# Required Information Release

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1109/CSF.2010.22</td>
</tr>
<tr>
<td>Accessed</td>
<td>December 6, 2017 5:29:16 PM EST</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:34401541">http://nrs.harvard.edu/urn-3:HUL.InstRepos:34401541</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>

(Article begins on next page)
Required Information Release

Stephen Chong

TR-04-10

Computer Science Group
Harvard University
Cambridge, Massachusetts
Required Information Release

Stephen Chong
School of Engineering and Applied Sciences
Harvard University
chong@seas.harvard.edu

Abstract
Many computer systems have a functional requirement to release information. Such requirements are an important part of a system’s information security requirements. Current information-flow control techniques are able to reason about permitted information flows, but not required information flows.

In this paper, we introduce and explore the specification and enforcement of required information release in a language-based setting. We define semantic security conditions that express both what information a program is required to release, and how an observer is able to learn this information. We also consider the relationship between permitted and required information release, and define bounded release, which provides upper- and lower-bounds on the information a program releases. We show that both required information release and bounded release can be enforced using a security-type system.

1. Introduction
Information-flow control holds the promise of strong, end-to-end, application-specific information security [Sabelfeld and Myers, 2003]. To date, most research on information-flow control has focused on what flows are permitted or prohibited in a system. For example, noninterference [Goguen and Meseguer, 1982] prohibits confidential inputs flowing to public outputs.

Many computer systems release (or declassify) confidential information as part of their intended functionality, and as such, violate noninterference. Much work in recent years has considered weakening noninterference to permit some flow of confidential inputs to public outputs (e.g., Li and Zdancewic [2005]; Chong and Myers [2004]; Clark et al. [2005]; Sabelfeld and Sands [2005]; van der Meyden [2007]; Swamy and Hicks [2008]; Banerjee et al. [2008]).

However, many systems have more than just permission to release information; they have an obligation to release information. In this work, we introduce and explore the specification and enforcement of required information release, or simply, required release.

Examples abound of systems with an obligation to release information. The Sarbanes-Oxley Act of 2002 is a United States federal law that was enacted after a series of corporate accounting scandals, and requires publicly held companies to report details of their finances to a government agency. Thus, financial systems of such companies are required to release sensitive financial information to the government agency. Pharmaceutical companies in many countries are required to report all results of clinical trials of new drugs to a government agency (such as the Food and Drug Administration) to receive approval. Computer systems that support the conduct of clinical trials must release all trial results, and not withhold negative results. In general, transparency of organizations and processes requires the release of sensitive information. Other systems that are required to release information include the following.

- Sealed bid auctions: at the end of the auction, the winning bid (and, depending on the auction, the winner’s identity) is required to be released.
- Information purchase: once a customer has paid for information (such as electronic media), the information is required to be available for download.
- Games: legal game-play often requires release of a player’s secret information, such as the cards in a player’s hand, or the location of battleships on a player’s board.
- Course management system: when a professor indicates that exam results are available, the system is required to allow students view their grades.
- Credit card sales: the receipt for a credit card purchase is required to show the final four digits of the customer’s credit card number.

In the examples above, the required release of information is an important aspect of each system’s information security. To gain assurance in the systems’ correct implementation, it is necessary both to specify the required release (and other information security requirements) and to verify that the implementation satisfies the specification.

However, the specification of required release is subtle. What does it mean for a program to satisfy the required release of information? How do we know if a program is successfully and correctly releasing information? It does not suffice for the output of a program to simply depend upon, or be influenced by, the information required for release. Surprisingly, even if the output contains the information required for release, the program may not satisfy the required release of information. We use epistemic logic, and algorithmic knowledge [Halpern et al., 1994] in particular, to guide our definition of required information release. Required information release must specify not just what information is to be released, but also how that information is to be learned by its intended recipient.

Required release is a functional requirement on a system: the output of a program must allow an observer to learn certain information. Required release makes explicit what information the observer learns, and how the observer obtains this knowledge. Noninterference [Goguen and Meseguer, 1982] and most other information flow security conditions are not functional requirements.

However, required release is an information flow security condition; it describes a mandatory flow of information to an observer. By contrast, most existing work in information flow considers permitted flows of information. In terms of dependence, permitted information flow conditions restrict how the output is permitted to

depend on the inputs. For example, noninterference requires that public outputs do not depend on private inputs—if a private input changes, the public output should not change. Required information release mandates that outputs must depend on inputs in a way that allows an observer to learn certain information—if an input changes, the output must change to allow an observer to learn the specified information.

Required release interacts with permitted information flows in more interesting ways than other functional requirements: if a system is required to release information, then the system must also be permitted to release it. Indeed, required information release and permitted information release can be combined to specify both upper and lower bounds on the information that a system releases. We do so by defining bounded release, a security condition that combines required release and delimited release [Sabelfeld and Myers, 2004], and thus specifies both what information a program is required to release, and what information it is allowed to release.

For some systems, bounds on information release should be tight. For example, a company producing reports in accordance with the Sarbanes-Oxley Act typically wishes to release no more information than is required by law; thus, the information that their financial system is permitted to release should be identical to the information it is required to release. In other systems, the bounds are not tight, such as in a poker game where some players are permitted, but not required, to reveal their cards at the end of a hand.

The remainder of the paper is structured as follows. Section 2 uses the example of a simple credit card sales system to examine what it means for a system to satisfy required information release. Section 3 presents an interactive imperative language that we use in Section 4 to formally define required release. We also define bounded release, a security condition that specifies what information a program is required and permitted to release. We show in Section 5 that required release and bounded release can be soundly enforced in an interactive language by a type system. Section 6 discusses related work, and Section 7 concludes.

2. What is required release?

Consider, as a running example, a (grossly simplified) credit card sales system that takes a credit card number as input from high confidentiality channel \( H \), and is required to release the last four digits to low confidentiality channel \( L \) (representing, for example, the printer, or an audit log). What does it mean for this system to satisfy the required release of the last four digits?

Noninterference [Goguen and Meseguer, 1982] is a strong information security condition that requires that public outputs reveal no information about confidential inputs. Any system that releases confidential information violates noninterference; the credit card sales system, which must release the last four digits of the confidential credit card number to a publicly observable printer, violates noninterference. However, just because a system violates noninterference does not mean it satisfies the required information release.

Consider the following attempt to implement the credit card sales system.

\[
P_1 : \quad \text{input } cc \text{ from } H; \\
\quad \text{if } (cc \mod 10,000) < 5,000 \text{ then} \\
\quad \quad \quad \text{output 0 to } L \\
\quad \text{else} \\
\quad \quad \quad \text{output 1 to } L \\
\]

The program inputs the credit card number from channel \( H \), and then outputs either 0 or 1 to channel \( L \). The output observed on channel \( L \) is influenced by the last four digits of the confidential input, and thus the program does not satisfy noninterference. However, the program fails to satisfy the required information release, as an observer of channel \( L \) does not learn the credit card number’s last four digits.

Surprisingly, even if a system outputs the information required for release, it may fail to satisfy the required information release. This is demonstrated in the following program, which is another attempt to implement the credit card sales system.

\[
P_2 : \quad \text{input } cc \text{ from } H; \\
\quad i := 0; \\
\quad \text{while } i < (cc \mod 10,000) \text{ do} \\
\quad \quad \quad \text{output } i \text{ to } L; \\
\quad \quad \quad i := i + 1; \\
\quad \quad \quad \text{output } (cc \mod 10,000) \text{ to } L; \\
\quad \quad \quad i := i + 1; \\
\quad \text{while } i < 10,000 \text{ do} \\
\quad \quad \quad \text{output } i \text{ to } L; \\
\quad \quad \quad i := i + 1 \\
\]

The command output \((cc \mod 10,000)\) to \(L\) in program \(P_2\) above explicitly outputs the credit card’s last four digits. However, every execution of the program outputs all integers from 0 to 9,999 in order. An observer of channel \( L \) always sees the same output, regardless of the credit card’s last four digits, and so the observer learns nothing.

These examples show that it is insufficient for observable output to be merely correlated with the information required for release, or even for the output to contain that information. The key insight is that to satisfy required release, the output must allow an observer to know what information was required for release.

In models of knowledge based on possible world semantics [Hintikka, 1962; Fagin et al., 1995], an agent has implicit knowledge (or, simply, knowledge) of fact \( \phi \) if in all possible worlds consistent with the agent’s observations, \( \phi \) is satisfied. In the credit card system, an observer of channel \( L \) knows the last four digits of the credit card if all credit cards that could have produced the observed output end in the same four digits. Programs \(P_1\) and \(P_2\) do not allow an observer of channel \( L \) to know the last four digits.

Standard logical approaches to knowledge suffer from the problem of logical omniscience: an agent knows all logical consequences of its knowledge. The following attempt to implement the credit card system demonstrates this problem. The program chooses two large primes, outputs their product, and the result of XOR-ing the smaller prime with the last four digits of the credit card number (padded with random bits to be the same length as the prime).

\[
P_3 : \quad \text{input } cc \text{ from } H; \\
\quad p := \text{generateLargePrime}(); \\
\quad q := \text{generateLargePrime}(); \\
\quad \text{output } p \times q \text{ to } L; \\
\quad \text{if } p < q \text{ then } t := p \text{ else } t := q; \\
\quad \text{output } t \text{ xor pad}(cc \mod 10,000) \text{ to } L \\
\]

A logically omniscient observer of the program’s output knows what the last four digits of the credit card number are. However, determining this requires factoring a large number, which is beyond the abilities of humans and current computer systems to perform in reasonable time.

Algorithmic knowledge [Halpern et al., 1994] was introduced to address the problem of logical omniscience, and we can use algorithmic knowledge to reason whether a system satisfies required release.

An agent has algorithmic knowledge, or explicit knowledge, of fact \( \phi \) if the agent has an algorithm that responds “Yes” when given input \( \phi \) and the agent’s observations. The agent’s knowledge algorithm is sound if whenever it responds “Yes” then the agent has
implicit knowledge of $\phi$, and if the knowledge algorithm responds “No” then the agent does not have implicit knowledge of $\phi$.

Rich classes of knowledge algorithms have been studied that can conservatively overestimate the computational ability of agents without giving the agents logical omniscience (e.g., Ramanujam [1999]; Pucella [2006]). However, we are interested in simple algorithms. Such algorithms may be described in user manuals, specified by a government agency or auditor, or may be inferred from self-explanatory output. In all cases, the aim is to make it easy for an observer to learn the released information. In this setting, the observer is not the adversary, and it is acceptable (even desirable) to underestimate the observer’s computational abilities, much as an instruction manual aims to be usable by as wide an audience as possible. For some programs (such as $P_3$), there may be sound knowledge algorithms that are beyond the ability of any observer to execute in reasonable time; such programs do not allow the observer to easily learn the released information, and are thus of no interest to us. For required release, we are concerned with the existence of sufficiently simple sound knowledge algorithms.

The following program does release the last four digits of the credit card number to channel $L$.

$$P_3 : \begin{align*}
  \text{input } cc & \text{ from } H; \\
  \text{output } "\text{Last 4 credit card digits:}" & \text{ to } L; \\
  \text{output } (cc \mod 10, 0,000) & \text{ to } L.
\end{align*}$$

Moreover, there is a simple sound knowledge algorithm to provide explicit knowledge for an observer of channel $L$: given fact $\phi \equiv (cc \mod 10, 000) = n$, respond “Yes” if and only if the second output is $n$. Because there is a simple sound algorithm, an observer can gain explicit (and implicit) knowledge of the last four credit card digits, and so the program satisfies the required information release.

To specify required release, we must specify not only what information is to be released, but also how that information is to be learned. We formalize this intuition by defining required information release in terms of a simple interactive programming language.

3. Language

In this section we present a simple interactive imperative programming language due to O’Neill et al. [2006]. We use an interactive language as it is more general than the batch model traditionally used to reason about language-based information flow, and it can more accurately model real word programs that interact with their external environment, such as server processes, and programs with user interfaces.

We assume set $L$ of security levels, ordered by a reflexive transitive relation $\sqsubseteq$ that indicates the relative restrictiveness of the levels. In this paper, our examples use the two element set $\{L, H\}$ where $L \sqsubseteq H$. Security level $L$ represents low confidentiality, and security level $H$ represents high confidentiality. More expressive security levels are possible (e.g., Myers and Liskov [1998]; Chen and Chong [2004]). Metavariable $\ell$ ranges over security levels.

