A Systematic Approach to Static Access Control

François Pottier¹, Christian Skalka², and Scott Smith²

¹ INRIA Rocquencourt, Francois.Pottier@inria.fr
² The Johns Hopkins University, {ces,scott}@cs.jhu.edu

Abstract. The Java JDK 1.2 Security Architecture includes a dynamic mechanism for enforcing access control checks, so-called stack inspection. This paper studies type systems which can statically guarantee the success of these checks. We develop these systems using a new, systematic methodology: we show that the security-passing style translation, proposed by Wallach and Felten as a dynamic implementation technique, also gives rise to static security-aware type systems, by composition with conventional type systems. To define the latter, we use the general HM(X) framework, and easily construct several constraint- and unification-based type systems. They offer significant improvements on a previous type system for JDK access control, both in terms of expressiveness and in terms of readability of inferred type specifications.

1 Introduction

The Java Security Architecture [2], found in JDK 1.2 and later, includes mechanisms to protect systems from operations performed by untrusted code. These access control decisions are enforced by dynamic checks. Our goal is to make some or all of these decisions statically, by extensions to the type system. Thus, access control violations will be caught at compile-time rather than run-time. Furthermore, types (whether inferred or programmer-supplied) will constitute a specification of the security policy.

A Brief Review of the JDK Security Architecture For lack of space, we cover the JDK security architecture in a cursory manner here; see [2, 13, 8] for more detailed background. To use the access control system, the programmer adds doPrivileged and checkPrivilege commands to the code. At run-time, a doPrivileged command adds a flag to the current stack frame, enabling a particular privileged operation. The flag is implicitly eliminated when the frame is popped. When a privilege is checked via a checkPrivilege command, the stack frames are searched most to least recent. If a frame is encountered with the desired flag, the search stops and the check succeeds. Additionally, each stack frame is annotated with its owner (the owner of the method being invoked), and all stack frames searched by the above algorithm must be owned by some principal authorized for the privilege being checked. This keeps illicit code, invoked by the trusted codebase when doPrivileged is on the stack, from performing the privileged operation.
Our Framework This paper follows up on an initial access control type system presented by the last two authors in [8] and places emphasis on a more modular approach to type system construction. The previous paper developed the security type system ab initio. In this paper, we reduce the security typing problem to a conventional typing problem using a translation-based method inspired by [5]. We use a standard language of row types [7] to describe sets of privileges. We also re-use the HM(X) framework [3, 9], which allows a wide variety of type systems to be defined in a single stroke, saves some proof effort, and (most importantly) shows that our custom type systems arise naturally out of a standard one.

In addition to these methodological enhancements, this paper improves upon its predecessor in several other ways. In particular, [8] was based on subtyping constraints, whereas one of the type systems presented here uses row unification alone; this makes it more efficient and leads to more concise types. Also, the calculus studied in this paper allows for dynamic test-and-branch on whether a privilege is enabled. Lastly, because our new approach relies on HM(X), we can easily provide let-polymorphism.

We begin by defining a simplified model of the Java JDK 1.2 security architecture. It is a λ-calculus, called λsec, equipped with a non-standard operational semantics that includes a specification of stack inspection. In order to construct a static type system for λsec, we translate it into a standard λ-calculus, called λset. The translation is a security-passing style transformation [13]: it implements stack inspection by passing around sets of privileges at run-time. For this purpose, λset is equipped with built-in notions of set and set operations.

Then, we define a type system for λset. Because λset is a standard λ-calculus, we are able to define our type system as a simple instance of the HM(X) framework [3]. In fact, by using this framework a whole family of type systems may be succinctly defined, each with different costs and benefits. In order to give precise types to λset’s built-in set operations, our instance uses set types, defined as a simplification of Rémy’s record types [7].

Lastly, we show that any type system for λset gives rise through the translation to a type system for λsec. The latter’s correctness follows immediately from the former’s, provided the translation itself is correct. This is quite easy to show, since the property does not involve types at all.

2 The Source Language λsec

This section defines λsec, a simplified model of the JDK 1.2 security architecture. It is a λ-calculus equipped with a notion of code ownership and with constructs for enabling or checking privileges. Its grammar is given in Fig. 1.

We assume given notions of principals and resources (the latter also known as privileges), taken from arbitrary sets P and R. We use p and r to range over principals and resources, respectively, and P and R to range over sets thereof.

