Enforcing More with Less: Formalizing Target-aware Run-time Monitors

Yannis Mallios, Lujo Bauer, Dilsun Kaynar, and Jay Ligatti

May 3, 2012
(revised September 1, 2013)

CMU-CyLab-12-009

CyLab
Carnegie Mellon University
Pittsburgh, PA 15213

Enforcing More with Less: Formalizing Target-aware Run-time Monitors

Yannis Mallios[1], Lujo Bauer[1], Dilsun Kaynar[1], and Jay Ligatti[2]

[1] Carnegie Mellon University, Pittsburgh, USA
{mallios,lbauer,dilsunk}@cmu.edu
[2] University of South Florida, Tampa, USA
ligatti@cse.usf.edu

Abstract. Run-time monitors ensure that untrusted software and system behavior adhere to a security policy. This paper defines an expressive formal framework, based on I/O automata, for modeling systems, policies, and run-time monitors in more detail than is typical. We explicitly model, for example, the environment, applications, and the interaction between them and monitors. The fidelity afforded by this framework allows us to study and explicitly formulate practical constraints on policy enforcement that were often only implicit in previous models, providing a more accurate view of what can be enforced by monitoring in practice. Moreover, we introduce two definitions of enforcement, target-specific and generalized, that allow us to reason about practical monitoring scenarios. Finally, we provide some meta-theoretical comparison of these definitions and we apply them to investigate policy enforcement in scenarios where the monitor designer has knowledge of the target application and show how this can be exploited for making more efficient design choices.

1 Introduction

Today's computing climate is characterized by increasingly complex software systems and networks, and inventive and determined attackers. Hence, one of the major thrusts in the software industry and in computer security research has become to devise ways to *provably guarantee* that software does not behave in dangerous ways or, barring that, that such misbehavior is contained and mitigated. Example guarantees could be that programs: only access memory that has been allocated to them (memory safety); only jump to and execute valid code (control-flow integrity); use no more than 10 MB of storage and 10 KB/sec network bandwidth for grid use (resource allocation); and never send secret data over the network (a type of information flow).

A common mechanism for enforcing security policies on untrusted software is run-time monitoring. Run-time monitors observe the execution of untrusted applications or systems and ensure that their behavior adheres to a security policy. This type of enforcement mechanism is pervasive, and can be seen in operating systems, web browsers, firewalls, intrusion detection systems, etc. A common

specific example of monitoring is system-call interposition (e.g., [27, 12]). Here, given an untrusted application and a set of security-relevant system calls, a monitor intercepts calls made by the application to the kernel, and enforces a security policy by taking remedial action when a call violates the policy. This idea is depicted in Fig. 1. In practice, there are several instantiations of monitors for this general concept. Understanding and formally reasoning about specific instantiations is as important as understanding the general concept, since in enables us to reason about important details that might be lost at a higher level of abstraction. Two dimensions along which instantiations can differ are: (1) *the monitored interface*: monitors can mediate different parts of the communication between the application and the kernel; e.g., an input sanitization monitor will mediate only inputs to the kernel (dashed lines in Fig. 1); and (2) *trace modification capabilities*: monitors may have a variety of enforcement capabilities, from being restricted to just terminating the application (e.g., when the application tries to write to the password file), to being able to perform additional remedial actions (e.g., suppress a write system call and log the attempt)[3].

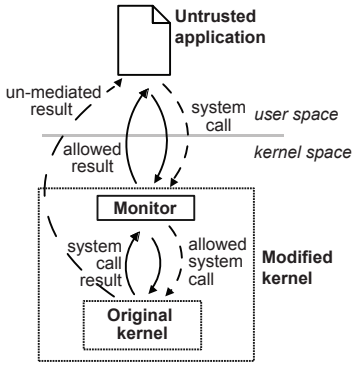


Fig. 1: System-call interposition: dashed line shows an input-mediating monitor; solid line an input/output-mediating monitor.

Despite the ubiquity of run-time monitors, their use has far outpaced theoretical work that makes it possible to formally and rigorously reason about monitors and the policies they enforce. Such theoretical work is necessary, however, if we are to have confidence that enforcement mechanisms are successfully carrying out their intended functions.

Several proposed formal models (e.g., [25, 19]) make progress towards this goal. They use formal frameworks to model monitors and their enforcement capabilities (e.g., whether the monitors can insert arbitrary actions into the stream of actions that the target wants to execute). These frameworks have

[3] In this paper we do not consider mechanisms that modify traces that arbitrarily modify the target application, such as by rewriting.

been used to analyze and characterize the policies that are enforceable by the various types of monitors.

However, such models typically do not capture many details of the monitoring process, including the monitored interface, leaving us with practical scenarios that we cannot reason about in detail. In our system-call interposition scenario, for example, without the ability to express the communication between the untrusted application, the monitor, and the kernel in a model, it might not be possible to differentiate between and compare monitors that can mediate all security-relevant communication between the application and the kernel (solid lines in Fig. 1) from monitors that can mediate only some of it (dashed lines in Fig. 1).

Some recent models (e.g., [20, 13]) make progress towards such more detailed reasoning by including in the model bi-directional communication between the monitor and its environment (e.g., application and kernel), but they do not explicitly reason about the application or system being monitored. In practice, however, monitors can enforce policies beyond their operational enforcement capabilities by exploiting knowledge about the component that they are monitoring. For example, a policy that requires that every file that is opened must be eventually closed cannot, in general, be enforced by any monitor, because the monitor does not know what the untrusted application will do in the future, and thus such a policy is outside its enforcement capabilities. However, if the monitored application always closes files that it opens, then this policy is no longer unenforceable for that particular application. Such distinctions are often relevant in practice—e.g., when implementing a patch for a specific type or version of an application—and, thus, there is a need for formal frameworks that will aid in making informed and provably correct design and implementation decisions.

In this paper, we propose a general framework, based on I/O automata, for more detailed reasoning about policies, monitoring, and enforcement. The I/O automaton model [22, 21] is a labeled transition model for asynchronous concurrent systems. Thus, we are using an automata-based formalism, similarly to many previous models of run-time enforcement mechanisms, with enough expressive power to model asynchronous systems (e.g., the communication between the application, the monitor, and the kernel). Our framework provides abstractions for reasoning about many practical details important for run-time enforcement, and, in general, yields a richer view of monitors and applications than is typical in previous analyses of run-time monitoring. For example, our framework supports modeling practical systems with security-relevant actions that the monitor cannot mediate, rather than assuming complete mediation [16, 5]. (We discuss more such examples in §3.)

We make the following specific contributions:

- We show how I/O automata can be used to faithfully model target applications, monitors, and the environments in which monitored targets operate, as well as various types of monitors and monitoring architectures (§3).
- We extend previous definitions of security policies and enforcement to support more fine-grained formal reasoning of policy enforcement (§4).

– We show that this more detailed model of monitoring forces explicit reasoning about concerns that are important for designing run-time monitors in practice, but that previous models often reasoned about only informally (§5.2). We formalize these results as a set of lower bounds on the policies enforceable by any monitor in our framework.
– We demonstrate how to use our framework to exploit knowledge about the target application to make design and implementation choices that may lead to more efficient enforcement (§6). For example, we exhibit constraints under which monitors with different monitoring interfaces (i.e., one can mediate more actions than the other) can enforce the same class of policies.

Roadmap We start by briefly reviewing I/O automata (§2). We then informally show how to model monitors and targets in our framework and discuss some of the benefits of this approach (§3). Next, we formally define policies and enforcement (§4). Then, we show several examples of the meta-theoretical analysis that our framework enables by (a) providing some lower bounds for enforceable policies (§5), and (b) exposing constraints under which seemingly different monitoring architectures can enforce the same classes of policies (§6).

2 I/O Automata

I/O automata are a labeled transition model for asynchronous concurrent systems [22, 21]. In this section we review aspects of I/O automata that we build on in the rest of the paper. We encourage readers familiar with I/O automata to skip to §3.

Given a function $f : A \to B$ we write $\mathsf{dom}(f)$ for A (i.e., the domain of f) and $\mathsf{range}(f)$ for B (i.e., the range of f).

I/O automata are typically used to describe the behavior of a system interacting with its environment. The interface between an automaton A and its environment is described by the *action signature* $sig(A)$ of A. The signature $sig(A)$ is a triple of disjoint sets—$input(A)$, $output(A)$, and $internal(A)$. We write $acts(A)$ for $input(A) \cup output(A) \cup internal(A)$. We sometimes refer to output and internal actions as *locally-controlled* actions.

Formally, an I/O automaton A consists of:

1. an action signature, $sig(A)$;
2. a (possibly infinite) set of *states*, $states(A)$;
3. a nonempty set of *start states*, $start(A) \subseteq states(A)$;
4. a transition relation, $trans(A) \subseteq states(A) \times acts(A) \times states(A)$, with the property that for every state q and input action a there is a transition $(q, a, q') \in trans(A)$; and
5. an equivalence relation $Tasks(A)$ partitioning the set $Local(A)$ into at most a countable number of equivalence classes.

If A has a transition (q, a, q') then we say that action a is *enabled* in state q. When only input actions are enabled in q, then q is called a *quiescent* state.

The set of all quiescent states of an automaton A is denoted by $quiescent(A)$. A task C is enabled in a state q if some action in C is enabled in q.

An *execution* e of A is a finite sequence, $q_0, a_1, q_1, \ldots, a_r, q_r$, or an infinite sequence $q_0, a_1, q_1, \ldots, a_r, q_r, \ldots$, of alternating states and actions such that $(q_k, a_{k+1}, q_{k+1}) \in trans(A)$ for $k \geq 0$, and $q_0 \in start(A)$. A *schedule* is an execution without states in the sequence, and a *trace* is a schedule that consists only of input and output actions. An *execution*, *trace*, or *schedule module* describes the behavior exhibited by an automaton. An execution module E consists of a set $states(E)$, an action signature $sig(E)$, and a set $execs(E)$ of executions. Schedule and trace modules are similar, but do not include states. The sets of executions, schedules, and traces of an I/O automaton (or module) X are denoted by $execs(X)$, $scheds(X)$, and $traces(X)$.Given a sequence s and a set X, $s|X$ denotes the sequence resulting from removing from s all elements that do not belong in X. Similarly, for a set of sequences S, $S|X = \{(s|X) \mid s \in S\}$. The symbol ϵ denotes the empty sequence. We write $\sigma_1; \sigma_2$ for the concatenation of two schedules or traces, the first of which has finite length. When σ_1 is a finite prefix of σ_2, we write $\sigma_1 \preceq \sigma_2$, and, if a strict finite prefix, $\sigma_1 \prec \sigma_2$. $\Sigma^\star$ denotes the set of finite sequences of actions and Σ^ω the set of infinite sequences of actions. The set of all finite and infinite sequences of actions is $\Sigma^\infty = \Sigma^\star \cup \Sigma^\omega$.

The I/O automata definition also includes the equivalence relation $Tasks(A)$. This is used in the definition of *fairness*, which essentially says that the automaton will give fair turns to each of its tasks while executing.

An execution e of an I/O automaton A is said to be *fair* if for each class C of $Tasks(A)$: (1) if e is finite, then C is not enabled in the final state of e, or (2) if e is infinite, then e contains either infinitely many events from C or infinitely many occurrences of states in which C is not enabled.

Fairness abstracts the need for modeling a scheduler in the system. Specifically, when reasoning about practical systems, instead of explicitly modeling a scheduler one can simply reason about a fair version of the system. The type of fairness that I/O automata define is called "weak fairness, and is only one of the many different types of fairness [17].

Given an automaton or a module A we denote the sets of fair executions, schedules and traces by $fairexecs(A)$, $fairscheds(A)$ and $fairtraces(A)$.

An automaton that models a complex system can be constructed by *composing* automata that model the system's components. When composing automata S_i, where $i \in I$, or modules, their signatures are called *compatible* if their output actions are disjoint and the internal actions of each automaton are disjoint with all actions of the other automata. More formally, The actions signatures $S_i : i \in I$ or called compatible if for all $i, j \in I$:

1. $output(S_i) \cap output(S_j) = \emptyset$
2. $internal(S_i) \cap acts(S_j) = \emptyset$

When the signatures are compatible we say that the corresponding automata and modules are compatible too. The composition $A = \prod_{i \in I} A_i$ of a set of compatible automata $\{A_i : i \in I\}$ is defined as:

1. $states(A) = \prod_{i \in I} states(A_i)$
2. $start(A) = \prod_{i \in I} start(A_i)$,
3. $sig(A) = \prod_{i \in I} sig(A_i) =$
   $\Big($ $output(A) = \cup_{i \in I} output(A_i)$,
   $internal(A) = \cup_{i \in I} internal(A_i)$,
   $input(A) = \cup_{i \in I} input(A_i) - \cup_{j \in I} output(A_j)\Big)$,
4. $trans(A)$ is equal to the set of triples (q, a, q') such that for all $i \in I$
   (a) if $a \in acts(A_i)$ then $(q_i, a, q_i') \in trans(A_i)$, and
   (b) if $a \notin acts(A_i)$ then $q_i = q_i'$
5. $Tasks(A) = \cup_{i \in I} Tasks(A_i)$

Similarly to the composition of automata, we can define the composition of execution, schedule, and trace modules. Given a countable collection of compatible schedule[4] (or trace) modules $\{S_i, i \in I\}$ we define the composed execution module $S = \prod_{i \in I} S_i$:

1. $sig(S) = \prod_{i \in I} sig(S_i)$,
2. $execs(S)$ is the set of schedules s such that the subsequence s' of s consisting of actions of S_i, is a schedule of S_i for every $i \in I$.

Unlike in models such as CCS [24], composing two automata that share some actions (i.e., outputs of one automaton may be inputs to the other) causes those actions to be regarded as output actions of the composition. Those that are required to be internal need to be explicitly classified as such using the *hiding* operation. If S is a signature and $\Phi \subseteq output(S)$, then $\mathsf{hide}_\Phi(S)$ is defined to be the new signature S', where $input(S') = input(S)$, $output(S') = output(S) - \Phi$, and $internal(S') = internal(S) \cup \Phi$. Given an I/O automaton A and $\Phi \subseteq output(A)$, $\mathsf{hide}_\Phi(A)$ is the automaton A' obtained by replacing $sig(A)$ with $sig(A') = \mathsf{hide}_\Phi(sig(A))$.

The operation of *renaming*, on the other hand, changes the names of actions, but not their types. An *action mapping* (or renaming) f is a total injective mapping between sets of actions. Such a renaming is said to be *applicable* to an automaton if the domain of f contains the actions of the automaton. If the renaming f is applicable to an automaton A, then the automaton $\mathsf{rename}(A)$ is the automaton with the states and start states of A; with the input, output and internal actions $\mathsf{rename}(input(A))$, $\mathsf{rename}(output(A))$, $\mathsf{rename}(internal(A))$ respectively; with the transition relation $\{(q, \mathsf{rename}(a), q') : (q, a, q') \in trans(A)\}$; and the equivalence relation $\{(\mathsf{rename}(a), \mathsf{rename}(a')) : (a, a') \in tasksA\}$. We overload a renaming function f to be applicable to schedules as follows: given a schedule $s = a_1; a_2; \ldots$ then $f(s) = f(a_1); f(a_2); \ldots$.

