# Automata, Logic and Games: Theory and Application Higher-Order Model Checking 1 / 2

# Luke Ong

#### University of Oxford http://www.cs.ox.ac.uk/people/luke.ong/personal/

### TACL Summer School University of Salerno, 14-19 June 2015

Interface between (Computer-Aided Formal) Verification and Semantics of Computation

# Logical and Algorithmic Foundations of Verification

- Automata on infinite trees as computational models of state-based systems
- Logical systems for describing correctness properties
- Two-person games as an abstract model of interaction between a reactive system and its environment

# Semantics of Higher-Order Computation

- Lambda calculus as a definitional device
- Game semantics as accurate, intensional model
- Type systems: compositional, syntax-directed inference systems of behavioural properties

Luke Ong (University of Oxford)

Higher-Order Model Checking

Model checking: an approach to program verification that promises accurate analysis with push-button automation.

Verification Problem: Given system *Sys* (e.g. OS), and correctness property *Spec* (e.g. deadlock freedom), does *Sys* satisfy *Spec*?

# The model checking approach:

- Find an abstract model  $\mathcal{M}$  of the system Sys.
- 2 Describe property Spec as a formula  $\varphi$  of a decidable logic.
- **③** Exhaustively check if  $\varphi$  is violated by  $\mathcal{M}$ .

Major progress in verification of 1st-order imperative programs. Many tools: SLAM, Blast, Terminator, SatAbs, etc.

Two key techniques: State-of-the-art tools use

- abstraction refinement techniques: CEGAR (Counter-Example Guided Abstraction Refinement)
- **2** acceleration methods such as SAT- and SMT-solvers.

Examples of Higher-Order / Functional Languages: OCaml, F#, Haskell, Lisp/Scheme, JavaScript, and Erlang; even C++.

# Why Higher-Order / Functional Languages?

- Functional programs are succinct, less error-prone, easy to write and maintain, good for prototyping.
- Lambdas (closures) and streams now standard in today's leading languages (TIOBE Index): Java8, C++11, C#5.0, Javascript, Perl5, Python, Scala.
- FL support domain-specific languages and organise data parallelism well; increasingly prevalent in scientific applications and financial modelling
- Attractive for concurrent programminng (multicore, GPU-processing and cloud computing), thanks to absence of mutable variables and monadic structuring principles

### Two Standard Approaches

### Type-Based Program Analysis.

- sound and scalable but often imprecise ("curse of false positives") E.g. type-and-effect system (region-based memory management), qualifier types, linear types, intersection types, resource usage (sized types), etc.

#### Theorem Proving and Dependent Types

- accurate, typically requires human intervention; does not scale well E.g. Coq, Agda, etc.

- Infinite-state and extremely complex: Even without recursion, higher-order programs over a finite base type are infinite-state.
- Many other sources of infinity: data structures, control structures (with recursion), concurrency, distribution and asynchronous communication, real-time and embedded systems, systems with parameters, etc.
- Models of higher-order features as studied in semantics are typically too "abstract" to support any algorithmic analysis.

A notable exception is game semantics.

Higher-Order Model Checking is the model checking of infinite trees which are generated by recursion schemes (equivalently,  $\lambda \mathbf{Y}$ -calculus), with a view to formally analysing higher-order computation.

### Outline of Part 2

- Relating Families of Generators of Infinite Trees / Graphs: Recursion Schemes and (Collapsible) Pushdown Automata
- Algorithmics and Expressivity
- Seducing Model Checking to Type Inference
- Compositional Model Checking of Higher-Type Böhm Trees
- In Practical Algorithms for Higher-Order Model Checking
- Onclusions and Further Directions

#### A review: Church's simple types

**Types** 
$$\kappa$$
 ::=  $\circ$  |  $(\kappa \rightarrow \kappa')$ 

Every type can be written uniquely as

$$\kappa_1 \to (\kappa_2 \to \dots \to (\kappa_n \to \mathsf{o}) \dots), \quad n \ge 0$$

often abbreviated to  $\kappa_1 \rightarrow \kappa_2 \rightarrow \cdots \rightarrow \kappa_n \rightarrow o$ . Arrows associate to the right.

