Implementation of the Composition-nominative Approach to Program Formalization in Mizar

Ievgen Ivanov, Artur Kornilowicz, Mykola Nikitchenko

Abstract

In this paper we describe an ongoing work on implementation of the composition-nominative approach to program formalization in Mizar proof assistant based on the first-order logic and axiomatic set theory. The further aim of this work is development of a formal verification tool for software which processes and communicates with complex forms of data.

Keywords: Formal methods, program semantics, semistructured data, formalization, proof assistant.

1 Introduction

Formal verification of software systems has been a topic of interest of researches in computer science for more than fifty years. During this period many formal software verification tools based on different theoretical frameworks (automata theory, first-order logic, dynamic logics, program logics, etc.) were developed, but most of the existing tools are still in research stage and their usage in software industry is negligible. Some reasons include:

- they do not integrate well into typical software development cycles;
- successful practical application of such tools requires specialized knowledge, is labour intensive, time consuming, and not cost effective for most software projects.

However, in industries related to development of safety-critical systems such as aerospace, automotive, health technology formal verification of software plays a more significant, but still limited role.
The well known tools that support or aid formal verification of software for safety-critical systems include:

- model checkers such as Simulink Design Verifier, Systerel Smart Solver;
- verified translators and compilers such as CompCert that generate code in a low-level language or machine code from the source code in a high-level language that is proven to be equivalent to the source code under the assumptions of the formalized source and target language semantics;
- microkernels and hypervisors such as seL4 and CertiKOS that are formally verified with respect to formal specifications of their application programming interfaces and formal models of microprocessor instruction set architectures.

These tools allow one to eliminate some sources of deviations of software implementation from its specification and the implied safety problems, however, they and their underlying theoretical frameworks focus on verification of relatively simple systems, primarily on systems of the following types:

- software which performs basic logical operations or numerical computations (e.g. software controllers);
- communication protocols which involve simple types of data.

Therefore such tools lack sufficiently easily usable methods of verification of

- software which performs complex processing of partially structured (semistructured) data;
- communication protocols which involve complex types of data.

Besides, their application is restricted because it requires specialized knowledge, is labour intensive, and time consuming.

These restrictions are a factor that may prevent expansion of the mentioned tools and theoretical frameworks outside of traditional safety-critical systems domains. Emerging high tech areas like the Internet of Things (IoT) rely on the idea of combining software systems and hardware devices and physical objects which involve complex interaction protocols, large-scale interaction, and processing of large volumes of semistructured data e.g. in home automation, smart buil-
ings, smart cities, etc. Errors in IoT software can impact the real world and lead to cyber security breaches or direct hazards to humans, but due to the nature of IoT systems in each particular case the potential impact of software errors is difficult to assess. Moreover, such errors are difficult to eliminate through testing, because IoT systems have to be able to function under a variety of circumstances which are costly to model or reproduce. Thus IoT and other relevant high-tech areas could benefit from introduction of formal software verification approaches to their systems development processes.

Software for the Internet of Things (IoT) and other emerging high-tech areas has some differences from traditional safety-critical software which make application of the state-of-the-art verification tools to it rather difficult. One of them is processing of complex, usually semistructured types of data, instead of simple types of data such as logical values or numbers. Usually such data are encoded in data formats like JSON and XML which have tree-like, hierarchical nature.

2 General considerations about implementation of the approach

The work described in this paper aims to implement a formal verification tool which may overcome some limitations of the existing verification tools which prevent them to deal with software which processes complex, semistructured types of data.

The implementation of this tool is based on the composition-nominative approach to program formalization [1] [2] [3], and the Mizar system [4] [5], a software for formalizing mathematical theories (proof assistant) based on first-order logic and Tarski-Grothendieck set theory and a library of already formalized theories (Mizar Mathematical Library).

Composition-nominative approach provides the means of formalization of data – the notion of nominative data which is able to uniformly represent common forms of data used in programming (e.g. lists, trees, tables, multidimensional arrays, etc.), mathematical models of software
which operates on such data based on generalization of Glushkov algorithmic algebras [6], and a logic for reasoning about properties and correctness of such software – a generalized Floyd-Hoare logic [7] [8] with partial pre- and post-conditions for programs which operate on nominative data [9] [10]. The Mizar system provides an environment where the notions, models and logics of the composition-nominative approach can be formalized and implemented. In more detail the plan of their implementation in Mizar is described in [11].