3 Specifying Targets and Monitors

We now show how monitors and targets can be modeled using I/O automata, building on the example of system-call interposition in Fig. 3. In §3.1 specify the

[4] In this paper, for brevity, our analyses focus on schedules and traces.

system-call interposition using I/O automata. Then, in §3.2 we discuss how the monitor designer can formalize different decision choices using I/O automata, and in Section 3.3 we explain how to encode monitors of previous frameworks (e.g., truncation and edit monitors) in our framework. Finally, in Section 3.4 we provide some concluding remarks on the advantages of using our framework for formalizing enforcement scenarios.

3.1 Modeling Targets and Monitors with I/O automata

We model targets (the entities to be monitored) and monitors as I/O automata. We let the metavariables T and M range over targets and monitors. Targets composed with monitors are called *monitored targets* or *monitored applications*; examples are the modified kernel and the safe application in Fig. 1. A monitored target might itself be a target for another monitor.

Suppose that the application's only actions are *OpenFile*, *WriteFile*, and *CloseFile* system calls; the kernel's actions are *FD* (to return a file descriptor) and the *Kill* system call. The application can make a request to open a file *fn*, and the kernel keeps track of the requests as part of its state. When a file descriptor *fd* is returned in response to a request for *fn*, *fn* is removed from the set of the requests. The application can then write *bytes* number of bytes to, or close, *fd*. Finally, a *Kill* action terminates the application and clears all requests. Such a formalization, where the target's actions depend on results returned by the environment, was outside the scope of original run-time monitors models, as identified also by more recent frameworks (e.g., [20, 13]).

Fig. 3a shows I/O automata interface diagrams of the monitored system consisting of the application and the monitored kernel.An I/O automaton definition for this kernel is shown in Fig. 2, using the standard precondition-effect style of writing transition relations for I/O automata.

The application's and the kernel's interfaces differ only in that the input actions of the kernel are output actions of the application, and vice versa. This models the communication between the application and the kernel when they are considered as a single system. The kernel's readiness to always accept file-open requests is modeled naturally by the input-enabledness of the I/O automaton. Paths (2) and (3) represent communication between the monitor and the kernel through the renamed actions of the kernel (using the renaming operation of I/O automata, §2): e.g., $\mathit{OpenFile}(x)$ becomes $\mathit{OpenFile\text{-}Ker}(x)$, and thus irrelevant to a policy that reasons about *OpenFile* actions. Renaming models changing the target's interface by adding hooks that allow the monitor to intercept the target's actions. In practice, this is often accomplished by rewriting the target in order to inline or interpose a monitor. Finally, we also *hide* the communication between monitor and the kernel so that it remains internal to the monitored target (denoted by the dotted line around the monitored kernel automaton). This is because we model a monitoring process that is transparent to the application (i.e., the application remains unaware that the kernel is monitored).

Signature: Input: *OpenFile*(*fn*), where *fn* is a *file_name*
(type *file_name* = nat)
WriteFile(*fd*, *bytes*), where *fd* is a *file_descriptor*
(type *file_descriptor* = nat)
(type *bytes* = nat)
CloseFile(*fd*), where *fd* is a *file_descriptor*
Output: *Kill*(), *FD*(*fd*,*fn*), where *fd* is a *file_descriptor*
(type *file_descriptor* = nat)
and *fn* is the corresponding *file_name*.

States: *req_list* : List of elements of type *file_name*
assigned_list : List of elements of type *file_descriptor*
kill : flag of type *bool*

Start States: *req_list* = *nil*
assigned_list = *nil*
kill = *false*

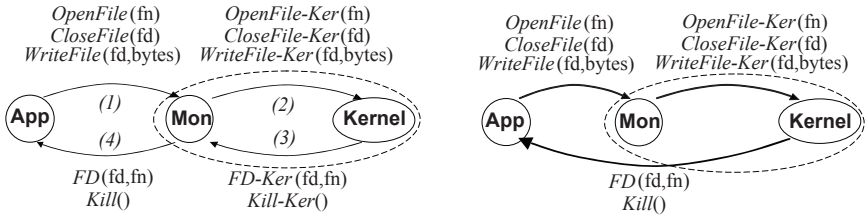
Transitions: *OpenFile*(*fn*)
Effect: *req_list* = *req_list*@[*fn*]
CloseFile(*fd*)
Effect: *assigned_list* = *assigned_list*\[*fn*], where
(*assigned_list**z*) denotes the function
that removes the element *z* from list *assigned_list*
WriteFile(*fd*, *bytes*)
Effect: Some lower level specification of write for writing
bytes on the actual file
FD(*fd*,*fn*)
Precondition: ¬*empty*(*req_list*) and ∃(*fn* : *file_name*) ∈ *req_list*,
where *empty* is a predicate on lists that returns *true* whenever
its argument is an empty list
Effect: *req_list* = (*req_list**fn*), where (*req_list**fn*) denotes the
function that removes the element *fn* from list *req_list*
assigned_list = *assigned_list*@[*fd*]
Kill()
Precondition: *kill* = *true*
Effect: *req_list* = *nil*
assigned_list = *nil*
and *kill* = *false*.

Fig. 2: Kernel I/O automaton definition

3.2 Modeling Monitoring Decisions with I/O automata

In our system-call interposition example from §1 we described some choices that a monitor designer can make, such as choosing the (1) interface to be monitored (e.g., mediate only input actions), and (2) the trace modification capabilities of the monitor. We next describe how to express the above choices in our model.

Modeling the Monitored Interface By appropriately restricting the renaming function applied to the target, we can model different monitoring architectures



(a) Input/output-mediating monitor (b) Input-mediating monitor

Fig. 3: I/O automata interface diagrams of kernel, application, and monitor

(e.g., input sanitization, §1). For example, in Fig. 3b, we renamed only the input actions of the kernel (i.e., *OpenFile*, *CloseFile*, and *WriteFile*). This allows us to model monitors that mediate inputs sent to the target and can prevent, for example, SQL injections attacks. Similarly, renaming only the outputs of the target we can model monitors that mediate only output actions (and can prevent, for example, cross-site scripting attacks).

Modeling Implementation Aspects of Monitors When defining monitors of different enforcement capabilities, as the ones of previous models (e.g., [25, 18]), one can realize that mapping transition functions of previous models to transition relations of I/O automata does not suffice to uniquely identify practical implementations of monitors. The added expressiveness of I/O automata allows us to model different implementations of monitors that might make different choices about: whether the monitor can edit input before forwarding it to the target application; the extent to which the monitor can ignore the application; and the extent to which the monitor can use the application as an oracle or simulator to discover, in a controlled way, how it would respond to different input actions.

However, if we focus on a uni-directional communication path from the target to the monitor to the environment, then all types of monitors defined in previous work are expressible in our framework (e.g., *security automata* [25] or *truncation automata* [19]), which halt targets that try to execute actions that violate the security policy; *suppression automata* [18], which can ignore some actions that the target wants to execute; and *edit automata* [19], which can both insert and remove actions from the trace that the target is producing.

Returning to our example: to model a truncation monitor that halts the kernel once the kernel outputs an *FD-Ker*(*fd*,*fn*) with an already-assigned file descriptor *fd*, we add a transition to the monitor that, upon receipt of the "bad" action, takes the monitor to a specific "halt state". Since the monitor is input enabled, that transition can be made regardless of which state the monitor is in. Once the monitor goes into this halt state, the only enabled output action will be a "halt" action to kill the kernel. The kernel will need to have this "halt" action as an input action, and an appropriate transition to stop its execution. Since the monitor is input enabled, it may, even when in the halt state, still receive

invalid actions from the kernel until the kernel is halted. In previous models, for any action that the target wanted to execute, the target would wait for the monitor to finish considering that action before trying to execute the following one; in other words, the target and the monitor were synchronized. However, in our framework the monitor does not in general have such control over the target. These issues affect the policies that are enforceable by the monitor, since the target might try to execute a series of invalid actions before the monitor gets a change to take corrective action; we will revisit this point in the next section. Next, in §3.3 we provide more technical details on how to encode truncation monitors using I/O automata. The section can be skipped without affecting the readability and understanding of the rest of the paper.

3.3 Encoding Truncation Monitors with I/O automata

More general, a translation of monitors from previous monitors to ours must involve three steps that account for the added expressiveness of I/O automata. The first two are "definitional"; i.e., the steps are related to the differences between the corresponding automata definitions. First, we need to extend the transition function of an automaton with transitions that model the input enabledness of the I/O automata. Second, in truncation, suppression, and edit automata the actions that the target was sending to the monitor belonged to the same action set as the ones that the monitor was forwarding to the environment. However, in I/O automata, the signature prohibits that, since the input and output actions must belong to disjoint sets. To account for that we need to define a bijection that will map the inputs of the truncation automata to fresh output actions. The third step involves the implicit assumptions made by previous models and exposed by our framework: a run-time monitor might not be able to control, in practice, how the target produces actions. More specifically, in previous models, for any action that the target wanted to execute, the monitor could decide and (perhaps) forward some action to the environment, and the target would wait for the monitor to finish before trying to execute the next action. In other words, the target and the monitor were synchronized. However, in our framework the monitor does not have such a control over the target. This means that the target might be producing actions without waiting for the monitor to make synchronized decisions. For that reason, our monitors need to have some data structure (e.g., a queue) to buffer the inputs from the target, and then, when given the chance (by fairness assumptions, for example), dequeue the corresponding actions and take appropriate action. Similarly, for truncation automata, the monitor might not have the ability to halt the target, unless we know that the target has some *halt* input action that will guarantee its termination (as in our system call interposition, with the *kill* system call). Next, we provide a translation of truncation automata [19] to I/O automata, to illustrate the above steps. Translations of other types of monitors can be defined similarly.

We will assume that the target can be terminated by a *stop* action. Given a truncation automaton $A_T = \langle Q, Q_0, \delta \rangle$ that is defined over some action set

Σ_{A_T}, we define a truncation monitor $M_T = \langle sig(M_T),\ states(M_T),\ start(M_T),\ R_{M_T},\ Tasks(M_T)\rangle$, where:

1. $sig(M_T) = \langle input(M_T), internal(M_T), output(M_T)\rangle$, where:
   (i) $input(M_T) = \Sigma_{A_T}$,
   (ii) $internal(M_T) = \emptyset$,
   (iii) $output(M_T) = f(input(M_T)) \cup \{stop\}$,
   where $f : input(M_T) \xrightarrow[onto]{1-1} (\Sigma \backslash input(M_T))$.
2. $states(M_T) = (Q \times (input(M_T))^\star) \cup \{\langle halt, \epsilon\rangle\}$; i.e., the state of automaton together with the queue to buffer inputs from the target, plus an additional halt state,
3. $start(M_T) = Q_0 \times \{\epsilon\}$,
4. $trans(M_T) =$
   $\{\langle\langle q, \sigma\rangle, \iota, \langle q, \sigma; \iota\rangle\rangle \mid \langle q, \sigma\rangle \in states(M_T)$ and $\iota \in input(M_T)\}$
   $\cup\ \{\langle\langle q, \alpha; \sigma\rangle, f(\alpha), \langle q', \sigma\rangle\rangle \mid \langle q, \alpha; \sigma\rangle \in states(M_T)$ and $\delta(q, \alpha) = q'\}$
   $\cup\ \{\langle\langle q, \alpha; \sigma\rangle, stop, \langle halt, \epsilon\rangle\rangle \mid \langle q, \alpha; \sigma\rangle \in states(M_T)$ and $\delta(q, \alpha) = halt\}$,
5. Each action in $local(M_T)$ defines a unique equivalence class.

3.4 Discussion

In this section we have discussed how to model more faithfully than usual practical enforcement scenarios (e.g., system-call interposition) by appropriately specifying targets, monitors, monitored targets, different monitoring architectures, and monitors of different enforcement capabilities using I/O automata. Concluding this section, we would like to point out certain more general benefits of using expressive frameworks like I/O automata for modeling enforcement scenarios. In particular, our model exposes many details that often remain implicit or informal when reasoning about enforcement, including the following.

1. *Monitored interface/Target modification:* The way a monitor is integrated with a target is typically not expressed in a model (e.g.,[25, 19, 20]). Our model makes this integration explicit through the renaming operation of I/O automata. Two of the benefits of this ability are: (a) acknowledgement that any target (or its interface) needs to be re-written, even minimally, so that the security relevant actions are directed to, and thus intercepted by, the monitor, and (b) provides an easy syntactic check for complete mediation; e.g., if not all actions have been renamed in the transition relation of the target, then there might exist a case where the execution of a security relevant action bypasses the monitor. This precisely captures the typical monitoring approach in which a monitor intercepts a target's security-relevant actions, but does not otherwise modify a target's state and behavior. Tighter integration could be modeled by changing the target's transition relation, but we do not explore that in this paper.
2. *Complete mediation/instrumentation:* Monitors typically assume that all security-relevant actions of a target can be mediated. Our model takes a

more nuanced view, which admits that there may exist security relevant actions that the monitor cannot observe or mediate. Such actions are labeled either as internal to a target, and the I/O automata formalism prevents them from being exposed (forwarded) to the monitor, or differently from any monitor's actions, and I/O automata composition prohibits communication between them. This allows us to accurately express some realistic scenarios; e.g., a monitor installed by a non-administrator may have only partial access to the target's system calls.

Reasoning explicitly about these details gives us reassurance that we are closing the gap between monitors in practice and their theoretical abstractions while receiving a better insight about their theoretical limits. From a practical perspective, it can shed light on issues related to the design of monitors. For example, attempting to encode a specific system in our framework may show that the monitor is not sufficiently protected from a target; conversely, a system faithfully implemented based on a model in our framework should inherit properties that are explicit in our model, including monitor integrity. From a theoretical perspective, it allows us to identify practical constraints in the theoretical limits of enforcement powers of monitors. For example, as we saw in the translation of the truncation automata above, and as we will further analyze in §5, we must acknowledge that there might be security relevant aspects of the enforcement process that our outside the monitor's control, such as the ability to control the target (either synchronize it with the monitor or control its local actions).

4 Policy Enforcement

In this section we define security policies and introduce two definitions of enforcement. The first defines enforcement with respect to a specific target, thus capturing scenarios in which the designer knows where the monitor is being installed (e.g., installing a system call interposition monitor to a specific version of a Linux kernel). The second one defines enforcement independently of the target, thus capturing scenarios in which the monitor designer might not know apriori the targets to which the monitor will be applied (e.g., when designing a system call interposition that enforces policies independently of the underlying kernel).

4.1 Security Policies

A *policy* is a set of (execution, schedule, or trace[5]) modules. We let the metavariables $\mathscr{P}$ and $\hat{P}$ range over policies and their elements (i.e., modules) respectively. The novelty of this definition of policy compared to previous ones (e.g., [25, 19]) is that each element of the policy is not a set of automaton runs, but, rather, a pair of a set of runs (i.e., schedules or traces) and a signature, which is a triple

[5] Our analyses equally apply to execution modules, but, for brevity, we discuss only schedule and trace modules.

consisting of a set of inputs, a set of outputs, and a set of internal actions. The signature describes explicitly which actions that do not appear in the set of runs are relevant to a policy. This is useful in a number of ways. When enforcing a policy on a system composed of several previously defined components, for example, the signatures can clarify whether a policy that is being enforced on one component also reasons about (e.g., prohibits or simply does not care about) the actions of another component. For our running example, if the signature contains only *Open*, *FD*, and *Kill*, then all other system calls are security irrelevant and thus permitted; if the signature contains other system calls (*SocketRead*), then any behaviors exhibiting those calls will be prohibited.

