The Modelling and Verification of PLC Program Based on Interactive Theorem Proving Tool COQ

Litian Xiaoa,b,+, Mengyuan Lib, Ming Gua and Jiaguang Suna

aNational Laboratory for Information Science and Technology, Key Laboratory for Information System Security, Ministry of Education, Department of Computer Science and Technology, School of Software, Tsinghua University, Beijing 100084, China.

bBeijing Special Engineering Design and Research Institute, Beijing 100028, China.

Abstract. COQ is an interactive theorem proving tool. The paper abstractly describes the feature of COQ, the architecture and working modes of PLC program with the example of typical PLC. It also introduces the first-order logic syntax and semantics of Intuitionistic Logic. It briefly introduces the main Gallina language syntax elements, the corresponding use methods and main theorem proving tactic on COQ. The work has modeled kernel data type and basic statements and the denotational semantics of PLC program with Gallina. It has given the correctness proof of PLC program based on theorem proving, i.e. based on semantics function the relationship of configuration between the before codes execution and the after is proved. The main purpose is to prove whether a PLC program satisfies certain nature within a scan period.

Keywords: PLC Program, Embedded System, Verification, Denotational Semantics, Theorem Proving, COQ

1. Introduction

With the increasing requirements for embedded applications, the complexity and the scale of PLC programs is also increasing and becoming larger. As the applications of embedded system in a wide range of security field, relevant to its correctness guarantee becomes more urgent. In terms of PLC, the hardware logic of its foundation has high reliability, but its software correctness is much difficultly guaranteed due to the function complexity of PLC program.

In order to eliminate bugs as possible and verify the correctness of a program, the main methods are test [1]-[3] and formal verification. Although the testing technology is an effective method to find out bugs, it has incompleteness. Finite test cases are difficultly utilized to cover all the accessible codes of a program. Thus, the researchers try to use the formal methods to verify the correctness of program.

One type of formal verification method is model checking which to be verified by state space search based on specified features (such as safety and activeness etc.) upon the abstract model. This method is reliable and complete. However, it faces the biggest problem of “state space explosion” [4]-[6]. The other is the formal verification method of theorem proving. It firstly describes the characteristics of system behaviors with a series of logical formulae. Then it proves the specified objectives correct or falsification by the means of deductive methods with inference rules provided by the logical system (or proof tools).

The theorem proving is well suitable for the verification of infinite state system. The most of the proof tools (such as PVS [7], COQ [8] and Nqthm [9], etc) provides the basic logic that is high-order logic possessed
of the function of describing the infinite data structure, i.e., insensitive to the scale of state space. But this type of method has relatively higher requirements to the verification personnel.

For PLC program, its formal verification has its own characteristics:

a) PLC program is used in the most embedded hardware environment, so its logic structure is relatively simple, the sentence categories is few in program, its program is relative short.

b) PLC program is written in hardware machine instruction (or ladder diagram). It is convenient for the abstracted process of model, even in some times it can be used directly as models for verification (similar to assembly language, Verilog, VHDL, etc.).

c) Although PLC program is simple, but it still has the most mechanism of high-level programming language.

d) PLC program completes in sequence within a scan period, and then out images are refreshed and next scan period is started. From the interior of the scan period, it has characteristics of sequence program. From the whole scan period, what showed by PLC is the output response to the various input signals. Though its response is not fixed — because the accumulation of many kinds of timers and counters is span-scan period, a PLC program can be simply regarded as the fixed transformation logic of input-output within a scan period.

Compared with the verification of general assembly or high-level language program, PLC programs' has the advantages that are relative small scale, relative simple structure, relative clear control logic and close to mathematical representation.

In order to verify the correctness of PLC program and the temporal sequence feature of it within a scan period, we address the research for verification approach based on COQ which is a theorem proving tool. To using the COQ, PLC program need to be modeled at first. The paper defines the grammatical components of the typical PLC program based on Gallina specification language of COQ for the program modeling, and expresses it with Gallina. Its main idea is the structural induction conversion of the semantics. It deals with the COQ descriptions of PLC program semantics and proof strategies for proving some program natures. Theses work has established the foundation of COQ verifying PLC program.

