DYNAMIC SECURITY LABELS AND NONINTERFERENCE

(Extended Abstract)

Lantian Zheng
Computer Science Department
Cornell University
zlt@cs.cornell.edu

Andrew C. Myers
Computer Science Department
Cornell University
andru@cs.cornell.edu

Abstract This paper presents a language in which information flow is securely controlled by a type system, yet the security class of data can vary dynamically. Information flow policies provide the means to express strong security requirements for data confidentiality and integrity. Recent work on security-typed programming languages has shown that information flow can be analyzed statically, ensuring that programs will respect the restrictions placed on data. However, real computing systems have security policies that vary dynamically and that cannot be determined at the time of program analysis. For example, a file has associated access permissions that cannot be known with certainty until it is opened. Although one security-typed programming language has included support for dynamic security labels, there has been no demonstration that a general mechanism for dynamic labels can securely control information flow. In this paper, we present an expressive language-based mechanism for reasoning about dynamic security labels. The mechanism is formally presented in a core language based on the typed lambda calculus; any well-typed program in this language is provably secure because it satisfies noninterference.

1. Introduction

Information flow control protects information security by constraining how information is transmitted among objects and users of various security classes. These security classes are expressed as *labels* associated with the information or its containers. Denning [5] showed how to use static analysis to ensure that

programs use information in accordance with its security class, and this approach has been instantiated in a number of languages in which the type system implements a similar static analysis (e.g., [23, 9, 27, 17, 2, 19]). These type systems are an attractive way to enforce security because they can be shown to enforce *noninterference* [8], a strong, end-to-end security property. For example, when applied to confidentiality, noninterference ensures that confidential information cannot be released by the program no matter how it is transformed.

However, security cannot be enforced purely statically. In general, programs interact with an external environment that cannot be predicted at compile time, so there must be a run-time mechanism that allows security-critical decisions to be taken based on dynamic observations of this environment. For example, it is important to be able to change security settings on files and database records, and these changes should affect how the information from these sources can be used. A purely static mechanism cannot enforce this.

To securely control information flow when access rights can be changed and determined dynamically, *dynamic* labels [14] are needed that can be manipulated and checked at run time. However, manipulating labels dynamically makes it more difficult to enforce a strong notion of information security for several reasons. First, changing the label of an object may convert sensitive data to public data, directly violating noninterference. Second, label changes (and changes to access rights in general) can be used to convey information covertly; some restriction has to be imposed to prevent covert channels [25, 20]. Some mandatory access control (MAC) mechanisms support dynamic labels but cannot prevent *implicit flows* arising from control flow paths not taken at run time [4, 11].

JFlow [13] and its successor, Jif [15] are the only implemented security-typed languages supporting dynamic labels. However, although the Jif type system is designed to control the new information channels that dynamic labels create, it has not been proved to enforce secure information flow. Further, the dynamic label mechanism in Jif has limitations that impair expressiveness and efficiency.

In this paper, we propose an expressive language-based mechanism for securely manipulating information with dynamic security labels. The mechanism is formalized in a core language (based on the typed lambda calculus) with first-class label values, dependent security types and run-time label tests. Further, we prove that any well-typed program of the core language is secure because it satisfies noninterference. This is the first noninterference proof for a security-typed language in which general security labels can be manipulated and tested dynamically, though a noninterference result has been obtained for a simpler language supporting the related notion of dynamic *principals* [22].

Some previous MAC systems have supported dynamic security classes as part of a downgrading mechanism [21]; in this work the two mechanisms are

considered orthogonal. While downgrading is important, it is useful to treat it as a separate mechanism so that dynamic manipulation of labels does not necessarily destroy noninterference.

The remainder of this paper is organized as follows. Section 2 presents some background on lattice label models and security type systems. Section 3 introduces the core language λ_{DSec} and uses sample λ_{DSec} programs to show some important applications of dynamic labels. Section 4 describes the type system of λ_{DSec} and the noninterference result. Section 5 covers related work, and Section 6 concludes.

2. Background

Static information flow analysis can be formalized as a security type system, in which security levels of data are represented by security type annotations, and information flow control is performed through type checking.

2.1 Security classes

We assume that security requirements for confidentiality or integrity are defined by associating *security classes* with users and with the resources that programs access. These security classes form a lattice $\mathcal{L}$. We write $k \sqsubseteq k'$ to indicate that security class k' is at least as restrictive as another security class k. In this case it is safe to move information from security class k to k', because restrictions on the use of the data are preserved. To control data derived from sources with classes k and k', the least restrictive security class that is at least as restrictive as both k and k' is assigned. This is the least upper bound, or join, written $k \sqcup k'$.

2.2 Labels

Type systems for confidentiality or integrity are concerned with tracking information flows in programs. Types are extended with security *labels* that denote security classes. A label ℓ appearing in a program may be simply a constant security class k, or a more complex expression that denotes a security class. The notation $\ell_1 \sqsubseteq \ell_2$ means that ℓ_2 denotes a security class that is at least as restrictive as that denoted by ℓ_1.

Because a given security class may be denoted by different labels, the relation $\sqsubseteq$ generates a lattice of *equivalence classes* of labels with $\sqcup$ as the *join* (least upper bound) operator. Two labels ℓ_1 and ℓ_2 are equivalent, written $\ell_1 \approx \ell_2$, if $\ell_1 \sqsubseteq \ell_2$ and $\ell_2 \sqsubseteq \ell_1$. The join of two labels, $\ell_1 \sqcup \ell_2$, denotes the security class that is the join of the security classes that ℓ_1 and ℓ_2 denote. For example, if x has label ℓ_x and y has label ℓ_y, then the sum x+y is given the label $\ell_x \sqcup \ell_y$.

