Towards Building Verifiable CPS using Lingua Franca*

Citation for published version:

Digital Object Identifier (DOI):
10.1145/3609134

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published in:
ACM Transactions on Embedded Computing Systems

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Towards Building Verifiable CPS using Lingua Franca

SHAOKAI LIN, University of California, Berkeley, USA
YATIN A. MANERKAR, University of Michigan, USA
MARTEN LOHSTROH, University of California, Berkeley, USA
ELIZABETH POLGREEN, University of Edinburgh, Scotland
SHENG-JUNG YU, University of California, Berkeley, USA
CHADLIA JERAD, University of Manouba, Tunisia
EDWARD A. LEE and SANJIT A. SESHIA, University of California, Berkeley, USA

Formal verification of cyber-physical systems (CPS) is challenging because it has to consider real-time and concurrency aspects that are often absent in ordinary software. Moreover, the software in CPS is often complex and low-level, making it hard to assure that a formal model of the system used for verification is a faithful representation of the actual implementation, which can undermine the value of a verification result.

To address this problem, we propose a methodology for building verifiable CPS based on the principle that a formal model of the software can be derived automatically from its implementation. Our approach requires that the system implementation is specified in Lingua Franca (LF), a polyglot coordination language tailored for real-time, concurrent CPS, which we made amenable to the specification of safety properties via annotations in the code. The program structure and the deterministic semantics of LF enable automatic construction of formal axiomatic models directly from LF programs. The generated models are automatically checked using Bounded Model Checking (BMC) by the verification engine Uclid5 using the Z3 SMT solver. The proposed technique enables checking a well-defined fragment of Safety Metric Temporal Logic (Safety MTL) formulas. To ensure the completeness of BMC, we present a method to derive an upper bound on the completeness threshold of an axiomatic model based on the semantics of LF. We implement our approach in the LF Verifier and evaluate it using a benchmark suite with 22 programs sampled from real-life applications and benchmarks for Erlang, Lustre, actor-oriented languages, and RTOSes. The LF Verifier correctly checks 21 out of 22 programs automatically.

CCS Concepts: • Computer systems organization → Embedded software; • Software and its engineering → Formal software verification; Model checking;

This article appears as part of the ESWEEK-TECS special issue and was presented in the International Conference on Embedded Software (EMSOFT), 2023.

The work in this paper was supported in part by the National Science Foundation (NSF), awards #CNS-1836601 (Reconciling Safety with the Internet) and #CNS-2233769 (Consistency vs. Availability in Cyber-Physical Systems) and the iCyPhy Research Center (Industrial Cyber-Physical Systems), supported by Denso, Siemens, and Toyota. This work was also supported in part by DARPA grant FA8750-20-C-0156 and by Intel.

Authors’ addresses: S. Lin, M. Lohstroh, S.-J. Yu, E. A. Lee, and S. A. Seshia, University of California, Berkeley, Cory Hall, Berkeley, CA, 94720-1770, USA; emails: {shaokai, marten, shengjunyu, eall}@berkeley.edu, sseshia@eecs.berkeley.edu; Y. A. Manerkar, University of Michigan, 2260 Hayward Street, Ann Arbor, MI 48109-2121, USA; email: manerkar@umich.edu; E. Polgreen, University of Edinburgh, 10 Crichton St, Edinburgh, Lothian, EH8 9AB, UK; email: elizabeth.polgreen@ed.ac.uk; C. Jerad, University of Manouba, Campus universitaire de la Manouba, Manouba, Manouba Governorate, 2010, Tunisia; email: chadlia.jerad@ensi-uma.tn.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2023 Copyright held by the owner/author(s).
1539-9087/2023/09-ART155 $15.00
https://doi.org/10.1145/3609134

1 INTRODUCTION

Cyber-physical systems (CPS) [47, 49, 66] play a major role in today’s society. From sensor networks to autonomous vehicles, from industrial automation to avionics, these systems have permeated every aspect of our daily lives. CPS applications are often safety-critical; needless to say, it is crucial that such systems are designed and implemented with assurance of correctness.

CPS design and verification can be challenging because one needs to capture the interactions between the digital world and the physical world, and these two worlds have properties that cannot be unified easily [47]. For example, the C language, a mainstream programming language for embedded applications, adopts a sequential programming model, whereas the physical world is intrinsically concurrent. In addition, timing is mostly considered as a side effect of the software implementation and most languages do not include time in their syntax or semantics; however, time is central to the evolution of physical processes and timing should be treated as a correctness criterion rather a performance metric. A CPS-oriented technology stack should address both concurrency and timing, which should also be addressed in CPS formal models.

CPS verification needs more automation. Although advances in model checking and automated theorem proving, such as bounded model checking (BMC) [12, 13] and satisfiability modulo theory (SMT) solvers [7], have gained popularity due to their capabilities to help automate verification, verification models are often generated by hand. This can be greatly time-consuming for the highly complex systems typical in CPS [47]. It is important to lower the burden of applying formal methods by fully automating verification, which requires not only automated verification algorithms but also generating verification models automatically from systems under verification, with an interface that is easy-to-use and tightly integrated with design and development. In the ideal case, the language used for design should also naturally support the verification process.

One natural formalism for CPS modeling is that of discrete-event systems (DES) [15]. However, formally modeling the behavior of DES using a conventional operational model (i.e., a transition system) is difficult because one needs to specify properties globally over execution traces, a task that cannot be easily achieved at the transition level. In addition, modeling concurrent execution often involves modeling partial order traces, which typically involves the use of nondeterminism. Therefore, when it comes to modeling real-time, concurrent CPS as DES, one needs to look for a modeling strategy more suitable than the operational approach. In this paper, we focus on building axiomatic models.

Our solution. Our solution for building verifiable CPS starts with the adoption of LINGUA FRANCA (LF) [51], a language suitable for modeling and implementing CPS applications. LINGUA FRANCA is a polyglot coordination language designed to have native support for time and concurrency and adopt a discrete-event semantics. LF endows popular programming languages like C (which we focus on in this paper), C++, and Rust with mechanisms to define and compose reactive and concurrent software components called reactors. While traditional CPS implementations in low-level languages (e.g., C) expose no easy way of extracting the interactions between subsystems, LF programs provide highly analyzable program structures and a well-defined...
deterministic semantics, exposing the interactions between subsystems clearly and explicitly. These features make it well-suited for formal verification. LF also blurs the boundary between models and implementations by allowing the user to specify both the software and the environment. LF has been used to implement real-life applications, including Autoware [8], AUTOSAR [56], distributed databases [48], and labs in embedded systems courses [38].

In this paper, we aim to ensure the functional correctness of LF programs and thus define a verification problem of checking Safety Metric Temporal Logic (Safety MTL) formulas [59] for LF programs. To perform verification, we present a methodology to encode the execution of an LF program as an axiomatic model, instead of an operational model that aims to build an abstraction of the system using explicit notions of states and transitions [1, 31]. Axiomatic models allow us to specify LF behavior based on a well-defined LF semantics. An axiomatic model can thus be sound w.r.t. all valid runtime implementations that correctly implement the same LF semantics. This is a significant advantage over operational models, which need to be carefully designed to be sound abstractions of a subset of the possible implementations. To make BMC complete, we propose an algorithm that calculates an upper bound on the BMC completeness threshold based on an LF program under verification and a Safety MTL property.

We implement our approach in the LF Verifier, which can verify an LF program fully automatically. For each LF implementation, the LF Verifier generates a verification model for Uclid5 [64, 70], a modeling and verification framework. The Uclid5 model is compiled into an SMT query and then checked automatically by the Z3 solver [18]. The Uclid5 model generation is enabled by the program structure exposed by LF and its deterministic semantics.

Related work. Outside of Lingua Franca, our work is influenced by the existing literature on verifying dataflow languages [22, 27, 32, 33, 73], Petri nets [23, 26], actors [6, 45, 68, 74], and other SMT-based techniques [1, 57]. The Uclid5 verifier we use is an evolution of the older UCLID system [14], one of the first SMT solvers and SMT-based verifiers. In terms of the underlying encoding and the verification strategies, we note that various works [32, 54, 57, 65] implement similar axiomatic models in other domains.

Within the scope of Lingua Franca, we are aware of two works that are directly related to verifying reactor systems at the time of this writing. Sirjani et al. [71] map LF programs onto models written in the Timed Rebeca language [67] and perform model checking using the Rebeca model checker [37], an explicit-state model checker. They model the reactor systems based on the event-based semantics and show that their approach can handle state properties. Deantoni et al. [19] integrate the MoCML [20] into LF’s Eclipse IDE and leverage the CADP toolbox [29] to check CCSL [3] properties, which can be considered a variant of dataflow properties. We will expand on the difference between state and dataflow properties in Section 4.

