Software Engineering using Formal Methods Reasoning about Programs with Dynamic Logic

Wolfgang Ahrendt

08 October 2015

Dynamic Logic

(JAVA) Dynamic Logic

Typed FOL

- + (JAVA) programs p
- + modalities $\langle p \rangle \phi$, [p] ϕ (p program, ϕ DL formula)

▶ + ... (later)

Remark on Hoare Logic and DL		
In Hoare logic {Pre} p {Post}	(Pre, Post must be FOL)	
In DL Pre \rightarrow [p]Post	(Pre, Post any DL formula)	

Proving DL Formulas

An Example

∀ int x;

$$(x = n \land x >= 0 \rightarrow$$

 $[i = 0; r = 0;$
while(i < n){i = i + 1; r = r + i;}
r = r + r - n;
]r = x * x)

How can we prove that the above formula is valid (i.e. satisfied in all states)?

Semantics of DL Sequents

 $\Gamma = \{\phi_1, \dots, \phi_n\}$ and $\Delta = \{\psi_1, \dots, \psi_m\}$ sets of DL formulas where all logical variables occur bound

Recall: $\mathcal{S} \models (\Gamma \Longrightarrow \Delta)$ iff $\mathcal{S} \models (\phi_1 \land \dots \land \phi_n) \rightarrow (\psi_1 \lor \dots \lor \psi_m)$

Define semantics of DL sequents identical to semantics of FOL sequents

Definition (Validity of Sequents over DL Formulas) A sequent $\Gamma \Rightarrow \Delta$ over DL formulas is valid iff

 $\mathcal{S} \models (\Gamma \Longrightarrow \Delta)$ in all states \mathcal{S}

Consequence for program variables

Initial value of program variables implicitly "universally quantified"

SEFM: Java DL

Symbolic Execution of Programs

Sequent calculus decomposes top-level operator in formula. What is "top-level" in a sequential program p; q; r; ?

Symbolic Execution

- Follow the natural control flow when analysing a program
- Values of some variables unknown: symbolic state representation

Example

Compute the final state after termination of

x=x+y; y=x-y; x=x-y;

Symbolic Execution of Programs Cont'd

General form of rule conclusions in symbolic execution calculus

 $\langle \texttt{stmt; rest} \rangle \phi, \qquad [\texttt{stmt; rest}] \phi$

- Rules symbolically execute *first* statement ("active statement")
- Repeated application of such rules corresponds to symbolic program execution

```
Example (symbolicExecution/simpleIf.key,
Demo, active statement only)
```

```
\programVariables {
    int x; int y; boolean b;
}
\problem {
    \<{ if (b) { x = 1; } else { x = 2; } y = 3; }\> y > x
}
```

Symbolic Execution of Programs Cont'd

$$\begin{split} \textbf{Symbolic execution of conditional} \\ \text{if} \ \frac{\Gamma, \text{b} = \textbf{true} \Rightarrow \langle \text{p} \text{; } \text{rest} \rangle \phi, \Delta \qquad \Gamma, \text{b} = \textbf{false} \Rightarrow \langle \text{q} \text{; } \text{rest} \rangle \phi, \Delta \\ \hline \Gamma \Rightarrow \langle \text{if (b) { p } else { q } ; } \text{rest} \rangle \phi, \Delta \end{split}$$

Symbolic execution must consider all possible execution branches

Symbolic execution of loops: unwind

$$\begin{array}{c} \text{unwindLoop} & \overline{\Gamma \Longrightarrow \langle \text{if (b) } \{ \text{ p; while (b) } p \}; \text{ rest} \rangle \phi, \Delta} \\ \hline & \Gamma \Longrightarrow \langle \text{while (b) } \{ p \}; \text{ rest} \rangle \phi, \Delta \end{array}$$

Updates for KeY-Style Symbolic Execution

Needed: a Notation for Symbolic State Changes

- Symbolic execution should "walk" through program in natural forward direction
- Need succint representation of state changes effected by each symbolic execution step
- Want to simplify effects of program execution early
- Want to apply state changes late (to branching conditions and post condition)

We use dedicated notation for state changes: updates

Explicit State Updates

Definition (Syntax of Updates, Updated Terms/Formulas)