Our definition of a policy as a set of modules resembles that of a hyperproperty [8] and previous definition of policies (modulo the signature of each schedule or trace module) and captures common types of policies such as access control, noninterference, information flow, and availability.

Since I/O automata can have infinite states and produce possibly infinite computations, we would like to avoid arguments and discussions about computability and complexity issues: they are outside the scope of this paper. Thus, we make the assumption that all policies $\mathscr{P}$ that we discuss are *implementable* [26], meaning that for each module $\hat{P}$ in $\mathscr{P}$ under discussion, there exists an I/O automaton A such that $sig(A) = sig(\hat{P})$ and $scheds(\hat{P}) \subseteq scheds(A)$.

4.2 Enforcement

In §3 we showed how monitoring can be modeled by renaming a target T so that its security-relevant actions can be observed by a monitor M and by hiding actions that represent communication unobservable outside of the monitored target. We now define enforcement formally as a relation between the behaviors allowed by the policy and the behaviors exhibited by the monitored target.

Definition 1 *(Target-specific enforcement) Given a policy $\mathscr{P}$, a target T, and a monitor M we say that $\mathscr{P}$ is* target-specifically soundly enforceable *on T by M if and only if there exists a module $\hat{P} \in \mathscr{P}$, a renaming function* rename*, and a hiding function* hide *for some set of actions Φ such that $(scheds(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))|acts(\hat{P})) \subseteq scheds(\hat{P})$.*

Here, $\mathsf{hide}_\Phi(M \times \mathsf{rename}(T))$ is the monitored target: the target T is renamed so that its security-relevant actions can be observed by the monitor M; *hide* is applied to their composition to prevent communication between the monitor and the target from leaking outside the composition[6]. If a target does not need renaming, rename can be the identity function; if we do not care about hiding all communication, the hiding function can apply to only some actions. For example, suppose the monitored target from our running example (node with

[6] Since Def. 11 reasons about schedules (i.e., internal actions as well as input and output), hide_Φ is redundant. We include it in this definition to expose the re-writing process that needs to happen for run-time enforcement in practical scenarios, but we will omit it in the rest of the paper.

dotted lines in Fig. 3b) is composed with an additional monitor that logs system-call requests and responses. We would then keep the actions for system-call requests and responses visible to the logging monitor by not hiding them in the initial monitored target.

Def. 11 binds the enforcement of a policy by a monitor to a specific target. We refer this type of enforcement as *target-specific enforcement* and to the corresponding monitor as a target-specific monitor. However, some monitors may be able to enforce a property on any target. One such example is a system-call interposition mechanism that operates independently of the target kernel's version or type (e.g., a single monitor binary that can be installed in both Windows and Linux). We call this type of enforcement *generalized enforcement*, and the corresponding monitor a generalized monitor[7]. More formally:

Definition 2 *(Generalized enforcement) Given a policy $\mathscr{P}$ and a monitor M we say that $\mathscr{P}$ is* generally soundly enforceable *by M if and only if for all targets T there exists a module $\hat{P} \in \mathscr{P}$, a renaming function* rename*, such that $(scheds(M \times \mathsf{rename}(T))|acts(\hat{P})) \subseteq scheds(\hat{P})$.*

Different instances of Def. 11 and Def. 12 can be obtained by replacing schedules with traces (trace enforcement), by fair schedules or fair traces (fair enforcement), or by replacing the subset relation with other set relations (e.g., equality) for comparing the behaviors of the monitored target with that of the policy [20, 5]. In this paper we focus on the subset and the equality relations, and refer to the corresponding enforcement definitions, respectively, as *sound* (e.g., Def. 11) and *precise enforcement*.

4.3 Comparing Enforcement Definitions

As a first example of meta-theoretic analysis in our framework, we compare these two definitions. More specifically, one might expect target-specific monitors to have an advantage in enforcement. If we have a monitor that enforces a policy for any target (i.e., a generalized monitor) then the monitor also specifically enforces the policy for some target T. However, a monitor that is "customized" for enforcing a policy on a specific target (e.g., Linux) might not enforce the policy on any target (e.g., Windows).

Proposition 1. *Given a monitor M then:*

1. $\forall \mathscr{P}$: $\mathscr{P}$ *is* generally soundly enforceable *by* $M \Rightarrow$
   $\forall T : \mathscr{P}$ *is* specifically soundly enforceable *on* T *by* M*, and*
2. $\exists \mathscr{P} \exists T$: ($\mathscr{P}$ *is* specifically soundly enforceable *on* T *by* M)$\wedge$
   $\neg$($\mathscr{P}$ *is* generally soundly enforceable *by* M).

Proof idea. For (1) we use the module $\hat{P} \in \mathscr{P}$ and renaming function with which M generally enforces $\mathscr{P}$ and we apply them to any target T that we are given.

[7] Monitors of previous models, such as [25] and [19], are generalized monitors.

Intuitively, the universal quantification over targets in the right hand side of the implication is internalized on the left hand side in the definition of generalized enforcement.

For (2) we choose a policy $\mathscr{P}$ that contains only one module $\hat{P}$. $\hat{P}$ has as signature all possible actions and contains only the schedules that the monitor M can exhibit. Next we choose as T the trivial automaton that exhibits no behavior. Clearly M specifically enforces $\mathscr{P}$ on T. But, for any target that exhibits some internal actions, since these actions are not part of $\hat{P}$, M cannot generally enforce $\mathscr{P}$. □

Complete proofs are given in App. B. The proofs use several key I/O automata theorems, such as associativity of composition, or that renaming of compatible components can be done before or after composition. These theorems are included in App. A.

Prop. 1 compares the two definitions of enforcement (Def. 11 and Def. 12) with respect to the same monitor and shows that our definitions capture the intuitive notions of enforcement; i.e., a monitor that enforces a policy without being tailored for a specific target should enforce the policy on any target, while the inverse should not be true in general.

However, we can get a deeper insight when trying to characterize the two definitions of enforcement in general (i.e., independently of a specific monitor). Surprisingly, in such a comparison the two definitions turn out to be equivalent.

Theorem 1. $\forall \mathscr{P} : \forall T\text{:}\ \exists M : \mathscr{P}$ *is* specifically soundly enforceable *on* T *by* M $\Leftrightarrow \exists M' : \mathscr{P}$ *is* generally soundly enforceable *by* M'.

Proof idea. For the left direction use M' as M. For the right direction, use as M' the monitored target $M \times \mathsf{rename}(T)$ (which we know that exhibits behaviors that belong to $\mathscr{P}$) and $\alpha-$rename any target that we try to enforce the policy on. □

The left direction of the theorem is straightforward: any generalized monitor can be used as a target-specific monitor. The right direction is more interesting since it suggests, perhaps surprisingly, that it is possible to construct a generalized monitor from a target-specific one. More specifically, once we have a monitor that enforces a policy on a specific target, we can use this *monitored target* as the basis for a monitor on any other target. In that case, the only security-relevant behavior of the system would be exhibited by the monitor (formally, every action in every other target would be renamed to become security irrelevant). For example, suppose we have different versions of a specific application installed on each of our machines. If we find a patch (i.e., monitor) for one version, then Thm. 1 implies that instead of finding patches for all other versions, we can simply distribute the patched version (i.e., monitored target) to all machines and modify the existing applications on those machines so that their behavior is ignored. This approach might be relevant when reinstalling the patched version of the application on top of other versions is more cost-efficient than finding patches for every other version.

Thm. 1 holds for two reasons: Def. 11 and 12 place no restrictions on (1) renaming functions (i.e., how a monitor is integrated with a target), and (2) the choice of elements of the policy $\mathscr{P}$; i.e., what are the exact semantics of the requirements we impose no the monitored target. In practice, the interactions between the monitor and the target and the description of the element of a policy that we want to enforce may be more constrained. Thus, one might argue that it would be more natural to have the only-if direction of the theorem fail, since it erases the distinction between target-specific and generalized enforcement. This happens only if we restrict some elements in our definition of enforcement.

Towards that end we introduce the notion of *maximal enforcement*. Intuitively, when talking about a monitor *maximally enforcing* a policy on a target, we imply the following restriction: when picking an element of the policy to compare our monitored system with, we always have to pick the one that (maximally) matches the signature of the monitor, the target, and the (range of the) renaming function; i.e., the element of the policy has to reason about how the monitor and the target are integrated together. Thus, we make explicit the semantics of $\alpha-$renaming: we cannot make security-relevant actions security-irrelevant, by $\alpha-$renaming them to actions that our outside the signature of the module. This is to ensure that we do not choose modules that do not care about the behavior of the original target, and thus implicitly allow everything the target wants to do. For example, assume we have a policy that reasons about file operations and networking events. Moreover, assume the policy has two modules: one reasons about file operations and the other reasons about networking events. Both modules allow empty traces; i.e., they allow scenarios where the (monitored) target does nothing. With maximal enforcement, to check whether the policy is enforced by a firewall on the networking interface of a target, we compare the behavior of the monitored target with the schedules of the network module, and not the file module. Thus, if the target tries to send a packet to a blacklisted address, and the firewall does not mediate that communication (i.e., incomplete mediation due to e.g., insufficient rights) then the firewall cannot enforce that policy. However, had we used the basic definition of enforcement, then one could argue that the policy is enforceable since the firewall trivially enforces it using the file module: no file operations are exhibited by the network interface and thus nothing bad can happen.

Def. 3 and 4 formally express the above constraints.

Definition 3 *(Target-specific Maximal enforcement) Given a policy $\mathscr{P}$, a monitor M, a target T, and a set of renaming functions $\mathcal{R}$ we say that $\mathscr{P}$ is* specifically maximally soundly enforceable *on T by M using $\mathcal{R}$ iff there exists $ren \in \mathcal{R}$ and $\hat{P} \in \mathscr{P}$ such that:*

1. $sig(\hat{P}) = sig(M) \cup sig(T) \cup range(ren)$, *and*
2. $\big(scheds(M \times ren(T)) \mid acts(\hat{P})\big) \subseteq scheds(\hat{P})$.

Definition 4 *(Generalized Maximal enforcement) Given a policy $\mathscr{P}$, a monitor M, a set of targets $\mathcal{T}$, and a set of renaming functions $\mathcal{R}$ we say that $\mathscr{P}$ is*

generally maximally soundly enforceable *on* $\mathcal{T}$ *by* M *using* $\mathcal{R}$ *iff* $\forall T \in \mathcal{T}$ *there exists* $ren \in \mathcal{R}$ *and* $\hat{P} \in \mathscr{P}$ *such that:*

1. $sig(\hat{P}) = sig(M) \cup sig(T) \cup range(ren)$, *and*
2. $\big(scheds(M \times ren(T)) \mid acts(\hat{P})\big) \subseteq scheds(\hat{P})$.

Using the definition of maximal enforcement we can show that the equivalence between general and specific enforcement is no longer true[8]:

Theorem 2. $\exists \mathscr{P} : \exists \mathcal{R} : \exists \mathcal{T} : \exists T \in \mathcal{T}$*:*
$\exists M : \mathscr{P}$ *is* specifically maximally soundly enforceable *on* T *by* M *using* $\mathcal{R} \wedge$
$\nexists M' : \mathscr{P}$ *is* generally maximally soundly enforceable *on* $\mathcal{T}$ *by* M' *using* $\mathcal{R}$.

Proof idea. Construct a policy $\mathscr{P}$ with two elements $\hat{P}_1$ and $\hat{P}_2$ that describe two different targets, T_1 and T_2. But although $\hat{P}_1$ is enforceable, $\hat{P}_2$ is not (we construct a non-enforceable policy by applying Thm. 3). Thus, by definition of maximal enforcement when trying to enforce $\mathscr{P}$ on T_1 we have to pick $\hat{P}_1$ which is enforceable, and thus specifically maximally soundly enforceable. Dually, when trying to enforce $\mathscr{P}$ on T_2 we have to pick $\hat{P}_2$ which is not enforceable, and thus $\mathscr{P}$ is not generally maximally soundly enforceable. □

Thm. 2 does not allow for general and specific enforcement to imply each other. Thus, the definition of maximal enforcement restricts the generality of the framework as was introduced in the previous subsections. Maximal enforcement can be useful in two types of scenarios: (1) we want to reason about how the monitor and the target are integrated and thus our policy needs to reason about the renaming function (e.g., we are not allowed to $\alpha-$rename), and (2) scenarios where the monitor cannot implement (i.e., substitute for) every behavior of the target. For example, if we have a policy that reasons about cryptographic operations and file operations, we might be able to build a monitor that faithfully reproduces the file operations that the target can perform, but not a monitor that reproduces the cryptographic operations of the target (e.g., due to lack of access to private cryptographic keys). In this case maximal enforcement can help us to precisely talk about the practical limitation of monitors when discussing about the enforceability of policies. On the other hand, the patching example above illustrates a practical scenario where the most general framework is appropriate for modeling and reasoning about enforcement. Since our goal is to introduce a framework that is general enough to accommodate as many practical scenarios as possible (even seemingly degenerate ones), we rely on the monitor designer to impose appropriate restrictions on renaming or monitors to better reflect on the (practical) monitors under scrutiny.

4.4 Transparency

Typically, definitions of enforcement include, besides soundness, the notion of *transparency* [19]. Defining enforcement as a subset relation (i.e., sound enforce-

[8] Note that Thm. 1 is implicitly universally quantified over all (sets of) renaming functions and targets. Thus, Thm. 2 is indeed the negation of Thm. 1.

ment) allows certain, not always practical, cases where monitors correctly (i.e., soundly) enforce policies by simply doing nothing. Thus, to capture practical monitors that do not inhibit the correct behavior of the target, the requirement of transparency was introduced: monitors are forced to output all the correct traces that the target wants to exhibit.

All definitions of transparency that have been introduced so far are within frameworks where policies reason only about the target's behavior [25, 19]: a policy $\mathfrak{P}$ is a predicate over traces of the target (i.e., a subset of the traces that the target might exhibit). In such frameworks one explicitly states transparency in the definition of enforcement by stating that if a target's behavior s is allowed by a policy $\mathfrak{P}$ (i.e., $s \in \mathfrak{P}$) then the monitor exhibits s.

In our framework, we take a more general view and we allow policies to describe how monitors are integrated with targets, and how monitors are allowed to react to target's requests. Thus, enforcement is now implicit in the definition of a policy (i.e., in the traces that the policy allows). The latter view has also been adopted by Mandatory Results Automata (MRA) [20]. As discussed there, in such more expressive frameworks we do not have to explicitly state transparency as a requirement in the definition of enforcement [20] since we can define it as a specific type of input/output relation (within the formal syntax of the framework).

We first show how to encode in our framework the notion of transparently enforcing a policy $\mathscr{P}_T$ over a target (i.e., a policy of previous frameworks) as a policy $\mathscr{P}$ over a monitored target. The idea is that for every behavior t that is allowed by $\mathscr{P}_T$ we will construct a schedule that belongs to $\mathscr{P}$ such that the renamed target exhibits (the renamed) t, and the monitor exhibits t. More formally:

From transparent enforcement to transparent policies Given a monitor M, a renaming function ren, a target T, and a policy $\mathscr{P}_T$ with $\mathscr{P}_T = sig(T)$, indicating the allowed behaviors of the target (i.e., policies of previous frameworks) we say that a module $\hat{P}$ of a policy $\mathscr{P}$ describing the monitored target $(M \times ren(T))$ is *transparent* if and only if for all $t \in \mathscr{P}_T$ there exists a schedule $s \in \hat{P}$ such that $s = (ren(t) \mid\mid t)$.

