K-ST: A Formal Executable Semantics of the Structured Text Language for PLCs

Kun Wang ©, Jingyi Wang ©, Christopher M. Poskitt ©, Xiangxiang Chen ©, Jun Sun ©, and Peng Cheng ©

Abstract—Programmable Logic Controllers (PLCs) are responsible for automating process control in many industrial systems (e.g., in manufacturing and public infrastructure), and thus it is critical to ensure that they operate correctly and safely. The majority of PLCs are programmed in languages such as Structured Text (ST). However, a lack of formal semantics makes it difficult to ascertain the correctness of their translators and compilers, which vary from vendor-to-vendor. In this work, we develop K-ST, a formal executable semantics for ST in the K framework. Defined with respect to the IEC 61131-3 standard and PLC vendor manuals, K-ST is a high-level reference semantics that can be used to evaluate the correctness and consistency of different ST implementations. We validate K-ST by executing 567 ST programs extracted from GitHub and comparing the results against existing commercial compilers (i.e., CODESYS, CX-Programmer, and GX Works2). We then apply K-ST to validate the implementation of the open source OpenPLC platform, comparing the executions of several test programs to uncover five bugs and nine functional defects in the compiler.

Index Terms—Formal executable semantics, PLC programming, Structured text, K framework, OpenPLC.

I. INTRODUCTION

PROGRAMMABLE Logic Controllers (PLCs) are responsible for automating process control in several modern industrial systems, e.g., in manufacturing and public infrastructure. It is critical to ensure that PLCs are operating correctly, as any functional or security-related defects may lead to serious incidents in the system. This has most famously been demonstrated by the Stuxnet worm [1], while many other less-known safety and security incidents [2], [3], [4] and potential hazards [5], [6] related to PLCs have resulted in significant consequences, with an estimated $350,000 in damage on average [7].

The majority of PLCs are programmed using languages defined in the IEC 61131-3 open international standard [8]. Programs can be written in graphical languages such as Function Block Diagrams (FBD), but the standard also defines Structured Text (ST), a fully textual language based on the idea of organizing code into “function blocks” and designed with a syntax similar to Pascal. ST is a particularly important IEC 61131-3 language given its utility for data processing [9], and the fact that snippets of ST are actually required in FBD and other graphical languages. It is therefore important that translators and compilers for ST are correctly implemented and exhibit only expected behaviors when the code is being run on a PLC.

This has motivated a surge of research on analyzing and verifying PLC programs [7], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], although few works focus on ST implementations/compilers. Zhang et al. [24] propose VetPLC, a temporal context-aware, program-based approach to produce timed event sequences that can be used for automatic safety vetting. McLaughlin et al. [21] propose TSV which translates assembly-level code into an intermediate language (ILIL) to verify safety-critical code executed on PLCs. Mader and Wupper [26] translate Instruction List (IL) code into timed automata [31]. Bauer et al. [25] similarly use timed automata as the formalism for Sequential Function Chart (SFC). In [27], the proposed method transforms IL to Petri-nets [32], and manually builds two additional Petri-nets for modeling the PLC and its environment. Xiong et al. [23] propose an algorithm based on variable state analysis for automatically extracting a Behavior Model (BM) from an ST program. These works attempt to transform PLC programs into an intermediate language or another programming language (i.e., C) which is suitable for verifying or detecting potential issues using existing associated verifiers or checkers. The issue of these approaches is that they lack analysis and proof of equivalence in the conversion process. In addition, the analyses they perform are often limited (since the existing tools are not designed for PLCs) and do not offer the feedback to the level of source code. Canet et al. [22] propose formal semantics for a significant fragment of the IL language, and a direct coding of this semantics into a model checking tool. Huuck [28] develops a formal operational

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
semantics and abstract semantics for IL, which allows approximating program simulation for a set of inputs in one simulation run. Blech et al. [10], [11], [30] attempted to define the formal semantics of the IL and SFC languages in Coq and NuSMV and, based on that, verify the safety properties in the code. However, IL is a low level assembly-like language that has been deprecated from the IEC61131-3 standard. Furthermore, these studies mainly concentrate on analyzing the functional aspects of the programs and may overlook potential vulnerabilities and security risks introduced during the compilation process.

While extensive research has been conducted on testing more "traditional" compilers (e.g. vulnerability detection for GCC and Clang [33], [34], [35]), compilers for PLC languages such as ST have received much less attention. The challenges associated with testing the implementation of a compiler arise from the inherent difficulties of ensuring its correctness. One particular challenge stems from the absence of a precise specification of the expected behavior of a compiler. For most popular programming languages, there exist multiple purportedly equivalent implementations of compilers. Compiler testing can take advantage of this by utilizing these implementations as oracles for conducting differential testing [36]. However, in the case of the domain-specific ST language, there is no specific implementation standard, and different vendors often develop their own compilers based on their specific requirements. Another challenge is the semantic complexity of the input and output languages that compilers handle. The fact that different vendors develop their own implementations further exacerbates this issue. Compiler testing methods based on formal semantics [37] have shown advantages in addressing these challenges. With a formal semantics of the ST language, the expected behavior of ST compilers can be precisely and unambiguously defined, which can greatly aid in testing and verifying their correctness.

To the best of our knowledge, a practical and complete semantics for the ST language does not exist, which makes it difficult to ascertain the correctness of ST translators and compilers (e.g. by comparing executions). There are a number of reasons why such a reference semantics is yet to emerge. First, there is insufficient documentation defining or describing the complete features of the ST language [9]. For instance, the official documentation introduces language features by only a few examples, based on which it is difficult for readers to fully understand the behavior of the language. Second, the ST compilers provided by different vendors (e.g. Allen-Bradley, Siemens) can implement the language differently, and their closed source solutions make it difficult to fully assess how they behave systematically (other than through manual observation). For example, CODESYS, CX-Programmer, and GX Works2 all produce negative numbers in the results of negative modulo operations, even though this behavior is undefined according to the official documentation. Furthermore, GX Works2 supports only 10 basic data types, whereas CODESYS supports 17 types. Thus, a formal semantics needs to be ‘concrete’ enough to be useful, but ‘high-level’ enough to be general/extendable to the different nuances of vendors’ compilers. A preliminary attempt at defining a high-level semantics for ST was made by Huang et al. [38]. However, it falls short of a full reference semantics as it misses several important features of the language, e.g. certain data types, and key sentences.

In this work, we develop K-ST, a formal executable reference semantics for ST in the $\kappa$ framework [39]. Our high-level semantics is both executable and machine readable, and can be used by the $\kappa$ framework to generate interpreters, compilers, state-space explorers, model checkers, and deductive program verifiers. Our principal goals for the design of K-ST are as follows:

1. **Validated reference semantics.** K-ST is designed to cover all the main features of ST, and is validated against hundreds of different real-world ST programs extracted from GitHub.
2. **General and extendable.** The semantics is high-level (rather than tied to a particular compiler), with the goal of supporting different ST implementations as well as extensions for vendor-specific functions.
3. **Analyses of ST compilers.** Most importantly, K-ST can be used to check the correctness and consistency of different ST implementations, and thus ensure that a compiler is not introducing an unintended behavior or compile-time threat [40], [41] into a critical industrial system.

Given the absence of complete feature descriptions for the ST language in official documentation, we not only refer to the definitions and code samples in the official documents, but also extensively consult the guidance manuals provided by multiple vendors to better define the semantics of the ST language. For example, there is no specific documentation on how integer overflow is handled in the official documents. Through investigating multiple instruction manuals, we found that existing ST compilers generally use truncation to handle integer overflow without any warning. In defining the semantics, we find that the rewriting rule of the $\kappa$ framework provides a good mechanism for capturing the unique features of ST. For example, we can rewrite REPEAT WHILE to achieve the execution effect of REPEAT.