2. The Features of Interactive Theorem Proving Tool Coq

COQ and some famous theorem proving tools are generally adopted with Intutionistic Logic \(^{[10]}\) for nature reduction. The logic has the corresponding relation between it and program. On the one hand, the concepts of formula, proof and checking in logic correspond to that of type, function and type checking in functional program. On the other hand, the process of writing functional program can be mutually conversed with the process of theorem proving, which is Curry-Howard correspondence or isomorphism \(^{[10][11]}\).

The Construction Theory of MartinLof et al and is integrated into the extended Automath system by Huet, Cousineau, Coquand and et al, i.e. Calculus of Constructions. Also Coquand gives the termination proof of the calculus \(^{[12]}\). COQ is developed by Inria (a research organization in France). It provides an integrated interactive platform with program description or definition and nature proving. It can be used in various industries for the development of trusted program, also be used for the specification and the verification of program \(^{[8]}\).

Gallina is the specification language of COQ, which is built-in a number of commonly used programming language syntax and type such as integer (\(\mathbb{Z}\)), natural number (\(\text{nat}\)), boolean variable (\(\text{bool}\)), character (\(\text{ascii}\)), string (\(\text{string}\)) and the types of list (\(\text{list}\)) and choice (\(\text{option}\)) etc. derived from the former construction.

In the COQ system, a term and an expression work under certain context and in some scope. The same expression has different meanings sometimes in different scope. After a term or an expression is declared, it immediately has an only type. For example, \(\text{Variables } X Y: \text{Prop}\) declares the type \(\text{Prop}\) (i.e. proposition type) of variable \(X\) and \(Y\). When the command \(\text{Check } X\) is used for examining the type corresponded to \(X\), COQ gives the response, i.e. \(\text{Prop}\). Any expressions have the types of them and cannot be deduced whether the expressions corresponded to types are illegal \(^{[13]}\).

Inductive method can be user-defined new data type \(^{[14]}\). For example, in the type library of COQ the definition of natural number can be

\[
\text{Inductive } \text{nat} : \text{Set} := \{ \text{O} | \text{S} : \text{nat} \rightarrow \text{nat} \},
\]

i.e. natural number type has two constructors \(O\) and \(S\). It can be equivalence to the following inductive
definition:
\[ O \in \text{nat}; \]
If \( m \in \text{nat} \), then \( S \, m \in \text{nat} \).

Therefore, \( O, S \, O, S \, (S \, O) \ldots \) are natural numbers.

Two definitions approach of term and function in COQ are often direct definition and inductive definition\(^{[12]}\) which the former keyword is \textit{Definition} and the latter keyword is \textit{Fixpoint}. Their difference is that the latter is allowed to recursion calling itself within function body. For example,

\[
\text{Fixpoint \ fac (n: nat) : nat :=}
\begin{align*}
\text{match } n \text{ with} \\
\mid O \Rightarrow 1 \\
\mid S \, m \Rightarrow n \,* \,(\text{fac} \, m)
\end{align*}
\]

It defines a recursion function \textit{fac} which receives a \textit{nat} type parameter and returns a \textit{nat} type value. In function body, \textit{match} statement matches form to \( n \), i.e. if the \( n \) value is \( O \) (natural number 0) then the function returns the value \( 1 \), if \( n \) forms such as \( S \, m \) then it returns \( n \,*(\text{fac} \, m) \). Where, there is a recursion calling.

The proved theorem and lemma etc. can be defined by keyword \textit{Theorem} and \textit{Lemma}. The following statements in COQ: Theorem \textit{add_perm}: \[ \forall P \, Q : \text{Prop}, \ P \land Q \leftrightarrow Q \land P. \]
is expressed that conjunction operation has exchangeable nature. Logic symbols \( \neg, \ \land, \ \lor, \ \rightarrow, \ \leftrightarrow, \ \exists, \ \forall \) are respectively expressed with \( \neg, \ \land, \ \lor, \ \rightarrow, \ \leftrightarrow, \ \exists, \ \forall \) in COQ.

Besides modeling language, COQ also provides a series of theorem proof tactics such as \textit{intro, split, unfold, simpl} and \textit{reflexivity} for accomplishing interactive theorem proving. A tactics is corresponding to converse application of rules in Intuitionistic Logic. Main proving tactics and other details about COQ can be referred to [8] [12] [15] [16].

