Control Flow Analysis of Mobile Ambients with Security Boundaries [1]

Chiara Braghin Agostino Cortesi Riccardo Focardi

Dipartimento di Informatica,
Università Ca' Foscari di Venezia,
Via Torino 155, 30173 Venezia – Mestre (Italy)
{dbraghin,cortesi,focardi}@dsi.unive.it

Abstract

A multilevel security policy is considered in the scenario of mobile systems, and modeled within "pure" Mobile Ambients calculus, in which no communication channels are present and the only possible actions are represented by the moves performed by mobile processes. The information flow property of interest is defined in terms of the possibility for a confidential ambient/data to move outside a security boundary. In a previous paper, we gave a very simple syntactic property that is sufficient to imply the absence of unwanted information flows. In this paper, a control flow analysis is defined, as a refinement of the Hansen-Jensen-Nielsons's CFA, that allows to capture boundary crossings with better accuracy.

Keywords: Mobile Ambients, Security, Static Analysis.

1 Introduction

When a user is identified and allowed to access some computer resources, an access control policy is imposed that guarantees that no information leak is possible. In particular, the system should detect "Trojan horses", i.e. (aware or unaware) malicious programs that hide their dangerous contents behind a trustworthy façade.

We focus on *Multilevel Security*, a particular *Mandatory Access Control* security policy: every entity is bound to a security level (for simplicity, we consider only two levels: high and low), and information may just flow from the low level to the high one. Typically, two access rules are imposed: (*i*) *No Read Up*, a low level entity cannot access information of a high level entity; (*ii*) *No Write Down*, a high level entity cannot leak information to a low level entity. Sometimes, these two access controls are not enough as information may be indirectly leaked, through, e.g., some system side-effect: a typical example is represented by a resource shared among the security levels which may be alternatively overloaded by some

[1] Partially supported by MURST Projects "Interpretazione Astratta, Type Systems e Analisi Control-Flow", and "Certificazione automatica di programmi mediante interpretazione astratta".

Trojan horse (causing, e.g., longer response time at all security levels) in order to transmit information to a malicious low level entity. These indirect ways of transmitting information are called *covert channels*. Figure1 summarizes this policy.

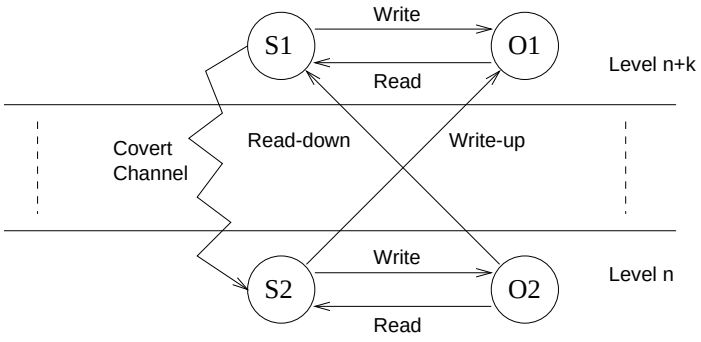

Fig. 1. Multilevel Security Policy.

In order to detect both direct and indirect information leakages, a typical approach (see, e.g., [2,6–8,10,11]) consists in directly defining what is an information flow from one level to another one. Then, it is sufficient to verify that, in any system execution, no information flow is possible from level high to level low. This is the approach we follow in this paper.

We will consider information flow security in the scenario of mobile systems. This particular setting, where code may migrate from one security level to another one, complicates even further the problem of capturing all the possible information leakages. As an example, confidential data may be read by an authorized agent which, moving around, could expose them to unexpected attacks. Moreover, the code itself could be confidential, and so not allowed to be read/executed by lower levels.

In order to study this problem as much abstractly as possible, we consider the "pure" Mobile Ambients calculus [4], in which no communication channels are present and the only possible actions are represented by the moves performed by mobile processes. This allows to study a very general notion of information flow which should be applicable also to more "concrete" versions of the calculus.

The information flow property of interest is defined in terms of the possibility for a confidential ambient/data to move outside a security boundary. In [1], a very simple syntactic property is introduced that it is sufficient to imply the absence of unwanted information flows. Here, we introduce a refinement of the control flow analysis defined in [9] that deals with the same property with improved accuracy.

The rest of the paper is organized as follows. In Section 2 we introduce the basic terminology on ambient calculus and we briefly report the control flow analysis of [9]. In Section 3, we present the model of multilevel security for mobile agents and we show how to guarantee absence of unwanted information flows. In Section 4, we introduce the Control Flow Analysis. Section 5 concludes the paper.

2 Background

In this section we introduce the basic terminology on ambient calculus and we briefly report the control flow analysis of [9].

2.1 Mobile Ambients

The Mobile Ambients calculus has been introduced in [4] with the main purpose of explicitly modeling mobility. Indeed, ambients are arbitrarily nested boundaries which can move around through suitable capabilities. The syntax of processes is given as follows, where n denotes an ambient name.

$$\begin{array}{rcll} P, Q & ::= & (\nu n)P & \text{restriction} \\ & | & \mathbf{0} & \text{inactivity} \\ & | & P \mid Q & \text{composition} \\ & | & !P & \text{replication} \\ & | & n^{\ell^a}[\, P \,] & \text{ambient} \\ & | & in^{\ell^t} n.P & \text{capability to enter } n \\ & | & out^{\ell^t} n.P & \text{capability to exit } n \\ & | & open^{\ell^t} n.P & \text{capability to open } n \end{array}$$

The labels $\ell^a \in \mathbf{Lab}^a$ on ambients and labels $\ell^t \in \mathbf{Lab}^t$ on transitions, have been introduced in the control flow analysis proposed in [9]. This is just a way of indicating "program points" and will be useful in the next section when developing the analysis.

Intuitively, the restriction $(\nu n)P$ introduces the new name n and limits its scope to P; $\mathbf{0}$ does nothing; $P \mid Q$ is P and Q running in parallel; replication provides recursion and iteration as $!P$ represents any number of copies of P in parallel. By $n^{\ell^a}[\, P \,]$ we denote the ambient named n with the process P running inside it. The capabilities $in^{\ell^t} n$ and $out^{\ell^t} n$ move their enclosing ambients in and out ambient n, respectively; the capability $open^{\ell^t} n$ is used to dissolve the boundary of a sibling ambient. $P \to Q$ denotes the reduction relation as defined in [4].

