

An Abstract Semantics for Inference of Types and Effects in a Multi-Tier Web Language

Letterio Galletta Giorgio Levi

Dipartimento di Informatica, Università di Pisa
{galletta, levi}@di.unipi.it

Workshop on Automated Specification and Verification of Web
Systems, 2011



Multi-tier architecture

Standard web applications have a multi-tier architecture



Each tier runs on a different computational environment characterized by its language and its data representation (**impedance mismatch problem**)



LINKS: a Multi-Tier Web Language

Multi-tier web languages allow one to blend server, client and database code and provide automatic mechanisms for the partition of the application over tiers



LINKS: a Multi-Tier Web Language

Multi-tier web languages allow one to blend server, client and database code and provide automatic mechanisms for the partition of the application over tiers

LINKS

LINKS is a functional multi-tier web language

- from a single source code the compiler generates code for each tier
- support an unified cross-tier programming model by exploiting **web continuations**



LINKS: a Multi-Tier Web Language

Multi-tier web languages allow one to blend server, client and database code and provide automatic mechanisms for the partition of the application over tiers

LINKS

LINKS is a functional multi-tier web language

- from a single source code the compiler generates code for each tier
- support an unified cross-tier programming model by exploiting **web continuations**

Web continuations in LINKS

Closures (expression to be executed plus bindings of free variables) stored in HTML pages



LINKS: security

Baltopoulos and Gordon have shown that

- storing web continuation in HTML page is not secure
- an attacker can violate
 1. Secrecy
 2. Data Integrity
 3. Control Integrity



LINKS: security

Solution

To overtake the security issues they have proposed a secure implementation that includes

1. a compilation strategy based on authenticated encryption
2. a types-and-effects system to enable source level reasoning about security of web applications



LINKS: security

Solution

To overtake the security issues they have proposed a secure implementation that includes

1. a compilation strategy based on authenticated encryption
2. a types-and-effects system to enable source level reasoning about security of web applications

The secure implementation has been formalized for TINYLINKS, a λ -calculus augmented with

1. XML values for representing web pages
2. `event e assert` annotation for expressing safety properties

TINYLINKS

Syntax

f, y, x

P

$c ::= \text{Unit} \mid \text{Zero} \mid \text{Succ} \mid \text{String}$
 $\mid \text{Nil} \mid \text{Cons} \mid \text{Tuple} \mid \text{Elem} \mid \text{Text}$

$g ::= + \mid - \mid * \mid /$

$L ::= p(V_1, \dots, V_n)$

$V, U ::= x \mid c(V_1, \dots, V_n) \mid \text{href}(E)$
 $\mid \lambda x_1. \dots, x_n. E \mid \text{form}([l_1, \dots, l_n], E)$

$E ::= V \mid \text{var } x = E_1; E_2 \mid g(E_1, E_2)$
 $\mid V(U_1, \dots, U_n) \mid \text{post}([l_1 = V_1, \dots, l_n = V_n], U)$
 $\mid \text{get}(V) \mid \text{event } L \mid \text{assert } L$
 $\text{switch}(V)\{$
 $\quad \text{case } c(x_1, \dots, x_n) \rightarrow E_1$
 $\quad _ \rightarrow E_2$
 $\}$

Types-and-effects system

A powerful extension of type systems which allows one to statically reason about program's execution

$$\frac{\Gamma \vdash M_1 : \tau_1 \& \phi_1 \quad \dots \quad \Gamma \vdash M_n : \tau_n \& \phi_n}{\Gamma \vdash E(M_1, \dots, M_n) : \tau \& \phi}$$

Types-and-effects system

A powerful extension of type systems which allows one to statically reason about program's execution

premise

$\Gamma \vdash M_1 : \tau_1 \ \& \ \phi_1 \quad \dots \quad \Gamma \vdash M_n : \tau_n \ \& \ \phi_n$