Our work differs from these prior approaches significantly:

1. LF Verifier is the first fully automated verifier for Lingua Franca and generates verification models directly from LF programs. Existing approaches require the user to first manually translate an LF program into an analyzable representation.
2. Our approach supports a well-defined fragment of Safety MTL, a temporal logic well-suited for real-time systems.
3. Our tooling uses an intermediate Uclid5 encoding that is much easier to manipulate and matches the LF semantics precisely.

Contributions. This paper makes the following key contributions:

- We advocate for the use of Lingua Franca for a "design-for-verification" approach to CPS design. We define a verification problem for checking Safety MTL properties of LF programs.
We present an approach to performing BMC on axiomatic models of LF programs, which models executions of LF programs at the semantics level and hence ensures soundness for all valid LF runtime implementations. To ensure the completeness of BMC, we present an algorithmic method to upper bound the completeness threshold based on an LF program and a Safety MTL property.

We present the LF Verifier, which can automatically generate axiomatic models directly from an LF program. The models are subsequently checked by the UCLID5 verification engine with the Z3 solver. The LF Verifier also provides an annotation system fully integrated into the LF development environment.

We developed a benchmark suite that consists of 22 LF programs sampled from real-life applications and benchmarks for Erlang, Lustre, actor-oriented languages, and RTOSes. And we evaluate the LF Verifier against the benchmark suite. The results show that the LF Verifier correctly checks 21 out of 22 programs automatically.

The remainder of the paper is organized as follows. Section 2 presents a running example of a simplified ADAS system encoded in LF. Section 3 presents background on Safety Metric Temporal Logic (Safety MTL) and Bounded Model Checking (BMC). In Section 4, we present our axiomatic approach for modeling the execution of an LF program and checking Safety MTL formulas using BMC. To ensure the completeness of BMC, Section 5 presents a method to compute an upper bound for a completeness threshold. In Section 6, we present the implementation details of the LF Verifier and work through verifying a property in the running example. In Section 7, we evaluate the LF Verifier against a benchmark suite. We conclude the paper and discuss the future work in Section 8.

2 Running Example

The reactor model [50, 53] is a deterministic model of concurrent computation for reactive systems. It adopts a discrete-event semantics and extends the actor model [35] with a deterministic order of handling concurrent timestamped messages. Reactors also share similarities with synchronous languages [9], logical execution time (LET) [34], dataflow [21, 58], and process networks [39].

To informally introduce the reactor model, we present a simplified model of an ADAS (Advanced Driver Assistance System) system [75], commonly found in modern vehicles. In this system, a camera and a LiDAR periodically collect data from the environment and send them to an ADAS processor, which performs image recognition on the given data. If an object is getting too close to the vehicle, a warning message is displayed on the dashboard. If the object is within some safety-critical distance, then the ADAS processor automatically applies the brakes. The user can also manually apply the brakes by pressing on the pedal.

We now explain how the diagram in Figure 1 represents this simplified ADAS system. A reactor is a stateful concurrent component. In the diagram, reactors are represented as rectangles with rounded corners, and they are Camera, LiDAR, Pedal, Brake, Dashboard, ADASProcessor, and ADASModel. A reactor can contain reactions, actions, timers, ports, and connections.

A reaction (e.g., line 6) is a routine that executes when any of its triggers is present, and it is rendered as a chevron in the diagram. The body of a reaction (e.g., lines 7–9) is written in a target language and represents the "business logic" of the application. In this example, the target language is C (line 1), and C code snippets are wrapped between the {$=}$ and {$=}$ brackets. In addition, the ADASProcessor reactor (lines 35–52) has two reactions, the one on the left labeled as 1 and the one on right as 2. These numbers correspond to the priorities of the reactions within ADASProcessor: when they are triggered simultaneously, reaction 1 will fire before reaction 2. It is important to note that reaction priorities establish a total order only within the same reactor.
A reaction can be triggered by a port (e.g., line 36), a timer (e.g., line 5), or an action (e.g., line 40). A port, rendered as a black triangle in the diagram, is used to communicate with other reactors. A timer, rendered as a clock in the diagram, is used to generate periodic signals. An action is used to schedule future events within a reactor. It is rendered as a white triangle with an “L” if it is a logical action, or a “P” if it is a physical action. A logical action is scheduled by a reaction, such as reaction 2 in ADASProcessor. A physical action is scheduled asynchronously by the environment, such as a press on the brake pedal that triggers an interrupt service routine (ISR). Connections can be drawn between two ports using connection statements such as lines 73–77. Logical connections are specified using -> with an optional logical delay specified using after (e.g., line 76). Without a delay specified, an event travels from an output port to an input port logically instantaneously.

All events in the system are handled in the order of tags, which use a superdense representation of time \( t \), encoded as tuples of the form \( g = (t, m) \), where \( t \in T \) is the time value and \( m \in \mathbb{N} \) is a microstep index. Reactions are logically instantaneous; logical time does not elapse during the execution of a reaction (physical time, denoted as \( T \), on the other hand, does elapse). If a reaction produces an output that triggers another reaction, then the two reactions execute logically simultaneously (i.e., at the same tag). Deadlines can be specified and are violated when \( T - t \) is greater than the specified intervals (e.g., lines 35 & 51).
**LF features used in this work.** The formal analyses from the proposed methodology operate on the *logical* behavior of the program, and the set of language features this work addresses includes all features introduced earlier except *physical* actions, deadlines, as well as advanced features such as modal models [69]. **LINGUA FRANCA** is a highly expressive language, and our proposed methodology makes use of most but not all of the language features available. We leave analyzing programs with these language features for future work.

**Safety properties in reactors.** Safety properties are properties that state, in loose terms, “something bad should not happen.” We focus on safety properties because they are generally considered the most important class of properties for safety-critical systems, and they tend to be amenable to automated techniques [33]. In addition, since bounded liveness properties can be converted into safety properties [10], we do not make a distinction between bounded liveness and safety properties.

In reactors, we identify three types of properties: (I) *state* properties, (II) *dataflow* properties, and (III) *mixed* properties. State properties focus on the values of variables. In the context of reactors, they include state variables, ports, and actions. Dataflow properties, on the other hand, specify the temporal behavior of *reaction invocations*. Mixed properties combine state and dataflow properties.

To illustrate the difference between these types of properties, consider the ADASModel reactor shown in Figure 1. The property “The ‘frame’ variable in Camera is initially set to an empty value” is a state property (Type I) that only refers to the state variable frame. The property “In ADASProcessor,Reaction 2 is invoked within 100 ms after Reaction 1 is invoked.” is a dataflow property (Type II). Notice that this property only refers to reaction invocations and does not refer to variables. Lastly, a mixed property (Type III) is shown at lines 60–65 in the LF program in Figure 1. In natural language, the property states “Globally from time 0 to 10 ms, when the LiDAR reaction invokes and the ADASProcessor requests stop at the same tag, the brakes will eventually be applied within 55 ms.” We write the property in MTL

\[
G[0, 10 \text{ ms}](\text{ADASModel}_l\_reaction_0 \&\& (F[0](\text{ADASModel}_p\_requestStop == 1))) \\
\text{implies} \ (F[0, 55 \text{ ms}](\text{ADASModel}_b\_brakesApplied == 1)))
\]

where ADASModel\_l\_reaction\_0 refers to the invocation of the reaction inside the LiDAR reactor, ADASModel\_p\_requestStop is the requestStop variable inside the ADASProcessor reactor, and ADASModel\_b\_brakesApplied is the brakesApplied variable inside the Brakes reactor. We wrap the property inside an @property annotation, which we will expand on and verify in Section 6.

At a high level, the key contribution of this paper is to give a methodology for verification where the only thing a designer needs to do is adding a specification of the safety requirements (as an @property annotation) to the implementation program. This contrasts with most verification methodologies, which require the designer to first construct a model of the implementation. Here, we work directly with the implementation, which is given in **LINGUA FRANCA**. The model of the implementation is generated automatically from the **LINGUA FRANCA** program, as is the low-level C code that realizes the application.

3 PRELIMINARIES

**Metric Temporal Logic.** Metric Temporal Logic (MTL) [41], a real-time extension of Linear Temporal Logic (LTL), has been a successful logic for specifying properties of reactive, embedded systems. The syntax and semantics of MTL are given below. We cite the definitions of MTL from Ouaknine and Worrell [59, 60]. The syntax and semantics of MTL are given as follows.

