\tau, \mathbf{And}(\mathbf{And}(t_a \leq \mathbf{v_}\tau, \mathbf{v_}\tau \leq t_b), \mathbf{h}(p)))$
$\mathbf{h}(\mathbf{exists} \ \sigma \ \mathbf{in} \ I_J \ \mathbf{such} \ \mathbf{that} \ p) =$ $\mathbf{Exists}(\mathbf{v_}\sigma, \mathbf{And}(\mathbf{And}(a \leq \mathbf{v_}\sigma, \mathbf{v_}\sigma \leq b), \mathbf{h}(p)))$
$\mathbf{h}(\mathbf{exists} \ \rho \ \mathbf{such} \ \mathbf{that} \ p) = \mathbf{Exists}(\mathbf{v_}\rho, \mathbf{h}(p))$

Figure 4: Rules for translating HLS formulae into \mathcal{L} .

form $(s \ @i \ it)$ is translated into an expression that retrieves the value of variable $\mathbf{v_s}$ at index $\mathbf{h}(it)$. Similarly, a value term of the form $(s \ @t \ tt)$ is translated into an expression that retrieves the value of variable $\mathbf{v_s}$ at the index obtained through the evaluation of $\mathbf{h}(t)(\mathbf{h}(tt))$.

The translation of function \mathbf{t} supports both definitions presented in section III-B. It consists of a rewriting of the definition into the equivalent syntax of the target logic. We remark that the size of the arithmetic expression to compute $\mathbf{h}(t^V)$ in the case of a variable sample rate is linear in the length of the trace and the number of timestamp variables. Evaluating the impact of our translation and of the selection of the definition of function \mathbf{t} on the performance of the trace-checking procedure is part of our evaluation.

The translation of HLS formulae is basically their rewriting into the equivalent syntax of the target logic, modulo the translation of the variables and of the sub-formulae. For example, a formula of the form $\mathbf{exists} \ \sigma \ \mathbf{in} \ I_J \ \mathbf{such} \ \mathbf{that} \ p$ is rewritten as $\mathbf{Exists}(\mathbf{v_}\sigma, \mathbf{And}(\mathbf{And}(a \leq \mathbf{v_}\sigma, \mathbf{v_}\sigma \leq b), \mathbf{h}(p)))$, where the target logic variable $\mathbf{v_}\sigma$ corresponds to variable σ in the HLS formula, a and b are the lower and upper bounds of the closed interval I_J , and $\mathbf{h}(p)$ is the translation of sub-formula p .¹

ThEodorE ensures that $(\pi, \mu) \models \phi$ iff $\mathbf{h}(\neg\phi) \wedge \mathbf{t}(\pi)$ is not satisfiable. The correctness of our procedure is based on two arguments: (i) \mathbf{t} translates the trace π into a set of array

¹Our translation also supports open intervals. In this case, the relational operator $<$ (instead of \leq) is used in the target logic formula to constrain the values $\mathbf{v_}\sigma$ can assume.

variables whose values are set according to the values of the original trace, and (ii) \mathfrak{h} rewrites the HLS formula into the target logic without applying any change (that could alter the semantics) to the structure of the formula.

C. Implementation

We implemented ThEodorE as an Eclipse plugin using Xtext [25] and Xtend [26] and made it publicly available [27, 28]. We selected Z3 [24] as SMT solver, since it is an award-winning [29, 30], industry-strength tool. As such, it is likely to satisfy goal $G2$ discussed above. Checking whether this conjecture holds is part of our evaluation.

V. EVALUATION

In this section, we report on the evaluation of our contributions. First, we evaluate the expressiveness of HLS, and compare it with state-of-the-art specification languages. Second, we evaluate the applicability of the ThEodorE trace checker, and compare it to state-of-the-art tools. Specifically, we aim to answer the following research questions:

RQ1 *To which extent can HLS express requirements from industrial CPS applications and how does it compare with state-of-the-art specification languages in terms of expressiveness?* (section V-A)

RQ2 *Can ThEodorE verify CPS requirements on real-world execution traces within practical time and how does it compare with state-of-the-art tools?* (section V-B)

A. Expressiveness of HLS (RQ1)

To answer RQ1, we collected a set of industrial CPS requirements expressed in plain English text, and verified whether they could be expressed in HLS and in other state-of-the-art specification languages.