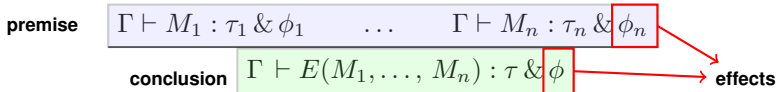
conclusion

$\Gamma \vdash E(M_1, \dots, M_n) : \tau \ \& \ \phi$

effects

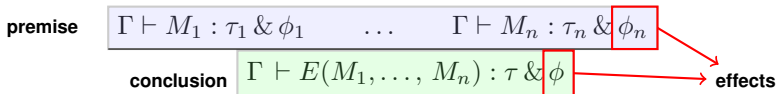
Types-and-effects system

A powerful extension of type systems which allows one to statically reason about program's execution



Types-and-effects system

A powerful extension of type systems which allows one to statically reason about program's execution



It compute for each program phrase its type augmented with a semantic property

- if $\Gamma \vdash M_i : \tau_i \& \phi_i \Rightarrow \tau_i$ is the type of the expression M_i and the semantic property ϕ_i holds
- then $\Gamma \vdash E(M_1, \dots, M_n) : \tau \& \phi \Rightarrow \tau$ is the type of the expression $E(M_1, \dots, M_n)$ and the semantic property ϕ holds



TINYLINKS's types-and-effects system

Goal

Whenever an assertion `assert L` occurs in the execution there exists a previous occurrence of an event `event L`

$$\Gamma; F \vdash E \xRightarrow{exp} \langle _ : T \rangle \{ F_1 \}$$

$$\Gamma; F \vdash E \xleftarrow{exp} \langle _ : T \rangle \{ F_1 \}$$

TINYLINKS's types-and-effects system

Goal

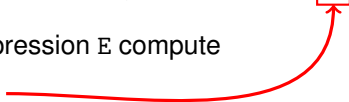
Whenever an assertion `assert L` occurs in the execution there exists a previous occurrence of an event `event L`

$$\Gamma; F \vdash E \xRightarrow{exp} \langle _ : T \rangle \{ F_1 \}$$

$$\Gamma; F \vdash E \xleftarrow{exp} \langle _ : T \rangle \{ F_1 \}$$

For each expression `E` compute

1. the type



TINYLINKS's types-and-effects system

Goal

Whenever an assertion `assert L` occurs in the execution there exists a previous occurrence of an event `event L`

$$\Gamma; \mathbf{F} \vdash E \xRightarrow{exp} \langle _ : T \rangle \{ F_1 \}$$

$$\Gamma; \mathbf{F} \vdash E \xleftarrow{exp} \langle _ : T \rangle \{ F_1 \}$$

For each expression `E` compute

1. the type
2. the preconditions

TINYLINKS's types-and-effects system

Goal

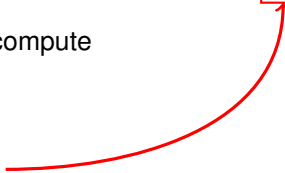
Whenever an assertion `assert L` occurs in the execution there exists a previous occurrence of an event `event L`

$$\Gamma; F \vdash E \xRightarrow{exp} \langle _ : T \rangle \{ F_1 \}$$

$$\Gamma; F \vdash E \xleftarrow{exp} \langle _ : T \rangle \{ F_1 \}$$

For each expression `E` compute

1. the type
2. the preconditions
3. the post-conditions



TINYLINKS's types-and-effects system

Some example of rules

$$\text{(T-Get)} \frac{\Gamma; F \vdash V \stackrel{val}{\Leftarrow} \text{xml}}{\Gamma; F \vdash \text{get}(V) \stackrel{exp}{\Rightarrow} \langle _ : \text{xml} \rangle \{ \}}$$