**Definition 3.1 (Syntax of MTL).** Given an alphabet Σ of events, the formulas of MTL are built up from Σ by Boolean connectives and time-constrained versions of the next operator X, the until
operator $U$, and the dual until operator $\tilde{U}_I$ as follows:

$$\psi ::= \text{true} \mid \text{false} \mid a \mid \psi_1 \land \psi_2 \mid \psi_1 \lor \psi_2 \mid X_I \psi \mid \psi_1 U_I \psi_2 \mid \psi_1 \tilde{U}_I \psi_2$$

where $a \in \Sigma$, and $I \subseteq \mathbb{R}_{\geq 0}$ is an open, closed, or half-open interval with end points in $\mathbb{R}_{\geq 0} \cup \{\infty\}$.

**Definition 3.2 (Semantics of MTL).** Given a (finite or infinite) time word $\pi = (\sigma, \tau)$ over alphabet $\Sigma$, a word position $i < |\pi|$, and an MTL formula $\psi$, the satisfaction relation $(\pi, i) \models \psi$ (read $\pi$ satisfies $\psi$ at position $i$) is defined as follows.

$$(\pi, i) \models a \iff \sigma_i = a$$

$$(\pi, i) \models \psi_1 \land \psi_2 \iff (\pi, i) \models \psi_1 \land (\pi, i) \models \psi_2$$

$$(\pi, i) \models \psi_1 \lor \psi_2 \iff (\pi, i) \models \psi_1 \lor (\pi, i) \models \psi_2$$

$$(\pi, i) \models X_I \psi \iff i < |\pi| \land \tau_{i+1} - \tau_i \in I \land (\pi, i + 1) \models \psi$$

$$(\pi, i) \models \psi_1 U_I \psi_2 \iff \exists j, i \leq j \leq |\pi| \land \tau_j - \tau_i \in I. [(\pi, j) \models \psi_2 \land \forall k, i \leq k < j.(\pi, k) \models \psi_1]$$

$$(\pi, i) \models \psi_1 \tilde{U}_I \psi_2 \iff \forall j, i \leq j \leq |\pi| \land \tau_j - \tau_i \in I. [(\pi, j) \models \psi_2 \lor \exists k, i \leq k < j.(\pi, k) \models \psi_1]$$

From here, additional temporal operators can be defined: constrained eventually $F_I$ and constrained globally $G_I$, where $F_I \psi = \text{true} U_I \psi$, and $G_I \psi = \text{false} \tilde{U}_I \psi$. In this work, we focus on a fragment of MTL called Safety MTL $[59, 61, 62]$, which requires the interval $I$ of the until operator $U_I$ to have finite length. Both the satisfiability problem and the model-checking problem for Safety MTL over infinite timed words are decidable. Within Safety MTL, in this work we additionally constrain the interval $I$ of the next operator $X_I$ and the dual until operator $\tilde{U}_I$ to have finite length. We leave the relaxation of this constraint for future work.

**Bounded Model Checking (BMC).** BMC $[12]$ is a well-known technique for bug finding and bounded verification. A model checker implementing BMC usually proceeds as follows: It first checks a property $\psi$ is valid along a path $\pi$ for some $k$ steps (denoted as $\pi \models_k \psi$). If a satisfying assignment is found at this point, the counterexample is returned to the user. If not, the model checker increments $k$ and checks $\pi \models_{k+1} \psi$. The procedure continues until checking $k$ steps becomes intractable, or a completeness threshold is reached. A completeness threshold $[42]$, denoted as $CT$, is a minimum number of $k$ such that $\pi \models_{CT} \psi \implies \pi \models \psi$, i.e. if $\pi$ satisfies $\psi$ up to $CT$ steps, then $\pi$ satisfies the property $\psi$. In Section 5, we present an algorithm for upper-bounding $CT$.

## 4 AXIOMATIC SEMANTICS FOR LF FOR VERIFICATION

Figure 2 shows the verification workflow of LF programs and the architecture of the LF Verifier. In this section, we zoom in on the “UCLID5 model” in the figure and present the method of performing BMC using axiomatic models derived from the semantics of Lingua Franca. We begin by describing the typical properties of interest in a system of reactors.

**Axiomatic semantics.** In this section, we give an axiomatic semantics for reactors and use the semantics to construct a model in UCLID5 to solve the verification problem defined below.

**Definition 4.1 (Verification Problem).** Given a reactor, denoted as $r$, and a Safety MTL formula $\psi$, determine whether all execution paths $\pi$ of $r$ satisfy the formula $\psi$, i.e., $\pi \models \psi$. 

Fig. 2. Verification workflow of LF programs and the architecture of the LF Verifier. Blue represents software modules implemented in this work. The other matching colors indicate correspondence between the input program and the generated model.

In the proposed axiomatic semantics, a reaction invocation defines a transition, which matches the event-based semantics in Sirjani et al. [71]. To show the correspondence between the LF semantics in prose and our axiomatic semantics, we will constantly refer to parts of the LF semantics in prose from prior works [50, 52, 53] as we present the axiomatic semantics. The proposed semantics include four segments: (1) logical time and path; (2) reactor semantics; (3) connections and triggers; (4) reaction code. This section only includes the core axioms used to construct the axiomatic model.

(1) Notational conventions. We use uppercase letters to denote sets defined in LF (e.g. \(N\) is the set of reactions in an LF program) and italicized words to denote sets in an axiomatic model (e.g. \(\text{rxns}\) is the set of reactions in a corresponding axiomatic model). A lowercase letter is always used to denote an element of a set (e.g. \(n \in N\)). For functions, we use the calligraphic font (e.g. \(G\)) and italicized words (e.g. \(\text{val}\)). In addition, we use the blackboard-bold font to denote common sets (e.g. \(\mathbb{N}\) for integers and \(\mathbb{B}\) for \(\{\text{true}, \text{false}\}\)).

(2) Logical time and path. The definitions covered in this segment are shown in Figure 3. A logical time tag is defined as \(\text{tag} := (t, m)\), which is a tuple of two natural numbers: \(t \in \mathbb{N}\) to represent a timestamp (i.e., the number of nanoseconds elapsed in logical time since the start of the program), and \(m \in \mathbb{N}\) to represent a microstep index, which is used to enforce an ordering on events with the...
Towards Building Verifiable CPS using Lingua Franca

same timestamp. To extract a timestamp or a microstep from a tag, we define projection functions $T : \text{tags} \rightarrow \mathbb{N}$ for mapping a time tag to its time value and $M : \text{tags} \rightarrow \mathbb{N}$ for mapping a time tag to its microstep index. To compare two tags, we define an ordering relation

$$< = \{(\text{tag}_1, \text{tag}_2) \in \text{tags}^2 | T(\text{tag}_1) < T(\text{tag}_2)$$

$$\lor (T(\text{tag}_1) = T(\text{tag}_2) \land M(\text{tag}_1) < M(\text{tag}_2))\}$$

and a relation $\leq = \{(\text{tag}_1, \text{tag}_2) \in \text{tags}^2 | \text{tag}_1 < \text{tag}_2 \lor \text{tag}_1 = \text{tag}_2\}$. We use $R$ to denote the set of reactors and $N$ the set of reactions defined in an LF program. In the axiomatic model, a reaction invocation is defined as $\text{rxn} \in \text{rxns} = \mathbb{N} \cup \{e\}$, where $e$ indicates an empty reaction. Let $S$ denote the union of state variables, ports, and actions defined in the LF program. A vector of values carried by elements in $S$ is defined as $\text{val} \in \text{vals} = \mathbb{V}^{\mid S\mid}$, where $\mathbb{V}$ is an abstract set of values. Let $T$ denote the union of ports and actions defined in the LF program. A boolean vector denoting the presence of triggers is defined as $\text{trig} \in \text{trigs} = \mathbb{B}^{\mid T\mid}$. We further let $A \subseteq T$ denote the set of actions specifically. A boolean vector denoting whether actions are scheduled is defined as $\text{sch} \in \text{schs} = \mathbb{B}^{\mid A\mid}$. This is different than the same actions in $\text{trigs}$, which are set to true when the scheduled actions actually become present at a future time tag. When actions are scheduled, their payloads are stored in $\text{pld} \in \text{plds} = \mathbb{B}^{\mid A\mid}$.