We assume given a fixed access credentials list A. It is a function which maps every principal p ∈ P to a subset of R. We let A−1 denote its “inverse”, that is,
\[ p \in P, P \subseteq P \] 
\[ r \in R, R \subseteq R \] 
\[ A \in P \rightarrow 2^R \] 
\[ v ::= \lambda x. f \] 
\[ e ::= x | \lambda x. f | e e | \text{let } x = e \text{ in } e | \text{letpriv } r \text{ in } e | \text{checkpriv } r \text{ for } e | \text{testpriv } r \text{ then } e \text{ else } e | f \] 
\[ f ::= p.e \] 
\[ E ::= [] | E e | v E | \text{let } x = E \text{ in } e | \text{letpriv } r \text{ in } E | p.E \] 

**Fig. 1. Grammar for \( \lambda_{sec} \)**

the function which maps a resource \( r \in R \) to \( \{ p \in P \mid r \in A(p) \} \). Without loss of generality, we assume the existence of a fixed principal \( p_0 \) such that \( A(p_0) = \emptyset \).

A signed expression \( p.e \) behaves as the expression \( e \) endowed with the authority of principal \( p \). Notice how the body of every \( \lambda \)-abstraction is required to be a signed expression — thus, every piece of code must be vouched for by some principal. The construct \( \text{letpriv } r \text{ in } e \) allows an authorized principal to enable the use of a resource \( r \) within the expression \( e \). The construct \( \text{checkpriv } r \text{ for } e \) asserts that the use of \( r \) is currently enabled. If \( r \) is indeed enabled, \( e \) is evaluated; otherwise, execution fails. The construct \( \text{testpriv } r \text{ then } e_1 \text{ else } e_2 \) dynamically tests whether \( r \) is enabled, branching to \( e_1 \) or \( e_2 \) if this holds or fails, respectively.

### 2.1 Stack Inspection

The JDK 1.2 determines whether a resource is enabled by literally examining the runtime stack, hence the name *stack inspection*. We give a simple specification of this process by noticing that stacks are implicitly contained in *evaluation contexts*, whose grammar is defined in Fig. 1. Indeed, a context defines a path from the term’s root down to its active redex, along which one finds exactly the security annotations which the JDK 1.2 would maintain on the stack, that is, code owners \( p \) and enabled resources \( r \).

To formalize this idea, we associate a finite string of principals and resources, called a *stack*, to every evaluation context \( E \). The right-most letters in the string correspond to the most recent stack frames.

\[
\begin{align*}
\text{stack}([]) &= \epsilon & \text{stack}(E e) &= \text{stack}(E) \\
\text{stack}(v E) &= \text{stack}(E) & \text{stack}(\text{let } x = E \text{ in } e) &= \text{stack}(E) \\
\text{stack}(\text{letpriv } r \text{ in } E) &= r,\text{stack}(E) & \text{stack}(p.E) &= p,\text{stack}(E)
\end{align*}
\]

Then, Fig. 2 defines stack inspection, with \( S \vdash r \) meaning access to resource \( r \) is allowed by stack \( S \), and \( S \vdash P \) meaning some principal in \( P \) is the most recent owner on \( S \). This specification corresponds roughly to Wallach’s [13, p. 71]. We write \( E \vdash r \) for \( \text{stack}(E) \vdash r \).
\[ \begin{array}{cccc}
  r \in \mathcal{A}(p) & S \vdash r & S \vdash r & S \vdash \mathcal{A}^{-1}(r) \\
  S.p \vdash r & S.r \vdash r & S.r \vdash P & S.p \vdash P
\end{array} \]

Fig. 2. Stack inspection algorithm

### 2.2 Operational Semantics for \( \lambda_{\text{sec}} \)

The operational semantics of \( \lambda_{\text{sec}} \) is defined by the following reduction rules:

\[
E[(\lambda x.f) v] \rightarrow E[f[v/x]] \\
E[\text{let } x = v \text{ in } e] \rightarrow E[e[v/x]] \\
E[\text{checkpriv } r \text{ for } e] \rightarrow E[e] \quad \text{if } E \vdash r \\
E[\text{testpriv } r \text{ then } e_1 \text{ else } e_2] \rightarrow E[e_1] \quad \text{if } E \vdash r \\
E[\text{testpriv } r \text{ then } e_1 \text{ else } e_2] \rightarrow E[e_2] \quad \text{if } \neg(E \vdash r) \\
E[\text{letpriv } r \text{ in } v] \rightarrow E[v] \\
E[p.v] \rightarrow E[v]
\]

The first two rules are standard. The next rule allows \text{checkpriv } r \text{ for } e \text{ to reduce into } e \text{ only if stack inspection succeeds (as expressed by the side condition } E \vdash r); \text{ otherwise, execution is blocked. The following two rules use stack inspection in a similar way to determine how to reduce } \text{testpriv } r \text{ then } e_1 \text{ else } e_2; \text{ however, they never cause execution to fail. The last two rules state that security annotations become unnecessary once the expression they enclose has been reduced to a value. In a Java virtual machine, these rules would be implemented simply by popping stack frames (and the security annotations they contain) after executing a method.}

This operational semantics constitutes a concise, formal description of Java stack inspection in a higher-order setting. It is easy to check that every closed term either is a value, or is reducible, or is of the form \( E[\text{checkpriv } r \text{ for } e] \) where \( \neg(E \vdash r) \). Terms of the third category are \textit{stuck}; they represent access control violations. An expression \( e \) is said to go \textit{wrong} if and only if \( e \rightarrow^* e' \), where \( e' \) is a stuck expression, holds.