2.2 Control Flow Analysis

The control flow analysis described in [9] aims at modeling which processes can be inside what other processes. It works on pairs $(\hat{I}, \hat{H})$, where:

- The first component $\hat{I}$ is an element of $\wp(\mathbf{Lab}^a \times (\mathbf{Lab}^a \cup \mathbf{Lab}^t))$. If a process contains an ambient labelled ℓ^a having inside either a capability or an ambient labelled ℓ, then (ℓ^a, ℓ) is expected to belong to $\hat{I}$.
- The second component $\hat{H}$ keeps track of the correspondence between names and labels. If a process contains an ambient labelled ℓ^a with name n, then (ℓ^a, n) is expected to belong to $\hat{H}$.
- The pairs are component-wise partially ordered.

$$
\begin{array}{lll}
(res) & \beta^{\mathrm{CF}}_{\ell}((\nu n)P) & = \beta^{\mathrm{CF}}_{\ell}(P) \\
(zero) & \beta^{\mathrm{CF}}_{\ell}(\mathbf{0}) & = (\emptyset, \emptyset) \\
(par) & \beta^{\mathrm{CF}}_{\ell}(P \mid Q) & = \beta^{\mathrm{CF}}_{\ell}(P) \sqcup \beta^{\mathrm{CF}}_{\ell}(Q) \\
(repl) & \beta^{\mathrm{CF}}_{\ell}(!P) & = \beta^{\mathrm{CF}}_{\ell}(P) \\
(amb) & \beta^{\mathrm{CF}}_{\ell}(n^{\ell^a}[\,P\,]) & = \beta^{\mathrm{CF}}_{\ell^a}(P) \sqcup (\{(\ell, \ell^a)\}, \{(\ell^a, n)\}) \\
(in) & \beta^{\mathrm{CF}}_{\ell}(in^{\ell^t} n.P) & = \beta^{\mathrm{CF}}_{\ell}(P) \sqcup (\{(\ell, \ell^t)\}, \emptyset) \\
(out) & \beta^{\mathrm{CF}}_{\ell}(out^{\ell^t} n.P) & = \beta^{\mathrm{CF}}_{\ell}(P) \sqcup (\{(\ell, \ell^t)\}, \emptyset) \\
(open) & \beta^{\mathrm{CF}}_{\ell}(open^{\ell^t} n.P) & = \beta^{\mathrm{CF}}_{\ell}(P) \sqcup (\{(\ell, \ell^t)\}, \emptyset)
\end{array}
$$

Fig. 2. Representation Function for the Control Flow Analysis

$$
\begin{array}{lll}
(res) & (\hat{I}, \hat{H}) \models^{\mathrm{CF}} (\nu n)P & \text{iff } (\hat{I}, \hat{H}) \models^{\mathrm{CF}} P \\
(zero) & (\hat{I}, \hat{H}) \models^{\mathrm{CF}} \mathbf{0} & \text{always} \\
(par) & (\hat{I}, \hat{H}) \models^{\mathrm{CF}} P \mid Q & \text{iff } (\hat{I}, \hat{H}) \models^{\mathrm{CF}} P \ \wedge \ (\hat{I}, \hat{H}) \models^{\mathrm{CF}} Q \\
(repl) & (\hat{I}, \hat{H}) \models^{\mathrm{CF}} !P & \text{iff } (\hat{I}, \hat{H}) \models^{\mathrm{CF}} P \\
(amb) & (\hat{I}, \hat{H}) \models^{\mathrm{CF}} n^{\ell^a}[\,P\,] & \text{iff } (\hat{I}, \hat{H}) \models^{\mathrm{CF}} P \\
(in) & (\hat{I}, \hat{H}) \models^{\mathrm{CF}} in^{\ell^t} n.P & \text{iff } (\hat{I}, \hat{H}) \models^{\mathrm{CF}} P \wedge \\
 & \quad \forall \ell^a, \ell^{a'}, \ell^{a''} \in \mathbf{Lab}^a : ((\ell^a, \ell^t) \in \hat{I} \ \wedge \ (\ell^{a''}, \ell^a) \in \hat{I} \ \wedge \ (\ell^{a''}, \ell^{a'}) \in \hat{I} & \\
 & \quad\quad \wedge \ (\ell^{a'}, n) \in \hat{H}) \implies (\ell^{a'}, \ell^a) \in \hat{I} & \\
(out) & (\hat{I}, \hat{H}) \models^{\mathrm{CF}} out^{\ell^t} n.P & \text{iff } (\hat{I}, \hat{H}) \models^{\mathrm{CF}} P \wedge \\
 & \quad \forall \ell^a, \ell^{a'}, \ell^{a''} \in \mathbf{Lab}^a : ((\ell^a, \ell^t) \in \hat{I} \ \wedge \ (\ell^{a'}, \ell^a) \in \hat{I} \ \wedge \ (\ell^{a''}, \ell^{a'}) \in \hat{I} & \\
 & \quad\quad \wedge \ (\ell^{a'}, n) \in \hat{H}) \implies (\ell^{a''}, \ell^a) \in \hat{I} & \\
(open) & (\hat{I}, \hat{H}) \models^{\mathrm{CF}} open^{\ell^t} n.P & \text{iff } (\hat{I}, \hat{H}) \models^{\mathrm{CF}} P \wedge \\
 & \quad \forall \ell^a, \ell^{a'} \in \mathbf{Lab}^a : ((\ell^a, \ell^t) \in \hat{I} \ \wedge \ (\ell^a, \ell^{a'}) \in \hat{I} \ \wedge \ (\ell^{a'}, n) \in \hat{H}) & \\
 & \quad\quad \implies \left\{(\ell^a, \ell') \mid (\ell^{a'}, \ell') \in \hat{I}\right\} \subseteq \hat{I} &
\end{array}
$$

Fig. 3. Specification of the Control Flow Analysis

The analysis is defined by a representation function and a specification. [2] They are recalled, respectively, in Figure 2 and Figure 3.