A path $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ is a bounded sequence of states, where the length of the sequence is equal to a bound $k$. This is modeled in SMT using the theory of arrays with the constraint that any element outside of the range $[1, k]$ carries an empty value. We also define an indices set as a sequence of natural numbers from 1 to the bound $k$, i.e., $\text{indices} = \{1, 2, \ldots, k\}$. Intuitively, a path can be interpreted as an unrolled execution trace over which the axioms will be enforced. To ensure that certain axioms, to be introduced below, also hold at the boundary of the path (i.e., at step $k$), we define an extended version of the path with a padding, where $\text{indices} = \text{indices} \cup (k+1, \ldots, k + |N|)$, $\bar{\pi} = \pi \cup (\pi_{k+1}, \ldots, \pi_{k+|N|}) \in \Pi$, and the length of the padding is set to $|N|$. Note that Safety MTL formulas are evaluated on $\bar{\pi}$ and not on $\pi$, the padded version, which is used here to prevent the axioms from being trivially unsatisfiable at the path boundary.

A state $s \in \text{states}$ is a 6-tuple $(\text{rxn}, \text{tag}, \text{val}, \text{trig}, \text{sch}, \text{pld})$, where $\text{rxn} \in \text{rxns}$ is the reaction invoked, $\text{tag} \in \text{tags}$ is the logical time at which the invocation occurs, $\text{val} \in \text{vals}$ is the variable values after the invocation, $\text{trig} \in \text{trigs}$ the presence of triggers at the moment of the invocation, $\text{sch} \in \text{schs}$ indicates whether actions are scheduled during the current reaction invocation, and $\text{pld} \in \text{plds}$ the payloads actions carry when they are scheduled.

We use subscripts to denote an individual elements in a vector. The $i^{th}$ entry of $\text{val}$ is denoted as $\text{val}_i$, and the $i^{th}$ entry of $\text{trig}$ is denoted by $\text{trig}_i$. We use the same symbols and define projection functions $\text{val}_i : \text{states} \rightarrow \mathbb{V}$ and $\text{trig}_i : \text{states} \rightarrow \mathbb{B}$, to extract the $i^{th}$ value or trigger at a particular state in a path. We define a projection function $N : \text{states} \rightarrow N$, which extracts a $\text{rxn}$ from a state, and a projection function $G : \text{states} \rightarrow \text{tags}$, which extracts a $\text{tag}$ from a state. We are now ready to start axiomatizing the core reactor semantics.

(3) Reactor semantics. One of the most important sets of axioms is the semantics of the reactor model. We first define an axiom enforcing that the ordering of reaction invocations must be in the order of time tags:

$$\forall i, j \in \text{indices}. \ G(\pi_i) < G(\pi_j) \implies i < j.$$  

(TimeOrder)

where $<$ is the ordering relation defined earlier. At a time tag, a reactor maintains a deterministic order of execution by executing its reactions in the order of their priorities.

$$\forall i, j \in \text{indices}. \ G(\pi_i) = G(\pi_j) \land \Psi(N(\pi_i)) < \Psi(N(\pi_j)) \implies i < j.$$  

(ReactionOrder)
where \( \Psi : N \to \mathbb{N} \) is a function that maps a reaction to a priority value, which is a natural number. Besides enforcing a tag order, we also require that a given reaction can only be invoked once at any logical time instant:

\[
\forall i, j \in \text{indices}. \ (N(\pi_i) = N(\pi_j) \land i \neq j) \implies \mathcal{G}(\pi_i) \neq \mathcal{G}(\pi_j).
\]  

(Uniqueness)

Next, we place constraints on the time tags, where both the timestamp and the microstep need to be non-negative.

\[
\forall i \in \text{indices}. \ \mathcal{T}(\mathcal{G}(\pi_i)) \geq 0 \land \mathcal{M}(\mathcal{G}(\pi_i)) \geq 0.
\]  

(ValidTag)

For an empty invocation \( \epsilon \), we require that it must always occur at the end of the trace. In other words, \( \epsilon \) invocations cannot occur in between real reaction invocations in a trace. This desired behavior as the purpose of \( \epsilon \) events is to fill in unused entries in the trace array when modeling an execution trace that is shorter than the verification bound.

\[
\forall j \in \text{indices}. \ N(\pi_j) \neq \epsilon \implies (\forall i \in \text{indices}, i < j. \ N(\pi_i) \neq \epsilon)
\]  

(EmptyLast)

(4) Connections and triggers. Since we represent the system axiomatically, we do not explicitly specify the transition relation between each state variable, but instead specify a set of invariants over the traces being examined.

Let \( p_u \) and \( p_d \) be a pair of upstream and downstream ports, with their values stored in \( \text{val}_u, \text{val}_d \in \text{val} \) and their presence in \( \text{trig}_u, \text{trig}_d \in \text{trig} \). We use a variable \( \text{physical} \in \mathbb{B} \) to denote whether the connection is a physical connection, and we define a variable \( \text{delay} \in \mathbb{N} \) to indicate the logical delay carried by the connection. We also define a schedule operator \( \oplus : \text{tags} \times \mathbb{N} \to \text{tags} \) as \( g \oplus t = g' \) s.t. \( \mathcal{T}(g') = \mathcal{T}(g) + t \) and \( \mathcal{M}(g') = 0 \) if \( t > 0 \) and \( \mathcal{M}(g') = \mathcal{M}(g) + 1 \) if \( t = 0 \). The behavior of a connection is specified as follows.

\[
\forall i \in \text{indices}. \ \text{trig}_u(\pi_i) \implies (\exists j \in \text{indices}. j > i \land \text{trig}_d(\pi_j) \\
\land \text{val}_d(\pi_j) = \text{val}_u(\pi_i) \land (\neg \text{physical} \implies \mathcal{G}(\pi_i) = \mathcal{G}(\pi_i) \oplus \text{delay})))
\]  

(SendToDownstream)

The SendToDownstream axiom specifies that when the upstream trigger is present at some step in the path, there either exists a step in the path with the downstream trigger present. We further specify the Correspondence axiom below to ensure a one-to-one correspondence between the presence of triggers in logical time.

\[
\forall i \in \text{indices}. \ \text{trig}_d(\pi_i) \implies (\exists j \in \text{indices}. j < i \land \text{trig}_u(\pi_j) \\
\land \text{val}_u(\pi_j) = \text{val}_d(\pi_i) \land (\neg \text{physical} \implies \mathcal{G}(\pi_i) = \mathcal{G}(\pi_j) \oplus \text{delay})))
\]  

(Correspondence)

Since the default behavior of variables in an axiomatic model is nondeterministic assignment, the SameIfAbsent axiom states that when a trigger is absent, the value corresponding to the trigger stays the same as the previous step.

\[
\forall i \in \text{indices}. \ \text{trig}_d(\pi_i) = \text{false} \implies (\text{val}_d(\pi_i) = \text{val}_d(\pi_{i-1}))
\]  

(SameIfAbsent)

If a reaction \( v \in N \) is sensitive to a startup trigger, then it will be invoked at tag \((0,0)\). To encode this behavior, we state a StartupTrigger axiom.

\[
\exists i \in \text{indices}. \ N(\pi_i) = v \land \mathcal{G}(\pi_i) = (0,0) \land (\exists j \in \text{indices}. \ N(\pi_j) = v \land j \neq i)
\]  

(StartupTrigger)

Besides ports, LF also supports actions and timers as triggers. Let us first start with actions. To axiomatize the behavior of actions, we let \( a \) be an action with its value stored in \( \text{val}_a \in \text{vals} \), its presence stored in \( \text{trig}_a \in \text{trigs} \), and a logical delay \( \text{delay}_a \in \mathbb{N} \). We further let \( \text{reaction}_a \) denote a
set of reactions that can schedule the action \( a \). We can state an axiom, \( \text{ScheduleAction} \), to encode the semantics of actions.

\[
\exists i \in \text{indices}. \bigwedge_{n \in \text{reactions}_a} \text{trig}_a(\pi_i) = \text{true} \implies (\exists j \in \text{indices}. j < i \land N(\pi_j) = n) \\
\land G(\pi_i) = G(\pi_j) \land \text{delay}_a \land \text{sch}_a(\pi_j) = \text{true} \land \text{val}_a(\pi_i) = \text{pld}_a(\pi_j) \quad (\text{ScheduleAction})
\]

Timers are another important feature for generating periodic events. To axiomatize their behavior, let \( t \) be a timer with an initial offset \( \text{offset} \in \mathbb{N} \), a period \( \text{period} \in \mathbb{N} \), and its presence stored in \( \text{trig}_t \in \text{trigs} \). Recall also that \( k \) is the largest index in \( \text{indices} \). Then, a timer must start firing at the pre-defined offset, as stated in \( \text{TimerOffset} \).

\[
((T(G(\pi_k)) \geq \text{offset}) \implies \exists i \in \text{indices}. \text{trig}_t(\pi_i) = \text{true} \land G(\pi_i) = (\text{offset}, 0))
\land ((T(G(\pi_k)) < \text{offset}) \implies \forall j \in \text{indices}. N(\pi_j) \neq \epsilon) \quad (\text{TimerOffset})
\]