If v is program variable, t FOL term type-conformant to v, t' any FOL term, and ϕ any DL formula, then

- $\{v := t\}$ is an update
- $\{v := t\}t'$ is DL term
- $\{v := t\}\phi$ is DL formula

Definition (Semantics of Updates)

State S interprets program variables v with $\mathcal{I}_{S}(v)$ β variable assignment for logical variables in t, define semantics ρ as:

 $\rho_{\beta}(\{\mathtt{v} := t\})(\mathcal{S}) = \mathcal{S}' \text{ where } \mathcal{S}' \text{ identical to } \mathcal{S} \text{ except } \mathcal{I}_{\mathcal{S}'}(\mathtt{v}) = \mathit{val}_{\mathcal{S},\beta}(t)$

Facts about updates $\{v := t\}$

- Update semantics similar to that of assignment
- ▶ Value of update also depends on S and logical variables in t, i.e., β
- Updates are not assignments: right-hand side is FOL term

 $\{\mathbf{x} := n\}\phi$ cannot be turned into assignment (n logical variable)

 $(x=i++;)\phi$ cannot (immediately) be turned into update

Updates are not equations: they change value of v

Computing Effect of Updates (Automated)

Rewrite rules for update followed by ...program variable
$$\begin{cases} x := t \} x & \rightsquigarrow & t \\ \{x := t \} y & \rightsquigarrow & y \end{cases}$$
logical variable $\{x := t \} w & \rightsquigarrow & w$ complex term $\{x := t \} f(t_1, \dots, t_n) \rightsquigarrow f(\{x := t \} t_1, \dots, \{x := t \} t_n)$
(because f is rigid)FOL formula $\begin{cases} \{x := t \} (\phi \& \psi) \rightsquigarrow \{x := t \} \phi \& \{x := t \} \psi \\ \dots \\ \{x := t \} (\forall \tau \ y; \phi) \rightsquigarrow \forall \tau \ y; (\{x := t \} \phi) \end{cases}$ program formulaNo rewrite rule for $\{x := t \} (\langle p \rangle \phi)$

Update rewriting delayed until p symbolically executed

Assignment Rule Using Updates

$\begin{array}{l} \mbox{Symbolic execution of assignment using updates} \\ \\ \mbox{assign} \ \hline \frac{\Gamma \Longrightarrow \{ {\tt x} := t \} \langle {\tt rest} \rangle \phi, \Delta}{\Gamma \Longrightarrow \langle {\tt x} \; = \; {\tt t}; \; {\tt rest} \rangle \phi, \Delta} \end{array}$

- Simple! No variable renaming, etc.
- Works as long as t has no side effects

Demo

updates/assignmentToUpdate.key

Parallel Updates

How to apply updates on updates?

Example

Symbolic execution of

t=x; x=y; y=t;

yields:

{t := x}{x := y}{y := t}

Need to compose three sequential state changes into a single one: parallel updates

Parallel Updates Cont'd

Definition (Parallel Update)

A parallel update is an expression of the form $\{v_1 := r_1 || \cdots || v_n := r_n\}$ where each $\{v_i := r_i\}$ is simple update

- All r_i computed in old state before update is applied
- Updates of all program variables v_i executed simultaneously
- ▶ Upon conflict $v_i = v_j$, $r_i \neq r_j$ later update $(\max\{i, j\})$ wins

Definition (Composition Sequential Updates/Conflict Resolution) $\{v_1 := r_1\}\{v_2 := r_2\} = \{v_1 := r_1 || v_2 := \{v_1 := r_1\}r_2\}$ $\{v_1 := r_1 || \cdots || v_n := r_n\} = \begin{cases} x & \text{if } x \notin \{v_1, \dots, v_n\} \\ r_k & \text{if } x = v_k, x \notin \{v_{k+1}, \dots, v_n\} \end{cases}$

