

# Adjoining Declassification and Attack Models by Abstract Interpretation

Roberto Giacobazzi and Isabella Mastroeni

Dipartimento di Informatica,  
Università di Verona,  
Strada Le Grazie 15, I-37134 Verona, Italy  
{roberto.giacobazzi, mastroeni}@sci.univr.it

**Abstract.** In this paper we prove that attack models and robust declassification in language-based security can be viewed as adjoint transformations of abstract interpretations. This is achieved by interpreting the well known Joshi and Leino's semantic approach to non-interference as a problem of making an abstraction complete relatively to a program's semantics. This observation allows us to prove that the most abstract property on confidential data which flows, here called private observation, and the most concrete harmless attacker observing public data, here called public observable, both modeled as abstractions of the program's semantics, are respectively the adjoint solutions of a completeness problem in standard abstract interpretation theory. In particular declassification corresponds to refining the given model of an attacker with the minimal amount of information in order to achieve completeness, which is non-interference, while the harmless attacker corresponds to remove this information. This proves an adjunction relation between two basic approaches to language-based security: declassification and the construction of suitable attack models, and allows us to apply relevant techniques for abstract domain transformation in language-based security.

**Keywords:** Abstract interpretation, language-based security, declassification, abstract non-interference, attack models, adjunction, completeness.

## 1 Introduction

Many security problems in language-based security are problems of confidentiality: If a user wants to keep some information confidential then he/she has to state a policy stipulating that no data visible from other users is affected by confidential data. This policy allows programs to manipulate private data, unless visible/public outputs of those programs do not improperly reveal information about the data [27]. The usual way used to show confidentiality is to prove that an attacker cannot observe any difference between the public outputs of any two executions differing only in their private inputs with the assumption that an attacker (or unauthorized user) is allowed to view only information that is

not confidential. In this case the program is said to satisfy *non-interference* [18], also referred as *secrecy* [27, 30]. In standard non-interference, the attacker can fully analyze concrete computations. In this case, any conservative type/data-flow/control-flow analysis of information flows would discard all the programs which may provide any explicit or implicit concrete flows from confidential to public resources. Standard non-interference is therefore often too strict for practical use in language-based security. In order to adapt security policies to practical cases, it would be essential to know how much an attacker may learn from a program by (statically) analyzing its input/output behavior. This idea has recently lead to the definition of the notion of *abstract non-interference* [13] and robust declassification [32]. Abstract non-interference provides a method for modeling attackers as abstract interpretations [7, 8] of the input/output program behavior, in particular it is used for characterizing the most powerful attacker which is not able to disclose confidential properties, in the following called *harmless*. Declassification corresponds to downgrade the sensitivity of data in order to accommodate with (intentional) information leakage. In [13] and [32] systematic methods have been designed for deriving and analyzing respectively attack models and declassification by characterizing what information flows from confidential to public variables. It is clear that the stronger is the attacker, the more information can be released by the program. Namely, the more concrete is the model of the harmless attacker, the more abstract is the confidential information that can be kept private. This observation gives an intuitive explanation of the adjoint relation existing between the actions of weakening attackers and declassifying private information. In particular, we can note that when we derive the most concrete attack model, then we are looking for the most concrete *public observer*, while when we derive the most abstract property the flows during computation, for characterizing abstract declassification, we are looking for the most abstract *private observable*. Indeed, the most concrete public observer is the model of the most powerful attacker that can observe only public data. While, the most abstract private observable is the minimal amount of information that a program releases during computation.

In this paper, we prove that this duality corresponds precisely to an adjunction in the lattice of abstract interpretations. This is achieved by considering abstract non-interference as a generalization of both declassification for passive attackers and attack models. In this setting we prove that, under non restrictive hypotheses, abstract non-interference corresponds precisely to making abstract interpretation complete (see [17]) relatively to the denotational semantics of programs. This derives directly from an abstract interpretation-based generalization of Joshi and Leino's approach to secure information flows [19], which makes this approach equivalent to a completeness problem. Abstract interpretation plays a key role here, providing the adequate framework where program properties can be compared by considering their relative precision. In particular, we prove that declassification and attack models are adjoint notions and they correspond respectively to the minimal complete refinement, providing the most concrete *public observer* property of the program and the minimal complete simplification, providing the most abstract *private observable* property of the program.

## 2 Basic Notions

If  $S$  and  $T$  are sets, then  $\wp(S)$  denotes the powerset of  $S$ ,  $S \times T$  denotes the Cartesian product of  $S$  and  $T$ ,  $S \setminus T$  denotes the set-difference between  $S$  and  $T$ ,  $S \subsetneq T$  denotes strict inclusion, and for a function  $f : S \rightarrow T$  and  $X \subseteq T$ ,  $f(X) \stackrel{\text{def}}{=} \{f(x) \mid x \in X\}$  and  $f^{-1}(X) \stackrel{\text{def}}{=} \{x \mid f(x) \in X\}$ . We will often denote  $f(\{x\})$  as  $f(x)$  and use lambda notation for functions. Function composition  $\lambda x. f(g(x))$  is denoted  $f \circ g$ .  $\langle P, \leq \rangle$  denotes a poset  $P$  with ordering relation  $\leq$ , while  $\langle P, \leq, \vee, \wedge, \top, \perp \rangle$  denotes a complete lattice  $P$ , with ordering  $\leq$ , *lub*  $\vee$ , *glb*  $\wedge$ , greatest element (top)  $\top$ , and least element (bottom)  $\perp$ . Often,  $\leq_P$  will be used to denote the underlying ordering of a poset  $P$ , and  $\vee_P, \wedge_P, \top_P$  and  $\perp_P$  denote the basic operations and elements if  $P$  is a complete lattice.  $\text{id} \stackrel{\text{def}}{=} \lambda x. x$  and  $\top \stackrel{\text{def}}{=} \lambda x. \top$ . If  $S \subseteq P$  then  $\downarrow S \stackrel{\text{def}}{=} \{x \in P \mid \exists y \in S. x \leq y\}$ .  $\downarrow x$  is a shorthand for  $\downarrow \{x\}$ .  $f : C \rightarrow A$  is (completely) additive if  $f$  preserves *lub*'s of all subsets of  $C$  (emptyset included). Continuity holds when  $f$  preserved *lubs*'s of chains. Co-additivity and co-continuity are dually defined.

It is well known that abstract domains can be equivalently formulated either in terms of Galois connections or closure operators [8]. A pair of functions  $\alpha : C \rightarrow A$  and  $\gamma : A \rightarrow C$  on posets, denoted  $\langle C, \alpha, A, \gamma \rangle$ , forms an *adjunction* or a *Galois connection* (GC) if for any  $x \in C$  and  $y \in A$ :  $\alpha(x) \leq_A y \Leftrightarrow x \leq_C \gamma(y)$ .  $\alpha$  (resp.  $\gamma$ ) is the *left- (right-)adjoint* to  $\gamma$  ( $\alpha$ ) and it is an additive (co-additive) function. Additive and co-additive functions  $f$  admit respectively right and left adjoint:  $f^+ \stackrel{\text{def}}{=} \lambda x. \vee \{y \mid f(y) \leq x\}$  and  $f^- \stackrel{\text{def}}{=} \lambda x. \wedge \{y \mid x \leq f(y)\}$  respectively. Remember that  $(f^+)^- = (f^-)^+ = f$  [2]. If in addition for any  $a \in A$ :  $\alpha(\gamma(a)) = a$ , then we call  $\langle C, \alpha, A, \gamma \rangle$  a Galois insertion (GI) of  $A$  in  $C$ . In GC-based abstract interpretation the concrete and abstract domains,  $C$  and  $A$ , are complete lattices [7]. An *upper (lower) closure operator*  $\rho : P \rightarrow P$  on a poset  $P$  is monotone, idempotent, and extensive:  $\forall x \in P. x \leq_P \rho(x)$  (reductive:  $\forall x \in P. x \geq_P \rho(x)$ ). The set of all upper (lower) closure operators on  $P$  is denoted by  $\text{uco}(P)$  ( $\text{lco}(P)$ ). Let  $\langle C, \leq, \vee, \wedge, \top, \perp \rangle$  be a complete lattice. Closure operators are uniquely determined by the set of their fix-points  $\rho(C)$ . For upper closures,  $X \subseteq C$  is the set of fix-points of  $\rho \in \text{uco}(C)$  iff  $X$  is a *Moore-family* of  $C$ , i.e.,  $X = \mathcal{M}(X) \stackrel{\text{def}}{=} \{\wedge S \mid S \subseteq X\}$  — where  $\wedge \emptyset = \top \in \mathcal{M}(X)$ , iff  $X$  is isomorphic to an abstract domain  $A$  in a GI  $\langle C, \alpha, A, \gamma \rangle$ , i.e.,  $A \cong \rho(C)$  with  $\iota : \rho(C) \rightarrow A$  and  $\iota^{-1} : A \rightarrow \rho(C)$  being an isomorphism, and  $\langle C, \iota \circ \rho, A, \iota^{-1} \rangle$  is the GI, i.e.,  $\rho = \gamma \circ \alpha$ . In this case  $\rho(C)$  is a complete sub-lattice of  $C$  iff  $\rho$  is additive. Dual properties can be derived for lower closures. Therefore  $\text{uco}(C)$  is isomorphic to the so called *lattice of abstract interpretations of  $C$*  [8]. If  $C$  is a complete lattice then  $\text{uco}(C)$  and  $\text{lco}(C)$  ordered point-wise are also complete lattices. For upper closures  $\langle \text{uco}(C), \sqsubseteq, \sqcup, \sqcap, \top, \text{id} \rangle$  where for every  $\rho, \eta \in \text{uco}(C)$ ,  $\{\rho_i\}_{i \in I} \subseteq \text{uco}(C)$  and  $x \in C$ :  $\rho \sqsubseteq \eta$  iff  $\forall y \in C. \rho(y) \leq \eta(y)$  iff  $\eta(C) \subseteq \rho(C)$ ;  $(\sqcap_{i \in I} \rho_i)(x) = \wedge_{i \in I} \rho_i(x)$ ; and  $(\sqcup_{i \in I} \rho_i)(x) = x \Leftrightarrow \forall i \in I. \rho_i(x) = x$ . Dual properties can be derived for  $\langle \text{lco}(C), \sqsubseteq, \sqcup, \sqcap, \lambda x. x, \lambda x. \perp \rangle$ . In the following we will find particularly convenient to identify closure operators (and therefore abstract domains) with their sets of fix-points. The *disjunctive completion* of an abstract