3.1 Users, channels, and strategies

Users interact with executing programs. We assume that security levels characterize users: the security level of a user indicates the most restrictive level of information the user is permitted to read. We assume that users with the same security level can freely collaborate, and so, without loss of generality, assume only a single user at each level.

Users communicate with executing programs via channels. We assume input on channels is blocking, and output is non-blocking. We assume that there is a single channel for each user, which, given the assumption of a single user for each security level, implies a single channel for each security level. We thus identify channels with security levels. An event is the transmission of an input or output on a channel. Event $in(\ell, v)$ denotes the input of value $v$ on channel $\ell$, and event $out(\ell, v)$ denotes the output of value $v$ on channel $\ell$. For simplicity we restrict values to integers.

We use $Ev_{in}$ and $Ev_{out}$ to denote, respectively, the set of all input and output events. We use $Ev(\ell)$ to denote the set of all events that could occur on channel $\ell$, and $Ev$ to denote the set of all events.

$$\begin{align*}
  Ev_{in} & \triangleq \bigcup_{\ell \in L, v \in Z} \{in(\ell, v)\} \\
  Ev_{out} & \triangleq \bigcup_{\ell \in L, v \in Z} \{out(\ell, v)\} \\
  Ev(\ell) & \triangleq \bigcup_{v \in Z} \{in(\ell, v), out(\ell, v)\} \\
  Ev & \triangleq \bigcup_{\ell \in L} Ev(\ell)
\end{align*}$$

Given $E \subseteq Ev$, an event trace on $E$ is a finite or infinite sequence $t = (\alpha_0, \alpha_1, \ldots)\ |\ t$ such that $\alpha_i \in E$ for all $i$ such that $0 \leq i < |t|$, where $|t|$ is the length of $t$. For infinite traces $t$, we define $|t| = \infty$. The $i$th element of event trace $t$ is denoted $t(i)$, for $i$ such that $0 \leq i < |t|$. The empty trace is denoted $\langle \rangle$. We write $t' \sqcup t$ for the concatenation of finite trace $t$ and trace $t'$. For traces $t$ and $t'$, we say that $t$ extends $t'$, written $t \sqsupseteq t'$, when $t'$ is a prefix of $t$. Note that if $t$ is an infinite trace, then $t$ is the only trace that extends it. The set of all traces on $E$ is denoted $Tr(E)$.

The restriction of trace $t$ to $E$, written $t\ |\ E$, is the trace obtained by removing from $t$ all events not contained in $E$. We write $t \parallel t'$ as shorthand for $t \sqcup (E \setminus Ev(t))$.

User strategies express the behavior of users by describing how users interact with a program. Given trace $t$, a user of a channel with security level $\ell$ observes the event trace $t \parallel Ev(\ell)$; a user’s observations may influence their subsequent interaction with the program. Formally, a user strategy for a channel with security level $\ell$ is a function of type $Tr(E) \rightarrow Z$, and expresses what input a user will provide given their previous observations.

Let $UserStrat$ be the set of all user strategies. A joint strategy is a collection of user strategies, one for each channel. Formally, a joint strategy $\omega$ is a function of type $L \rightarrow UserStrat$, that is, a function from security levels to user strategies. Let $Strat$ be the set of all joint strategies.

User strategies are sensitive information. In general, we want to ensure that lower security users do not learn about strategies employed by higher security users: user $\ell$ should not learn anything about the strategy of user $\ell'$, where $\ell' \sqsupseteq \ell$. However, information release will violate this, revealing some information about the strategies of higher security users. In Section 4 we will discuss security requirements, and formally define semantic security conditions.

3.2 Syntax and semantics

We use a simple imperative language, extended with input, output, and declassification commands. The syntax of this language is:

$$\begin{align*}
  (\text{expressions}) & \quad e ::= n \ | \ x \ | \ e_0 \oplus e_1 \\
  (\text{commands}) & \quad c ::= \text{skip} \ | \ x ::= e \ | \ c_0; c_1 \ | \ \text{if } e \text{ then } c_0 \text{ else } c_1 \ | \\
  & \quad \text{while } e \text{ do } c \ | \\
  & \quad \text{input } x \text{ from } \ell \ | \\
  & \quad \text{output } c \text{ to } \ell \ | \\
  & \quad x ::= \text{declassify}(e \to \ell)
\end{align*}$$

Metavariable $x$ ranges over $Var$, the set of all program variables. Variables take integer values, and literal values $n$ also range over integers. Metavariable $\oplus$ ranges over total binary operations on the integers. A state $\sigma$ maps variables to values, and so is a function of type $Var \rightarrow Z$. A configuration is a 4-tuple $(e, \sigma, t, \omega)$.
representing a system about to execute \( c \) with state \( \sigma \) and joint strategy \( \omega \). Finite trace \( t \) is the history of events produced by the system so far. Terminal configurations have the form \((\text{skip}, \sigma, t, \omega)\). Metavariants \( m \) ranges over configurations.

We define a small-step operational semantics for our language, using the relation \( \longrightarrow \) over configurations. If \((c, \sigma, t, \omega) \longrightarrow (c', \sigma', t', \omega)\) then execution of command \( c \) can take a single step to command \( c' \), while updating the state from \( \sigma \) to \( \sigma' \). Trace \( t' \) extends \( t \) with any events that were produced during the step. Joint strategy \( \omega \) is unchanged when a configuration takes a step, and is included in configurations to simplify notation and presentation.

Figure 1 presents inference rules for the operational semantics. We use \( \sigma(e) \) to denote the evaluation of expression \( e \) where each variable \( x \) is replaced with the integer \( \sigma(x) \). Input command \( \text{input} x \) from \( \ell \) takes the next input value \( v \) as defined by the user strategy for \( \ell \), assigns it to variable \( x \), and updates the trace with input event \( \text{in}(\ell, v) \). Similarly, output command \( \text{output} e \) to \( \ell \) evaluates \( e \) to \( v \), and updates the trace with output event \( \text{out}(\ell, v) \). Declassification \( x := \text{declassify}(e \text{ to } \ell), \sigma, t, \omega \) is semantically equivalent to assignment \( x := e \); the declassify annotation and security level \( \ell \) are used in the type system, described in Section 5.

We use \( \longrightarrow^* \) to denote the reflexive transitive closure of \( \longrightarrow \). For finite trace \( t \), we say configuration \( c \text{ emits } t \), written \( c \not\rightarrow t \), if there is some configuration \((c, \sigma, t, \omega)\) such that \( m \not\rightarrow^* (c, \sigma, t, \omega) \). For infinite trace \( t \), \( m \) emits \( t \) if \( m \) emits all finite prefixes of \( t \). Note that emitted events may include both input and output events.

4. Security definitions

In this section we define the security conditions \textit{weak required release} and \textit{strong required release}, which formally express what it means for a program to satisfy the required release of information. We also present the security conditions \textit{noninterference} [Goguen and Meseguer, 1982] and \textit{delimited release} [Sabelfeld and Myers, 2004]. Noninterference requires that a program does not release any confidential information. Delimited release weakens noninterference by specifying what confidential information a program is allowed to release. We combine delimited release and required release to define bounds on what a program is permitted and required to release.

4.1 Required release

To formally define required release, we must be able to express \textit{what} information is to be released, and \textit{how} that information is to be learned by an observer. We introduce \textit{input expressions} and \textit{output expressions} to express each of these respectively. Input expressions are expressions over input values supplied on channels; output expressions are expressions over values output on a single channel.

The syntax for input and output expressions is:

\[
\begin{align*}
\text{(input expressions)} & : f ::= n \mid f_0 \oplus f_1 \mid \text{in}_i[n] \\
\text{(output expressions)} & : g ::= n \mid g_0 \oplus g_1 \mid \text{out}_i[x]
\end{align*}
\]

Input expression \( \text{in}_i[n] \) refers to the \( i \)th input event on channel \( \ell \), for \( i \in \mathbb{N} \). Input expressions may also contain integer constants and binary operations. Input expressions are evaluated against a trace. The judgment \( t \models_{\text{in}} f \downarrow v \) means that with trace \( t \), input expression \( f \) evaluates to value \( v \). Evaluation rules for input expressions are given in Figure 2. If \( t \) does not have an \( i \)th input event on channel \( \ell \), then \( \text{in}_i[n] \) evaluates to \( \bot \), that is, \( t \not\models_{\text{in}} \text{in}_i[n] \downarrow \bot \). We assume that any binary operator \( \oplus \) is defined is total over \( \mathbb{Z}_+ \) and strict, meaning that for all \( n, m \in \mathbb{Z}_+ \), \( m \oplus n \) is defined, and if \( m \) or \( n \) is \( \bot \), then \( m \oplus n = \bot \).

Output expressions are also evaluated against a trace. The judgment \( t \models_{\text{out}} g \downarrow v \) means that output expression \( g \) evaluates to value \( v \) using trace \( t \) restricted to channel \( \ell \) events. Output expression \( \text{out}_i[x] \) refers to the \( i \)th output event on channel \( \ell \), for \( i \in \mathbb{N} \). Figure 2 also presents the evaluation rules for output expressions. Similar to input expressions, if there is no \( i \)th output event on channel \( \ell \), then \( \text{out}_i[x] \) evaluates to \( \bot \).

Intuitively, user \( \ell \) learns input expression \( f \) from command \( c \) using output expression \( g \), if in every execution that \( g \) evaluates to an integer value (using the output provided to \( t \)), then \( f \) evaluates to the same integer. Thus, input expression \( f \) indicates \textit{what} information the user is to learn, and output expression \( g \) indicates \textit{how}
| $t \vdash_{in} f_0 \downarrow v_0$ | $t \vdash_{in} f_1 \downarrow v_1$ | $v = v_0 \land v_1$ | $t' = t \mid (Ev(\ell) \cap Ev_{in})$ | $t'(i) = in(\ell, v)$ | $t \vdash_{in} \text{in}[i] \downarrow v$ | $t \vdash_{in} (Ev(\ell) \cap Ev_{in})$ | $\neg(0 \leq i < |t'|)$ |
| $t \vdash_{out} g_0 \downarrow v_0$ | $t \vdash_{out} g_1 \downarrow v_1$ | $v = v_0 \land v_1$ | $t' = t \mid (Ev(\ell) \cap Ev_{out})$ | $t'(i) = out(\ell, v)$ | $t \vdash_{out} \text{out}[i] \downarrow v$ | $t \vdash_{out} (Ev(\ell) \cap Ev_{out})$ | $\neg(0 \leq i < |t'|)$ |

Figure 2. Evaluation rules for input and output expressions

the user learns it—g provides a sound knowledge algorithm. This leads us to our first definition of required release.

Command $c$ satisfies weak required release of input expression $f$ to user $\ell$ using output expression $g$ if for any trace that can be emitted by executing $c$, if $t$ provides enough output to $\ell$ to evaluate $g$, then $f$ and $g$ evaluate to the same value.

**Definition 1** (Weak required release). Command $c$ satisfies weak required release of input expression $f$ to user $\ell$ using output expression $g$ exactly when:

For all configurations $m = (c, \sigma, (\ell, \omega))$ (for all traces $t$ such that $m \sim t$, if $t \vdash_{out} g \downarrow v$ and $v \neq \bot$, then $t \vdash_{in} f \downarrow v$.

Program $P_3$ satisfies weak required release of $in_H[0] \mod 10,000$ to $L$ using output $[1]$, the second output to $L$ is the last four digits of the first $H$ input (the credit card number). By contrast, programs $P_1$ and $P_2$ do not satisfy weak required release of $in_H[0] \mod 10,000$ to $L$ for any output expression.

Weak required release is "weak" in that there is no requirement that command $c$ provide sufficient output to $f$ for $g$ to evaluate to an integer value. For example, the program skip satisfies weak required release of any input expression to $L$ using output expression $out[0]$, since no output is ever given to $L$, and output expression $out[0]$ never evaluates to an integer value.

We can strengthen weak required release to ensure that command $c$ always eventually provides sufficient output to $f$ for $g$ to evaluate to an integer value. Command $c$ satisfies strong required release of input expression $f$ to user $\ell$ using output expression $g$ if for any trace $t$ that can be emitted by executing $c$, there is a trace $t'$ that extends $t$, can be emitted by executing $c$, and provides sufficient output to $\ell$ to evaluate $g$, and $f$ and $g$ evaluate to the same value.