We validate K-ST by extracting 567 real-world ST code samples from GitHub and comparing their executions in our semantics against their executions resulting from various commercial compilers (i.e., CODESYS, CX-Programmer, and GX Works2). We find that K-ST is sufficiently complete to support 509 of these programs (consisting of 26,137 lines of code) and executes those programs correctly (i.e., producing the same outputs as the corresponding existing compiler), with the remaining programs only unsupported due to the use of certain vendor-specific or hardware-related functions that we did not yet formalize. Furthermore, to evaluate the utility of K-ST for testing ST compilers, we compared the executions of the 567 programs (and several mutants) under K-ST and OpenPLC [42], a popular open source PLC program compiler. Through this semantics-based testing, we are able to uncover five bugs and nine functional defects in the OpenPLC compiler, all of them are previously unknown. Fig. 1 summarises the high-level workflow of this process.
In summary, we make three main contributions.

• We propose an executable formal reference semantics for ST;
• We collect a set of 567 complete ST program samples, and validate the correctness of our executable semantics by running those programs in the semantics and via existing compilers (CODESYS, CX-Programmer, and GX Works2), comparing the results.
• We test OpenPLC, an open source PLC program compiler, using our proposed semantics, and find five bugs and nine functional defects.

The remaining part of this paper is organized as follows. Section II introduces the background of ST and the $K$ framework.

The proposed executable operational semantics of ST formalized in $K$ is introduced in Section III. Section IV shows some practical applications of our formal semantics. The evaluation results of the proposed semantics are introduced in Section V. Section VI summarises some related work, and Section VII concludes this work.

II. BACKGROUND

In this section, we briefly introduce the background of the Structured Text (ST) language and the $K$ framework.

A. Structured Text

The Programmable Logic Controller, invented in 1969 by Dick Morley, is specially designed for applications in industrial environments, e.g. assembly lines, robotic devices, or public infrastructure. These kinds of applications all require high reliability and ease of programming.

Early PLCs were represented as a series of logic expressions in some kind of Boolean format. With the development of programming terminals and the complexity of existing control procedures, Ladder Diagrams (LD) were developed to program PLCs. As of 1993, the IEC 61131-3 standard developed by the International Electrotechnical Commission (IEC) defined five programming languages, including two textual programming languages—ST and IL—as well as three graphical languages—LD, FBD, and SFC. A simple example in Fig. 2 [43] shows a ST code example which can be used for linear scaling of an analog sensor signal.

ST is a high-level PLC programming language which is similar to Pascal [44] (widely used from 1980 to 2000), C/C++ and Java. While it contains common constructs from modern programming languages such as FUNCTION, IF/ELSIF/ELSE and CASE branches, WHILE and FOR loops, it has its own characteristics, such as the lack of recursion, capitalized keywords, REPEAT statement, and FUNCTION_BLOCK structure. For instance, FUNCTION_BLOCK as an important part of ST, and has its own state. Its main purpose is to modularize and structure a straightforwardly defined portion of the program. It is similar to the class-object manifestation in object-oriented programming. Function blocks exist in two forms: as a type or as an instance, but only the instance can be called. For each function block, the local variables retain their values between each ‘call’. Table I shows the common elements of ST.

ST, as the only textual programming language supported by the new IEC standard, has a number of advantages compared to other PLC languages. First, being textual, ST programs can be copied relatively easily. Second, compared with the other four languages, it is more convenient for mathematical calculations, formulas and algorithms, and for managing large amounts of data [9]. Third, compared with 20 years ago, PLC solutions are more in demand today and ST can better adapt to this change. Finally, LD, SFC and FBD also require parts of the program to be written in ST anyway [45], [46].

Unfortunately, the absence of documents defining or describing the complete features of the ST language and the differing customizations of vendors can lead to inconsistent implementations of ST. In addition, understanding the semantics of the ST language, and ensuring that it is formally defined is difficult for end users accustomed to graphical programming. A formal executable semantics of ST not only provides a standard, but also

```
1 FUNCTION_BLOCK Scale
2 VAR_INPUT
3 ValueIn : REAL;
4 ScaleInMin : REAL;
5 ScaleInMax : REAL;
6 ScaleOutMin : REAL;
7 ScaleOutMax : REAL;
8 END_VAR
9 VAR
10 a : REAL;
11 b : REAL;
12 Error : BOOL := FALSE;
13 END_VAR
14 IF ScaleOutMin >= ScaleOutMax THEN
15 Error := TRUE; END_IF;
16 IF ScaleInMin >= ScaleInMax THEN
17 Error := TRUE; END_IF;
18 IF ValueIn < ScaleInMin OR ValueIn > ScaleInMax THEN
19 Error := TRUE; END_IF;
20 IF Error = FALSE THEN
21 a := (ScaleOutMax - ScaleOutMin) / (ScaleInMax - ScaleInMin);
22 b := ScaleOutMax - (a * ScaleInMax);
23 Scale := a * ValueIn + b;
24 ELSE
25 Scale := 0;
26 END_IF;
27 END_FUNCTION_BLOCK
```
TABLE I
COMMON ELEMENTS OF THE ST LANGUAGE

| Type                   | Element         | Type                           | Element         | Type                      | Element       |
|------------------------|-----------------|-------------------------------|-----------------|---------------------------|---------------|
| Program Organization Unit | FUNCTION_BLOCK | INT                           | DINT            | SINT                      | ARRAY         |
| PROGRAM                | CASE            | LINT                          | UINT            | UDINT                     | VAR_GLOBAL    |
|                       | WHILE           | UINT                          | USINT           | VAR_INPUT                 | VAR           |
|                       | FOR             | USINT                         | VAR_OUTPUT      | VAR_IN_OUT                | VAR_TEMP      |
|                       | REPEAT          | ULINT                         | REAL            | AT                        | RETAIN        |
|                       | EXIT            | REAL                          | BOOL            | PERSISTENT                | CONSTANT      |
|                       | RETURN          | LREAL                         |                |                           |               |
| Main Statement         |                 |                               |                |                           |               |
| User Data Type         | ENUM            |                               | STRING          |                           |               |
|                       | STRUCT          |                               | WSTRING         |                           |               |
|                       | TIME            |                               | TIME_OF_DAY     |                           |               |
|                       | DATE            |                               | DATE_AND_TIME   |                           |               |
| Built-in Data Type     |                 |                               |                |                           |               |

helps PLC engineers verify the completeness and correctness of these implementations.

B. The \( \mathbb{K} \) Framework

\( \mathbb{K} \) is a formal logic framework based on rewriting logic [47]. It was developed with the overarching goal of pursuing the ideal language framework, where all programming languages have formal semantic definitions and all language tools are automatically derived in a correct-by-construction manner at no additional cost. The \( \mathbb{K} \) backends, such as the Isabelle theory generator, the model checker, and the deductive verifier, can be utilized to prove properties based on the semantics and generated verification tools [48]. Several executable semantics in \( \mathbb{K} \) have been developed for mainstream programming languages, including C [49], Java [50], JavaScript [51], Rust [52], Solidity [53], and IMP [54].

A language semantics definition in \( \mathbb{K} \) consists of three parts: the language syntax, the configuration, and a set of semantics constructed based on the syntax and the configuration. Given the semantics definition for a programming language and some source programs, \( \mathbb{K} \) executes these programs like a translator. For illustration, in the following we take a strict subset of the ST language, i.e., \( \text{ST}_{\text{demo}} \) shown in Fig. 2 as an example to illustrate how to define language semantics in \( \mathbb{K} \).