domain  $\rho \in uco(C)$  is the most abstract domain able to represent the concrete disjunction of its objects:  $\Upsilon(\rho) = \sqcup\{\eta \in uco(C) \mid \eta \sqsubseteq \rho \text{ and } \eta \text{ is additive}\}$ .  $\rho$  is disjunctive iff  $\Upsilon(\rho) = \rho$  (cf. [8]). Closure operators and partitions are related concepts. If  $\pi$  is a partition (viz. an equivalence relation), then  $[\cdot]_\pi$  is the corresponding equivalence class. A closure  $\eta \in uco(\wp(S))$  induces a partition on  $S$ :  $\{ [x]_\eta \mid x \in S \}$ , where  $[x]_\eta \stackrel{\text{def}}{=} \{ y \mid \eta(x) = \eta(y) \}$ . The most concrete closure that induces the same partition of values as  $\eta$  is  $\mathcal{P}(\eta) \stackrel{\text{def}}{=} \Upsilon(\{ [x]_\eta \mid x \in S \})$ .  $\eta$  is *partitioning* if  $\eta = \mathcal{P}(\eta)$  [24]. The idea is that  $\mathcal{P}(\eta)$  is the most concrete closure such that for any  $y \in \mathcal{P}(\eta(x))$ :  $\mathcal{P}(\eta(x)) = \mathcal{P}(\eta(y))$ , while in general  $\eta(y) \subseteq \eta(x)$ .

In abstract interpretation there are two equivalent ways to express the soundness of an abstraction [7]. Let  $C$  be a complete lattice,  $f : C \rightarrow C$ ,  $(C, \alpha, A, \gamma)$  be a Galois insertion, and  $f^\# : A \rightarrow A$ . Then  $(C, \alpha, A, \gamma)$  and  $f^\#$  provide a sound abstraction of  $f$  if  $\alpha \circ f \leq f^\# \circ \alpha$ , or equivalently (by adjunction) if  $f \circ \gamma \leq \gamma \circ f^\#$ . While these two definitions of soundness are equivalent, they are not equivalent when equality is required, i.e., when we consider completeness [8, 17, 15]. In the first case  $\alpha \circ f = f^\# \circ \alpha$  means that no loss of precision is accumulated by approximating the input arguments of a given semantic function; while  $f \circ \gamma = \gamma \circ f^\#$  means that no loss of precision is accumulated by approximating the result of computations on abstract objects. We follow [15] where the first is called *backward* ( $\mathcal{B}$ ) and the second is called *forward* ( $\mathcal{F}$ ) completeness. The problem of making abstract domains  $\mathcal{B}$ -complete has been solved in [17]. These results have been extended to  $\mathcal{F}$ -completeness in [15]. The key point in this construction is that there exists an either  $\mathcal{B}$  or  $\mathcal{F}$ -complete abstract function  $f^\#$  in an abstract domain  $A$  iff the best correct approximation  $\alpha \circ f \circ \gamma$  of  $f$  in  $A$  is respectively either  $\mathcal{B}$  or  $\mathcal{F}$  complete. This means that both  $\mathcal{F}$  and  $\mathcal{B}$  completeness are properties of the underlying abstract domain  $A$  relatively to the concrete function  $f$ . In a more general setting let  $f : C_1 \rightarrow C_2$  be a function on complete lattices  $C_1$  and  $C_2$ , and  $\rho \in uco(C_2)$  and  $\eta \in uco(C_1)$  be abstract domains.  $\langle \rho, \eta \rangle$  is a pair of  $\mathcal{B}(\mathcal{F})$ -complete abstractions for  $f$  if  $\rho \circ f = \rho \circ f \circ \eta$  ( $f \circ \eta = \rho \circ f \circ \eta$ ). In the following we denote by  $\mathcal{F}(C_1, C_2, f) \stackrel{\text{def}}{=} \{ \langle \rho, \eta \rangle \mid f \circ \eta = \rho \circ f \circ \eta \}$  and  $\mathcal{B}(C_1, C_2, f) \stackrel{\text{def}}{=} \{ \langle \rho, \eta \rangle \mid \rho \circ f = \rho \circ f \circ \eta \}$ . A pair of domain transformers can be associated with any completeness problem. We follow [11, 16] by defining a *domain refinement* and *simplification* as any monotone function  $\tau : uco(L) \rightarrow uco(L)$  such that  $X \subseteq \tau(X)$  and  $\tau(X) \subseteq X$  respectively. In [17] and [15], a constructive characterization of the most abstract refinement, called *complete shell*, and of the most concrete simplification, called *complete core*, of any domain, making it  $\mathcal{F}$  or  $\mathcal{B}$  complete, for a given continuous function  $f$ , is given as a solution of a simple domain equation. Consider the following basic operators on closures:

$$\boxed{ \begin{array}{l} R_f^{\mathcal{F}} \stackrel{\text{def}}{=} \lambda X. \mathcal{M}(f(X)) \quad \Bigg| \quad R_f^{\mathcal{B}} \stackrel{\text{def}}{=} \lambda X. \mathcal{M}(\bigcup_{y \in X} \max(f^{-1}(\downarrow y))) \\ C_f^{\mathcal{F}} \stackrel{\text{def}}{=} \lambda X. \{ y \in L \mid f(y) \subseteq X \} \quad \Bigg| \quad C_f^{\mathcal{B}} \stackrel{\text{def}}{=} \lambda X. \{ y \in L \mid \max(f^{-1}(\downarrow y)) \subseteq X \} \end{array} }$$

Let  $\ell \in \{ \mathcal{F}, \mathcal{B} \}$ . In [17] the authors proved that the only interesting cases, as far as the refinement and simplification towards  $\ell$ -completeness are concerned, are respectively the most concrete  $\beta \sqsupseteq \rho$  such that  $\langle \beta, \eta \rangle$  is  $\ell$ -complete and the most abstract  $\beta \sqsubseteq \eta$  such that  $\langle \rho, \beta \rangle$  is  $\ell$ -complete. In particular given  $\rho \in uco(C_2)$  the  $\ell$ -complete shell of  $\eta \in uco(C_1)$  is  $\mathcal{R}_f^{\ell, \rho}(\eta) \stackrel{\text{def}}{=} \eta \sqcap R_f^\ell(\rho)$  and given

$\eta \in uco(C_1)$  the  $\ell$ -complete core of  $\rho \in uco(C_2)$  is  $C_f^{\ell, \eta}(\rho) \stackrel{\text{def}}{=} \rho \sqcup C_f^{\ell}(\eta)$ . Note that, when  $f$  is additive  $\max \{ x \mid f(x) \leq y \} = \bigvee \{ x \mid f(x) \leq y \} = f^{\perp}$ , and therefore  $\mathcal{B}(C_1, C_2, f) = \mathcal{F}(C_2, C_1, f^{\perp})$  (cf. [15]). Clearly, when we consider  $f : C \rightarrow C$  and the constraint  $\eta = \rho$ , the above construction requires a fixpoint iteration on abstract domains:  $\mathcal{R}_f^{\ell}(\rho) = \text{gfp}(\lambda X. \rho \sqcap R_f^{\ell}(X))$  and  $C_f^{\ell}(\rho) = \text{lfp}(\lambda X. \rho \sqcup C_f^{\ell}(X))$  are called respectively the *absolute  $\ell$ -complete shell* and *core* of  $\rho$  for  $f$ . Note that  $\mathcal{R}_f^{\ell} \in lco(uco(C))$  and  $C_f^{\ell} \in uco(uco(C))$  (see [17]). It is worth noting that  $\ell$ -complete cores and shells are adjoint abstract domain transformers, i.e., adjoint functions on the lattice of abstract interpretations. For any  $\eta \in uco(C_1)$  and  $\rho \in uco(C_2)$ :  $C_f^{\ell}(\eta) \sqsubseteq \rho \Leftrightarrow \eta \sqsubseteq R_f^{\ell}(\rho)$ , which, by definition, implies that  $C_f^{\ell, \eta}(\rho) \sqsubseteq \rho \Leftrightarrow \eta \sqsubseteq \mathcal{R}_f^{\ell, \rho}(\eta)$ .