TINYLINKS's types-and-effects system

Some example of rules

$$\text{(T-Post)} \frac{\Gamma; F \vdash V_i \stackrel{val}{\Leftarrow} \text{string} \quad \forall i \in \{1, \dots, n\} \quad \Gamma; F \vdash U \stackrel{val}{\Leftarrow} \text{xml}}{\Gamma; F \vdash \text{post}([\langle l_1 = V_1, \dots, l_n = V_n \rangle], U) \stackrel{exp}{\Rightarrow} \langle _ : \text{xml} \rangle \{ \}}$$

TINYLINKS's types-and-effects system

Some example of rules

$$\text{(T-Assert)} \frac{\Gamma \vdash \diamond \quad fv(\mathbf{F}, \mathbf{L}) \subseteq dom(\Gamma) \quad \mathbf{L} \in \mathbf{F} \quad \mathbf{L} = \mathbf{p}(V_1, \dots, V_n) \quad \Gamma; \mathbf{F} \vdash V_i \xRightarrow{val} T_i \quad \forall i \in \{1, \dots, n\}}{\Gamma; \mathbf{F} \vdash \text{assert } \mathbf{L} \xRightarrow{exp} \langle _ : \text{unit} \rangle \{ \mathbf{L} \}}$$

TINYLINKS's types-and-effects system

Some example of rules

$$\text{(T-App)} \frac{
 \begin{array}{l}
 \Gamma; F \vdash U \xRightarrow{val} T \quad T = \langle x_1 : T_1, \dots, x_n : T_n \rangle \{ F_1 \} \rightarrow W \quad fv(T) = \emptyset \\
 \Gamma; F \vdash V_i \xleftarrow{val} T_i \quad \forall i \in \{ 1, \dots, n \} \quad F_1 [V_1/x_1] \dots [V_n/x_n] \subseteq F
 \end{array}
 }{
 \Gamma; F \vdash U(V_1, \dots, V_n) \xRightarrow{exp} W[V_1/x_1] \dots [V_n/x_n]
 }$$



TINYLINKS's types-and-effects system

Safe Web Application

A web application E is safe if and only if there exists a proof within the types-and-effects system of the judgment

$$\emptyset; \emptyset \vdash E \stackrel{exp}{\Leftarrow} \langle _ : \text{xml} \rangle \{ \}$$



Types-and-effects system

Usually the definition of a types-and-effects analysis requires

1. Definition of rules
2. State and prove the soundness of analysis
3. Definition of inference algorithm
4. Prove that the algorithm is correct (soundness/completeness)

Types-and-effects system

Usually the definition of a types-and-effects analysis requires

1. Definition of rules
2. State and prove the soundness of analysis
3. Definition of inference algorithm
4. Prove that the algorithm is correct (soundness/completeness)

The approach used for the TINYLINKS's types-and-effects system is different

- each expression is translated in an expression of F7
 - this translation hides the property of the analysis



Our goal

Reconstruct the TINYLINKS's types-and-effects system by **abstract interpretation**

Our goal

Reconstruct the TINYLINKS's types-and-effects system by **abstract interpretation**

Benefits

1. precise definition of relation between analysis and semantics
2. analysis is correct by construction

Our goal

Reconstruct the TINYLINKS's types-and-effects system by **abstract interpretation**

Benefits

1. precise definition of relation between analysis and semantics
2. analysis is correct by construction

We follow Cousot's methodology for type systems

1. we define a denotational semantics for TINYLINKS (concrete semantics)
2. we define a abstract semantics that computes types augmented by effects



Concrete semantics

A denotational semantics that considers

Concrete semantics

A denotational semantics that considers

- TINYLINKS as an untyped λ -calculus

$$Eval = (\dots + \underbrace{EEnv \rightarrow (Eval \times EEnv)}_{Href} + \dots)_{\perp}$$


Concrete semantics

A denotational semantics that considers

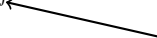
- TINYLINKS as an untyped λ -calculus

$$Eval = (\dots + \underbrace{EEnv \rightarrow (Eval \times EEnv)}_{Href} + \dots)_{\perp}$$

Domain of values



Injection (Text, Form, Fun, ...)