A benefit of usage of Mizar as such an environment is that the Mizar Mathematical Library includes a large amount of notions and facts about continuous mathematics which allow formalization of mathematical models of physical aspects of IoT systems. More details on the link between the composition-nominative approach and the mathematical systems theory can be found in [12] [13].

3 Mizar Formalization

We formalized the basic notions of composition-nominative approach in a series of Mizar papers entitled NOMIN_1.MIZ [14]–NOMIN_4.MIZ and PARTPR_1.MIZ [15]. More specifically, we formalized

- the notion of a nominative data with simple names and complex values called the nominative data of type $TND_{SC}$ [6];

- the main operations on nominative data of type $TND_{SC}$ – naming, denaming, overlapping [6]; together with their domain they form an algebra called the nominative data algebra;

- the notion of a partial predicate on an arbitrary nonempty set;

- the logical compositions of negation, conjunction, disjunction [6] on partial predicates defined in accordance with the truth tables of Kleene’s strong logic of indeterminacy [16]; together with the set of partial predicates they form an algebra which belongs to the class of Kleene algebras [17];
On Implementation of the Composition-nominative ...

- the notion of a binominative function [6] which is a partial function mapping nominative data to nominative data; such functions serve as semantic models of sequential programs which process nominative data;

- the main compositions on binominative functions and partial predicates on nominative data [6] – sequential composition (of two binominative functions), superposition (of binominative functions into a binominative function or a predicate on nominative data), assignment (of the result of a binominative function to a name), branching (“if” operator), loop (“while” operator); together with the sets of binominative functions and predicates on nominative data and several chosen constants (null-ary compositions) they form an algebra called the nominative algorithmic algebra [18] which generalizes Glushkov algorithmic algebras [19];

- the notion of a Hoare triple [9], [7], [8] consisting of a precondition, a program, and a postcondition; the precondition and postcondition are partial predicates on nominative data; the program is a binominative function;

- the Floyd-Hoare composition [9];

- the rules of the extended Floyd-Hoare logic [9] for programs over nominative data (on semantic level) which allow reasoning about Hoare triples with partial pre- and post-conditions.

The mentioned elements allow one to

- represent semantics of sequential programs which process hierarchically organized data (modeled as nominative data) in the form of binominative functions in Mizar;

- formulate program properties in the form of Hoare triples with pre- and postconditions represented by partial predicates on nominative data;
prove program properties using the rules of the extended Floyd-Hoare logic in Mizar.

The main elements of the mentioned Mizar formalization are described below. The details on syntax and semantics of the Mizar language can be found in [20].

The set of nominative data with simple names and complex values [6] over given (arbitrary, but fixed) nonempty sets of names \( V \) and atomic values \( A \) is the set

\[
ND(V, A) = \bigcup_{k=0}^{\infty} ND_k(V, A),
\]

where

\[
ND_0(V, A) = A \cup \{\emptyset\},
\]

\[
ND_{k+1}(V, A) = A \cup \left(V \xrightarrow{n} ND_k(V, A)\right), \quad k \in \{0, 1, 2, \ldots\}.
\]

Here \( V \xrightarrow{n} X \) denotes the set of all partial functions with finite graph from \( V \) to a set \( X \).

The set \( ND(V, A) \) and the associated notions are formalized in Mizar as follows (the detailed definitions can be found in [14] [21]):

- The Mizar mode \( \text{NonatomicND} \) of \( V, A \) represents the type of nonatomic nominative data, i.e. the type of elements of the set \( ND(V, A) \setminus A \). Here \( V, A \) are set-valued parameters of the type.