**Definition 2** (Strong required release). Command $c$ satisfies strong required release of input expression $f$ to user $\ell$ using output expression $g$ exactly when:

For all configurations $m = (c, \sigma, (\ell, \omega))$ and for all traces $t$ such that $m \sim t$, there exists trace $t'$ such that $t' \triangleright t$, $m \sim t'$, and $t' \vdash_{out} g \downarrow v$ and $t' \vdash_{in} f \downarrow v$ for some $v \neq \bot$.

Strong required release is strictly stronger than weak required release: if command $c$ satisfies strong required release of $f$ to $\ell$ using $g$, then it satisfies weak required release of $f$ to $\ell$ using $g$. Indeed, weak required release is a form of partial correctness, whereas strong required release is a total correctness condition. The following program satisfies weak required release, but not strong required release of $in_H[0] \mod 10,000$ to $L$ using output $[1]$, because in some cases it will never produce sufficient output to $L$. (For presentation purposes, we assume that constant strings, such as “Last 4 credit card digits: ” can be converted to appropriate constant integer values, and output to channels.)

$P_5$:
- input $cc$ from $H$;
- output “Last 4 credit card digits: ” to $L$;
- if $cc \mod 10 = 0$ then (while 1 do skip) else skip;
- output ($cc \mod 10,000$) to $L$

Program $P_5$ satisfies strong required release of $in_H[0] \mod 10,000$ to $L$ using output $[1]$, because it always produces appropriate output to channel $L$.

**Connection to explicit knowledge**

If a program satisfies (weak or strong) required release of input expression $f$ to user $\ell$ using output expression $g$, then output expression $g$ provides a sound knowledge algorithm for $\ell$ to learn the value of $f$. The knowledge algorithm takes as input a formula $\phi$ and the sequence of events that $\ell$ has observed. The knowledge algorithm is straightforward:

If $\phi \equiv f = n$ and $t \vdash_{out} g \downarrow n$ then respond ‘Yes’. Otherwise, respond ‘?”.

Note that the algorithm never responds “No”, and if the algorithm responds “Yes”, then, because the program satisfies required release of $f$ to $\ell$ using $g$, $t \vdash_{out} g \downarrow n$ implies $f = n$. Thus, the knowledge algorithm is sound.

Strong and weak required release are both parameterized by output expression $g$. As discussed in Section 2, the output expression $g$ may be specified by the consumer of the output (such as an auditor or government agency), an instruction manual, or may be described by the program’s output (as in Program $P_5$, where the text “Last 4 credit card digits” is output just before the last four credit card digits). In practice, there may be additional requirements on the form of the output function, such as a limit on the number of steps required to evaluate it (analogous to requiring that the instructions for a task be no more than two pages).

**Integrity, availability, and properties**

Although required information release is primarily concerned with the confidentiality of information, it is also related to the integrity and availability of information. Weak information release is an integrity requirement: if the output expression evaluates to an integer value, it must equal the evaluation of the input expression. Strong information release contains an availability requirement: the output expression must eventually evaluate to an integer value. Information security requirements are not always easily separable into confidentiality, integrity, and availability requirements.

Weak and strong required release can be defined as *properties*: predicates over single execution traces. Weak required release is a safety property, and strong required release is neither safety nor
liveness [Alpern and Schneider, 1985]. Recent work by Clarkson and Schneider [2008] expresses some information-flow conditions as hyperproperties: predicates of sets of traces. They note that all information-flow conditions they considered were hyperproperties and not properties. Although weak and strong required release are properties, they are clearly constitute part of a systems information flow requirements, so some information-flow conditions of interest are properties. Indeed, in Section 4.3 below, we discuss the relationship between required release and delimited release, an information flow security condition for permitted information release that is a hyperproperty and not a property.

4.2 Noninterference

Noninterference [Goguen and Meseguer, 1982] is a well-known semantic security condition that requires that public observations reveal no secrets. Applied to the interactive setting used here, noninterference ensures that user strategy of any such user does not gain any knowledge about the strategy employed by any user such that that strategy extends, or vice-versa. This notion of trace indistinguishability is suitable given the observational model, which does not allow a user to directly distinguish a terminated program from a program in a non-terminating loop, or from a program that may eventually produce additional output.

Definition 3 (Noninterference). A command c satisfies noninterference exactly when for all levels l ∈ L:

For all m = (c, σ, (s), ω) and m′ = (c, σ, (s), ω′)

such that ω(l′) = ω′(l′) for all l′ ⊆ ℓ, and for all traces t, t′ such that m ≒ t and m′ ≒ t′,

we have t ≒ t′.

Note that both weak and strong required release of f to ℓ violate noninterference if the input expression f contains an input expression in_{i} such that ℓ′ ∉ ℓ (and the evaluation of f depends nontrivially on in_{i}). For example, any program that satisfies (weak or strong) required release of in_{i} mod 10,000 to L (such as program Pk) must violate noninterference, since H ⊆ L, and L learns something about the strategy of H, to wit, the last four digits of the credit card number that H entered.

4.3 Delimited and bounded release

Noninterference is a very restricted security condition. Many real-world programs must violate noninterference in order to satisfy functional requirements that require or allow the release of information.

The security condition delimited release [Sabelfeld and Myers, 2004] weakens noninterference by specifying what information a program is permitted to release. An escape hatch is a pair (f, ℓ) of input expression f and security level ℓ. Intuitively, given escape hatch (f, ℓ), a program is permitted to release information f to security level ℓ. Thus, given escape hatches (f₀, ℓ₀), . . . , (fₖ, ℓₖ), user ℓ is permitted to learn the evaluation of fᵢ for any escape hatch (fᵢ, ℓᵢ) such that ℓᵢ ⊆ ℓᵢ, in addition to the user strategies of any user ℓ such that ℓ ⊆ ℓ.

A program satisfies delimited release by escape hatches (f₀, ℓ₀), . . . , (fₖ, ℓₖ) if, for any security level ℓ and configurations m and m′ that have the same user strategy for any user ℓ such that ℓ′ ⊆ ℓ, if m and m′ respectively emit traces t and t′ that agree on the evaluation of all escape hatches that may release information to level ℓ, then the traces emitted by m and m’ are indistinguishable to user ℓ. Formally, we say that traces t and t’ agree up to ℓ on escape hatches (f₀, ℓ₀), . . . , (fₖ, ℓₖ) if for all i ∈ 0..k such that ℓᵢ ⊆ ℓ, we have t ⊑ₘ fᵢ ⇐ vᵢ and t’ ⊑ₘ fᵢ ⇐ vᵢ for some vᵢ ≠ ⊥.

Definition 4 (Delimited release). Command c satisfies delimited release by escape hatches (f₀, ℓ₀), . . . , (fₖ, ℓₖ) exactly when for all levels ℓ ∈ L:

For all m = (c, σ, (s), ω) and m′ = (c, σ, (s), ω′)

such that ω(l′) = ω′(l′) for all ℓ′ ⊆ ℓ,

and for all traces t, t′ such that m ≒ t and m′ ≒ t′,

if t and t’ agree up to ℓ on escape hatches (f₀, ℓ₀), . . . , (fₖ, ℓₖ),

then t ≒ t’.

Delimited release generalizes noninterference: if command c satisfies delimited release by an empty set of escape hatches, then c satisfies noninterference.

Both delimited release and required release are concerned with information flow, and with the knowledge an observer acquires. Required release specifies what information, at a minimum, a program must release. It specifies what an observer must be able to (explicitly) know, and can be viewed as specifying “lower bounds” on what information a program releases. By contrast, delimited release specifies what an observer is permitted to (implicitly) know, and can be seen as specifying “upper bounds”, the maximum information a program is permitted to release. We can combine the security conditions of delimited release and required release to obtain both upper and lower bounds on a program’s information release.

An escape chute is a tuple (f, ℓ, g) of input expression f, security level ℓ, and output expression g. We define bounded release by combining delimited release by escape hatches, and required release by escape chutes.

Definition 5 (Bounded release). Command c satisfies weak (strong) bounded release by escape chutes

(f₀, ℓ₀, g₁), . . . , (fₖ, ℓₖ, gₖ)

and escape hatches

(f₀′, ℓ₀′), . . . , (fₖ′, ℓₖ′)

effectively when

1. for all i ∈ 0..k, c satisfies weak (strong) required release of fᵢ to ℓᵢ using gᵢ; and

2. c satisfies delimited release by escape hatches (f₀′, ℓ₀′), . . . , (fₖ′, ℓₖ′).

Program Pk satisfies strong bounded release by escape chute (in_{i} mod 10,000, L, out[1]) and escape hatch (in_{i} mod 10,000, L). Thus, bounded release tells us not only that Pk releases the input expression in_{i} mod 10,000, but also that this is the only information released by Pk.

The following program has different upper and lower bounds. It satisfies strong bounded release by escape chute (in_{i} mod 10,000, L, out[1]) and escape hatch (in_{i} mod 10,000, L) and (in_{i} div 10^15, L). It always releases the last four digits of the credit card number (via output expression out[1]) and it may in
bounded release by escape chutes

The escape hatches, the “upper bounds” of information release. More precisely, if $t$ and $t'$ are traces that can be produced by a command satisfying bounded release, and $t$ and $t'$ agree on all input and output events on all channels up to some level $\ell$, and agree on the value of all escape hatches that declassify to $\ell$ or below, then for each escape chute at level $\ell$ or below, either $t$ and $t'$ agree on the value of the escape chute, or $t$ and $t'$ do not have sufficient input events to evaluate the escape chute. We say that traces $t$ and $t'$ agree on escape chute $(f_i, \ell_i, g_i)$ if $t \equiv_{in} f_i \Downarrow v_i$ and $t' \equiv_{in} f_i \Downarrow v_i$ for some $v_i \neq \bot$.

**Property 1** (Consistency). If command $c$ satisfies (weak or strong) bounded release by escape chutes $(f_0, \ell_0, g_0), \ldots, (f_n, \ell_n, g_n)$ and escape hatches $(f'_0, \ell'_0), \ldots, (f'_n, \ell'_n)$, then

- For all $\ell \in L$, and for all configurations $m = (e, \sigma, ()), \omega)$ and $m' = (e, \sigma, ()'), \omega')$, and for all traces $t, t'$ such that $m \leadsto t$ and $m' \leadsto t'$,
  - if $t$ and $t'$ agree up to $\ell$ on escape hatches $(f'_0, \ell'_0), \ldots, (f'_n, \ell'_n)$, and
  - for all $t' \subseteq \ell$ we have $t|\ell = t'|\ell$
  - then for all $i \in \ell$, such that $\ell_i \subseteq $, either $t \equiv_{in} f_i \Downarrow v_i$ or $t' \equiv_{in} f_i \Downarrow v_i$, or $t$ and $t'$ agree on escape chute $(f_i, \ell_i, g_i)$.

**Proof.** (Sketch) Given $\ell$ and $t$ and $t'$ such that $t$ and $t'$ agree up to $\ell$ on escape hatches $(f'_0, \ell'_0), \ldots, (f'_n, \ell'_n)$, and $t \equiv t' \equiv \ell'$ for all $\ell' \subseteq \ell$, then we can construct joint user strategies $\omega$ and $\omega'$ such that $m \leadsto t$ and $m' \leadsto t'$, and $\omega(\ell') = \omega'(\ell')$ for all $\ell' \subseteq \ell$.

For any escape chute $(f_i, \ell_i, g_i)$ such that $\ell_i \subseteq \ell$, suppose $t \equiv_{in} f_i \Downarrow v_i \infer{t'}{t_i' \equiv_{in} f_i' \Downarrow v'_i}$ for some $v_i, v'_i \neq \bot$. By delimited release, $t \equiv_{\ell_i} t'$, and so $t$ and $t'$ agree on the values of all output expressions required to evaluate $g_i$ to an integer value. Therefore, $v_i = v'_i$. By bounded release, the evaluation of $f_i$ in $t$ and $t'$ also equal $v_i$, and so $t$ and $t'$ agree on escape chute $(f_i, \ell_i, g_i)$.

**5. Enforcement**

In this section we show that weak bounded release can be soundly enforced with a security type system. Weak bounded release is the conjunction of weak required release, and delimited release. Since weak required release is a safety property, clearly other enforcement mechanisms could also be used to enforce it, including dynamic mechanisms such as execution monitors. However, due to the similarity of escape chutes and escape hatches, a type system to enforce delimited release can easily be adapted to enforce weak bounded release as well.

Our type system conservatively tracks both the security level of information as it flows through a program, and what input expressions have been output and declassified. This allows us to ensure that (i) confidential information is never output to non-confidential channels; (ii) only appropriate escape hatches are declassified; and (iii) appropriate escape chutes are output to the correct channel in the correct order.