### 3 Information Flows in Language-Based Security

Confidential data are considered *private*, labeled with H (high-level of secrecy), while all other data are public, labeled with L (low-level of secrecy) [10]. Non-interference can be naturally expressed by using semantic models of program execution. This idea goes back to Cohen's work on *strong dependency* [6], which uses denotational semantics for modeling how information can be transmitted among variables during the execution of programs. Therefore non-interference for programs essentially means that "*a variation of confidential (high or private) input does not cause a variation of public (low) output*" [27]. When this happens, we say that the program has only *secure information flows* [1, 6, 9, 10, 19, 30]. This situation has been modeled by considering the denotational (input/output) semantics  $\llbracket P \rrbracket$  of the program  $P$ . In particular we consider programs where data are typed as private (H) or public (L). Program states in  $\Sigma$  are functions (represented as tuples) mapping variables in the set of values  $\mathbb{V}$ . Finite traces on  $\Sigma$  are denoted  $\Sigma^+$ . If  $\mathbf{T} \in \{\mathbf{H}, \mathbf{L}\}$ ,  $n = |\{x \in \text{Var}(P) \mid x : \mathbf{T}\}|$ , and  $v \in \mathbb{V}^n$ , we abuse notation by denoting  $v \in \mathbb{V}^{\mathbf{T}}$  the fact that  $v$  is a possible value for the variables with security type  $\mathbf{T}$ . Moreover, we assume that any input  $s$ , can be seen as a pair  $(h, l)$ , where  $s^{\mathbf{H}} = h$  is a value for private data and  $s^{\mathbf{L}} = l$  is a value for public data. In this case, *non-interference* can be formulated as follows.

A program $P$ is <i>secure</i> if $\forall \text{ input } s, t. s^{\mathbf{L}} = t^{\mathbf{L}} \Rightarrow (\llbracket P \rrbracket(s))^{\mathbf{L}} = (\llbracket P \rrbracket(t))^{\mathbf{L}}$
---

This problem has been formulated also as a *Partial Equivalence Relation* (PER) [28]. In this case we have that if the input data are equivalent under a given equivalent relation, then also the outputs are equivalent w.r.t. a corresponding output equivalence relation. The result is a PER on the domain of semantic functions which can be used to model non-interference as above, where the equality on public data can be generalized by considering any equivalence relation. McLean [23] treats possibilistic notions of non-interference for even non-deterministic programs in the context of trace semantics: A program is secure if the set of its traces is closed under a function *purge*, i.e., it is insensible by varying private

inputs. In [22] the different notions of possibilistic non-interference are modeled in a modular way. Ryan [25], Focardi and Gorrieri [12] all provide a comprehensive treatment of non-interference for concurrent programs in process algebras, where attackers are modeled as view relations on computation trees. The standard methods for checking non-interference are based on security-type systems and data-flow/control-flow analysis. Type-based approaches are designed in such a way that well-typed programs do not leak secrets. In a security-typed language, a type is inductively associated at compile-time with program statements in such a way that any statement showing a potential flow disclosing secrets is rejected [29, 31]. Similarly, data-flow/control-flow analysis techniques are devoted to statically discover flows of secret data into public variables [3, 4, 19, 20, 28]. All these approaches are characterized by the way they model attackers (or unauthorized users).

### 3.1 Joshi and Leino’s Semantic-Based Approach

As we said above, a program is secure if any observation of the initial and final values of  $l : L$  do not provide any information about the initial value of  $h : H$  [19]. Assume that the adversary has knowledge of the program text and of the initial and final values of  $l$ . The idea of Joshi and Leino’s semantic-based approach to language-based security is that of characterizing secure information flow as program equivalence, denoted by  $\doteq$ . They introduce a program  $\text{HH} \stackrel{\text{def}}{=} \text{“assign to } h \text{ an arbitrary value”}$ . Consider a program  $P$  for which we want to prove non-interference. The program  $\text{HH}; P$  corresponds to run  $P$  after having set  $h$  to an arbitrary value; while the program  $P; \text{HH}$  discards the final value of  $h$  resulting from the execution of  $P$ . Then a program  $P$  is said to be *secure* if

$$\text{HH} ; P ; \text{HH} \doteq P ; \text{HH} \quad (1)$$

where  $\doteq$  is the relational input/output semantic equality between programs, namely for each possible input the two programs have to show the same public output behavior. In order to understand this characterization, note that the occurrence of  $\text{HH}$  after  $P$  on both the sides of the equality indicates that only the final values of  $l$  are of interest, whereas the occurrence of  $\text{HH}$  before  $P$  on the left side of the equality indicates that the program starts with an arbitrary assignment to  $h$ . Clearly, the two programs are input/output equivalent provided that the final value of  $l$ , produced by  $P$ , does not depend on the initial value of  $h$ , which is indeed standard non-interference.

### 3.2 Robust Declassification

Declassifying information means downgrading the sensitivity of data in order to accommodate with (intentional) information leakage. Robust declassification has been introduced in [32] as a systematic method to drive declassification by characterizing what information flows from confidential to public variables. In particular the observational attacker’s capability is modeled by using equivalence relations as in PER models, and declassification of private data is obtained by

**Table 1.** Narrow and Abstract Non-Interference

$[\eta]P(\rho) \text{ if } \forall h_1, h_2 \in \mathbb{V}^{\mathbb{H}}, \forall l_1, l_2 \in \mathbb{V}^{\mathbb{L}}. \eta(l_1) = \eta(l_2) \Rightarrow \rho(\llbracket P \rrbracket(h_1, l_1)^{\perp}) = \rho(\llbracket P \rrbracket(h_2, l_2)^{\perp})$
$(\eta)P(\phi \rightsquigarrow \rho) \text{ if } \forall h_1, h_2 \in \mathbb{V}^{\mathbb{H}}, \forall l \in \mathbb{V}^{\mathbb{L}}. \rho(\llbracket P \rrbracket(\phi(h_1), \eta(l))^{\perp}) = \rho(\llbracket P \rrbracket(\phi(h_2), \eta(l))^{\perp})$

manipulating these relations in a semantic-driven way. The semantics considered is the operational semantics, defined on a transition system. The authors provide a systematic method for identifying what the attacker could observe of the concrete execution traces, by iteratively refining the initial equivalence relation on the states of the program. At this point they declassify private data in order to make the attacker *blind*, i.e., they declassify all the information that the attacker can get from the execution of the program.

### 3.3 Abstract Non-interference: Attack Models and Declassification

In [13], we introduced the notion of abstract non-interference modeling weaker information flows, attack models, and declassification. The idea is that an attacker can observe only some properties, modeled as abstract interpretations of program semantics, of public concrete values. The *model of an attacker*, also called *attacker*, is therefore a pair of abstractions  $\langle \eta, \rho \rangle$ , with  $\eta, \rho \in uco(\wp(\mathbb{V}^{\mathbb{L}}))$ , representing what an observer can see about, respectively, the input and output of a program. The notion of *narrow (abstract) non-interference* (NNI) represents the first weakening of standard non-interference relatively to a given model of an attacker. When a program  $P$  satisfies narrow non-interference we write  $[\eta]P(\rho)$ , see Table 1. The problem with this notion is that it introduces *deceptive flows* [13]. Consider, for instance,  $l := l * h^2$ , and consider the public input property of being an even number, then we can observe a variation of the output's sign due to the existence of both negative and positive even numbers, revealing flows which does not depend on the private data, here called *deceptive*. In order to avoid deceptive interference we introduce a weaker notion of non-interference, having no deceptive flows, yet modeling properties of informations flows. Namely, such that, when the attacker is able to observe the property  $\eta$  of public input, and the property  $\rho$  of public output, then no information flow concerning the property  $\phi$  of the private input is observable from the public output. Namely,  $\phi$  represents the confidential information that we want to keep secret. We call this notion *abstract non-interference* (ANI). When a program  $P$  satisfies abstract non-interference we write  $(\eta)P(\phi \rightsquigarrow \rho)$ , where  $\phi \in uco(\wp(\mathbb{V}^{\mathbb{H}}))$ , see Table 1. Note that  $[\text{id}]P(\text{id})$  models exactly (standard) non-interference. Moreover, we have that abstract non-interference is a weakening of both, standard and narrow non-interference:  $\forall \eta, \rho \in uco(\wp(\mathbb{V}^{\mathbb{L}})), \phi \in uco(\wp(\mathbb{V}^{\mathbb{H}}))$  we have  $[\text{id}]P(\text{id}) \Rightarrow (\eta)P(\phi \rightsquigarrow \rho)$  and  $[\eta]P(\rho) \Rightarrow (\eta)P(\phi \rightsquigarrow \rho)$ , while standard non-interference is not stronger than the narrow one due to deceptive interference. A proof-system has been introduced, in [14], for checking both narrow and abstract non-interference inductively on program's syntax. Moreover, in [13], two methods for deriving the most concrete output observation for a program, given

the input one, for both narrow and abstract non-interference are provided. In particular the idea is that of collecting in the same abstract object all the elements that, if distinguished, would generate a visible flow. These most concrete output observations that are not able to get information from the program  $P$ , observing  $\eta$  in input, are, respectively, denoted  $[\eta][P](\text{id})$  and  $(\eta)[P](\phi \rightsquigarrow \text{id})$ , both in  $uco(\wp(\mathbb{V}^L))$ . The following theorem is proved in [13].

**Theorem 1.**  $[\eta][P](\text{id}) \sqsubseteq \rho \Leftrightarrow [\eta]P(\rho)$ ,  $(\eta)[P](\phi \rightsquigarrow \text{id}) \sqsubseteq \rho \Leftrightarrow (\eta)P(\phi \rightsquigarrow \rho)$ .