- The mode \( \text{TypeSCNominativeData} \) of \( V, A \) represents the type of elements of the set \( ND(V, A) \) (i.e. of either atomic or nonatomic nominative data). It is defined as

```mizar
definition
let V, A be set;
mode TypeSCNominativeData of V, A -> set means
it in A or it is NonatomicND of V, A;
end;
```

- The functor \( \text{ND}(V, A) \to \text{set} \) represents the set \( ND(V, A) \) mentioned above. It is defined using \( \text{TypeSCNominativeData} \) as
On Implementation of the Composition-nominative...

definition
let V,A be set;
func ND(V,A) -> set equals
the set of all D
where D is TypeSCNominativeData of V,A;
end;

The main operations on nominative data are formalized in two ways (the detailed definitions can be found in [14]).
Firstly, each operation is defined as a functor which receives a nominative data as an input and returns a nominative data.

• The functor naming(V,A,v,D) -> NonatomicND of V,A (where D is of type TypeSCNominativeData of V,A) represents the naming operation $\Rightarrow v$ on $ND(V,A)$; here $V,A,v$ are the operation's parameters (assumed to be fixed) and $D$ is the operation's input. The operation's result is a nonatomic nominative data.

• The functor denaming(v,D) -> TypeSCNominativeData of V,A (where $D$ is of type NonatomicND of $V,A$) represents the denaming operation $v \Rightarrow$ on $ND(V,A)$; here $V,A,v$ are the operation's parameters (assumed to be fixed) and $D$ is the operation's input. The operation's result is a nominative data.

• The functor global_overlapping(V,A,d1,d2) -> TypeSCNominativeData of V,A (where $d1,d2$ are data of type TypeSCNominativeData of $V,A$) represents the binary overlapping operation $\nabla$ on $ND(V,A)$. Here $V,A$ are the operation's parameters (assumed to be fixed) and $d1,d2$ are the operation's left and right input respectively.

Secondly, each operation is defined using a functor which returns the corresponding partial function on nominative data:

definition
let V,A be set; let v be object;
func naming(V,A,v)
-> Function of ND(V,A),ND(V,A) means
for D being TypeSCNominatorData of V,A holds
  it.D = naming(V,A,v,D);
end;

definition
  let V,A be set; let v be object;
  func denaming(V,A,v) -> PartFunc of ND(V,A),ND(V,A) means
    dom it = ND(V,A) \ A &
    for D being NonatomicND of V,A st not D in A
      holds it.D = denaming(v,D);
end;

definition
  let V,A be set; let v be object;
  func local_overlapping(V,A,v) -> PartFunc of [:ND(V,A),ND(V,A):],ND(V,A) means
    dom it = [: ND(V,A) \ A , ND(V,A) \ A :] &
    for d1,d2 being NonatomicND of V,A
      st not d1 in A & not d2 in A holds
      it. [d1,d2] = local_overlapping(V,A,d1,d2,v);
end;

Depending on the way of usage, one of these two formalizations of
operations may be more convenient than another one.

The types of partial predicates on nominative data and binominative
functions are defined as follows [18]:

definition
  let V,A;
  mode SCPartialNominatorPredicate of V,A is PartFunc of ND(V,A),BOOLEAN;
  mode SCBinominatorFunction of V,A is PartFunc of ND(V,A),ND(V,A);
end;

Also, the sets of partial predicates on nominative data (Pr(V,A))
and binominative functions (FPrg(V,A)) are defined as follows:
Binominative functions represent semantics of sequential programs which process nominative data (i.e. receive input data, perform certain computations and produce output data) and partial predicates represent semantics of conditions on nominative data.

Binominative functions corresponding to programs are constructed from the basic functions using operations (compositions) corresponding to programming language constructs like sequential execution, branching, loop.

These compositions were defined mathematically in [6]. The most important of them are formalized in Mizar as follows (the detailed formal definitions can be found in [18]).

- The functor
  \[ \text{SCassignment}(V,A,v) \rightarrow \text{Function of } \text{FPrg}(V,A),\text{FPrg}(V,A) \]
  formalizes the unary assignment composition \( \text{Asg}^v \) which acts on the set of binominative functions \( ND(V,A) \rightarrow ND(V,A) \) on data of type \( TND_{SC} \). Here \( v \) is a parameter (name) to which a value is assigned. The result of application of the assignment composition \( \text{Asg}^v(f) \) (where \( f : ND(V,A) \rightarrow ND(V,A) \) is given) represents the semantics of the program consisting of the assignment statement \( v := f \), where \( f \) is the semantics of the expression in program variables on the right hand side of the assignment statement.