For command $c$, type judgments have the form

$$pc, \Gamma, C, \Delta, E, H \vdash c : \Gamma' ; C', \Delta', E', H'$$

where entities to the left of the turnstile ($\vdash$) indicate the context before the execution of $c$, and primed versions on the right hand side of the judgment indicate how the contexts change as a result of the execution of $c$.

Security level typing context $\Gamma$ maps variables to security levels, and indicates an upper bound on the information stored in each variable. Program counter level $pc$ is an upper bound on the information that may cause command $c$ to be executed, and is used to track implicit information flows [Denning and Denning, 1977]. Typing context $\Gamma$ and program counter level $pc$ are standard in security type systems. Our type system is flow-sensitive, as it allows command $c$ to modify $\Gamma$, and is based on the flow-sensitive security type system of Hunt and Sands [2006]. The remaining entities in the context $(C, \Delta, E, H)$ are used to track what input expressions have been output and declassified. Specifically, we conservatively track how many input and output events have been produced on each channel, what input expression (if any) is stored in each variable, what input expressions (if any) have been output, and what input expressions have been declassified.

- $C : L \rightarrow \mathbb{Z} \times \mathbb{Z}$ counts the input and output events that have occurred on each channel. If $C(\ell) = (i, j)$, then the program has received $i$ input events from channel $\ell$, and produced $j$ output events to channel $\ell$. If $i = \bot$, then an unknown number of input events have been received on channel $\ell$, and similarly, if $j = \bot$, an unknown number of output events have been produced.
- $\Delta : \mathcal{V} \rightarrow \mathcal{I} \times \mathcal{I}$ indicates what input expression is stored in each variable. For any variable $x$, if $\Delta(x) = f$ then the value stored in $x$ is equivalent to input expression $f$. If $\Delta(x) = \bot$ then nothing is known about the value stored in $x$.
- $E : L \times \mathbb{Z} \rightarrow \mathcal{I}$ indicates which input expressions have been output to channels.\(^{3}\) If $E(\ell, n) = f$ then the $n$th value output on channel $\ell$ was equal to the evaluation of input expression $f$. If $E(\ell, n) = \bot$ then either the $n$th output to channel $\ell$ has not yet been produced, or nothing is known about the $n$th output to channel $\ell$.
- $H : \vartriangleright$ is a set of escape chutes that may have been declassified.

Figure 3 presents inference rules for the typing judgment. Given a function $h$, we write $h[a \rightarrow b]$ for the function that evaluates to $b$ on input $a$, and otherwise behaves like $h$. We use $\Gamma(e)$ to denote an upper bound of all levels $\Gamma(x)$ for variables $x$ occurring in $e$; if $L$ is a join semi-lattice, then this is the join of all $\Gamma(x)$ for $x \in E$. We extend function $\Delta$ to a homomorphism on program expressions, and write $\Delta(e)$ for the result of applying the homomorphism to expression $e$.

In the typing rules, security level context $\Gamma$ and program counter level $pc$ do not interact with other parts of the context, and by themselves form a standard flow-sensitive information-flow security type system, similar to that of Hunt and Sands [2006]. In the following discussion of the typing rules, we focus on the type system’s novel aspects.

For assignment $x := e$, $\text{T-Assign}$ updates input expression context $\Delta(x)$ for $\Delta(e)$, which is either $\bot$ or an input expression equal to $e$ at this program point. The typing rule $\text{T-Declassify}$ for declassification $x := \text{declassify}(e \text{ to } \ell)$ is similar to assignment, but escape hatch $(\Delta(e), \ell)$ is added to declassification effect $H$. Note that the rule implicitly requires $\Delta(e) \neq \bot$ since $H$ must

\(^{3}\)“$E$” is the uppercase “$e$”, the Greek letter epsilon.
contain escape hatches. Rule T-SEQ simply threads the context through a sequence of commands. A skip command has no effect on the context, shown in rule T-Skip.

Command input \(x\) from \(\ell\) assigns the next input from channel \(\ell\) to variable \(x\). Rule T-IN updates input expression context \(\Delta\) using helper function \(\text{recordInput}(\Delta, x, \ell)\), which updates \(\Delta(x)\) either to \(\perp\) if the number of input events on channel \(\ell\) is not known, or to input expression \(\text{in}_i[x]\), where \(i\) is the number of input events received on channel \(\ell\). If known, the number of input events on channel \(\ell\) is incremented using the helper function \(\text{inc}_{\text{in}}(\ell, C, \ell)\).

Command output \(e\) to \(\ell\) outputs expression \(e\) to channel \(\ell\). Using helper function \(\text{recordOutput}(E, C, \ell, f)\), rule T-OUT records that the \(j\)th output on channel \(\ell\) is equal to input expression \(\Delta(e)\), where \(j\) is the number of output events produced on channel \(\ell\), and increments the number of output events produced on channel \(\ell\) with helper function \(\text{inc}_{\text{out}}(C, \ell)\). If the number of output events produced on channel \(\ell\) is unknown (i.e., \(\perp\)), then no update to \(E\) or \(C\) is made.

The subsumption rule T-SUB allows the context to be weakened, or made less precise. It uses the flat ordering \(\succeq\): for any lifted set \(\mathbf{S}_{\perp}\), and for any \(a, b \in \mathbf{S}_{\perp}\), \(a \succeq b\) iff \(a = b\) or \(b = \perp\). We extend the \(\succeq\) relation in the obvious way to pairs, and to a pointwise relation over functions. For example, \(\Delta_0 \succeq \Delta_1\) iff for all \(x \in \mathbf{Var}\), \(\Delta_0(x) \succeq \Delta_1(x)\). Similarly, we extend the binary relation \(\sqsubseteq\) over \(\mathcal{L}\) to a pointwise relation over functions with codomain \(\mathcal{L}\).

The rules for if and while commands (T-IF and T-WHILE respectively) type check their sub-commands with a program counter level bounded below by \(pc\) and \(\Gamma(e)\), since \(c\) controls the execution of the sub-commands. Rule T-WHILE requires that context is unchanged by the execution of the while command; for any channel \(\ell\), this means either that the loop body performs no input or output on \(\ell\), or that the context cannot precisely track the number on inputs or outputs received on channel \(\ell\), i.e., \(\Delta(\ell) = (i, \perp)\) and \(\perp \in \{i, j\}\). Similarly for an if command, the context will lose track of the number on inputs or outputs received on channel \(\ell\) unless both branches always perform the same number of inputs and outputs on \(\ell\).

The type system can easily be converted into an algorithmic type system, using the same technique as Hunt and Sands [2006]. If the security levels \(\mathcal{L}\) and binary relation \(\sqsubseteq\) form a join-semi lattice,
then type checking and type inference with the algorithmic type system is decidable in time polynomial in the size of the program.

If command $c$ is well-typed, then it satisfies both weak required release, and delimited release. Theorem 1 below states this claim formally. To state Theorem 1 concisely, we first introduce a helper function and additional notation.

Helper function $\text{substOutExp}(E, \ell, g)$ takes output context $E$, security level $\ell$, and output expression $g$, and substitutes any occurrence of $\text{out}[i]$ with input expression $E(\ell, i)$, that is, the input expression that $E$ claims was the $i$th output on channel $\ell$. For example, if $E((L, 2)) = \text{inu}_{\{0\}}[1]$, then $\text{substOutExp}(E, L, 42 + \text{out}[2]) = 42 + \text{inu}_{\{0\}}[1]$. Rules for $\text{substOutExp}(E, \ell, g)$ are given in Figure 4.

We assume there is a notion of equivalence between input expressions, denoted by $\equiv$. We require that if $f_0 \equiv f_1$, then for all traces $t$ and $v \in \mathbb{Z}_+$, $t \vdash_m f_0 \downarrow v$ if $t \vdash_m f_1 \downarrow v$. The equivalence relation could be syntactic identity, or syntactic identity up to commutativity and associativity of operators, or, (depending on the operators in the language) a deeper semantic equivalence.

Finally, for any set $S$ and $v \in S$, we use $\tau$ as shorthand for a constant function that always returns $v$. For example, $(0, 0)$ is a function that always returns the pair $(0, 0)$.

**Theorem 1.** If $p, \Gamma_0; (0, 0), \top, \top, \emptyset \vdash c \triangleright \Gamma; C, \Delta, E, H$ for some $\Gamma_0$ and $pc$, then
1. $c$ satisfies weak required release of input expression $f$ to user $\ell$ using output expression $g$ if $\text{substOutExp}(E, \ell, g) \equiv f$.
2. $c$ satisfies delimited release by escape hatches $(f_0, \ell_0), \ldots, (f_k, \ell_k, g_k)$ and escape hatches $(f_0', \ell_0'), \ldots, (f_n', \ell_n')$ if for all $(f, \ell) \in H$ there exists $i \in 0..k$ such that $f \equiv f_i$ and $\ell_i \subseteq \ell$.

A proof of Theorem 1 appears in Appendix A.

If command $c$ is well-typed, then it satisfies both weak required release, and delimited release, it satisfies weak bounded release.

**Corollary 1.** Command $c$ satisfies weak bounded release by escape chutes $(f_0, \ell_0, g_0), \ldots, (f_k, \ell_k, g_k)$ and escape hatches $(f_0', \ell_0'), \ldots, (f_n', \ell_n')$ if
\[
\text{pc}, \Gamma_0; (0, 0), \top, \top, \emptyset \vdash c \triangleright \Gamma; C, \Delta, E, H
\]
for some $\Gamma_0$ and $pc$, and $\text{substOutExp}(E, \ell, g_i) \equiv f_i$ for all $i \in 0..k$ and for all $(f, \ell) \in H$ there exists $i \in 0..n$ such that $f \equiv f'_i$ and $\ell'_i \subseteq \ell$.

**Proof.** Immediate from Theorem 1.

Although program $P_\delta$ satisfies bounded release, it does not type-check: it attempts to release information from $H$ to $L$ but does not have any declassify annotations. Program $P_\delta$ does type-check. The judgment
\[
\Gamma; \top; (0, 0), \top, \top, \emptyset \vdash P_\delta \triangleright \Gamma; C, \Delta, E, H
\]
holds for
\[
\Gamma = \top[cc \mapsto H, x \mapsto L, y \mapsto L]
\]
\[
C = (\langle 0, 0 \rangle | H \mapsto (1, 0), L \mapsto (0, \bot))
\]
\[
\Delta = \top[cc \mapsto \text{inu}_{\{0\}}[0], x \mapsto \text{inu}_{\{0\}}[0] \mod 10,000, y \mapsto \text{inu}_{\{0\}}[0] \div 10^{15}]
\]
\[
E = \top[(L, 0) \mapsto \text{""""}, (L, 1) \mapsto \text{inu}_{\{0\}}[0] \mod 10,000]
\]
\[
H = \top[(\text{inu}_{\{0\}}[0] \mod 10,000, L), (\text{inu}_{\{0\}}[0] \div 10^{15}, L)].
\]

Thus, by Corollary 1, $P_\delta$ satisfies weak bounded release by escape chute $(\text{inu}_{\{0\}}[0] \mod 10,000, L, \text{out}[1])$ and escape hatches $(\text{inu}_{\{0\}}[0] \mod 10,000, L)$ and $(\text{inu}_{\{0\}}[0] \div 10^{15}, L)$.

A more sophisticated static analysis (or a more restrictive language) could enforce strong required release, by reasoning about the termination of loops, and the eventual production of outputs.

### 6. Related work

Much recent work has considered information release. Sabelfeld and Sands [2005] present a survey of work on (permitted) information release, and introduce four aspects of declassification: who, what, where and when. These aspects are also relevant to required information release. This work is primarily concerned with what information is required for release, expressed using input expressions. Strong required release relates to the when aspect: it mandates that information is eventually released, whereas weak required release places no requirements on when (if ever) information is released.

Sabelfeld and Sands [2005] also introduce several prudent principles of declassification. Of these principles, semantic consistency is directly applicable to required release, and is satisfied by weak and strong required release, and bounded release: semantically equivalent programs satisfy the same security conditions. The other principles do not seem directly applicable to required release.

Previous work on specification and enforcement of information release focuses on permitted information release. To the best of our knowledge, this work is the first to address required information release. Appropriate extensions to the policies and enforcement mechanisms of previous work may allow them to reason about required release. The automata for information release of Swamy and Hicks [2008], and the flowspecs used by Banerjee et al. [2008] seem particularly suitable for extending for required release.

