

Theories of Information Hiding in Lambda-Calculus

*Logical Relations and Bisimulations
for Encryption and Type Abstraction*

Eijiro Sumii

University of Pennsylvania



Main Results

- ◆ **Proof methods for two forms of information hiding in computer programs**
 - **Logical relations for perfect encryption**
 - **Bisimulations for perfect encryption**
 - **Bisimulations for type abstraction**
 - **First solution to a problem of 20 years**

Background

- ◆ Information hiding (or abstraction) is crucial for building large systems
 - ...including computer software!
- ◆ Type abstraction is the primary method of information hiding in programming languages
 - The basis of objects, modules, components, etc.

A Classical Example

```
interface Complex =  
  type t  
  fun make : real ' real ® t  
  fun mul : t ' t ® t  
  fun re : t ® real
```

Cartesian Implementation

```
module Cartesian : Complex =  
  type t = real ^ real  
  fun make(x,y) = (x,y)  
  fun mul((x1,y1), (x2,y2)) =  
    (x1*x2-y1*y2 , x1*y2+x2*y1)  
  fun re(x,y) = x
```

Polar Implementation

```
module Polar : Complex =  
  type t = real * real  
  fun make(x,y) =  
    (sqrt(x*x+y*y), atan2(y,x))  
  fun mul((r1,q1), (r2,q2)) =  
    (r1*r2, q1+q2)  
  fun re(r,q) = r*cos(q)
```

The Abstraction Property

Contextual equivalence:

Cartesian \circ **Polar** : **Complex**

I.e., the different implementations give the same result under any well-typed context in the language

- In this study, "result" means only the final value (or possible divergence)
 - Time, energy, rounding errors, etc. are out of scope

How to Prove it?

Logical relations [Reynolds 83]:

Induction on the type of the interface
(**Complex** in previous example)

◆ So far, so good.

The Problems

- ◆ Logical relations become complex for expressive languages (e.g. with recursion or concurrency)
 - ...yet attackers must be expressed in the language (as well as users)
- ◆ Type abstraction doesn't work against untyped users/attackers
 - ...but we cannot "type-check the Internet"

Our Solutions

- ◆ Use bisimulations in place of logical relations
- ◆ Use encryption instead of type abstraction

Outline

- 1. Logical relations for type abstraction**
[Reynolds 83] [Mitchell 91]
- 2. Bisimulations for type abstraction**
[Sumii-Pierce, POPL'05]
- 3. Bisimulations for encryption**
[Sumii-Pierce, POPL'04 & TCS]
 - **Cf. Logical relations for encryption**
[Sumii-Pierce, CSFW'01 & JCS]

Logical Relations (without Type Abstraction)

Relations between programs, defined by induction on their types

- ◆ **Constants are related iff they are equal**
- ◆ **Tuples are related iff their elements are related**
- ◆ **Functions are related iff they map related arguments to related results**

Examples

- ◆ 123 and 123 are related at type Int
- ◆ $\lambda x. x+1+2$ and $\lambda x. x+3$ are related at type $\text{Int} \textcircled{\text{R}} \text{Int}$
- ◆ $\lambda x. (123, x+1+2)$ and $\lambda x. (123, x+3)$ are related at type $\text{Int} \textcircled{\text{R}} (\text{Int} \text{ ' } \text{Int})$

Logical Relations for Type Abstraction

In addition to the previous cases:

- ◆ Abstract data of type a are related iff they are related by $j(a)$
 - where j is a relation environment mapping abstract types to the relation between their implementations

Example

Take

$$j(\text{Complex.t}) = \{((\mathbf{x}, \mathbf{y}), (\mathbf{r}, \mathbf{q})) \mid \mathbf{x} = r \cos q, \mathbf{y} = r \sin q\}$$

Then

j **Cartesian** ~ **Polar** : **Complex**

That is, **Cartesian** and **Polar** are
logically related at type **Complex** under j

Soundness of Logical Relations

Logical relations imply contextual equivalence

- ◆ **Corollary of the "fundamental property" (a.k.a. the "basic lemma")**
 - **Proved by induction on the typing of terms, with assumptions on their free variables**