### 3 The Target Calculus \( \lambda_{\text{set}} \)

We now define a standard calculus, \( \lambda_{\text{set}} \), to be used as the target of our translation. It is a \( \lambda \)-calculus equipped with a number of constants which provide set operations, and is given in Fig. 3. We will use \( e,r; e \lor R \) and \( e \land R \) as syntactic sugar for \( (e,r)_1, (\lor_R e) \) and \( (\land_R e) \), respectively.

The constant \( R \) represents a constant set. The construct \( e,r \) asserts that \( r \) is an element of the set denoted by \( e \); its execution fails if that is not the case. The construct \( e \lor R \) (resp. \( e \land R \)) allows computing the union (resp. intersection) of the set denoted by \( e \) with a constant set \( R \). Lastly, the expression \( ?_r e f g \)
dynamically tests whether \( r \) belongs to the set \( R \) denoted by \( e \), and accordingly invokes \( f \) or \( g \), passing \( R \) to it. The operational semantics for \( \lambda_{\text{set}} \) is as follows:

\[
\begin{align*}
(\lambda x.e) v & \rightarrow e[v/x] \\
\text{let } x = v \text{ in } e & \rightarrow e[v/x] \\
R x & \rightarrow R \quad \text{if } r \in R \\
?_r R & \rightarrow \lambda f.\lambda g.(f R) \quad \text{if } r \in R \\
?_r R & \rightarrow \lambda f.\lambda g.(g R) \quad \text{if } r \not\in R \\
R_1 \lor R_2 & \rightarrow R_1 \cup R_2 \\
R_1 \land R_2 & \rightarrow R_1 \cap R_2 \\
E[e] & \rightarrow E[e'] \quad \text{if } e \rightarrow e'
\end{align*}
\]

Again, an expression \( e \) is said to go wrong if and only if \( e \rightarrow^* e' \), where \( e' \) is a stuck expression, holds.

## 4 Source-to-Target Translation

A translation of \( \lambda_{\text{sec}} \) into \( \lambda_{\text{set}} \) is defined in Fig. 4. The distinguished identifiers \( s \) and \( _* \) are assumed not to appear in source expressions. Notice that \( s \) may appear free in translated expressions. Translating an (unsigned) expression requires specifying the current principal \( p_0 \).

One will often wish to translate an expression under minimal hypotheses, i.e. under the initial principal \( p_0 \) and a void security context. To do so, we define \( \{ e \} = \{ e \}_p [\emptyset / s] \). Notice that \( s \) does not appear free in \( \{ e \} \). If \( e \) is closed, then so is \( \{ e \} \).

\[
\begin{align*}
[x]_p & = x \\
[\lambda x.y]_p & = \lambda x.\lambda s.[y] \\
[e_1 e_2]_p & = [e_1]_p [e_2]_p \cdot s \\
[\text{let } x = e_1 \text{ in } e_2]_p & = [e_1]_p [e_2]_p \\
[\text{letpriv } r \text{ in } e]_p & = \text{let } s = s \lor ((r) \cap A(p)) \text{ in } [e]_p \\
[\text{checkpriv } r \text{ for } e]_p & = \text{let } s = s \land \text{in } [e]_p \\
[\text{testpriv } r \text{ then } e_1 \text{ else } e_2]_p & = ?_r s (\lambda s.\{ e_1 \}_p) (\lambda s.\{ e_2 \}_p) \\
[f]_p & = [f] \\
[p.e] & = \text{let } s = s \land A(p) \text{ in } [e]_p
\end{align*}
\]

Fig. 4. Source-to-Target Translation
The idea behind the translation is simple: the variable $s$ is bound at all times to the set of currently enabled resources. Every function accepts $s$ as an extra parameter, because it must execute within its caller's security context. As a result, every function call has $s$ as its second parameter. The constructs $\text{letpriv } r \text{ in } e$ and $p.e$ cause $s$ to be locally bound to a new value, reflecting the new security context; more specifically, the former enables $r$, while the latter disables all privileges not in $A(p)$. The constructs $\text{checkpriv } r$ for $e$ and $\text{testpriv } r$ then $e_1$ else $e_2$ are implemented simply by looking up the current value of $s$. In the latter, $s$ is re-bound, within each branch, to the same value. This may appear superfluous at first sight, but has an important impact on typing, because it allows $s$ to be given a different (more precise) type within each branch.