Dataset. We considered 212 industrial requirements from our satellite case study, coming from three different sources:

$S1$: 61 requirements were randomly selected from 745 requirements contained in the requirement specification document of the satellite on-board software (OBSW). Due to the prohibitive effort (more than 20 hours spanned across several working days) involved, both on our part and that of the domain experts who helped us formalize these requirements, we could only process a subset. Such requirements mostly refer to the software dynamics of the satellite, as in “*When the satellite switches to “Idle Mode”, the OBSW shall checkout the GPS, wait 50 ms, and then checkout the sun sensors*”.

$S2$: 101 requirements were provided by the authors of SB-TemPsy-DSL [7]. They mostly refer to the physical dynamics of the satellite, as in “*the beta angle [31] shall show an oscillatory behavior with a maximum period of 2500 s*”.

$S3$: 50 requirements were extracted from the design and architectural documents of the satellite. These documents describe the relations and interactions among the different components of the satellite. They contain cyber-physical requirements that relate the software and the physical dynamics of the satellite, as in “*if the satellite mode switches from “Idle Mode” to “Safe Spin Mode” and the satellite is not in eclipse,*

Table I: Number of requirements expressible in each of the languages for each set of requirements.

	$S1$	$S2$	$S3$	Total
HLS	61/61	101/101	50/50	212/212 (100%)
SB-TemPsy-DSL	34/61	92/101*	19/50	145/212 (68%)
STL	38/61	51/101*	13/50	102/212 (48%)

the magnetic field recorded by the magnetometer shall contain a spike with a maximum amplitude of 0.02 T”.

Methodology. We tried to express the requirements from our dataset using HLS and two state-of-the-art specification languages, namely SB-TemPsy-DSL [7] and STL [2]. We selected these languages because they are both supported by trace checking tools. We assessed the extent to which requirements were expressible in each language.

Results. Table I reports² the number of requirements that we were able to express in each of the languages, for each set of requirements ($S1$, $S2$, and $S3$). HLS was able to express 100% (212/212) of the requirements, while SB-TemPsy-DSL and STL were able to express 68% (145/212) and 48% (102/212) of the requirements, respectively. These results confirm that HLS is highly expressive and much more so than alternatives. We remark that all the HLS constructs were useful to express at least some of the considered CPS requirements, though in very different proportions.

The answer to RQ1 is that HLS could express *all* the requirements of our case study, many more than SB-TemPsy-DSL (+67) and STL (+110).

B. Applicability of ThEodorE (RQ2)

To answer RQ2, we (i) assessed to which extent ThEodorE can be applied to check the execution traces of our case study; (ii) compared, in terms of applicability, ThEodorE with SB-TemPsy-Check [7] and Breach [17]. SB-TemPsy-Check is the trace checker for SB-TemPsy-DSL; Breach is a trace checker for STL. We chose Breach among other similar tools listed in a recent survey [32] (i.e., AMT [33, 34] and S-TaLiRo [35]), because AMT 2.0, in contrast to Breach, is not publicly available, and because Breach is faster than S-TaLiRo [17]. Furthermore, we excluded from our comparison tools for online trace checking (e.g., SOCRaTEs [4] and RTAMT [36]).

Dataset. Our industrial partner provided 20 traces, obtained by simulating the behavior of the satellite in different scenarios; the simulation time ranged from four to six hours. Their size (in number of entries) ranges from 41844 to 1202241 entries ($avg = 389771$, $sd = 393718$); the corresponding file size ranges from ≈ 1.7 MB to ≈ 58.9 MB ($avg \approx 17.6$ MB, $sd \approx 19.4$ MB). The traces have a considerably large (yet variable) number of records and size.