Note, that the above construction does not tell us anything about what additional schedules $\hat{P}$ contains. This is because we have multiple ways to interpret, or express, in our framework what it means to transparently enforce a policy. A question that illustrates this point is whether $\hat{P}$ should contain the schedule $s' = ren(t)$; i.e., the schedule that the monitored target is exhibiting. Under the view of old frameworks this is fine, since $\hat{P}$ considers the target's behavior t to be valid. However, one might say that s' should not be part of $\hat{P}$ since it contradicts the intended semantics: $s' \in \hat{P}$ states that the monitor can react to $ren(t)$ by doing nothing, whereas $s \in \hat{P}$ states that the monitor can react by exhibiting t: but only the latter describes transparency.

As we will see in the next section, this distinction, and choice of semantics, is strongly related to the fairness assumptions that we make in our model. If we assume fairness we should not include s' in $\hat{P}$. But, if we don't, then we might

accept including s' as long as s belongs in $\hat{P}$: essentially we are saying that it is ok for the monitor to do less due to external reasons, as long as the monitor wants to do the right thing (i.e., transparent enforcement).

We formalize these two different approaches through the notions of *weak* and *strong transparency.*

Definition 5 *(Weakly transparent module) Given a monitored target* $(M \times ren(T))$*, and a schedule module* $\hat{P}_T$ *with* $sig(\hat{P}_T) = sig(T)$*, a schedule module* $\hat{P}$*, with* $sig(\hat{P}) = sig(M \times ren(T))$*, is* weakly transparent *for* $(M \times ren(T))$ *w.r.t.* $\hat{P}_T$ *if and only if* $\forall t \in \hat{P}_T : \exists s \in scheds(\hat{P}) : ren^{-1}(s|\mathsf{range}(ren)) = (s|\mathsf{dom}(ren)) = t$*.*

Def. 5 states that a module is weakly transparent, w.r.t. some policy on the target, if and only if for every schedule that belongs to these target's behaviors there is some schedule in the module such that if we take the subsequence that consists of actions of the monitored target and we reverse the renaming then we will get another subsequence of that schedule that is exhibited by the monitor.

Definition 6 *(Strongly transparent module) Given a monitored target* $(M \times ren(T))$*, and a schedule module* $\hat{P}_T$ *with* $sig(\hat{P}_T) = sig(T)$*, a schedule module* $\hat{P}$*, with* $sig(\hat{P}) = sig(M \times ren(T))$*, is* strongly transparent *for* $(M \times ren(T))$ *w.r.t.* $\hat{P}_T$ *if and only if* $\forall t \in \hat{P}_T : \forall s \in scheds(\hat{P}) :$ *if* $ren^{-1}(s|\mathsf{range}(ren)) = t$*, then* $(s|\mathsf{dom}(ren)) = t$*.*

Def. 6 states that a module is strongly transparent, w.r.t. some policy on the target, if and only if for every schedule s that belongs to the acceptable target's behaviors there is no schedule that the monitored target exhibits where the (renamed) target tries to execute (renamed) s, and the monitor does not exhibit s as well. The correspondence is as the one in weak transparency: for every schedule in the module if we take the subsequence that consists of actions of the monitored target and we reverse the renaming then we will get another subsequence of that schedule that is exhibited by the monitor (which is an acceptable behavior of the target).

Def. 5 and 6 defined transparency as a specific type of policy $\mathscr{P}$ over a monitored target given a policy $\mathscr{P}_T$ over the original target (that is now monitored). However, we did not relate $\mathscr{P}_T$ and $\mathscr{P}$ w.r.t. "soundness". For example, if $\mathscr{P}_T$ does not contain a schedule s, and thus disallows it, we did not require from $\mathscr{P}$ to also exclude schedules in which the monitor exhibits s. The next definition shows how to achieve this goal, assuming monitors that completely mediate target's security relevant actions:

Definition 7 *(Sound module) Given a monitored target* $(M \times ren(T))$*, and a schedule module* $\hat{P}_T$ *with* $sig(\hat{P}_T) = sig(T)$*, a schedule module* $\hat{P}$*, with* $sig(\hat{P}) = sig(M \times ren(T))$ *and* $sig(T) \subseteq sig(M)$*, is* sound *for* $(M \times ren(T))$ *w.r.t.* $\hat{P}_T$ *if and only if* $\forall t \notin \hat{P}_T : \nexists s \in scheds(\hat{P}) : (s|\mathsf{dom}(ren)) = t$*.*

Similar definitions can be expressed for partially-mediating monitors, but we do not pursue them here further. Using Def. 7, 5 (or, 6), and the construction

principles in §3.3, one can fully embed previous frameworks (e.g., definitions of enforcement and edit automata [19]) in ours.

5 Generally Enforceable Policies

The definitions and abstractions described thus far enable rigorous, detailed analyses of practical monitored systems, and also facilitate meta-theoretic reasoning that furthers our understanding of general limitations of practical monitors that fit this model. In this section we begin our analysis by focusing on general enforcement and derive several such meta-theoretic results. In the next section, we further our analysis by focusing on target-specific enforcement.

5.1 Auxiliary Definitions

I/O automata are input enabled—all input actions are enabled at all states. Several arguments can be made in favor of or against input-enabledness. For example, one might argue that input-enabledness may lead to better design of systems because one has to consider all inputs that may be received from the environment [21]. On the other hand, this constraint might be too restrictive for practical systems [1].

In our context, we believe that input-enabledness is a useful characteristic, since run-time monitors are by nature input-enabled systems: the monitor may receive input at any time both from the target and from the environment (e.g., keyboard or network). However, a monitor modeled as an input-enabled automaton can enforce only those policies that allow the arrival of inputs at any point during execution. This is reasonable: a policy that prohibits certain inputs cannot be enforced by a monitor that cannot control those inputs. We later combine this and several other constraints to describe the lower bound of enforceability in our setting.

We say that a module (or policy) is *input forgiving* (respectively, *internal* and *output* forgiving) if and only if it allows the empty sequence and allows each valid sequence to be extended to another valid sequence by appending any (possibly infinite) sequence of inputs.

Definition 8 *A schedule module $\hat{P}$ is* input forgiving *if and only if:*
(1) $\epsilon \in \mathit{scheds}(\hat{P})$*; and*
(2) $\forall s_1 \in \mathit{scheds}(\hat{P}) : \forall s_2 \preceq s_1 : \forall s_3 \in (\mathit{input}(\hat{P}))^{\infty} : (s_2; s_3) \in \mathit{scheds}(\hat{P})$.

I/O automata's definition of executions allows computation to stop at any point. Thus, the behavior of an I/O automaton is *safe*: any prefix of a schedule exhibited by an automaton is also a schedule of that automaton, and all successive extension of schedules are limit closed [21]:

Definition 9 *A schedule module $\hat{P}$ is a* safety module *if and only if:*

1. $\mathit{scheds}(\hat{P}) \neq \emptyset$.

2. *scheds*$(\hat{P})$ *is* prefix closed*; i.e., if* $s \in scheds(\hat{P})$ *and* $s' \preceq s$*, then* $s' \in scheds(\hat{P})$*.*
3. *scheds*$(\hat{P})$ *is* limit closed*; i.e., if* $s_1, s_2, \ldots$ *is an infinite sequence of finite sequences in* $scheds(\hat{P})$*, and for each* i*,* s_i *is a prefix of* s_{i+1}*, then the unique schedule* s *that is the limit of* s_i *under the successive extension ordering is also in* $scheds(\hat{P})$*.*

These two characteristics are unsurprising from the standpoint of models for distributed computation, but describe practically relevant details that are typically absent from models of run-time enforcement. Our model, instead of making assumptions that might not hold in every practical scenario (e.g., that all actions can be mediated) takes a more nuanced view, which admits that there are aspects of enforcement outside the monitor's control, such as security-relevant actions that the monitor cannot observe or mediate (labeled as internal to a target), or the existence (or lack of) scheduling strategies that might not favor the monitor. The definitions above help us explicate these assumptions when reasoning about enforceable policies, as we see next.

5.2 Lower Bounds of Enforceable Policies

Another constraint that affects the lower bounds of enforceability and is semantically specific to monitoring is that, in practice, a monitored system cannot always ignore *all* behavior of the target application. Some realistic monitors decide what input actions the application sees, but otherwise do not interfere with the application's behavior—firewalls belong to this class of monitors. In such cases, a monitor can soundly enforce a policy only if the policy allows all the behaviors that the target can exhibit even if it receives no input. We call these policies *quiescent forgiving* (recall the definition of a quiescent state from §2). Modules contained is such policies are also called quiescent forgiving. This definition captures one type of limitation that was understood to be present in run-time monitoring, but that typically was not formally expressed. Quiescent forgiving modules can be defined more formally as follows:

Definition 10 *A schedule module* $\hat{P}$ *is* quiescent forgiving *for some* T *if and only if:*
$\forall e \in execs(T)$ *such that* $e = q_0, a_1, \ldots, q_n$:
$\Big(q_n \in quiescent(T) \wedge \big(\forall i \in \mathbf{N} : 0 \leq i < n : q_i \notin quiescent(T)\big)\Big) \Rightarrow$
$\big(sched(e)|acts(\hat{P})\big) \in scheds(\hat{P}) \wedge (\forall i \in \mathbf{N} : 0 \leq i < n : (sched(q_0, \ldots, q_i) \mid acts(\hat{P})) \in scheds(\hat{P}))$.

The following theorem formalizes a lower bound: a policy that is not quiescent forgiving, input forgiving, and prefix closed cannot be (precisely) enforced by any monitor.

Theorem 3. $\forall \mathscr{P} : \forall \hat{P} \in \mathscr{P} : \forall T : \forall \mathsf{rename} :$
$\exists M : (scheds(M \times \mathsf{rename}(T))|acts(\hat{P})) = scheds(\hat{P}) \Rightarrow$
$\hat{P}$ *is input forgiving, safety, and quiescent forgiving for* $\mathsf{rename}(T)$.

Proof idea. The proof is by contradiction. If $\hat{P}$ does not allow certain inputs then it is not enforceable because the monitored target is an I/O automaton, and I/O automata cannot block inputs: they always appear in the schedules of the monitored target. The same argument holds for safety: I/O automata's schedules are prefix closed (and limit closed), thus $\hat{P}$ has to be as well. Finally, since the monitor does not control the local actions that the target will execute at the beginning of its execution, if the property forbids some of these schedules then the monitor cannot enforce the policy. □

Thm. 3 reveals that monitors, regardless of their editing power, can enforce only safety properties. Thus, in our context, even the equivalent of an edit monitor cannot enforce renewal properties (as opposed to [19]), since when renewal properties are constrained by prefix closure they collapse to safety. This is because, as mentioned above, our model of executions allows computation to stop at any point. This is another helpful characteristic of our model; i.e., it highlights that on systems in which execution may cease at any moment (e.g., due to a power outage), only safety properties can be enforced.

In practice, monitors typically reproduce at most a subset of a target's functionality. Hence, if a monitor composed with an application is to exhibit the same range of behaviors as the unmonitored application, it will have to consult the target application in order to generate these behaviors. In the system-call interposition example, for instance, the monitor cannot return correct file descriptors without consulting the kernel. Such monitors, which regularly consult an application, cannot precisely enforce (with respect to schedules) arbitrary policies even if they are quiescent forgiving, input forgiving, and prefix-closed. This is because an input forwarded by the monitor to an application might cause the application to execute internal or output actions (e.g., a buffer overflow) that are not allowed by the policy and that the monitor cannot prevent, since these are outside of the interface between the monitor and the target.

On the other hand, in practice it is also common for the monitor (or system designer) to have some knowledge about the target, even if it does not have access to its state. This knowledge can be exploited to use simpler-than-expected monitors to enforce of (seemingly) complex policies. Although similar observations have been made before (e.g., program re-writing [15], non-uniformity [19], use of static analysis [7]), our framework can be used to formally extend them, as we demonstrate in §6.

5.3 Lower Bounds of Transparently Enforceable Policies

In §4.4 we discussed how to encode the notion of transparent enforcement as a specific type, or instantiation, of policy in our framework. Here we take a closer look at the constraints under which such policies are enforceable.

As discussed in the previous section, in our basic framework monitors can only enforce safety policies. Thus, if a monitor needs to exhibit more than one actions to (strongly) transparently enforce a policy, there is no guarantee that it will be able to do so. In contrast, previous models (e.g., [19]) assumed that

enabled actions of a monitor would always be performed. In our framework, to achieve equivalent results we can either explicitly add similar guarantees about the runs of the system through appropriate fairness constraints [17], or relax the requirements in the definition of transparent enforcement (cf. weak and strong transparency). This is another instance of our framework making explicit the (practical) assumptions and constraints that affect the enforcement of policies.

First, we provide definitions of target-specific and general *fair enforcement*. The following definitions of enforcement compare the schedules of the policy with the fair schedules of the monitored target; i.e., we only care about the "final" behaviors that the monitored target wants to exhibit and not all steps that it has to take to reach them.

Definition 11 *(Fair target-specific enforcement) Given a policy $\mathscr{P}$, a target T, and a monitor M we say that $\mathscr{P}$ is* fairly specifically soundly enforceable *on T by M if and only if there exists a module $\hat{P} \in \mathscr{P}$, a renaming function* rename*, and a hiding function* hide *for some set of actions Φ such that $(fairscheds(\mathsf{hide}_{\Phi}(M \times \mathsf{rename}(T)))|acts(\hat{P})) \subseteq scheds(\hat{P})$.*

Definition 12 *(Fair generalized enforcement) Given a policy $\mathscr{P}$ and a monitor M we say that $\mathscr{P}$ is* fairly generally soundly enforceable *by M if and only if for all targets T there exists a module $\hat{P} \in \mathscr{P}$, a renaming function* rename*, such that $(fairscheds(M \times \mathsf{rename}(T))|acts(\hat{P})) \subseteq scheds(\hat{P})$.*

Using these definitions we can now characterize the policies that are transparently enforceable by monitors, or equivalently, under which constraints monitors can enforce policies that encode transparency.

If a policy $\mathscr{P}$ contains a module $\hat{P}$ that is strongly transparent, then there is no monitor that can precisely enforce $\mathscr{P}$ using $\hat{P}$ without taking into account fairness: the schedules of the monitored target will be prefix closed, whereas the schedules of the policy won't; i.e., the equality relation won't hold. On the other hand, if a monitor fairly specifically precisely enforces *policy* using $\hat{P}$ then the $\hat{P}$ must be be input and quiescent forgiving. It is easy to show that the last statement is non-trivial; i.e., there exists some strongly transparent $\hat{P}$ that is fairly specifically enforceable by some monitor M. These observations are formalized in Thm. 4.

Theorem 4. $\forall \mathscr{P} : \forall \hat{P} \in \mathscr{P} : \forall T : \forall M : \forall \hat{P}_T : \forall \mathsf{rename} :$
if $\hat{P}$ is strongly transparent for $M \times \mathsf{rename}(T)$ w.r.t. $\hat{P}_T$ then: (1) there is no monitor M that specifically precisely enforces $\mathscr{P}$ on T using $\hat{P}$, and (2) if there exists M that fairly specifically precisely enforces $\mathscr{P}$ on T using $\hat{P}$ then $\hat{P}$ is input forgiving, and quiescent forgiving for $\mathsf{rename}(T)$.