Symbolic Execution with Updates (by Example)

$$\begin{array}{c} x < y \implies x < y \\ \vdots \\ x < y \implies \{x:=y \mid\mid y:=x\} \langle \rangle \; y < x \\ \vdots \\ x < y \implies \{t:=x \mid\mid x:=y \mid\mid y:=x\} \langle \rangle \; y < x \\ \vdots \\ x < y \implies \{t:=x \mid\mid x:=y\} \{y:=t\} \langle \rangle \; y < x \\ \vdots \\ x < y \implies \{t:=x\} \{x:=y\} \langle y=t; \rangle \; y < x \\ \vdots \\ x < y \implies \{t:=x\} \langle x=y; \; y=t; \rangle \; y < x \\ \vdots \\ \Rightarrow \; x < y \implies \langle int \; t=x; \; x=y; \; y=t; \rangle \; y < x \end{array}$$

Parallel Updates Cont'd

Example

symbolic execution of x=x+y; y=x-y; x=x-y; gives

KeY automatically deletes overwritten (unnecessary) updates

Demo

updates/swap2.key

Parallel updates to store intermediate state of symbolic computation

SEFM: Java DL

If you would like to quantify over a program variable ...

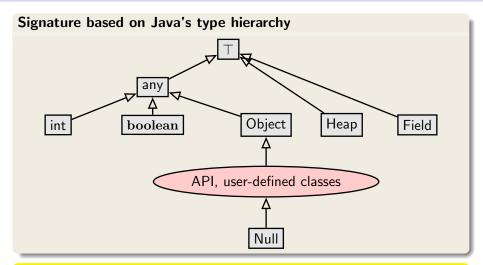
Not allowed: $\forall \tau i; \langle \dots i \dots \rangle \phi$ (program variables \cap logical variables $= \emptyset$)

Instead

Quantify over value, and assign it to program variable:

 $\forall \tau \mathbf{x}; \{ \mathbf{i} := \mathbf{x} \} \langle \dots \mathbf{i} \dots \rangle \phi$

Modelling Java in FOL: Fixing a Type Hierarchy



Each interface and class in API and in target program becomes type with appropriate subtype relation

SEFM: Java DL

CHALMERS/GU

Modelling the Heap in FOL

The Java Heap

Objects are stored on (i.e., in) the heap.

- Status of heap changes during execution
- Each heap associates values to object/field pairs

The Heap Model of KeY-DL

Each element of data type Heap represents a certain heap status. Two functions involving heaps:

in F_Σ: Heap store(Heap, Object, Field, any);
 store(h, o, f, v) returns heap like h, but with v associated to (o, f)

in F_∑: any select(Heap, Object, Field); select(h, o, f) returns value associated to (o, f) in h

Modelling the Heap in FOL

Modelling instance fields

Person		
int int	age id	
	<pre>setAge(int newAge) getId()</pre>	

- ► for each JAVA reference type C there is a type $C \in T_{\Sigma}$, for example, Person
- for each field f there is a unique constant f of type Field, for example, id
- domain of all Person objects: D^{Person}
- a heap relates objects and fields to values

Reading Field id of Person p

FOL notation select(h,p,id)
KeY notation p.id@h (abbreviating select(h,p,id))
p.id (abbreviating select(heap,p,id))^a

^aheap is special program variable for "current" heap; mostly implicit in o.f

Modelling the Heap in FOL

Modelling instance fields

Person		
${f int}{{f int}}$	age id	
	<pre>setAge(int newAge) getId()</pre>	

- For each JAVA reference type C there is a type C ∈ T_Σ, for example, Person
- for each field f there is a unique constant f of type Field, for example, id
- domain of all Person objects: D^{Person}
- a heap relates objects and fields to values

Writing to Field id of Person p

FOL notation store(h, p, id, 6238) KeY notation h[p.id := 6238] (notation for store, not update)

The Algebra of Heaps

We do *not* formalise the *structure* (implementation) of heaps. We formalise the *behaviour*, with an algebra of heap operations:

$$\texttt{select}(\texttt{store}(h, o, f, v), o, f) = v$$

 $(o \neq o' \lor f \neq f') \rightarrow \texttt{select}(\texttt{store}(h, o, f, x), o', f') = \texttt{select}(h, o', f')$