Askarov and Sabelfeld [2007b] introduce localized delimited release, which refines delimited release by restricting not only what information may be released, but where it may be released (at an appropriate declassify command). The type system used by Sabelfeld and Myers [2004] to enforce delimited release also enforces localized delimited release, so we speculate that the type system used in this paper to enforce bounded release would also enforce an appropriately defined localized bounded release.

Askarov and Sabelfeld [2007a] define the semantic security condition gradual release in terms of attacker knowledge: an attacker’s knowledge of secrets may become more precise only at specified declassification events. Gradual release restricts permitted information release, and as such it suffices to use implicit knowledge; since we are concerned with required release, we use algorithmic knowledge to ensure that knowledge can be obtained with reasonable resources. The use of algorithmic knowledge leads us to specify how an observer learns released information, in addition to what information they learn.

Askarov and Sabelfeld [2009] present semantic security conditions that generalize localized delimited release and gradual release, and consider their enforcement using a combination of static and dynamic techniques in an interactive language. Since weak required release is a safety property, it could be enforced using dynamic techniques, such as execution monitors [Schneider, 2000].
O’Neill [2006] presents many information flow conditions in an epistemic framework, but doesn’t consider algorithmic knowledge or required information release.

As discussed in Section 4.1, required information release is concerned with the integrity and availability of information, in addition to its confidentiality. Zheng and Myers [2005] study the end-to-end enforcement of availability policies, and present a policy framework for specifying confidentiality, integrity, and availability policies. Their policies are based on the decentralized label model [Myers and Liskov, 2000], and specify who may affect the confidentiality, integrity, and availability of information. By contrast, this work focuses on what aspects of availability and integrity: what information must be available, and what it must be equal to.

7. Conclusion
As part of their correct functionality, many systems are required (not just permitted) to release information. This paper introduces the problem of required information release: specifying, reasoning about, and enforcing, the information security of systems that must release information.

We have defined semantic conditions for required information release. Inspired by work on algorithmic knowledge, the semantic conditions must specify both what information is to be released, and how that information is to be learned by an observer. Input expressions specify what information is to be released, and output expressions specify how an observer learns the information. A program satisfies weak required release of input expression $f$ to user $\ell$ using output expression $g$ if whenever user $\ell$ is able to evaluate $g$, then $f$ evaluates to the same value. A program satisfies strong required release if it satisfies weak required release, and eventually produces sufficient output for user $\ell$ to evaluate $g$.

We investigated the relationship between a system’s required and permitted information release, and defined bounded release, which combines required release with delimited release. Bounded release specifies upper and lower bounds on the information a system releases. For many systems, these bounds should be tight: the system should release all and only information it is required to release. We have shown that (weak) bounded release can be conservatively enforced by a type system.

Both weak and strong required release are properties: predicates over single execution traces. Noninterference, delimited release, and many other information security requirements, are hyperproperties, but not properties. One may thus be concerned whether required information release is an information security requirement. We believe that required information release, while a property, is clearly concerned with the flow of information in a system: it requires that, at a minimum, certain information flows to an observer. We have shown a connection between required information release and delimited release: whereas required information release specifies the minimum information flow from high security inputs to low security outputs that a system must satisfy, delimited release specifies maximum information flow. Thus, we believe that required information release is part of a system’s information security requirements.

There is still much left to understand with respect to required information release. There are systems with information release requirements that cannot be expressed using the policies presented in this paper. For example, financial reports of a company should be released to all shareholders, not a subset: if Alice and Bob are the shareholders, the system must release reports to Alice if and only if it releases that information to Bob. In terms of enforcing required information release, it may be impractical to explicitly specify the knowledge algorithm by which an observer may learn the released information; static analyses may allow the automatic discovery of the knowledge algorithm, thus reducing the burden of proving a system satisfies required information release.

To build trustworthy computer systems, it is important to understand and provably enforce a system’s information security requirements. By introducing the concept of required information release, and providing mechanisms to specify and enforce these requirements, this work brings us closer to the goal of strong, end-to-end, application-specific information security.

Acknowledgments
We thank Andrew Myers for very useful discussions and feedback about this work. We also thank Michael Clarkson, Allan Friedman, Tyler Moore, Kevin O’Neill, Fred Schneider, and Jeff Vaughan for interesting and helpful discussions related to this work, and the anonymous reviewers for their useful comments.

References


A. Proof of Theorem 1

In this appendix we prove Theorem 1. Section A.1 shows that a well-typed program satisfies required release, and Section A.2 shows that a well-typed program satisfies delimited release. We first introduce and prove some useful lemmas and theorems about the type system.

We say that a configuration \((c, \sigma, t, \omega)\) satisfies context \(C, \Delta, E\) if the context \(C, \Delta, E\) accurately reflects the configuration. That is, the entity \(C\) records how many input and output events have been received and sent on each channel, and must agree trace \(t\). Similarly, \(\Delta\) records input expressions that are equivalent to values stored in the state, and the state \(\sigma\) and trace \(t\) must agree on these values. Finally, \(E\) records input expressions that are equivalent to output values, and \(t\) must satisfy these relationships.

**Definition 6.** We say that configuration \((c, \sigma, t, \omega)\) satisfies context \(C, \Delta, E,\) written \(C, \Delta, E, \vdash (c, \sigma, t, \omega)\) if all of the following conditions hold.

1. For all \(\ell \in \mathcal{L}\), let \(C(\ell) = (i, j)\).
   (a) Either \(i = \bot\) or \(|\{\ell | (\text{Ev}_{\text{in}} \cap \text{Ev}(\ell))\}| = i\); and
   (b) Either \(j = \bot\) or \(|\{\ell | (\text{Ev}_{\text{out}} \cap \text{Ev}(\ell))\}| = j\).
2. For all \(x \in \text{Var}\), either \(\Delta(x) = \bot\) or \(t \vdash_{\text{in}} \Delta(x) \nvdash \sigma(x)\).
3. For all \(\ell \in \mathcal{L}\) and \(i \in \mathbb{N}\), either \(E(\ell, i) = \bot\) or \(t \vdash_{\text{in}} E(\ell, i) \nvdash \nu\) where \((\{\ell | (\text{Ev}(\ell) \cap \text{Ev}_{\text{out}})(i) = \text{out}(\ell, \nu)\}) \). If a configuration satisfies context \(C, \Delta, E\), then it also satisfies a weaker, or less precise, context \(C', \Delta', E', \).

**Lemma 1.** If \(C, \Delta, E, \vdash (c, \sigma, t, \omega)\) and \(C \supseteq C'\) and \(\Delta \supseteq \Delta'\) and \(E \supseteq E'\) then \(C', \Delta', E', \vdash (c, \sigma, t, \omega)\).

**Proof.** For any \(\ell\), let \(C'(\ell) = (i', j')\) and \(C(\ell) = (i, j)\). If \(i' \neq i\) then \(i' = \bot\), and similarly if \(j' \neq j\) then \(j' = \bot\). Therefore condition 1 is satisfied.

Similarly, if \(\Delta(x) \neq \Delta'(x)\), then \(\Delta'(x) = \bot\), and so condition 2 is satisfied.

Finally, for all \(\ell, t\), and \(i\), if \((E(\ell, i)) \neq (E'(\ell, i))\) then \(E'(\ell, i) = \bot\) and so condition 3 is satisfied. \(\square\)

The operational semantics preserve typings.

**Lemma 2 (Type preservation).** If

\[
C, \Delta, E, \vdash (c, \sigma, t, \omega)
\]

and

\[
\text{pc}, \Gamma; C, \Delta, E, H \vdash c \triangleright \Gamma'; C', \Delta', E', H'
\]

and

\[
(c, \sigma, t, \omega) \longrightarrow (c', \sigma', t', \omega),
\]

then there exists \(C'', \Delta'', E'', \Gamma''\) and \(\text{pc}''\) such that

\[
\text{pc}''; \Gamma''; C'', \Delta'', E'', H'' \vdash c' \triangleright \Gamma''; C', \Delta', E', H'
\]

and

\[
C'', \Delta'', E'', \vdash (c', \sigma', t', \omega).
\]

**Proof.** By induction on the typing judgment \(\text{pc}., \Gamma; C, \Delta, E, H \vdash c \triangleright \Gamma'; C', \Delta', E', H'\).

- **T-Skip**
  Trivial, as there is no \((c', \sigma', t', \omega)\) such that \((\text{skip}, \sigma, t, \omega) \longrightarrow (c', \sigma', t', \omega)\).

- **T-Assign**
  In this case, \(c = x := e\), \(c' = \text{skip}\), \(\sigma' = \sigma[x \mapsto \sigma(e)]\), \(t' = t\), \(C' = C\), \(\Delta' = \Delta[x \mapsto \Delta(e)]\), \(E' = E\), \(H' = H\), \(\Gamma' = \Gamma[x \mapsto \ell]\).

  Let \(C'' = C', \Delta'' = \Delta', E'' = E', H'' = H', \Gamma'' = \Gamma'\). Then \(\text{pc}, \Gamma''; C'', \Delta'', E'', H'' \vdash c' \triangleright \Gamma''; C', \Delta', E', H'\). Since \(C, \Delta, E, \vdash (c, \sigma, t, \omega)\) we have for all \(y \in \text{Var}\) such that \(y \neq x\) either \(\Delta''(y) = \bot\) or \(t \vdash_{\text{in}} \Delta''(y) \nvdash \sigma(y)\). We have \(\Delta''(x) = \Delta(e)\), and if \(\Delta(e) \neq \bot\) then, since \(\sigma''(x) = \sigma(x)\), we can show by induction on the structure of \(c\) that \(t \vdash_{\text{in}} \Delta''(x) \nvdash \sigma(x)\). Moreover, since \(C'' = C\) and \(E'' = E\) and \(t' = t\) we have \(C'', \Delta'', E'', \vdash (c', \sigma', t', \omega)\).

- **T-Seq**
  In this case, \(c = c_0; c_1\). If \(c_0 \neq \text{skip} \) then the result follows trivially from the inductive hypothesis. If \(c_0 = \text{skip}\) then \(c' = c_1\), and \(\sigma' = \sigma\) and \(t' = t\). The result follows easily from \(\text{pc}, \Gamma; C, \Delta, E, H \vdash c' \triangleright \Gamma''; C', \Delta', E', H'\).

- **T-In**
  In this case, \(c = \text{input} \ell\) from \(x, c' = \text{skip}, \sigma' = \sigma[x \mapsto v], t' = t'(\text{in}(\ell, v))\), \(C' = \text{inc}_{\text{in}}(C, \ell), \Delta' = \text{recordInput}(\Delta, x, C, \ell), E' = E, H' = H, \Gamma' = \Gamma[x \mapsto \ell]\).

  Let \(C'' = C', \Delta'' = \Delta', E'' = E', H'' = H', \Gamma'' = \Gamma'\). Then \(\text{pc}, \Gamma''; C'', \Delta'', E'', H'' \vdash c' \triangleright \Gamma''; C', \Delta', E', H'\). If \(C'' \neq C\) then \(C''(\ell) = (i, j)\) for some \(i, j\) and \(C(\ell) = (i - 1, j)\). Since \(|\{\ell | (\text{Ev}_{\text{in}} \cap \text{Ev}(\ell))\}| = i - 1\), we have \(|\ell' | (\text{Ev}_{\text{in}} \cap \text{Ev}(\ell))\}| = i\). Also, either \(j = \bot\) or \(|\ell' | (\text{Ev}_{\text{out}} \cap \text{Ev}(\ell))\}| = j\).

  If \(\Delta''(x) \neq \bot\) then \(i \neq \bot\) and \(\Delta''(x) = \text{in}_{\text{in}}[i]\). Since \(\sigma'(x) = v\) and \((t' | (\text{Ev}(\ell) \cap \text{Ev}_{\text{in}}))(i) = \text{in}(\ell, v)\) we have \(t' \vdash_{\text{in}} \Delta(x) \nvdash \sigma'(x)\). Thus we have \(C'', \Delta'', E'', \vdash (c', \sigma', t', \omega)\) as required.
Lemma 3.
Proof.

The type system ensures that the count of input and output events on each channel can only increase, or lose precision.

Similarly, the type system ensures that for typing context E (which records input expressions that are equivalent to output values), the post-context may be less precise than the pre-context, but otherwise agrees with it. That is, once the type system has recorded that a given input expression is equivalent to a given output value, the type system can not change it to a different input expression.