Configuration. The whole configuration cell \( T \) of \( \text{ST}_{\text{demo}} \) contains two cells, namely \( k \) and \( \text{state} \). The cell \( k \) is used to store the source program \( \$PGM \) for execution, and the cell \( \text{state} \) is used to record the mapping from a variable identifier to its value. The configuration simulates the memory status and environmental changes during runs of the program.

\[
\langle \langle \$PGM : Pgm \rangle_k \langle \text{Map} \rangle_{\text{state}} \rangle_T
\]

With the configuration defined, we present the syntax of \( \text{ST}_{\text{demo}} \) in Fig. 3, which includes some numerical operations, logic operations and commonly used statements. Based on the configuration and the syntax of \( \text{ST}_{\text{demo}} \), we introduce some basic rules in the semantics. The role of the semantics is to tell \( \mathbb{K} \) how to execute the source code, where \( \mathbb{K} \) executes the code and updates the configuration sentence-by-sentence after parsing the source program.

Here, we show the semantics of Allocate, Lookup and Assignment in Fig. 4 as they are the most commonly used constructs in programming languages. Table II describes some common semantic notations. Take Allocate as an example: when \( \mathbb{K} \) runs to lines 9–13 in Fig. 2, the content in the \( k \) cell is \( \langle \text{VAR} a : \text{REAL}; V B s \text{ END_VAR} \cdots \rangle_k \).
The beginning of a semantic rule.

The execution of

Mapping from

Execute

The symbol

The type of variable

VBs

where
demo

Fig. 4. The partial semantics of ST
demo.

rule Variable_Allocate

\[
\text{VAR } X : \text{REAL}; \text{VBs : VarBody} \in \text{VARIABLES END_VAR} \ldots \}
\]

\[
\text{VAR VB} \in \text{VARIABLES END_VAR}
\]

\[
\text{Rho : Map}
\]

\[
\text{(Rho (X \mapsto 0.0))}_{\text{state}}
\]

\text{requires notBool} \ (X \text{ in keys (Rho)})

rule Variable_Finish_Allocate

\[
\text{VAR VarBody} \in \text{VARIABLES END_VAR} \ldots \}
\]

rule Variable_Lookup

\[
\text{X : I \ldots} \}
\]

\[
\langle \ldots X \mapsto I \ldots \rangle_{\text{state}}
\]

rule Variable_Assignment

\[
\text{X := 1 : Float; \ldots} \}
\]

\[
\langle \ldots X \mapsto \text{I} \ldots \rangle_{\text{state}}
\]

where \ VBs \ stands \ for \ b : \text{REAL; Error} : \text{BOOL} := \text{FALSE};.\ Then, \ K \ will \ rewrite \ \langle \text{VAR a : REAL; VB}s \ \text{END_VAR} \ldots \rangle_{k} \ to \ \langle \text{VAR VB}s \ \text{END_VAR} \ldots \rangle_{k}, \ which \ means \ that \ a : \text{REAL} \ has \ been \ executed \ according \ to \ \text{rule Variable_Allocate}. \ Meanwhile, \ it \ adds \ the \ mapping \ between \ the \ variable \ name \ and \ the \ corresponding \ value \ (a \mapsto 0.0) \ in \ the \ current \ \text{state} \ cell \ \text{Rho}. \ In \ addition, \ \text{requires notBool} \ (X \text{ in keys (Rho)}) \ guarantees \ that \ the \ variable \ will \ not \ be \ re-declared. \ Similarly, \ variables \ b \ and \ \text{Error} \ will \ be \ allocated \ separately. \ After \ that, \ the \ content \ in \ the \ k \ cell \ is \ \langle \text{VAR VarBody} \in \text{VARIABLES END_VAR} \ldots \rangle_{k}, \ where \ \text{VarBody} \ represents \ an \ empty \ variable \ declaration \ list, \ that \ is, \ no \ additional \ variable \ needs \ to \ be \ allocated. \ The \ \text{rule Variable_Finish_Allocate} \ will \ be \ called \ to \ convert \ \langle \text{VAR VarBody} \in \text{VARIABLES END_VAR} \ldots \rangle_{k} \ to \ \langle \ldots \rangle_{\text{state}}, \ which \ means \ that \ there \ is \ no \ more \ code \ to \ execute \ in \ the \ \text{VAR block} \ and \ \text{K} \ will \ continue \ to \ execute \ the \ subsequent \ code.

III. FORMAL SEMANTICS OF STRUCTURED TEXT

IN THE $\mathbb{K}$ FRAMEWORK

In this section, we introduce K-ST, the executable operational semantics of ST formalized in $\mathbb{K}$. Note that in practice the PLC programming environment is provided by specific PLC manufacturers including CODESYS and Siemens’s TIA portal (TIA, Structured Control Language (SCL)). As a consequence, the implementations of different manufacturers can vary and may also include their own unique functions or structures.

Our approach is therefore to focus on the common features, allowing other unique functions of the environment to be implemented by extending the operational semantics. Specifically, the syntax of ST is constructed based on the official IEC 61131-3 standard [46]. The configuration is specifically designed for ST. Based on the syntax and the configuration, we then formalize the semantic rules for the language features with rewriting logic. Next, we present each component of the semantic one by one.

A. The Syntax of ST

Table III presents the syntax of ST defined in K-ST, which covers most of the core syntax. We remark that Table III only contains the main part of K-ST while omitting others, e.g., some built-in functions (LEN, DELETE and so on) for space reasons. The syntax is specified by a grammar in a dialect of Extended Backus-Naur Form (EBNF) [55], where "*" means zero or more repetitions. In ST, the top-level grammatical structures include user-defined types (TYPE statements) and three Program Organization Units (POUs): FUNCTION, FUNCTION BLOCK and PROGRAM. Other syntactical elements are derived within these top-level grammatical structures.

B. The Configuration of ST

The execution of an ST program needs to update the following kinds of state: data segment, code segment and stack. Among them, the data segment is used to store global variables, the code segment is used to store program execution code, and the stack is used to store local variables of the program. Note that runtime environment switching caused by function calls is also achieved by the operation of stack. The overall runtime configuration of ST in $\mathbb{K}$ is presented in Fig. 5. We highlight our careful design choices as follows.

Overview. There are 11 main cells in the configuration $T$, i.e., $k$, control, allenv, genv, genv, store, type, constant input, output and next Loc. The value of each cell is initialized according to its specified type. For instance, for cells with a mapping relationship, their values are initialized to Map type, and for cells that store a collection, they are initialized to List type. A '*' followed by any type means an empty set of this type. For instance, *Map in the cell genv represents that genv is initialized with an empty map.
Enumeration type. By default, when an enumeration type is defined in ST, PLC compilers will automatically associate a number (indexed from 0 and incremented by 1 each time) to each variable in the enumeration. For repeated declarations, we use count cell to record the value of the current enumeration.

Global variables. There are two types of global variables. First, the POUs and customized types that users define. These variables can be accessed anywhere in the program. We store these variables in the genv cell as the basis for program operation. Second, the variables defined in VAR_GLOBAL. These variables cannot be directly accessed in the program unless they are declared with VAR_EXTERNAL. We store these variables in the genv cell and provide them on demand.

Program execution. The source code parsed by the syntax SourceUnit, called $PGM$, is stored in the cell $k$ for execution. Then the $PGM$ will be executed unit by unit. If the program terminates normally, there will be a `' in the $k$ cell, denoting that no more units need to be executed. In the
preprocessing phase (the first pass of K), the k cell only contains the token `execute`. Afterwards, K will start executing from the MAIN program.