This translation can be viewed as a generalization of Wallach's security-passing style transformation [13] to a higher-order setting. Whereas they advocated this idea as an implementation technique, with efficiency in mind, we use it only as a vehicle in the proof of our type systems. Here, efficiency is not at stake. Our objective is only to define a correct translation, that is, to prove the following:

**Theorem 4.1.** If $e \rightarrow^* v$, then $\{e\} \rightarrow^* \{v\}$. If $e$ goes wrong, then $\{e\}$ goes wrong. If $e$ diverges, then $\{e\}$ diverges.

The proof is divided in two steps. First, we define a new stack inspection algorithm, which walks the stack forward instead of backward, and computes, at each step, the set of currently enabled resources. Then, we show that the translation implements this algorithm, interleaved with the actual code. Both proof steps are straightforward, and we omit them here for brevity.

## 5 Types for $\lambda_{\text{set}}$

We define a type system for the target calculus as an instance of the parametric framework $\text{HM}(X)$ [3, 9]. $\text{HM}(X)$ is a generic type system in the Hindley-Milner tradition, parameterized by an abstract constraint system $X$. Sect. 5.1 briefly recalls its definition. Sect. 5.2 defines a specific constraint system called SETS, yielding the type system $\text{HM}(\text{SET})$. Sect. 5.3 extends $\text{HM}(\text{SET})$ to the entire language $\lambda_{\text{set}}$ by assigning types to its primitive operations. Sect. 5.4 states type safety results and discusses a couple of choices.

### 5.1 The System $\text{HM}(X)$

The system $\text{HM}(X)$ is parameterized by a sound term constraint system $X$, i.e. by notions of types $\tau$, constraints $C$, and constraint entailment $\vdash$, which must satisfy a number of axioms [3].

Then, a type scheme is a triple of a set of quantifiers $\bar{\alpha}$, a constraint $C$, and a type $\tau$ (which, in this paper, must be of kind $\text{Type}$; see Sect. 5.2), written $\sigma := \forall \bar{\alpha}[C] \cdot \tau$. A type environment $\Gamma$ is a partial mapping of program variables to type schemes. A judgement is a quadruple of a satisfiable constraint $C$, a
type environment $\Gamma$, an expression $e$ and a type scheme $\sigma$, written $C, \Gamma \vdash e : \sigma$, derivable using the rules of Fig. 5. These rules correspond to those given in [9]. The following type safety theorem is proven in [3] with respect to a denotational presentation of the call-by-value $\lambda$-calculus with let. We have proved a syntactic version of it, in the style of [14], which better suits our needs.

**Theorem 5.1.** If $C, \Gamma \vdash e : \sigma$ holds, then $e$ does not go wrong.

### 5.2 The Constraint System SETS

In order to give precise types to the primitive set operations in $\lambda_{set}$, we need specific types and constraints. Together with their logical interpretation, which defines their meaning, these form a constraint system called SETS.

The syntax of types and constraints is defined in Fig. 6. The type language features a set type constructor \{\}, the two standard row constructors [7], and four capability constructors. Capabilities tell whether a given element may appear in a set (Pre), may not appear in it (Abs), may or may not appear in it (Either), or whether this information is irrelevant, because the set itself is unavailable (NA). For instance, the singleton set \{t\} will be one (and the only)
value of type \( r : \text{Pre} \cup \partial \text{Abs} \). The constraint language offers standard equality and subtyping constraints, as well as a form of conditional constraints. Sample uses of these types and constraints will be shown in Sect. 5.3.

To ensure that only meaningful types and constraints can be built, we immediately equip them with kinds, defined by \( k := \text{Cap} \mid \text{Row}_R \mid \text{Type} \), where \( R \) ranges over finite subsets of \( \mathcal{R} \). For every kind \( k \), we assume given a distinct, denumerable set of type variables \( \mathcal{V}_k \). We use \( \alpha, \beta, \gamma, \ldots \) to represent type variables. From here on, we consider only well-kinded types and constraints, as defined in Fig. 7. The purpose of these rules is to guarantee that every constraint has a well-defined interpretation within our model, whose definition follows.

To every kind \( k \), we associate a mathematical structure \( [k] \). \( [\text{Cap}] \) is the set of all four capabilities. Given a finite set of resources \( R \subseteq \mathcal{R} \), \( [\text{Row}_R] \) is the set of total, almost constant functions from \( \mathcal{R} \setminus R \) into \( [\text{Cap}] \). (A function is almost constant if it is constant except on a finite number of inputs.) In short, \( \text{Row}_R \) is the kind of rows which do not carry the fields mentioned in \( R \); \( \text{Row}_\emptyset \) is the kind of complete rows. \( [\text{Type}] \) is the free algebra generated by the constructors \( - \mapsto \), with signature \( [\text{Type}] \times [\text{Type}] \to [\text{Type}] \), and \{\}, with signature \( [\text{Row}_\emptyset] \to [\text{Type}] \).