2.3 Security type systems for information flow

Security type systems can be used to enforce security information flows statically. Information flows in programs may be explicit flows such as assignments, or *implicit flows* [5] arising from the control flow of the program. Consider an assignment statement `x=y`, which contains an information flow from `y` to `x`. Then the typing rule for the assignment statement requires that $\ell_y \sqsubseteq \ell_x$, which means the security level of `y` is lower than the security level of `x`, guaranteeing the information flow from `y` to `x` is secure.

One advantage of static analysis is more precise control of implicit flows. Consider a simple conditional:

$$\texttt{if b then x} = \texttt{true else x} = \texttt{false}$$

Although there is no direct assignment from `b` to `x`, this expression has an implicit flow from `b` into `x`. A standard technique for controlling implicit flows is to introduce a *program-counter label* [4], written *pc*, which indicates the security level of the information that can be learned by knowing the control flow path taken thus far. In this example, the branch taken depends on `b`, so the *pc* in the `then` and `else` clauses will be joined with ℓ_b, the label of `b`. The type system ensures that any effect of expression e has a label at least as restrictive as its *pc*. In other words, an expression e cannot generate any effects observable to users who should not know the current program counter. In this example, the assignments to `x` will be permitted only if $pc \sqsubseteq \ell_x$, which ensures $\ell_b \sqsubseteq \ell_x$.

3. The λ_{DSec} language

The core language λ_{DSec} is a security-typed lambda calculus that supports first-class dynamic labels. In λ_{DSec}, labels are terms that can be manipulated and checked at run time. Furthermore, label terms can be used as statically analyzed type annotations. Syntactic restrictions are imposed on label terms to increase the practicality of type checking, following the approach used by Xi and Pfenning in $\mathtt{ML}_0^{\Pi}(C)$ [26].

From the computational standpoint, λ_{DSec} is fairly expressive, because it supports both first-class functions and state, which together are sufficient to encode recursive functions.

3.1 Syntax

The syntax of λ_{DSec} is given in Figure 1. We use the name k to range over a lattice of label values $\mathcal{L}$ (more precisely, a join semi-lattice with bottom element $\bot$), x, y to range over variable names $\mathcal{V}$, and m to range over a space of memory addresses $\mathcal{M}$.

Base Labels	k	$\in$	$\mathcal{L}$
Variables	x, y	$\in$	$\mathcal{V}$
Locations	m	$\in$	$\mathcal{M}$
Labels	ℓ, pc	$::=$	$k \mid x \mid \ell_1 \sqcup \ell_2$
Constraints	C	$::=$	$\ell_1 \sqsubseteq \ell_2 \,,\, C \mid \epsilon$
Base Types	β	$::=$	$\texttt{int} \mid \texttt{label} \mid \texttt{unit} \mid (x{:}\tau_1)[C] * \tau_2 \mid \tau\ \texttt{ref} \mid (x{:}\tau_1) \xrightarrow{C\,;pc} \tau_2$
Security Types	τ	$::=$	β_ℓ
Values	v	$::=$	$x \mid n \mid m^\tau \mid \lambda(x{:}\tau)[C\,;pc].\, e \mid () \mid k \mid (x{=}v_1[C],\, v_2{:}\tau)$
Expressions	e	$::=$	$v \mid \ell_1 \sqcup \ell_2 \mid e_1\, e_2 \mid !e \mid e_1 := e_2 \mid \texttt{ref}^\tau e$
		$\mid$	$\texttt{if}\ \ell_1 \sqsubseteq \ell_2\ \texttt{then}\ e_1\ \texttt{else}\ e_2 \mid \texttt{let}\ (x, y){=}e_1\ \texttt{in}\ e_2$

Figure 1. Syntax of λ_{DSec}

To make the lattice explicit, we write $\mathcal{L} \models k_1 \sqsubseteq k_2$ to mean that k_2 is at least as restrictive as k_1 in $\mathcal{L}$, and $\mathcal{L} \models k = k_1 \sqcup k_2$ to mean k is the join of k_1 and k_2 in $\mathcal{L}$. The least and greatest elements of $\mathcal{L}$ are $\bot$ and $\top$. Any non-trivial label lattice contains at least two points L and H where $H \not\sqsubseteq L$. Intuitively, the label L describes what information is observable by *low-security users* who are to be prevented from seeing confidential information. Thus, *low-security* data has a label bounded above by L; *high-security* data has a label (such as H) not bounded by L.

In λ_{DSec}, a label can be either a label value k, a variable x, or the join of two other labels $\ell_1 \sqcup \ell_2$. For example, L, x, and $L \sqcup x$ are all valid labels, and $L \sqcup x$ can be interpreted as a security policy that is as restrictive as both L and x. The security type $\tau = \beta_\ell$ is the base type β annotated with label ℓ. The base types include integers, unit, labels, functions, references and products.

The function type $(x : \tau_1) \xrightarrow{C\,;pc} \tau_2$ is a dependent type since τ_1, τ_2, C and *pc* may mention x. The component C is a set of *label constraints* each with the form $\ell_1 \sqsubseteq \ell_2$; they must be satisfied when the function is invoked. The *pc* component is a lower bound on the memory effects of the function, and an upper bound on the *pc* label of the caller. Consequently, a function is not able to leak information about where it is called. Without the annotations C and *pc*, this kind of type is sometimes written as $\Pi x{:}\tau_1.\tau_2$ [12].

The product type $(x : \tau_1)[C] * \tau_2$ is also a dependent type in the sense that occurrences of x can appear in τ_1, τ_2 and C. The component C is a set of label constraints that any value of the product type must satisfy. If τ_2 does not contain x and C is empty, the type may be written as the more familiar $\tau_1 * \tau_2$. Without the annotation C, this kind of type is sometimes written $\Sigma x{:}\tau_1.\tau_2$ [12].