The representation function mainly collects information about all ambient nestings yielded by a process, in its initial state. The representation of a process P is defined as $\beta^{\mathrm{CF}}_{\ell^a_*}(P)$, where label ℓ^a_* is a special label corresponding to the environment.

The specification mostly amounts to recursive checks of subprocesses. The *open*-capability says that if some ambient labelled ℓ^a has an *open*-capability ℓ^t on an ambient n that may apply due to the presence of a sibling ambient labelled $\ell^{a'}$ whose name is just n, then the result of performing that capability should also be recorded in $\hat{I}$. The *in* and *out* capabilities behave similarly.

The correctness of the analysis is proven by showing that every reduction of the semantics is properly mimicked in the analysis:

Theorem 2.1 *Let P and Q be two processes such that* $\beta^{\mathrm{CF}}_{\ell^a_*}(P) \sqsubseteq (\hat{I}, \hat{H}) \quad \wedge \quad (\hat{I}, \hat{H}) \models^{\mathrm{CF}} P \ \wedge \ P \to Q$ *then* $\beta^{\mathrm{CF}}_{\ell^a_*}(Q) \sqsubseteq (\hat{I}, \hat{H}) \quad \wedge \quad (\hat{I}, \hat{H}) \models^{\mathrm{CF}} Q$

Intuitively, whenever $(\hat{I}, \hat{H}) \models^{\mathrm{CF}} P$ and the representation of P is contained in $(\hat{I}, \hat{H})$, we are assured that every nesting of ambients and capabilities in every possible derivative of P is also captured in $(\hat{I}, \hat{H})$.

It is important to recall also that the resulting control flow analysis applies to any process, and that every process enjoys a *least* analysis.

3 Information Flow

In this section, we present a formalization of multilevel security in the setting of Mobile Ambients. Then, a simple syntactical property is given which allows to verify the absence of unwanted information flows.

3.1 Modelling Multilevel Security

In order to define Multilevel security in Mobile Ambients we first need to classify information into different levels of confidentiality. We do that by exploiting the labelling of ambients. In particular, we partition the set of ambient labels $\mathbf{Lab}^a$ into three disjoint sets $\mathbf{Lab}^a_H, \mathbf{Lab}^a_L$ and $\mathbf{Lab}^a_B$, which stand for *high*, *low* and *boundary* labels.

Given a process, the multilevel security policy may be established by deciding which ambients are the ones responsible for confining confidential information. These are all labelled with boundary labels from set $\mathbf{Lab}^a_B$ and we will refer to them as *boundary ambients*. Thus, all the high level ambients must be contained in a boundary ambient and labelled with labels from set $\mathbf{Lab}^a_H$. On the other side, all the external ambients are considered low level ones and consequently labelled with labels from set $\mathbf{Lab}^a_L$. This is how we will always label processes, and it corresponds to defining the security policy (what is secret, what is not, what

[2] In ambient calculus bound names may be α-converted. For the sake of simplicity, here we are assuming that ambient names are *stable*, i.e., n is indeed a representative for a class of α-convertible names. See [9] for more details on how this can be handled.

is a container of secrets). In all the examples, we will use the following notation for labels: $b \in \mathbf{Lab}^a_B, h \in \mathbf{Lab}^a_H$, $m, m' \in \mathbf{Lab}^a_L$ and $c, ch, cl, cm, cm' \in \mathbf{Lab}^t$.

As an example consider the following process:

$$P = container^b[\, hdata^h[\, out^c container.\mathbf{0}\,]\,] \mid Q$$

where Q contains some low level ambients. Ambient *container* is a boundary for the high level data *hdata* (note that data are abstractly represented as ambients). This process is an example of a direct information flow. Indeed, P may evolve to $container^b[\,] \mid hdata^h[\,] \mid Q$, where the high level *hdata* is out of any boundary ambient, thus vulnerable and potentially accessible by any ambient or process in Q. [3] This flow of high level data/ambients outside the security boundaries is exactly what we intend to control and avoid.

In distributed and mobile systems, it is unrealistic to consider a unique boundary, containing all the confidential information. As an example consider two different sites *venice* and *twente*, each with some set of confidential information that need to be protected. This can be modeled by just defining two boundary ambients, one for each site: $venice^b[\, P_1\,] \mid twente^b[\, P_2\,] \mid Q$. In order to make the model applicable, it is certainly needed a mechanism for moving confidential data from one boundary to another one. This is achieved through another boundary ambient which moves out from the first protected area and into the second one. An example follows:

$$venice^b[\, send^b[\, out^c venice.in^c twente \mid hdata^h[\,]\,]\,] \mid twente^b[\, open^c send\,] \mid Q$$

that may evolve to:

$$venice^b[\,] \mid twente^b[\, open^c send \mid send^b[\, hdata^h[\,]\,]\,] \mid Q$$

and finally to:

$$venice^b[\,] \mid twente^b[\, hdata^h[\,]\,] \mid Q$$

Note that *send* is labelled as a boundary ambient. Thus, the high level data *hdata* is always protected by boundary ambients, during the whole execution.

3.2 Verifying Absence of Information Leakage

In this section, we study how to verify that no leakage of secret data/ambients outside the boundary ambients is possible. A natural approach could be the direct application of the control flow of [9] reported in section 2.2. As a matter of fact, consider again the example presented above:

$$venice^b[\, send^b[\, out^c venice.in^c twente \mid hdata^h[\,]\,]\,] \mid twente^b[\, open^c send\,]$$

The least analysis for this process can be easily shown to be the following:

[3] Note that the presence of an ambient may be tested by trying to open it or by entering and then exiting from it. A low level ambient may thus test if *hdata* is present. This may be seen as reading such high level information.

$$\hat{I} = \{(l^a_*, b), (b, b), (b, h), (b, c)\}$$
$$\hat{h} = \{(b, venice), (b, send), (b, twente), (h, hdata)\}$$

The important thing is that h is always contained inside b, i.e., a boundary ambient. This basically proves that the system is secure and no leakage of h data may happen.