**Stack operations.** The cell `control` contains seven subcells—fstack, env, temp, count, gvid, print and break—which record the operating environment of the currently running code segment. Specifically, the function stack `fstack` is a list used to store the environment before executing other POUs, including variables in the current environment and the subsequent program. Next, the cell `env` is used to store the mapping relationship between variables and indexes in the current environment during program execution. Furthermore, cells `temp` and `count` are used in ENUM and STRUCT, where `temp` is for temporary mapping and `count` is used as a counting pointer. The cell `gvid` records all identifiers of global variables to assist in the generation of global variables. The cell `print` records variables which need to be output. Finally, `break` stores the program after the loop in order to support the implementation of the EXIT statement in FOR, WHILE and REPEAT loops.

**Execution environment.** The `allenv` cell is used to cache the execution environment before function calls (for strict type checking of parameter passing in function calls). The cell `genv` records the result of the pre-processing (including POUs and custom types) and will be copied to `env` when `env` is refreshed. The last cell related to the environment is called `genv` and is used to index global variables.

**Memory operation.** The `store` cell is used to simulate memory to record the mapping relationships of indexes and variable values. After that, the cells `input` and `output` are used to realize external inputs and external output respectively. The last cell, `nextLoc`, ensures that the index of a variable can always be incremented without duplication. The design consideration behind this is that for complex languages, it is more effective to explicitly manage arbitrarily large memory than use garbage collection [56].

C. Semantics of the Core Features

We implement the executable semantics covering most core features of ST and leave the vendor-specific functionalities as potential extensions. For example, some compilers would use additional keywords to distinguish the declaration part and the execution part of the program. In the following, we provide an overview of four core semantic features of ST, including 1) data types, 2) main control statements, 3) declarations and calls of POUs and 4) memory operations. Before diving into the details, we present the notations as follows.

1) Extended Data Types: The K framework supports diverse data types including identifiers (Id), integers (Int), bools (Bool), floats (Float) and strings (String), which cover most of the requirements. However, there are still some unsupported data types needing additional implementation in K-ST, which we call extended data types. These extended data types can be categorized into two kinds: 1) elementary types (TIME, BYTE, DWORD, TIME_OF_DAY, DATE and DATE_AND_TIME) and 2) compound types (ENUM and STRUCT). We implement these extended data types by the composition of built-in types and methods in K as follows.

We take TIME_OF_DAY as an example to introduce elementary types. There are two types of TIME_OF_DAY in ST, e.g., TIME_OF_DAY#23 : 45 : 56.30 and TOD#23 : 45 : 56.30. Fig. 6 shows our implementation of TIME_OF_DAY type together with its relevant operations. Lines 1 and 2 respectively define the syntax of TIME_OF_DAY and how to parse it (Get_TIME_OF_DAY). Line 3 is used to convert Get_TIME_OF_DAY to TIME_OF_DAY, which is achieved by two

---

1This is optional but recommended for ST compilers.
steps—Gtd2Td and Standardization—where Gtd2Td realizes the conversion of the format and Standardization realizes content conversion, e.g., replacing 60 minutes with 1 hour. Lines 4–11 define some arithmetic and relational operations of TIME_OF_DAY.

For compound types, we take STRUCT as an example and show its semantics in Fig. 7, including both STRUCT declaration and instantiation. Declarations are shown in ruleStruct_Declaration, where we allocate memory for each defined data structure. The instantiation of STRUCT consists of four main steps in ruleStruct_Instantiation: 1) CreatStruct allocates memory for I1, 2) StructInits generates each variable in turn according to Vds in STRUCT, 3) Set assigns values to the corresponding variables according to Idvs, and finally, 4) Update stores the mapping relationship of variables related to I1 into the memory of I1 to facilitate subsequent use.

2) Main Control Statements: Control statements are important in ST for achieving complex program logic (as in most other programming languages). We show the rules for CASE, REPEAT and EXIT in Fig. 8 (as the semantics of IF, WHILE

```plaintext
rule Case
CASE E1 : Expression OF E2 : Expression,
Es : Expressions : S : Statements
CB : CaseBlock END_CASE;
 IF E1 = E2 THEN S ELSE
 CASE E1 OF Es : CB END_CASE; END_IF;
```

```plaintext
rule Repeat
REPEAT S : Statements UNTIL
 E : Expression END_REPEAT;
 S \while \not E \do S END_WHILE;
```

```plaintext
rule Exit
EXIT; \break
 List \( K \) \break
```

and FOR are typical). A CASE statement can be rewritten as a combination of an IF and CASE through rule Case. The rule Repeat is implemented as follows. We first store the subsequent statements outside the loop (recorded as \( K \)) in cell \( \text{break} \) to deal with the EXIT statement that may appear, and then rewrite it into the form of \( \text{WHILE} \) for further execution. During the execution of the loop body, once \( \text{EXIT} \) is executed, all the statements in the current cell \( k \) are discarded and rewritten to \( K \) (storing the subsequent statements), as shown in rule Exit.

3) The Declaration and Call of POUs: In ST programs, statements are inside Program Organization Units (POUs), i.e., FUNCTION, FUNCTION_BLOCK or PROGRAM. A FUNCTION is a stateless POU type, comparing to a FUNCTION_BLOCK which stores its own state after execution. The design of the FUNCTION_BLOCK is similar to the concept of class-object manifestation in object-oriented programming (OOP), which aims to achieve better modularization. FUNCTION_BLOCKS exist in two forms: as a type or as an instance, and only the instance can be called. For a FUNCTION_BLOCK instance, the local variables retain their values between each ‘call’. PROGRAMs are defined by the IEC 61131-3 standard as a “logical assembly of all the programming language elements and constructs necessary for the intended signal processing required for the control of a machine or process by a PLC-system” [46]. Due to space constraints, we show the declaration, call and return operation of FUNCTION_BLOCKS in Fig. 9 as an example for illustration (FUNCTION and PROGRAM are shown in Fig. 10 and explained only when necessary).

**Declaration.** The declaration of FUNCTION_BLOCK is similar to that of STRUCT. As shown in rule Function_Block_Declaration, we first assign an index in memory for FUNCTION_BLOCK \( X \), set the type to the built-in FunctionBlock, and convert the entire declaration statement to the built-in type \( \text{funblambda}(X, \text{void}, Vds, S) \) for storage, where \( \text{void} \) means no return value, \( Vds \) and \( S \) are variable declarations and operations in \( X \) respectively.
The purpose of setting `const` to `true` is to prevent it from being modified. Note that `FUNCTION` and `PROGRAM` set `type` and `store` to `Function`, `funlambda(X,T,Vds,S)` and `Program`, `plambda(X, void, Vds, S, .Map)`.

**Instantiation.** The instantiation of `FUNCTION_BLOCK` is achieved through variable declarations, as shown in rule `Function_Block_Initialization`. However, the value is set to `runfunblambda(X, void, Vds, S, .Map)` to distinguish it from `funlambda` and `.Map` is designed to store the `FUNCTION_BLOCK` environment for next call and external query. This is because a `FUNCTION_BLOCK` can only be called after instantiation, i.e., `runfunblambda` can be executed but `funlambda` cannot. Since `FUNCTION` and `PROGRAM` have no such restrictions, `funlambda` and `plambda` can be directly called and executed.