In λ_{DSec}, values include integers n, typed memory locations m^τ, functions $\lambda(x{:}\tau)[C\,;pc].\, e$, the unit value $()$, constant labels k, and pairs $(x{=}v_1[C],\, v_2{:}\tau)$. A function $\lambda(x : \tau)[C\,;pc].\, e$ has one argument x with type τ, and the

components C and *pc* have the same meanings as those in function types. The empty constraint set C or the top *pc* can be omitted. A pair $(x = v_1[C],\ v_2 : \tau)$ contains two values v_1 and v_2. The second element v_2 has type τ and may mention the first element v_1 by the name x. The component C is a set of label constraints that the first element of the pair must satisfy.

Expressions include values v, variables x, the join of two labels $\ell_1 \sqcup \ell_2$, applications $e_1\ e_2$, dereferences $!e$, assignments $e_1 := e_2$, references $\texttt{ref}^{\tau} e$, label-test expressions $\texttt{if}\ \ell_1 \sqsubseteq \ell_2\ \texttt{then}\ e_1\ \texttt{else}\ e_2$, and product destructors $\texttt{let}\ (x, y) = v\ \texttt{in}\ e_2$.

The label-test expression $\texttt{if}\ \ell_1 \sqsubseteq \ell_2\ \texttt{then}\ e_1\ \texttt{else}\ e_2$ is used to examine labels. At run time, if the value of ℓ_2 is a constant label at least as restrictive as the value of ℓ_1, then e_1 is evaluated; otherwise, e_2 is evaluated. Consequently, the constraint $\ell_1 \sqsubseteq \ell_2$ can be assumed when type-checking e_1.

The product destructor $\texttt{let}\ (x, y) = e_1\ \texttt{in}\ e_2$ unpacks the result of e_1, which is a pair, assigns the first element to x and the second to y, and evaluates e_2.

3.2 Operational Semantics

The small-step operational semantics of λ_{DSec} is given in Figure 2. Let M represent a memory that is a finite map from typed locations to closed values, and let $\langle e,\ M\rangle$ be a machine configuration. Then a small evaluation step is a transition from $\langle e,\ M\rangle$ to another configuration $\langle e',\ M'\rangle$, written $\langle e,\ M\rangle \longmapsto \langle e',\ M'\rangle$.

It is necessary to restrict the form of $\langle e,\ M\rangle$ to avoid using undefined memory locations. Let $\mathit{loc}(e)$ represent the set of memory locations appearing in e. A memory M is well-formed if every address m appears at most once in $\mathit{dom}(M)$, and for any m^{τ} in $\mathit{dom}(M)$, $\mathit{loc}(M(m^{\tau})) \subseteq \mathit{dom}(M)$. By induction we can prove that evaluation preserves memory well-formedness.

The notation $e[v/x]$ indicates capture-avoiding substitution of value v for variable x in expression e. Unlike in the typed lambda calculus, $e[v/x]$ may generate a syntactically ill-formed expression if x appears in type annotations inside e, and v is not a label. However, this is not a problem because the type system of λ_{DSec} guarantees that a well-typed expression can only be evaluated to another well-typed and thus well-formed expression.

The notation $M(m^{\tau})$ denotes the value of location m^{τ} in M, and the notation $M[m^{\tau} \mapsto v]$ denotes the memory obtained by assigning v to m^{τ} in M.

The evaluation rules are standard. In rule (E3), notation *address-space*(M) represents the set of location names in M, that is, $\{m \mid \exists \tau\ s.t.\ m^{\tau} \in \mathit{dom}(M)\}$. In rule (E8), v_2 may mention x, so substituting v_2 for y in e is performed before substituting v_1 for x. The variable name in the product value matches x so that no variable substitution is needed when assigning v_1 and v_2 to x and y. In rule

$$[E1]\ \frac{\mathcal{L} \models k = k_1 \sqcup k_2}{\langle k_1 \sqcup k_2,\ M\rangle \longmapsto \langle k,\ M\rangle} \qquad [E3]\ \frac{m \notin \textit{address-space}(M)}{\langle \mathtt{ref}^{\tau} v,\ M\rangle \longmapsto \langle m^{\tau},\ M[m^{\tau} \mapsto v]\rangle}$$

$$[E2]\ \langle !m^{\tau},\ M\rangle \longmapsto \langle M(m^{\tau}),\ M\rangle \qquad [E4]\ \langle m^{\tau} := v,\ M\rangle \longmapsto \langle (),\ M[m^{\tau} \mapsto v]\rangle$$

$$[E5]\ \langle (\lambda(x{:}\tau)[C\,;pc].\,e)\ v,\ M\rangle \longmapsto \langle e[v/x],\ M\rangle$$

$$[E6]\ \frac{\mathcal{L} \models k_1 \sqsubseteq k_2}{\langle \mathtt{if}\ k_1 \sqsubseteq k_2\ \mathtt{then}\ e_1\ \mathtt{else}\ e_2,\ M\rangle \longmapsto \langle e_1,\ M\rangle}$$

$$[E7]\ \frac{\mathcal{L} \models k_1 \not\sqsubseteq k_2}{\langle \mathtt{if}\ k_1 \sqsubseteq k_2\ \mathtt{then}\ e_1\ \mathtt{else}\ e_2,\ M\rangle \longmapsto \langle e_2,\ M\rangle}$$

$$[E8]\ \langle \mathtt{let}\ (x,y) = (x = v_1[C],\ v_2{:}\tau)\ \mathtt{in}\ e,\ M\rangle \longmapsto \langle e[v_2/y][v_1/x],\ M\rangle$$