Lemma 4. Given \( \text{pc}, \Gamma; \Delta, E, H \vdash e \triangleright \Gamma'; C', \Delta', E', H' \) and \( \ell \in L \), let \( C(\ell) = (i, j) \) and \( C'(\ell) = (i', j') \). Either \( i \neq \perp \) and \( i' \neq \perp \) and \( i < i' \) or \( i' = \perp \).

Proof. By induction on \( \text{pc} \), \( \Gamma; \Delta, E, H \vdash e \triangleright \Gamma'; C', \Delta', E', H' \).

- **T-Skip, T-Assign, T-While, T-Declassify**
  Trivial, as \( C' = C \).

- **T-Seq, T-If**
  Follows easily from the inductive hypothesis.

- **T-Sub**
  Follows easily from the inductive hypothesis, and ordering \( \geq \).

- **T-In**
  Here \( C' \) is identical to \( C \) with one possible exception for some \( \ell' \), in which case \( C'(\ell') = (i + 1, j) \) where \( C(\ell') = (i, j) \).

- **T-Out**
  Here \( C' \) is identical to \( C \) with one possible exception for some \( \ell' \), in which case \( C'(\ell') = (i, j + 1) \) where \( C(\ell') = (i, j) \).

The type system ensures that the count of input and output events on each channel can only increase, or lose precision.

**Lemma 3.** Given \( \text{pc}, \Gamma; \Delta, E, H \vdash e \triangleright \Gamma'; C', \Delta', E', H' \) and \( \ell \in L \), let \( C(\ell) = (i, j) \) and \( C'(\ell) = (i', j') \). Either \( i \neq \perp \) and \( i' \neq \perp \) and \( i < i' \) or \( i' = \perp \).

**Proof.** By induction on \( \text{pc} \), \( \Gamma; \Delta, E, H \vdash e \triangleright \Gamma'; C', \Delta', E', H' \).
• **T-SEQ**
  Here $c = c_0, c_1$ and $pc, \Gamma; C, \Delta, E, H \vdash c_0 \triangleright \Gamma''; C''', \Delta''', E'', H''$ and $pc''', \Gamma'''; C''', \Delta''', E'', H'' \vdash c_1 \triangleright \Gamma'\; C', \Delta', E', H'$.
  If $j = \bot$ then by the inductive hypothesis, we have for all $k \in \mathbb{N}$, we have $E(\ell, k) \geq E''(\ell, k)$. Moreover, by Lemma 3, $C''''(\ell) = (i'', \bot)$ for some $i''$. Thus, by another application of the inductive hypothesis, we have for all $k \in \mathbb{N}$, we have $E''''(\ell, k) \geq E''(\ell, k)$, and so for all $k \in \mathbb{N}$, we have $E(\ell, k) \geq E''(\ell, k)$ as required.
  If $j \in \mathbb{N}$ then by the inductive hypothesis, for all $0 \leq k < j$, we have $E(\ell, k) \geq E''''(\ell, k)$. By Lemma 3, $C''''(\ell) = (i'', j')$ and either $j' > j$ or $j' = \bot$. Either way, by another application of the inductive hypothesis for all $0 \leq k < j$, we have $E''''(\ell, k) \geq E''(\ell, k)$, and so for all $0 \leq k < j$, we have $E(\ell, k) \geq E''(\ell, k)$ as required.

• **T-SUB**
  Here we have $pc_1, \Gamma_1; C_1, \Delta_1, E_1, H_1 \vdash c \triangleright \Gamma''_1; C''_1, \Delta''_1, E''_1$ where $E \geq E_1$ and $E'' \geq E''_1$.
  If $j = \bot$ then $E_1(\ell) = (i_1, \bot)$ for some $i_1$, and by the inductive hypothesis, for all $k \in \mathbb{N}$ we have $E_1(\ell, k) \geq E''(\ell, k)$, and since $E'' \geq E''_1$, we have for all $k \in \mathbb{N}$ we have $E(\ell, k) \geq E''(\ell, k)$ as required.
  If $j \bot$ then $E_1(\ell) = (i_1, j_1)$ for some $i_1, j_1$ where either $j_1 = j$ or $j_1 = \bot$. Either way, by the inductive hypothesis, for all $0 \leq k < j$, we have $E_1(\ell, k) \geq E''(\ell, k)$, and since $E'' \geq E'$, we have for all $k \in \mathbb{N}$ we have $E(\ell, k) \geq E''(\ell, k)$ as required.

• **T-IF**
  Trivial, by inductive hypothesis.

• **T-OUT**
  Here, $E' = \text{recordOutput}(E, C, \ell', f)$ for some $\ell'$ and $f$. If $\ell \neq \ell'$, then $E(\ell, k) = E''(\ell, k)$ for all $k \in \mathbb{N}$. If $\ell = \ell'$ and $j = \bot$, then $E = E'$. If $\ell = \ell'$ and $j \neq \bot$, then $E(\ell, k) = E''(\ell, k)$ for all $0 \leq k < j$. In all cases, the required result holds.

The substitution function $\text{substOutExp}(E, \ell, g)$ is correct, in that if output expression $g$ evaluates to a value $v$, then replacing all $\text{out}[i]$ subexpressions with appropriate input expressions stored in $E$ will result in an input expression that also evaluates to $v$.

**Lemma 5.** If $C, \Delta, E, \vdash (c, \sigma, t, \omega)$ and $\ell \vdash \text{out}[i] \downarrow v$ and $E(\ell, i) \not\equiv \bot$ for all $i$ such that $\text{out}[i]$ appears in $g$ (that may affect the evaluation of $g$), then $\ell \vdash \text{out}[i] \downarrow v$.

**Proof.** By induction on $\text{substOutExp}(E, \ell, g)$. The only interesting case is $g = \text{out}[i]$ (where $\text{out}[i]$ may affect the evaluation of the whole output expression). In that case, $\text{substOutExp}(E, \ell, g) = E(\ell, i) \not\equiv \bot$. We have $(t \vdash (E_{\text{out}} \cap E_{\ell}))[i] = \text{out}(\ell, v)$, and so, since $C, \Delta, E, \vdash (c, \sigma, t, \omega)$, we have $\ell \vdash E(\ell, i) \downarrow v$.

Another useful lemma is that the program counter level $pc$ is a lower bound on the side effects of a command.

**Lemma 6.** If $pc, \Gamma; C, \Delta, E, H \vdash c \triangleright \Gamma'; C', \Delta', E', H'$ and 

$$(c, \sigma, t, \omega) \rightarrow (c', \sigma', t', \omega),$$

then for all $\ell$ such that $pc \not\subseteq \ell$, we have

- $t \not\ell = t' \not\ell$ and
- $\text{if } t \not\ell \text{ then } \sigma(x) = \sigma'(x)$.

**Proof.** By induction on $pc, \Gamma; C, \Delta, E, H \vdash c \triangleright \Gamma'; C', \Delta', E', H'$. The interesting cases are $\text{T-ASSIGN}, \text{T-DECLASSIFY}, \text{T-IN},$ and $\text{T-OUT}$. In each of these cases $pc$ is a lower bound on the side effect.

**A.1 Required release**

Having shown several useful lemmas, we are now ready to prove that well-typed programs satisfy required release.

**Lemma 7.** If $pc, \Gamma_0; (0,0), E, \not\subseteq \emptyset \vdash c \triangleright \Gamma; C, \Delta, E, H$

for some $\Gamma_0$ and $pc$, and $\text{substOutExp}(E, \ell, g) \equiv f$, then $c$ satisfies weak required release of input expression $f$ to user $\ell$ using output expression $g$.

**Proof.** Assume $pc, \Gamma_0; (0,0), E, \not\subseteq \emptyset \vdash c \triangleright \Gamma; C, \Delta, E, H$ for some $\Gamma_0$ and $pc$, and $\text{substOutExp}(E, \ell, g) \equiv f$. Let $m = (c, \sigma, (\ell, \omega))$ for some $\sigma$ and $\omega$. Suppose that $m \rightarrow (c', \sigma', t, \omega)$ and $\ell \vdash \text{out}[i] \downarrow v$ for some $v \not\equiv \bot$. We need to show that $\ell \vdash \text{out}[i] \downarrow v$.

By repeated application of Lemma 2, there exists $C', \Delta', E', H', \Gamma'$ and $pc'$ such that $pc' \Gamma' \vdash c' \triangleright \Gamma; C', \Delta', E', H'$.

Suppose that $m \rightarrow (c', \sigma', t, \omega)$. Since $\ell \vdash \text{out}[i] \downarrow v$ for $v \not\equiv \bot$, there have been sufficient output events to channel $\ell$ to evaluate $g$ to a non-$\bot$ value. Since $C', \Delta', E', \vdash (c', \sigma', t, \omega)$, we have $C'(\ell) = (i, \bot)$ and either $j = \bot$ or $j = \ell \vdash (E_{\text{out}} \cap E_{\ell})[i] = \text{out}(\ell, v)$. Either way, by Lemma 4, we have $\text{substOutExp}(E', \ell, g) = \text{substOutExp}(E', \ell, g) \equiv f$, then $E'(\ell, i) \not\equiv \bot$ for all $i$ such that $\text{out}(\ell, i)$ appears in $g$ (that may affect the evaluation of $g$). By Lemma 5, $\ell \vdash \text{out}[i] \downarrow v$. From the definition of $\equiv$, we have $\ell \vdash \text{out}[i] \downarrow v$ as required.

**A.2 Delimited release**

We prove that the type system enforces delimited release user a proof technique based on the technique of Pottier and Simonet [2002] for showing noninterference in the ML programming language. We define a new language, IMPl2, that can represent two executions of a program. We show that type preservation in IMPl2 implies that the program satisfies delimited release. (For convenience, we use IMPl to refer to the interactive imperative language presented in Section 3.)
A.2.1 Syntax and semantics

The language IMP² extends the interactive language with pair constructs for commands \( \langle c_1 \mid c_2 \rangle \), integers \( \langle v_1 \mid v_2 \rangle \), and events \( \langle \alpha_1 \mid \alpha_2 \rangle \). The pair constructs represent different commands, integers, and events that may arise in two different executions of a program. A command pair cannot be nested inside another command pair, but can otherwise appear nested at arbitrary depth. Integer pairs are used to represent different input values that may be provided by different user strategies, and to track how states differ in different executions of a program: user strategies in IMP² are functions from (IMPI) traces to integers and integer pairs, and stores in IMP² are functions from variables to integers and integer pairs. We introduce the special event void, and allow elements of event pairs to range over events and void. The constant void is used to indicate that an event occurred in only one of the two executions. We also allow input and output values to range over integer pairs and integers.

\[
\begin{align*}
(\text{expressions}) & \quad : c::=\ldots\mid \langle v_1 \mid v_2 \rangle \\
(\text{commands}) & \quad : c::=\ldots\mid \langle c_1 \mid c_2 \rangle
\end{align*}
\]

For an extended command \( c \), let the projections \( [c]_1 \) and \( [c]_2 \) represent the two commands that \( c \) encodes. The projection functions satisfy \( \langle [c_1]_1, [c_2]_1 \rangle = [c]_1 \), and are homomorphisms on other commands. Similarly for integer pairs, \( \langle [v_1]_1, [v_2]_1 \rangle = [v]_1 \). The projection functions are extended to states, so that

\[
[s]_i(x) = \begin{cases} v & \text{if } s(x) = n \\ v_i & \text{if } s(x) = \langle v_1 \mid v_2 \rangle \end{cases}
\]

The evaluation of expressions are also extended, so that binary operations \( \oplus \) are homomorphic on integer pairs. Thus, \( \sigma(e) \), the evaluation of expression \( e \) using state \( s \), may be either an integer \( n \) or an integer pair \( \langle v_1 \mid v_2 \rangle \).

We extend projection to event pairs \( \langle [\alpha_1 \mid \alpha_2]_i \rangle = \alpha_i \) and define projection homomorphically on events. We define projection on traces inductively, as follows.

\[
\begin{align*}
\langle \rangle_i & = \langle \rangle \\
\langle \alpha, \alpha_0, \alpha_1, \ldots \rangle_i & = \begin{cases} \langle \alpha_0, \alpha_1, \ldots \rangle_i & \text{if } [\alpha_0]_i = \text{void} \\ \langle \alpha_0 \rangle_i \cdot \alpha \cdot \langle \alpha_1, \alpha_2, \ldots \rangle_i & \text{otherwise} \end{cases}
\end{align*}
\]

Finally, we extend projection to joint strategies, so that for any \( \ell \) and \( t \), \( [\alpha]_i(\ell)(t) = [\omega]_i(\ell)(t) \).