Proof idea. Since $\hat{P}$ is strongly transparent then every schedule that belongs to $\hat{P}$ must be of even length (containing the behavior that the target wants to execute and the behavior that the monitor forwards to the environment). But, the monitored target is an I/O automaton which means that its schedules are prefix closed and thus it contains schedules of odd length. Thus, there is no way

for the set of schedules of the monitored target to be equal to set of schedules of a strongly transparent $\hat{P}$.

On the other hand, the fair schedules of the monitored target can be of even length, and thus the only constraints that $\hat{P}$ must adhere to is input and quiescence forgiveness (for reasons described in Thm. 3). □

The requirement for fairness in order to enforce strongly transparent modules (and policies) is not tight to the definition of precise enforcement (i.e., equality relation between sets of schedules). It is required even when talking about sound enforcement (i.e., subset relation). However, there is a corner case where we can soundly enforce a strongly transparent module without the use of fairness: the monitored target exhibits no behavior at all and the module allows it. This is an important point for the specification of security policies since there might be cases where transparency is not correctly captured: if a target is not violating the policy but is blocking; i.e., it does exhibits good behaviors but only if the monitor initiates the computation, then one might argue that we are not transparent since we are prohibiting the target from performing good actions. But, it is up to the specific scenario and semantics of transparency and policies that the designer wants to capture to decide whether this corner case should be allowed or not. The following proposition formalizes this idea:

Proposition 2. $\forall \mathscr{P} : \forall \hat{P} \in \mathscr{P}$, *such that* $\epsilon \notin \mathit{scheds}(\hat{P})$:
$\forall T : \forall M : \forall \hat{P}_T$ *with* $\mathit{scheds}(\hat{P}_T) \neq \emptyset$*:* $\forall \mathsf{rename}$:
if $\hat{P}$ *is strongly transparent for* $M \times \mathsf{rename}(T)$ *w.r.t.* $\hat{P}_T$ *then there is no monitor* M *that specifically soundly enforces* $\mathscr{P}$ *on* T *using* $\hat{P}$.

Proof idea. As explained in the proof idea of Thm. 4, a monitored target may exhibit schedules of odd length that can not belong in a strongly transparent module. The only corner case is when the monitored target exhibits no schedules at all, which means that it exhibits the empty schedules (which is of even length). Since the strongly transparent modules of the theorem do not contain the empty schedule, then the policy is not even soundly enforceable. □

As discussed above, we can transparently enforce a larger class of policies if we allow for a weaker definition of transparency (i.e., weak transparency). A weakly transparent module may range between the two extremes: strongly transparent modules and safety modules. These two bounds are found when we take into account fairness, or the lack of it:

Theorem 5. $\forall \mathscr{P} : \forall \hat{P} \in \mathscr{P} : \forall T : \forall M : \forall \hat{P}_T : \forall \mathsf{rename}$:
if $\hat{P}$ *is weakly transparent for* $M \times \mathsf{rename}(T)$ *w.r.t.* $\hat{P}_T$ *then: (1) if there exists monitor* M *that specifically precisely enforces* $\mathscr{P}$ *on* T *using* $\hat{P}$*, then* $\hat{P}$ *is input forgiving, safety, and quiescent forgiving for* $\mathsf{rename}(T)$*, and (2) if there exists monitor* M *that fairly specifically precisely enforces* $\mathscr{P}$ *on* T *using* $\hat{P}$ *then* $\hat{P}$ *is input forgiving, and quiescent forgiving for* $\mathsf{rename}(T)$.

Proof idea. Derived easily from Thm. 3, for (1), and 4, for (2). □

6 Target-specifically Enforceable Policies

As discussed in §3, the expressiveness of our model allows multiple ways to define monitors (e.g., a truncation monitor) that had a single natural definition in previous models. Due to space limitations, rather than comprehensively analyzing the policies enforceable by specific monitors, as done in previous work [25, 18–20, 15], we demonstrate how our framework enables formal results that can be exploited by designers of run-time monitors who have knowledge about the target application: a novel analysis of how some knowledge of the target can compensate (in terms of enforceability) for a narrower monitoring interface (i.e., incomplete mediation).

We begin by showing how partially mediating monitors can be formally defined in our framework. We will first define what it means for a monitored target to be input/output mediating and input mediating. The definitions formalize the constraints on the renaming functions of the monitored target, as they were described in §3.

Definition 13 *A monitor M is input/output mediating iff:* $\forall T$*:* $\exists \mathsf{rename}$*:*

1. $output(\mathsf{rename}(T)) \subseteq input(M)$
2. $input(\mathsf{rename}(T)) \subseteq output(M)$
3. $internal(\mathsf{rename}(T)) = internal(T)$
4. $output(T) \subseteq output(M)$
5. $input(T) \subseteq input(M)$

Constraints (1-3) force the renaming function to match the interfaces of the target and the monitor (i.e., it does not allow to arbitrarily rename the target interface) and ensure that all security relevant input/output behavior of the target is completely mediated by the monitor. In particular, constraint (1) ensures that all security relevant outputs will be received by the monitor, while constraint (2) ensures that all the security relevant inputs to the target will come from the monitor. Constraint (3) ensures that the security relevant actions of the target are not renamed so that we can capture the fact that there are actions that are outside the monitor's control: if we could rename them to some internal actions of the monitor, then since the monitor controls its own internal actions, it would be possible to not exhibit invalid internal actions. Finally, constraints (4) and (5) ensure that the monitor has the ability to input and output the actions that the original target could.

Similarly to Definition 13 we can define a monitored target to be *input mediating*:

Definition 14 *A monitor M is input mediating iff:* $\forall T$*:* $\exists \mathsf{rename}$*:*

1. $input(\mathsf{rename}(T)) \subseteq output(M)$
2. $internal(\mathsf{rename}(T)) = internal(T)$
3. $output(\mathsf{rename}(T)) = output(T)$
4. $input(T) \subseteq input(M)$

Constraint (1) ensures that all the security relevant inputs to the target will come for the monitor. Constraints (2) and (3) ensure that the security relevant actions of the target are not renamed so that we can capture the fact that there are actions that are outside the monitor's control, this time including the output actions of the target. Finally, constraint (4) ensures that the monitor has the ability to input all the actions that the original target could.

In §3 we described two monitoring architectures: one in which the monitor mediates the inputs and the outputs of the target, and another in which it mediates just the inputs. Intuitively, an input/output-mediating monitor should be able to enforce a larger class of policies than an input-mediating one, since the former is able to control (potentially) more security-relevant actions than the latter (i.e., the outputs of the target). In other words, there exist policies that are enforceable by input/output mediating monitors, but not by input mediating monitors. This can be expressed as follows:

Theorem 6. $\exists \mathscr{P}$:
($\mathscr{P}$ *is generally precisely enforceable by some input/output-mediating* M_1)$\wedge$
$\neg$($\mathscr{P}$ *is generally precisely enforceable by some input-mediating* M_2)*, if the policy does not reason about the communication between the monitor and the target*[9].

Proof idea. Construct a policy that prohibits certain (targets') output actions. It is easy to see that there exists a target (i.e., an I/O automaton) that exhibits exactly the actions that the policy disallows. An input/output-mediating monitor can enforce that policy since it mediates the output actions that the (renamed) target wants to execute and if they violate the policy does not forward them to the environment. But an input-mediating monitor cannot (by definition) prohibit the bad output actions from happening, and thus it cannot precisely enforce the policy. □

The constraint in Thm. 6 identifies those policies for which input/output mediating architectures are at least as powerful as input mediating architectures.If the policy reasons about, and prohibits, (some) communication between the target and the monitor, then the two architectures are equivalent.

For the proof we pick a policy whose elements (i.e., modules) disallow all output actions (excluding the ones used for target-monitor communication) and a target that performs only output actions. An input mediating monitor cannot enforce that policy on that target since it does not mediate its output, and thus it cannot generally enforce the policy. However, an input/output mediating monitor will be able to enforce the policy, since whenever it receives any (renamed) actions from the target, it will just suppress them.

It follows from Thm. 6 that an input/output-mediating monitor enforces strictly more policies than an input-mediating monitor.

Corollary 1. $\{\mathscr{P} \mid \mathscr{P}$ *is reasonable and generally precisely enforceable by some input-mediating* $M_1\} \subsetneq \{\mathscr{P} \mid \mathscr{P}$ *is reasonable and generally precisely enforceable by some input/output-mediating* $M_2\}$.

[9] If we used trace enforcement, the constraint would be superfluous. We used schedules to remain consistent with other theorems in the paper.

The subset direction is straightforward: for any input-mediating monitor M_1 we can construct an input-output mediating monitor M_2 that echoes to the environment every output action that it intercepts. It is clear that for any target T, M_1 and M_2 will have equivalent behaviors. The not-equal direction follows from Thm. 6.

If we instantiate in Cor. 1 monitor M_1 with a truncation monitor T_{M_1} and monitor M_2 with a truncation monitor T_{M_2}, we get the following result.

Corollary 2. *Given an input-mediating truncation monitor T_{M_1} and input/output-mediating truncation monitor T_{M_2}, $\{\mathscr{P} \mid \mathscr{P}$ is reasonable and generally precisely enforceable by $T_{M_1}\} \subsetneq \{\mathscr{P} \mid \mathscr{P}$ is reasonable and generally precisely enforceable by $T_{M_2}\}$.*

Cor. 2 illustrates how differences in the implementation of monitors with seemingly equal "power", or capabilities, (according to certain previous models, e.g., [19]), affect the enforceability of security policies.

In fact, when taking into consideration implementation details of monitors we have to revisit previous results on enforceability of policies by monitors with different operational semantics. For example, it has been proven that edit monitors can enforce a strict superset of policies that truncation monitors enforce [19]. However, this relation does not hold if we consider how the monitors are implemented (i.e., what interface of the target they monitor):

Corollary 3. *$\exists$ safety policy $\mathscr{P}$: $\exists$ input/output mediating truncation monitor M_T : $\exists$ input-mediating edit monitor M_E : ($\mathscr{P}$ is generally precisely enforceable by M_T)$\wedge$*
$\neg$($\mathscr{P}$ is generally precisely enforceable by M_E), if the policy does not reason about the communication between the monitor and the target.

Thm. 6 and Corollaries 1- 3 establish that some policies are generally enforceable by input/output mediating monitors but not by input mediating monitors. However, for some targets the two architectures are equivalent in enforcement power. The following theorem characterizes the targets for which this equivalence holds.

Theorem 7. $\forall\mathscr{P} : \forall T :$
$\mathscr{P}$ is specifically precisely enforceable on T by some input/output-mediating M_1 iff $\mathscr{P}$ is specifically precisely enforceable on T by some input-mediating M_2 given that:

- ***(C1)*** *$\mathscr{P}$ does not reason about the communication between the monitor and the target,*
- ***(C2)*** *$\forall\hat{P} \in \mathscr{P} : scheds(\hat{P})$ is quiescent forgiving for T,*
- ***(C3)*** *$\forall\hat{P} \in \mathscr{P} : scheds(\hat{P}) \subseteq scheds(T)$, and*
- ***(C4)*** *$\forall\hat{P} \in \mathscr{P}$: $\forall s \in scheds(T)$:*
  $s \notin scheds(\hat{P}) \Rightarrow \exists s' \preceq s$:
  $(s' \in scheds(\hat{P})) \wedge (s' = s''; a) \wedge (a \in input(T))$
  $\wedge\ (\forall t \succeq s' : t \in scheds(T) \Rightarrow t \notin scheds(\hat{P}))$.

Proof idea. The constraints of the theorem are explained below. The main proof of the theorem is by construction. For the right direction we construct M_2 by removing from M_1 any transitions that deal with outputs that the target wants to execute, and connect together the transition graph, if disconnected components were created from the removal of the transitions. For the left direction, we construct M_1 by using M_2 as a basis and adding transition that simply intercept output actions that the target wants to execute, and forward them to the environment. □

Constraint **C2** ensures that whatever the target chooses to output in the beginning of its execution, and until it blocks for some input, obeys the policy. Constraint **C3** ensures that the input/output-mediating monitor does not have an "unfair" advantage over the input mediating one, just because the policy requires from the monitor to output actions, even if the target would never perform them. Constraint **C4** ensures that whenever the target receives some input (from the monitor), then no behavior that it exhibits (until it blocks to wait for another input) will violate the policy, or, if it does, then that behavior can be suppressed without affecting the target's future behavior.

Thm. 7 is an illustration of how our framework can help in making sound decisions for designing and implementing run-time monitors in practice. For example, suppose we have a Unix kernel and want to enforce the policy that that secret file cannot be (a) deleted or (b) displayed to guest users. A monitor designer who wants to *precisely* enforce that policy cannot in general use an input-mediating monitor: although it can enforce (a) by not forwarding commands like "rm secret-file", it cannot enforce (b), because it does not know whether the kernel can, for example, correctly identify guest users and not display secret files to them. However, the designer can check if the specific kernel meets the constraints of Thm. 7. If it does, e.g., the kernel does not display any secret files while booting (i.e., **C2**), and does not display secret files to guest users, e.g., through a correct access-control mechanism (i.e., **C4**), then an input-mediating monitor suffices to enforce the policy. The correctness of such design choices might not always be obvious, and the above example demonstrates how our framework can aid in making more informed decisions. Moreover, such decisions can have benefits both in efficiency (by not monitoring the kernel's output sequence at run time), and in security (since the TCB/attack surface of the monitor is smaller).

Precise enforcement is a useful notion in situations where we want to guarantee that the monitor exhibits certain behaviors that may not be exhibited by the monitored target; e.g., always append to a file the date of modification before closing it (where append and close actions belong to the signature of the target), or log certain security relevant events (where logging actions do not belong to the signature of the target). In the first case, if we allow for transparent and sound enforcement, a monitor could simply buffer the open and writes to the file, and once a close action was issued it could simply discard the particular sequence of actions.

7 Related Work

The first model of run-time monitors, *security automata*, was based on Büchi Automata and introduced by Schneider [25]. Since then, several similar models have been proposed that extend or refine the class of enforceable policies based on the enforcement capabilities (i.e., operational semantics) and computational powers (e.g., finite or infinite state) of monitors (e.g., [19, 11, 4, 10]).

Another track of extensions of Schneider's work includes frameworks that model additional aspects of the monitoring and enforcement process, as opposed to the abilities and powers of the monitors themselves. These frameworks are orthogonal to the models of "computational" extensions. We can summarize the majority of these frameworks in three main axes: (1) Static Information, (2) Interaction, and (3) Incomplete Mediation.

Static Information Some frameworks extend Schneider's work to account for information that is not available to the monitor at run-time, e.g., information about the target obtained by, for example, static analysis (also identified as *non-uniformity* [19]). Hamlen et al. described a model based on Turing Machines [15], with which they compared the classes of policies enforceable by several types of enforcement mechanisms, such as static analysis and inlined monitors. Chabot et al. used Rabin automata to derive in-lined monitors that enforce policies on specific targets, and showed that non-uniform truncation monitors (i.e., monitors that consider only a subset of all possible executions that a target might exhibit) are strictly more powerful than uniform truncation monitors [7]. Our framework is more general than the above as it allows to formally reason about the communication between the target and the monitor.