**Order** of a type: measures "nestedness" on LHS of  $\rightarrow$ .

$$\mathsf{order}(\mathsf{o}) := 0$$
  
 $\mathsf{order}(\kappa \to \kappa') := \max(\mathsf{order}(\kappa) + 1, \mathsf{order}(\kappa'))$ 

**Examples.**  $\mathbb{N} \to \mathbb{N}$  and  $\mathbb{N} \to (\mathbb{N} \to \mathbb{N})$  both have order 1;  $(\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$  has order 2.

**Notation**.  $e: \kappa$  means "expression e has type  $\kappa$ ". Applications associate to the left: write f g h a to mean ((f g) h) a.

Luke Ong (University of Oxford)

An order-*n* recursion scheme = closed, ground-type term definable in order-*n* fragment of  $\lambda$ **Y**-calculus (i.e. simply-typed  $\lambda$ -calculus with recursion and uninterpreted order-1 constant symbols).

We use recursion schemes to define infinite trees.

**Example**: An order-1 recursion scheme. Fix an alphabet of 1st-order constants  $\Sigma = \{ f : o \to o \to o, g : o \to o, a : o \}.$ 

$$G : \left\{ \begin{array}{ccc} S & \to & F \, a \\ F \, x & \to & f \, x \, (F \, (g \, x)) \end{array} \right.$$

Unfolding from the start symbol S:

$$\begin{array}{rcl} S & \rightarrow & F \, a \\ & \rightarrow & f \, a \, (F \, (g \, a)) \\ & \rightarrow & f \, a \, (f \, (g \, a) \, (F \, (g \, (g \, a)))) \\ & \rightarrow & \cdots \end{array}$$

The term-tree thus generated, tree(G), is  $f a (f (g a) (f (g (g a))(\cdots)))$ .

Luke Ong (University of Oxford)

Higher-Order Model Checking

A recursion scheme is a quadruple  $G = \langle \Sigma, \mathcal{N}, \mathcal{R}, S \rangle$  where

- $\Sigma$  is a set of first-order constant symbols (tree constructors); write elements of  $\Sigma$  as  $a, b, \ldots$
- $\mathcal N$  is a set of function symbols; write elements of  $\mathcal N$  as  $F, H, \ldots$
- $S \in \mathcal{N}$  is a distinguished *start symbol* where S: o
- $\mathcal{R}$  is a set of rewrite rules of the form, one for each  $F \in \mathcal{N}$ :

$$F x_1 \cdots x_k \rightarrow e$$

where  $F: \kappa_1 \to \cdots \to \kappa_k \to o$ , and e: o is an applicative term built up from  $\mathcal{N} \cup \{x_1, \cdots, x_k\}$ .

The order of G is the highest order of the function symbols in  $\mathcal{N}$ .

#### **Representing the term-tree** tree(G) as a $\Sigma$ -labelled tree

 $\mathsf{tree}(G) = f\,a\,(f\,(g\,a)\,(f\,(g\,(g\,a))(\cdots)))$  is the term-tree



Formally the term-tree, tree(G), is a map  $T \longrightarrow \Sigma$ , where T is a prefix-closed subset of  $\{1, \dots, m\}^*$ , and m is the maximal arity of symbols in  $\Sigma$ .

tree(G) is ranked and ordered trees.

(Think of tree(G) as the Böhm tree of G.)

#### An Order-3 Example: Fibonacci Numbers

fib generates an infinite spine, with each member (in unary) of the Fibonacci sequence appearing in turn as a left branch from the spine.