²The values in Table I marked with an asterisk are slightly different from those reported in [7]. In the latter, quantification on real-valued variables (not supported in STL and SB-TemPsy-DSL) was handled by artificially selecting a value for the quantified variables within their quantification range. In this work, we marked such requirements as not specifiable.

For each trace in our dataset, our industrial partner indicated which requirements to check. Indeed, since only a subset of the satellite signals is recorded in each simulation scenario, not all the requirements have to be checked on each trace. In total, we considered 747 trace-requirement combinations: 320 obtained from requirements in $\mathcal{S}1$, 178 obtained from requirements in $\mathcal{S}2$, and 249 obtained from traces in $\mathcal{S}3$. We remark that, out of these 747 combinations, 337 involve a requirement that can be expressed neither in SB-TemPsy-DSL nor in STL.

Our industrial partner used a variable sample-rate for generating the trace records; hence not all the signal values were recorded at each sample index. Since our approach assumes that all the signals are assigned a value at each sample index, we pre-processed the traces. First, for each trace-requirement combination, we filtered out from the trace all the records that contained only signals that were not used in the HLS specification of the requirement. This step prevents the trace checker from handling an unnecessarily large set of records. Then, we transformed the traces using both pre-processing strategies $\mathcal{A}1$ and $\mathcal{A}2$ presented in section III-B; in both cases, the interpolation function to use for each signal was indicated by the engineers of our industrial partner.

By applying the $\mathcal{A}1$ and $\mathcal{A}2$ strategies on the original 747 trace-requirement combinations, the final dataset contains 1494 trace-requirement combinations (with half of them obtained using one of the two strategies). The size of the traces obtained using $\mathcal{A}1$ ranges from 2 to 17321 entries ($avg = 2071$, $sd = 3840$); the corresponding file size ranges from ≈ 15 B to ≈ 5.9 MB ($avg \approx 0.1$ MB, $sd \approx 0.4$ MB). The size of the traces obtained using $\mathcal{A}2$ ranges from 2 to 2360674 entries ($avg = 52406$, $sd = 185875$); the file size ranges from ≈ 15 B to ≈ 90.0 MB ($avg \approx 2.3$ MB, $sd \approx 8.4$ MB).

Methodology. We ran ThEodorE over the 1494 trace-requirements combinations in our dataset. When translating the HLS properties in the target logic, we used function ι^V for the trace-requirement combinations generated using strategy $\mathcal{A}1$ (since the pre-processed traces have a variable sample rate), and function ι^F for those generated using strategy $\mathcal{A}2$ (since the pre-processed traces have a fixed sample rate).

We conducted our evaluation on a high-performance computing platform, using nodes equipped with Dell C6320 units (2 Xeon E5-2680v4@2.4 GHz, 128 GB).³ Each run (checking a distinct combination of a trace and a property) was repeated 10 times, to account for variations in the performance of the HPC platform and of the SMT solver. In total, we executed $1494 \times 10 = 14940$ runs of ThEodorE. We allocated 4 GB of memory for each run and considered a timeout of one hour. We recorded whether the trace-checking procedure ended within the timeout, the trace checking result, and the time required to yield a verdict.

As for the comparison with SB-TemPsy-Check and Breach, we only considered the requirements from $\mathcal{S}2$ since it has the highest number of requirements expressible in SB-TemPsy-

³We executed our experiments on the HPC facilities of the University of Luxembourg [37].

Table II: Output of ThEodorE (percentage and execution time) when using the pre-processing strategies $\mathcal{A}1$ and $\mathcal{A}2$.