However, the fact that the analysis simply collects all the potential nesting without considering the temporal ordering of the events, may sometimes be too approximated. As an example, consider again the previous process and suppose that high level data is willing to enter some filter process, which could possibly be low level code:

$$venice^b[\, send^b[\, out^c venice.in^c twente \mid hdata^h[\, in^{ch} filter\,]\,]\,] \mid$$
$$\mid twente^b[\, open^c send\,] \mid filter^m[\, in^c send\,] \mid open^{cl} filter$$

Note that the filter behaves correctly with respect to multilevel security rules, i.e., it only enters boundaries. In particular, this means that it will never transport high level data outside the security boundaries. However, if we perform the control flow analysis we obtain the following least solution:

$$\hat{I} = \{(l^a_*, b), (l^a_*, h), (l^a_*, m), (l^a_*, cl), (l^a_*, c), (b, b), (b, h), (b, m), (b, c),$$
$$(h, ch), (m, h), (m, c)\}$$
$$\hat{H} = \{(b, venice), (b, send), (b, twente), (h, hdata), (m, filter)\}$$

Note that h appears at the environment level (i.e. the pair (l^a_*, h) occurs in $\hat{I}$), showing a potential attack. However, as observed before, there is no execution leading to such a situation. The reason why the analysis looses precision here, is due to the fact that h enters a m ambient which might be opened at the environment level, but the analysis does not capture the fact that h enters m only after it has crossed the boundary and can never return back.

In [1] we studied a (syntactic) condition on processes that is sufficient to prove the absence of leakage of secret data/ambients outside the boundary ambients. Moreover, such a condition properly deals with the situation discussed before. Let us briefly recall the main results presented in [1].

The idea is to control the $out^l n$ and $open^l n$ capabilities executed on a boundary ambient n. In particular, we require that such capabilities may only be performed by boundary ambients.

First, we characterize a subset of capability labels, in order to mark *out* and *open* capabilities that refer to boundary ambients. Let $\mathbf{Lab}^t_O \subseteq \mathbf{Lab}^t$ be the subset of labels that refer to *out* and *open* capabilities, and let $\mathbf{Lab}^t_{BM} \subseteq \mathbf{Lab}^t_O$ be a subset of this set of *out* and *open* capability labels. BM stands for *boundary moves* capabilities. Let also $\phi : \mathbf{Lab}^t \to \wp(\mathbf{Amb})$ be a function that given a capability label ℓ^t, returns the set of ambient names on which all the capabilities labelled with ℓ^t operate.

Given a process P, the conditions that should be imposed on $\beta^{\mathrm{CF}}_{\ell^a_*}(P)$ to guarantee absence

of information leakage are the following.

$$(i)\quad (\ell^a, n) \in \hat{H}, \ell^a \in \mathbf{Lab}^a_B, n \in \phi(\ell^t), \ell^t \in \mathbf{Lab}^t_O \Rightarrow \ell^t \in \mathbf{Lab}^t_{BM}$$

$$(ii)\quad (\ell, \ell') \in \hat{I}, \ell' \in \mathbf{Lab}^t_{BM} \Rightarrow \ell \in \mathbf{Lab}^a_B$$

Observe that condition (i) results in a well-formedness labelling. It requires that all the *out* and *open* capabilities that operate on boundary ambients are labelled as boundary moves (i.e., with labels in set $\mathbf{Lab}^t_{BM}$). If this condition is initially satisfied by P (i.e., by $\beta^{\mathrm{CF}}_{\ell^a_*}(P)$), then it will hold also for every derivative of P, as the labelling cannot change during process execution.

Condition (ii) requires that every *out* and *open* boundary move is executed inside a boundary ambient. Note that, in general, this may be not preserved when P evolves. Indeed, the following theorem states that also condition (ii) above is preserved, in every execution of P.

Theorem 3.1 *If the representation function $\beta^{\mathrm{CF}}_{\ell^a_*}(P)$ initially fulfills conditions $(i) - (ii)$, then the least solution $(\hat{I}, \hat{H}) \models^{\mathrm{CF}} P$ to the control flow analysis in [9] enjoys these conditions as well.*

Condition (ii) basically states two important properties on P execution: every time a boundary ambient is opened, this is done inside another boundary ambient; the only ambients that may exit from a boundary ambients are boundary ambients. By induction on reduction rules of Mobile Ambients it is easy to prove the following information flow theorem:

Theorem 3.2 *If $\beta^{\mathrm{CF}}_{\ell^a_*}(P)$ fulfills conditions $(i) - (ii)$, then, in every Q s.t. $P \to Q$, every high level ambient is always inside at least one boundary ambient.*

Note that the conditions are really simple to check. As an example consider again the two example presented above. In particular,

$$P = container^b[\, hdata^h[\, out^c container.\mathbf{0}\,]\,] \mid Q$$

does not satisfy condition (ii) as $out^c container$, by condition (i), should be labelled as a boundary move. However this makes a boundary move executable in a high level ambient, invalidating condition (ii). On the other side, the second example

$$venice^b[\, send^b[\, out^c venice.in^c twente \mid hdata^h[\,]\,]\,] \mid twente^b[\, open^c send\,] \mid Q$$

fulfills both the conditions, with $c \in \mathbf{Lab}^t_{BM}$. This proves that *hdata*, in every execution, is always inside a boundary ambient.

The syntactic conditions successfully applies also to the extended example with *hdata* entering the filter:

$$venice^b[\, send^b[\, out^c venice.in^c twente \mid hdata^h[\, in^{ch} filter\,]\,]\,] \mid$$
$$\mid twente^b[\, open^c send\,] \mid filter^m[\, in^c send\,] \mid open^{cl} filter$$

Also in this case, we are able to prove that *hdata*, in every execution, is always inside a boundary ambient. Note that this was not provable through the presented control flow analysis.