- The functor \( \text{SCIF}(V,A) \rightarrow \text{Function of } \lfloor :\text{Pr}(V,A),\text{FPrg}(V,A),\text{FPrg}(V,A) : \rceil,\text{FPrg}(V,A) \) formalizes the ternary branching composition \( IF \). The result
of application of branching composition \( IF(p, f, g) \), where \( p : ND(V, A) \rightarrow \{T, F\} \) is a partial predicate, \( T, F \) are logical values (\{True, False\}), and \( f, g : ND(V, A) \rightarrow ND(V, A) \) are binominative functions, represents semantics of the program consisting of the statement “\( \text{if } p \text{ then } f \text{ else } g \)”.

- The functor \( SC\text{while}(V,A) \rightarrow \text{Function of } [[:Pr(V,A),FPrg(V,A):],FPrg(V,A)] \) formalizes the binary While loop composition \( WH \). The result of application of the While loop composition \( WH(p, f) \), where \( p : ND(V, A) \rightarrow \{T, F\} \) is a partial predicate, \( T, F \) are logical values (\{True, False\}), and \( f : ND(V, A) \rightarrow ND(V, A) \) is a binominative function, represents semantics of the program consisting of the statement “\( \text{while } p \text{ do } f \text{ end} \)”.

The necessary compositions of partial predicates on arbitrary sets are formalized in PARTPR_1.MIZ [15].

Generally, the set of partial predicates (on an arbitrary nonempty set \( D \)) is formalized as follows:

\[
\text{definition}
\begin{align*}
\text{let } D; \\
\text{func } Pr(D) \rightarrow \text{set equals} \\
P\text{Funcs}(D,\text{BOOLEAN}); \\
\text{coherence;}
\end{align*}
\]

end;

The compositions are formalized as follows:

- The functor \( \text{PPnegation}(D) \rightarrow \text{Function of } Pr(D),Pr(D) \) represents the negation on partial predicates on a set \( D \).

- The functor \( \text{PPdisjunction}(D) \rightarrow \text{Function of } [:Pr(D),Pr(D):],Pr(D) \) represents the disjunction on partial predicates on a set \( D \).

Alternatively, the negation and disjunction operations on predicates are formalized as functors \( \text{PP not} \) and \( \text{PP or} \) as follows:
definition
let D,p,q;
func PP_or(p,q) -> PartialPredicate of D equals
PPdisjunction(D).(p,q);
coherence;
commutativity;
  idempotence;
end;

definition
let D,p;
func PP_not(p) -> PartialPredicate of D equals
PPnegation(D).p;
coherence;
involutiveness;
end;

There are also specializations of these notions for predicates on nominative data.

- The functor
  \( SC\_not(p) \rightarrow SCPartialNominativePredicate of V,A \)
  where \( p \) is of type \( SCPartialNominativePredicate of V,A \), represents the negation on partial predicates on \( ND(V,A) \rightarrow \{T,F\} \).

- The functor
  \( SC\_or(p,q) \rightarrow SCPartialNominativePredicate of V,A \)
  where \( p,q \) are of type \( SCPartialNominativePredicate of V,A \), represents the disjunction on partial predicates on \( ND(V,A) \rightarrow \{T,F\} \).

Conjunction and implication are formalized as a derived compositions using negation and disjunction:

definition
let V,A,p,q;
Ievgen Ivanov, et al.

func SC_and(p,q) -> SCPartialNominativePredicate of V,A equals
    SC_not SC_or(SC_not(p),SC_not(q));
coherence;
commutativity;
idempotence;
func SC_imp(p,q) -> SCPartialNominativePredicate of V,A equals
    SC_or(SC_not(p),q);
coherence;
end;

For reasoning about properties of programs (binominative functions) on nominative data an extended Floyd-Hoare logic [9] is used. Semantically, it is based on the notion of a Hoare triple formalized in Mizar as follows.