**Call.** There are two cases when a `FUNCTION_BLOCK` is called. The first case is that the `FUNCTION_BLOCK` is called for the first time, as shown in rule `Function_Block_Call_First`. Since there is no initial environment (the last value of `runfunblambda` is `.Map`), we will first store the current execution environment `info` in `fstack`, including subsequent statements `K`, the `Allenv` of the current environment, and the parameters `C` in cell `control`. Then, we reset parameters `C` through `renew`. After that, `K` executes the variable declaration `Vds` (including index application, initialization and assignment) and statements `S` in the function block. In addition,
Update is used to update the .Map in runfunblambda to record the current environment. Finally, RETURN can return to the calling program and configure the corresponding environment. In other cases (not called for the first time), as shown in rule Function_Block_Call_Others, there is already a mapping relationship between related variables and values in cell store, and the mapping relationship between identifiers and indexes is also stored in the runfunblambda. Therefore, no new memory allocation will be made during the execution process and the existing environment will be used. Note that the value of the variable in the FUNCTION_BLOCK will not be initialized, which means that the execution result for the same input may be different.

Regardless of whether RETURN appears in the FUNCTION_BLOCK, we add a RETURN by default for each FUNCTION_BLOCK as a sign that the FUNCTION_BLOCK has finished running and returned to the calling POUs. Since FUNCTION_BLOCKS and PROGRAMs do not have a return value, we set null as the return value. Note that a FUNCTION has a return value, and the returned value is the value corresponding to the function identifier, so we need to use Clearenv to clean up the memory environment corresponding to the function identifier after calling procedure renew and add the declaration of the function identifier variable in Vds.

4) Memory Operations: Here, we present the rules for memory operations on elementary types in ST, such as built-in types and extended elementary types. What elementary types have in common is that they take only one memory slot. For complex types, such as enums, structs, arrays, etc., which are compositions of elementary types, the memory operation can be regarded as a set of memory operations on elementary types. For instance, the assignment to struct can be equivalent to assigning variable values to each member of the struct.

Similar to ST demo, main memory operations in ST are still composed of Allocation, Lookup, Assignment and additional Clearenv. Where Allocation implements the allocation of memory for variables in the store, Lookup is used to find variable values in store cell, Assignment implements the assignment of variables, and Clearenv implements the recovery of memory in the store. However, because the complete ST semantics has a more complex type design, they will involve more cells in configurations, and are more complicated, as shown in Fig. 11.

Note that HOLE is just a variable, but it has special meaning in the context of sentences with the ‘heat’ or ‘cool’ attribute. In short, ‘heat’ is to lookup the corresponding content of the HOLE in the formula, and ‘cool’ is to put the recheck results back into the formula. For example, in expression $a + b$ where $a$ is represented as HOLE, ‘heat’ is to take $a$ out of the formula and find its corresponding value. If it is 3, and ‘cool’ puts 3 back into the original formula, then the formula becomes $3 + b$.

Let us start with the Assignment operation (we omit Lookup as it is straightforward). The Assignment of ST divides the Assignment of ST demo into two steps, where context and rule Find Index are used to determine the index $L$ of the assigned variable $X$ in store, and rule Assignment implements the update of the store at index $L$. The purpose of this division is to make the Assignment operation better applicable to complex types, because in some cases the index of the assigned variable can not be directly obtained and multiple queries are required. For instance, when assigning a value to $A[3, 5, 7]$, where $A$ is a multi-dimensional array, we need to look up each dimension one by one to finally determine the index. In addition, we refer to the state of $X$ in type and constant during the assignment process. On the one hand, we use Limit to ensure that the assigned value meets the type requirements, and on the other hand, we use constant to ensure that the constant cannot be modified. Although the memory cleaning operation is not necessary for ST in $\mathbb{K}$, a simple Clearenv operation can effectively reduce repetitive code and improve code readability. For rule Clearenv, what needs attention is the operation on cell env: it replaces the index $L$ of variable $X$ with undef which means null in the map supported by $\mathbb{K}$.

ST has relatively complex and strict type definitions, therefore the rule Allocation of ST involves more cells and operations, such as type and constant for storing variable types and whether they are constants, where Undefined is used to generate the default of the specified type. In addition, according to the content in Table I, not only VAR will be used in the variable declaration process, but also other keywords, such as VAR_INPUT, VAR_IN_OUT, etc. In order to reduce the complexity of the code, we also implement these declarations through VAR declarations. For instance, Fig. 12 shows the implementation of VARLOBAL.
rule Global_Declaration
\[ \text{VAR}_{\text{GLOBAL}} X : \text{Id} : T : \text{EleType}; \text{END}_{\text{VAR}} \]
\[ \text{VAR} X : T; \text{END}_{\text{VAR}} \sim \text{let} \text{go}(X) \sim \text{Clear} \text{Env}(X) \]

rule Constant_Declaration
\[ \text{VAR} \text{CONSTANT} X : \text{Id} : T : \text{EleType}; \text{END}_{\text{VAR}} \]
\[ \text{VAR} X : T; \text{END}_{\text{VAR}} \sim \text{SetConstant}(X) \]

Fig. 12. The partial semantics of variable declarations.

We observe that K-ST covers 259 core features with 876 rules in total, using 2315 lines of \( K \) code. The complete code can be accessed through \url{https://github.com/wkyml/K-ST}. It has also been included by the \( K \) team in their projects, which can be found at the following link: \url{https://github.com/runtimeverification/k/blob/master/web/pages/projects.md}.

IV. TESTING AND ANALYSING ST COMPILEDERS

In addition to providing formal references for defined lan-

guages, our formal semantics also has several applications that
use language-independent tools provided by \( K \), such as state
space exploration, model checking, symbol execution and de-
ductive program validation. We omit demonstration of these
applications in this paper since they have been well-illustrated
in related works [51], [52]. In this work, we introduce the
testing of ST implementations/compilers based on our executable
semantics, K-ST.

As discussed earlier, because ST compiles are typically
provided by vendors, the execution behavior of compiles may
be different, and may even be inconsistent with respect to the
high-level semantics [37]. One of the main applications of the
proposed semantics is to define the ‘reference’ execution behav-
or of ST, which can help programmers detect bugs in existing
ST compilers.

To explore this application (and given the closed nature of
commercial compilers), we choose OpenPLC\(^2\) as our test ob-
ject, which is open source and supports ST compiled.

The overall workflow of our testing approach is depicted in Fig. 13.
It includes three parts: program variation, program execution
and result comparison. First, seed programs are mutated to
improve the diversity of test samples. Next, we use the mutated
program as input to run OpenPLC and our executable semantics
respectively. Finally, the result comparison part compares the
consistency of the two execution results. It should be noted
that we use a policy similar to [33], that is, the program does
not need input, and the category of result consistency compar-
ison includes the values of all variables in the program.

The comparison of results is performed to analyze potential
inconsistencies between K-ST and OpenPLC. By comparing
the final execution state of the program with its variable state,
we can identify potential inconsistencies. The execution state

\(^2\)\url{https://www.openplcproject.com/}

focuses on determining whether the program has completed its
execution or terminates at the same statement. On the other
hand, the variable state captures the values of all variables in the
program, including input, output, and intermediate variables,
after the program has finished running. Table IV shows our
measure of consistency, where \( Q \) and \( Q' \) represent the values
of each variable after the program executes, \( I \) and \( I' \) repre-
tent the commands corresponding to the exception termination,
and \( \checkmark \) and \( \times \) represent consistency and inconsistencys
respectively. As a result, unless K-ST and OpenPLC exhibit identi-
cal execution and memory states, their behavior will be de-
emed inconsistent.