The approach above may also be adapted to the case in which the external environment (e.g. any malicious process put in parallel with the analyzed process P) does not fulfill the required conditions. This is indeed reasonable in a distributed open system. The idea is to suitably restrict the scope of boundary ambients and provide low level ambients with some "taxi" processes that, once entered, bring the client inside restricted boundaries. Let $b_1, \ldots, b_n$ represent all the boundary ambients of process P. Then consider process

$$(\nu b_1, \ldots, b_n)(P \mid !t_1[in^l b_1] \mid \ \ldots \ \mid !t_n[in^l b_n]) \mid Q$$

As $b_1, \ldots, b_n$ are restricted names, they may not appear in Q. As a consequence, if P fulfills the conditions $(i) - (ii)$, this is sufficient to prove that the whole system (whatever Q is considered) satisfies such conditions, too. It is indeed simple to prove the following:

Proposition 3.3 *If $\beta^{\mathrm{CF}}_{\ell^a_*}(P)$ fulfills conditions $(i) - (ii)$, then, for all Q (labelled in $\mathbf{Lab}^a_L \cup \mathbf{Lab}^t \setminus \mathbf{Lab}^t_{BM}$),*

$$\beta^{\mathrm{CF}}_{\ell^a_*}((\nu b_1, \ldots, b_n)(P \mid !t_1[in^l b_1] \mid \ \ldots \ \mid !t_n[in^l b_n]) \mid Q)$$

fulfills conditions $(i) - (ii)$.

Note that processes $!t_i[in^l b_i]$ allow any low level ambient to enter boundary b_i. So, legitimate flows from level to high level are possible even if boundaries are restricted. Note also that the condition on the labelling of Q simply means that Q just contains low level ambients and its capabilities are not (incorrectly) labelled as boundary moves.

4 Refining the Control Flow Analysis

In this section we introduce a refinement of the Control Flow Analysis of [9] mentioned above, in order to incorporate into the analysis the ideas discussed in Section 3, thus yielding to a more accurate tool for detecting unwanted boundary crossings. The resulting analysis improves also w.r.t. the syntactic properties in [1].

The main idea is to split the $\hat{I}$ component of the abstract domain in two (not necessarily distinct) sets, in order to keep information about the nesting of boundaries, and about "unprotected" ambients.

The refined control flow analysis works on triplet $(\hat{I_B}, \hat{I_E}, \hat{H})$, where:

- The first component $\hat{I_B}$ is an element of $\wp(\mathbf{Lab}^a \times (\mathbf{Lab}^a \cup \mathbf{Lab}^t))$. If a process contains either a capability or an ambient labelled ℓ inside an ambient labelled ℓ^a which is a boundary or an ambient nested inside a boundary (referred as *protected ambient*) then (ℓ^a, ℓ) is expected to belong to $\hat{I_B}$. As long as a high level datum is contained inside a protected ambient there is no unwanted information flow.
- The second component $\hat{I_E}$ is still an element of $\wp(\mathbf{Lab}^a \times (\mathbf{Lab}^a \cup \mathbf{Lab}^t))$. If a process

$$
\begin{array}{lll}
 & \beta^{\mathrm{CF}}(P_*) & = \beta^{\mathrm{CF}}_{\ell,False}(P_*) \\
(res) & \beta^{\mathrm{CF}}_{\ell,Protected}((\nu n)P) & = \beta^{\mathrm{CF}}_{\ell,Protected}(P) \\
(zero) & \beta^{\mathrm{CF}}_{\ell,Protected}(\mathbf{0}) & = (\emptyset,\emptyset,\emptyset) \\
(par) & \beta^{\mathrm{CF}}_{\ell,Protected}(P \mid Q) & = \beta^{\mathrm{CF}}_{\ell,Protected}(P) \sqcup \beta^{\mathrm{CF}}_{\ell,Protected}(Q) \\
(repl) & \beta^{\mathrm{CF}}_{\ell,Protected}(!P) & = \beta^{\mathrm{CF}}_{\ell,Protected}(P) \\
(amb) & \beta^{\mathrm{CF}}_{\ell,Protected}(n^{\ell^a}[\,P\,]) & = \text{case } \mathit{Protected} \text{ of} \\
 & & \quad \text{True} : \beta^{\mathrm{CF}}_{\ell^a,Protected}(P) \sqcup (\{(\ell,\ell^a)\},\emptyset,\{(\ell^a,n)\}) \\
 & & \quad \text{False: if } (\ell_a \in \mathbf{Lab}^a_B) \text{ then} \\
 & & \qquad \text{let Protected}' = \text{True in} \\
 & & \qquad \beta^{\mathrm{CF}}_{\ell^a,Protected'}(P) \sqcup (\emptyset,\{(\ell,\ell^a)\},\{(\ell^a,n)\}) \\
(in) & \beta^{\mathrm{CF}}_{\ell,Protected}(in^{\ell^t} n.P) & = \text{case } \mathit{Protected} \text{ of} \\
 & & \quad \text{True} : \beta^{\mathrm{CF}}_{\ell,Protected}(P) \sqcup (\{(\ell,\ell^t)\},\emptyset,\emptyset) \\
 & & \quad \text{False: } \beta^{\mathrm{CF}}_{\ell,Protected}(P) \sqcup (\emptyset,\{(\ell,\ell^t)\},\emptyset) \\
(out) & \beta^{\mathrm{CF}}_{\ell,Protected}(out^{\ell^t} n.P) & = \text{case } \mathit{Protected} \text{ of} \\
 & & \quad \text{True} : \beta^{\mathrm{CF}}_{\ell,Protected}(P) \sqcup (\{(\ell,\ell^t)\},\emptyset,\emptyset) \\
 & & \quad \text{False: } \beta^{\mathrm{CF}}_{\ell,Protected}(P) \sqcup (\emptyset,\{(\ell,\ell^t)\},\emptyset) \\
(open) & \beta^{\mathrm{CF}}_{\ell,Protected}(open^{\ell^t} n.P) & = \text{case } \mathit{Protected} \text{ of} \\
 & & \quad \text{True} : \beta^{\mathrm{CF}}_{\ell,Protected}(P) \sqcup (\{(\ell,\ell^t)\},\emptyset,\emptyset) \\
 & & \quad \text{False: } \beta^{\mathrm{CF}}_{\ell,Protected}(P) \sqcup (\emptyset,\{(\ell,\ell^t)\},\emptyset)
\end{array}
$$

Fig. 4. Representation Function for the Control Flow Analysis

contains either a capability or an ambient labelled ℓ inside an ambient labelled ℓ^a which is not protected, then (ℓ^a, ℓ) is expected to belong to $\hat{I_E}$.

- The third component $\hat{H}$ keeps track of the correspondence between names and labels. If a process contains an ambient labelled ℓ^a with name n, then (ℓ^a, n) is expected to belong to $\hat{H}$.