$$[E9]\ \frac{\langle e,\ M\rangle \longmapsto \langle e',\ M'\rangle}{\langle E[e],\ M\rangle \longmapsto \langle E[e'],\ M'\rangle}$$

$$\begin{array}{rcl} E[\cdot] & ::= & [\cdot]\ e \mid v\ [\cdot] \mid [\cdot] := e \mid v := [\cdot] \mid\ !\,[\cdot] \mid \mathtt{ref}^{\tau}\ [\cdot] \mid [\cdot] \sqcup \ell_2 \mid k_1 \sqcup [\cdot] \\ & \mid & \mathtt{if}\ [\cdot] \sqsubseteq \ell_2\ \mathtt{then}\ e_1\ \mathtt{else}\ e_2 \mid \mathtt{if}\ k_1 \sqsubseteq [\cdot]\ \mathtt{then}\ e_1\ \mathtt{else}\ e_2 \\ & \mid & \mathtt{let}\ (x,y) = [\cdot]\ \mathtt{in}\ e \end{array}$$

Figure 2. Small-step operational semantics of λ_{DSec}

(E9), E represents an evaluation context, a term with a single hole in redex position, and the syntax of E specifies the evaluation order.

3.3 Example: multilevel I/O channels

As discussed in Section 1, dynamic labels are vital for precisely controlling information flows between security-typed programs and the external environment. When information is exported outside a program through an I/O channel, the receiver might want to know the exact label of the information, which calls for *multilevel communication channels* [6] unambiguously pairing the information sent or received with its corresponding security label. Supporting multilevel channels is one of the basic requirements for a MAC system [6].

In λ_{DSec}, a multilevel channel can be encoded by a memory reference of type $((x : \mathtt{label}_x) * \mathtt{int}_x)_{\perp}\ \mathtt{ref}$, which stores a pair composed of an integer value and its label. The confidentiality of the integer component is protected by the label component, since extracting the integer from such a pair requires testing the label component:

$$\lambda z{:}((x{:}\mathtt{label}_x) * \mathtt{int}_x)_{\perp}.\,\mathtt{let}\ (x,y) = z\ \mathtt{in}\ \mathtt{if}\ x \sqsubseteq L\ \mathtt{then}\ m^{\mathtt{int}_L} := y\ \mathtt{else}\ ()$$

In the above code, the constraint $x \sqsubseteq L$ must be satisfied in order to store the integer component in $m^{\mathtt{int}_L}$. Since the readability of the integer depends on the value of x, letting x recursively label itself ensures that all the authorized readers of the integer component can test x and retrieve the integer.

$$[C1]\ \frac{\mathcal{L} \models k_1 \sqsubseteq k_2}{C \vdash k_1 \sqsubseteq k_2} \qquad [C2]\ \frac{\ell_1 \sqsubseteq \ell_2 \in C}{C \vdash \ell_1 \sqsubseteq \ell_2}$$

$$[C3]\ C \vdash \ell \sqsubseteq \top \qquad [C4]\ C \vdash \bot \sqsubseteq \ell \qquad [C5]\ C \vdash \ell \sqsubseteq \ell \sqcup \ell'$$

$$[C6]\ \frac{C \vdash \ell_1 \sqsubseteq \ell_2 \quad C \vdash \ell_2 \sqsubseteq \ell_3}{C \vdash \ell_1 \sqsubseteq \ell_3} \qquad [C7]\ \frac{C \vdash \ell_1 \sqsubseteq \ell_3 \quad C \vdash \ell_2 \sqsubseteq \ell_3}{C \vdash \ell_1 \sqcup \ell_2 \sqsubseteq \ell_3}$$

Figure 3. Relabeling rules

Sending an integer through a multilevel channel is encoded by pairing the integer and its label and storing the pair in the reference representing the channel:

$$\lambda z{:}(((x{:}\mathtt{label}_x)) * \mathtt{int}_x)_\bot\ \mathtt{ref})_\bot.\, \lambda w{:}\mathtt{label}_w.\, \lambda(y{:}\mathtt{int}_w)[\bot].\, z := (x{=}w,\ y{:}\mathtt{int}_x)$$

Like other I/O channels, a multilevel channel may have a label that is an upper bound of the security levels of the information that can be sent through the channel. Product label constraints can be used to specify the label of a multilevel channel. For example, a bounded multilevel channel can be represented by a memory reference with type $((x{:}\mathtt{label}_x)[x \sqsubseteq \ell] * \mathtt{int}_x)_\bot\ \mathtt{ref}$, where ℓ is the label of the channel, and the constraint $x \sqsubseteq \ell$ guarantees any information stored in the reference has a security label at most as high as ℓ. Sending information through a bounded multilevel channel often needs a run-time check as in the following code:

$$\begin{array}{l}\lambda z{:}(((x{:}\mathtt{label}_x))[x \sqsubseteq \ell] * \mathtt{int}_x)_\bot\ \mathtt{ref})_\bot.\, \lambda w{:}\mathtt{label}_w. \\ \quad \lambda(y{:}\mathtt{int}_w)[\bot].\, \mathtt{if}\ w \sqsubseteq \ell\ \mathtt{then}\ z := (x{=}w,\ y{:}\mathtt{int}_x)\ \mathtt{else}\ ()\end{array}$$

4. Type system and noninterference

This section describes the type system of λ_{DSec} and formalizes the noninterference result (any well-typed program has the noninterference property), whose proof is presented in the full version of this paper [29].