	Output	%	<i>avg</i>	<i>min</i>	<i>max</i>	<i>sd</i>
$\mathcal{A}1$	<i>satisfied</i>	53.9	80.2	0.01	2693.0	334.7
	<i>violated</i>	12.1	14.2	0.01	513.9	57.9
	<i>unknown</i>	1.6	6.5	5.8	7.4	0.6
	<i>timeout</i>	0.5	-	-	-	-
	<i>max_depth_exceeded</i>	13.0	-	-	-	-
	<i>out_of_memory</i>	18.9	-	-	-	-
$\mathcal{A}2$	<i>satisfied</i>	53.8	102.5	0.01	3432.9	331.7
	<i>violated</i>	20.7	96.5	0.01	3143.5	379.8
	<i>unknown</i>	2.2	8.7	5.4	12.3	2.1
	<i>timeout</i>	23.3	-	-	-	-

DSL and STL, and it was recently used for comparing SB-TemPsy-DSL with STL [7]. More specifically, we considered the 162 trace-requirement combinations (with requirements from the set $\mathcal{S}2$) expressible in SB-TemPsy-DSL, and the 103 trace-requirement combinations expressible in STL. We ran the tools following the same methodology described above. Since each run was repeated ten times, in total we considered 1620 runs of SB-TemPsy-Check and 1030 runs of Breach.

Results - Applicability of ThEodorE. Table II shows the different types of output returned by ThEodorE for checking the 7470 trace-requirement combinations generated using the variable sample rate interpolation (row $\mathcal{A}1$) and the fixed sample rate interpolation (row $\mathcal{A}2$). Column “%” indicates the percentage of cases in which each type of verdict was returned. For each of the cases in which ThEodorE finished within the timeout (i.e., it yielded a *satisfied*, *violated*, or *unknown* verdict), Table II also provides the average (*avg*), minimum (*min*), maximum (*max*) and standard deviation (*sd*) of the ThEodorE execution time (s).

The results in row $\mathcal{A}1$ show that ThEodorE finished within the timeout in 67.6% of the cases. In 66.0% of the cases, ThEodorE produced a definitive verdict (i.e., *satisfied* or *violated*); in 0.5% of the cases, ThEodorE timed out. ThEodorE returned a “*max_depth_exceeded - maximum recursion depth exceeded during compilation*” error in 13.0% of the cases, and an “*out_of_memory*” error in 18.9% of the cases; both errors are generated by the Z3 solver. The root cause of these errors is the translation of function ι^V , used in the case of variable sample rate traces: the size of the arithmetic expression resulting from the translation is linear in the length of the trace. As expected, ThEodorE inherits the limitations of SMT solvers and its applicability is expected to improve along with the quick pace of progress in that field.

The results in row $\mathcal{A}2$ show that ThEodorE finished within the timeout in 76.7% of the cases. In 74.5% of the cases, ThEodorE produced a definitive verdict; in 23.3% of the cases, ThEodorE timed out. When using strategy $\mathcal{A}2$, the number of times ThEodorE reached the timeout was higher than when using $\mathcal{A}1$. Indeed, many trace-requirement runs that generated *max_depth_exceeded* and *out_of_memory* errors in the case of $\mathcal{A}1$, timed out when using $\mathcal{A}2$. As discussed for the case of $\mathcal{A}1$, the applicability of ThEodorE when using $\mathcal{A}2$ is

determined by the scalability of the underlying SMT solver.

To evaluate whether ThEodorE is applicable in cases in which neither SB-TemPsy-Check nor Breach is applicable, we considered the subset of 3370 runs associated with the 337 trace-requirement combinations that involve a requirement that can be expressed neither in SB-TemPsy-DSL nor in STL. For those combinations, ThEodorE was able to produce a verdict in 67.9% of the cases.

To evaluate the impact of the trace accuracy (as determined by the application of the pre-processing strategies $\mathcal{A}1$ and $\mathcal{A}2$) on the correctness of the trace-checking procedure, we considered the 449 runs in which ThEodorE returned a definitive verdict both when using $\mathcal{A}1$ and when using $\mathcal{A}2$, and we compared the verdicts. In 95.1% of the cases (427 over 449), the verdicts coincided. For the 22 cases in which the verdicts were different, we manually inspected the generated traces and confirmed that differences in verdicts were caused by the pre-processing strategies.