Shortcoming of Logical Relations

Don't "scale" to more expressive languages

- ◆ **Recursive functions (or while-loops) complicate the fundamental property**
- ◆ **Recursive types complicate the definition of logical relations**

**Extra work required in every use of logical relations
(cannot be done once and for all in meta theory)**

Outline

1. **Logical relations for type abstraction**
[Reynolds 83] [Mitchell 91]
2. **Bisimulations for type abstraction**
[Sumii-Pierce, POPL'05]
3. **Bisimulations for encryption**
[Sumii-Pierce, POPL'04 & TCS]

Bisimulations (without Type Abstraction)

Applicative bisimulations [Abramsky 90]:

Relations between values that satisfy some conditions to exclude inequivalent values

- Bisimilar constants are equal
- Bisimilar tuples have bisimilar elements
- Bisimilar functions return bisimilar results when applied to the same argument

Examples

- ◆ $\{(123, 123)\}$ is a bisimulation
- ◆ $\{(4, 4), (5, 5)\}$ is a bisimulation
- ◆ $\{((4, 5), (4, 5)), (4, 4), (5, 5)\}$ is a bisimulation
- ◆ $\{(lx.x+1+2, lx.x+3), (i, j) \mid i = k+1+2, j = k+3, k : \text{int}\}$ is a bisimulation
- ◆ Union of bisimulations is a bisimulation

Shortcoming of Applicative Bisimulations

Don't extend to type abstraction

- ◆ **Cartesian.re** and **Polar.re** do not return bisimilar results (i.e. the same real number) when applied to the same argument
 - They expect different representations

Bisimulations for Type Abstraction: First Try

**Bisimilar functions return bisimilar results
when applied to bisimilar arguments**

- No condition for abstract data themselves, as long as the other conditions are satisfied

THIS IS UNSOUND!

- ◆ Because contexts (users or attackers) can combine bisimilar values to make more complex arguments

Counter-Example

- ◆ $(1, \text{fst})$ and $(2, \text{fst})$ are not contextually equivalent at type $a' (a' a \textcircled{R} \text{int})$

But

$\{((1, \text{fst}), (2, \text{fst}), a' (a' a \textcircled{R} \text{int})),$
 $(1, 2, a),$
 $(\text{fst}, \text{fst}, a' a \textcircled{R} \text{int})\}$

satisfies all the bisimulation conditions so far!

Bisimulations for Type Abstraction: Second Try

Bisimilar functions return bisimilar results

when applied to $C[v_1, \dots, v_n]$ and $C[v_1', \dots, v_n']$

- ◆ **for any bisimilar values v_1, \dots, v_n and v_1', \dots, v_n' ,
and**
- ◆ **for any value context C .**

Example

$$\mathbf{R} = \{ (\text{Cartesian}, \text{Polar}, \text{Complex}), \\ (\text{Cartesian.make}, \text{Polar.make}, \\ \text{real} \text{ } \dot{\text{real}} \text{ } \textcircled{\mathbb{R}} \text{Complex.t}), \\ (\text{Cartesian.mul}, \text{Polar.mul}, \\ \text{Complex.t} \text{ } \dot{\text{Complex.t}} \text{ } \textcircled{\mathbb{R}} \text{Complex.t}), \\ (\text{Cartesian.re}, \text{Polar.re}, \\ \text{Complex.t} \text{ } \textcircled{\mathbb{R}} \text{real}), \\ ((\mathbf{x}, \mathbf{y}), (\mathbf{r}, \mathbf{q}), \text{Complex.t}), \\ (\mathbf{z}, \mathbf{z}, \text{real}) \mid \mathbf{x} = \mathbf{r} \cos \mathbf{q}, \mathbf{y} = \mathbf{r} \sin \mathbf{q} \}$$

The Last Problem

The previous definition is sound,
but completeness is unclear

- ◆ Because the union of two bisimulations wouldn't always be a bisimulation
 - Counter-example: the union $R \dot{\cup} R^{-1}$ of the previous R and its inverse R^{-1}
- ⊢ Standard co-inductive method wouldn't apply

Our Solution

Consider sets of relations as bisimulations

- ◆ **Intuition: Each relation represents a "world"**
- ◆ **Another intuition: Each relation represents the knowledge of an attacker, which increases by time (but nevertheless stays in the bisimulation)**
- ◆ **Also gives a natural account for the generativity of existential types**