Example

$$\begin{aligned} & \texttt{select}(\texttt{store}(h, \texttt{o}, \texttt{f}, \texttt{15}), \texttt{o}, \texttt{f}) \rightsquigarrow \texttt{15} \\ & \texttt{select}(\texttt{store}(h, \texttt{o}, \texttt{f}, \texttt{15}), \texttt{o}, \texttt{g}) \rightsquigarrow \texttt{select}(h, \texttt{o}, \texttt{g}) \\ & \texttt{select}(\texttt{store}(h, \texttt{o}, \texttt{f}, \texttt{15}), \texttt{u}, \texttt{f}) \rightsquigarrow \\ & \texttt{if} (\texttt{o} = \texttt{u}) \texttt{then} (\texttt{15}) \texttt{else} (\texttt{select}(h, \texttt{u}, \texttt{f})) \end{aligned}$$

Pretty Printing

Shorthand Notations for Heap Operations

o.f@h	is	select(h, o, f)
h[o.f := v]	is	store(h, o, f, v)
therefore:		
u.f@h[o.f := v]	is	select(store(h, o, f, v), u, f)
$\mathtt{h}[\mathtt{o.f}:=\mathtt{v}][\mathtt{o}'.\mathtt{f}':=\mathtt{v}']$	is	store(store(h, o, f, v), o', f', v')

Very-Shorthand Notations for Current Heap

Modelling the Heap in FOL—The Full Story

Is formula select(h, p, id) >= 0 type-safe?

- 1. Return type is any-need to 'cast' to int
- 2. There can be many fields with name id

Real Field Access

int::select(h, p, Person::\$id) >= 0 is type-safe

- int::select is a function name, not a cast
- can be understood intuitively as (int)select

General

For each T typed field f of class C, F_{Σ} contains

- a constant declared as Field C::\$f
- ► a function declared as T T::select(Heap, C, Field)

Everything blue is a function name

Modelling the Heap in FOL—The Full Story

Writing to Fields

We stick to the above:

Declaration: Heap store(Heap, Object, Field, any);

Usage: store(h, p, Person::\$id, 42)

Changing the value of fields

How to translate assignment to field, for example, p.age=17; ?

assign
$$\frac{\Gamma \Longrightarrow \{\texttt{o.f} := t\} \langle \texttt{rest} \rangle \phi, \Delta}{\Gamma \Longrightarrow \langle \texttt{o.f} = \texttt{t}; \texttt{rest} \rangle \phi, \Delta}$$

Admit on left-hand side of update JAVA location expressions

Changing the value of fields

How to translate assignment to field, for example, p.age=17; ?

$$\begin{array}{l} \text{assign} & \frac{\Gamma \Longrightarrow \{\texttt{heap} := \texttt{store}(\texttt{heap},\texttt{p},\texttt{age},\texttt{17})\} \langle \texttt{rest} \rangle \phi, \Delta}{\Gamma \Longrightarrow \langle \texttt{p.age} = \texttt{17}; \ \texttt{rest} \rangle \phi, \Delta} \end{array}$$

Admit on left-hand side of update JAVA location expressions

Changing the value of fields

How to translate assignment to field, for example, p.age=17; ?

assign
$$\frac{\Gamma \Longrightarrow \{p.age := 17\} \langle rest \rangle \phi, \Delta}{\Gamma \Longrightarrow \langle p.age = 17; rest \rangle \phi, \Delta}$$

Admit on left-hand side of update JAVA location expressions

Dynamic Logic: KeY input file

KeY reads in all source files and creates automatically the necessary signature (types, program variables, field constants)

Demo

updates/firstAttributeExample.key

Does abrupt termination count as normal termination? No! Need to distinguish normal and exceptional termination

- ⟨p⟩φ: p terminates normally and formula φ holds in final state (total correctness)
- ▶ [p] φ: If p terminates normally then formula φ holds in final state (partial correctness)

Abrupt termination on top-level counts as non-termination!

Example Reconsidered: Exception Handling

```
\javaSource "path to source code";
```

```
\programVariables {
    ...
}
\problem {
        p != null -> \<{        p.age = 18;   }\> p.age = 18
}
```

Only provable when no top-level exception thrown

Demo

updates/secondAttributeExample.key

Modeling reference this to the receiving object

Special name for the object whose JAVA code is currently executed:

in JML: Object this;

in Java: Object this;

in KeY: Object self;

Default assumption in JML-KeY translation: self != null

How to model object creation with **new**?