Each of these structures is then equipped with an ordering. Here, a choice has to be made. If we do not wish to allow subtyping, we merely define the ordering on every \( [k] \) as equality. Otherwise, we proceed as follows. First, a lattice over \( [\text{Cap}] \) is defined, whose least (resp. greatest) element is \( \text{NA} \) (resp. \( \text{Either} \)), and where \( \text{Abs} \) and \( \text{Pre} \) are incomparable. This ordering is then extended, pointwise and covariantly, to every \( [\text{Row}_R] \). Finally, it is extended inductively to \( [\text{Type}] \) by viewing the constructor \( \{\} \) as covariant, and the constructor \( - \mapsto \) as contravariant (resp. covariant) in its first (resp. second) argument.

We may now give the interpretation of types and constraints within the model. It is parameterized by an assignment \( \rho \), i.e. a function which, for every kind \( k \), maps \( \mathcal{V}_k \) into \([k]\). The interpretation of types is obtained by extending \( \rho \) so as to map every type of kind \( k \) to an element of \([k]\), as follows:

\[
\begin{align*}
\rho(\tau \to \tau') & = \rho(\tau) \to \rho(\tau') \\
\rho(\tau) & = \rho(\tau) \\
\rho(\text{true}) & = \rho(\text{true}) \\
\rho(\text{if } \tau_1 \text{ then } \tau_2) & = \rho(\tau_1) \\
\rho(\exists \alpha. C) & = \rho(\exists \alpha. C) \\
\rho(\partial \tau) & = \rho(\partial \tau) \\
\rho(c) & = c
\end{align*}
\]

![Fig. 7. Kinding rules](image-url)
Fig. 8. Interpretation of constraints

Fig. 8 defines the constraint satisfaction predicate \( \vdash \), whose arguments are an assignment \( \rho \) and a constraint \( C \). (The notation \( \rho = \rho' [\alpha] \) means that \( \rho \) and \( \rho' \) coincide except possibly on \( \alpha \).) Entailment is defined as usual: \( C \vdash C' \) (read: \( C \) entails \( C' \)) holds iff, for every assignment \( \rho \), \( \rho \vdash C \) implies \( \rho \vdash C' \).

We refer to the type and constraint logic, together with its interpretation, as SETS. More precisely, we have defined two logics, where \( \leq \) is interpreted as either equality or as a non-trivial subtype ordering. We will refer to them as SETS\(^-\) and SETS\(^\leq\), respectively. Both are sound term constraint systems [3].

5.3 Dealing with the Primitive Operations in \( \lambda_{\text{set}} \)

The typing rules of HM(X) cover only the \( \lambda \)-calculus with let. To extend HM(SETS) to the whole language \( \lambda_{\text{set}} \), we must assign types to its primitive operations. Let us define an initial type environment \( I_1 \) as follows:

\[
\begin{align*}
R & : \{ R : \text{Pre}; \partial \text{Abs} \} \\
\vdash & : \forall \beta.\{ r : \text{Pre}; \beta \} \to \{ r : \text{Pre}; \beta \} \\
\forall_R & : \forall \beta \gamma.\{ R : \gamma; \beta \} \to \{ R : \text{Pre}; \beta \} \\
\land_R & : \forall \beta \gamma.\{ R : \gamma; \beta \} \to \{ R : \gamma; \partial \text{Abs} \} \\
\land_r & : \forall \alpha \beta \gamma.\{ r : \gamma; \beta \} \to (\{ r : \text{Pre}; \beta \} \to \alpha) \to (\{ r : \text{Abs}; \beta \} \to \alpha) \to \alpha
\end{align*}
\]

Here, \( \alpha, \beta, \gamma \) range over type variables of kind Type, Row, Cap, respectively. We abuse notation: if \( R \) is \( \{ r_1, \ldots, r_n \} \), then \( R : c \) denotes \( r_1 : c; \ldots; r_n : c \), and \( R : \gamma \) denotes \( r_1 : \gamma_1; \ldots; r_n : \gamma_n \).