Examples

For the previous R ,

- ◆ $\{ R \}$ is a bisimulation
- ◆ $\{ R^{-1} \}$ is another bisimulation
- ◆ $\{ R, R^{-1} \}$ is also a bisimulation
- ◆ $\{ R \dot{\cup} R^{-1} \}$ is not a bisimulation

Soundness and Completeness

- ◆ Contextual equivalence is also generalized as a set of relations
- ◆ Then, it coincides with bisimilarity (the largest bisimulation)
 - Everything is formalized and proved in λ -calculus with full universal, existential, and recursive types (first result in 20 years!)

Other Examples

- ◆ **Object encodings**
- ◆ **ML-like functors**
- ◆ **Higher-order polymorphic functions**
(the "dual" of abstract types)

Outline

1. **Logical relations for type abstraction**
[Reynolds 83] [Mitchell 91]
2. **Bisimulations for type abstraction**
[Sumii-Pierce, POPL'05]
3. **Bisimulations for encryption**
[Sumii-Pierce, POPL'04 & TCS]

Idea: Abstraction by Encryption

Reinvention of dynamic sealing [Morris 73]

- ◆ Secret key is generated for each abstract type
- ◆ Abstract data are encrypted when exported out of a module
- ◆ ...and decrypted when imported back

Cartesian Implementation of Complex with Encryption

```
module Cartesian =  
  fun make(x,y) = encryptk(x,y)  
  fun mul(c1,c2) =  
    let (x1,y1) = decryptk(c1) in  
    let (x2,y2) = decryptk(c2) in  
    encryptk(x1*x2-y1*y2 , x1*y2+x2*y1)  
  fun re(c) =  
    let (x,y) = decryptk(c) in x
```

Polar Implementation of Complex with Encryption

```
module Polar =  
  fun make(x,y) =  
    encryptk.(sqrt(x*x+y*y),atan2(y,x))  
  fun mul(c1,c2) =  
    let (r1,q1) = decryptk.(c1) in  
    let (r2,q2) = decryptk.(c2) in  
    encryptk.(r1*r2,q1+q2)  
  fun re(c) =  
    let (r,q) = decryptk.(c) in r*cos(q)
```

The Abstraction Property

Untyped contextual equivalence:

Cartesian \circ **Polar**

- ◆ Abstraction holds against any context in the language, even if untyped

How to Prove it?

Bisimulations!

- ◆ Conditions for constants, tuples, functions are the same as before

Bisimulations for Encryption

- ◆ Bisimulation respects equality of keys
 - I.e., if k_1 and k_1' are bisimilar, and if k_2 and k_2' are bisimilar, then $k_1 = k_2 \hat{=} k_1' = k_2'$
- ◆ For any bisimilar ciphertexts $\text{encrypt}_k(v)$ and $\text{encrypt}_{k'}(v')$,
 - Neither k nor k' is in the relation, or
 - v and v' are bisimilar

Example

$$\mathbf{R} = \{ (\text{Cartesian}, \text{Polar}),$$
$$(\text{Cartesian.make}, \text{Polar.make}),$$
$$(\text{Cartesian.mul}, \text{Polar.mul}),$$
$$(\text{Cartesian.re}, \text{Polar.re}),$$
$$(\text{encrypt}_k(x, y), \text{encrypt}_k(r, q)),$$
$$(\mathbf{z}, \mathbf{z}) \mid$$
$$\mathbf{x} = \mathbf{r} \cos q, \mathbf{y} = \mathbf{r} \sin q, \mathbf{z} : \text{real} \}$$

Formalization

- ◆ Defined untyped λ -calculus extended with:
 - Keys k and fresh key generation $\text{new } x \text{ in } e$
 - Encryption $\{e_1\}_{e_2}$ and decryption $\text{let } \{x\}_{e_1} = e_2 \text{ in } e_3 \text{ else } e_4$
 - Assumes perfect encryption
- ◆ Proved soundness and completeness of our bisimulations in this language

Operational Semantics of the Language (1/2)

- ◆ Big-step evaluation $(s)e \beta (t)v$ from terms e to values v
 - Annotated with the set of keys s and t before and after the evaluation

$$\frac{k \notin s \quad (s \setminus \{k\}) [k/x]e \beta (t)v}{(s) \text{ new } x \text{ in } e \beta (t)v}$$