Constant Domain Assumption

Assume that domain \mathcal{D} is the same in all states of LTS $\mathcal{K} = (S, \rho)$

Desirable consequence:

Validity of rigid FOL formulas unaffected by programs containing new()

 $\models \forall T x; \phi \rightarrow [p](\forall T x; \phi)$ is valid for rigid ϕ

Object Creation

Realizing Constant Domain Assumption

- Implicitly declared field boolean <created> in class Object
- Equal to true iff argument object has been created
- Object creation modeled as {heap := create(heap, o)} for not (yet) created o (essentially sets <created> field of o to true)
- Normal heap function store "cannot" set value of field <created>

ObjectCreation(simplified)

 $\begin{array}{l} \label{eq:relation} \Gamma, \ \{u\}(\texttt{select}(\texttt{heap},\texttt{ob},\texttt{Object}::<\texttt{created}>) = \texttt{FALSE}) \Longrightarrow \\ \\ \hline \\ \frac{\{u\}(\{\texttt{heap}:=\texttt{create}(\texttt{heap},\texttt{ob})\}\{\texttt{o}:=\texttt{ob}\}\langle\pi\texttt{o}.<\texttt{init}>(\textit{param})\texttt{; }\omega\rangle\phi), \ \Delta \\ \hline \\ \hline \\ \hline \\ \Gamma \Longrightarrow \{u\}(\langle\pi\texttt{ o}=\texttt{new } \texttt{T}(\textit{param})\texttt{; }\omega\rangle\phi), \Delta \end{array}$

ob is a fresh program variable

Object Creation Round Tour Java Programs Arrays Side Effects Abrupt Termination Aliasing Method Calls Null Pointers API

Summary

Literature

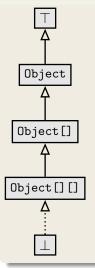
Dynamic Logic to (almost) full Java

KeY supports full sequential Java, with some limitations:

- Limited concurrency
- No generics
- No I/O
- No floats
- No dynamic class loading or reflexion
- ► API method calls: need either JML contract or implementation

Java Features in Dynamic Logic: Arrays

Arrays



- JAVA type hierarchy includes array types that occur in given program (for finiteness)
- Types ordered according to JAVA subtyping rules
- Value of entry in array T[] a; defined in class C depends on reference a to array in C and index
- \blacktriangleright Function arr : int \rightarrow Field injective mapping from indices to fields
- Store array elements on heap, e.g., the value of a[i] on the heap store(heap, a, arr(i), 17) is 17
- Arrays a and b can refer to same object (aliases)
- KeY implements simplification and evaluation rules for array locations

Java Features in Dynamic Logic: Complex Expressions

Complex expressions with side effects

- ► JAVA expressions may contain assignment operator with side effect
- JAVA expressions can be complex, nested, have method calls
- FOL terms have no side effect on the state

Example (Complex expression with side effects in Java)
int i = 0; if ((i=2)>= 2) i++; value of i ?

Complex Expressions Cont'd

Decomposition of complex terms by symbolic execution Follow the rules laid down in JAVA Language Specification

Local code transformations

evalOrderIteratedAssgnmt
$$\frac{\Gamma \Longrightarrow \langle \mathbf{y} = \mathbf{t}; \mathbf{x} = \mathbf{y}; \omega \rangle \phi, \Delta}{\Gamma \Longrightarrow \langle \mathbf{x} = \mathbf{y} = \mathbf{t}; \omega \rangle \phi, \Delta} \quad \mathbf{t} \text{ simple}$$

Temporary variables store result of evaluating subexpression

$$\label{eq:Fval} \begin{array}{c} \Gamma \Longrightarrow \langle \mathbf{boolean} \ \mathbf{v0}; \ \mathbf{v0} = \mathbf{b}; \ \mathbf{if} \ (\mathbf{v0}) \ \mathbf{p}; \ \omega \rangle \phi, \Delta \\ \hline \Gamma \Longrightarrow \langle \mathbf{if} \ (\mathbf{b}) \ \mathbf{p}; \ \omega \rangle \phi, \Delta \end{array} \quad \mathbf{b} \ \mathrm{complex}$$

Guards of conditionals/loops always evaluated (hence: side effect-free) before conditional/unwind rules applied