*Example 1.* Consider the properties  $Sign$  and  $Par$ , observing, respectively, the sign and the parity of integers, and the program fragment:  $P \stackrel{\text{def}}{=} l := l * h^2$ . with security typing:  $h : \mathbb{H}$  and  $l : \mathbb{L}$  and  $\mathbb{V} = \mathbb{Z}$ . Let us check if  $(\text{id})P(\text{id} \rightsquigarrow Par)$ . Note that  $Par([\![P]\!](2, 1)^L) = Par(4) = 2\mathbb{Z}$  while  $Par([\![P]\!](3, 1)^L) = Par(9) = 2\mathbb{Z} + 1$ , which are clearly different, therefore in this case  $(\text{id})P(\text{id} \rightsquigarrow Par)$  doesn't hold. Consider  $(\text{id})P(Sign \rightsquigarrow Par)$ . Note that  $Par([\![P]\!](Sign(2), 1)^L) = Par([\![P]\!](Sign(3), 1)^L) = Par(0+) = \mathbb{Z}$ . In this case it is simple to check that  $(\text{id})P(Sign \rightsquigarrow Par)$  holds.

The PER model of non-interference can be easily viewed as a narrow non-interference, where both input and output closures are partitioning, i.e., equivalence relations. This corresponds to narrow non-interference because in PERs the equivalence is checked on the program outputs of concrete computations. Abstract non-interference provides also an abstraction of declassification. The idea is to find the most abstract property on confidential data which has to be declassified in order to guarantee secrecy. For this reason we define the set:

$$\Pi_P(\eta, \rho) \stackrel{\text{def}}{=} \{ \langle \{ h \in \mathbb{V}^H \mid \rho([\![P]\!](\langle h, \eta(l) \rangle)^L) = A \}, \eta(l) \rangle \mid l \in \mathbb{V}^L, A \in \rho \}$$

This is the set of all the pairs  $\langle H, L \rangle \in \wp(\mathbb{V}^H) \times \wp(\mathbb{V}^L)$ , such that, whenever  $\eta(l) = L$ , then for any  $h_1, h_2 \in H$ , no information flows, from private to public, are revealed. We use this set for deriving a partition of private data that guarantees secrecy. For each  $L \in \eta$ , we define  $\Pi_P(\eta, \rho)|_L \stackrel{\text{def}}{=} \{ H \mid \langle H, L \rangle \in \Pi_P(\eta, \rho) \}$ . The partition on private data corresponds to the most abstract property that flows when the property observed of the public input is  $L$ :  $\mathcal{P}(\prod_{L \in \eta} \mathcal{M}(\Pi_P(\eta, \rho)|_L))$ . In particular, it is the most abstract property that contains all the possible variations of private inputs that generate insecure information flows, and the most concrete such that each variation generates a flow. namely it uniquely represents the confidential information that flows into the public output.

*Example 2.* Consider the program fragment:  $P = l := l * h^2$ .  $\Pi_P(\text{id}, Par)$  is the set  $\{ \langle \mathbb{Z}, l \rangle \mid l \in 2\mathbb{Z} \} \cup \{ \langle 2\mathbb{Z}, l \rangle \mid l \in 2\mathbb{Z} + 1 \} \cup \{ \langle 2\mathbb{Z} + 1, l \rangle \mid l \in 2\mathbb{Z} + 1 \}$ . Therefore by using the notation above we have that if  $l \in 2\mathbb{Z}$  then  $\Pi_P(\text{id}, Par)|_l = \mathbb{Z}$  and if  $l \in 2\mathbb{Z} + 1$  then  $\Pi_P(\text{id}, Par)|_l = \{2\mathbb{Z}, 2\mathbb{Z} + 1\}$ . Therefore, the most abstract partition on private data that can be declassified is  $\{2\mathbb{Z}, 2\mathbb{Z} + 1\}$ . In other words we have that by looking at the low variables the only information that leaks about the high variables is its parity.

In order to adapt robust declassification in [32] to the abstract non-interference case, we consider passive attackers only and a semantics observing the initial and the final states of computations. We follow [32] in defining the information leaked by an equivalence relation transformer  $S[\eta, \rho]$  on  $\Sigma$  for each



$\eta, \rho \in uco(\wp(\mathbb{V}^L))$ :  $s_1 S[\eta, \rho] s_2$  iff  $s_1^L \approx_\eta s_2^L$  and  $(\forall \sigma, \delta \in \Sigma^+ . \sigma_{\vdash} = s_1 \wedge \delta_{\vdash} = s_2 \Rightarrow \sigma_{\vdash}^L \approx_\rho \delta_{\vdash}^L)$ , where  $s_1, s_2 \in \Sigma$  and given  $\sigma \in \Sigma^+$  such that  $|\sigma| = n \in \mathbb{N}$ , then  $\sigma_{\vdash} \stackrel{\text{def}}{=} \sigma_0$  and  $\sigma_{\vdash} \stackrel{\text{def}}{=} \sigma_{n-1}$ . It is simple to verify that  $s_1 S[\eta, \rho] s_2$  iff  $\eta(s_1^L) = \eta(s_2^L)$  and  $\rho(\llbracket P \rrbracket(s_1)^L) = \rho(\llbracket P \rrbracket(s_2)^L)$ . This means that abstract robust declassification *a la* [32] characterizes the information leaked in narrow abstract non-interference.

## 4 Abstract Non-interference as Completeness

Joshi and Leino’s semantic-based approach to information flows [19] provides a way to interpret abstract non-interference as the problem of making an abstraction complete [17]. By considering the denotational semantics of a program  $P$ ,  $\llbracket P \rrbracket$ , the Equation (1) becomes a *backward* completeness problem if the semantics of  $\mathbb{H}\mathbb{H}$  could be described as an abstraction. Indeed the program that associates with private variables an arbitrary value can be interpreted as the closure that abstracts the private value to the “*don’t know*” abstract value, i.e., the set of all the possible values for private variables. Therefore, we define the function  $\mathcal{H} : \wp(\mathbb{V}) \rightarrow \wp(\mathbb{V})$  in the following way (recall that  $\wp(\mathbb{V}) = \wp(\mathbb{V}^{\mathbb{H}} \times \mathbb{V}^L)$ ):  $\mathcal{H} = \lambda X. \langle \mathbb{V}^{\mathbb{H}}, X^L \rangle$ , where  $X^L \stackrel{\text{def}}{=} \{ l \mid \langle h, l \rangle \in X \}$ . It is straightforward to prove its monotonicity, idempotence and extensivity. So we can finally conclude that

$$\llbracket \mathbb{H}\mathbb{H} ; P ; \mathbb{H}\mathbb{H} \rrbracket = \mathcal{H} \circ \llbracket P \rrbracket \circ \mathcal{H} \text{ and } \llbracket P ; \mathbb{H}\mathbb{H} \rrbracket = \mathcal{H} \circ \llbracket P \rrbracket$$

Hence, non-interference can be equivalently formalized as  $\mathcal{H} \circ \llbracket P \rrbracket \circ \mathcal{H} = \mathcal{H} \circ \llbracket P \rrbracket$ . The idea is to transform  $\mathcal{H}$  in order to either refine or simplify the abstraction in order to get completeness, and therefore, abstract non-interference. This can be achieved by observing that  $\mathcal{H} = \lambda X. \langle \mathbb{T}(X^{\mathbb{H}}), \text{id}(X^L) \rangle = \lambda X. \langle \mathbb{V}^{\mathbb{H}}, X^L \rangle$  where  $X^{\mathbb{H}} \stackrel{\text{def}}{=} \{ h \mid \langle l, h \rangle \in X \}$ , i.e.,  $\mathcal{H}$  is the product of respectively the top and the bottom abstractions in the lattice of abstract interpretations. This means that the private component of  $\mathcal{H}$  can only be refined as well as we can only abstract its public one. In this context we prove that shell and core have two different and precise meanings: The core abstracts the public component, viz. characterizes the most concrete attacker that cannot disclose private properties; The shell refines the private component, viz. characterizes the most abstract property that flows.