Overall, these results show that ThEodorE, when configured with the pre-processing strategy based on a fixed sample rate ($\mathcal{A}2$), produced a definitive verdict for a considerable number of trace-requirement combinations (74.5%), thus confirming ThEodorE’s applicability in practical scenarios. Relying on the $\mathcal{A}2$ strategy led to a significantly wider applicability of ThEodorE than with the $\mathcal{A}1$ strategy (74.5% vs 66.0%), while resulting in negligible differences in trace accuracy. Therefore, for comparing ThEodorE with other tools, we resorted to using the $\mathcal{A}2$ pre-processing strategy.

Finally, we remark that ThEodorE detected an issue in the satellite design: some of the traces exhibited an unexpected spike in a signal related to the physical dynamics of the satellite, which was caused by a change in a signal related to its software dynamics.

Results - Comparison with other tools. Table III reports the percentage of cases in which ThEodorE, SB-TemPsy-Check, and Breach provided a verdict within the timeout and the minimum, maximum, average and standard deviation of the time required to yield the verdict.

The results show that, when the requirements are expressible in SB-TemPsy-DSL and STL, SB-TemPsy-Check and Breach are faster than ThEodorE. However, given the usage scenario considered in our work (offline trace checking), the difference in execution times reported in Table III does not have significant practical consequences since the average trace-checking time (less than two minutes) is significantly lower than the time required to collect the traces (several hours). Note that all tools were consistent in terms of verdicts: when ThEodorE returned a definitive verdict, it matched the verdict returned by SB-TemPsy-Check and Breach (when they did not time out).

The answer to RQ2 is that ThEodorE could compute a definitive verdict, within one hour, for 74.5% of the trace-requirement combinations of our industrial case study, and produced a verdict for 67.9% of the 337 trace-requirement combinations that could not be checked by the other tools.

Table III: Comparison of ThEodorE, SB-TemPsy-Check, and Breach in terms of the execution time.

Tool	%	avg	min	max	sd
ThEodorE	72.2	69.6	0.01	2506.2	317.6
SB-TemPsy	94.1	30.1	0.09	3440.0	310.1
ThEodorE	95.1	81.4	0.01	2506.2	345.7
Breach	100	0.03	0.02	0.1	0.007

C. Discussion and Threats to Validity

Based on results, we recommend the following workflow. Developers should initially use ThEodorE since its language (HLS) is the most expressive, and it is generally difficult to know in advance which requirement types engineers will need to specify. If the property to be verified does not contain the **t2i** HLS operator, which causes the generation of large arithmetic expressions, engineers should use ThEodorE with the pre-processing strategy based on a variable sample rate ($\mathcal{A}1$). If the property contains the **t2i** operator, engineers should use the pre-processing strategy based on a fixed sample rate ($\mathcal{A}2$). If ThEodorE was not able to produce a definitive verdict, and the requirement is expressible in SB-TemPsy-DSL or STL, engineers should use SB-TemPsy-Check or Breach.

Threats to validity. The requirements and traces we used in our evaluation come from a single case study in the satellite domain. Although this could influence the generalization of our results, our industrial case study is representative of what can be found in other cyber-physical domains, where the system requirements are complex properties related to the software system, its environment and their interactions, and traces are obtained by simulating (or executing) the behavior of the CPS in many different scenarios.

VI. RELATED WORK

Our contribution is mainly related to work done in the area of hybrid specification languages.

STL-MX [12] extends STL to define properties both on discrete time and on dense time. The language includes two layers, one based on LTL to express properties of discrete-time Boolean signals (sampled at a fixed sample rate), and another one based on STL, to express properties on dense-time real-valued signals. Time mapping operators define the conversion between dense-time and discrete-time signals and formulae. A trace-checking procedure has been proposed for STL-MX, but its implementation is not available. Compared with HLS, STL-MX restricts discrete-time Boolean signals to be sampled at a fixed sample rate, and lacks first-order quantifiers on real-valued variables.

HyLTL [13], HRELTL [14], and HTL [16] extend existing languages (e.g., LTL) with operators to express constraints on certain behaviors of signals (e.g., derivatives or limits). In contrast to HLS, they cannot express properties that refer to specific time instants and to the distance between them.