In order to better mutate seed programs to improve the diver-
sity of test samples, we propose specific mutation operations in
Table V to generate mutated test samples. These mutation op-

ertions can enrich the test samples while minimizing program
errors. Our method for generating mutant ST programs is shown
in Algorithm 1. Given an ST program \( S \), the algorithm makes a
copy, randomly assigns initial values to all variables at the time
of declaration, and applies some applicable mutation operators
to randomly selected lines in the program. The test is done
by comparing results of these samples in K-ST and OpenPLC.
It should be noted that correct and erroneous programs in the
test sample are both meaningful for checking the consistency
of execution behavior. This is because K-ST and OpenPLC
report program errors at the same time, allowing us to verify a
stronger notion of consistency. In addition, considering the lag
of OpenPLC updates, we also tested it on the latest Beremiz\(^3\)
which uses the same underlying implementation (MATIEC\(^4\))
as OpenPLC. The specific results of the test are shown in Section V.

V. EVALUATION

In order to evaluate the semantics of ST we defined in \( K \),
we deployed K-ST on \( K \) version 5.1.11 (Intel(R) Core(TM) i7-
9750H CPU @ 2.60GHz). In the following, we design multiple
experiments to systematically answer the following research
questions (RQs).

- \textbf{RQ1: How much of the ST language is K-ST covering?}
  Completeness of the semantics is an important indica-
tor to measure executable formal semantics. The lack of
  key semantics will seriously affect the usefulness of for-
  mal semantics.

- \textbf{RQ2: Is K-ST correct?} Semantic correctness is the basis
  for ensuring the usability of executable formal semantics,
  so we need to analyze the correctness of formal semantics
  implemented.

- \textbf{RQ3: Can K-ST be used to discover bugs in a compiler?}
  This is important since a key application of executable
  formal semantics is to identify compiler bugs.

A. Test Sets

For the purpose of evaluating the coverage and the correct-
ness of K-ST, the test data set that we used comes from GitHub.

\(^3\)\url{https://beremiz.org/}

\(^4\)\url{https://github.com/thiagorales/OpenPLC_Editor/tree/master/matiec}
We searched 4853 programs in GitHub through keywords in the ST language. Then, we automatically screened out samples containing other programming languages (2516) and XML forms (1542). After that, we manually splice the remaining programs and remove samples that lack the components required for operation (such as POUs). After screening, 567 complete programs written in pure ST formed our test set. In other words, these 567 samples contain all the components required for operation and do not use other languages, such as C and Python.

With the aim of comprehensively testing the correctness of the execution behavior of OpenPLC, we use two sample sets, including test samples collected from GitHub (GitHub set) and test samples obtained through mutation (Mutated set). The GitHub set is the sample set with 567 test samples mentioned before. The Mutated set is generated by Algorithm 1. We selected 30 high-quality samples from GitHub set as initial mutant seeds. These 30 samples contain all the key features of ST. Then, three rounds of iterative mutation are carried out through Algorithm 1. Each round of iteration produces 10 mutation samples per seed. Except for the initial seed used in the first round, the seeds of each round of mutation are the result of the previous round of mutation. We get a set containing 33,330 mutation samples.
B. Experiment Results and Analyses

1) Semantic Completeness (RQ1): We executed K-ST on 567 test samples collected from GitHub. Among these 567 test samples, K-ST supports the execution of 509 of them. For these 509 tests which K-ST can support, Fig. 14 lists the number of tests for some important features (based on Table I) used in the evaluation. Specifically, there are six kinds of features, namely FUNCTION, FUNCTION_BLOCK, PROGRAM, Declaration types, Datatypes and Statements. For Declaration types, we list the number of tests for CONSTANT, VAR, VAR_GLOBAL, VAR, VAR_INPUT, VAR_OUTPUT, VAR_IN_OUT, VAR_TEMP and VAR_EXTERNAL. For Datatypes, we list the number of tests for elementary types signed integer (INT, DINT, SINT, LINT), unsigned integer (UINT, UDINT, USINT, ULINT), float (REAL, LREAL), Boolean (BOOL), byte (BYTE, WORD, DWORD), string (STRING, WSTRING), and time (TIME, DATE, TIME_OF_DAY, DATE_AND_TIME); compound types enum (ENUM) and struct (STRUCT); and finally, the array type array. For Statements, we list the number of tests for main control statements: IF, CASE, FOR, WHILE, REPEAT, EXIT and RETURN.

As indicated in Fig. 14, compared with FUNCTION, the FUNCTION_BLOCK is more favored by ST programmers (PROGRAM is necessary for ST program operation). For Declaration types, the most used is VAR (with a ratio of 470/509), followed by VAR_INPUT (386/509), VAR_OUTPUT (360/509) and VAR_IN_OUT (313/509). Among all the Datatypes, BOOL is the most used, followed by unsinged integer and ARRAY, constituting 322/509 and 311/509 respectively. For the Datatypes, BOOL is the most common type. In addition, we must remark that we do not count the type of array members. Finally, IF is the most common statement in all the tests considered. This is also in line with the main working scenarios of PLCs.

We remark that we do not consider the vendor-based functions in this experiment as these functions vary not only from vendor to vendor, but even from product to product. In particular, Mitsubishi PLCs provide completely different data types, including Bit, Word[signed/unsigned], Double Word[signed/unsigned], Bit STRING[16-bit/32-bit], FLOAT, STRING[32] and Time. Siemens PLCs support keyword BEGIN to represent the end of variable declaration and the beginning of operation instructions. In addition, there are also obvious differences between different products of the same vendor. For example, the S7-1500 and the S7-1200 from Siemens support different type conversion methods5, where the former only provides explicit conversions of types, and the latter provides both explicit and implicit conversions.

2) Semantics Correctness (RQ2): On the other hand, in order to evaluate the correctness of K-ST, we compared the execution results of K-ST against those of vendor compilers CODESYS, CX-Programmer and GX Works2. We consider the proposed semantics correct if the execution behaviors of K-ST are consistent with the ones of the CODESYS, CX-Programmer and GX Works2 compilers. The consistency criteria described in Section IV are utilized to evaluate the consistency of behavior between K-ST and the compilers provided by vendors.

5https://support.industry.siemens.com/dl/dl-media/272/109742272/art_918238/v6/9351699991/zh-CHS/index.html#ae443583b99950f7c9a8f7237fe81ad4
TABLE VI

| Feature                      | Coverage | Feature                      | Coverage | Coverage       |
|------------------------------|----------|------------------------------|----------|---------------|
| **POUs(core)**               |          | Data types(core)             | Enum instantiation | FC            |
| **POU’s declaration**       |          | SINT                         | FC       |                |
| **FUNCTION**                 | FC       | INT                          | FC       |                |
| **FUNCTION_BLOCK**           | FC       | DINT                         | FC       |                |
| **PROGRAM**                  | FC       | LINT                         | FC       |                |
| **POU’s calls**              |          | USINT                        | FC       |                |
| **FUNCTION**                 | FC       | UINT                         | FC       | Array          |
| **FUNCTION_BLOCK**           | FC       | UDIMT                        | FC       | One-dimensional array | C |
| **PROGRAM**                  | FC       | USTRING                      | FC       | Multi-dimensional array | C |
| **Variable Declaration(core)** |       | REAL                         | FC       | Statements(core) |
| **CONSTANT**                 | FC       | LREAL                        | FC       | Assignment statement |
| **VAR_GLOBAL**               | FC       | BOOL                         | FC       | :=             |
| **VAR**                      | FC       | BYTE                         | FC       | =>             |
| **VAR_INPUT**                | FC       | WORD                         | FC       | Branch statement |
| **VAR_OUTPUT**               | FC       | DWORD                        | FC       | IF             |
| **VAR_IN_OUT**               | FC       | STRING                       | FC       | CASE           |
| **VAR_EXTERNAL**             | FC       | WSTRING                      | FC       | Loop statement |
| **VAR_TEMP**                 | FC       | TIME                         | FC       | WHILE          |
| **AT**                       | C        | DATE                         | FC       | FOR            |
| **RETAIN**                   | N        | TIME_OF_DAY                  | FC       | REPEAT         |
| **PERSISTENT**               | N        | DATE_AND_TIME                | FC       | Break statement |
| **Typed constant**           |          | Enum declaration             | FC       | EXIT           |
| **Type # Data**              | FC       |                              | FC       |                |