None of the type schemes in \( I_1 \) carry constraints. If we wish to take advantage of conditional constraints, we must refine the type of \( \land_r \). Let \( I_2 \) be the initial type environment obtained by replacing the last binding in \( I_1 \) with

\[
\land_r : \forall \alpha \beta \gamma.\{ r : \gamma; \beta \} \to (\{ r : \text{Pre}; \beta_1 \} \to \alpha_1) \to (\{ r : \text{Abs}; \beta_2 \} \to \alpha_2) \to \alpha
\]

where \( C = \)

\[
\begin{align*}
\text{if } \text{Pre} & \leq \gamma \text{ then } \beta \leq \beta_1 \land \text{ if } \text{Abs} & \leq \gamma \text{ then } \beta \leq \beta_2 \\
\land & \text{ if } \text{Pre} \leq \gamma \text{ then } \alpha_1 \leq \alpha \land \text{ if } \text{Abs} \leq \gamma \text{ then } \alpha_2 \leq \alpha
\end{align*}
\]

Here, the input and output of each branch (represented by \( \beta_i \) and \( \alpha_i \), respectively) are linked to the input and output of the whole construct (represented
by \( \beta \) and \( \alpha \) through conditional constraints. Intuitively, this means that the security requirements and the return type of a branch may be entirely ignored unless the branch seems liable to be taken. (For more background on conditional constraints, the reader is referred to \([1, 4]\).)

### 5.4 The Type Systems \( S_i^{rel} \)

Sect. 5.2 describes two constraint systems, \( \text{SETS}^- \) and \( \text{SETS}^\leq \). Sect. 5.3 defines two initial typing environments, \( T_1 \) and \( T_2 \). These choices give rise to four related type systems, which we refer to as \( S_i^{rel} \), where \( rel \) and \( i \) range over \{\( =, \leq \)\} and \{1, 2\}, respectively. Each of them offers a different compromise between accuracy, readability and cost of analysis. In each case, Theorem 5.1 may be extended to the entire language \( \lambda_{set} \) by proving a simple \( \delta\)-typability [14] lemma, i.e. by checking that \( T_i \) correctly describes the behavior of the primitive operations. The proofs are straightforward and are not given here.

Despite sharing a common formalism, these systems may call for vastly different implementations. Indeed, every instance of \( \text{HM}(X) \) must come with a constraint resolution algorithm. \( S_i^- \) is a simple extension of the Hindley-Milner type system with rows, and may be implemented using unification \([6]\). \( S_i^\leq \) is similar, but requires conditional (i.e. delayed) unification constraints, adding some complexity to the implementation. \( S_i^- \) and \( S_i^\leq \) require maintaining subtyping constraints, usually leading to complex implementations.

In the following, we lack the space to describe all four variants. Therefore, we will focus on \( S_1^- \). Because it is based on unification, it is efficient, easy to implement, and yields readable types. We conjecture that, thanks to the power of row polymorphism, it is flexible enough for many practical uses (see Sect. 7.3).

### 6 Types for \( \lambda_{sec} \)

#### 6.1 Definition

Sect. 5 defined a type system, \( S_i^{rel} \), for \( \lambda_{set} \). Sect. 4 defined a translation of \( \lambda_{sec} \) into \( \lambda_{set} \). Composing the two automatically gives rise to a type system for \( \lambda_{sec} \), also called \( S_i^{rel} \) for simplicity, whose safety is a direct consequence of Theorems 4.1 and 5.1.

**Definition 6.1.** Let \( e \) be a closed \( \lambda_{sec} \) expression. By definition, \( C, \Gamma \vdash e : \sigma \) holds if and only if \( C, \Gamma \vdash \{e\} : \sigma \) holds.

**Theorem 6.2.** If \( C, \Gamma \vdash e : \sigma \) holds, then \( e \) does not go wrong.

Turning type safety into a trivial corollary was the main motivation for basing our approach on a translation. Indeed, because Theorem 4.1 concerns untyped terms, its proof is straightforward. (The \( \delta\)-typability lemmas mentioned in Sect. 5.3 do involve types, but are typically very simple.) A direct type safety proof would be non-trivial and would duplicate most of the steps involved in proving \( \text{HM}(X) \) correct.
### 6.2 Reformulation: Derived Type Systems

Definition 6.1, although simple, is not a direct definition of typing for $\lambda_{sec}$. We thus will give rules which allow typing $\lambda_{sec}$ expressions without explicitly translating them into $\lambda_{set}$. These so-called derived rules can be obtained in a rather systematic way from the definition of $S^m_l$ and the definition of the translation. (In fact, it would be interesting to formally automate the process.)

In these rules, the symbols $\tau$ and $\varsigma$ range over types of kind $\text{Type}$; more specifically, $\varsigma$ is used to represent some security context, i.e., a set of available resources. The symbols $\varphi$ and $\psi$ range over types of kind $\text{Row}_s$ and $\text{Cap}$, respectively. The $\ast$ symbol in the rules indicates an irrelevant principal. In the source-to-target translation, all functions are given an additional parameter, yielding types of the form $\tau_1 \rightarrow \varphi_2 \rightarrow \tau_2$. To recover the more familiar and appealing notation proposed in [8], we define the macro $\tau_1 \rightarrow \varphi_2 \rightarrow \tau_2 \to_{\varsigma} \tau_1 \rightarrow \varphi_2 \rightarrow \tau_2$.