Interaction Another line of frameworks focuses on modeling the interaction and communication interface between the target and the monitor. Such frameworks, either revise the "computational" models, or adopt alternate ones, such as the Calculus of Communicating Systems (CCS) [24] and Communicating Sequential Processes (CSP) [6], to more conveniently reason about applications, the interaction between applications and monitors, and enforcement in distributed systems. An example of revising existing models is Ligatti and Reddy's Mandatory Results Automata, which model the (synchronous) communication between the monitor and the target [20]. MRA's, however, do not model the target explicitly, and thus results about enforceable policies in target-specific environments might be difficult to derive. Among the works building on CCS or CSP is Martinelli and Matteucci's model of run-time monitors based on CCS [23]. Like ours, their model captures the communication between the monitor and the target, but their main focus is on synthesizing run-time monitors from policies. In contrast, we focus on a meta-theoretical analysis of enforcement in a more expressive framework. Basin et al. proposed a practical language, based on CSP and Object-Z (OZ), for specifying security automata [3]. This work focuses on the synchronization between a single monitor and target application, although the language is expressive enough to capture many other enforcement scenarios. Our

work is similar to Basin's, however we focus more on showing how such a more expressive framework can be used to derive meta-theoretical results on enforceable policies in different scenarios, instead of focusing on the (complementary aspect) of showing how to faithfully translate and model practical scenarios in such frameworks. Gay et al. introduced *service automata*, a framework based on CSP for enforcing security requirements in distributed systems at run time [13]. Although CSP provides the abstractions necessary to reason about specific targets and the communication with the monitor, such investigation and analysis is not the focus of that work.

Incomplete Mediation Concurrently with our work, Basin et al. introduced a model to reason about actions that (truncation) monitors cannot modify [2]. For instance, a monitor cannot control time, and thus if we model time as a specific type of actions (e.g., clock tick actions) a monitor can observe but not suppress them. In our context, an instance of uncontrollable actions are input actions to an I/O automaton: since I/O automata are input enabled they cannot prevent inputs from arriving. Both Basin and we have derived to similar theoretical results, namely that a policy is enforceable only if it does not prohibit traces that terminate in uncontrollable actions (cf. input forgiveness, Def. 8). Basin's framework can be embedded to ours by encoding clock ticks as input actions to the monitored target from the environment, and using a fairness definition that interleaves clock ticks appropriately with other actions. In addition, our framework allows for: (1) reasoning about (partial) knowledge about the target's behavior, (2) encoding more powerful monitors than truncation automata (e.g., edit automata), and (3) modeling actions that the monitor cannot even observe (either through incomplete re-writing/mediation or internal actions of the target). Although there are scenarios where encoding unobservable actions as uncontrollable ones suffices for certain formal analyses, faithful formal modeling of practical scenarios may require both types of actions explicitly.

In summary, the above research has focused individually on each of these axes. Our work uses I/O automata to establish an automata-based framework that allows reasoning about all these three axes, and presents several results that weave reasoning and results from these axes, thus extending each of the above previous work individually.

8 Conclusion

Formal models for run-time monitors have helped improve our understanding of the powers and limitations of enforcement mechanisms [25, 19], and aided in their design and implementation [9, 14]. However, these models often fail to capture many details and complexity relevant to real-world run-time monitors, such as how monitors integrate with targets, and the extent to which monitors can control targets and their environment.

In this paper, we propose a general framework, based on I/O automata, for reasoning about policies, monitoring, and enforcement. This framework provides

abstractions for reasoning about many practically relevant details important for run-time enforcement, and, in general, yields a richer view of monitors and applications than is typical in previous analyses of run-time monitoring. Moreover, we show how this framework can be used for meta-theoretic analysis of enforceable security policies. In particular, we derive results that describe lower bounds on enforceable policies that are independent of the particular choice of monitor (Thm. 3). We also identify constraints under which monitors with different monitoring and enforcement capabilities (i.e., monitors that see only a subset of the target's actions; and monitors that have more or less ability to correct a target's invalid behavior) can enforce the same classes of policies (Thm. 7).

References

1. de Alfaro, L., Henzinger, T.A.: Interface automata. In: Proceedings of the 8th European software engineering conference. pp. 109–120 (2001)
2. Basin, D., Jugé, V., Klaedtke, F., Zălinescu, E.: Enforceable security policies revisited. ACM Trans. Inf. Syst. Secur. 16(1), 3:1–3:26 (Jun 2013)
3. Basin, D., Olderog, E.R., Sevinc, P.E.: Specifying and analyzing security automata using CSP-OZ. In: Proceedings of the 2nd ACM symposium on Information, computer and communications security. pp. 70–81. ASIACCS '07 (2007)
4. Bielova, N., Massacci, F.: Do you really mean what you actually enforced? IJIS pp. 1–16 (2011), http://dx.doi.org/10.1007/s10207-011-0137-2
5. Bishop, M.: Computer Security: Art and Science. Addison-Wesley Professional (2002)
6. Brookes, S.D., Hoare, C.A.R., Roscoe, A.W.: A theory of communicating sequential processes. Journal of the ACM 31, 560–599 (June 1984)
7. Chabot, H., Khoury, R., Tawbi, N.: Extending the enforcement power of truncation monitors using static analysis. Computers and Security 30(4), 194 – 207 (2011)
8. Clarkson, M.R., Schneider, F.B.: Hyperproperties. In: IEEE Computer Security Foundations Symposium (2008)
9. Erlingsson, U., Schneider, F.B.: SASI enforcement of security policies: a retrospective. In: Workshop on New security paradigms (2000)
10. Falcone, Y., Fernandez, J.C., Mounier, L.: What can you verify and enforce at runtime? International Journal on Software Tools for Technology Transfer (STTT) pp. 1–34 (2011)
11. Fong, P.W.L.: Access control by tracking shallow execution history. In: Proceedings of the 2004 IEEE Symposium on Security and Privacy. pp. 43–55 (2004)
12. Garfinkel, T.: Traps and pitfalls: Practical problems in in system call interposition based security tools. In: Network and Distributed Systems Security Symposium (2003)
13. Gay, R., Mantel, H., Sprick, B.: Service automata. In: 8th International Workshop on Formal Aspects of Security and Trust (2011)
14. Hamlen, K.: Security policy enforcement by automated program-rewriting. Ph.D. thesis, Cornell University (2006)
15. Hamlen, K.W., Morrisett, G., Schneider, F.B.: Computability classes for enforcement mechanisms. ACM Trans. Program. Lang. Syst. 28(1), 175–205 (2006)
16. H.Salzer, J., Schroeder, M.D.: The protection of information in computer systems. In: Fourth ACM Symposium on Operating System Principles (Mar 1973)

17. Kwiatkowska, M.: Survey of fairness notions. Information and Software Technology 31(7), 371 – 386 (1989)
18. Ligatti, J., Bauer, L., Walker, D.: Enforcing non-safety security policies with program monitors. In: European Symposium on Research in Computer Security (ESORICS). vol. 3679, pp. 355–373 (2005)
19. Ligatti, J., Bauer, L., Walker, D.: Run-time enforcement of nonsafety policies. ACM Transactions on Information and System Security 12(3) (2009)
20. Ligatti, J., Reddy, S.: A theory of runtime enforcement, with results. In: European Symposium on Research in Computer Security (ESORICS) (2010)
21. Lynch, N.A.: Distributed Algorithms. Morgan Kaufmann Publishers Inc. (1996)
22. Lynch, N.A., Tuttle, M.R.: Hierarchical correctness proofs for distributed algorithms. In: PODC '87: Proceedings of the 6th ACM Symposium on Principles of Distributed Computing (1987)
23. Martinelli, F., Matteucci, I.: Through modeling to synthesis of security automata. Electron. Notes Theor. Comput. Sci. 179, 31–46 (July 2007)
24. Milner, R.: A Calculus of Communicating Systems. Springer-Verlag New York, Inc. (1982)
25. Schneider, F.B.: Enforceable security policies. ACM Trans. Inf. Syst. Secur. 3(1), 30–50 (2000)
26. Tuttle, M.R.: Hierarchical correctness proofs for distributed algorithms. Master's thesis, Dept. of Electrical Engineering and Computer Science, MIT (1987)
27. Wagner, D.A.: Janus: an approach for confinement of untrusted applications. Tech. Rep. UCB/CSD-99-1056, EECS, University of California, Berkeley (1999)

Appendix A. I/O Automata Theorems

This appendix contains the formal expression of a collection of several important I/O automata theorems. Their discussion and proofs can be found in [22, 26, 21].

Theorem 8. *An execution of a composition induces executions of the component automata*
Let $\{A_i\}_{i\in I}$ *be a collection of compatible automata and let* $A = \prod_{i\in I} A_i$*. If* $\alpha \in execs(A)$ *then* $\alpha|A_i \in execs(A_i)$ *for every* $i \in I$*. Moreover, the same results holds for scheds(), traces() and their fair versions.*

Theorem 9. *Executions of component automata can often be pasted together to form an execution of the composition*
Let $\{A_i\}_{i\in I}$ *be a collection of compatible automata and let* $A = \prod_{i\in I} A_i$*. Suppose* α_i *is an execution of* A_i *for every* $i \in I$*, and suppose* β *is a sequence of actions in* $acts(A)$ *such that* $\beta|A_i = sched(\alpha_i)$ *for every* $i \in I$*. Then there is an execution* α *of* A *such that* $\beta = sched(\alpha)$ *and* $\alpha_{=}\alpha|A_i$ *for every* $i \in I$*. Moreover the same holds for external() and traces() instead of acts() and sched(), and for fair executions.*

Theorem 10. *Schedules and traces of component automata can be pasted together to form schedules and traces of the composition*
Let $\{A_i\}_{i\in I}$ *be a collection of compatible automata and let* $A = \prod_{i\in I} A_i$*. Let* β *be a sequence of actions in* $acts(A)$*. If* $\beta|A_i \in scheds(A_i)$ *for every* $i \in I$*, then* $\beta \in scheds(A)$*.Moreover the same holds for external() and traces() instead of acts() and sched(), and for their fair versions.*

Theorem 11. *Composition of modules correspond to composition of automata*
Let $\{A_i\}_{i\in I}$ *be a collection of compatible automata. Then* $execs(\prod_{i\in I} A_i) = \prod_{i\in I} execs(A_i)$, $scheds(\prod_{i\in I} A_i) = \prod_{i\in I} scheds(A_i)$, $traces(\prod_{i\in I} A_i) = \prod_{i\in I} traces(A_i)$, $fairexecs(\prod_{i\in I} A_i) = \prod_{i\in I} fairexecs(A_i)$, $fairscheds(\prod_{i\in I} A_i) = \prod_{i\in I} fairscheds(A_i)$, $fairtraces(\prod_{i\in I} A_i) = \prod_{i\in I} fairtraces(A_i)$. *Moreover the same hold for execution, schedule and trace modules.*

Theorem 12. *Commutativity and Associativity of composition of single components*
Let A, B *and* C *I/O automata. Then:*

1. $A \times B = B \times A$
2. $(A \times B) \times C = A \times (B \times C) = A \times B \times C$

Theorem 13. *Commutativity, Associativity and Congruence of automata and modules*
Let $A = \prod_i A_i$, $B = \prod_i B_i$ *and* $C = \prod_i C_i$ *and* $D = \prod_i D_i$ *where* A_i, B_i, C_i *and* D_i *are either I/O automata or modules. Then:*

1. $A \times B = B \times A$
2. $(A \times B) \times C = A \times (B \times C) = A \times B \times C$
3. *if* $A = B$ *and* $C = D$, *then* $A \times C = B \times D$ *whenever* $A \times B$ *and* $C \times D$ *are defined.*

Theorem 14. *Hiding on Automata passes to modules*
For all automata A, *execution modules* E, *schedule modules* S *and sets of actions* Σ*:*

1. $execs(\mathsf{hide}_\Sigma(A)) = \mathsf{hide}_\Sigma(execs(A))$
2. $scheds(\mathsf{hide}_\Sigma(E)) = \mathsf{hide}_\Sigma(scheds(E))$
3. $traces(\mathsf{hide}_\Sigma(S)) = \mathsf{hide}_\Sigma(traces(S))$

Notice that the last two versions, also hold for automata besides execution and trace modules.

Theorem 15. *Hiding of composition corresponds to hiding of components*
Let $\{M_i : i \in I\}$ *be a collection of compatible automata or modules, and let* $\{\Sigma_i : i \in I\}$ *be a collection of sets of actions. If* $acts(M_i)$ *and* Σ_j *are disjoint for all* $i \neq j$ *then:* $\mathsf{hide}_{\cup_i \Sigma_i}(\prod_{i\in I} M_i) = \prod_{i\in I} \mathsf{hide}_{\Sigma_i}(M_i)$.

Theorem 16. *Renaming of composition corresponds to renaming of components*
Let $\{M_i : i \in I\}$ *be a collection of compatible automata or modules, and let* $\{\mathsf{rename}_i : i \in I\}$ *be compatible action mappings. If* rename_i *is applicable to* M_i *for every* $i \in I$, *then* $(\prod_{i\in I} \mathsf{rename}_i)(\prod_{i\in I} M_i) = \prod_{i\in I} \mathsf{rename}_i M_i$.

Theorem 17. *Renaming on Automata passes to modules*
Let rename *be a renaming applicable to the automaton* A, *execution module* E *and schedule module* S*:*

1. $execs(\mathsf{rename}(A)) = \mathsf{rename}(execs(A))$
2. $scheds(\mathsf{rename}(E)) = \mathsf{rename}(scheds(E))$
3. $traces(\mathsf{rename}(S)) = \mathsf{rename}(traces(S))$.

Theorem 18. *Sequence of hiding and renaming does not matter* $\mathsf{hide}_{f(\Sigma)}(f(M)) = f(\mathsf{hide}_{\Sigma}(M))$ *for any automaton or module* M *and applicable renaming* f.

Appendix B. Proofs of Theorems

Proposition 1. Given a monitor M then:

1. $\forall \mathscr{P}$: $\mathscr{P}$ is *generally soundly enforceable* by $M \Rightarrow$
   $\forall T$: $\mathscr{P}$ is *specifically soundly enforceable* on T by M, and
2. $\exists \mathscr{P} \exists T$: ($\mathscr{P}$ is *specifically soundly enforceable* on T by M)$\wedge$
   $\neg$($\mathscr{P}$ is *generally soundly enforceable* by M).

Proof sketch. For *(1)*, we assume that we have an arbitrary $\mathscr{P}$ that is generally soundly enforceable by some M. By Def. 12, this means that:

> for all targets T there exists a module $\hat{P} \in \mathscr{P}$, a renaming function rename, such that $(scheds(M \times \mathsf{rename}(T))|acts(\hat{P})) \subseteq scheds(\hat{P})$. **(A)**

We have to show that for all targets T', $\mathscr{P}$ is specifically soundly enforceable on T' by M, which by Def. 11 means that we have to show that for some arbitrary T' there exists a module $\hat{P}' \in \mathscr{P}$, a renaming function rename', such that $(scheds(M \times \mathsf{rename}'(T'))|acts(\hat{P}')) \subseteq scheds(\hat{P}')$.

By **(A)** we know that there are $\hat{P}$ and rename that correspond to any T, and thus for T'. Use the corresponding choices of $\hat{P}$ and rename for T' and our claim follows from **(A)** immediately.