Operational Semantics of the Language (2/2)

- ◆ Success of decryption:

$$\frac{(s)e_1 \beta (s_1)k \quad (s_1)e_2 \beta (s_2)\{v\}_k}{(s_2) [v/x]e_3 \beta (t)w}$$

$$(s) \text{ let } \{x\}_{e_1} = e_2 \text{ in } e_3 \text{ else } e_4 \beta (t)w$$

- ◆ Failure of decryption:

$$\frac{(s)e_1 \beta (s_1)k \quad (s_1)e_2 \beta (s_2)\{v\}_{k'} \quad k \neq k'}{(s_2)e_4 \beta (t)w}$$

$$(s) \text{ let } \{x\}_{e_1} = e_2 \text{ in } e_3 \text{ else } e_4 \beta (t)w$$

Other Examples

- ◆ Generative functors
- ◆ Non-generative functors
- ◆ **Encodings of security protocols**

Our Protocol Encoding

- ◆ A protocol is encoded as a tuple of participants (and their public keys)
 - Senders are encoded as the values being sent
 - Receivers are encoded as functions from received values to returned values
- ◆ Then, contexts play the role of the network, scheduler, and attackers by applying the receivers to the senders

Example

1. $A \circledR B : \{N\}_K$
2. $B \circledR A : N \bmod 2$

$\text{Sys}_N = \text{new } x \text{ in}$
 $(\text{encrypt}_x(N),$
 $\quad | y. (\text{decrypt}_x(y) \bmod 2))$

Non-Interference: Secrecy as Equivalence

With our bisimulations, it is easy to prove

$$\text{Sys}_M \circ \text{Sys}_N$$

for any M and N with

$$M \bmod 2 = N \bmod 2$$

Which means:

Sys_N keeps N secret (except for its least significant bit) under any context

Example

$$\mathbf{R} = \{ (\mathbf{Sys}_M, \mathbf{Sys}_N), \\ (\mathbf{encrypt}_k(M), \mathbf{encrypt}_{k'}(N)), \\ (l y. (\mathbf{decrypt}_k(y) \bmod 2), \\ l y. (\mathbf{decrypt}_{k'}(y) \bmod 2)), \\ (\mathbf{M} \bmod 2, \mathbf{N} \bmod 2) \}$$

Protocols Encoded and Proved (or Disproved)

- ◆ Needham-Schroeder (**insecure**)
- ◆ Needham-Schroeder-Lowe (**secure**)
- ◆ "ffgg" protocol [Millen 99] (**insecure**)
 - Attack is "necessarily parallel" but can be simulated in λ -calculus via interleaving

Various properties (such as integrity) can be checked as long as expressed as equivalence

Needham-Schroeder-Lowe Protocol

1. $B \rightarrow A : B$
2. $A \rightarrow B : \{N_A, A\}_{k_B}$
3. $B \rightarrow A : \{N_A, N_B, B\}_{k_A}$
4. $A \rightarrow B : \{N_B\}_{k_B}$
5. $B \rightarrow A : \{i\}_{N_B}$

Encoding of Needham-Schroeder-Lowe Protocol

$$W = \langle \lambda x. \{x\}_{k_A}, \lambda x. \{x\}_{k_B}, k_E, U, V \rangle$$

$$U = \langle B, \lambda \{ \langle x, y \rangle \}_{k_B}. \text{assert}(y = A); \\ \nu z. \langle \{ \langle x, z, B \rangle \}_{k_A}, \\ \lambda \{ z_0 \}_{k_B}. \text{assert}(z_0 = z); \\ \{i\}_z \rangle \rangle$$

$$V = \lambda x. \text{let } k_x = (\text{if } x = B \text{ then } k_B \text{ else} \\ \text{if } x = E \text{ then } k_E \text{ else } \perp) \text{ in} \\ \nu y. \langle \{ (y, A) \}_{k_x}, \\ \lambda \{ \langle y_0, z, x_0 \rangle \}_{k_A}. \text{assert}(y_0 = y); \\ \text{assert}(x_0 = x); \\ \{z\}_{k_x} \rangle$$