SEFM: Java DL

Java Features in Dynamic Logic: Abrupt Termination

Abrupt Termination: Exceptions and Jumps Redirection of control flow via return, break, continue, exceptions

 $\langle \pi \operatorname{try} \{p\} \operatorname{catch}(e) \{q\} \operatorname{finally} \{r\} \omega \rangle \phi$

Rules ignore inactive prefix, work on active statement, leave postfix

Rule tryThrow matches try-catch in pre-/postfix and active throw

 $\Rightarrow \langle \pi \text{ if (e instance of T) } \{ try \{ x=e;q \} \text{ finally } \{ r \} \} else \{ r; throw e; \} \omega \rangle \phi$ $\Rightarrow \langle \pi try \{ throw e; p \} \text{ catch}(T x) \{ q \} \text{ finally } \{ r \} \omega \rangle \phi$

Demo

exceptions/try-catch.key

SEFM: Java DL

Java Features in Dynamic Logic: Aliasing

Demo

aliasing/attributeAlias1.key

Reference Aliasing

Naive alias resolution causes proof split (on o = u) at each access

$$\Rightarrow$$
 o.age = 1 \rightarrow (u.age = 2;)o.age = u.age

Java Features in Dynamic Logic: Method Calls

Method Call

First evaluate arguments, leading to:

$$\{\operatorname{arg}_0 := t_0 || \cdots || \operatorname{arg}_n := t_n || c := t_c\} \langle c.m(\operatorname{arg}_0, \ldots, \operatorname{arg}_n); \rangle \phi$$

Actions of rule methodCall

- ▶ for each formal parameter p_i of m: declare and initialize new local variable τ_i p#i =arg_i;
- look up implementation class C of m and split proof if implementation cannot be uniquely determined
- ► create concrete method invocation c.m(p#0,...,p#n)@C

Method Body Expand

- **1.** Execute code that binds actual to formal parameters $\tau_i p \# i = arg_i$;
- 2. Call rule methodBodyExpand

$$\begin{split} & \Gamma \Longrightarrow \langle \pi \; \texttt{method-frame(source=C, this=c) { body } \omega \rangle \phi, \Delta \\ & \Gamma \Longrightarrow \langle \pi \; \texttt{c.m(p#0,...,p#n)@C; } \omega \rangle \phi, \Delta \end{split}$$

Demo

methods/instanceMethodInlineSimple.key

Localisation of Fields and Method Implementation

JAVA has complex rules for localisation of fields and method implementations

- Polymorphism
- Late binding
- Scoping (class vs. instance)
- Context (static vs. runtime)
- Visibility (private, protected, public)

Proof split into cases when implementation not statically determined

Null pointer exceptions

There are no "exceptions" in FOL: \mathcal{I} total on FSym Need to model possibility that o = null in o.a

▶ KeY branches over o != null upon each field access

A Round Tour of Java Features in DL Cont'd

Formal specification of Java API

How to perform symbolic execution when JAVA API method is called?

 API method has reference implementation in JAVA Call method and execute symbolically Problem Reference implementation not always available

Problem Breaks modularity

- 2. Use JML contract of API method:
 - 2.1 Show that requires clause is satisfied
 - 2.2 Obtain postcondition from ensures clause
 - 2.3 Delete updates with modifiable locations from symbolic state

Java Card API in JML or DL

DL version available in KeY, JML work in progress See W. Mostowski

```
http://limerick.cost-ic0701.org/home/
```

```
verifying-java-card-programs-with-key
```

- Most JAVA features covered in KeY
- Several of remaining features available in experimental version
 - Simplified multi-threaded JMM
 - Floats
- Degree of automation for loop-free programs is very high
- Proving loops requires user to provide invariant
 - Automatic invariant generation sometimes possible
- Symbolic execution paradigm lets you use KeY w/o understanding details of logic

Essential

KeY Book Verification of Object-Oriented Software (see course web page), Chapter 10: Using KeY

KeY Book Verification of Object-Oriented Software (see course web page), Chapter 3: Dynamic Logic, Sections 3.1, 3.2, 3.4, 3.5, 3.6.1, 3.6.2, 3.6.3, 3.6.4, 3.6.5, 3.6.7