Intuitively, if the last state of \( \pi \) (i.e., \( \pi_k \)) has reached the timestamp specified by \( \text{offset} \), then there must be a state in which the timer trigger is present and the timestamp matches \( \text{offset} \). Otherwise, if the last state does not reach the timestamp specified by \( \text{offset} \), all the reaction invocations along the path must be non-empty. And the subsequent firing should occur repeatedly at an interval specified by the period. We state two axioms, \( \text{TimerPeriod} \) and \( \text{TimerSpacing} \) to encode this behavior.

\[
\forall i \in \text{indices}. \text{trig}_t(\pi_i) = \text{true} \implies (\exists j \in \text{indices}. j > i \land \text{trig}_t(\pi_j) = \text{true} \land G(\pi_j) = G(\pi_i) \land \text{period}) \quad (\text{TimerPeriod})
\forall i \in \text{indices}. \text{trig}_t(\pi_i) = \text{true} \implies ((\exists x \in \mathbb{N}. x \geq 0 \land (T(G(\pi_i)) = \text{offset} + x \times \text{period}) \land M(G(\pi_i)) = 0) \quad (\text{TimerSpacing})
\]

The LF semantics specifies that a reaction is triggered when any of its triggers is present. To encode this behavior, we state a \( \text{ReactionTriggers} \) axiom. Let \( n \in \mathbb{N} \) be a reaction and let \( \text{triggers}_n \) be the set of triggers the reaction \( n \) is sensitive to.

\[
\forall i \in \text{indices}. (\exists t \in \text{triggers}_n. t = \text{true}) \iff N(\pi_i) = n \quad (\text{ReactionTriggers})
\]

(5) Reaction code. To axiomatize reaction body code written in C, we adopt the preprocessing method presented by Armando et al. [4], in which a block of C code is transformed into a representation called If Normal Form. Starting from a C abstract syntax tree (AST) in If Normal Form, we then translate each if block into an implication where the if condition becomes the antecedent and the then branch becomes the consequent. Since axiomatizing C code is not our contribution, we will not expand this here but will later revisit it in Section 7.

5 BOUNDING THE COMPLETENESS THRESHOLD FOR BMC

Recall that the completeness threshold, denoted as \( C^T \), is an attribute of a system \( M \) with respect to a property \( p \) and a translation scheme such that checking a trace up to \( C^T \) steps without errors proves \( M \models p \) [13]. In the context of our work, the system \( M \) is a reactor system. The property \( p \) is a Safety MTL formula. The translation scheme is the translation of a reactor system to an axiomatic model (Section 4) and the translation of a Safety MTL formula to a First-Order Logic formula over a timed word \( \pi \) based on the semantics of MTL (Definition 3.2).

Here, we do not seek to compute the precise completeness threshold since such computation is as hard as the BMC problem itself [43]. Instead, we efficiently compute an upper bound on the completeness threshold, and we denote this bound as \( C^T \). Since we adopt the event-based semantics...
introduced by Sirjani et al. [71] (i.e., modeling an LF execution as a sequence of reaction invocations), computing a bound for $\mathcal{CT}$ requires finding the maximum number of reaction invocations within a logical time interval specified by the Safety MTL property. To do this, we construct a state space diagram for an LF program.

**State space diagram.** Intuitively, each node in the state space diagram consists of a time tag, a set of reactions invoked at this tag, and a set of pending events set to appear at some future time tags. The state space diagram of the running example is shown in Figure 4.

Formally, a state space diagram is a directed graph $(V, E)$ where $V \subseteq \text{tags} \times \mathcal{P}(\text{rxns}) \times \mathcal{P}(\text{events})$ and $E \subseteq V \times V$, where $\mathcal{P}$ denotes a power set. The symbols tags and rxns are defined previously in Figure 3. Here we introduce a new set events, which is defined as $\text{events} = \text{tags} \times \text{trigs}$, where trigs is also defined in Figure 3. To extract a tag, a set of reactions, or a set of events from a node, we define projection functions $\text{tag}_n : V \rightarrow \mathcal{P}(\text{events})$, $\text{rxns} : V \rightarrow \mathcal{P}(\text{rxns})$, and $\text{events} : V \rightarrow \mathcal{P}(\text{events})$. Similarly, to extract a tag or a trigger from an event, we define projection functions $\text{tag}_e : \text{events} \rightarrow \mathcal{P}(\text{events})$ and $\text{trigs} : \text{events} \rightarrow \mathcal{P}(\text{rxns})$.

Constructing a state space diagram requires a light-weight simulation of the worst-case execution until some time horizon, denoted as $\text{horizon} \in \mathbb{N}$ is reached, which is the time interval specified by a Safety MTL formula. We provide a precise definition of a theoretical worst-case execution below.

**Definition 5.1 (Theoretical Worst-case Execution of an LF Program).** A theoretical worst-case execution of an LF program is a execution path $\pi$ permitted by the LF program structure such that

$$\forall \pi' \in \Pi. \forall g \in \text{tags}. |\{\pi_i \in \pi \mid G(\pi_i) = g \land N(\pi_i) \neq \epsilon\}| \geq |\{\pi'_i \in \pi' \mid G(\pi'_i) = g \land N(\pi'_i) \neq \epsilon\}|$$

Intuitively, the theoretical worst-case execution path has the most possible number of reaction invocations at each logical tag among all possible execution paths defined by the LF program structure.

During the simulation, the algorithm keeps track of three quantities (as defined earlier): a current time tag (denoted as currentTag), a list of reactions invoked at the current tag (denoted as reactionInvoked), and a priority queue of pending events ordered by time tags (denoted as pendingEvents). The algorithm is listed in Algorithm 1 with specific implementation details abstracted away.

The main exploration logic happens inside a while loop 5. Before the actual exploration begins, the algorithm requires a horizon argument and identifying a set of initial events from which the execution unfolds. The horizon argument is derived based on the given MTL formula as follows. Let $\psi$ be an MTL formula in Negation Normal Form, and let $h : \text{MTL} \rightarrow \mathbb{R}$ be a function mapping an MTL formula to its horizon. The function $h$ can then be defined recursively.

$$h(\psi) = \begin{cases} 
\max(h(\psi_1), h(\psi_2)) & \text{if } \psi \equiv \psi_1 \land \psi_2 \text{ or } \psi \equiv \psi_1 \lor \psi_2 \\
\max(I) + h(\psi_1) & \text{if } \psi \equiv G_I \psi_1 \text{ or } \psi \equiv F_I \psi_1 \text{ or } \psi \equiv X_I \psi_1 \\
\max(I) + \max(h(\psi_1), h(\psi_2)) & \text{if } \psi \equiv \psi_1 U_I \psi_2 \text{ or } \psi \equiv \psi_1 \bar{U}_I \psi_2 \\
0 & \text{otherwise}
\end{cases}$$

To determine the set of initial events, the algorithm identifies the startup trigger used in the program and the initial firings of timers and add them to pendingEvents.

Then, the procedure starts simulating the execution by stepping into a while loop, which can only be exited when the following stop conditions are reached:

1. there are no more pending events (i.e., $|\text{pendingEvents}| = 0$);
2. the simulation completes the time horizon (i.e., $\mathcal{T}(\text{currentTag}) > \text{horizon}$);
3. a cycle is found in the state space diagram.
During each cycle iteration, the simulation pops all the pending events that are set to appear at the current tag off the $\text{pendingEvents}$ queue. Then, reactions triggered by these events are added to the $\text{reactionsInvoked}$ list. Note that since this is a lightweight simulation, the C bodies of these reactions will not be simulated, but instead, they are assumed to produce outputs at all the ports and schedule all the actions available to them. This is an over-approximation of the actual behaviour, and leads to the worst-case execution we seek. The set of new outputs and actions produced are inserted back to $\text{pendingEvents}$, waiting to be processed at some future loop iterations. When all the events at the current tag have been processed, the execution advances to the nearest future tag. A new node in the state space diagram is then created for the previous tag, storing the $\text{reactionsInvoked}$ list and the $\text{pendingEvents}$ queue. There is also an edge from the previous node to this new node.

We prove that the state space diagram generated using this algorithm is indeed the worst case.