Bisimulation for Needham-Schroeder-Lowe Protocol

$$\begin{aligned}
 & \{(U, U'), (V, V'), (W, W')\}, \\
 & (\bar{k}, \bar{k}'), (A, A), (B, B), (E, E), (\lambda x. \{x\}_{k_A}, \lambda x. \{x\}_{k'_A}), (\lambda x. \{x\}_{k_B}, \lambda x. \{x\}_{k'_B}), (k_E, k'_E), (\bar{w}, \bar{w}'), (\{\bar{w}\}_{k_A}, \{\bar{w}'\}_{k'_A}), (\{\bar{w}\}_{k_B}, \{\bar{w}'\}_{k'_B}), \\
 & (\lambda \{(x, y)\}_{k_B}. \text{assert}(y = A); \nu z. (\{(x, z, B)\}_{k_A}, \lambda \{z_0\}_{k_B}. \text{assert}(z_0 = z); \{i\}_{\bar{k}_B}), \lambda \{(x, y)\}_{k'_B}. \text{assert}(y = A); \nu z. (\{(x, z, B)\}_{k'_A}, \lambda \{z_0\}_{k'_B}. \text{assert}(z_0 = z); \{j\}_{\bar{k}'_B})), \\
 & (\{(\bar{k}_{AB}, A)\}_{k_B}, \lambda \{(y_0, z, x_0)\}_{k_A}. \text{assert}(y_0 = \bar{k}_{AB}); \text{assert}(x_0 = B); \{z\}_{k_B}), \{(\bar{k}'_{AB}, A)\}_{k'_B}, \lambda \{(y_0, z, x_0)\}_{k'_A}. \text{assert}(y_0 = \bar{k}'_{AB}); \text{assert}(x_0 = B); \{z\}_{k'_B}), \\
 & (\{(\bar{k}_{AB}, A)\}_{k_B}, \{(\bar{k}'_{AB}, A)\}_{k'_B}), \\
 & (\lambda \{(y_0, z, x_0)\}_{k_A}. \text{assert}(y_0 = \bar{k}_{AB}); \text{assert}(x_0 = B); \{z\}_{k_B}, \lambda \{(y_0, z, x_0)\}_{k'_A}. \text{assert}(y_0 = \bar{k}'_{AB}); \text{assert}(x_0 = B); \{z\}_{k'_B}), \\
 & (\{(\bar{k}_{AB}, \bar{k}_B, B)\}_{k_A}, \lambda \{z_0\}_{k_B}. \text{assert}(z_0 = \bar{k}_{AB}); \{i\}_{\bar{k}_B}), \{(\bar{k}'_{AB}, \bar{k}'_B, B)\}_{k'_A}, \lambda \{z_0\}_{k'_B}. \text{assert}(z_0 = \bar{k}'_{AB}); \{j\}_{\bar{k}'_B}), \\
 & (\{(\bar{k}_{AB}, \bar{k}_B, B)\}_{k_A}, \{(\bar{k}'_{AB}, \bar{k}'_B, B)\}_{k'_A}), \\
 & (\lambda \{z_0\}_{k_B}. \text{assert}(z_0 = \bar{k}_{AB}); \{i\}_{\bar{k}_B}, \lambda \{z_0\}_{k'_B}. \text{assert}(z_0 = \bar{k}'_{AB}); \{j\}_{\bar{k}'_B}), \\
 & (\{\bar{k}_B\}_{k_B}, \{\bar{k}'_B\}_{k'_B}), \\
 & (\{i\}_{\bar{k}_B}, \{j\}_{\bar{k}'_B}), \\
 & (\{(\bar{k}_{AE}, A)\}_{k_B}, \lambda \{(y_0, z, x_0)\}_{k_A}. \text{assert}(y_0 = \bar{k}_{AE}); \text{assert}(x_0 = E); \{z\}_{k_E}), \{(\bar{k}'_{AE}, A)\}_{k'_B}, \lambda \{(y_0, z, x_0)\}_{k'_A}. \text{assert}(y_0 = \bar{k}'_{AE}); \text{assert}(x_0 = E); \{z\}_{k'_E}), \\
 & (\{(\bar{k}_{AE}, A)\}_{k_B}, \{(\bar{k}'_{AE}, A)\}_{k'_B}), \\
 & (\lambda \{(y_0, z, x_0)\}_{k_A}. \text{assert}(y_0 = \bar{k}_{AE}); \text{assert}(x_0 = E); \{z\}_{k_E}, \lambda \{(y_0, z, x_0)\}_{k'_A}. \text{assert}(y_0 = \bar{k}'_{AE}); \text{assert}(x_0 = E); \{z\}_{k'_E}), \\
 & (\{(\bar{k}_{AE}, A), (\bar{k}'_{AE}, A)\}), \\
 & (\bar{k}_{AE}, \bar{k}'_{AE}), \\
 & (\{(\bar{w}, \bar{k}_B, B)\}_{k_A}, \lambda \{z_0\}_{k_B}. \text{assert}(z_0 = \bar{k}_B); \{i\}_{\bar{k}_B}), \{(\bar{w}', \bar{k}'_B, B)\}_{k'_A}, \lambda \{z_0\}_{k'_B}. \text{assert}(z_0 = \bar{k}'_B); \{j\}_{\bar{k}'_B}), \\
 & (\{(\bar{w}, \bar{k}_B, B)\}_{k_A}, \{(\bar{w}', \bar{k}'_B, B)\}_{k'_A}), \\
 & (\lambda \{z_0\}_{k_B}. \text{assert}(z_0 = \bar{k}_B); \{i\}_{\bar{k}_B}, \lambda \{z_0\}_{k'_B}. \text{assert}(z_0 = \bar{k}'_B); \{j\}_{\bar{k}'_B}), \\
 & (\{\bar{w}\}_{k_B}, \{\bar{w}'\}_{k'_B})
 \end{aligned}$$