Built – in function

**Numerical function** (30)
ADD, SUB, MUL, SQR, INC, DEC, MAX, MIN, MUX, ABS, SQRT, TRUNC, FRAC, FLOOR, LN, LOG, EXP, SIN COS, TAN, COS, TAN ASIN, ACOS, ATAN, NEG, EXPT, DIV, MOD, LIMIT

**Logical function** (9)
GT, LT, GE, LE, EQ, NE, AND, OR, SEL

**String function** (9)
CONCAT, INSERT, DELETE, REPLACE, FIND, LEN, LEFT, RIGHT, MID

**Translate function** (160)
□ FC: Fully Covered and Consistent with Compilers (256/262) □ C: Covered and Consistent with Compilers (3/262) □ N: Not Covered (3/262)

Specifically, if K-ST and these compilers demonstrate identical execution and variable states for the same program, their behavior is deemed consistent. We list the coverage of the K-ST semantics in Table VI from the perspective of each feature specified by the official ST documentation, where FC, C and N mean “Fully Covered and Consistent with Compilers”, “Covered and Consistent with Compilers” and “Not Covered”, respectively.

From Table VI, we can see clearly that for POUs, we fully cover the declaration and call. In variable declarations, AT is related to input and output. We remark, however, that the storage mode of variables in \( \mathbb{K} \) is very different from that in real PLCs, so we just support simple computer-side input and output. In addition, RETAIN and PERSISTENT are related to the actual situation in the PLC, so they are not implemented. For instance, AT is used to bind the actual point of the PLC; RETAIN and PERSISTENT support the preservation of variable values after a power failure or power loss. Array is the only one which is covered but not fully covered in all data types. Limited by the realization of arrays, it is temporarily impossible to achieve the array for enum and struct, and to assign values to multi-dimensional arrays as a whole. In statements, \( \Rightarrow \) has been used in \( \mathbb{K} \) and can be replaced by \( := \). For built-in functions, we show a list which we supported, including 30 numerical functions, 9 logical functions, 9 string functions and 160 translate functions.

In the process of comparing with CODESYS, CX-Programmer and GX Works2, the following points need to be explained. Firstly, due to the closed nature of these compilers, they cannot be simply called, so we have to manually fill the
code in the specified way into the compiler to compile and run, and compare the results, which is laborious and tedious work. This also hinders us from testing these commercial compilers in an extensively large scale. After that, different vendors have obvious differences in the implementation of compilers, so the source code needs to be adapted to a certain extent. For example, only 10 basic data types—Bit, Word[Signed/Unsigned], Double Word[Signed/Unsigned], Bit STRING[16-bit/32-bit], FLOAT, STRING[32] and Time—are provided in the GX Works2 compiler, so we need to adapt the variable types of the source program.

3) Finding Bugs in OpenPLC (RQ3): We execute OpenPLC and K-ST with the GitHub set and Mutated set as input. The execution results of the two data sets are shown in Table VII. Here, $K_pO_j$ is the number of programs that K-ST can execute normally but OpenPLC cannot compile and run; $K_pO_p$ is the number of programs that K-ST cannot run normally but OpenPLC can.

For the GitHub set, K-ST supports 509 of them, and OpenPLC supports 490. Through analysis, we found that the reason for this phenomenon is that OpenPLC has some functional deficiencies. For example, OpenPLC does not support the initialization of variables using formulas at the time of declaration; numerical calculations of BYTE, WORD, DWORD types are not supported, etc.

For the Mutated set, there is a big difference between the execution results of K-ST and OpenPLC. First of all, we filter 2,271 timeout programs that timed out both in OpenPLC and K-ST with 10 seconds as the time limit. After that, we manually analyzed these samples with inconsistent results to determine the causes. For the large $K_pO_j$ value, functional deficiencies remain the main reason.

We found an interesting bug in OpenPLC. The bug is a “VAR” parsing exception in OpenPLC. If the first operation instruction starts with “VAR”, such as “VAR0 := 1;”, OpenPLC terminates abnormally. The interesting phenomenon is when an error statement appears in an unexecuted part of the program, such as after the “RETURN:”; K-ST can execute such a program, but OpenPLC cannot. The main reason for this phenomenon is that K ST adopts an operation-based detection mechanism. Because the error code will not be executed, it will not lead to the termination of our executable semantics. The case study is shown in APPENDIX A.

After that, by analyzing those programs that have different results on K-ST and OpenPLC, we find that the reasons for the different results are mainly due to the differences in underlying implementations between K-St and OpenPLC. For example, for integer mode operation $−7 \mod 3$, the execution result of K-ST is $−1$, whereas the result for OpenPLC is $2$. From a mathematical point of view, both results are correct, but they will have a completely different impact on any following operations. When we run the program again in CODESYS, the results of CODESYS were the same as K-ST.

For those samples that K-ST cannot run normally but OpenPLC can execute normally, our analysis found some bugs in OpenPLC. For example, while OpenPLC can check explicit divide-by-zero operations, it allows the execution of implicit divide-by-zero operations. Table VIII details all functional deficiencies and bugs we found in OpenPLC. We show some relevant case studies in APPENDIX B. Considering that Beremiz can be regarded as an updated version of OpenPLC, we have retested the inconsistencies we found in Beremiz. We found that in the latest Beremiz, it fixes some problems, including negative MOD operation results and “VAR” parsing exceptions. But other bugs and shortcomings still exist. In response to these problems in OpenPLC, we have submitted them to the OpenPLC and Beremiz developers and are waiting for their confirmation.

VI. RELATED WORK

In this section, we discuss some other PLC program analysis techniques, summarize their characteristics, and distinguish them from our work.

Keliris et al. [19] propose a framework (ICSREF) which can automate the reverse engineering process for PLC binaries. They instantiate ICSREF modules for reversing binaries compiled with CODESYS and getting the complete Control Flow Graph (CFG), and they provide an end-to-end case study of dynamic payload generation and attack deployment. Tychlas et al. [7] analyze the binary files generated by all control system programming languages in CODESYS to understand the differences and even the vulnerabilities introduced during the program compilation process. Based on this analysis, they provide a fuzzing framework (ICSFuzz) to perform security evaluation of the PLC binaries. Our work differs from them because we focus on the source code and do not rely on any specific compilation environment.