Constants:  $b: o \to o \to o, s: o \to o, z: o$ Functions: Write *Church* as a shorthand for  $(o \to o) \to o \to o$ 

$$\begin{array}{rcl}S &: o\\Zero &: Church\\One &: Church\\Show &: Church \rightarrow Church \rightarrow o\\Add &: Church \rightarrow Church \rightarrow (o \rightarrow o) \rightarrow o \rightarrow o\end{array}$$

$$\texttt{fib} \left\{ \begin{array}{ccc} S & \rightarrow & Show \ Zero \ One \\ Zero \ \varphi \ x & \rightarrow & x \\ One \ \varphi \ x & \rightarrow & \varphi \ x \\ Show \ n_1 \ n_2 \ \rightarrow & b \ (n_1 \ s \ z) \ (Show \ n_2 \ (Add \ n_1 \ n_2)) \\ Add \ n_1 \ n_2 \ \varphi \ x & \rightarrow & n_1 \ \varphi \ (n_2 \ \varphi \ x) \end{array} \right.$$



Idea: A word is just a linear tree.

Represent a finite word "a b c" (say) as the applicative term a (b (c e)), viewing a, b and c as  $\Sigma$ -symbols of arity 1, where e is a distinguished nullary end-of-word marker.

- A word language is regular iff it is generated by an order-0 recursion scheme.
- A word language is context-free iff it is generated by an order-1 recursion scheme.

What class of word languages do order-2 recursion schemes define?

#### Order-2 pushdown automata

A 1-stack is an ordinary stack. A 2-stack (resp. (n + 1)-stack) is a stack of 1-stacks (resp. n-stack).

**Operations on 2-stacks**:  $s_i$  ranges over 1-stacks.

Idea extends to all finite orders: an order-n PDA has an order-n stack, and has  $push_i$  and  $pop_i$  for each  $1 \le i \le n$ .

Luke Ong (University of Oxford)

Higher-Order Model Checking

#### **Example:** $L := \{ a^n b^n c^n : n \ge 0 \}$ is recognisable by an order-2 PDA

L is not context free—thanks to the "uvwxy Lemma".

Idea: Use top 1-stack to process  $a^n b^n$ , and height of 2-stack to remember n.



Higher-Order Model Checking

# Some properties of the Maslow Hierarchy (Maslov 74, 76)

- In HOPDA define an infinite hierarchy of word languages.
- Output Context State of the second state of
- So For each n ≥ 0, the order-n languages form an abstract family of languages (closed under +, ·, (−)\*, intersection with regular languages, homomorphism and inverse homo.)
- **④** For each  $n \ge 0$ , the emptiness problem for order-n PDA is decidable.

#### A recent breakthrough

Theorem (Inaba + Maneth FSTTCS08)

All languages of the Maslov Hierarchy are context-sensitive.

Proof uses macro tree transducers (Engelfriet); order-n languages  $\subseteq$  image of n iterates of MTTs

# Theorem (Engelfriet 1991)

Let  $s(n) \ge \log(n)$ .

- (i) For k ≥ 0, word acceptance problem of non-det. order-k PDA augmented with a two-way work-tape with s(n) space is k-EXPTIME complete.
- (ii) For  $k \ge 1$ , the acceptance problem of alternating order-k PDA augmented with a two-way work-tape with s(n) space is (k-1)-EXPTIME complete.
- (iii) For  $k \ge 0$ , the acceptance problem of alternating order-k PDA is k-EXPTIME complete.
- (iv) For  $k \ge 1$ , the emptiness problem of non-det. order-k PDA is (k-1)-EXPTIME complete.

### Theorem (Equi-expressivity)

For each  $n \ge 0$ , the three formalisms

- order-*n* pushdown automata (Maslov 76)
- **a** order-*n* safe recursion schemes (Damm 82, Damm + Goerdt 86)
- order-n indexed grammars (Maslov 76)

generate the same class of word languages.

What is safety?

#### Summary

Higher-order pushdown automata can be used as recognising/generating device for

- **(**) finite-word languages (Maslov 74) and  $\omega$ -word languages
- of such trees
  (KNU01) and, more generally, languages of such trees
- possibly infinite graphs (Muller+Schupp 86, Courcelle 95, Cachat 03), qua configuration graphs of these pushdown systems

Higher-order recursion schemes can also be used to generate word languages, potentially-infinite trees (and languages there of) and graphs.