4.1 Subtyping

The subtyping relationship between security types plays an important role in enforcing information flow security. Given two security types $\tau_1 = \beta_{1\ell_1}$ and $\tau_2 = \beta_{2\ell_2}$, suppose τ_1 is a subtype of τ_2, written as $\tau_1 \leq \tau_2$. Then any data of type τ_1 can be treated as data of type τ_2. Thus, data with label ℓ_1 may be treated as data with label ℓ_2, which requires $\ell_1 \sqsubseteq \ell_2$.

In λ_{DSec}, label terms have a restricted syntactic form so that they can be used as type annotations, and constraints on label terms are also type-level information that the type checker can use. Indeed, label constraints introduced in

$$[S1]\ \frac{C \vdash \tau_1 \le \tau_2 \quad C \vdash \tau_2 \le \tau_1}{C \vdash \tau_1\ \mathtt{ref} \le \tau_2\ \mathtt{ref}} \qquad [S2]\ \frac{\begin{array}{c} C \vdash \tau_2 \le \tau_1 \quad C \vdash \tau_1' \le \tau_2' \\ C \vdash pc_2 \sqsubseteq pc_1 \quad C, C_2 \vdash C_1 \end{array}}{C \vdash (x{:}\tau_1) \xrightarrow{C_1\,;\,pc_1} \tau_1' \le (x{:}\tau_2) \xrightarrow{C_2\,;\,pc_2} \tau_2'}$$

$$[S3]\ \frac{C \vdash \tau_1 \le \tau_2 \quad C \vdash \tau_1' \le \tau_2' \quad C, C_1 \vdash C_2}{C \vdash (x{:}\tau_1)[C_1] * \tau_1' \le (x{:}\tau_2)[C_2] * \tau_2'} \qquad [S4]\ \frac{C \vdash \beta_1 \le \beta_2 \quad C \vdash \ell_1 \sqsubseteq \ell_2}{C \vdash (\beta_1)_{\ell_1} \le (\beta_2)_{\ell_2}}$$

Figure 4. Subtyping rules

label-test expressions, functions and pairs are critical for precise static analysis of dynamic labels.

The type system keeps track of the set of label constraints that can be used to prove relabeling relationships between labels. Let $C \vdash \ell_1 \sqsubseteq \ell_2$ denote that $\ell_1 \sqsubseteq \ell_2$ can be inferred from the set of constraints C. The inference rules are shown in Figure 3; they are standard and consistent with the lattice properties of labels. Rule (C2) shows that all the constraints in C are assumed to be true.

Since the subtyping relationship depends on the relabeling relationship, the subtyping context also needs to include the C component. The inference rules for proving $C \vdash \tau_1 \le \tau_2$ are the rules shown in Figure 4 plus the standard reflexivity and transitivity rules.

Rules (S1)–(S3) are about subtyping on base types. These rules demonstrate the expected covariance or contravariance. In λ_{DSec}, function types contain two additional components *pc* and C, both of which are contravariant because they restrict where a function can be invoked. In rules (S2) and (S3), variable x is bound in the function and product types. For simplicity, we assume that x does not appear in C, since α-conversion can always be used to rename x to another fresh variable. This assumption also applies to the typing rules.

Rule (S4) is used to determine the subtyping on security types. The premise $C \vdash \beta_1 \le \beta_2$ is natural. The other premise $C \vdash \ell_1 \sqsubseteq \ell_2$ guarantees that coercing data from τ_1 to τ_2 does not violate information flow policies.

4.2 Typing

The type system of λ_{DSec} prevents illegal information flows and guarantees that well-typed programs have a noninterference property. The typing rules are shown in Figure 5. The notation $label(\beta_\ell) = \ell$ is used to obtain the label of a type, and the notations $\ell \sqsubseteq \tau$ and $\tau \sqsubseteq \ell$ are abbreviations for $\ell \sqsubseteq label(\tau)$ and $label(\tau) \sqsubseteq \ell$, respectively.

The typing context includes a *type assignment* Γ, a set of constraints C and the program-counter label *pc*. Γ is a finite *ordered* list of $x{:}\tau$ pairs in the order that they came into scope. For a given x, there is at most one pair $x{:}\tau$ in Γ.

$$[INT]\ \ \Gamma; C;\ pc \vdash n : \mathtt{int}_{\bot} \qquad [UNIT]\ \ \Gamma; C;\ pc \vdash () : \mathtt{unit}_{\bot}$$

$$[LABEL]\ \ \Gamma; C;\ pc \vdash k : \mathtt{label}_{\bot} \qquad [LOC]\ \ \frac{FV(\tau) = \emptyset}{\Gamma; C;\ pc \vdash m^{\tau} : (\tau\ \mathtt{ref})_{\bot}}$$

$$[VAR]\ \ \frac{x{:}\tau \in \Gamma}{\Gamma; C;\ pc \vdash x : \tau} \qquad [JOIN]\ \ \frac{\Gamma; C;\ pc \vdash \ell_1 : \mathtt{label}_{\ell_1'} \quad \Gamma; C;\ pc \vdash \ell_2 : \mathtt{label}_{\ell_2'}}{\Gamma; C;\ pc \vdash \ell_1 \sqcup \ell_2 : \mathtt{label}_{\ell_1' \sqcup \ell_2'}}$$

$$[REF]\ \ \frac{\Gamma; C;\ pc \vdash e : \tau \qquad C \vdash pc \sqsubseteq \tau}{\Gamma; C;\ pc \vdash \mathtt{ref}^{\tau} e : (\tau\ \mathtt{ref})_{\bot}} \qquad [DEREF]\ \ \frac{\Gamma; C;\ pc \vdash e : (\tau\ \mathtt{ref})_{\ell}}{\Gamma; C;\ pc \vdash !e : \tau \sqcup \ell}$$