**Proposition 5.2.** Algorithm 1 produces a state space diagram representing the theoretical worst-case execution of an LF program.
Proof. We prove this lemma by induction on the logical tag. In the base case, we focus on the first time tag and show that the algorithm computes the theoretical worst-case execution path up to the conclusion of the first time tag. When the algorithm begins, it adds initial events (line 3) including the startup trigger (set to occur at tag $(0, 0)$) and the initial firings of timers (could happen at any time tags depending on the offsets) to pendingEvents. At this point, these initial events are not dependent on any prior state, but their addition to pendingEvents ensures that their effects will be considered when computing the next state, thus establishing state-dependency from the very beginning. If there are no initial events, the algorithm terminates and we arrive at a correct state space diagram that is empty. If there are initial events, then we enter the while loop (line 5) and pop off the earliest events at line 6. Reactions triggered by these events are added to reactionsInvoked and all of their effects are added back to pendingEvents. This means that the effects of all the reactions invoked at the current state are taken into account when constructing the next state. These effects can influence both the state at the next time tag and the states at subsequent time tags, which ensures the propagation of state-dependency across different time tags. If the newly added effects happen at some future tag, then we advance the current tag (line 24) and are done with the first tag, leaving us with a correct state space diagram constructed up to the conclusion of the first tag with as many reactions invoked as possible at the first tag. On the other hand, if the added events are instantaneous, e.g., setting output ports connected to a downstream input port with zero delay, then line 24 maintains the current tag, runs the loop iteration until we are done with the first tag. The resulting state space diagram is representative of the theoretical worst-case execution because during the loop iterations, all the possible reaction invocations have been accounted for. We thus conclude the base case.

For the inductive step, we make the inductive hypothesis that the algorithm has correctly computed a state space diagram up to some time tag later than the first tag, we seek to show that the algorithm can correctly compute the next tag. At this point, there must be unprocessed events in pendingEvents and the current tag must not have reached horizon, otherwise the algorithm would have been terminated at line 26. By our inductive hypothesis, pendingEvents contains the effects of all the reactions invoked at the previous states. Hence, the current state is dependent on all the previous states, and any reaction invoked at the current state will also take into account the effects of the reactions invoked at the previous states. We begin handling the next tag by popping the earliest events, checking the set of reactions invoked by these popped events, and again add all the effects of the invoked reactions back to pendingEvents. By adding all the effects of the reactions invoked at the current state to pendingEvents, we ensure that these effects are propagated to the next state and states at subsequent time tags. This maintains state-dependency across different time tags. If the newly added events are in the future, then we update currentTag (line 24) and are done with the next tag. If not, line 24 does not update currentTag, and we run the loop iteration until the next tag is finished, at which point, we are left with a state space diagram correctly constructed up to the conclusion of this next tag since all the possible reaction invocations have been accounted for. Invoking all possible reactions at each time tag results in the maximal number of reactions invoked over the entire execution, which constitutes the worst-case scenario. □

Cycle detection. As an optimization, the simulation seeks to find a cycle in a state space diagram. LF programs often exhibit periodic behaviors (e.g., the execution of our running example shown in Figure 4). For such programs, the simulated execution does not need to simulate until the completion of a horizon but only needs to simulate until the completion of a period (i.e., a cycle). The rest of the execution up to the horizon can be inferred from the periodic behavior. To check whether an execution completes a period, when the execution advances to a new tag, the
Towards Building Verifiable CPS using Lingua Franca

Fig. 4. An abstract view of the statespace diagram for the running example in Figure 1. In this abstract view, except for S0, which is shown in full, each node consists of (from left to right) an identifier, the number of reactions invoked at the current tag, the number of events pending, and (at the bottom) the current tag. Simulation checks whether the newly added node is equivalent to any of the previous nodes. Two nodes, $n_1$ and $n_2$, are considered equivalent if

1. $\text{rxns}(n_1) = \text{rxns}(n_2)$;
2. $\{\text{trigs}(e_1) \subseteq \text{trigs} | e_1 \in \text{events}(n_1)\} = \{\text{trigs}(e_2) \subseteq \text{trigs} | e_2 \in \text{events}(n_2)\}$;
3. $\{\text{tag}_e(e_1) - \text{tag}_n(n_1) | e_1 \in \text{events}(n_1)\} = \{\text{tag}_e(e_2) - \text{tag}_n(n_2) | e_2 \in \text{events}(n_2)\}$.

The cycle detection approach is similar to [30], which constructs state space diagrams for synchronous dataflow graphs [46].

From diagram to $\hat{CT}$. Once the state space diagram is generated, we can then compute $\hat{CT}$ from it. For this, we state two theorems.

**Theorem 5.3 ($\hat{CT}$ w.r.t. a cycle-free diagram).** The completeness threshold w.r.t. an LF program that generates a cycle-free state space diagram $(V, E)$ is upper bounded by $\hat{CT}$ such that

$$\hat{CT} = \sum_{n \in V} |\text{rxns}(n)|.$$

**Proof.** A cycle-free diagram is by construction a finite linear sequence of nodes. For a node $n \in V$, the number of reaction invocations is given by $|\text{rxns}(n)|$. The maximum possible number of reaction invocations, i.e., $\hat{CT}$, can be computed by summing up the number of reaction invocations of each node. We then obtain the desired expression. \[\square\]

Intuitively, this $\hat{CT}$ counts the number of reaction invocations up to the completion of the horizon. If a cycle is found, we use its periodicity to infer $\hat{CT}$. 

Theorem 5.4 (\(\hat{C}T\) w.r.t. a Diagram with a Cycle). The completeness threshold w.r.t. an LF program that generates a state space diagram \((V, E)\) with a cycle is upper bounded by \(\hat{C}T\) such that

\[
\hat{C}T = \sum_{n'' \in V'' \cup \{n_c\}} |\text{rxns}(n'')| + \sum_{n' \in V'} |\text{rxns}(n')| \ast \text{iterations},
\]

where \(V'\) denotes the set of nodes that form the cycle, \(V'' = V \setminus V'\) denotes the rest of the nodes, \(n_c \in V'\) denotes the node where the cycle begins, \(\text{iterations} = \lceil (\text{horizon} - T(tag_n(n_c))) / \rho \rceil\), and \(\rho\) denotes the cycle period.

Proof. A state space diagram with a cycle has an initialization phase, consisting of nodes \(V'' \cup \{n_c\}\), and a periodic phase, consisting of nodes \(V'\). Note that the first time of reaching \(n_c \in V'\) counts toward the initialization phase since the periodic phase cannot start without first reaching \(n_c\). The overall \(\hat{C}T\) is the sum of reaction invocations during both the initialization phase and the periodic phase. The first expression, \(\sum_{n'' \in V'' \cup \{n_c\}} |\text{rxns}(n'')|\), follows directly from the proof of Theorem 5.3. The second expression over-approximates the number of reaction invocations during the periodic phase by multiplying the number of invocations during one iteration, given by \(\sum_{n' \in V'} |\text{rxns}(n')|\) (also directly following the proof of Theorem 5.3), and, \(\text{iterations}\), the minimum number of cycle iterations required to guarantee completing the time horizon. The value of \(\text{iterations}\) can be obtained by taking the remaining horizon after entering the periodic phase, i.e., \(\text{horizon} - T(tag_n(n_c))\), dividing it by the period of the cycle, \(\rho\), and applying the ceiling operation. We then obtain the desired expression. \(\square\)

6 IMPLEMENTATION

We now describe the implementation details of the LF Verifier, and walk through the steps to verify the safety property in the running example. The LF Verifier is embedded inside the Lingua Franca compiler, which is open-source on Github.\(^1\) From the user’s perspective, using the LF Verifier simply requires extending the original program with \@property annotations, and the rest is invoking the compiler.

Recall from Figure 2 that the user first provides an LF program along with annotations specifying temporal properties of the program in Safety MTL. Once the input files are provided, they are then passed into the LF Verifier, which parses LF code, reaction code (written in C, in our case), and the Safety MTL annotations. Based on the parsed LF code, LF Verifier analyzes the topology of the system and generates an axiomatic model that can be processed by Uclid5. In the rest of the section, we will focus on the modules used by the LF Verifier to verify properties and ignore the modules used for compiling the LF program.

Annotations. The user specifies an MTL property using an \@property annotation. At the time of writing, the \@property annotation requires three parameters: (1) the name of the property, (2) the tactic used to check the property (only bmc is supported for now), and (3) the MTL formula. An valid MTL formula follows the syntax shown in Figure 5. Safety MTL is enforced by ensuring “int” to be finite non-negative integers. Reaction invocations (resp., values of LF components) are specified using fully qualified names of the reactions (resp., the components).

Parsing. After opening the LF file provided by the user, the LF Verifier parses the LF code using an “LF Parser” based on the Xtext framework \[11, 25\], and it parses the reaction body code (in C) and the MTL annotations using ANTLR4 parsers \[63\] (shown in Figure 2 as “C Parser” and “MTL Parser”), generating a parse tree for C and another parse tree for MTL.