The two families are closely related.

#### A challenge problem in verification of higher-order computation

**Example**: Consider tree(G) on the right

- $\varphi_1 =$  "Infinitely many *f*-nodes are reachable".
- $\varphi_2 =$  "Only finitely many g-nodes are reachable".

Every node on the tree satisfies  $\varphi_1 \lor \varphi_2$ .

Monadic second-order (MSO) logic can describe properties such as  $\varphi_1 \lor \varphi_2$ .

MSO Model-Checking Problem for Order-n Recursion Schemes

- INSTANCE: An order-n recursion scheme G, and an MSO formula  $\varphi$
- QUESTION: Does the  $\Sigma$ -labelled tree tree(G) satisfy  $\varphi$ ?

Is the above problem decidable?

where  $\Sigma$  is a ranked alphabet.

First-order variables: x, y, z, etc. (ranging over nodes)

Second-order variables: X, Y, Z, etc. (ranging over sets of nodes i.e. monadic relations)

MSO formulas are built up from atomic formulas:

- Parent-child relationship between nodes:  $\mathbf{d}_i(x, y)$ , for  $1 \leq i \leq m$
- Node labelling:  $\mathbf{p}_f(x)$ , for  $f \in \Sigma$
- Set-membership:  $x \in X$

and closed under boolean connectives, first and second-order quantifications.

Because it is the gold standard of logics for describing correctness properties of reactive systems.

• MSO is *very* expressive.

Over graphs, MSO is more expressive than the modal mu calculus, into which all standard temporal logics (e.g. LTL, CTL, CTL\*, etc.) can embed.

• It is hard to extend MSO meaningfully without sacrificing decidability where it holds.

- Rabin 1969: Infinite binary trees and regular trees. "Mother of all decidability results in algorithmic verification."
- Muller and Schupp 1985: Configuration graphs of PDA.
- Caucal 1996 Prefix-recognisable graphs (*e*-closures of configuration graphs of pushdown automata, Stirling 2000).
- Knapik, Niwiński and Urzyczyn (TLCA 2001, FOSSACS 2002): **PushdownTree**<sub>n</sub> $\Sigma$  = Trees generated by order-*n* pushdown automata.

**SafeRecSchTree**<sub>n</sub> $\Sigma$  = Trees generated by order-*n* safe rec. schemes.

• Subsuming all the above: Caucal (MFCS 2002). CaucalTree<sub>n</sub> $\Sigma$  and CaucalGraph<sub>n</sub> $\Sigma$ .

Theorem (KNU-Caucal 2002)

For  $n \ge 0$ , PushdownTree<sub>n</sub> $\Sigma$  = SafeRecSchTree<sub>n</sub> $\Sigma$  = CaucalTree<sub>n</sub> $\Sigma$ ; and they have decidable MSO theories.

Luke Ong (University of Oxford)

Higher-Order Model Checking

#### What is the safety constraint on recursion schemes? (1)

There is another notion of "higher-order" types:

Safe types

$$\begin{array}{ll} D_0 &:= \; \{ \, trees \, \} \\ & \text{Order-0 objects are trees.} \\ D_{i+1} &:= \; \bigcup_{k \geq 0} [\underbrace{D_i \times \cdots \times D_i}_k \to D_i] \\ & \text{Order } (i+1) \text{-objects are functions from (tuples of) order-} i \text{ objects } \\ & \text{to order-} i \text{ objects.} \\ & \text{Define s-order}(t) := i \quad \text{if } t \in D_i \end{array}$$

#### Safe (or "Derived") types

- OI-Hierarchy (Damm 82)
- Higher-level tree transducer (Engelfriet & Vogler 88)

#### What is the safety constraint on recursion schemes? (2)

Safety is a set of constraints on where variables may occur in a term.

Definition (Damm TCS 82, KNU FoSSaCS'02)

No order-k subterm of a safe term can contain free variables of order <k.