$$[ABS]\ \ \frac{\Gamma, x{:}\tau'; C';\ pc' \vdash e : \tau}{\Gamma; C;\ pc \vdash \lambda(x{:}\tau')[C';\ pc'].\, e : ((x{:}\tau') \xrightarrow{C';\ pc'} \tau)_{\bot}}$$

$$[ASSIGN]\ \ \frac{\Gamma; C;\ pc \vdash e_1 : (\tau\ \mathtt{ref})_{\ell} \quad \Gamma; C;\ pc \vdash e_2 : \tau \quad C \vdash pc \sqcup \ell \sqsubseteq \tau}{\Gamma; C;\ pc \vdash e_1 := e_2 : \mathtt{unit}_{\bot}}$$

$$[L\text{-}APP]\ \ \frac{\begin{array}{c}\Gamma; C;\ pc \vdash e_1 : ((x{:}\mathtt{label}_{\ell'}) \xrightarrow{C';\ pc'} \tau)_{\ell} \qquad \Gamma; C;\ pc \vdash \ell_2 : \mathtt{label}_{\ell'[\ell_2/x]} \\ C \vdash pc \sqcup \ell \sqsubseteq pc'[\ell_2/x] \quad C \vdash C'[\ell_2/x] \quad x \in FV(\tau) \cup FV(\ell') \cup FV(C') \cup FV(pc')\end{array}}{\Gamma; C;\ pc \vdash e_1\ \ell_2 : \tau[\ell_2/x] \sqcup \ell}$$

$$theorem^{[APP]}\ \ \frac{\begin{array}{c}\Gamma; C;\ pc \vdash e_1 : ((x{:}\tau') \xrightarrow{C';\ pc'} \tau)_{\ell} \qquad \Gamma; C;\ pc \vdash e_2 : \tau' \\ C \vdash pc \sqcup \ell \sqsubseteq pc' \qquad C \vdash C' \qquad x \notin FV(\tau) \cup FV(\tau') \cup FV(C') \cup FV(pc')\end{array}}{\Gamma; C;\ pc \vdash e_1\ e_2 : \tau \sqcup \ell}$$

$$[IF]\ \ \frac{\begin{array}{c}\Gamma; C;\ pc \vdash \ell_i : \mathtt{label}_{\ell_i'} \quad i \in \{1,2\} \\ \Gamma;\ C, \ell_1 \sqsubseteq \ell_2;\ pc \sqcup \ell_1' \sqcup \ell_2' \vdash e_1 : \tau \qquad \Gamma; C;\ pc \sqcup \ell_1' \sqcup \ell_2' \vdash e_2 : \tau\end{array}}{\Gamma; C;\ pc \vdash \mathtt{if}\ \ell_1 \sqsubseteq \ell_2\ \mathtt{then}\ e_1\ \mathtt{else}\ e_2 : \tau \sqcup \ell_1' \sqcup \ell_2'}$$

$$[UNPACK]\ \ \frac{\Gamma; C;\ pc \vdash e_1 : ((x{:}\tau_1)[C'] * \tau_2)_{\ell} \qquad \Gamma, x{:}\tau_1 \sqcup \ell, y{:}\tau_2 \sqcup \ell;\ C, C';\ pc \vdash e_2 : \tau}{\Gamma; C;\ pc \vdash \mathtt{let}\ (x,y) = e_1\ \mathtt{in}\ e_2 : \tau}$$

$$[PROD]\ \ \frac{\begin{array}{c}\Gamma; C;\ pc \vdash v_1 : \tau_1[v_1/x] \qquad \Gamma, x{:}\tau_1 \vdash \tau_2 \\ \Gamma; C;\ pc \vdash v_2[v_1/x] : \tau_2[v_1/x] \qquad C \vdash C'[v_1/x]\end{array}}{\Gamma; C;\ pc \vdash (x = v_1[C'],\ v_2 : \tau_2) : ((x{:}\tau_1)[C'] * \tau_2)_{\bot}} \qquad [SUB]\ \ \frac{\begin{array}{c}\Gamma; C;\ pc \vdash e : \tau \\ C \vdash \tau \leq \tau'\end{array}}{\Gamma; C;\ pc \vdash e : \tau'}$$

Figure 5. Typing rules for the λ_{DSec} language

A variable appearing in a type must be a label variable. Therefore, a type τ is well-formed with respect to type assignment Γ, written $\Gamma \vdash \tau$, if Γ maps all the variables in τ to label types. The definition of well-formed labels ($\Gamma \vdash \ell$) is the same. Consider $\Gamma = x_1 : \tau_1, \ldots, x_n : \tau_n$. For any $0 \leq i \leq n$, the type τ_i may only mention label variables that are already in scope: x_1 through x_i. Therefore, Γ is well-formed if for any $0 \leq i \leq n$, τ_i is well-formed with respect to $x_1 : \tau_1, \ldots, x_i : \tau_i$. For example, "$x : \mathtt{label}_L, y : \mathtt{int}_x$" is well-formed, but "$y{:}\mathtt{int}_x, x{:}\mathtt{label}_L$" is not.

The typing assertion $\Gamma ; C ; pc \vdash e : \tau$ means that with the type assignment Γ, current program-counter label as pc, and the set of constraints C satisfied, expression e has type τ.

Rules (INT), (UNIT), (LABEL) and (LOC) are used to check values. Value v has type $\beta_\perp$ if v has base type β. Rule (LOC) requires typed location m^τ contain no label variables so that m^τ remains a constant during evaluation. This is enforced by the premise $FV(\tau) = \emptyset$, where $FV(\tau)$ denotes the set of free variables appearing in τ.