Concrete semantics

A denotational semantics that considers

- TINYLINKS as an untyped λ -calculus

$$Eval = (\dots + \underbrace{EEnv \rightarrow (Eval \times EEnv)}_{Href} + \dots)_{\perp}$$

Domain of values

Injection (Text, Form, Fun, ...)

- the occurrence and assertion of the events

$$EEnv = Pred \rightarrow (Dval \times Mark)$$

Concrete semantics

A denotational semantics that considers

- TINYLINKS as an untyped λ -calculus

$$Eval = (\dots + \underbrace{EEnv \rightarrow (Eval \times EEnv)}_{Href} + \dots)_\perp$$

Diagram illustrating the definition of $Eval$:

- $Eval$ is the domain of values.
- $EEnv$ is the environment.
- $Href$ is the injection (Text, Form, Fun, ...).

- the occurrence and assertion of the events

$$EEnv = Pred \rightarrow (Dval \times Mark)$$

Diagram illustrating the definition of $EEnv$:

- $EEnv$ is the events environment.
- $Pred$ is the predicate.
- $Dval$ is the domain of values, where values in the events are integers only.
- $Mark$ is the mark, with elements $\{ E, EA, A \}$.

Concrete semantics

Two semantic functions

1. for values

$$\mathcal{V}[-] : \text{VAL} \rightarrow Env \rightarrow EEnv \rightarrow Eval$$

2. for expressions

$$[-] : \text{EXP} \rightarrow Env \rightarrow EEnv \rightarrow (Eval \times EEnv)$$

Concrete semantics

Two semantic functions

1. for values

$$\mathcal{V}[-] : \text{VAL} \rightarrow \text{Env} \rightarrow \text{EEnv} \rightarrow \text{Eval}$$

2. for expressions

$$[-] : \text{EXP} \rightarrow \text{Env} \rightarrow \text{EEnv} \rightarrow (\text{Eval} \times \text{EEnv})$$

Examples of semantic equation

$$\begin{aligned} \llbracket \text{get}(v) \rrbracket \rho \phi &= \text{let}^* v' = \mathcal{V}[\mathbf{V}] \rho \phi \text{ in} \\ &\text{case } v' \text{ of} \\ &\quad H \text{ref}(f) \rightarrow f \phi \\ &\quad _ \rightarrow (\llbracket \text{WrongValue}() \rrbracket, \iota) \end{aligned}$$

Concrete semantics

Two semantic functions

1. for values

$$\mathcal{V}[-] : \text{VAL} \rightarrow \text{Env} \rightarrow \text{EEnv} \rightarrow \text{Eval}$$

2. for expressions

$$[-] : \text{EXP} \rightarrow \text{Env} \rightarrow \text{EEnv} \rightarrow (\text{Eval} \times \text{EEnv})$$

Examples of semantic equation

$$\begin{aligned} \llbracket \text{assert } q(\mathbf{V}) \rrbracket \rho \phi &= \text{let}^* \text{ ev} = \text{evalToDval}(\mathcal{V}[\mathbf{V}]\rho \phi) \text{ in} \\ &\quad \text{let } (ev', m) = \phi q \\ &\quad \text{if } ev = ev' \text{ then} \\ &\quad \quad (\llbracket \text{Unit}() \rrbracket, \phi[(ev', EA)/q]) \\ &\quad \text{else} \\ &\quad \quad (\llbracket \text{WrongValue}() \rrbracket, \iota) \end{aligned}$$



Unsoundness in TINYLINKS analysis

Consider the expression

```
get(Text("Hello World!"))
```

Unsoundness in TINYLINKS analysis

Consider the expression

```
get(Text("Hello World!"))
```

Semantics

$$\llbracket \text{get}(\text{Text}(\text{"Hello World!"})) \rrbracket \rho \phi = (\lfloor \text{WrongValue}(), \iota \rfloor)$$