# Example (unsafe rule).

$$F:(o \to o) \to o \to o \to o, \ f:o \to o \to o, \ x,y:o.$$

$$F \varphi x y = f \left( F (F \varphi y) y (\varphi x) \right) a$$

The subterm  $F \varphi y$  has order 1, but the free variable y has order 0.

Safety does have an important algorithmic advantage!

# Theorem (KNU 02, Blum + O. TLCA 07, LMCS 09)

Substitution (hence  $\beta$ -reduction) in safe  $\lambda$ -calculus can be safely implemented without renaming bound variables! Hence no fresh names needed!

# Theorem (Expressivity)

- (Schwichtenberg 76) The numeric functions representable by simply-typed λ-terms are multivariate polynomials with conditional.
- (Blum + O. LMCS 09) The numeric functions representable by simply-typed safe  $\lambda$ -terms are the multivariate polynomials.

(See (Blum + O. LMCS 09) for a study on the safe lambda calculus .)

- MSO decidability: Is safety a genuine constraint for decidability? I.e. do trees generated by (arbitrary) recursion schemes have decidable MSO theories?
- Machine characterisation: Find a hierarchy of automata that characterise the expressive power of recursion schemes.
- Expressivity: Is safety a genuine constraint for expressivity?
   I.e. are there inherently unsafe word languages / trees / graphs?
- Graph families:
  - Definition: What is a good definition of "graphs generated by recursion schemes"?
  - Model-checking properties: What are the decidable theories of the graph families?

#### A Tale of Two Higher-Order Systems

<b>Damm's Safe Types (TCS 82)</b> $D_{i+1} := \bigcup_{k \ge 0} [\underbrace{D_i \times \cdots \times D_i}_k \to D_i]$	<b>Church's Simple Types</b> (JSL 40) $\kappa := o \mid \kappa \to \kappa'$
MSO model checking is decidable	<b>Q1</b> ?
Safe RS equi-expressive with HOPDA (Damm 82, KNU 02)	<b>Q2</b> : Equi-expressive with HOPDA++?
	<b>Q3</b> : Are there inherently unsafe word languages / trees / graphs?
Hierarchy is strict (Damm 82)	?
Word languages are context-sensitive (Inaba & Maneth 08)	?

# Theorem (Aehlig, de Miranda + O. TLCA 2005)

Trees generated by order-2 recursion schemes (whether safe or not) have decidable MSO theories.

Theorem (Knapik, Niwinski, Urczyczn + Walukiewicz, ICALP 2005)

Modal mu-calculus model checking problem for homogenously-typed order-2 schemes (whether safe or not) is 2-EXPTIME complete.

#### What about higher orders?

Yes: MSO decidability extends to all orders (O. LICS06).

- (Rabin 69): MSOL and parity tree automata are effectively equi-expressive for tree languages.
- (EJ 91): mu-calculus and alternating parity tree automata (APT) are effectively equi-expressive for tree languages.
- Mu-calculus and MSOL are equi-expressive for tree languages. (JW 96): mu-calculus is the bisimulation-invariant fragment of MSOL.

# Theorem (O. LICS 2006)

For  $n \ge 0$ , the mu-calculus model-checking problem for trees generated by order-n recursion schemes is n-EXPTIME complete. Thus these trees have decidable MSO theories.

 $\lambda(G)$  is the tree-unravelling of the underlying (finite) syntax graph of G.

# Theorem (Transfer)

Given  $\Sigma$ , there is an effective transformation of APT,  $\alpha$ , such that for every HORS G, we have  $\mathcal{B}$  accepts tree(G) iff  $\alpha(\mathcal{B})$  accepts  $\lambda(G)$ .

 $\lambda(G)$  is regular; and APT acceptance problem of regular trees is decidable. Hence:

# Corollary

The modal mu-calculus model checking problem for HORS is decidable.

 $\lambda(G)$  is the tree-unravelling of the underlying (finite) syntax graph of G.