3. The Architecture Definition of PLC Program

The cyclic scanning mode is adopted in PLC. When in operation, each scan period is divided into five stages such as internal processing, program communication, input scanning, program performing and output processing. Among the stages internal processing, program communication are auxiliary. The primary units composed in PLC system hardware are input relay unit (\( X \)), output relay unit (\( Y \)), auxiliary relay unit (\( M \)), status register (\( S \)), timer (\( T \)), counter (\( C \)), data register (\( D \)), PLC pointer (\( P \)) and so on. The various series of PLC typical instructions are basically similar, including the normal opened/closed contacts connecting to the control bus (I/O: \( LD/LDI \)), series connection to the normal opened/closed contacts (\( AND / ANDI \)), parallel connection to the normal opened / closed contacts (\( OR/ORI \)), block operation (\( ANDB/ORB \)), negating (\( INV \)), special element set 1/0 and remaining (\( SET/RST \)), non-operation (\( NOP \)), main program end (\( END \)), stepping (\( STL \) or \( RET \)), push stack / read stack / pop stack (\( MPS / MRD / MPP \)), condition jump (\( CJ \)), loop (\( For...NEXT \)), subprogram call and return (\( CALL/SRET \)), arithmetic operation (\( ADD, SUB, MUL, DIV, INC, DEC \)), digitwise logical operation (\( WAND, WOR, WXOR, NEG \)), comparison (\( CMP \)), data transfer (\( MOV \)), etc. The detailed instructions can be referred to the related manuals and reference [17].

The exact mathematical description of PLC program architecture can be defined as following:

\begin{enumerate}
\item[a)] To introduce an argument \( V_z \in \{0,1\} \) to every I/O relay unit, aux relay unit and status register \( Z \), e.g. the input unit X0 corresponding to \( V_{x0} \).
\item[b)] To introduce a corresponding argument \( V_z \in \mathbb{N} \) (the domain of natural numbers) to every data register unit and counter unit \( Z \).
\item[c)] To regarding the pointer value (such as subprogram jump entry address, conditional jump statement address) as constant value.
\item[d)] Not considering the mathematical description of the timer due to the verification technology in this paper not related to the real-time problem.
\item[e)] The above arguments are divided into two types: global argument and local argument. Among them, the unit operated by \( SET/RST \) command and the counter unit are global arguments, and the rest are local arguments. \( GV \) and \( LV \) are separately denoted as the sets of global and local arguments.
\item[f)] In addition, for the counter units, because a scheduled counting value is required and this value will be set for many times. A temporary argument \( E_c \) will be set for each counter \( C \), which \( E_c \) value is effected by
OUT command. For example, \textit{OUT C0 K200}, when $E_C$ is assigned as 200.

g) Due to stack operation exiting, a structure $S_M \in \{0,1\}^*$ used to record the current changes of bus stack. To consider the following codes,

\begin{verbatim}
LD   X0
MPS
AND  X1
OUT  Y1
MRD
AND  X2
MPS
AND  X3
MPP
\end{verbatim}

Supposed $V_{x0}=1$, $V_{x1}=1$, $V_{x2}=0$, $V_{x3}=1$, then after implementing the codes stack values are respectively $e,1,1,1,1,1,0,0,1$. where empty string $e$ is expressed that stack is empty.

h) There are a large number of commands similar to \textit{AND/OR/ANDB/ORB/ANDI/ORI}, which results of the implementation rely on the current bus connecting status. Meanwhile, the implementation of these commands can change the current bus status, as well as the block series and parallel commands rely on bus connection status prior to some steps. Therefore, it is necessary to set a stack for the current bus connecting status.

Generally $V$ is expressed as an argument and $C$ is constant, which may be with subscript. A quad $<\omega_M, \omega_T, \sigma_G, \sigma_L>$ is expressed as the current Configuration. Where, $\omega_M$ is bus stack and $\omega_T$ is current connecting stack, $\sigma_G$ is assignment functions of global variables and $\sigma_L$ is local, i.e. the current value of every global variable $V$ is assigned as \sigma_G(V) and the value of every local variable $V$ is \sigma_L(V).