Also in this case, the analysis is defined by a representation function and a specification. They are depicted, respectively, in Figure 4 and Figure 5.
A pictorial representation of the most interesting application of the *in*-rule (i.e. a boundary crossing) is provided by Figure 6: the state of $\hat{I_E}$ and $\hat{I_B}$ before and after the move of ambient k into ambient n is represented by graphs (a) and (b), respectively.

The result of the analysis should be read, as expected, in terms of information flows.

Theorem 4.1 *No leakage of secret data/ambients outside the boundary ambients is possible if in the analysis h (high level datum) does not appear in any of the pairs belonging to* $\hat{I_E}$.

(res) $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} (\nu n)P$ iff $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} P$

$(zero)$ $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} \mathbf{0}$ always

(par) $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} P \mid Q$ iff $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} P \ \wedge \ (\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} Q$

$(repl)$ $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} !P$ iff $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} P$

(amb) $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} n^{\ell^a}[\,P\,]$ iff $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} P$

(in) $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} in^{\ell^t} n.P$ iff $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} P \wedge$

$\forall \ell^a, \ell^{a'}, \ell^{a''} \in \mathbf{Lab}^a : ((\ell^a, \ell^t) \in \hat{I_E} \cup \hat{I_B} \ \wedge \ (\ell^{a''}, \ell^a) \in \hat{I_E} \cup \hat{I_B}$
$\wedge \ (\ell^{a''}, \ell^{a'}) \in \hat{I_E} \cup \hat{I_B} \ \wedge \ (\ell^{a'}, n) \in \hat{H}) \implies$
if $(\mathrm{Protected}(\ell^{a''}) \vee (\neg\mathrm{Protected}(\ell^{a''}) \ \wedge \ \ell^{a'} \in \mathbf{Lab}^a_B \ \wedge \ \ell^a \in \mathbf{Lab}^a_B))$
then$(\ell^{a'}, \ell^a) \in \hat{I_B}$
and
if $(\neg\mathrm{Protected}(\ell^{a''}) \ \wedge \ \ell^{a'} \notin \mathbf{Lab}^a_B)$ then$(\ell^{a'}, \ell^a) \in \hat{I_E}$
and
if $(\neg\mathrm{Protected}(\ell^{a''}) \ \wedge \ \ell^{a'} \in \mathbf{Lab}^a_B \ \wedge \ \ell^a \notin \mathbf{Lab}^a_B)$
then$(\ell^{a'}, \ell^a) \in \hat{I_B} \ \wedge \ \left\{(\ell, \ell') \in \hat{I_E} \mid path_E(\ell^a, \ell)\right\} \subseteq \hat{I_B}$

(out) $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} out^{\ell^t} n.P$ iff $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} P \wedge$

$\forall \ell^a, \ell^{a'}, \ell^{a''} \in \mathbf{Lab}^a : ((\ell^a, \ell^t) \in \hat{I_E} \cup \hat{I_B} \ \wedge \ (\ell^{a'}, \ell^a) \in \hat{I_E} \cup \hat{I_B}$
$\wedge \ (\ell^{a''}, \ell^{a'}) \in \hat{I_E} \cup \hat{I_B} \ \wedge \ (\ell^{a'}, n) \in \hat{H}) \implies$
if $(\mathrm{Protected}(\ell^{a''}))$ then$(\ell^{a''}, \ell^a) \in \hat{I_B}$
and
if $(\neg\mathrm{Protected}(\ell^{a''}) \ \wedge \ (\ell^a \in \mathbf{Lab}^a_B \vee (\ell^a \notin \mathbf{Lab}^a_B \ \wedge \ \ell^{a'} \notin \mathbf{Lab}^a_B)))$
then$(\ell^{a''}, \ell^a) \in \hat{I_E}$
and
if $(\neg\mathrm{Protected}(\ell^{a''}) \ \wedge \ \ell^a \notin \mathbf{Lab}^a_B \ \wedge \ \ell^{a'} \in \mathbf{Lab}^a_B)$
then$(\ell^{a''}, \ell^a) \in \hat{I_E} \ \wedge \ \left\{(\ell, \ell') \in \hat{I_B} \mid path_B(\ell^a, \ell)\right\} \subseteq \hat{I_E}$

$(open)$ $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} open^{\ell^t} n.P$ iff $(\hat{I_B}, \hat{I_E}, \hat{H}) \models^{\mathrm{CF}} P \wedge$

$\forall \ell^a, \ell^{a'} \in \mathbf{Lab}^a : ((\ell^a, \ell^t) \in \hat{I_E} \cup \hat{I_B} \ \wedge \ (\ell^a, \ell^{a'}) \in \hat{I_E} \cup \hat{I_B}$
$\wedge \ (\ell^{a'}, n) \in \hat{H}) \implies$
if $(\mathrm{Protected}(\ell^a))$ then $\left\{(\ell^a, \ell) \mid (\ell^{a'}, \ell) \in \hat{I_B}\right\} \subseteq \hat{I_B}$
and
if $(\neg\mathrm{Protected}(\ell^a) \ \wedge \ \ell^{a'} \in \mathbf{Lab}^a_B)$
then $\left\{(\ell, \ell') \ \wedge \ (\ell^a, \ell'') \mid (\ell, \ell') \in \hat{I_B} \ \wedge \ (\ell^{a'}, \ell'') \in \hat{I_B} \ \wedge \ path_B(\ell'', \ell)\right\} \subseteq \hat{I_E}$
and
if $(\neg\mathrm{Protected}(\ell^a) \ \wedge \ \ell^{a'} \notin \mathbf{Lab}^a_B)$ then $\left\{(\ell^a, \ell) \mid (\ell^{a'}, \ell) \in \hat{I_E}\right\} \subseteq \hat{I_E}$