Differential Dynamic Logic [15] differs from HLS since it is designed for specifying properties of systems expressed using

the hybrid system [38] modeling formalism. As such, its modal operators enable references to the states that are reachable after firing the transitions of the hybrid system model.

The approach of reducing the trace-checking problem to the verification of the satisfiability of a logical formula has been also used in other works ([14, 39, 40]). However, our approach supports HLS, a more widely applicable language, and developed an efficient translation for it.

SOCRaTEs [4], Striver [41], TeSSLa [42], and RTLola [43] and a tool recently proposed by Arrieta et al. [44] are also related to our work. Unlike ThEodorE, which supports offline trace checking, these tools support online run-time verification.

To summarise, in our context and given our goal, in addition to the lack of trace-checking tools, none of the languages discussed above is as expressive as HLS. Taking into account the expressiveness limitations of state-of-the-art languages like SB-TemPsy-DSL and STL, which were not able to express many of our requirements (see section V), the development of a new language (and of the corresponding trace-checking tool) was indeed necessary.

VII. CONCLUSION

Software verification and validation requires specification-driven trace-checking techniques that strike a balance between the expressiveness of the specification language and the efficiency of its trace-checking procedures. In this paper, we specifically address this problem in the CPS domain. We proposed the Hybrid Logic of Signals (HLS), a specification language tailored to the specifics of CPS requirements. HLS allows engineers to specify complex CPS requirements related to its cyber and the physical components, as well as their interactions. Additionally, we developed ThEodorE, an efficient SMT-based trace-checking procedure for HLS.

We evaluated our solutions through a large-scale, complex industrial case study involving an on-board satellite system. Results show that our approach achieves a better trade-off between expressiveness and performance than existing solutions. HLS was able to express all system requirements in contrast to existing languages. As a result, ThEodorE supports a much wider set of property types than other trace checkers. In most cases, ThEodorE was able to check those properties within practical time limits. Furthermore, the applicability of ThEodorE is expected to improve in the future along with the underlying SMT technology. Last, based on results, we suggest a way to effectively combine various trace-checking tools.

As part of future work, we plan to develop trace diagnostics methods for HLS, inspired by existing work [45, 46], to explain the violations found by ThEodorE.

VIII. DATA AVAILABILITY

ThEodorE is publicly available [27, 28] under the Apache License 2.0. The entry on Zenodo.org [28] contains, in addition to the software, the files containing the results produced by ThEodorE, and the scripts to compute the aggregated results presented in the paper. The traces and requirements used in the experiments cannot be publicly released because they are subject to a non-disclosure agreement.

ACKNOWLEDGMENT

This work has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 694277), from the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Discovery and CRC programs.

The experiments presented in this paper were carried out using the HPC facilities of the University of Luxembourg [37] — see hpc.uni.lu.