*Example 3.* Consider  $P \stackrel{\text{def}}{=} l := 2 * h$ , where  $l : \mathbb{L}$  and  $h : \mathbb{H}$ .  $P$  violates non-interference, e.g.,  $\mathcal{H} \circ \llbracket P \rrbracket \circ \mathcal{H}(\langle 2, 3 \rangle) = \mathcal{H} \circ \llbracket P \rrbracket(\langle \mathbb{Z}, 3 \rangle) = \mathcal{H}(\langle \mathbb{Z}, 2\mathbb{Z} \rangle) = \langle \mathbb{Z}, 2\mathbb{Z} \rangle$  while  $\mathcal{H} \circ \llbracket P \rrbracket(\langle 2, 3 \rangle) = \mathcal{H}(\langle 2, 4 \rangle) = \langle \mathbb{Z}, 4 \rangle$ , where  $2\mathbb{Z} \neq 4$ . We can derive the complete core of  $\mathcal{H}$ , which makes the program secure. From [17] we have to keep only those elements whose inverse image is a fix-point of  $\mathcal{H}$ :  $\mathcal{C}_{\llbracket P \rrbracket}^{\mathbb{B}}(\mathcal{H}) = \{ \langle \mathbb{Z}, L \rangle \mid \{ \langle h, l \rangle \mid \langle h, 2h \rangle \subseteq \langle \mathbb{Z}, L \rangle \} \subseteq \mathcal{H} \}$ . Note that  $\langle H, L \rangle \in \mathcal{H}$  iff  $H = \mathbb{Z}$  and  $\langle \mathbb{Z}, L' \rangle \supseteq \langle \mathbb{Z}, 2\mathbb{Z} \rangle$  iff  $L' \supseteq 2\mathbb{Z}$ . More generally, if  $L' \subseteq 2\mathbb{Z} + 1$  then  $\langle h, 2h \rangle \in \langle \mathbb{Z}, L' \rangle$  is false for each possible  $L'$ , namely  $\{ \langle h, l \rangle \mid \llbracket P \rrbracket(\langle h, l \rangle) \subseteq \langle H', L' \rangle \} = \emptyset \subseteq \mathcal{H}$  which means that in this case  $\langle \mathbb{Z}, L' \rangle$  is kept. Therefore,  $\mathcal{C}_{\llbracket P \rrbracket}^{\mathbb{B}}(\mathcal{H}) = \{ \langle \mathbb{Z}, L \rangle \mid L \cap 2\mathbb{Z} \in \{ 2\mathbb{Z}, \emptyset \} \}$ , which corresponds to abstracting the public output in the domain that is not able to distinguish even numbers. Let  $\overline{\mathcal{H}} \stackrel{\text{def}}{=} \mathcal{C}_{\llbracket P \rrbracket}^{\mathbb{B}}(\mathcal{H})$ , then in the previous case, we have  $\overline{\mathcal{H}} \circ \llbracket P \rrbracket \circ \overline{\mathcal{H}}(\langle 2, 3 \rangle) = \overline{\mathcal{H}} \circ \llbracket P \rrbracket(\langle \mathbb{Z}, 3 \rangle) = \overline{\mathcal{H}}(\langle \mathbb{Z}, 2\mathbb{Z} \rangle) = \langle \mathbb{Z}, 2\mathbb{Z} \rangle$  and  $\overline{\mathcal{H}} \circ \llbracket P \rrbracket(\langle 2, 3 \rangle) = \overline{\mathcal{H}}(\langle 2, 4 \rangle) = \langle \mathbb{Z}, 2\mathbb{Z} \rangle$ .

*Example 4.* Consider  $P \stackrel{\text{def}}{=} l := (2h + 1) \bmod 2$ , where  $l : L$  and  $h : H$ . The program violates non-interference, since, for instance,  $\mathcal{H} \circ [P] \circ \mathcal{H}(\langle 2, 3 \rangle) = \mathcal{H} \circ [P](\langle \mathbb{Z}, 3 \rangle) = \mathcal{H}(\langle \mathbb{Z}, \{-1, 1\} \rangle) = \langle \mathbb{Z}, \{-1, 1\} \rangle$  while  $\mathcal{H} \circ [P](\langle 2, 3 \rangle) = \mathcal{H}(\langle 2, 1 \rangle) = \langle \mathbb{Z}, 1 \rangle$  and  $\{-1, 1\} \neq 1$ . We compute the complete shell of  $\mathcal{H}$ , characterizing the flowing property of private information, namely we add all the inverse images of the elements in  $\mathcal{H}$ .

$$\mathcal{R}_{[P]}^{\mathcal{B}}(\mathcal{H}) = \mathcal{H} \sqcap \mathcal{M}(\bigcup_{L' \in \wp(\mathbb{V}^L)} \{ \langle h, l \rangle \mid \langle h, 2h + 1 \bmod 2 \rangle \in \langle \mathbb{Z}, L' \rangle \})$$

If  $-1 \notin L'$ , then  $\{ \langle h, l \rangle \mid \langle h, 2h + 1 \bmod 2 \rangle \in \langle \mathbb{Z}, L' \rangle \} = \langle \mathbb{Z}_0^+, \mathbb{Z} \rangle$ ,  $\mathbb{Z}_0^+ \stackrel{\text{def}}{=} \mathbb{Z}^+ \cup \{0\}$ . If  $1 \notin L'$ , then  $\{ \langle h, l \rangle \mid \langle h, 2h + 1 \bmod 2 \rangle \in \langle \mathbb{Z}, L' \rangle \} = \langle \mathbb{Z}^-, \mathbb{Z} \rangle$ . Finally, if  $1, -1 \notin L'$  we have  $\{ \langle h, l \rangle \mid \langle h, 2h + 1 \bmod 2 \rangle \in \langle \mathbb{Z}, L' \rangle \} = \emptyset$ .

Hence  $\bigcup_{L' \in \wp(\mathbb{V}^L)} \{ \langle h, l \rangle \mid \langle h, 2h + 1 \bmod 2 \rangle \in \langle \mathbb{Z}, L' \rangle \} = \{ \langle \mathbb{Z}_0^+, \mathbb{Z} \rangle, \langle \mathbb{Z}^-, \mathbb{Z} \rangle, \emptyset \}$ , which implies  $\mathcal{R}_{[P]}^{\mathcal{B}}(\mathcal{H}) = \mathcal{H} \cup \{ \langle H, L \rangle \mid H \in \{ \mathbb{Z}_0^+, \mathbb{Z}^- \}, L \in \wp(\mathbb{V}^L) \}$ . Let  $\overline{\mathcal{H}} \stackrel{\text{def}}{=} \mathcal{R}_{[P]}^{\mathcal{B}}(\mathcal{H})$ , then  $\overline{\mathcal{H}} \circ [P] \circ \overline{\mathcal{H}}(\langle 2, 3 \rangle) = \overline{\mathcal{H}} \circ [P](\langle \mathbb{Z}_0^+, 3 \rangle) = \overline{\mathcal{H}}(\langle \mathbb{Z}_0^+, \{1\} \rangle) = \langle \mathbb{Z}_0^+, \{1\} \rangle$  and  $\overline{\mathcal{H}} \circ [P](\langle 2, 3 \rangle) = \overline{\mathcal{H}}(\langle 2, 1 \rangle) = \langle \mathbb{Z}_0^+, \{1\} \rangle$ .

The idea is to embed the model of an attacker as given in ANI, i.e., as a pair of input/output abstractions, in  $\mathcal{H}$ . Consider  $\langle \wp(\mathbb{V}^H) \times \wp(\mathbb{V}^L), \emptyset, \langle \mathbb{V}^H, \mathbb{V}^L \rangle, \sqcup, \cap, \subseteq \rangle$ , where  $\langle H_1, L_1 \rangle \sqcup \langle H_2, L_2 \rangle \stackrel{\text{def}}{=} \langle H_1 \cup H_2, L_1 \cup L_2 \rangle$ . It is well known that there exists an obvious GI of  $\wp(\mathbb{V}^H) \times \wp(\mathbb{V}^L)$  in  $\wp(\mathbb{V}^H \times \mathbb{V}^L)$ , corresponding to the closure:  $\text{Split} \stackrel{\text{def}}{=}} \lambda X. \{ \langle x_1, x_2 \rangle \mid \exists y. \langle x_1, y \rangle \in X, \exists z. \langle z, x_2 \rangle \in X \}$ . Consider the closure  $\rho \in \text{uco}(\wp(\mathbb{V}^L))$ . We define  $\mathcal{H}_\rho \in \text{uco}(\wp(\mathbb{V}^H) \times \wp(\mathbb{V}^L))$ :

$$\mathcal{H}_\rho \stackrel{\text{def}}{=}} \lambda X. \langle \mathbb{V}^H, \rho(X^L) \rangle$$

Note that  $\mathcal{H} = \mathcal{H}_{\text{id}}$ ,  $\mathcal{H}_\rho \in \text{uco}(\wp(\mathbb{V}^H \times \mathbb{V}^L))$ , and for any pair of disjunctive closures  $\eta, \rho \in \text{uco}(\mathbb{V}^L)$  and for all  $\langle h, l \rangle \in \mathbb{V}$ :  $\mathcal{H}_\rho \circ [P] \circ \mathcal{H}_\eta(\langle h, l \rangle) = \mathcal{H}_\rho \circ [P](\langle h, l \rangle) \Leftrightarrow \mathcal{H}_\rho \circ [P] \circ \mathcal{H}_\eta = \mathcal{H}_\rho \circ [P]$ .

**Theorem 2.** *Let  $\rho, \eta \in \text{uco}(\wp(\mathbb{V}^L))$ .*

1.  $[\eta]P(\rho) \Leftrightarrow \mathcal{H}_\rho \circ [P] \circ \mathcal{H}_\eta = \mathcal{H}_\rho \circ [P]$ ;
2. *If  $\rho$  is disjunctive and  $\eta$  is partitioning:  $[\eta]P(\rho) \Rightarrow \mathcal{H}_\rho \circ [P] \circ \mathcal{H}_\eta = \mathcal{H}_\rho \circ [P]$ .*

This result proves that narrow abstract non-interference is weaker than the generalization of Joshi and Leino's semantics-based approach to non-interference, which is a problem of completeness. Moreover, when both  $\eta$  and  $\rho$  are equivalence relations on public data as in the PER model, i.e., partitioning closures, then the narrow abstract non-interference is equivalent to the PER model of non-interference, which is in turn an instance of a completeness problem. Theorem 2 gives a slightly weaker condition, because all partitioning closures (viz. equivalence relations) are disjunctive, but the converse does not hold in general. In order to extend Theorem 2 to model abstract non-interference we have to modify the program semantics. The idea is to consider an abstract semantics that is applied to abstract (public and private) data. Consider the closures  $\eta \in \text{uco}(\wp(\mathbb{V}^L))$  and  $\phi \in \text{uco}(\wp(\mathbb{V}^H))$ . We define the abstract semantics as  $[P]^{n, \phi} \stackrel{\text{def}}{=}} \lambda \langle h, l \rangle. [P](\phi(h), \eta(l))$ . Note that, for any pair of disjunctive closures  $\eta, \rho \in \text{uco}(\mathbb{V}^L)$  and for all  $\langle h, l \rangle \in \mathbb{V}$ :  $\mathcal{H}_\rho \circ [P]^{n, \phi} \circ \mathcal{H}_\eta(\langle h, l \rangle) = \mathcal{H}_\rho \circ [P]^{n, \phi}(\langle h, l \rangle) \Leftrightarrow \mathcal{H}_\rho \circ [P]^{n, \phi} \circ \mathcal{H}_\eta = \mathcal{H}_\rho \circ [P]^{n, \phi}$ .