Fig. 5. Specification of the Control Flow Analysis

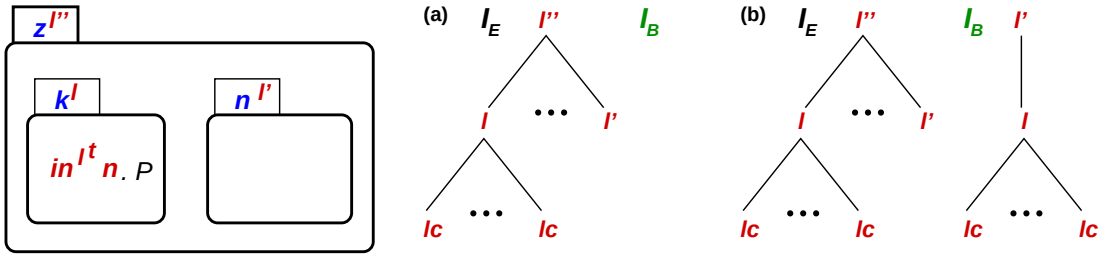

Fig. 6. The *in*-rule: boundary crossing

Observe that the abstraction and concretization functions $(\alpha^{CF}, \gamma^{CF})$ can be defined in terms of the occurrence counting domain in [9] (where a third component $\hat{A}$ maintains multiplicity information), as follows:

Let $\eta^B : \hat{I} \mapsto \hat{I_B}$ and $\eta^E : \hat{I} \mapsto \hat{I_E}$ be functions splitting $\hat{I}$ according to the boundary nestings, in [4]

$$\alpha^{CF}(\mathcal{C}) = \bigsqcup\{(\eta^B(\hat{I}), \eta^E(\hat{I}), \hat{H}) \mid (\hat{I}, \hat{H}, \hat{A}) \in \mathcal{C}\}$$

$$\gamma^{CF}(\hat{I_B}, \hat{I_E}, \hat{H}) = \{(\hat{I}', \hat{H}', \hat{A}') \mid (\eta^B(\hat{I}'), \eta^E(\hat{I}'), \hat{H}') \sqsubseteq (\hat{I_B}, \hat{I_E}, \hat{H}) \ \wedge \ (\hat{I}', \hat{H}', \hat{A}') \text{ is compatible } \}$$

The two functions η^B and η^E are defined in terms of protected and unprotected path predicates as follows:

- $p\_path(\ell^a) = \begin{cases} True & \text{iff } \exists \ell_0 \in \mathbf{Lab}^a_B \ \wedge \ \exists \ell_1, \ell_2, ..., \ell_n : n \geq 0 \\ & (\ell_0, \ell_1), (\ell_1, \ell_2), ..., (\ell_n, \ell^a) \in \hat{I}, \\ False & \text{otherwise.} \end{cases}$
- $u\_path(\ell^a) = \begin{cases} True & \text{iff } \exists \ell_1, \ell_2, ..., \ell_n : n \geq 0 \\ & (env, \ell_1), (\ell_1, \ell_2), ..., (\ell_n, \ell^a) \text{ s.t.} \forall k \ \ell_k \notin \mathbf{Lab}^a_B, \\ False & \text{otherwise.} \end{cases}$

$$\eta^B : \wp(\mathbf{Lab}^a \times (\mathbf{Lab}^a \cup \mathbf{Lab}^t)) \rightarrow \wp(\mathbf{Lab}^a \times (\mathbf{Lab}^a \cup \mathbf{Lab}^t))$$
$$\eta^B(\hat{I}) = \{(\ell^a, \ell) |\ p\_path(\ell^a) \vee \neg u\_path(\ell^a)\}$$

$$\eta^E : \wp(\mathbf{Lab}^a \times (\mathbf{Lab}^a \cup \mathbf{Lab}^t)) \rightarrow \wp(\mathbf{Lab}^a \times (\mathbf{Lab}^a \cup \mathbf{Lab}^t))$$
$$\eta^E(\hat{I}) = \{(\ell^a, \ell) |\ u\_path(\ell^a)\}$$

The abstraction and concretization functions $(\alpha^{CF}, \gamma^{CF})$ form a Galois connection; i.e.

- both functions are monotone;

[4] According to [9], a triplet $(\hat{I_B}, \hat{I_E}, \hat{H})$ is *compatible* whenever the labels in $\hat{I_E} \cup \hat{I_B}$ are consistent with the mapping $\hat{H}$. More formally, if the following condition is satisfied: if $(\ell^a, \ell) \in \hat{I_E} \cup \hat{I_B}$ or $(\ell, \ell^a) \in \hat{I_E} \cup \hat{I_B}$ then there exists n such that $(\ell^a, n) \in (\hat{I}, \hat{H})$.

- $\mathcal{C} \subseteq \gamma^{\mathrm{CF}}(\alpha^{\mathrm{CF}}(\mathcal{C}))$ for any $\mathcal{C} \in \mathbf{CountSet}$;
- $\alpha^{\mathrm{CF}}(\gamma^{\mathrm{CF}}(\hat{I_B}, \hat{I_E}, \hat{H})) \sqsubseteq (\hat{I_B}, \hat{I_E}, \hat{H})$ for any triplet $(\hat{I_B}, \hat{I_E}, \hat{H})$.

This leads to a hierarchy of abstractions that very well fits in the Abstract Interpretation theory, yielding an expected trade-off between accuracy and efficiency of the analyses.

Observe that within the specification of the analysis (depicted in Figure 5), some predicates are introduced that simplify the notation, namely

- $path_B(\ell^a, \ell) = \begin{cases} True & \text{if } \ell^a = \ell \vee \exists \ell_1, \ell_2, ..., \ell_n \notin \mathbf{Lab}^a_B : n \geq 0 \\ & (\ell^a, \ell_1), (\ell_1, \ell_2), ..., (\ell_n, \ell) \in \hat{I_B} \ \wedge \ \ell^a, \ell \notin \mathbf{Lab}^a_B, \\ False & \text{otherwise.} \end{cases}$
- $path_E(\ell^a, \ell) = \begin{cases} True & \text{if } \ell^a = \ell \vee \exists \ell_1, \ell_2, ..., \ell_n \notin \mathbf{Lab}^a_B : n \geq 0 \\ & (\ell^a, \ell_1), (\ell_1, \ell_2), ..., (\ell_n, \ell) \in \hat{I_E} \ \wedge \ \ell^a, \ell \notin \mathbf{Lab}^a_B, \\ False & \text{otherwise.} \end{cases}$
- $Protected(\ell^a) = \begin{cases} True & \text{if } \nexists \ell^{a'} : (\ell^{a'}, \ell^a) \in \hat{I_E} \ \wedge \ \ell^a \neq env, \\ True & \text{if } \ell^a \in \mathbf{Lab}^a_B, \\ False & \text{otherwise.} \end{cases}$

The correctness proof of the Control Flow Analysis can be obtained by structural induction along the lines of the proof in [9].