Rules (VAR), (JOIN), (REF), (DEREF), (ASSIGN), (ABS) and (SUB) are standard for a security type system [27, 17]. Due to the space limitation, we do not include the detailed descriptions of these rules, which can be found in the full paper [29].

Rule (L-APP) is used to check applications of dependent functions. Expression e_1 has a dependent function type $((x : \mathtt{label}_{\ell'}) \xrightarrow{C' ; pc'} \tau)_\ell$, where x does appear in ℓ', C', pc' or τ. As a result, rule (L-APP) needs to use $\ell'[\ell_2/x]$, $C'[\ell_2/x]$, $pc'[\ell_2/x]$ and $\tau[\ell_2/x]$, which are well-formed since ℓ_2 is a label. That also explains why e_1, with its dependent function type, cannot be applied to an arbitrary expression e_2: substituting e_2 for x in ℓ', C', pc' and τ may generate ill-formed labels or types. The expressiveness of λ_{DSec} is not substantially affected by the restriction, because the function can be applied to a variable that receives the result of an arbitrary expression. Rule (APP) applies when x does not appear in C', pc' or τ. In this case, the type of e_1 is just a normal function type, so e_1 can be applied to arbitrary terms.

Rule (PROD) is used to check product values. To check v_2, the occurrences of x in v_2 and τ_2 are both replaced by v_1, since x is not in the domain of Γ. If v_1 is not a label, then x cannot appear in τ_2. Thus, $\tau_2[v_1/x]$ is always well-formed no matter whether v_1 is a label or not. Rule (UNPACK) checks product destructors straightforwardly. After unpacking the product value, those product label constraints in C' are in scope and used for checking e_2.

Rule (IF) checks label-test expressions. The constraint $\ell_1 \sqsubseteq \ell_2$ is added into the typing context when checking the first branch e_1.

This type system satisfies the subject reduction property and the progress property. The proof is standard.

4.3 Noninterference theorem

This section formalizes the noninterference result: any well-typed program in λ_{DSec} satisfies the noninterference property (see the full paper [29] for the proof). Consider an expression e in λ_{DSec}. Suppose e has one free variable x, and $x : \tau \vdash e : \mathtt{int}_L$ where $H \sqsubseteq \tau$. Thus, the value of x is a high-security input to e, and the result of e is a low-security output. Then noninterference requires that for all values v of type τ, evaluating $e[v/x]$ in the same memory

must generate the same result, if the evaluation terminates. For simplicity, we only consider that results are integers because they can be compared outside the context of λ_{DSec}. Let $\longmapsto^*$ denote the transitive closure of the $\longmapsto$ relationship. The following theorem formalizes the claim that the type system of λ_{DSec} enforces noninterference:

THEOREM 1 (NONINTERFERENCE) *Suppose* $x : \tau \vdash e : \mathtt{int}_L$, *and* $H \sqsubseteq \tau$. *Given two arbitrary values* v_1 *and* v_2 *of type* τ, *and an initial memory* M, *if* $\langle e[v_i/x],\ M\rangle \longmapsto^* \langle v_i',\ M_i'\rangle$ *for* $i \in \{1,2\}$, *then* $v_1' = v_2'$.

The noninterference property discussed here is *termination insensitive* [19] because $e[v/x]$ is required to generate the same result only if the evaluation terminates. The type system of λ_{DSec} does not attempt to control termination and timing channels. Control of these channels is largely an orthogonal problem. Some recent work [1, 18, 28] partially addresses timing channels.

5. Related Work

Dynamic information flow control mechanisms [24, 25] track security labels dynamically and use run-time security checks to constrain information propagation. These mechanisms are transparent to programs, but cannot prevent illegal implicit flows arising from control flow paths not taken at run time.

Various general security models [10, 21, 7] have been proposed to incorporate dynamic labeling. Unlike noninterference, these models define what it means for a system to be secure according to a certain relabeling policy, which may allow downgrading labels.

Using static program analysis to check information flow was first proposed by Denning and Denning [5]; later work phrased the analysis as type checking (e.g., [16]). Noninterference was later developed as a more semantic characterization of security [8], followed by many extensions. Volpano, Smith and Irvine [23] first showed that type systems can be used to enforce noninterference, and proved a version of noninterference theorem for a simple imperative language, starting a line of research pursuing the noninterference result for more expressive security-typed languages [9, 27, 17, 3]. A more complete survey of language-based information-flow techniques can be found in [19, 29].

The Jif language [13, 15] extends Java with a type system for analyzing information flow, and aims to be a practical language for developing secure applications. However, there is not yet a noninterference proof for the type system of Jif, because of its complexity.

Banerjee and Naumann [3] proved a noninterference result for a Java-like language with simple access control primitives. Unlike in λ_{DSec}, run-time access control in this language is separate from the static label mechanism. In their language, the label of a method result may depend in limited ways on the (implicit) security state of its caller; however, it does not seem to be possible

in the language to control the flow of information from an I/O channel or file based on permissions discovered at run time.

Concurrent to our work, Tse and Zdancewic proved a noninterference result for a security-typed lambda calculus (λ_{RP}) with run-time principals [22], which can be used to construct dynamic labels. However, λ_{RP} does not support references or existential types, which makes it unable to represent dynamic security policies that may be changed at run time, such as file permissions. In addition, support for references makes λ_{DSec} more powerful than λ_{RP} computationally.

6. Conclusions

This paper formalizes computation and static checking of dynamic labels in the type system of a core language λ_{DSec} and proves a noninterference result: well-typed programs have the noninterference property. The language λ_{DSec} is the first language supporting general dynamic labels whose type system provably enforces noninterference.