In order to give the denotational semantics of PLC program, BNF paradigm is introduced as following:

\begin{verbatim}
e::= V | C | e+e \mid e-e \mid e\times e \mid e\div e \mid ee \mid ee \mid e
\end{verbatim}

Where $+, -, \times$ and $\div$ are corresponding arithmetic operation symbols of “plus, minus, multiplication, division”. Symbol $\&$, $|$, $\oplus$ and $\neg$ are logic operation of “and, or, XOR, not”. Symbol $\&$, $|$, $\oplus$ and $\neg$ are bit operation of “bit and, bit or, bit XOR, bit not”.

Given configuration $CF=<\omega_M, \omega_T, \sigma_G, \sigma_L>$, the semantics value under an expression $e$ is denoted by $\llbracket e \rrbracket_{CF}$. The inductive definition is as follows.

- $\llbracket C \rrbracket_{CF} = C$.
- $\llbracket V \rrbracket_{CF} = \{ \sigma_G(V), V \in GV \} \cap \{ \sigma_L(V), V \in LV \}$
- $\llbracket e_1* e_2 \rrbracket_{CF} = \llbracket e_1 \rrbracket_{CF} \times \llbracket e_2 \rrbracket_{CF}$, where $* \in \{+, -, \times, \div, \&$, $|$, $\oplus, \neg, \} \}$, its operation definition is same to normal.
- $\llbracket \neg e \rrbracket_{CF} = \neg \llbracket e \rrbracket_{CF}$, where $\neg \in \{\neg\}$, its operation definition is same to normal.

4. PLC Program Modeling Based on COQ

In PLC program the components need to be divided into two categories: one category is the component of storage one bit such as input relay unit ($X$), another is that of storage data which is 16 or 32 bit such as counter ($C$). Therefore, $\sigma_G$ and $\sigma_L$ need to refine further which is separated into the different functions composed of “bit type” and “data type” registers.

The following structure data type can model the configuration.

\begin{verbatim}
Record config := MkConfig {
M_Stack : list bool;
T_Stack : list bool;
L_Bit_Map : string \rightarrow bool;
L_Data_Map : string \rightarrow Z;
G_Bit_Map : string \rightarrow bool;
G_Data_Map : string \rightarrow Z;
}\end{verbatim}
Where, \( \text{Record} \) is the keyword of COQ which its data type is equivalence to the structure type of high-level language. Local arguments assigned function \( \sigma_L \) is divided into two fields: \( L_{\text{Bit\_Map}} \) and \( L_{\text{Data\_Map}} \). Global arguments assigned function \( \sigma_G \) is divided into two fields: \( G_{\text{Bit\_Map}} \) and \( G_{\text{Data\_Map}} \).

The COQ description of PLC program statements can be defined nil, unitary, binary and ternary operations of PLC program with type declaration statement.

\[
\text{Inductive PLC\_Z\_OP} := \text{ORB} | \text{ANDB} | \text{INV} | \text{NOP} | \text{MPS} | \text{MRD} | \text{MPP}.
\]

\[
\text{Inductive PLC\_U\_OP} := \text{LD} | \text{LDI} | \text{AND} | \text{ANFI} | \text{OR} | \text{ORI} | \text{OUT} | \text{SET} | \text{RST} | \text{INC} | \text{DEC} | \text{NEG}.
\]

\[
\text{Inductive PLC\_B\_OP} := \text{OUTC} | \text{MOV} | \text{XCH}.
\]

\[
\text{Inductive PLC\_T\_OP} := \text{ADD} | \text{SUB} | \text{MUL} | \text{DIV} | \text{WAND} | \text{WOR} | \text{WXOR} | \text{CMP}.
\]

Where, \( \text{PLC\_Z\_OP} \), \( \text{PLC\_U\_OP} \), \( \text{PLC\_B\_OP} \) and \( \text{PLC\_T\_OP} \) separately describe the commands of which don’t need any operands, need an operand or two operands and three operands.