REFERENCES

- [1] A. Platzer, *Logical foundations of cyber-physical systems*. Springer, 2018.
- [2] O. Maler and D. Nickovic, “Monitoring temporal properties of continuous signals,” in *Formal Techniques, Modelling and Analysis of Timed and Fault-Tolerant Systems*. Springer, 2004, pp. 152–166.
- [3] L. Brim, P. Dluhoš, D. Šafránek, and T. Vejpustek, “STL*: Extending signal temporal logic with signal-value freezing operator,” *Information and computation*, vol. 236, pp. 52–67, 2014.
- [4] C. Menghi, S. Nejati, K. Gaaloul, and L. C. Briand, “Generating automated and online test oracles for simulink models with continuous and uncertain behaviors,” in *European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*. ACM, 2019.
- [5] A. Bakhirkin, T. Ferrère, T. A. Henzinger, and D. Ničković, “The first-order logic of signals: keynote,” in *International Conference on Embedded Software*. IEEE Press, 2018, p. 1.
- [6] R. Alur and T. A. Henzinger, “Real-time logics: Complexity and expressiveness,” *Information and Computation*, vol. 104, no. 1, pp. 35–77, 1993.
- [7] C. Boufaied, C. Menghi, D. Bianculli, L. Briand, and Y. Isasi-Parache, “Trace-checking signal-based temporal properties: A model-driven approach,” in *International Conference on Automated Software Engineering (ASE 2020)*. IEEE, 2020, pp. 1004–1015.
- [8] E. A. Emerson and J. Y. Halpern, ““sometimes” and “not never” revisited: on branching versus linear time temporal logic,” *Journal of the ACM (JACM)*, vol. 33, no. 1, pp. 151–178, 1986.
- [9] A. W. Ficarek, L. G. Wagner, J. A. Hoffman, B. D. Rodes, M. A. Aiello, and J. A. Davis, “SpeAR v2. 0: Formalized past LTL specification and analysis of requirements,” in *NASA Formal Methods Symposium*. Springer, 2017, pp. 420–426.
- [10] D. Giannakopoulou, T. Pressburger, A. Mavridou, and J. Schumann, “Generation of formal requirements from structured natural language,” in *Requirements Engineering: Foundation for Software Quality (REFSQ 2020)*. Springer, 2020, pp. 19–35.
- [11] A. Champion, A. Gurfinkel, T. Kahsai, and C. Tinelli, “CoCoSpec: A mode-aware contract language for reactive systems,” in *International Conference on Software*