Fig. 9 gives derived rules for $S^1_l$, the simplest of our type systems. There, all

![Typing rules for $\lambda_{sec}$ derived from $S^1_l$](image)

Fig. 9. Typing rules for $\lambda_{sec}$ derived from $S^1_l$
constraints are equations. As a result, all type information can be represented in term form, rather than in constraint form [9]. We exploit this fact to give a simple presentation of the derived rules. Type schemes have the form \( \forall \alpha . \tau \), and judgements have the form \( p : \gamma ; \Gamma \vdash e : \sigma \).

To check that these derived rules are correct, we prove the following lemmas:

**Lemma 6.3.** \( p : \gamma ; \Gamma \vdash e : \sigma \) holds iff \( \text{true} \), (\( \Gamma_1 ; \Gamma ; s : \gamma \) \( \vdash \llbracket e \rrbracket_p : \sigma \) holds.

**Lemma 6.4.** \( \langle \beta \rangle \text{Abs} \), \( \Gamma \vdash e : \sigma \) holds iff \( \text{true} \), (\( \Gamma_1 ; \Gamma \) \( \vdash \| e \| : \sigma \) holds.

Together, Theorem 6.2 and Lemma 6.4 show that, if a closed \( \lambda_{\text{sec}} \) expression \( e \) is well-typed according to the rules of Fig. 9, under the initial principal \( p_0 \) and the empty security context \( \{ \beta \text{Abs} \} \), then \( e \) cannot go wrong.

Derived rules for each member of the \( S_{\text{rel}} \) family can be given in a similar way. The same process can also be used to yield type inference rules, rather than the logical typing rules shown here.

7 Examples

7.1 Basic Use of Security Checks

Imagine an operating system with two kinds of processes, root processes and user processes. Killing a user process is always allowed, while killing a root process requires the privilege killing. At least one distinguished principal root has this privilege. The system functions which perform the killing are implemented by root, as follows:

\[
\begin{align*}
\text{kill} &= \lambda (p : \text{process}). \text{root}. \text{checkpriv}\ \text{killing}\ for\ \ldots () & \text{kill the process} \\
\text{killIfUser} &= \lambda (p : \text{process}). \text{root} \ldots () & \text{kill the process if it is user-level}
\end{align*}
\]

In system \( S_1 \), these functions receive the following (most general) types:

\[
\begin{align*}
\text{kill} : \forall \beta . \text{process} & \rightarrow \text{unit} \\
\text{killIfUser} : \forall \beta . \text{process} & \rightarrow \text{unit}
\end{align*}
\]

The first function can be called only if it can be statically proven that the privilege killing is enabled. The second one, on the other hand, can be called at any time, but will never kill a root process. To complement these functions, it may be desirable to define a function which provides a “best attempt” given the current (dynamic) security context. This may be done by dynamically checking whether the privilege is enabled, then calling the appropriate function:

\[
\begin{align*}
\text{tryKill} &= \lambda (p : \text{process}). \text{root} . \\
& \text{testpriv killing then kill}\ (p) \ \text{else killIfUser}\ (p)
\end{align*}
\]

This function is well-typed in system \( S_1 \). Indeed, within the first branch of the testpriv construct, it is statically known that the privilege killing must be enabled; this is why the sub-expression kill\( (p) \) is well-typed. The inferred type shows that \( \text{tryKill} \) does not have any security requirements:

\[
\text{tryKill} : \forall \beta . \text{process} \rightarrow \text{unit}
\]
7.2 Security Wrappers

A library writer often needs to surround numerous internal functions with "boilerplate" security code before making them accessible. To avoid redundancy, it seems desirable to allow the definition of generic security wrappers. When applied to a function, a wrapper returns a new function which has the same computational meaning but different security requirements.

Assume given a principal \( p \) such that \( \mathcal{A}(p) = \{r, s\} \). Here are two wrappers likely to be of use to this principal:

\[
\begin{align*}
\text{enable}_r &= \lambda f.p.\lambda x.p.\text{letpriv } r \text{ in } f \ x \\
\text{require}_r &= \lambda f.p.\lambda x.p.\text{checkpriv } r \text{ for } f \ x
\end{align*}
\]