In PLC control circuit, also an only quaternary operation is \( \text{ZCP} \) statement. So, type declaration is

\[
\text{Inductive PLC\_F\_OP} := \text{ZCP}.
\]

The following declaration statements is defined as main PLC program statements.

\[
\text{Inductive PLC\_Sent} := | \text{Z\_Cmd: PLC\_Z\_OP} \rightarrow \text{PLC\_Sent} | \text{U\_Cmd: PLC\_U\_OP} \rightarrow \text{string} \rightarrow \text{PLC\_Sent} | \text{B\_Cmd: PLC\_B\_OP} \rightarrow \text{string} \rightarrow \text{string} \rightarrow \text{PLC\_Sent} | \text{T\_Cmd: PLC\_T\_OP} \rightarrow \text{string} \rightarrow \text{string} \rightarrow \text{string} \rightarrow \text{PLC\_Sent} | \text{Conseq: PLC\_Sent} \rightarrow \text{PLC\_Sent} \rightarrow \text{PLC\_Sent} | \text{Cond: PLC\_Sent} \rightarrow \text{PLC\_Sent} \rightarrow \text{PLC\_Sent} | \text{Loop\_Z: string} \rightarrow \text{PLC\_Sent} \rightarrow \text{PLC\_Sent} | \text{Loop\_C: nat} \rightarrow \text{PLC\_Sent} \rightarrow \text{PLC\_Sent}.
\]

Where, \( \text{Z\_Cmd\ op} \), \( \text{U\_Cmd\ op\ X} \), \( \text{B\_Cmd\ op\ X\ Y\ Z} \) and \( \text{T\_Cmd\ op\ X\ Y\ Z\ W} \) are separately corresponding to nil, unitary, binary and ternary operation commands. \( \text{Conseq, Cond} \) and \( \text{Loop\ constructor} \) are separately constructed synthetic sequence, conditional branch and loop commands.

For example, considering the PLC program,

\[
\begin{align*}
\text{LD} & \quad \text{X1} \\
\text{JC} & \quad \text{L} \\
\text{AND} & \quad \text{X2} \\
\text{L:} & \quad \text{AND} \quad \text{X3} \\
\text{OUT} & \quad \text{Y1}
\end{align*}
\]

in COQ the program can be expressed as

\[
\text{Conseq (Conseq (U\_Cmd \ LD \ "X1") (\text{Cond (U\_Cmd \ AND \ "X2") (U\_Cmd \ AND \ "X3")}) (U\_Cmd \ OUT \ "Y1")}.\]

Based on formal denotational semantics, the denotational semantics of PLC program is a mapping from a configuration to another. The declaration is described as followings.

\[
\text{Fixpoint Prog\_Semantics (St : PLC\_Sent) (C: config): config := match St with}
\]

\[
| \text{Z\_Cmd\ op} \Rightarrow (\text{Z\_Cmd\ Semantics\ C\ op}) \\
| \text{U\_Cmd\ op\ X} \Rightarrow (\text{U\_Cmd\ Semantics\ C\ op\ X}) \\
| \text{B\_Cmd\ op\ X\ Y} \Rightarrow (\text{B\_Cmd\ Semantics\ C\ op\ X\ Y}) \\
| \text{T\_Cmd\ op\ X\ Y\ Z} \Rightarrow (\text{T\_Cmd\ Semantics\ C\ op\ X\ Y\ Z}) \\
| \text{Conseq\ St\ St′} \Rightarrow (\text{Prog\_Semantics\ (St\ (Prog\_Semantics\ St′\ C))}) \\
| \text{Cond\ St\ St′} \Rightarrow (\text{match\ (T\_Stack\ C)\ with}) \\
| \text{I:: i \Rightarrow Prog\_Semantics\ St\ C} \\
| \text{\_ \Rightarrow Prog\_Semantics\ (Conseq\ St\ St′)\ C} \\
| \text{Loop\ Z\ X\ St′} \Rightarrow
\]

\[
\begin{align*}
\text{if (Is\ glb\ var\ name\ X)\ then} \\
\text{if \ (G\ Data\_Map\ C\ X <= 0)\ then} \\
\text{let \ C′ := Prog\_Semantics\ St′\ C\ in} \\
\text{C″ := MkConfig\ (M\ Stack\ C)\ (T\ Stack\ C)\ (L\ Bit\ Map\ C)}
\end{align*}
\]
(L_Data_Map C') (G_Bit_Map C') (Adapt (G_Data_Map C') X (G_Data_Map X)-1) in Prog_Semantics (Loop_Z X St') C''
else C
else if (L_Data_Map C X <> 0) then
let C' := Prog_Semantics St' C in
let C'' := MkConfig (M_Stack C') (T_Stack C') (L_Bit_Map C') (Adapt (L_Data_Map C') X (L_Data_Map C X)-1) (G_Bit_Map C) (G_Data_Map C) in
Prog_Semantics (Loop_Z X St') C''
else C
| Loop_C  K St' ֜  if (K <> 0) then let C' := Prog_Semantics St' C in Prog_Semantics(Loop_C K-1 St') C' else C end.