For *(2)*, we must exhibit a $\mathscr{P}$ and a T such that $\mathscr{P}$ is *specifically soundly enforceable* on T by M and it is not the case that $\mathscr{P}$ is *generally soundly enforceable* by M.

Let $\mathscr{P} = \{scheds(M) \cup \{\langle a \rangle \mid a \in \Sigma - acts(M)\}\}$. Also, let T be the trivial automaton, i.e., the I/O automaton with the empty set for actions and just a single start state. Thus, $scheds(T) = \{\epsilon\}$. It is easy to see that $\mathscr{P}$ is specifically soundly enforceable on T by M, i.e., that there exists a module $\hat{P} \in \mathscr{P}$, a renaming function rename, such that $(scheds(M \times \mathsf{rename}(T))|acts(\hat{P})) \subseteq scheds(\hat{P})$. $\mathscr{P}$ contains only one element, so $\hat{P} = scheds(M) \cup \{\langle a \rangle \mid a \in \Sigma - acts(M)\}$ which contains all schedules that M can produce. Moreover, let rename be the identity function. From Thm. 8 we know that $scheds(M \times \mathsf{rename}(T))$ will be the pasting of the schedules of the two components, and since the schedules of the component $\mathsf{rename}(T)$ is just the empty sequence, $scheds(M \times \mathsf{rename}(T)) = scheds(M)$. So we have to show that $scheds(M)|acts(\hat{P})) \subseteq scheds(M) \cup \{\langle a \rangle \mid a \in \Sigma - acts(M)\}$, which is trivially true.

To prove the second conjunct of the claim, i.e., that it is not the case that $\mathscr{P}$ is *generally soundly enforceable* by M, pick any T' that has as a signature only one output action, and produces some finite sequence of repetitions of this action of length greater than 1; i.e., $scheds(T') = \{(a;a)^n \mid n \geq 1 \text{ and } a \in output(T')\}$. Note that no matter how we rename T', its renamed output actions will still be an action of $\hat{P}$, since we added all actions that are not actions of the monitor $(\Sigma - acts(M))$. Using Thm. 8 again, we see that the schedules of the composition will contain schedules of the component $\mathsf{rename}(T')$, which means that there is some sequence $s = (a;a) \in scheds(T')$, where $a \in acts(\mathsf{rename}(T'))$. But $s \notin scheds(\hat{P})$ because $s \notin scheds(M)$ (since M and T' have disjoint sets of output actions by definition of composition of I/O automata), and $s \notin \{\langle a \rangle \mid a \in \Sigma - acts(M)\}$ since s has length > 1. This concludes the proof of our claim. □

Theorem 1.

$$\forall \mathscr{P} : \forall T\text{: } \exists M : \mathscr{P} \text{ is } \textit{specifically soundly enforceable} \text{ on } T \text{ by } M \Leftrightarrow \exists M' : \mathscr{P} \text{ is } \textit{generally soundly enforceable} \text{ by } M'.$$

Proof sketch. ($\Rightarrow$ *direction*) We assume that we are given a policy $\mathscr{P}$ and a target T such that $\mathscr{P}$ is soundly enforceable on T by some monitor M. That is, we assume that there exists a module $\hat{P} \in \mathscr{P}$, a renaming function rename, and a hiding function hide for some set of actions Φ such that $(scheds(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))|acts(\hat{P})) \subseteq scheds(\hat{P})$.

We have to show that $\mathscr{P}$ is generally soundly enforceable by some monitor M', or, by definition, that there exists monitor M' such that for all targets T' there exists a module $\hat{P'} \in \mathscr{P}$, a renaming function rename', and a hiding function hide' such that $(scheds(\mathsf{hide}'_\Phi(M' \times \mathsf{rename}'(T')))|acts(\hat{P'})) \subseteq scheds(\hat{P'})$.

Let:

1. $M' = \mathsf{hide}_\Phi(M \times \mathsf{rename}(T))$,
2. $\hat{P'} = \hat{P}$,
3. rename' be a function that maps a to a' where $a \in acts(T'), a' \notin acts(\hat{P})$,
4. $\mathsf{hide}'_\Phi = \mathsf{hide}_\emptyset$.

Now it is easy to see that:

$(scheds(\mathsf{hide}'_\Phi(M' \times \mathsf{rename}'(T')))|acts(\hat{P'})) \subseteq scheds(\hat{P'})$
$\Leftrightarrow (scheds(\mathsf{hide}_\emptyset(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)) \times \mathsf{rename}'(T')))|acts(\hat{P})) \subseteq scheds(\hat{P})$ (by substitution)
$\Leftrightarrow (scheds(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)) \times \mathsf{rename}'(T'))|acts(\hat{P})) \subseteq scheds(\hat{P})$ (by definition of hiding and the fact that $\Phi = \emptyset$)
$\Leftrightarrow (scheds(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))|acts\hat{P}) \times (scheds(\mathsf{rename}'(T'))|acts(\hat{P})) \subseteq scheds(\hat{P})$ (by Theorems 5 and 7 in App. A)
$\Leftrightarrow (scheds(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))|acts\hat{P}) \times \epsilon \subseteq scheds(\hat{P})$ (by definition of rename' and operator $|$)
$\Leftrightarrow (scheds(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))|acts\hat{P}) \subseteq scheds(\hat{P})$ (by Theorem 7 in App. A)
Note that the last line is true from our assumption, so we are done.

($\Leftarrow$ *direction*) We assume that we are given a policy $\mathscr{P}$ and a target T. Moreover we assume that $\mathscr{P}$ is generally soundly enforceable by some monitor M'.

That is, by definition, we assume that for all targets T there exists a module $\hat{P} \in \mathscr{P}$, a renaming function rename, and a hiding function hide such that $(scheds(\mathsf{hide}_{\Phi}(M \times \mathsf{rename}(T)))|acts(\hat{P})) \subseteq scheds(\hat{P})$

We have to show that $\mathscr{P}$ is soundly enforceable on T by some monitor M. That is, we have to show that there exists a module $\hat{P}' \in \mathscr{P}$, a renaming function rename, and a hiding function hide for some set of actions Φ such that $(scheds(\mathsf{hide}_{\Phi}(M \times \mathsf{rename}(T)))|acts(\hat{P}')) \subseteq scheds(\hat{P}')$.

This is trivially true, since we can use the module, renaming function, hiding function, and monitor from our assumptions. Since the subset relationship is satisfied for every target, it is also trivially satisfied by T.

□

Theorem 2.

$\exists \mathscr{P} : \exists \mathcal{R} : \exists \mathcal{T} : \exists T \in \mathcal{T}$:
$\exists M : \mathscr{P}$ is *specifically maximally soundly enforceable* on T by M using $\mathcal{R} \wedge$
$\nexists M' : \mathscr{P}$ is *generally maximally soundly enforceable* on $\mathcal{T}$ by M' using $\mathcal{R}$.

Proof sketch. This is a proof by construction. We will construct a policy given a set of targets and renaming functions such that the main body of the the theorem holds.

Given an I/O automaton A, let $\mathcal{X}(A) = \{s \mid s$ contains only internal actions of $A\}$.

Let T_1 and T_2 be targets with $sig(T_1) \neq sig(T_2)$, $scheds(T_1) \neq \emptyset$, $scheds(T_2) \neq \emptyset$, and $\mathcal{X}(T_2) \neq \emptyset$.

Let $\mathcal{T} = \{T_1, T_2\}$, $\mathcal{R} = \{id\}$, i.e., the only renaming function allowed is the identity function.

Now we construct the policy $\mathscr{P} = \{\hat{P}_1, \hat{P}_2\}$, where $\hat{P}_1 = \langle sig(T_1), scheds(T_1)\rangle$, and $\hat{P}_2 = \langle sig(T_2), scheds(T_2) - \mathcal{X}(T_2)\rangle$, i.e., $\hat{P}_2$ disallows any schedule of T_2 that contains only internal actions of T_2.

Now, pick $T_1 \in \mathcal{T}$. It is easy to see that there exists a monitor M, the trivial monitor with the empty signature that does nothing, such that (1) $sig(\hat{P}) = sig(T_1) \cup sig(M) \cup range(id)$, and (2) $scheds(M \times id(T_1)) \subseteq scheds(\hat{P}_1)$. By simple syntactic manipulations we get (1) $sig(T_1) = sig(T_1)$, and (2) $scheds(T_1) \subseteq scheds(T_1)$, which is trivially true.

Now, it suffices to show that there is no monitor M' such that $\mathscr{P}$ is generally maximally soundly enforceable by M'. When trying to see whether $\mathscr{P}$ is enforceable for T_2, the first constraint of the definition of general maximal enforcement forces us to choose $\hat{P}_2$: it is the only module that matches the signature or T_2.

But, by construction, $\hat{P}_2$ disallows any schedule that contains internal actions of T_2, and T_2 can exhibit such schedules. Moreover, since our only renaming function available is the identity, there is no monitor M' that can prohibit these internal actions from happening: I/O automata composition do not allow for one component automaton to control the local actions of any other component. Thus, for any M', the monitored target $M' \times id(T_2)$ will exhibit schedules that belong to $\mathcal{X}(T_2)$ and thus the subset relation $scheds(M' \times id(T_2)) \subseteq scheds(\hat{P}_2)$

does not hold. Thus $\mathscr{P}$ is not generally maximally soundly enforceable, and this completes the proof of the theorem.

□

Theorem 3. $\forall \mathscr{P} : \forall \hat{P} \in \mathscr{P} : \forall T : \forall \mathsf{rename} :$
$\exists M : (\mathit{scheds}(M \times \mathsf{rename}(T))|\mathit{acts}(\hat{P})) = \mathit{scheds}(\hat{P}) \Rightarrow$
$\hat{P}$ is input forgiving, safety, and quiescent forgiving for $\mathsf{rename}(T)$.

Proof sketch. We fix a policy $\mathscr{P}$, a module $\hat{P}$, a hiding function $\mathsf{hide}_\Phi()$, and a renaming function $\mathsf{rename}()$, and we assume that there exists a monitor M such that $\Big(\mathit{scheds}(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))|\mathit{acts}(\hat{P}) \subseteq \mathit{scheds}(\hat{P})\Big)$. We have to show that $\hat{P}$ is input forgiving, prefix closed, and quiescent forgiving for $\mathsf{rename}(T)$.

For the sake of contradiction, assume that $\hat{P}$ is not input forgiving, or not prefix closed, or not quiescent forgiving for $\mathsf{rename}(T)$.

Case: $\hat{P}$ is not input forgiving:

Since $\hat{P}$ is not input forgiving, then either $\epsilon \notin \mathit{scheds}(\hat{P})$ or there exists an $s_1 \in \mathit{scheds}(\hat{P})$, a finite prefix s_2 of s_1, and some sequence of input actions s_3 such that $(s_2; s_3) \notin \mathit{scheds}(\hat{P})$. If we assume the first case of $\epsilon \notin \mathit{scheds}(\hat{P})$ we derive a contradiction since the empty sequence belongs to the schedules of any I/O automaton by definition of executions and schedules of I/O automata. If we assume the latter case, then we know that $s_2 \in \mathit{scheds}(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))$ by assumption. Let q_n be the state that the monitored target is at after executing the last action of s_2. By definition, every state of an I/O automaton is input enabled. Thus q_n is input enabled, which means that $\forall s' \in (\mathit{input}(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T))))^\infty$: $(s_2; s') \in \mathit{scheds}(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))$ (remember we assume no fairness thus it does not matter whether q_n is quiescent or not). But for $s' = s_3$ we get that $(s_2; s_3) \in \mathit{scheds}(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))$ and that $(s_2; s_3) \notin \mathit{scheds}(\hat{P})$ which contradicts our assumption. Thus, in both cases we derived a contradiction, and thus $\hat{P}$ must be input forgiving.

Case: $\hat{P}$ is not prefix closed:

Since $\hat{P}$ is not prefix closed, then there exists some schedule s_1 that belongs to the schedules of $\hat{P}$, but there exists some prefix s_2 of s_1 that does not belong to the schedule of $\hat{P}$, or more formally: $\exists s_1 \in \Sigma^\infty : \big(s_1 \in \mathit{scheds}(\hat{P})\big) \wedge \big(\exists s_2 \in \Sigma^\star : s_2 \preceq s_1 : s_2 \notin \mathit{scheds}(\hat{P})\big)$.

Without loss of generality, assume s_2 is the longest strict prefix of s_1, i.e., it is the longest prefix of s_1 that does not belong to the schedules of $\hat{P}$, and that all prefixes of s_2 belong to the schedules of $\hat{P}$. If $s_2 = a_1, \ldots, a_{n-1}, a_n$ then let $s_2^- = a_1, \ldots, a_{n-1}$ and $s_2^+ = a_1, \ldots, a_{n-1}, a_n, a_{n+1} \preceq s_1$. We know that $s_2^- \in \mathit{scheds}(\hat{P})$ by assumption, $s_2^+ \in \mathit{scheds}(\hat{P})$ because if it was not in the schedules of the property this would be the longest invalid prefix of s_1 which contradicts our choice of s_2, and thus by assumption they both also belong to the schedules of the monitored target $\mathsf{hide}_\Phi(M \times \mathsf{rename}(T))$. Let q_n be the state that the monitored target is after executing a_{n-1}, and q_{n+1} be the state before

executing a_{n+1}. In order for the automaton to transition from q_n to q_{n+1} it must execute some a_n. But then $s_2^-; a_n = s_2 \in \mathit{scheds}(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))$, while we assumed that $s_2 \notin \mathit{scheds}(\hat{P})$. This contradicts our original assumption, and thus $\hat{P}$ must be prefix closed.

Case: $\hat{P}$ is not quiescent forgiving:

Since $\hat{P}$ is not quiescent forgiving for $\mathsf{rename}(T)$, then there exists some execution $e = q_0, a_1, \ldots, q_n$ of $\mathsf{rename}(T)$ with $q_n \in \mathit{quiescent}(T)$ and $q_i \notin \mathit{quiescent}(T)$ for $0 \leq i < n$ such that either $(\mathit{sched}(e)|\mathit{acts}(\hat{P})) \notin \mathit{scheds}(\hat{P})$ or some prefix t of $(\mathit{sched}(e)|\mathit{acts}(\hat{P}))$ does not belong to the schedules of $\hat{P}$.

By Theorem 7 we know that if $\mathit{sched}(e) \in \mathit{scheds}(\mathsf{rename}(T))$, then $\mathit{sched}(e) \in \mathit{scheds}(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))$. Thus, $(\mathit{sched}(e)|\mathit{acts}\hat{P}) \in \mathit{scheds}(\mathsf{hide}_\Phi(M \times \mathsf{rename}(T)))$. But the fact that $(\mathit{sched}(e)|\mathit{acts}(\hat{P})) \notin \mathit{scheds}(\hat{P})$ contradicts our assumption. With the same argument we can show that even if we assume some prefix t of $\mathit{sched}(e)$ we also derive a contradiction. Thus $\hat{P}$ must be quiescent forgiving.

□

Theorem 4. $\forall \mathscr{P} : \forall \hat{P} \in \mathscr{P} : \forall T : \forall M : \forall \hat{P}_T : \forall \mathsf{rename} :$
if $\hat{P}$ is strongly transparent for $M \times \mathsf{rename}(T)$ w.r.t. $\hat{P}_T$ then: (1) there is no monitor M that specifically precisely enforces $\mathscr{P}$ on T using $\hat{P}$, and (2) if there exists M that fairly specifically precisely enforces $\mathscr{P}$ on T using $\hat{P}$ then $\hat{P}$ is input forgiving, and quiescent forgiving for $\mathsf{rename}(T)$.