\(^1\)GitHub link: https://github.com/lf-lang/lingua-franca
AST transformations. Once the C parse tree is generated by ANTLR4, the “C AST Builder” is used to build an abstract syntax tree (AST). Our implementation uses the technique presented by Armando et al. [4], which supports a subset of the C language. For language features we do not currently support, including pointers, loops, and macros, the AST builder marks these AST nodes as *opaque* and returns a warning to the user. While the support of C is incomplete, the LF Verifier currently can analyse arithmetic operations, variable assignments, LF runtime APIs, and if statements (including nesting if statements), which are arguably expressive enough to cover many real-time systems applications. We will discuss our plan to expand the support for C in Section 8.

Once the C AST is built, it is sent to the “INF Converter”, where INF stands for *If Normal Form*, a term introduced in Armando et al. [4]. The main idea is to flatten the nested if statements until an if condition is associated with each atomic statement. For example, the two reaction bodies in ADASProcessor in the running example (Figure 1) effectively turn into

```c
/* (C) Reaction body 1 */
if (true) lf_schedule(a, 0);
if (true) self->requestStop = 1;
/* (C) Reaction body 2 */
if (self->requestStop == 1) lf_set(out1, 1);
```

The latter reaction body does not change since it follows the If Normal Form before transformation.

Computing $\hat{CT}$. Let us come back to the LF code. After the LF program passes through the “LF Parser”, it goes into a “Dependency Analyzer,” which extracts the dependency relations between reactions. When the dependency analysis is done, the LF Verifier invokes a “$\hat{CT}$ Calculator” which executes Algorithm 1 in Section 5 to obtain an upper bound on the completeness threshold.

To check the property in the running example, the LF Verifier first computes a logical time horizon from the property, which, in this case, is 65 ms: the outer $G[0, 10 \text{ ms}]$ operator requires 10 ms while the inner $F[0, 55 \text{ ms}]$ operator requires another 55 ms. The $F[0]$ does not contribute to the horizon due to the interval of 0. Therefore, to ensure the completeness of BMC, the model needs to unroll until 60 ms is reached.

The state space diagram of the program (Figure 4) shows that within an interval of 65 ms, the program reaches six states (S0-S5). $\hat{CT}$, i.e., the maximum sum of reaction invocations along the six states, is $\hat{CT} = 3 + 2 + 3 + 2 + 2 + 1 = 13$. The “$\hat{CT}$ Calculator” passes this number to the downstream model generators.
Model generation. During the model generation phase, three generators work together to produce a single Uclid5 model: an "LF Axiom Generator", a "Reaction Axiom Generator", and an "MTL2FOL Generator." The "LF Axiom Generator" generates the definitions and axioms (except the reaction axioms) presented in Section 4. Here is a snippet of the generated Uclid5 code for the preliminary definitions. In this paper, we do not explain the Uclid5 syntax in detail, but instead refer the reader to [64, 70, 72].

```c
/* (Uclid5) Preliminary definitions */
const START : integer = 0; // The start index of the trace
const END : integer = 13; // The end index of the trace (ie, CT bound)
const END_TRACE : integer = 19; // The end index of the trace with padding
type state_t = {
    rxn_t, tag_t, val_t,
    trig_t, sch_t, pld_t}; // A state is a 6-tuple.
type path_t = [integer]state_t; // A path is a sequence of states indexed by integers.
```

The "Reaction Axiom Generator" generates Uclid5 axioms from the C AST in the If Normal Form. Again, we use the two reaction bodies in ADASProcessor as examples:

```c
/* (Uclid5) Axioms of reaction 1 of ADASProcessor */
axiom (finite_forall (i : integer) in indices :: (i > START && i <= END) ==> (ADASModel_p_reaction_0(rxn(i))) ==> (((true ==> ((finite_exists (j : integer) in indices :: (j > i && j <= END_TRACE) &&
&& tag_same(g(j), tag_schedule(g(i), (50000000+0)))
&& ADASModel_p_a(s(j)) == 0)) // Closes finite_exists
&& ADASModel_p_a_scheduled(d(i)))))
&& (true ==> ((ADASModel_p_requestStop(s(i))) == 1)))))
```

As mentioned in Section 4, the generated reaction axioms are derived from a C AST in the If Normal Form and work in the same way as the other axioms by specifying relationships between variables across multiple time steps. For example, line 16 of the listing above is a direct translation of line 5 (i.e., if (self->requestStop==1) lf_set(out1, 1);) in the C code listing above. The if-statement becomes an implication where the if condition is the antecedent and the then branch of the if statement becomes the consequent.

The “MTL2FOL Generator” generates a First-Order Logic (FOL) property over the path in Uclid5 from the given MTL property. For the property in the running example, the generator produces the following:

```c
/* (Uclid5) The FOL property translated from user-defined MTL property */
define p(i : step_t) : boolean = ((((finite_forall (j0 : integer) in indices ::
    (j0 > i && j0 <= END && !isNULL(j0))
    ((pi1(g(j0)) >= (pi1(g(i)) + 0)) && (pi1(g(j0)) <= (pi1(g(i)) + 10000000))))
  ))) ==> (((((ADASModel_l_reaction_0(rxn(j0)))
    && (ADASModel_p_requestStop(s(j))) == 1) ==> ((ADASModel_p_out1(s(j))) == 1) && (ADASModel_p_out1_is_present(t(j)))))))
```

The “MTL2FOL Generator” recursively traverses the subformulas of the given MTL formula and builds up an FOL formula in a bottom-up direction. The generated FOL formula implements the semantics of MTL shown in Section 3.

**Checking and reporting results.** Once the model is generated, it is then passed into UCLID5, which generates an SMT file and invokes the Z3 SMT solver [18]. If the result is UNSAT, then the temporal properties are valid w.r.t. the formal model of the system. If the result is SAT, a counterexample trace is returned by UCLID5, which can help guide the user to revise the LF program. The process iterates until the unsafe behaviors are eliminated from the program. For the property in the running example, the LF VERIFIER reports UNSAT with no counterexamples found, hence the program satisfies the MTL property.

On the other hand, there are a few locations in the program where bugs can be introduced to invalidate the property.

**Incorrect Functional Behavior.** The property could be invalidated if reactions have incorrect functional behavior, e.g., the second reaction of **ADASProcessor** (lines 48–51) produces an output to the wrong port.

```
48 reaction(a) -> out1, out2 =
49   if (self->requestStop==1)
50     lf_set(out2, 1);  // Original: lf_set(out1, 1);
51   =} deadline(20ms) {=...
```

**Incorrect Composition.** The property is also invalidated when the components are not correctly composed, e.g., when LiDAR is not connected to **ADASProcessor**.

```
73 // l.out -> v.in1    // Originally connected
74 c.out -> v.in2
75 v.out2 -> d.in
76 v.out1 -> b.inADAS after 5ms
77 p.out -> b.inPedal
```

**Incorrect Timing.** Timing behavior also plays a crucial role in determining the truth value of the property. The timer in LiDAR having a large enough offset, the logical action in **ADASProcessor** having a larger delay, or the connection between **ADASProcessor** and **Brakes** having a larger delay could invalidate the given property.

```
5 timer t(11ms,17ms)    // Original: timer t(0,17ms)
40 logical action a(51ms) // Original: logical action a(50ms)
76 v.out1 -> b.inADAS after 6ms  // Original: v.out1 -> b.inADAS after 5ms
```

If any of the bugs (or their combinations) occurs, the LF VERIFIER reports SAT and displays a counterexample path showing that the reaction in **Brakes** does not get triggered within 13 steps of the generated model. The user can then work through the counterexample to identify the specific bugs in the program.

7 EVALUATION

**Benchmarks.** We evaluate our approach against a set of 22 LF programs drawn from real-life applications [2, 17, 28, 33, 49, 53, 71, 75], most of which are CPS applications with real-time properties, as well as benchmarks for Erlang [22], Lustre [16, 33, 40], actors [36] and RTOSes [24]. Thirteen out of the 22 are safe examples and 9 are unsafe examples. The benchmark focuses on verifying
the three types of properties introduced in Section 4: State properties are named Type I, dataflow properties Type II, and mixed properties Type III. The benchmark suite is also open-source on GitHub.²

Results. Table 1 shows the results of evaluating our approach against the verification benchmark suite. The experiments are conducted on a personal laptop running macOS version 11.7 with a 2.3 GHz 8-Core Intel Core i9 and 16GB RAM. The versions of tools used are Uclid5 0.9.5 and Z3 4.8.8.