Where, Adapt accomplishes the single substitution function (i.e. \( f\{x:=t\} \)). string_eq_bool is used for the function that is compared two strings whether to equal. prefix is a function pre-defined in String library, which it is used to judge whether a string is the prefix of another string. In Prog_Semantics function definition, some assistant functions Z_Cmd_Semantics, U_Cmd_Semantics, B_Cmd_Semantics, T_Cmd_Semantics and F_Cmd_Semantics is the semantics that calculate nil, unitary, binary, ternary and quaternary commands. Specific definition of these functions can be expanded, which we shall discuss in another paper because of limited length.

5. Nature Proving of PLC Program Based on COQ

Some typical PLC programs will be examples to illustrate how to prove the nature of PLC programs in COQ. At first, the example starts from a basic module which accomplishes simply data exchange. The program is showed as following.

MOV  D1  D3
MOV  D2  D1
MOV  D3  D2

Its proved nature is described with the first-order logic formula of Intuitionistic Logic:

\[ C: \text{config.} (L\_Data\_Map) C \; \text{"D1"} = L\_Data\_Map (Prog\_Semantics Cod C) \; \text{"D2"} \wedge (L\_Data\_Map) C \; \text{"D2"} = L\_Data\_Map (Prog\_Semantics Cod C) \; \text{"D1"}. \]

Where, Prog_Semantics Cod C is the COQ description of above three line codes, i.e.

Conseq ((B\_Cmd MOV "D1""D3")
Conseq (B\_Cmd MOV"D2""D1") (B\_Cmd MOV "D3""D2").

The formula can be proved with the tactics of intro, split, unfold, simpl and reflexivity.

Considering that the value is summed from 1 to n, \( \sum_{i=1}^{n} \) which corresponding codes is showed as following.

FOR D1
ADD D1 D2 D2
NEXT

The Cod’ description in COQ is

LOOP_Z (T\_Cmd ADD "D1""D2" "D2").

The assistant function is defined as following for describing accumulation.

Fixpoint sum_up (n: nat) : nat :=
match n with
| O ⇒ O
| S m ⇒ n + (sum_up m)
end.

Then proved nature is described with the first-order logic formula:

\[ C: \text{config.} (L\_Data\_Map C \; \text{"D2"})=0 \rightarrow (L\_Data\_Map (Prog\_Semantics Cod'C) \; \text{"D2"}) = (sum\_up (L\_Data\_Map C \; \text{"D1"})). \]

Induction need to use for proving the nature. Inductive object should be L_Data_Map C "D1", i.e. after tactic intro the tactic induction L_Data_Map C "D1" should be used. Where, the first proving object can be eliminated by tactic simpl and reflexivity. The second proving object can be proved by tactic rewrite or change, simpl and reflexivity.
6. Conclusion and Further Work

With the example of typical PLC, the paper abstractly describes the feature of COQ, the architecture and working modes of PLC program. It also introduces the first-order logic syntax and semantics of Intuitionistic Logic. Further with theorem proving tool COQ, it briefly introduces the main Gallina language syntax elements, the corresponding use methods and main theorem proving tactic.

The work has modeled kernel data type and basic statements and and the denotational semantics of PLC program with Gallina. It has given the correctness proof of PLC program based on theorem proving, i.e. based on function Prog_Semantics the relationship of configuration between the before codes execution and the after is proved. The main purpose is to prove that a program should satisfy certain nature within a scan period. It is not much on the related work about correctness verification of PLC program based on theorem proving, which there mainly are [18] and [19].

It is the foundation of verification and theorem proving for PLC program. Given the denotational semantics of PLC program, we have further studied that the semantics is transformed into other mathematical model, i.e. the transition system (model) need to use in model checking technology. Specifically, the research further develops PLC program modeling relation based on Gallina language of COQ platform and makes the COQ description for each PLC syntax elements automatically defined. Then the development of a definition tool can really help PLC program verification.

7. Acknowledgements

This research is sponsored by NSFC Program (No.90718039 & No.91018015) and 973 Program (No.2010CB328003) of China.

8. References