Error: the denotation of `Text ("Hello World! ")` is not a link

Unsoundness in TINYLINKS analysis

Consider the expression

`get(Text("Hello World!"))`

Semantics

$\llbracket \text{get}(\text{Text}(\text{"Hello World!"})) \rrbracket \rho \phi = (\lfloor \text{WrongValue}(), \iota \rfloor)$

Error: the denotation of `Text("Hello World!")` is not a link

TINYLINKS's types-and-effects system

$\emptyset; \emptyset \vdash \text{get}(\text{Text}(\text{"Hello World!"})) \stackrel{exp}{\Leftarrow} \langle _ : \text{xml} \rangle \{ \}$

the expression is type checked and the computed type is `xml`

Abstract domain

Values

Trouble

- types have annotations

$$\text{integer} \{ \} \rightarrow \text{integer} \{ q : 5, p : 3 \}$$

Abstract domain

Values

Trouble

- types have annotations

$$\text{integer} \{ \} \rightarrow \text{integer} \{ q : 5, p : 3 \}$$

- annotated types are not a free algebra

$$\text{integer} \{ \} \rightarrow \text{integer} \{ q : 5, p : 3 \}$$
$$\text{integer} \{ \} \rightarrow \text{integer} \{ p : 3, q : 5 \}$$

Abstract domain

Values

Trouble

- types have annotations

$$\text{integer} \{ \} \rightarrow \text{integer} \{ q : 5, p : 3 \}$$

- annotated types are not a free algebra

$$\text{integer} \{ \} \rightarrow \text{integer} \{ q : 5, p : 3 \}$$

$$\text{integer} \{ \} \rightarrow \text{integer} \{ p : 3, q : 5 \}$$

Solution: simple type and constraints

Abstract domain

Values

Trouble

- types have annotations

$$\text{integer} \{ \} \rightarrow \text{integer} \{ q : 5, p : 3 \}$$

- annotated types are not a free algebra

$$\text{integer} \{ \} \rightarrow \text{integer} \{ q : 5, p : 3 \}$$

$$\text{integer} \{ \} \rightarrow \text{integer} \{ p : 3, q : 5 \}$$

Solution: simple type and constraints

- we substitute the annotations in the types with annotation variables

Abstract domain

Values

Trouble

- types have annotations

$$\text{integer} \{ \} \rightarrow \text{integer} \{ q : 5, p : 3 \}$$

- annotated types are not a free algebra

$$\text{integer} \{ \} \rightarrow \text{integer} \{ q : 5, p : 3 \}$$

$$\text{integer} \{ \} \rightarrow \text{integer} \{ p : 3, q : 5 \}$$

Solution: simple type and constraints

- we substitute the annotations in the types with annotation variables
- we introduce constraints to restrict annotation variables

$$\text{integer}(\gamma_1) \rightarrow \text{integer}(\gamma_2) \quad \gamma_1 \supseteq \emptyset \quad \gamma_2 \supseteq \{ p : 3, q : 5 \}$$

Abstract domain

Values

- the events depend on the concrete values ($Dval$)

Abstract domain

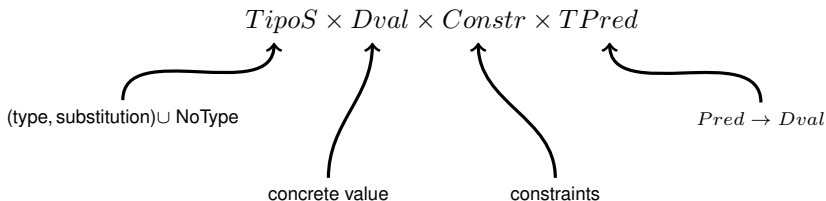
Values

- the events depend on the concrete values ($Dval$)
- the abstract domain need to include the concrete values

Abstract domain

Values

- the events depend on the concrete values ($Dval$)
- the abstract domain need to include the concrete values