Proof sketch. By contradiction. First, pick a policy $\mathscr{P}$ with only one non-trivial element $\hat{P}$ and a non-trivial module $\hat{P}_T$, i.e., $\mathit{scheds}(\hat{P}) \neq \emptyset$ and $\mathit{scheds}(\hat{P}_T) \neq \emptyset$. Now, let $\hat{P}$ contain only "transparent" schedules, i.e., $\forall t \in \hat{P}_T : \exists s \in \mathit{scheds}(\hat{P}) : \mathit{ren}^{-1}(s|\mathsf{range}(\mathit{ren})) = (s|\mathsf{dom}(\mathit{ren})) = t$, and $\hat{P}$ does not contain any other schedules besides the ones specified above.

Proof of (1): Assume there exists monitor M that specifically precisely enforces $\mathscr{P}$ on T using $\hat{P}$. By Thm. 3, $\hat{P}$ must be prefix closed. But also, by construction, $\forall s \in \mathit{scheds}(\hat{P}) : \exists t \in \hat{P}_T : \mathit{ren}^{-1}(s|\mathsf{range}(\mathit{ren})) = (s|\mathsf{dom}(\mathit{ren})) = t$.

Since $\mathit{scheds}(\hat{P}_T)$ and $\mathit{scheds}(\hat{P})$ are non-empty pick $s \in \mathit{scheds}(\hat{P})$ such that $\exists t \in \hat{P}_T$ such that $\mathit{ren}^{-1}(s|\mathsf{range}(\mathit{ren})) = (s|\mathsf{dom}(\mathit{ren})) = t$. Now since we assumed that $\hat{P}$ is prefix closed, then if $s = a_1; a_2; \ldots; a_n; a_{n+1}$, then $s' = a_1; a_2; \ldots; a_n$ must also belong to $\hat{P}$. But then it is easy to see by a simple counting argument that there is no $t' \in \hat{P}_T$ such that $\mathit{ren}^{-1}(s'|\mathsf{range}(\mathit{ren})) = (s'|\mathsf{dom}(\mathit{ren})) = t'$: s' contains one less action than s, whereas it should contain two less actions for the above equality to hold.

Thus we have a contradiction, and no such M can exist that specifically precisely enforce $\mathscr{P}$ on T using $\hat{P}$.

Proof of (2): Similar to the proof of Thm. 3 for the cases of input-forgiving and quiescent forgiving, but substituting occurrences of $\mathit{scheds}()$ with $\mathit{fairscheds}()$.

□

Proposition 2. $\forall \mathscr{P} : \forall \hat{P} \in \mathscr{P}$, such that $\epsilon \notin \mathit{scheds}(\hat{P})$:
$\forall T : \forall M : \forall \hat{P}_T$ with $\mathit{scheds}(\hat{P}_T) \neq \emptyset$: $\forall \mathsf{rename}$:
if $\hat{P}$ is strongly transparent for $M \times \mathsf{rename}(T)$ w.r.t. $\hat{P}_T$ then there is no monitor M that specifically soundly enforces $\mathscr{P}$ on T using $\hat{P}$.

Proof sketch. By Def. 6, of strong transparency, it is easy to see that every schedule that belongs to $\hat{P}$ and contains some behavior of the target that belongs to $\hat{P}_T$ must have even length. But as described in the proof of Thm. 4 since the schedules of the monitored target are prefix closed (i.e., safety), the monitored target will exhibit schedules (of odd length) that do not belong to $\hat{P}$. In Thm. 4 that was enough to contradict the theorem statement because we were proving precise enforcement, i.e., equality.

Here, there is one case that the monitor can soundly enforce the policy (i.e., subset relation): the monitored target does nothing, i.e., it exhibits a schedule with no actions (which has an even length, i.e., 0). But, by assumption $\epsilon \notin \hat{P}$. Thus, this corner case cannot happen under the assumptions of the theorem statement and we conclude that there is no monitor that specifically soundly enforces $\mathscr{P}$ on T using $\hat{P}$.

□

Theorem 5. $\forall \mathscr{P} : \forall \hat{P} \in \mathscr{P} : \forall T : \forall M : \forall \hat{P}_T : \forall \mathsf{rename}$:
if $\hat{P}$ is weakly transparent for $M \times \mathsf{rename}(T)$ w.r.t. $\hat{P}_T$ then: (1) if there exists monitor M that specifically precisely enforces $\mathscr{P}$ on T using $\hat{P}$, then $\hat{P}$ is input forgiving, safety, and quiescent forgiving for $\mathsf{rename}(T)$, and (2) if there exists monitor M that fairly specifically precisely enforces $\mathscr{P}$ on T using $\hat{P}$ then $\hat{P}$ is input forgiving, and quiescent forgiving for $\mathsf{rename}(T)$.

Proof sketch. Proof of (1): Direct application of Thm. 3.

Proof of (2): Similar to the proof of statement (2) of Thm. 4.

□

Theorem 6. $\exists \mathscr{P}$:
($\mathscr{P}$ is generally precisely enforceable by some input/output-mediating M_1)$\wedge$
$\neg$($\mathscr{P}$ is generally precisely enforceable by some input-mediating M_2), if the policy does not reason about the communication between the monitor and the target[10].

Proof sketch. Take $\mathscr{P} = \{\hat{P}\}$, where $\mathit{acts}(\hat{P}) = \mathit{output}(\hat{P}) = \bigcup_{i \in I} \mathit{output}(T_i) \cup \bigcup_{i \in I} \mathsf{rename}_{\mathsf{j} \in \mathsf{J}}(\mathit{output}(T_i))$, $\mathit{scheds}(\hat{P}) = \{\epsilon\} \cup \{\langle a \rangle \mid a \in \mathit{acts}(\hat{P})\}$, and $\hat{P}$ does not reason about the communication between the monitor and the target where I is the set of all targets, and J the set of all renaming functions (note that in the rest of the proof, for purposes of brevity of presentation, we are assuming that the universes of Input, Output, Internal actions are disjoint, and that renaming functions always map actions to fresh actions that are distinct from the Input, Output, and Internal actions of the targets).

[10] If we used trace enforcement, the constraint would be superfluous. We used schedules to remain consistent with other theorems in the paper.

For proving the left conjunct of the theorem statement, we have to prove that there exists an input/output-mediating M_1 such that for all targets T there exists a module $\hat{P} \in \mathscr{P}$, a renaming function rename, such that $(\mathit{scheds}(M_1 \times \mathsf{rename}(T))|\mathit{acts}(\hat{P})) = \mathit{scheds}(\hat{P})$.

Let M_1 be the input/output-mediating monitor that has as elements of its signature the following sets: $\mathit{input}(M_1) = \{\bigcup_{i\in I} \mathit{input}(T_i)\} \cup \{\bigcup_{i\in I} \mathsf{rename}_{\mathsf{j}\in\mathsf{J}}(\mathit{output}(T_i))\}$, $\mathit{output}(M_1) = \{\bigcup_{i\in I} \mathit{output}(T_i)\} \cup \{\bigcup_{i\in I} \mathsf{rename}_{\mathsf{j}\in\mathsf{J}}(\mathit{input}(T_i))\}$, $\mathit{internal}(M_1) = \emptyset$, where I is the set of all targets, and J the set of all renaming functions.

Moreover, let $\mathit{scheds}(M_1)$ contain no schedules that include more than one output actions from the subset $\{\bigcup_{i\in I} \mathit{output}(T_i)\}$, i.e., the monitor does not exhibit any output behavior to the environment that contains more than one action. This is easy to do: just exhibit the first valid output action that the target wants to execute, and suppress all future attempts. Now it is easy to see that for all targets T there exists a module $\hat{P} \in \mathscr{P}$, a renaming function rename, such that $(\mathit{scheds}(M_1 \times \mathsf{rename}(T))|\mathit{acts}(\hat{P})) = \mathit{scheds}(\hat{P})$: assume otherwise, i.e., there exists a schedule $s \in \mathit{scheds}(M_1 \times \mathsf{rename}(T))$ that is not an element of $\mathit{scheds}(\hat{P})$. Since $\mathit{scheds}(\hat{P})$ contains all possible sequences of length one that contain the output actions of all targets (and all their possible renamings), the only way for $(s \mid \mathit{acts}(\hat{P}))$ not to be an element of the schedules of $\hat{P}$ is to contain output actions and have length larger than 1. However, this is impossible by (1) construction of the monitor, and (2) assumption that the policy does not reason about the communication between the monitor and the target (by Def. 13, all output actions of the target are mediated by the monitor and thus considered part of their communication).

For proving the right conjunct of the theorem statement, we have to prove that it is not the case that there exists an input-mediating M_2 such that for all targets T there exists a module $\hat{P} \in \mathscr{P}$, a renaming function rename, such that $(\mathit{scheds}(M_2 \times \mathsf{rename}(T))|\mathit{acts}(\hat{P})) = \mathit{scheds}(\hat{P})$. In other words, we have to prove that for all input-mediating M_2, there exists a target T, such that for all modules $\hat{P} \in \mathscr{P}$, and for all renaming functions rename: $(\mathit{scheds}(M_2 \times \mathsf{rename}(T))|\mathit{acts}(\hat{P})) \neq \mathit{scheds}(\hat{P})$.

To prove the claim, take any target T such that $\exists s \in \mathit{scheds}(T)$, and s contains more than two output actions. Let s' be the schedule of the renamed target $\mathsf{rename}(T)$ that corresponds to s. Then, by Thm. 8 s' is contained in $(\mathit{scheds}(M_2 \times \mathsf{rename}(T))$ (since the output actions of the target and the monitor are disjoint). Also, s, and thus s' contain more than two output actions. Moreover, by definition of $\hat{P}_i$, $\mathit{acts}(\hat{P}) = \mathit{output}(\hat{P}_i) \supseteq \mathit{output}(\mathsf{rename}(T))$, for any rename function. And since every element $\hat{P}_i$ of $\mathscr{P}$ does not contain any schedules with more than two output actions, $(s' \mid \mathit{acts}(\hat{P})) \notin \mathit{scheds}(\hat{P})$. This concludes the proof of the claim.

□

Theorem 7. $\forall \mathscr{P} : \forall T :$
$\mathscr{P}$ is specifically precisely enforceable on T by some input/output-mediating M_1 iff $\mathscr{P}$ is specifically precisely enforceable on T by some input-mediating M_2

given that:

(**C1**) $\mathscr{P}$ does not reason about the communication between the monitor and the target,

(**C2**) $\forall \hat{P} \in \mathscr{P} : \mathit{scheds}(\hat{P})$ is quiescent forgiving for T,

(**C3**) $\forall \hat{P} \in \mathscr{P} : \mathit{scheds}(\hat{P}) \subseteq \mathit{scheds}(T)$, and

(**C4**) $\forall \hat{P} \in \mathscr{P}$: $\forall s \in \mathit{scheds}(T)$:
$s \notin \mathit{scheds}(\hat{P}) \Rightarrow \exists s' \preceq s$:
$(s' \in \mathit{scheds}(\hat{P})) \wedge (s' = s''; a) \wedge (a \in \mathit{input}(T))$
$\wedge\ (\forall t \succeq s' \colon t \in \mathit{scheds}(T) \Rightarrow t \notin \mathit{scheds}(\hat{P}))$.

Proof sketch. ($\Rightarrow$ direction) We assume that we have some arbitrary policy $\mathscr{P}$ and target T, and an input/output-mediating M_1 that specifically precisely enforces $\mathscr{P}$ on T. We need to show that there exists an input-mediating M_2 that specifically precisely enforces $\mathscr{P}$ on T.

We construct M_2 by (a) taking the restriction of M_1 that deals only with inputs from the environment, i.e., we ignore the part of M_1 that receives input actions from T and outputs output actions to environment, and (b) for every input i that belongs to the set of inputs that invalidate extensions we simply remove the corresponding transitions: in other words, for every i that invalidates executions, and for every transition of the form $\langle q, i, q' \rangle$, we remove all transitions of the form $\langle q', a, q'' \rangle$, where a is a local action (we keep input actions because of input-enabledness). We will show that if M_1 specifically precisely enforces $\mathscr{P}$ on T under the above constraints, then M_1 specifically precisely enforces $\mathscr{P}$ on T also.

First we will show that the part of M_1 that receives inputs from T and outputs actions to the environment does not do anything "non-trivial" (under the given constraints), i.e., it either outputs nothing or it simply forwards valid actions that T wants to execute. We do a case analysis on the actions that T might execute and prove that the output-mediating part of M_1 is trivial. First, observe that by (**C3**), M_1 cannot arbitrarily add actions that T might not execute. Second, if T wants to output some action a that obeys the policy, then because of the precise enforcement constraint, M_1 will have to (eventually) output it. Thus in the case of M_1, a is eventually exhibited by itself, whereas in M_2, a will be exhibited directly by T. Finally,if T wants to output some action a that disobeys the policy, then a can either be preceded by some input or not. If it is not preceded by some input, then it must be part of quiescent behavior. But since $\mathscr{P}$ is enforceable, then by Thm. 3 it must be quiescent forgiving, i.e., it must be valid – contradiction. So a must be preceded by some input. But, by (**C2**), there is some input i that precedes a after which all extensions are invalid. Thus, i will be the last action appearing on the schedule. This means that since T can still communicate with M_1, M_1 will suppress all the security relevant behavior following i (i.e., it will trivially output nothing).

Note that the latter is equivalent to not forwarding i to T (or any future inputs) and just continue execution by receiving inputs from the environment.

This is exactly the construction that corresponds to (b). So under the given constraints the two monitors will both precisely enforce $\mathscr{P}$ on T.
($\Leftarrow$ direction)

We assume that we have some arbitrary policy $\mathscr{P}$ and target T, and an input-mediating M_2 that specifically precisely enforces $\mathscr{P}$ on T. We need to construct an input/output-mediating M_1 that specifically precisely enforces $\mathscr{P}$ on T.

We use M_2 to construct M_1. Specifically, we use the same transition relation as M_2 which we extend in a manner similar to the construction of a truncation monitor from a truncation automaton (§3), i.e., we add a special state and a queue that buffers the inputs that the M_1 receives from T. Specifically, once M_1 receives some output $\mathsf{rename}(a)$ from T, it records the state it was before starting to receive inputs. If more inputs follow, it adds them to the queue. Once it has finished forwarding to the environment all the outputs that T wanted to execute (i.e., forward a for each $\mathsf{rename}(a)$ received), it returns to the original state to continue execution (as M_2).

Since $\mathscr{P}$ does not reason about the communication between the monitor and the target, it is easy to see that $(\mathit{scheds}(M_2 \times \mathsf{rename}(T))|\mathit{acts}(\hat{P})) = (\mathit{scheds}(M_1 \times \mathsf{rename}(T))|\mathit{acts}(\hat{P}))$. Every schedule that $M_2 \times \mathsf{rename}(T)$ produces is a schedule of $M_1 \times \mathsf{rename}(T)$, since by construction M_2 does not add any new schedules, and dually, every schedule that $M_1 \times \mathsf{rename}(T)$ produces is a schedule of $M_2 \times \mathsf{rename}(T)$, since by construction M_2 does not remove any schedules.

□