Outline

- 1. Logical relations for type abstraction**
[Reynolds 83] [Mitchell 91]
- 2. Bisimulations for type abstraction**
[Sumii-Pierce, POPL'05]
- 3. Bisimulations for encryption**
[Sumii-Pierce, POPL'04 & TCS]

Related Work (1/5): Logical Relations

- ◆ **Semantic logical relations**
[Tait 67, Plotkin 73, Reynolds 83, Mitchell 93, etc.]
 - **Suffers from the complexity and imprecision of denotational semantics for recursion**
- ◆ **Syntactic logical relations**
[Pitts 98, Birkedal-Harper 97, etc.]
 - **Still suffers from complications for recursive functions/types**

Related Work (2/5): Applicative Bisimulations

- ◆ For untyped λ -calculus [Abramsky 90]
- ◆ For object calculi with universal and subtyping polymorphism [Gordon-Rees]

**None can deal with type abstraction
(i.e., existential polymorphism)**

Related Work (3/5): Bisimulations for μ -Calculi

- ◆ **For polymorphic μ -calculus**
[Pierce-Sangiorgi 97, Berger-Honda-Yoshida 03]
- ◆ **For spi-calculus**
[Abadi-Gordon 98, Boreale-DeNicola-Pugliese 99, Borgstrom-Nestmann 02, Abadi-Fournet 01, etc.]

**Incomplete, or completeness claimed
but proof unpublished (or found wrong)**

Related Work (4/5): Operational Models of Types

- ◆ Indexed models [Appel et al.]
- ◆ Operational ideal models [Voillon-Melliès 04, etc.]

**Only unary case (safety) considered;
Binary case (equivalence) left open**

- Non-trivial in the presence of existential types

Related Work (5/5)

- ◆ **Categorical reformulation of our logical relations for perfect encryption**
[Goubault-Larrecq-Lasota-Nowak-Zhang, CSL 04]

Future Directions (1/3)

- ◆ **Translations between different forms of information hiding**
 - E.g. from type abstraction to encryption
 - Challenge: How to prove full abstraction (preservation of equivalence)?
- ◆ **Extension to even more expressive languages (e.g. higher-order π -calculus)**

Future Directions (2/3)

- ◆ **Adaptation/generalization for other forms of information hiding (e.g. security typing)**
 - Generative security levels?
- ◆ **A system allowing unchecked, dynamically checked, and statically typed code without losing abstraction**
 - C, Perl, ML, etc. coexist in peace?

Future Directions (3/3)