Abstract semantics

Example of semantic equation

$$\begin{aligned}
 \llbracket \text{get}(V) \rrbracket^a \rho \phi = & \\
 & \gamma \in V_a \text{ fresh} \\
 & \text{let } (ts, d, C, f) = \mathcal{V}[\llbracket V \rrbracket^a] \rho \phi \text{ in} \\
 & \text{if } ts \neq \text{NoType} \text{ then} \\
 & \quad \text{case } \text{mgu}(\{ ts.t = \text{link}(\gamma) \} \cup ts.\theta) \text{ of} \\
 & \quad \quad S(\theta) \rightarrow \text{let } C' = \{ (\theta(\gamma), q) \in \theta(C) \} \text{ in} \\
 & \quad \quad \quad \text{if } \text{check}(\theta(f \leftarrow C'), \phi) \text{ then} \\
 & \quad \quad \quad \quad ((\theta(\text{xml}(\gamma)), \theta), \\
 & \quad \quad \quad \quad \quad \text{nodval}, \theta(C) \setminus C', \theta(f \downarrow C)), \phi) \\
 & \quad \quad \text{else} \\
 & \quad \quad \quad (\text{Error}, \iota) \\
 & \quad \quad _ \rightarrow (\text{Error}, \iota) \\
 & \text{else} \\
 & \quad (\text{Error}, \iota)
 \end{aligned}$$



Analyzer

Both the concrete and abstract semantics have been implemented as OCaml programs



Analyzer

Both the concrete and abstract semantics have been implemented as OCaml programs

- TINYLINKS programs are represented in abstract syntax

Analyzer

Both the concrete and abstract semantics have been implemented as OCaml programs

- TINYLINKS programs are represented in abstract syntax
- the implementation have an unique semantic function parametrized with respect to

Analyzer

Both the concrete and abstract semantics have been implemented as OCaml programs

- TINYLINKS programs are represented in abstract syntax
- the implementation have an unique semantic function parametrized with respect to
 - the primitive operations

Analyzer

Both the concrete and abstract semantics have been implemented as OCaml programs

- TINYLINKS programs are represented in abstract syntax
- the implementation have an unique semantic function parametrized with respect to
 - the primitive operations
 - the semantic domain



Example 1

Expression

```
fun buy(value, dbpass) {  
    var _ = assert PriceIs(value);  
    Text("a")  
}
```

Example 1

Expression

```
fun buy(value, dbpass) {  
  var _ = assert PriceIs(value);  
  Text("a")  
}
```

Abstract semantics

```
(type - :  
Function(_#value#var0_, Integer(), _annvar0_,  
  Function(_#dbpass#var1_, _typevar1_, _annvar2_,  
    Xml(_annvar4_), _annvar3_),  
  _annvar1_)  
No_dval [(_annvar2_, PriceIs)]  
  {PriceIs -> _#value#var0_}, {})
```



Example 2

Expression

buy 5

Example 2

Expression

buy 5

Abstract semantics

```
(type - :  
  Function(_#dbpass#var3_, _typevar3_, _annvar7_,  
           Xml(_annvar9_), _annvar8_)  
  Unknown [(_annvar7_, PriceIs)] {PriceIs -> 5}, {})
```




Example 3

Expression

buy 5 "a"

Example 3

Expression

buy 5 "a"

Abstract semantics

Exception: No_type "apply_fun: no preconditions"

Conclusions

We have reconstructed the TINYLINKS's types-and-effects system by abstract interpretation

- we have precisely defined relationship between semantics and analysis
- we have shown unsoundness of TINYLINKS's types-and-effects system

Conclusions

We have reconstructed the TINYLINKS's types-and-effects system by abstract interpretation

- we have precisely defined relationship between semantics and analysis
- we have shown unsoundness of TINYLINKS's types-and-effects system

Future work

- Consider a type system with sub-types (`link <: xml`, `form <: xml`)
- Extend the class of value that can be used in the events
- Generalize the methodology and apply it to further analysis