The LF Verifier correctly checks 21 out of 22 programs automatically. The one program the LF Verifier fails to check is TrafficLight because the C implementation used in one of the reaction bodies exceeds the subset of C currently supported by the LF Verifier at the time of writing. The results show that our tool is sound but incomplete. We expect a future version of LF Verifier supporting a larger subset of C will be able handle this program. We discuss our plan to achieve this in Section 8.

We find that the majority of the computation time is spent on model generation: the median generation time for the Uclid5 model is 2.93 seconds, and the median generation time for the SMT formulas from the Uclid5 model is 12.15 seconds. However, once the SMT models are generated, solving them is very fast: the median solve time by Z3 is 0.31 seconds. The median total time is 15.9 seconds. Figure 6 shows the distributions of our benchmark result. The total time scales almost linearly as the LF lines of code (LF LOC) and $CT^{\gamma}$ increase.

8 CONCLUSION AND FUTURE WORK

In this paper, we present a methodology for building verifiable Cyber-Physical Systems using Lingua Franca. We specify CPS applications in LF, which provide language primitives for time and concurrency, and automatically generate axiomatic models in Uclid5 directly from LF programs to perform BMC. The properties of interest in this work are safety properties specified as Safety MTL formulas. Our axiomatic approach is sound w.r.t. all valid LF runtime implementations, and to ensure completeness of BMC, we present an algorithmic method to compute an upper bound on the completeness threshold. We implement our automated verification workflow in the LF Verifier, and our evaluation shows that the LF Verifier can effectively check a suite of benchmark programs fully automatically.

For the future work of the LF Verifier, it is our plan to include larger programs (in terms of lines of LF code) in the benchmark suite and make the LF Verifier more scalable. To tame the explosion of complexity, we plan to explore two strategies: (i) detecting reactors in an LF program that do not affect the truth value of a given property using static analysis and remove them from the axiomatic model, and (ii) using assume-guarantee reasoning to support compositional verification based on the modular structure of LF. In addition, we plan to support more complex C reaction bodies by expanding our “C AST Builder” or by offloading C code to external frameworks such as CBMC [44]. We also aim to support unbounded verification by potentially leveraging temporal induction [5].

²GitHub link: https://github.com/lf-lang/lf-verifier-benchmarks
Table 1. Benchmark Results

<table>
<thead>
<tr>
<th>Program</th>
<th>Type</th>
<th>#Rxns</th>
<th>$C^T$</th>
<th>Safe?</th>
<th>LOC LF Uclid5</th>
<th>GenUclid</th>
<th>GenSMT</th>
<th>Z3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADASModel [75]</td>
<td>III</td>
<td>6</td>
<td>13</td>
<td>yes</td>
<td>78 875</td>
<td>3.35</td>
<td>17.72</td>
<td>1.72</td>
<td>22.79</td>
</tr>
<tr>
<td>AircraftDoor [53]</td>
<td>III</td>
<td>3</td>
<td>3</td>
<td>yes</td>
<td>37 402</td>
<td>3.59</td>
<td>2.98</td>
<td>0.02</td>
<td>6.59</td>
</tr>
<tr>
<td>Alarm [33]</td>
<td>III</td>
<td>2</td>
<td>2</td>
<td>no</td>
<td>24 338</td>
<td>2.93</td>
<td>2.43</td>
<td>0.01</td>
<td>5.37</td>
</tr>
<tr>
<td>CoopSchedule [24]</td>
<td>I</td>
<td>2</td>
<td>18</td>
<td>no</td>
<td>29 659</td>
<td>2.75</td>
<td>18.35</td>
<td>0.31</td>
<td>21.41</td>
</tr>
<tr>
<td>Elevator [16]</td>
<td>I1</td>
<td>13</td>
<td>33</td>
<td>yes</td>
<td>118 1899</td>
<td>3.09</td>
<td>322.05</td>
<td>1038.02</td>
<td>1363.16</td>
</tr>
<tr>
<td>Election [17]</td>
<td>I1</td>
<td>9</td>
<td>18</td>
<td>yes</td>
<td>73 1045</td>
<td>3.47</td>
<td>38.18</td>
<td>3.27</td>
<td>44.92</td>
</tr>
<tr>
<td>Election2 [17]</td>
<td>I1</td>
<td>6</td>
<td>9</td>
<td>no</td>
<td>61 700</td>
<td>3.55</td>
<td>8.66</td>
<td>0.17</td>
<td>12.38</td>
</tr>
<tr>
<td>Factorial [16]</td>
<td>III</td>
<td>3</td>
<td>33</td>
<td>yes</td>
<td>38 461</td>
<td>2.81</td>
<td>9.48</td>
<td>0.74</td>
<td>13.97</td>
</tr>
<tr>
<td>Fibonacci [16]</td>
<td>III</td>
<td>5</td>
<td>55</td>
<td>yes</td>
<td>48 649</td>
<td>3.73</td>
<td>17.73</td>
<td>13.15</td>
<td>34.61</td>
</tr>
<tr>
<td>PingPong [36]</td>
<td>II</td>
<td>4</td>
<td>16</td>
<td>no</td>
<td>39 510</td>
<td>3.69</td>
<td>9.73</td>
<td>0.14</td>
<td>13.56</td>
</tr>
<tr>
<td>Pipe [22, 40]</td>
<td>I1</td>
<td>5</td>
<td>24</td>
<td>no</td>
<td>54 2059</td>
<td>2.89</td>
<td>139.52</td>
<td>12.15</td>
<td>154.56</td>
</tr>
<tr>
<td>ProcessMsg [24]</td>
<td>II</td>
<td>4</td>
<td>19</td>
<td>no</td>
<td>33 487</td>
<td>3.75</td>
<td>9.48</td>
<td>0.74</td>
<td>13.97</td>
</tr>
<tr>
<td>ProcessSync [24]</td>
<td>I1</td>
<td>1</td>
<td>3</td>
<td>no</td>
<td>12 284</td>
<td>2.78</td>
<td>2.21</td>
<td>0.01</td>
<td>5.00</td>
</tr>
<tr>
<td>Railroad [2]</td>
<td>I2</td>
<td>10</td>
<td>20</td>
<td>no</td>
<td>115 2333</td>
<td>2.93</td>
<td>155.21</td>
<td>4.33</td>
<td>162.47</td>
</tr>
<tr>
<td>Ring [22]</td>
<td>I2</td>
<td>4</td>
<td>8</td>
<td>yes</td>
<td>39 897</td>
<td>3.68</td>
<td>12.15</td>
<td>0.07</td>
<td>15.9</td>
</tr>
<tr>
<td>RoadsideUnit [28]</td>
<td>I1</td>
<td>8</td>
<td>21</td>
<td>yes</td>
<td>70 1536</td>
<td>3.78</td>
<td>70.29</td>
<td>0.62</td>
<td>74.69</td>
</tr>
<tr>
<td>SafeSend [22]</td>
<td>II</td>
<td>4</td>
<td>4</td>
<td>yes</td>
<td>37 479</td>
<td>2.70</td>
<td>3.78</td>
<td>0.03</td>
<td>6.51</td>
</tr>
<tr>
<td>Subway [33]</td>
<td>II</td>
<td>10</td>
<td>13</td>
<td>no</td>
<td>70 1097</td>
<td>2.85</td>
<td>27.11</td>
<td>0.31</td>
<td>30.27</td>
</tr>
<tr>
<td>Thermostat [49]</td>
<td>II</td>
<td>1</td>
<td>10</td>
<td>yes</td>
<td>59 652</td>
<td>2.80</td>
<td>10.61</td>
<td>0.14</td>
<td>13.55</td>
</tr>
<tr>
<td>TrafficLight †</td>
<td>I1</td>
<td>4</td>
<td>14</td>
<td>no</td>
<td>69 –</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>TrainDoor [71]</td>
<td>I1</td>
<td>3</td>
<td>3</td>
<td>no</td>
<td>31 398</td>
<td>2.62</td>
<td>3.03</td>
<td>0.01</td>
<td>5.66</td>
</tr>
<tr>
<td>UnsafeSend [22]</td>
<td>II</td>
<td>4</td>
<td>4</td>
<td>no</td>
<td>38 478</td>
<td>2.84</td>
<td>3.66</td>
<td>0.02</td>
<td>6.52</td>
</tr>
</tbody>
</table>

† No result is produced because the C code in this program exceeds the analyzable subset.

Furthermore, we hope to support other LF features including physical actions, deadlines, modal models, etc. We leave these improvements for future work.

REFERENCES


Received 23 March 2023; revised 2 June 2023; accepted 13 July 2023