**Theorem 3.** Consider  $\eta, \rho \in uco(\wp(\mathbb{V}^L))$  and  $\phi \in uco(\wp(\mathbb{V}^H))$ :

1.  $(\rho)P(\phi \rightsquigarrow \eta) \Leftarrow \mathcal{H}_\rho \circ \llbracket P \rrbracket^{\eta, \phi} \circ \mathcal{H}_\eta = \mathcal{H}_\rho \circ \llbracket P \rrbracket^{\eta, \phi}$ ;
2. If  $\rho$  and  $\eta$  are disjointive then:  $(\eta)P(\phi \rightsquigarrow \rho) \Rightarrow \mathcal{H}_\rho \circ \llbracket P \rrbracket^{\eta, \phi} \circ \mathcal{H}_\eta = \mathcal{H}_\rho \circ \llbracket P \rrbracket^{\eta, \phi}$ .

Once again abstract non-interference is weaker than the generalization of Joshi and Leino's approach to non-interference. Additivity is here sufficient in order to let these two approaches equivalent. Because the only difference in the corresponding completeness problems is due to the program semantics:  $\llbracket P \rrbracket^{\eta, \phi}$  in the case of abstract non-interference and  $\llbracket P \rrbracket^{\text{id}, \text{id}}$  in the narrow case, in the following, without loss of generality we consider the abstract non-interference case being more general. In the following we omit the apex  $\mathcal{B}$  from shells and cores, since we will consider always backward completeness.

## 5 The Most Concrete *Observer* as Completeness Core

In [13] we gave a method for systematically deriving the most concrete harmless attacker (canonical attacker) associated with a given program. By Theorem 3, the most concrete public observer, which is the canonical attacker, can be derived as the most concrete abstraction satisfying the following completeness problem:

$$\mathcal{H}_\circ \llbracket P \rrbracket^{\eta, \phi} \circ \mathcal{H}_\eta = \mathcal{H}_\circ \llbracket P \rrbracket^{\eta, \phi} \quad (2)$$

Then we have the following result which allows us to specify the canonical attacker as the fix-point of an abstract domain simplification.

**Theorem 4.** Let  $\eta \in uco(\wp(\mathbb{V}^L))$  be disjointive and  $\phi \in uco(\mathbb{V}^H)$ . Then we have  $\mathcal{C}_{\llbracket P \rrbracket^{\eta, \phi}}^{\mathcal{H}_\eta}(\mathcal{H}) = \{ \langle \mathbb{V}^H, L \rangle \mid \{ \langle h, l \rangle \mid \llbracket P \rrbracket(\langle \phi(h), \eta(l) \rangle) \subseteq \langle \mathbb{V}^H, L \rangle \} \in \mathcal{H}_\eta \}$  and

$$\left\{ L \in \wp(\mathbb{V}^L) \mid \langle \mathbb{V}^H, L \rangle \in \mathcal{C}_{\llbracket P \rrbracket^{\eta, \phi}}^{\mathcal{H}_\eta}(\mathcal{H}) \right\} = (\eta) \llbracket P \rrbracket(\phi \rightsquigarrow \text{id}).$$

*Example 5.* Consider the following program fragment, with  $l : L$  and  $h : H$ .

$$P \stackrel{\text{def}}{=} \text{while } h \text{ do } l := 2l; h := 0 \text{ endw} \quad \llbracket P \rrbracket(\langle h, l \rangle) = \begin{cases} \langle h, l \rangle & \text{if } h = 0 \\ \langle h, 2l \rangle & \text{otherwise} \end{cases}$$

We look for the core in order to make  $\langle \mathcal{H}, \mathcal{H} \rangle$  complete for the map  $\llbracket P \rrbracket^{\text{id}, \text{id}} = \llbracket P \rrbracket$ .

$$\mathcal{C}_{\llbracket P \rrbracket}^{\mathcal{H}}(\mathcal{H}) = \{ \langle \mathbb{Z}, L \rangle \mid \forall l \in \mathbb{V}^L. l \in L \Leftrightarrow 2l \in L \}$$

It is straightforward to show that  $\mathcal{C}_{\llbracket P \rrbracket}^{\mathcal{H}}(\mathcal{H})$  is the domain that abstracts the public data in the domain  $\Upsilon(\{ n\{2\}^N \mid n \in 2\mathbb{Z} + 1 \})$ , where  $\{2\}^N \stackrel{\text{def}}{=} \{ 2^k \mid k \in \mathbb{N} \}$ . Let  $\overline{\mathcal{H}} \stackrel{\text{def}}{=} \mathcal{C}_{\llbracket P \rrbracket}^{\mathcal{H}}(\mathcal{H})$ , then,  $\overline{\mathcal{H}} \circ \llbracket P \rrbracket \circ \mathcal{H}(\langle 3, 5 \rangle) = \overline{\mathcal{H}} \circ \llbracket P \rrbracket(\mathbb{Z}, 5) = \overline{\mathcal{H}}(\langle \mathbb{Z}, \{5, 10\} \rangle) = \langle \mathbb{Z}, 5\{2\}^N \rangle$ , and  $\overline{\mathcal{H}} \circ \llbracket P \rrbracket(\langle 3, 5 \rangle) = \overline{\mathcal{H}}(\langle \mathbb{Z}, \{10\} \rangle) = \langle \mathbb{Z}, 5\{2\}^N \rangle$  while we have that  $\mathcal{H}_\circ \llbracket P \rrbracket \circ \mathcal{H}(\langle 3, 5 \rangle) = \mathcal{H}_\circ \llbracket P \rrbracket(\mathbb{Z}, 5) = \mathcal{H}(\langle \mathbb{Z}, \{5, 10\} \rangle) = \langle \mathbb{Z}, \{5, 10\} \rangle$  and, on the other hand,  $\mathcal{H}_\circ \llbracket P \rrbracket(\langle 3, 5 \rangle) = \mathcal{H}(\langle \mathbb{Z}, \{10\} \rangle) = \langle \mathbb{Z}, \{10\} \rangle$ .

## 6 The Most Abstract *Observable* as Completeness Shell

We are now interested in applying the same construction for characterizing the most abstract private observable, used for defining abstract declassification as a solution of a completeness problem in abstract interpretation. Namely, we are interested in the most abstract property that can be declassified in order to guarantee abstract non-interference. By Theorem 3, this information can be obtained by solving the following completeness problem:

$$\mathcal{H}_\rho \circ \llbracket P \rrbracket^{\eta, \text{id}} \circ \mathcal{H}_\eta = \mathcal{H}_\rho \circ \llbracket P \rrbracket^{\eta, \text{id}} \tag{3}$$

**Lemma 1.** *Let  $\rho, \eta \in \text{uco}(\wp(\mathbb{V}^L))$ . Then we have*

$$\mathcal{R}_{\llbracket P \rrbracket^{\eta, \text{id}}}^{\mathcal{H}_\rho}(\mathcal{H}_\eta) = \mathcal{H}_\eta \sqcap \mathcal{M}(\{ \{ \langle h, l \rangle \mid \rho(\llbracket P \rrbracket(\langle h, \eta(l) \rangle)^L) \subseteq L \} \mid L \in \rho \} ).$$

Moreover, let  $\mathcal{R} \stackrel{\text{def}}{=} \mathcal{R}_{\llbracket P \rrbracket^{\eta, \text{id}}}^{\mathcal{H}_\rho}(\mathcal{H}_\eta)$ , then for all  $l, l' \in \mathbb{V}^L, h, h' \in \mathbb{V}^H$  we have

$$\mathcal{R}(\langle h, l \rangle) = \mathcal{R}(\langle h', l' \rangle) \quad \text{iff} \quad \rho(\llbracket P \rrbracket(\langle h, \eta(l) \rangle)^L) = \rho(\llbracket P \rrbracket(\langle h', \eta(l') \rangle)^L)$$

It is worth noting that, by Lemma 1, the partition induced by the complete shell of  $\mathcal{H}_\eta$  on  $\wp(\mathbb{V}^H \times \mathbb{V}^L)$  for Equation 3 does not affect the closure  $\eta$ . This means that the only component which is actually refined is the abstraction on private data, and this corresponds to the most abstract partitioning of private data which can be declassified. This means that any change between equivalent elements does not produce insecure flows, as stated in the following theorem.

**Theorem 5.** *Let  $\rho, \eta \in \text{uco}(\wp(\mathbb{V}^L))$  then for each  $l, l' \in \mathbb{V}^L, h, h' \in \mathbb{V}^H$  we have  $\eta(l) = \eta(l') = Y \Rightarrow (\mathcal{R}(\langle h, l \rangle) = \mathcal{R}(\langle h', l' \rangle))$  iff  $h' \in [h]_{\Pi_{\mathbb{F}}(\eta, \rho)_Y}$ .*

Next examples show how declassification can be obtained as solutions of completeness problems.