In system \( S^- \), these wrappers receive the following (most general) types:

\[
\begin{align*}
\text{enable}_r : \forall \ldots (\alpha_1 \xrightarrow{r: \text{Pre} \triangleright \text{Abs}} \beta_1) \rightarrow \alpha_2 \rightarrow (\alpha_1 \xrightarrow{r: \gamma_2 \triangleright \text{Abs}} \beta_2) \rightarrow \alpha_2 \\
\text{require}_r : \forall \ldots (\alpha_1 \xrightarrow{r: \text{Pre} \triangleright \text{Abs}} \beta_1) \rightarrow \alpha_2 \rightarrow (\alpha_1 \xrightarrow{r: \gamma_2 \triangleright \text{Abs}} \beta_2) \rightarrow \alpha_2
\end{align*}
\]

These types are very similar; they may be read as follows. Both wrappers expect a function \( f \) which allows that \( r \) be enabled \( (r: \text{Pre}) \), i.e. one which either requires \( r \) to be enabled, or doesn’t care about its status. (Indeed, as in ML, the type of the actual argument may be more general than that of the formal.) They return a new function with identical domain and codomain \( (\alpha_1, \alpha_2) \), which works regardless of \( r \)'s status \( (\text{enable}_r \) yields \( r: \gamma_2 \) or requires \( r \) to be enabled \( (\text{require}_r \) yields \( r: \text{Pre} \)). The new function retains \( f \)'s expectations about \( s \) \( (s: \gamma_1). f \) must not require any further privileges \( (\text{Abs}) \), because it is invoked by \( p \), which enjoys privileges \( r \) and \( s \) only.

These polymorphic types are very expressive. Our main concern is that, even though the privilege \( s \) is not mentioned in the code of these wrappers, it does appear in their type. More generally, every privilege in \( \mathcal{A}(p) \) may show up in the type of a function written on behalf of principal \( p \), which may lead to very verbose types. An appropriate type abbreviation mechanism may be able to address this problem; this is left as a subject for future work.

7.3 Advanced Examples

We lack space to cover numerous more subtle features of the type systems; let us give only some brief comments.

In Sect. 7.1, our use of testpriv was easily seen to be correct, because the sensitive action \( \text{kill}(p) \) was performed within its lexical scope. Matters become more delicate when testpriv is used to yield a function (or, in Java, an object), whose security requirements depend on the test’s outcome, and which is later invoked outside its scope. Conditional constraints are then required to track the dependency and prove that the function invocation is safe. It is not clear whether this idiom is a critical one to support in practice, and the question may be answerable only through experiment.
In Sect. 7.2, we pointed out that it is legal to pass \textit{enable}, a function \( f \) which doesn’t care about the status of \( r \), provided the type of \( f \) is \textit{polymorphic} in \( r \)’s status, as in

\[
\forall \gamma, \alpha_1 \xrightarrow{\{\gamma; \beta\}} \alpha_2
\]

If, on the other hand, it is monomorphic (because \( f \) is \( \lambda \)-bound rather than \textit{let}-bound), as in

\[
\alpha_1 \xrightarrow{\{\text{\texttt{Pre; \beta}}\}} \alpha_2
\]

then the application \((\text{\textit{enable}}; f)\) becomes well-typed only if subtyping is available, i.e. if \texttt{Pre} is a subtype of \texttt{Either}. We expect this situation to be infrequent, although this remains to be confirmed.

8 Discussion

\textit{Extension to a Full-Featured Language} Many features of the Java language or environment are not addressed in this theoretical study. In particular, Java views privileges as first-class objects, making static typing problematic. In our model, privileges are identifiers, and expressions cannot compute privileges. In the case of Java, it is an open question whether a completely static mechanism can be devised. If not, it may be desirable to take a soft typing approach [1].

\textit{Related Work} The security-passing style translation described in Sect. 4 is monadic. Monadic type systems have been used to analyze the use of impure language features in otherwise pure languages [11]. However, as deployed in [11], there is still “a need to create a new effect system for each new effect”. In other words, we apparently cannot readily re-use the work on monadic type systems in our setting. In fact, our work may be viewed as a \textit{systematic} construction of an “effect” type system adapted to our particular effectful programming language.

Several researchers have proposed ways of defining efficient, provably correct compilation schemes for languages whose security policy is expressed by a \textit{security automaton}. Walker [12] defines a source language, equipped with such a security policy, then shows how to compile it into a dependently-typed target language, whose type system, by encoding assertions about security states, guarantees that no run-time violations will occur. Walker first builds the target type system, then defines a typed translation. On the opposite, our approach consists in first defining an untyped translation, then letting the source type system arise from it. Thieman’s approach to security automata [10] is conceptually much closer to ours: he also starts with an untyped security-passing translation, whose output he then feeds through a standard program specializer, in order to automatically obtain an optimizing translation.

Our paper shares some motivations with these works; however, our aim was not only to gain performance by eliminating many dynamic checks, but also to define a programming discipline. This requires security types to be available not only at the level of compiled code, as in Walker’s work, but also in the source code itself.
References