Acknowledgements

The authors would like to thank Greg Morrisett, Steve Zdancewic and Amal Ahmed for their insightful suggestions. Steve Chong, Nate Nystrom, and Michael Clarkson also helped improve the presentation of this work.

References

[1] Johan Agat. Transforming out timing leaks. In *Proc. 27th ACM Symp. on Principles of Programming Languages (POPL)*, pages 40–53, Boston, MA, January 2000.

[2] Anindya Banerjee and David A. Naumann. Secure information flow and pointer confinement in a Java-like language. In *IEEE Computer Security Foundations Workshop (CSFW)*, June 2002.

[3] Anindya Banerjee and David A. Naumann. Using access control for secure information flow in a java-like language. In *Proc. 16th IEEE Computer Security Foundations Workshop*, pages 155–169, June 2003.

[4] Dorothy E. Denning. *Cryptography and Data Security*. Addison-Wesley, Reading, Massachusetts, 1982.

[5] Dorothy E. Denning and Peter J. Denning. Certification of programs for secure information flow. *Comm. of the ACM*, 20(7):504–513, July 1977.

[6] Department of Defense. *Department of Defense Trusted Computer System Evaluation Criteria*, DOD 5200.28-STD (The Orange Book) edition, December 1985.

[7] Simon Foley, Li Gong, and Xiaolei Qian. A security model of dynamic labeling providing a tiered approach to verification. In *IEEE Symposium on Security and Privacy*, pages 142–154, Oakland, CA, 1996. IEEE Computer Society Press.

[8] Joseph A. Goguen and Jose Meseguer. Security policies and security models. In *Proc. IEEE Symposium on Security and Privacy*, pages 11–20, April 1982.

[9] Nevin Heintze and Jon G. Riecke. The SLam calculus: Programming with secrecy and integrity. In *Proc. 25th ACM Symp. on Principles of Programming Languages (POPL)*, pages 365–377, San Diego, California, January 1998.

[10] John McLean. The algebra of security. In *IEEE Symposium on Security and Privacy*, pages 2–7, Oakland, California, 1988.

[11] Catherine Meadows. Policies for dynamic upgrading. In *Database Security, IV: Status and Prospects*, pages 241–250. North Holland, 1991.

[12] John C. Mitchell. *Foundations for Programming Languages*. The MIT Press, Cambridge, Massachusetts, 1996.

[13] Andrew C. Myers. JFlow: Practical mostly-static information flow control. In *Proc. 26th ACM Symp. on Principles of Programming Languages (POPL)*, pages 228–241, San Antonio, TX, January 1999.

[14] Andrew C. Myers and Barbara Liskov. A decentralized model for information flow control. In *Proc. 17th ACM Symp. on Operating System Principles (SOSP)*, pages 129–142, Saint-Malo, France, 1997.

[15] Andrew C. Myers, Lantian Zheng, Steve Zdancewic, Stephen Chong, and Nathaniel Nystrom. Jif: Java information flow. Software release. Located at `http://www.cs.cornell.edu/jif`, July 2001–2003.

[16] Jens Palsberg and Peter Ørbæk. Trust in the λ-calculus. In *Proc. 2nd International Symposium on Static Analysis*, number 983 in Lecture Notes in Computer Science, pages 314–329. Springer, September 1995.

[17] François Pottier and Vincent Simonet. Information flow inference for ML. In *Proc. 29th ACM Symp. on Principles of Programming Languages (POPL)*, pages 319–330, 2002.

[18] Andrei Sabelfeld and Heiko Mantel. Static confidentiality enforcement for distributed programs. In *Proceedings of the 9th International Static Analysis Symposium*, volume 2477 of *LNCS*, Madrid, Spain, September 2002. Springer-Verlag.

[19] Andrei Sabelfeld and Andrew Myers. Language-based information-flow security. *IEEE Journal on Selected Areas in Communications*, 21(1):5–19, January 2003.

[20] Ravi S. Sandhu and Sushil Jajodia. Honest databases that can keep secrets. In *Proceedings of the 14th National Computer Security Conference*, Washington, DC, 1991.

[21] Ian Sutherland, Stanley Perlo, and Rammohan Varadarajan. Deducibility security with dynamic level assignments. In *Proc. 2nd IEEE Computer Security Foundations Workshop*, Franconia, NH, June 1989.

[22] Stephen Tse and Steve Zdancewic. Run-time principals in information-flow type systems. In *IEEE Symposium on Security and Privacy*, Oakland, CA, May 2004.

[23] Dennis Volpano, Geoffrey Smith, and Cynthia Irvine. A sound type system for secure flow analysis. *Journal of Computer Security*, 4(3):167–187, 1996.

[24] Clark Weissman. Security controls in the ADEPT-50 time-sharing system. In *AFIPS Conference Proceedings*, volume 35, pages 119–133, 1969.

[25] John P. L. Woodward. Exploiting the dual nature of sensitivity labels. In *IEEE Symposium on Security and Privacy*, pages 23–30, Oakland, California, 1987.

[26] Hongwei Xi and Frank Pfenning. Dependent types in practical programming. In *Proc. 26th ACM Symp. on Principles of Programming Languages (POPL)*, pages 214–227, San Antonio, TX, January 1999.

[27] Steve Zdancewic and Andrew C. Myers. Secure information flow via linear continuations. *Higher Order and Symbolic Computation*, 15(2–3):209–234, September 2002.

[28] Steve Zdancewic and Andrew C. Myers. Observational determinism for concurrent program security. In *Proc. 16th IEEE Computer Security Foundations Workshop*, pages 29–43, Pacific Grove, California, June 2003.

[29] Lantian Zheng and Andrew C. Myers. Dynamic security labels and noninterference. Technical Report 2004–1924, Cornell University Computing and Information Science, 2004.