# Theorem (Transfer)

Given  $\Sigma$ , there is an effective transformation of APT,  $\alpha$ , such that for every HORS G, we have  $\mathcal{B}$  accepts tree(G) iff  $\alpha(\mathcal{B})$  accepts  $\lambda(G)$ .

### Extension to infintary HORS

Given finite sets  $\mathcal{T}$  and  $\mathcal{V}$  of types and variables respecively,  $G = \langle \Sigma, \mathcal{N}, \mathcal{R}, S \rangle \in \mathbf{RS}^{\infty}(\mathcal{T}, \mathcal{V})$  just if types of all subterms are in  $\mathcal{T}$ , and rules may only use variables from  $\mathcal{V}$ , but  $\mathcal{N}$  and  $\mathcal{R}$  may be infinite.

# Theorem (O. 2013)

Given  $\Sigma, \mathcal{T}, \mathcal{V}$ , there is an effective transformation of APT,  $\alpha$ , s.t. for every  $G \in \mathbf{RS}^{\infty}(\mathcal{T}, \mathcal{V})$ , we have  $\mathcal{B}$  accepts tree(G) iff  $\alpha(\mathcal{B})$  accepts  $\lambda(G)$ .

Cf. Same result in [Salvati & Walukiewicz 13] but they use infinitary  $\lambda Y\text{-calculus.}$ 

Luke Ong (University of Oxford)

 $\lambda(G)$  is the tree-unravelling of the underlying (finite) syntax graph of G.

# Theorem (Transfer)

Given  $\Sigma$ , there is an effective transformation of APT,  $\alpha$ , such that for every HORS G, we have  $\mathcal{B}$  accepts tree(G) iff  $\alpha(\mathcal{B})$  accepts  $\lambda(G)$ .

### Proof Idea. Two key ingredients:

APT  $\mathcal{B}$  has accepting run-tree over tree(G)

↔ { I. Traversal-Path Correspondence}

APT  $\mathcal{B}$  has accepting traversal-tree over  $\lambda(G)$ 

 $\iff \{ \text{ II. Simulation of traversals by paths } \}$ APT  $\alpha(\mathcal{B})$  has an accepting run-tree over  $\lambda(G)$ 

#### Transference principle, based on a theory of traversals

$$G: \left\{ \begin{array}{cccc} S &=& FH \\ F\varphi &=& \varphi(F\varphi) \\ Hz &=& fzz \end{array} \right. \mapsto \overline{G}: \left\{ \begin{array}{cccc} S &=& \lambda.@F(\lambda x.@H\lambda.x) \\ F &=& \lambda\varphi.\varphi(\lambda.@F(\lambda y.\varphi(\lambda.y)))) \\ H &=& \lambda z.f(\lambda.z)(\lambda.z) \end{array} \right. \\ \underbrace{\mathsf{tree}(G) \\ f & f & f & f \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \ddots & \ddots & \ddots & \ddots \\ \varphi & \varphi & \varphi \\ \lambda & \lambda \\ \vdots & y \end{array} \right. \left. \begin{array}{c} \lambda (G) \\ \lambda (G)$$

Idea:  $\beta$ -reduction is global (i.e. substitution changes the term being evaluated); game semantics gives an equivalent but local view. A traversal (over the computation tree  $\lambda(G)$ ) is a trace of the local computation that produces a path (over tree(G)).

### Theorem (Path-traversal correspondence)

Let G be an order-n recursion scheme.

- (i) There is a 1-1 correspondence between maximal paths p in (Σ-labelled) generated tree tree(G) and maximal traversals t<sub>p</sub> over computation tree λ(G).
- (ii) Further for each p, we have  $p = t_p \upharpoonright \Sigma$ .

Proof is by game semantics.

#### Explanation (for game semanticists):

- Term-tree tree(G) is (a representation of) the game semantics of G.
- Paths in tree(G) correspond to P-views in the strategy-denotation.
- Traversals t<sub>p</sub> over computation tree λ(G) are just (representations of) the uncoverings of the P-views (= path) p in the game semantics of G.