We indicate IMP² configurations with a bullet (•) subscript: \( \langle c, \sigma, t, \omega \rangle \). A IMP² configuration represents a pair of IMPI configurations.

The complete operational semantics of IMP² are shown in Figure 5. Note that rules OS²-ASSIGN, OS²-SEQ-1, OS²-SEQ-2, OS²-IF, OS²-OUT, OS²-IF-1, OS²-IF-2, OS²-DECLASSIFY, and OS²-WHILE are similar to their counterparts in the language IMP. Rules OS²-IF-1 and OS²-DECLASSIFY have been modified to be restricted to apply only to integer results of evaluating expression \( e \).

The rule OS²-PAIR-LIFT evaluates one of the two subcommands of a pair command \( \langle c_1 \mid c_2 \rangle \). The memory and trace are update to indicate that only one of the two projections represented by the configuration made progress. Thus, \( \sigma_j = \sigma'_j \) and \( t_j = t'_j \), where \( j \in \{1, 2\} \) is the execution that did not make progress. Note that the small step relation used in the premise is small step relation of language IMP.

The rule OS²-PAIR-SKIP applies when the two commands represented by a command pair both finished executing. This rule removes the command pair.

The rule OS²-PAIR-IF applies when the evaluation of a conditional expression differs in the two executions represented by the IMP² configuration. This rule introduces a command pair, representing the different branches that may be taken by the two executions. This is the only rule that introduces command pairs.

The rule OS²-PAIR-DECLASSIFY applies when an expression is declasified in both executions represented by the evaluation, and the evaluation of the expression is the same in both executions.

A.2.2 Adequacy

The language IMP² is adequate to express the execution of two IMPI programs. We show that the execution of a IMP² program is sound (a step taken by a IMP² program corresponds to one or zero steps taken by its projections) and complete (given two IMPI executions, there is a IMP² execution whose projection agrees with at least one of them). We write \( \rightarrow^* \) for the reflexive closure of \( \rightarrow \).

Lemma 8 (Soundness). If \( \langle c, \sigma, t, \omega \rangle \rightarrow \langle c', \sigma', t', \omega \rangle \), then \( \langle [c]_i, [\sigma]_i, [t]_i, [\omega]_i \rangle \rightarrow^* \langle [c']_i, [\sigma']_i, [t']_i, [\omega]_i \rangle \) for \( i \in \{1, 2\} \).

Proof. By induction on the derivation \( \langle c, \sigma, t, \omega \rangle \rightarrow \langle c', \sigma', t', \omega \rangle \). The interesting cases are the new rules introduced for IMP²: OS²-PAIR-LIFT, OS²-PAIR-SKIP, OS²-PAIR-IF, and OS²-PAIR-DECLASSIFY. For a reduction using OS²-PAIR-LIFT, clearly one of the two projections takes a step, while the other projection remains unchanged. For OS²-PAIR-SKIP, both projections remain unchanged. For both OS²-PAIR-IF, and OS²-PAIR-DECLASSIFY, both projections take a step. \( \square \)

If an IMP² configuration is stuck, it is because one of the two projections is stuck.

Lemma 9 (Stuck configurations). If \( \langle c, \sigma, t, \omega \rangle \) is stuck (i.e., cannot be reduced and \( c \neq \text{skip} \)), then \( \langle [c]_i, [\sigma]_i, [t]_i, [\omega]_i \rangle \) is stuck for some \( i \in \{1, 2\} \).

Proof. By structural induction on command \( c \). \( \square \)

Given two IMPI evaluations, there is an IMP² evaluation that represents the same IMPI evaluation for at least one of the two evaluation.

Lemma 10 (Completeness). If \( \langle [c]_i, [\sigma]_i, [t]_i, [\omega]_i \rangle \rightarrow^* \langle c'_i, \sigma'_i, t'_i, \omega'_i \rangle \) for \( i \in \{1, 2\} \), then there exists an IMP² configuration \( \langle c', \sigma', t', \omega \rangle \) and \( \langle [c'_i]_i, [\sigma'_i]_i, [t'_i]_i, [\omega]_i \rangle = \langle c'_i, \sigma'_i, t'_i, [\omega]_i \rangle \) for some \( i \in \{1, 2\} \).
Proof. Let \( \tau_i = ([c]_i, [\sigma]_i, [t]_i, [\omega]_i) \ldots ([c']_i, [\sigma']_i, [t']_i, [\omega]_i) \) be the sequence of configurations that witnesses \( ([c]_i, [\sigma]_i, [t]_i, [\omega]_i) \rightarrow^* (\{c'_i, \sigma'_i, t'_i, \omega'_i\}) \).

Let \( n_i \) be the length of \( \tau_i \). For a sequence IMPI\(^2\) configurations \( \tau = (c, \sigma, t, \omega)_* \ldots (c', \sigma', t', \omega)_* \) that witnesses \( (c, \sigma, t, \omega)_* \rightarrow^* (c', \sigma', t', \omega)_* \), let \( f_i(\tau) \) be \( n_i \) minus the number of reduction steps in \( \tau \) that reduce the \( i \)th projection. Note that \( f_i(\tau) \) is non-negative. Consider \( g(\tau) = \min(f_1(\tau), f_2(\tau)) \). If \( g(\tau) = 0 \), then \( \tau \) is a sequence that satisfies the conditions.

Suppose \( g(\tau) > 0 \). Consider the function

\[
h(\tau) = \left(g(\tau), \left| f_1(\tau) - f_2(\tau)\right|, \text{numPairs}(\tau[[|\tau| - 1]])\right)
\]

where \( |\tau| - 1 \) refers to the last configuration in the sequence \( \tau \), and \( \text{numPairs}((c, \sigma, t, \omega)_*) \) returns the number of pair commands in \( c \). Note that all elements of the triple returned by \( h(\tau) \) are non-negative. If we can extend \( \tau \) by one step to a sequence \( \tau' \) such that \( h(\tau') < h(\tau) \) under lexicographic ordering, then, by repeated applications, eventually we will produce a sequence \( \tau'' \) such that \( g(\tau'') = 0 \).

We now show how to extend sequence \( \tau \) to a sequence \( \tau' \) such that \( h(\tau') < h(\tau) \). By assumption, \( g(\tau) > 0 \), so neither last configuration of \( \tau_1 \) or \( \tau_2 \) is stuck. By Lemma 9, we can extend \( \tau \) by one more step, producing trace \( \tau' \). By Lemma 8, either \( f_i(\tau') = f_i(\tau) - 1 \) for some \( i \in \{1, 2\} \), or \( f_i(\tau') = f_i(\tau) \) for all \( i \in \{1, 2\} \). If the former, then \( h(\tau') < h(\tau) \). If the latter, then the rule OS\(^2\)-PAIR-SKIP was used in the reduction, and the last configuration of \( \tau' \) has one fewer pair commands than the last configuration of \( \tau \), and so \( h(\tau') < h(\tau) \).

\[ \square \]
A.2.3 Type preservation

We extend the type system to IMPL commands and configurations. Typing judgments are now of the form

\[ \text{pc}, \Gamma; C, \Delta, E, H \vdash_\ell c \triangleright \Gamma'; C', \Delta', E', H' \]

where \( \ell \in \mathcal{L} \). Intuitively, if IMPL command \( c \) is well-typed with typing parameter \( \ell \), then \( c \) represents the two IMPL commands that are indistinguishable from the point of view of any user \( \ell' \) such that \( \ell' \supseteq \ell \).

Because our type system is flow-dependent, we need to extend the typing context entities \( \Gamma, \Delta, \) and \( E \) so that they range over pairs. This allows the expression of different typing contexts for the two IMPL commands represented by a single IMPL command. Thus, \( \Gamma \) ranges over security levels \( \mathcal{L} \) and pairs of security levels (written \( \langle \ell_1 | \ell_2 \rangle \)). Similarly, \( \Delta \) and \( E \) range over elements of \( \text{InputExp}_\ell \) and pairs of elements of \( \text{InputExp}_\ell \). We also extend the entity \( C \) so that its range is \( (\mathcal{L} \cup (\mathcal{L}_\perp \times \mathcal{L}_\perp)) \times (\mathcal{L}_\perp \cup (\mathcal{L}_\perp \times \mathcal{L}_\perp)) \). That is, for any \( \ell \in \mathcal{L} \), we have \( C(\ell)(i, j) \), where \( i \) is either an integer (indicating the number of inputs received on channel \( \ell \)), \( \perp \) (indicating an unknown number of inputs received on channel \( \ell \)), or a pair \((i_1, i_2)\), where \( i_1 \) indicates inputs received on channel \( \ell \) in the first execution, and \( i_2 \) indicates inputs received on channel \( \ell \) in the second execution. Similarly, \( j \) describes the outputs sent on channel \( \ell \) in both executions. We define projection operations \([\cdot]_1\) and \([\cdot]_2\) for all the extended entities. We extend relations \( \subseteq \) and \( \geq \) such that

\[ \ell \subseteq \langle \ell_1 | \ell_2 \rangle \iff \ell \subseteq \ell_1 \text{ and } \ell \subseteq \ell_2 \]

\[ \langle \ell_1 | \ell_2 \rangle \subseteq \ell \iff \ell_1 \subseteq \ell \text{ and } \ell_2 \subseteq \ell \]

\[ \langle \ell_1 | \ell_2 \rangle \subseteq \langle \ell'_1 | \ell'_2 \rangle \iff \ell_1 \subseteq \ell'_1 \text{ and } \ell_2 \subseteq \ell'_2 \]

\[ v \geq \langle v_1 | v_2 \rangle \iff v \geq v_1 \text{ and } v \geq v_2 \]

\[ \langle v_1 | v_2 \rangle \geq v \iff v_1 \geq v \text{ and } v_2 \geq v \]

Typing rules for IMPL (given in Figure 5) are made into typing rules for IMPL by adding the typing parameter \( \ell \) to every rule. In addition, we severely restrict when the typing context entities may differ for the two different IMPL commands represented by a single IMPL command. We require for judgment \( \text{pc}, \Gamma; C, \Delta, E, H \vdash_\ell c \triangleright \Gamma'; C', \Delta', E', H' \) that (a) the image of \( C', \Delta', E' \) and \( \Gamma' \) do not contain any pairs (i.e., they are suitable IMPL entities); and (b) if \( c \) does not contain a command pair, then \( C, \Delta, E \) and \( \Gamma \) do not contain any pairs. We use the predicate \( \text{noPairs}(\Gamma, C, \Delta, E) \) to indicate that the image of \( C, \Delta, E \) and \( \Gamma \) do not contain any pairs.

All typing rules for IMPL are presented in Figure 6. The typing rule for the new pair command, \( T^2\text{-PAIR} \), requires that both projections type check using IMPL typing rules, for a program counter level \( \text{pc} \) that is at least as restrictive as typing parameter \( \ell \). Intuitively, this will ensure that any side-effects of a command pair will not be observable at level \( \ell \) or below. Note that the premise of \( T^2\text{-PAIR} \) uses the typing judgment for IMPL, i.e., without the typing parameter \( \ell \). This is because well-formed commands do not have nested command pairs. All typing rules other than \( T^2\text{-PAIR} \) correspond closely to their IMPL counterpart.

We define a notion of satisfaction for IMPL configurations. Intuitively, an IMPL configuration satisfies context \( C, \Delta, E, \Gamma \) for \( \ell \) if the two IMPL configurations represented by the IMPL configuration are identical at all levels \( \ell' \), such that \( \ell' \subseteq \ell \), and each IMPL configuration satisfies the appropriate IMPL context. We also require that no command pair appears as a subcommand of an if or while command.

**Definition 7.** We say that configuration \( (c, \sigma, t, \omega)* \), satisfies context \( C, \Delta, E, \Gamma \) for \( \ell \), written \( \Gamma, C, \Delta, E \models_\ell (c, \sigma, t, \omega)* \) if all of the following conditions hold.