- Engineering and Formal Methods*. Springer, 2016, pp. 347–366.
- [12] T. Ferrère, O. Maler, and D. Ničković, “Mixed-time signal temporal logic,” in *Formal Modeling and Analysis of Timed Systems*. Springer, 2019, pp. 59–75.
- [13] D. Bresolin, “HyLTL: a temporal logic for model checking hybrid systems,” *arXiv preprint arXiv:1308.5336*, 2013.
- [14] A. Cimatti, M. Roveri, and S. Tonetta, “HRETL: A temporal logic for hybrid systems,” *Information and Computation*, vol. 245, pp. 54 – 71, 2015.
- [15] A. Platzer, “Differential dynamic logic for hybrid systems,” *Journal of Automated Reasoning*, vol. 41, no. 2, pp. 143–189, 2008.
- [16] T. A. Henzinger, Z. Manna, and A. Pnueli, “Towards refining temporal specifications into hybrid systems,” in *Hybrid systems*. Springer, 1992, pp. 60–76.
- [17] A. Donzé, T. Ferrère, and O. Maler, “Efficient robust monitoring for STL,” in *Computer Aided Verification Conference (CAV)*. Springer, 2013, pp. 264–279.
- [18] (2020) Luxspace. [Online]. Available: <https://luxspace.lu/>
- [19] “The European Space Agency (ESA),” 2020. [Online]. Available: <https://www.esa.int/>
- [20] “exactEarth,” 2020. [Online]. Available: <https://www.exactearth.com/>
- [21] (2020) Satellite development phases. [Online]. Available: https://www.esa.int/Science_Exploration/Space_Science/Building_and_testing_spacecraft
- [22] C. Barrett, A. Stump, and C. Tinelli, “The SMT-LIB standard - version 2.6,” Department of Computer Science, The University of Iowa, Tech. Rep., 2017.
- [23] T. Weber, S. Conchon, D. Déharbe, M. Heizmann, A. Niemetz, and G. Reger, “The SMT competition 2015-2018,” *J. Satisf. Boolean Model. Comput.*, vol. 11, no. 1, pp. 221–259, 2019.
- [24] (2020) Z3. [Online]. Available: <https://github.com/Z3Prover/z3>
- [25] “Xtext,” <https://www.eclipse.org/Xtext/>, 2020.
- [26] “Xtend,” <https://www.eclipse.org/xtend/>, 2020.
- [27] “ThEodorE,” <https://github.com/SNTSVV/ThEodorE>, 2020.
- [28] C. Menghi, E. Viganò, D. Bianculli, and L. C. Briand, “Theodore - trace-checker,” Feb. 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.4506795>
- [29] “ACM SIGPLAN - Programming Languages Software Award,” <http://www.sigplan.org/Awards/Software/>, 2020.
- [30] “ETAPS 2018 Test of Time Award,” <https://etaps.org/about/test-of-time-award/test-of-time-award-2018>, 2020.
- [31] (2020) Beta angle. [Online]. Available: https://en.wikipedia.org/wiki/Beta_angle
- [32] E. Bartocci, J. Deshmukh, A. Donzé, G. Fainekos, O. Maler, D. Ničković, and S. Sankaranarayanan, “Specification-based monitoring of cyber-physical systems: a survey on theory, tools and applications,” in *Lectures on Runtime Verification*. Springer, 2018, pp. 135–175.
- [33] D. Ničković, O. Lebeltel, O. Maler, T. Ferrère, and D. Ulus, “AMT 2.0: Qualitative and quantitative trace analysis with extended signal temporal logic,” in *Tools and Algorithms for the Construction and Analysis of Systems*. Springer, 2018, pp. 303–319.
- [34] D. Ničković, O. Lebeltel, O. Maler, T. Ferrère, and D. Ulus, “AMT 2.0: qualitative and quantitative trace analysis with extended signal temporal logic,” *International Journal on Software Tools for Technology Transfer*, pp. 1–18, 2020.
- [35] Y. Annpureddy, C. Liu, G. Fainekos, and S. Sankaranarayanan, “S-taliro: A tool for temporal logic falsification for hybrid systems,” in *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*. Springer, 2011, pp. 254–257.
- [36] D. Ničković and T. Yamaguchi, “RTAMT: Online robustness monitors from STL,” in *International Symposium on Automated Technology for Verification and Analysis (ATVA)*. Springer, 2020, pp. 564–571.
- [37] S. Varrette, P. Bouvry, H. Cartiaux, and F. Georgatos, “Management of an academic hpc cluster: The ul experience,” in *Proc. of the 2014 Intl. Conf. on High Performance Computing & Simulation (HPCS 2014)*. Bologna, Italy: IEEE, July 2014, pp. 959–967.
- [38] R. Alur, *Principles of cyber-physical systems*. MIT Press, 2015.
- [39] M. M. Bersani, D. Bianculli, C. Ghezzi, S. Krstić, and P. San Pietro, “SMT-based checking of SOLOIST over sparse traces,” in *International Conference on Fundamental Approaches to Software Engineering (FASE)*, vol. 8411. Springer, 2014, pp. 276–290.
- [40] D. Bianculli, C. Ghezzi, S. Krstić, and P. San Pietro, “Offline trace checking of quantitative properties of service-based applications,” in *International Conference on Service Oriented Computing and Application (SOCA)*. IEEE, 2014, pp. 9–16.
- [41] F. Gorostiaga and C. Sánchez, “Striver: Stream runtime verification for real-time event-streams,” in *International Conference on Runtime Verification*. Springer, 2018, pp. 282–298.
- [42] L. Convent, S. Hungerecker, M. Leucker, T. Scheffel, M. Schmitz, and D. Thoma, “TeSSLa: temporal stream-based specification language,” in *Brazilian Symposium on Formal Methods*. Springer, 2018, pp. 144–162.
- [43] P. Faymonville, B. Finkbeiner, S. Schirmer, and H. Torfah, “A stream-based specification language for network monitoring,” in *Runtime Verification*. Springer, 2016, pp. 152–168.
- [44] A. Arrieta, J. A. Agirre, and G. Sagardui, “A tool for the automatic generation of test cases and oracles for simulation models based on functional requirements,” in *International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*. IEEE, 2020, pp. 1–5.
- [45] W. Dou, D. Bianculli, and L. Briand, “Model-driven trace diagnostics for pattern-based temporal specifications,” in

International Conference on Model Driven Engineering Languages and Systems (MODELS). ACM/IEEE, 2018, pp. 278–288.

- [46] T. Ferrère, O. Maler, and D. Ničković, “Trace diagnostics using temporal implicants,” in *International Symposium on Automated Technology for Verification and Analysis (ATVA)*. Springer, 2015, pp. 241–258.