*Example 6.* Consider the program fragment:  $P \stackrel{\text{def}}{=} l := l * h^2$ , with  $l : L$  and  $h : H$ . We want to find the shell in order to make  $\langle \mathcal{H}, \mathcal{H}_{P_{ar}} \rangle$  complete for the map  $\llbracket P \rrbracket^{\text{id}, \text{id}} = \llbracket P \rrbracket$ .

$$\mathcal{R}_{\llbracket P \rrbracket}^{\mathcal{H}_{P_{ar}}}(\mathcal{H}) = \mathcal{H} \sqcap \left( \left\{ \langle \mathbb{Z}, \mathbb{Z} \rangle, \langle \mathbb{Z}, 2\mathbb{Z} \rangle \cup \langle 2\mathbb{Z}, 2\mathbb{Z} + 1 \rangle, \langle 2\mathbb{Z} + 1, 2\mathbb{Z} + 1 \rangle, \right. \right. \\ \left. \left. \langle 2\mathbb{Z} + 1, 2\mathbb{Z} \rangle, \emptyset \right\} \right)$$

This means that the reduced product generates also  $\langle 2\mathbb{Z}, 2\mathbb{Z} + 1 \rangle$  and therefore  $\langle 2\mathbb{Z}, l \rangle$  for each  $l \in 2\mathbb{Z} + 1$ . Let  $\overline{\mathcal{H}} \stackrel{\text{def}}{=} \mathcal{R}_{\llbracket P \rrbracket}^{\mathcal{H}_{P_{ar}}}(\mathcal{H})$ , then for instance, we have  $\mathcal{H}_{P_{ar}} \circ \llbracket P \rrbracket \circ \overline{\mathcal{H}}(\langle 2, 3 \rangle) = \mathcal{H}_{P_{ar}} \circ \llbracket P \rrbracket(\langle 2\mathbb{Z}, 3 \rangle) = \langle \mathbb{Z}, 2\mathbb{Z} \rangle$ , and  $\mathcal{H}_{P_{ar}} \circ \llbracket P \rrbracket(\langle 2, 3 \rangle) = \langle \mathbb{Z}, 2\mathbb{Z} \rangle$ , while  $\mathcal{H}_{P_{ar}} \circ \llbracket P \rrbracket \circ \mathcal{H}(\langle 2, 3 \rangle) = \mathcal{H}_{P_{ar}} \circ \llbracket P \rrbracket(\mathbb{Z}, 3) = \langle \mathbb{Z}, \mathbb{Z} \rangle$ . As in abstract declassification [13], this means that it is the variation of parity of the private input that generates the flow.

*Example 7.* Consider  $\rho \stackrel{\text{def}}{=} \{\mathbb{Z}, 2\mathbb{Z}, 4\mathbb{Z}, 2\mathbb{Z}+1, \emptyset\}$  and  $\eta \stackrel{\text{def}}{=} \{\mathbb{Z}, 2\mathbb{Z}, 5\mathbb{Z}, 10\mathbb{Z}, \emptyset\}$ , and consider  $P \stackrel{\text{def}}{=} \mathbf{if} (h \bmod 4) = 0 \mathbf{ then } l := l * h \mathbf{ else } l := l * (h + 1) \mathbf{ fi}$ . Compute, first, the abstract robust declassification, as was introduced in [13]:

$$II_{\mathbb{P}}(\eta, \rho) = \left\{ \begin{array}{l} \langle 4\mathbb{Z} \cup 4\mathbb{Z} + 3, 10\mathbb{Z} \rangle, \langle 4\mathbb{Z} + 1 \cup 4\mathbb{Z} + 2, 10\mathbb{Z} \rangle, \\ \langle 4\mathbb{Z} \cup 4\mathbb{Z} + 3, 5\mathbb{Z} \rangle, \langle 4\mathbb{Z} + 1, 5\mathbb{Z} \rangle, \langle 4\mathbb{Z} + 2, 5\mathbb{Z} \rangle, \\ \langle 4\mathbb{Z} \cup 4\mathbb{Z} + 3, 2\mathbb{Z} \rangle, \langle 4\mathbb{Z} + 1 \cup 4\mathbb{Z} + 2, 2\mathbb{Z} \rangle, \\ \langle 4\mathbb{Z} \cup 4\mathbb{Z} + 3, \mathbb{Z} \rangle, \langle 4\mathbb{Z} + 1, \mathbb{Z} \rangle, \langle 4\mathbb{Z} + 2, \mathbb{Z} \rangle \end{array} \right\}$$

Therefore we obtain  $II_{\mathbb{P}}(\eta, \rho)_{10\mathbb{Z}} = II_{\mathbb{P}}(\eta, \rho)_{2\mathbb{Z}} = \{4\mathbb{Z} \cup 4\mathbb{Z} + 3, 4\mathbb{Z} + 1 \cup 4\mathbb{Z} + 2\}$  and  $II_{\mathbb{P}}(\eta, \rho)_{5\mathbb{Z}} = II_{\mathbb{P}}(\eta, \rho)_{\mathbb{Z}} = \{4\mathbb{Z} \cup 4\mathbb{Z} + 3, 4\mathbb{Z} + 1, 4\mathbb{Z} + 2\}$ . Consider now the completeness shell:

$$\begin{aligned} \{ \langle h, l \rangle \mid \llbracket P \rrbracket(\langle h, \eta(l) \rangle)^L \subseteq 4\mathbb{Z} \} &= \langle 4\mathbb{Z} \cup 4\mathbb{Z} + 3, \mathbb{Z} \rangle \\ \{ \langle h, l \rangle \mid \llbracket P \rrbracket(\langle h, \eta(l) \rangle)^L \subseteq 2\mathbb{Z} \} &= \langle \mathbb{Z} \setminus 4\mathbb{Z} + 2, \mathbb{Z} \rangle \cup \langle 4\mathbb{Z} + 2, 2\mathbb{Z} \rangle \\ \{ \langle h, l \rangle \mid \llbracket P \rrbracket(\langle h, \eta(l) \rangle)^L \subseteq 2\mathbb{Z} + 1 \} &= \emptyset \end{aligned}$$

Then we have:

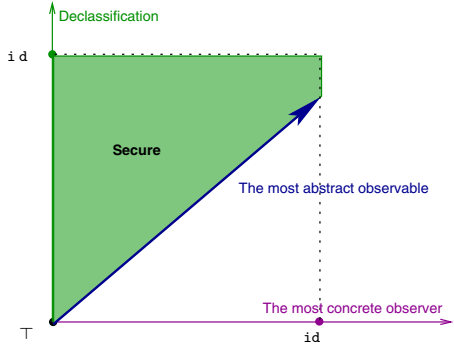
$$\mathcal{R}_{\llbracket P \rrbracket}^{\mathcal{H}_\rho}(\mathcal{H}_\eta) = \mathcal{H}_\eta \cup \left\{ \begin{array}{l} \langle \mathbb{Z} \setminus 4\mathbb{Z} + 2, \mathbb{Z} \rangle \cup \langle 4\mathbb{Z} + 2, 2\mathbb{Z} \rangle, \langle 4\mathbb{Z} \cup 4\mathbb{Z} + 3, \mathbb{Z} \rangle, \\ \langle \mathbb{Z} \setminus 4\mathbb{Z} + 2, 5\mathbb{Z} \rangle \cup \langle 4\mathbb{Z} + 2, 10\mathbb{Z} \rangle, \langle 4\mathbb{Z} \cup 4\mathbb{Z} + 3, 5\mathbb{Z} \rangle, \\ \langle 4\mathbb{Z} \cup 4\mathbb{Z} + 3, 2\mathbb{Z} \rangle, \langle 4\mathbb{Z} \cup 4\mathbb{Z} + 3, 10\mathbb{Z} \rangle \end{array} \right\}$$

For instance, consider  $5, 9 \in 4\mathbb{Z} + 1$ ,  $6, 10 \in 4\mathbb{Z} + 2$ , and note that  $\eta(10) = \eta(30) = 10\mathbb{Z}$ , and  $\eta(5) = \eta(15) = 5\mathbb{Z}$ . Note that, 5 and 6 are in the same equivalence class in the partition induced by  $II_{\mathbb{P}}(\eta, \rho)_{10\mathbb{Z}}$ , written  $5 \in [6]_{10\mathbb{Z}}$ , and indeed  $\mathcal{R}(\langle 5, 10 \rangle) = \mathcal{R}(\langle 6, 30 \rangle) = \langle \mathbb{Z}, 10\mathbb{Z} \rangle \in \mathcal{H}_\eta$ . While  $5 \in [9]_{5\mathbb{Z}} \neq [6]_{5\mathbb{Z}}$ , namely the partition induced by  $II_{\mathbb{P}}(\eta, \rho)_{5\mathbb{Z}}$  distinguishes 5 and 6, while 5 is together with 9 and 6 is together with 10, i.e.,  $10 \in [6]_{5\mathbb{Z}}$ . On the other hand, we have  $\mathcal{R}(\langle 5, 5 \rangle) = \mathcal{R}(\langle 9, 15 \rangle) = \langle \mathbb{Z} \setminus 4\mathbb{Z} + 2, 5\mathbb{Z} \rangle \cup \langle 4\mathbb{Z} + 2, 10\mathbb{Z} \rangle$  and  $\mathcal{R}(\langle 6, 5 \rangle) = \mathcal{R}(\langle 10, 15 \rangle) = \langle \mathbb{Z}, 5\mathbb{Z} \rangle \in \mathcal{H}_\eta$ .