A semantic Floyd-Hoare triple is a triple \((p,f,q)\), where

- \(p, q : ND(V,A) \rightarrow \{T,F\}\) are partial predicates on \(ND(V,A)\) called the pre- and post-condition respectively;

- \(f : ND(V,A) \rightarrow ND(V,A)\) is a binominative function (representing semantics of a program),

such that for each data \(d \in ND(V,A)\), if \(p(d)\) is defined and true, \(f(d)\) is defined, and \(q(f(d))\) is defined, then \(q(d)\) is true.

This means that whenever the precondition \((p)\) is satisfied \((p(d)\) is true) on the input data \((d)\) of the program \((f)\), and the program terminates on the input data \((f(d)\) is defined), and the postcondition \((q)\) is defined on the program’s output \((f(d))\), the postcondition is satisfied on the program’s output \((i.e. q(f(d))\) is true).

The set of all semantic Floyd-Hoare triples (for pre-/postconditions and binominative functions over \(ND(V,A)\)) is formalized in Mizar as follows:

definition
let V,A;
func SemanticFloydHoareTriples(V,A) -> set equals
\{ <p,f,q> where p,q is SCPartialNominativePredicate of V,A,
    f is SCBinominativeFunction of V,A :
for d holds d in dom p & p.d = TRUE & d in dom f & f.d in dom q implies q.(f.d) = TRUE ); coherence; end;

notation let V,A; synonym SFHTs(V,A) for SemanticFloydHoareTriples(V,A);
end;

The type of semantic Floyd-Hoare triples is introduced as follows:

definition let V,A;
mode SemanticFloydHoareTriple of V,A is Element of SemanticFloydHoareTriples(V,A);
mode SFHT of V,A is Element of SFHTs(V,A);
end;

Then semantic versions of the rules of the extended Floyd-Hoare logic for programs over nominative data [11] are formalized in the form of theorems which state that under certain conditions, the results of application of certain compositions to predicates and binominative functions form a semantic Floyd-Hoare triple.

For example, the composition rule $R.IF$ described in [11] is formalized as follows:

theorem$<*SC_and(r,p),f,q*>$ is SFHT of $V,A$ & $<*SC_and(SC_not(r),p),g,q*>$ is SFHT of $V,A$ implies $<*p,SC_IF(r,f,g),q*>$ is SFHT of $V,A$

This theorem formalizes the fact that $(p,IF(r,f,g),q)$ is a semantic Floyd-Hoare triple, if both $(r \land p, f, q)$ and $(\neg r \land p, g, q)$ are semantic Floyd-Hoare triples.

Informally, this means that in order to prove that from the assumption that on the input data of the program
if $r$ then $f$ else $g$

the precondition $p$ holds, after execution of this program the postcondition $q$ holds, it is sufficient to show that

- if $r \land p$ holds before execution of $f$ (the first branch), then $q$ holds after execution of $f$

- if $\neg r \land p$ holds before execution of $g$ (the second branch), then $q$ holds after execution of $g$.

Other rules of the extended Floyd-Hoare logic for programs on nominative data are formalized in a similar way.

4 Conclusions

We have described the ongoing work on implementation of the composition-nominative approach to program formalization in Mizar proof assistant. In particular, we described the way in which hierarchical data, sequential programs and conditions on hierarchically organized data are formalized and the way in which rules of an extended Floyd-Hoare logic for reasoning about properties of such programs are represented in Mizar. We plan to use the obtained results as a basis for development of a formal verification tool for software which processes complex forms of data.

References


On Implementation of the Composition-nominative ...


On Implementation of the Composition-nominative ...


Ievgen Ivanov¹, Artur Korniłowicz², Mykola Nikitchenko³

¹ Taras Shevchenko National University of Kyiv, Ukraine
E-mail: ivanov.eugen@gmail.com

² University of Biały Stok, Poland
E-mail: arturk@math.uwb.edu.pl

³ Taras Shevchenko National University of Kyiv, Ukraine
E-mail: nikitchenko@unicyb.kiev.ua