What about accuracy? The analysis just introduced is a refinement of the CFA in [9] and it properly deals with boundary nestings, in the spirit of Section 3. In particular, it strongly improves in accuracy with respect to the mentioned syntactic property introduced in [1].

Consider, for instance, the following example, where the process discussed in the previous sections is extended by allowing an *application* (say an applet) to be downloaded from the *web* within *twente*; then, the application may open the ambient *send* and disappear.

$$venice^b[\, send^b[\, out^c venice.in^c twente \mid hdata^h[\, in^{ch} filter\,]\,]\,] \mid$$
$$\mid twente^b[\, download^{m'}[\, out^{cm'} twente.in^{cm'} web.in^{cm'} twente\,] \mid$$
$$\mid open^c web.open^c application\,] \mid$$
$$web^m[\, application^m[\, open^{cm} send.filter^m[\ ]\,] \mid open^{cm} download\,]$$

Observe that in this case there is no information flow, as the application is not exporting any data out of the twente boundary. In this case, our refined CFA yields to positive information (see the least solution reported below), whereas the syntactic property cannot be successfully applied. In fact, the (untrusted) application downloaded from the net is not a boundary, its *open* capability is labeled BM by the first rule, and thus the second rule cannot be satisfied.

$$\hat{I_B} = \{(b,b),(b,c),(b,h),(b,m),(b,m^{'}),(h,ch),(m^{'},cm^{'}),(m,h)\}$$
$$\hat{I_E} = \{(l^a_*,b),(l^a_*,m),(l^a_*,m^{'}),(m,m),(m,m^{'}),(m,cm),(m^{'},cm^{'})\}$$
$$\hat{H} = \{(b,venice),(b,send),(b,twente),(h,hdata),(m^{'},download),$$
$$(m,web),(m,application),(m,filter)\}$$

Observe that the result is also better than the Hansen-Jensen-Nielsons's CFA [9] in which the following least solution is obtained:

$$\hat{I} = \{(l^a_*,b),(l^a_*,m),(l^a_*,m^{'}),(b,b),(b,c),(b,h),(b,m),(b,cm),(b,m^{'}),$$
$$(b,cm^{'}),(h,ch),(m,m),(m,m^{'}),(m,b),(m,cm),(m,cm^{'}),$$
$$(m,c),(m,h),(m^{'},cm^{'})\}$$
$$\hat{H} = \{(b,venice),(b,send),(b,twente),(h,hdata),(m^{'},download),$$
$$(m,web),(m,application),(m,filter)\}$$

Note that h appears inside a m ambient that at the beginning of the process is at the environmental level, but the analysis does not capture the fact that h enters m only after it has crossed the boundary and can never return back.

5 Conclusions

The main novelty of the approach presented in this paper is that we model multilevel information flow within a "pure" mobile ambient setting, without considering either channels or recently proposed restrictions of Mobile Ambients designed for security issues (like Secure Safe Ambients [3]).

As a future work, we intend to extend the approach to other versions of Mobile Ambients, and, in particular, to the full calculus with communication channels. It is our opinion that if only communication within ambients is considered, the approach should carry on very naturally. We also intend to compare our approach with other control flow analyses proposed for particular versions of Mobile Ambients, like, e.g., the one for Safe Ambients [5], aimed at capturing access control more than information flow. It would be also interesting to study whether our approach could be applied, via some suitable encoding, also to "classical" calculi, like pi-calculus.

References

[1] A. Cortesi, and R. Focardi. Information Flow Security in Mobile Ambients. In Proc. of International Workshop on Concurrency and Coordination CONCOORD'01, Lipari Island, July 2001, volume 54 of *Electronic Notes in Theoretical Computer Science*, Elsevier, 2001.

[2] C. Bodei, P. Degano, F. Nielson, and H.R.Nielson. Static Analysis of Processes for No Read-Up and No-Write-Down. In Proc. FoSSaCS'99, volume 1578 of *Lecture Notes in Computer Science*, pages 120–134, Springer-Verlag, 1999.

[3] M. Bugliesi and G. Castagna. "Secure Safe Ambients". Proc. 28th ACM Symposium on Principles of Programming Languages (POPL'01), pp. 222-235, London. 2001.

[4] L. Cardelli and A. Gordon. "Mobile Ambients". In Proc. FoSSaCS'98, volume 1378 of *Lecture Notes in Computer Science*, pages 140–155, Springer-Verlag, 1998.

[5] P. Degano, F. Levi, C. Bodei. Safe Ambients: Control Flow Analysis and Security. In proceedings of *ASIAN'00*, LNCS 1961, 2000, pages 199-214.

[6] R. Focardi and R. Gorrieri. "A Classification of Security Properties for Process Algebras", Journal of Computer Security, 3(1): 5-33, 1995.

[7] R. Focardi and R. Gorrieri, "The Compositional Security Checker: A Tool for the Verification of Information Flow Security Properties, IEEE Transactions on Software Engineering, Vol. 23, No. 9, September 1997.

[8] R. Focardi, R. Gorrieri, F. Martinelli, "Information Flow Analysis in a Discrete Time Process Algebra", in Proc. of 13th IEEE Computer Security Foundations Workshop (CSFW13), (P.Syverson ed), IEEE CS Press, 170-184, 2000.

[9] R. R. Hansen, J. G. Jensen, F. Nielson, and H. R. Nielson, Abstract Interpretation of Mobile Ambients. In Proc. Static Analysis Symposium SAS'99, volume 1694 of *Lecture Notes in Computer Science*, pages 134–148, Springer-Verlag, 1999.

[10] M. Hennessy, J. Riely. "Information Flow vs. Resource Access in the Asynchronous Pi-Calculus". ICALP 2000: 415-427.

[11] G. Smith, D.M. Volpano, "Secure Information Flow in a Multi-Threaded Imperative Language". In Proc. of POPL 1998: 355-364.