Kuzmin et al. [57] propose to use linear-time temporal logic (LTL) to guide program behavior and check whether ST programs satisfy the corresponding temporal logic through Cadence SMV. Darvas et al. [15] propose rule-based reductions and a Cone of Influence (COI) reduction variant for state explosion problems that may be encountered in the formal analysis of ST code, and use the NuSMV model checker to verify temporal logic. After that, they [58] provide a state machine and data-flow-based formal specification method for PLC modules. In addition, they [43] analyze the feasibility of converting between the 5 PLC programming languages provided by Siemens, and point out that the extended SCL (a vendor-defined ST) can be

| TABLE VII | THE RESULTS OF K-ST AND OPENPLC |
|-------------|---------------------|
| **Data Set** | **GitHub Set** | **Mutated Set** |
| Number of samples | 567 | 31059 (2271) |
| Number of programs run completely | K-ST | 509 | 15850 |
| | OpenPLC | 490 | 11581 |
| Inconsistent | $K_pO_j$ | 30 | 5664 |
| | $K_pO_p$ | 11 | 1395 |
| Diff. Result | 0 | 735 |
used as the target language for conversion. Adiego et al. [59] propose an intermediate model-based method which can transform PLC programs written in different modeling languages of verification tools to facilitate checking temporal logic. Haileselassie et al. [60] propose UBIS, which converts ST programs with potential intrusions as well as trusted versions of programs into attributed graphs through UPPAAL, and compares their nodes and edges to detect stealthy code injections. Bohlender et al. [61] apply formal verification and falsification of temporal logic specifications to analyze chemical plant automation systems. Rawlings et al. [62] use symbolic model checking tools st2smv and SynthSMV to verify and falsify a ST program controlling batch reactor systems. Xiong et al. [23] use the behavior model (BM) to specify the behavior of ST programs, and provide an method based on automatic theoretical to verify LTL attributes on BM. Our work differs from the aforementioned works because they attempt to transform PLC programs into intermediate languages or other programming languages which are suitable for verifying or detecting potential issues, and lack analysis in the conversion process. In addition, these methods do not offer feedback at the level of source code.

Huang et al. [38] is the closest work to ours. They first defined the executable semantics of the ST language in $K$ and use it to check some security properties. Our work differs because we cover a more complete ST language, and we can use it to discover errors in ST compilers.

**VII. CONCLUSION**

In this paper, we introduced an executable operational semantics of ST formalized in the $K$ framework. We presented the semantics of the core features of ST, namely data types, memory operations, its main control statements, and function calls. Our experimental results show that the proposed ST semantics has already covered the main core language features and correctly implements 26,137 lines of public ST code on GitHub. Furthermore, the application of the proposed semantics in testing and analyzing PLC compilers is discussed. By comparing and analyzing the execution results of OpenPLC and K-ST, we found five bugs and some functional deficiencies in OpenPLC. In the future, we hope to further extend K-ST to support the programming environments provided by different vendors. For example, vendors may customize keywords (Bit STRING of GX Works2), add additional structures (LABEL of Siemens), or even widely extend ST (ExST of CODESYS).

**ACKNOWLEDGMENT**

The authors thank the reviewers for their constructive feedback.

**REFERENCES**

[1] R. Langner, “Stuxnet: Dissecting a cyberwarfare weapon,” IEEE Secur. Privac, vol. 9, no. 3, pp. 49–51, May/Jun. 2011.
[2] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, “The 2015 Ukraine blackout: Implications for false data injection attacks,” IEEE Trans. Power Syst., vol. 32, no. 4, pp. 3317–3318, Jul. 2017.
[3] K. Zeitler, “The Ukrainian power grid was hacked again,” Motherboard, 2017.
[4] N. Perlroth and C. Krauss, “A cyberattack in Saudi Arabia had a deadly goal. Experts fear another try,” The New York Times, 2018. [Online]. Available: https://www.nytimes.com/2018/03/15/technology/saudiarabia-hacks-cyberattacks.html
[5] D. Tychals and M. Maniatakos, “Open platform systems under scrutiny: A cybersecurity analysis of the device tree,” in Proc. 25th IEEE Int. Conf. Electron., Circuits Syst. (ICECS), Piscataway, NJ, USA: IEEE, 2018, pp. 477–480.
[6] A. Nochvay, “Security research: CODESYS runtime, a PLC control framework,” Kaspersky ICS CERT, 2019. [Online]. Available: https://ics-cert.kaspersky.com/publications/reports/2019/09/18/security-researchcodesys-runtime-a-plc-control-framework-part-1
[7] D. Tychals, H. Benkraouda, and M. Maniatakos, “ICSFuzz: Manipulating I/Os and repurposing binary code to enable instrumented fuzzing in ICS control applications,” in Proc. 30th USENIX Secur. Symp. (USENIX Security), 2021, pp. 2847–2862.
[8] Programmable Controllers—Part 3: Programming Languages, International Electrotechnical Commission Standard, 2013.
[59] B. F. Adiego et al., “Applying model checking to industrial-sized PLC programs,” IEEE Trans. Ind. Inform., vol. 11, no. 6, pp. 1400–1410, 2015.

[60] M. Hailesellasie and S. R. Hasan, “Intrusion detection in PLC-based industrial control systems using formal verification approach in conjunction with graphs,” J. Hardware Syst. Secur., vol. 2, no. 1, pp. 1–14, 2018.

[61] D. Bohlender and S. Kowalewski, “Compositional verification of PLC software using horn clauses and mode abstraction,” IFAC-PapersOnLine, vol. 51, no. 7, pp. 428–433, 2018.

[62] B. C. Rawlings, J. M. Wassick, and B. E. Ydstie, “Application of formal verification and falsification to large-scale chemical plant automation systems,” Comput. Chem. Eng., vol. 114, pp. 211–220, 2018.

Kun Wang received the B.S. degree in information and computing sciences from Chongqing University of Posts and Telecommunications of China, in 2017. He received the M.Eng. degree in cyber-space security from the Xidian University of China, in 2020. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Industrial Control Technology, within the Group of Networked Sensing and Control at the School of Control Science and Engineering, Zhejiang University. His research interests include control system security and formal methods.

Jingyi Wang received the bachelor’s degree in information engineering from Xi’an Jiaotong University, in 2013, and the Ph.D. degree from the Singapore University of Technology and Design, in 2018. He is currently a Tenure-Track Assistant Professor with the College of Control Science and Engineering, Zhejiang University, China. He was a Research Fellow with the School of Computing, National University of Singapore during 2019–2020 and with Information Systems Technology and Design Pillar, Singapore University of Technology and Design during 2018–2019. His research interests include formal methods, software engineering, cyber-security, and machine learning.

Christopher M. Poskitt received the Ph.D. degree in computer science from the University of York, in 2014. He is an Associate Professor in computer science (education) at Singapore Management University (SMU), where he is a part of the Centre for Research on Intelligent Software Engineering. Prior to SMU, he held Postdoctoral Research positions at ETH Zürich and SUITD. His research broadly addresses the problem of engineering correct and secure software, especially in the context of cyber-physical systems (e.g. industrial control systems and autonomous vehicles). In addition to software engineering, his research interests include formal methods, cybersecurity, and computer science education.

Xiangxiang Chen received the B.Eng. degree in mechanical engineering from Xi’an Jiaotong University, Xi’an, China, in 2021. He is working toward the Ph.D. degree in cyber-space security with the IS2 Lab at the School of Control Science and Engineering, Zhejiang University, Hangzhou, China. His research interests include fuzzing and AI system testing.

Jun Sun received the bachelor’s and Ph.D. degrees in computing science from the National University of Singapore (NUS), in 2002 and 2006, respectively. He is currently a Tenured Professor with the School of Information Systems, Singapore Management University. From 2010 to 2019, he was an Assistant/Associate Professor with the Singapore University of Technology and Design. He was a Visiting Scholar at MIT from 2011 to 2012. His research interests include software engineering, formal methods, program analysis, and cyber-security.

Peng Cheng received the B.Sc. degree in automation and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively. From 2012 to 2013, he worked as a Research Fellow with Information System Technology and Design Pillar, Singapore University of Technology and Design. He is currently a Professor with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China. His research interests include networked sensing and control, cyber-physical systems, and control system security.