## 7 Adjoining Observer and Observable Properties

Modeling attackers means characterizing the maximal power of an harmless attacker, i.e., an attacker which cannot disclose confidential information. Declassification, instead, means characterizing the information revealed to a fixed attacker. As we have seen in the previous sections, the model of the most concrete harmless attacker corresponds to the most concrete public observer, while abstract declassification is characterized by the most abstract private observable. Clearly there is a strong relation between these two notions, since the more powerful is the attacker and the less is the confidential information that can be kept

private. In other words the index of the partition of private data for declassification is proportional to the cardinality of the abstract domain which models the precision of the property that the attacker can observe. This phenomenon can be precisely characterized in the lattice of abstract interpretations as an adjunction. In the picture on the right we provide a graphical representation of the relation existing between the most concrete property modeling the public



observer and the most abstract property modeling the private observable. In particular, this picture represents the fact that the more powerful is the attacker, i.e., the more concrete is the observer property, the less confidential information can be kept private, i.e., the more concrete is the private observable. The picture also shows that, if the arrow represents the most abstract private observable, then when we declassify a confidential property which lays in the white area we cannot guarantee the secrecy of the program, since we are declassifying less than what is released by the semantics. When we declassify a property in the filled area instead, then we guarantee that no confidential information leakage may happen. Moreover, note that, even if the attacker is able to observe the value of public variables, then the observable property can be more abstract than the identity since the program itself can behave as a firewall for certain confidential properties, such as the square operation hides the sign. In Section 5 and 6 we proved that both problems can be viewed as instances of the problem of making abstractions complete. While the private observable for declassification is obtained by computing the completeness shell, the public observer, modeling the attacker, is obtained by computing the completeness core in the same completeness problem. These abstract domain transformers have been proved in [17] to be adjoint functions (see Sect. 2) on the lattice of abstract interpretations. The following result is therefore a consequence of Theorem 4 and 5.

**Theorem 6.** *Let  $\eta \in uco(\wp(\mathbb{V}^L))$  be a disjunctive property, and  $P$  a program. Then we have that  $\text{id} \sqsubset (\eta)[[P]](\text{id} \sim \text{id}) \Leftrightarrow \mathcal{P}(\bigcap_{L \in \eta} \mathcal{M}(\Pi_P(\eta, \text{id})|_L)) \sqsubset \mathbb{T}$ .*

This result provides a precise mathematical framework where declassification and attack models can be systematically derived and compared with each other in the lattice of abstract interpretations by applying well known methods for abstract domain design. This framework can be the basis for applying quantitative methods and metrics [5] for measuring the amount of information leaked relatively to a given attack model, or by adjunction, the precision of an attacker under the hypothesis that some information can be declassified. Recently, several papers treated the problem of defining non-interference for programs where confidential information can be explicitly declassified [26, 21]. In these cases the authors define weaker notions of non-interference in order to model confinement

problem for programs where there are intentional releases of information. Abstract non-interference, instead does not consider explicit declassification, but allows to characterize which confidential information should be declassified, at least, in order to guarantee only secure information flows.

## References

1. D. E. Bell and L. J. LaPadula. Secure computer systems: Mathematical foundations and model. Technical Report M74-244, MITRE Corp. Bedford, MA, 1973.
2. T.S. Blyth and M.F. Janowitz. *Residuation theory*. Pergamon Press, 1972.
3. C. Bodei, P. Degano, F. Nielson, and H.R. Nielson. Static analysis for secrecy and non-interference in networks of processes. In *Proc. of PaCT'01*, volume 2127 of *Lecture Notes in Computer Science*, pages 27–41. Springer-Verlag, 2001.
4. D. Clark, C. Hankin, and S. Hunt. Information flow for algol-like languages. *Computer Languages*, 28(1):3–28, 2002.
5. D. Clark, S. Hunt, and P. Malacaria. Quantitative analysis of the leakage of confidential data. In *Workshop on Quantitative Aspects of Programming Languages (QAPL '01)*, volume 59 of *Electronic Notes in Theoretical Computer Science*. Elsevier, Amsterdam, 2001.
6. E. S. Cohen. Information transmission in computational systems. *ACM SIGOPS Operating System Review*, 11(5):133–139, 1977.
7. P. Cousot and R. Cousot. Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints. In *Proc. of Conf. Record of the 4th ACM Symp. on Principles of Programming Languages (POPL '77)*, pages 238–252. ACM Press, New York, 1977.
8. P. Cousot and R. Cousot. Systematic design of program analysis frameworks. In *Proc. of Conf. Record of the 6th ACM Symp. on Principles of Programming Languages (POPL '79)*, pages 269–282. ACM Press, New York, 1979.
9. D. E. Denning. A lattice model of secure information flow. *Communications of the ACM*, 19(5):236–242, 1976.
10. D. E. Denning and P. Denning. Certification of programs for secure information flow. *Communications of the ACM*, 20(7):504–513, 1977.
11. G. Filé, R. Giacobazzi, and F. Ranzato. A unifying view of abstract domain design. *ACM Comput. Surv.*, 28(2):333–336, 1996.
12. R. Focardi and R. Gorrieri. A classification of security properties for process algebras. *Journal of Computer security*, 3(1):5–33, 1995.
13. R. Giacobazzi and I. Mastroeni. Abstract non-interference: Parameterizing non-interference by abstract interpretation. In *Proc. of the 31st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '04)*, pages 186–197. ACM-Press, NY, 2004.
14. R. Giacobazzi and I. Mastroeni. Proving abstract non-interference. In *Annual Conference of the European Association for Computer Science Logic (CSL'04)*, volume 3210, pages 280–294. Springer-Verlag, 2004.
15. R. Giacobazzi and E. Quintarelli. Incompleteness, counterexamples and refinements in abstract model-checking. In P. Cousot, editor, *Proc. of The 8th Internat. Static Analysis Symp. (SAS'01)*, volume 2126 of *Lecture Notes in Computer Science*, pages 356–373. Springer-Verlag, 2001.

16. R. Giacobazzi and F. Ranzato. Refining and compressing abstract domains. In P. Degano, R. Gorrieri, and A. Marchetti-Spaccamela, editors, *Proc. of the 24th Internat. Colloq. on Automata, Languages and Programming (ICALP '97)*, volume 1256 of *Lecture Notes in Computer Science*, pages 771–781. Springer-Verlag, Berlin, 1997.
17. R. Giacobazzi, F. Ranzato, and F. Scozzari. Making abstract interpretations complete. *J. of the ACM.*, 47(2):361–416, 2000.
18. J. A. Goguen and J. Meseguer. Security policies and security models. In *Proc. IEEE Symp. on Security and Privacy*, pages 11–20. IEEE Computer Society Press, 1982.
19. R. Joshi and K. R. M. Leino. A semantic approach to secure information flow. *Science of Computer Programming*, 37:113–138, 2000.
20. P. Laud. Semantics and program analysis of computationally secure information flow. In *Programming Languages and Systems, 10th European Symp. On Programming, ESOP*, volume 2028 of *Lecture Notes in Computer Science*, pages 77–91. Springer-Verlag, 2001.
21. P. Li and S. Zdancewic. Downgrading policies and relaxed noninterference. In *Proc. of the 32st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '05)*. ACM-Press, NY, 2005. To appear.
22. H. Mantel. Possibilistic definitions of security – an assembly kit –. In *Proc. of the IEEE Computer Security Foundations Workshop*, pages 185–199. IEEE Computer Society Press, 2000.
23. J. McLean. Proving noninterference and functional correctness using traces. *Journal of Computer security*, 1(1):37–58, 1992.
24. F. Ranzato and F. Tapparo. Strong preservation as completeness in abstract interpretation. In D. Schmidt, editor, *Proc. of the 13th European Symposium on Programming (ESOP'04)*, volume 2986 of *Lecture Notes in Computer Science*, pages 18–32. Springer-Verlag, 2004.
25. P. Ryan. Mathematical models of computer security – tutorial lectures. In R. Focardi and R. Gorrieri, editors, *Foundations of Security Analysis and Design*, volume 2171 of *Lecture Notes in Computer Science*, pages 1–62. Springer-Verlag, 2001.
26. A. Sabelfeld and A. C. Myers. A model for delimited information release. In *Proc. of the International Symp. on Software Security (ISSS'03)*, *Lecture Notes in Computer Science*. Springer-Verlag, 2004.
27. A. Sabelfeld and A.C. Myers. Language-based information-flow security. *IEEE J. on selected areas in communications*, 21(1):5–19, 2003.
28. A. Sabelfeld and D. Sands. A PER model of secure information flow in sequential programs. *Higher-Order and Symbolic Computation*, 14(1):59–91, 2001.
29. C. Skalka and S. Smith. Static enforcement of security with types. In *ICFP'00*, pages 254–267. ACM Press, New York, 2000.
30. D. Volpano. Safety versus secrecy. In *Proc. of the 6th Static Analysis Symp. (SAS'99)*, volume 1694 of *Lecture Notes in Computer Science*, pages 303–311. Springer-Verlag, 1999.
31. D. Volpano, G. Smith, and C. Irvine. A sound type system for secure flow analysis. *Journal of Computer Security*, 4(2,3):167–187, 1996.
32. S. Zdancewic and A. C. Myers. Robust declassification. In *Proc. of the IEEE Computer Security Foundations Workshop*, pages 15–23. IEEE Computer Society Press, 2001.