1. For all \( x \in \text{Var} \), if \( \sigma(x) \) is a pair value then \( \Gamma(x) \not\supseteq \ell \).
2. For all \( i \) such that \( 0 \leq i < |t| \), if \( \text{value}(t(i)) \) is a pair value, then \( \text{level}(t(i)) \not\supseteq \ell \), where
   \[ \text{value}([\alpha_1 | \alpha_2]) = [0 | 0] \]
   \[ \text{value}(\text{in}(\ell, v)) = v \]
   \[ \text{value}(\text{out}(\ell, v)) = v \]
   \[ \text{level}([\text{void} | \alpha]) = \text{level}(\alpha) \]
   \[ \text{level}([\alpha | \text{void}]) = \text{level}(\alpha) \]

3. For all \( \ell', t' \), if \( \omega(\ell')(t') \) is a pair value, then \( \ell' \not\supseteq \ell \).
4. \( |C[i]|, |\Delta[i]|, |E[i]|, |\ell[i]|, |\sigma[i]|, |t[i]|, |\omega[i]| \) for \( i \in \{1, 2\} \).
5. No command pair appears as a subcommand of an if or while command of \( c \).

If a configuration satisfies context \( C, \Delta, E, \Gamma \) for \( \ell \), then it also satisfies a weaker, or less precise, context \( C', \Delta', E', \Gamma' \).

**Lemma 11.** If \( \Gamma, C, \Delta, E \vdash_\ell (c, \sigma, t, \omega)* \), and \( C \triangleright C' \) and \( \Delta \triangleright \Delta' \) and \( E \triangleright E' \) and \( \Gamma \triangleright \Gamma' \) then \( \Gamma', C', \Delta', E' \vdash_\ell (c, \sigma, t, \omega)* \).

**Proof.** If \( \Gamma(x) \) is a pair value, then \( \Gamma(x) \not\supseteq \ell \), and \( \Gamma(x) \not\supseteq \Gamma'(x) \). If \( \Gamma'(x) \not\supseteq \ell \), then \( \Gamma(x) \not\supseteq \ell \), a contradiction.

We have \( |C[i]|, |\Delta[i]|, |E[i]|, |\ell[i]|, |\sigma[i]|, |t[i]|, |\omega[i]| \) for \( i \in \{1, 2\} \) from Lemma 1.

The remaining requirements do not depend on the typing context.

If IMPL command \( c \) is well-typed, then both projections of \( c \) are well-typed (in the IMPL type system).
Lemma 12. If

\[ pc, \Gamma; C, \Delta, E, H \vdash_{\ell} e \vdash_{t} c \triangleright \Gamma'; C', \Delta', E', H' \]

then for \( i \in \{1, 2\} \) we have

\[ pc, [\Gamma_i]; [C_i]; [\Delta_i]; [E_i], H \vdash [c_i] \triangleright \Gamma'; C', \Delta', E', H' \]

Proof. By induction on the judgment \( pc, \Gamma; C, \Delta, E, H \vdash_{\ell} e \triangleright \Gamma'; C', \Delta', E', H' \), and the fact that IMPL\(^2\) typing rules are at least as restrictive as their IMPL counterparts. \( \square \)

The execution of a IMPL\(^2\) program preserves typings. The following type-preservation theorem is key to showing that well-typed IMPL programs satisfy delimited release.

**Theorem 2** (Type preservation). Let \( c \) be an IMPL\(^2\) command such that

\[ pc, \Gamma; C, \Delta, E, H \vdash_{\ell} e \triangleright \Gamma'; C', \Delta', E', H' \]

and let \( (c, \sigma, t, \omega)_* \) be an IMPL\(^2\) configuration such that

\[ \Gamma, C, \Delta, E \Rightarrow_{\ell} (c, \sigma, t, \omega)_* \]

If

\[ (c', \sigma', t', \omega)_* \Rightarrow (c', \sigma', t', \omega)_* \]

then

\[ pc, \Gamma; C, \Delta, E, H \vdash_{\ell} c \triangleright \Gamma'; C', \Delta', E', H' \]

**Figure 6.** Typing rules
and \( [t'] \) agree up to \( \ell \) on escape hatches \( H' \) then there exists \( C'', \Delta'', E'', H'', \Gamma'' \), and \( \rho'' \) such that \( \rho''(C'', \Delta'', E'', H'', \Gamma'') \models \ell \) \( C', \Delta, E', \Gamma \), and \( \rho'' \) that

\[
\Gamma'' \models C'', \Delta'', E'', H'' \models \ell \rightarrow \Gamma', C', \Delta', E', \Gamma'.
\]

Proof. By induction on the typing judgment \( \rho(C; \Gamma; \Delta, E, H \vdash_e c \triangleright \Gamma'; C'; \Delta', E', H') \).

First, note that

\[
\{(C', \Delta, E, H \vdash_e c \triangleright \Gamma'; C'; \Delta', E', H')\} \models \ldots
\]

for \( i \in \{1, 2\} \) by Lemmas 12 and 2 (since the witnesses chosen in the proof of Lemma 2 are the same as the witnesses chosen in this proof).

- **T1** - **SKIP**

  Trivial, as there is no \((c', \sigma', \ell', \omega)\) such that \((\text{skip}, \sigma, t, \omega) \rightarrow (c', \sigma', \ell', \omega)\).

- **T1** - **ASSIGN**

  In this case, \( \sigma' = \sigma[x \mapsto \sigma(v)] \), \( \ell' = t \), \( C' = C \), \( \Delta' = \Delta[x \mapsto \Delta(\ell)] \), \( E' = E \), \( H' = H \), \( \Gamma' = \Gamma[x \mapsto \ell] \).

  Let \( \rho''(C'', \Delta'', E'', H'', \Gamma'') \models \ell \rightarrow \Gamma', C', \Delta', E', \Gamma' \).

  Then \( \rho(C'; \Gamma; \Delta', E', H') \models \ell \rightarrow \Gamma'; C', \Delta', E', H' \).

  Since \( \Gamma, \Delta, E \vdash_e \ell \), we have for all \( y \in \text{Var} \) such that \( y \neq x \), if \( \sigma(y) \) is an integer pair, then \( \Gamma'(y) = \Gamma(y) \not\models \ell \). Also, if \( \sigma(x) \) is an integer pair, then \( \sigma(e) \) is an integer pair, and so there is some variable \( z \in \text{Var} \) such that \( \Gamma(z) \not\models \ell \). By **T2** - **ASSIGN**, we have \( \Gamma(z) \not\models \Gamma(x) \), and so \( \Gamma'(x) \not\models \ell \).

  Moreover, since \( \rho'' \models \Delta'' = C'' \) and \( E'' = E \) and \( \ell' = t \) we have \( \Gamma', C', \Delta', E' \models \ell \).

- **T1** - **SEQ**

  In this case, \( c = c_0; c_1 \). If \( c_0 \neq \text{skip} \) then the result follows trivially from the inductive hypothesis. If \( c_0 = \text{skip} \) then \( c' = c_1 \) and \( \sigma' = \sigma \) and \( \ell' = t \), and the result follows easily.

- **T1** - **OUT**

  In this case, \( c = \text{out}(e) \). \( \sigma' = \sigma \), \( \ell' = \text{out}(e), \sigma(e) \), \( C'' = \text{inc}_{\text{out}}(C, \ell), \Delta'' = \Delta, E'' = \text{rec}_{\text{out}}(E, C, \ell, \Delta(e)) \).

  Then \( \rho(C'; \Gamma; \Delta', E', H') \models \ell \rightarrow \Gamma'; C', \Delta', E', H' \).

  If \( \sigma(e) \) is a pair value, then \( v \) is a pair value, and so, by \( \Gamma, \Delta, E \models_e (c, \sigma, t, \omega) \), we must have \( \ell' \not\models \ell \). Similarly, if \( \text{value}(\text{out}(e), v) \) is a pair value, then \( \Gamma' \not\models \ell \). Thus, we have \( \Gamma', C', \Delta', E' \models_e (c', \sigma', \ell', \omega) \).

- **T1** - **IF**

  In this case, \( c = \text{if} e \text{ then } c_0 \text{ else } c_1 \). Suppose that \( \sigma(e) \) is not a pair value, i.e., \( \sigma(e) \subseteq \mathbb{Z} \). Then \( c' = c_0, \sigma = \sigma, \ell' = t, C' = C, \Delta' = \Delta, E' = E, H' = H, \Gamma' = \Gamma \).

  Let \( \rho''(C'', \Delta'', E'', H'', \Gamma'') \models \ell \rightarrow \Gamma', C'', \Delta'', E'', H'', \Gamma'' \).

  Then \( \rho(C'; \Gamma; \Delta', E', H') \models \ell \rightarrow \Gamma'; C', \Delta', E', H' \).

  Otherwise, suppose \( \sigma(e) = \text{pair} \). Then \( c' = c_0, c_1 \), \( \sigma = \sigma, \ell' = t, C' = C, \Delta' = \Delta, E' = E, H' = H, \Gamma' = \Gamma \).

  Let \( \rho''(C'', \Delta'', E'', H'', \Gamma'') \models \ell \rightarrow \Gamma', C'', \Delta'', E'', H'', \Gamma'' \).

  By Lemma 12 and **T2** - *IF* we have

  \[
  \Gamma(\rho''(C'', \Delta'', E'', H'', \Gamma'')) \subseteq \rho''(C', \Delta', E', H', \Gamma') \text{ for } i \in \{1, 2\} \text{ and some } \rho'' \text{ such that } \Gamma'C', \Delta', E', H', \Gamma' \text{ as required.}
  \]

- **T1** - **WHILE**

  In this case, \( c = e; c_0; c_1 \) if \( e \text{ then } e \text{ while } c_0 \text{ else } c_1 \) if \( e \) then \( c_0 \) while \( e \text{ do } c_1 \) else skip. \( \sigma = \sigma, \ell' = \ell, C' = C, \Delta' = \Delta, E' = E, H' = H, \Gamma' = \Gamma \).

  Let \( \rho''(C'', \Delta'', E'', H'', \Gamma'') \models \ell \rightarrow \Gamma', C'', \Delta'', E'', H'', \Gamma'' \).

  We can easily construct a judgment \( \rho(C'); \Gamma; \Delta, E, H \models \ell \rightarrow \Gamma'; C', \Delta', E', H', \Gamma' \).

  Note that \( \text{noPairs}(\Gamma, \Delta), \Delta(e) \) is not a pair of input expressions. By assumption, \( \{t'\}_1 \) and \( \{t'\}_2 \) agree up to \( \ell \) on escape hatches \( H' \). If \( \ell' \not\models \ell \), since \( \{\Delta(e), t'\}_1 \in H' \) then \( \{t'\}_1 \models \Delta(e) \not\models \ell \) and \( \{t'\}_2 \models \Delta(e) \not\models \ell \) for some integer \( v \). Since \( C, \Delta, E \vdash_e (c, \sigma, t, \omega) \), we have \( \sigma(e) = v \), and so \( \sigma'(x) = v \).}

19

2010/4/28
then for all \( \ell \) we have 
\[
\text{pc}, \Gamma; C, \Delta, E, H \vdash \tau (c', \sigma', t, \omega),
\]
then for all \( \ell \) we have 
\[
\text{pc}, \Gamma; C, \Delta, E, H \vdash \tau (c', \sigma', t, \omega),
\]

\[\]
Theorem 2, we have $\Gamma', C', \Delta', E' \models (c', \sigma', t, \omega)$ for some $C', \Delta', E'$, and $\Gamma'$. Thus, by Lemma 13, we have $|t|_i | \ell = |t|_j | \ell$, and so $t_j | \ell \succeq t_i | \ell$ and thus $t_1 \approx t_2$ as required.

Let one or both of $t_1$ or $t_2$ be an infinite trace. Suppose that it is not the case that either $t_1 | \ell \succeq t_2 | \ell$ or $t_2 | \ell \succeq t_1 | \ell$. Therefore there is some index $n$ such that $(t_1 | \ell)(n) \neq (t_2 | \ell)(n)$. Consider finite traces $t'_1$ and $t'_2$ such that $t_1 \succeq t'_1$ and $t_2 \succeq t'_2$, and $|t'_1 | \ell| = |t'_2 | \ell| = n+1$. Note that $(c, \sigma, \langle \rangle, \omega_1) \rightarrow t'_1$ and $(c, \sigma, \langle \rangle, \omega_2) \rightarrow t'_2$. By a similar argument above, we derive that $t'_1 \approx t'_2$. But this implies that $(t'_1 | \ell)(n) = (t'_2 | \ell)(n)$, a contradiction! Therefore, either $t_1 | \ell \succeq t_2 | \ell$ or $t_2 | \ell \succeq t_1 | \ell$, and so $t_1 \approx t_2$ as required.

A.3 Bounded release

The proof of Theorem 1 follows immediately from Lemmas 7 and 15.