



**HAL**  
open science

# Mechanized Proofs of Adversarial Complexity and Application to Universal Composability

Manuel Barbosa, Gilles Barthe, Benjamin Grégoire, Adrien Koutsos,  
Pierre-Yves Strub

► **To cite this version:**

Manuel Barbosa, Gilles Barthe, Benjamin Grégoire, Adrien Koutsos, Pierre-Yves Strub. Mechanized Proofs of Adversarial Complexity and Application to Universal Composability: Journal pre-print: full version. ACM Transactions on Privacy and Security, 2023, 26 (3), pp.1-34. 10.1145/3589962 . hal-04048217

**HAL Id: hal-04048217**

**<https://inria.hal.science/hal-04048217>**

Submitted on 27 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Mechanized Proofs of Adversarial Complexity and Application to Universal Composability

Journal pre-print: full version\*

MANUEL BARBOSA, University of Porto (FCUP) & INESC TEC, Portugal

GILLES BARTHE, MPI-SP, Germany and IMDEA Software Institute, Spain

BENJAMIN GRÉGOIRE, Inria & Université Côte d'Azur, France

ADRIEN KOUTSOS, Inria, France

PIERRE-YVES STRUB<sup>†</sup>, Meta, France

In this paper we enhance the EasyCrypt proof assistant to reason about computational complexity of adversaries. The key technical tool is a Hoare logic for reasoning about computational complexity (execution time and oracle calls) of adversarial computations. Our Hoare logic is built on top of the module system used by EasyCrypt for modeling adversaries. We prove that our logic is sound w.r.t. the semantics of EasyCrypt programs — we also provide full semantics for the EasyCrypt module system, which was previously lacking.

We showcase (for the first time in EasyCrypt and in other computer-aided cryptographic tools) how our approach can express precise relationships between the probability of adversarial success and their execution time. In particular, we can quantify existentially over adversaries in a complexity class, and express general composition statements in simulation-based frameworks. Moreover, such statements can be composed to derive standard concrete security bounds for cryptographic constructions whose security is proved in a modular way. As a main benefit of our approach, we revisit security proofs of some well-known cryptographic constructions and we present a new formalization of Universal Composability (UC).

CCS Concepts: • **Theory of computation** → **Interactive proof systems**; • **Security and privacy** → **Logic and verification**.

Additional Key Words and Phrases: Verification of Cryptographic Primitives; Formal Methods; Interactive Proof System; Complexity Analysis

## ACM Reference Format:

Manuel Barbosa, Gilles Barthe, Benjamin Grégoire, Adrien Koutsos, and Pierre-Yves Strub. 2023. Mechanized Proofs of Adversarial Complexity and Application to Universal Composability **Journal pre-print: full version**. *ACM Trans. Priv. Sec.* 1, 1, Article 1 (March 2023), 57 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUCTION

Cryptographic designs are typically supported by mathematical proofs of security. Unfortunately, these proofs are error-prone and subtle flaws can go unnoticed for many years, in spite of careful and extensive scrutiny from experts. Therefore, it is desirable that cryptographic proofs are formally verified using computer-aided tools [27]. Over the last decades, many formalisms and tools have been developed for mechanizing cryptographic proofs [5]. In this paper we focus on the EasyCrypt proof assistant [8, 11], which has been used to prove security of a diverse set of cryptographic constructions in the computational model of cryptography [2, 3]. In this setting, cryptographic

\*An earlier version of this work has been presented at ACM CCS'21.

<sup>†</sup>This work was conducted while the author was at École Polytechnique, Institut Polytechnique de Paris, France

---

Authors' addresses: Manuel Barbosa, University of Porto (FCUP) & INESC TEC, Porto, Portugal, [mbb@fc.up.pt](mailto:mbb@fc.up.pt); Gilles Barthe, MPI-SP, Bochum, Germany; IMDEA Software Institute, Madrid, Spain, [gbarthe@mpi-sp.org](mailto:gbarthe@mpi-sp.org); Benjamin Grégoire, Inria & Université Côte d'Azur, Sophia Antipolis, France, [benjamin.gregoire@inria.fr](mailto:benjamin.gregoire@inria.fr); Adrien Koutsos, Inria, Paris, France, [adrien.koutsos@inria.fr](mailto:adrien.koutsos@inria.fr); Pierre-Yves Strub, Meta, Paris, France, [pierre-yves@strub.nu](mailto:pierre-yves@strub.nu).

designs and their corresponding security notions are modeled as probabilistic programs. Moreover, security proofs provide an upper bound on the probability that an adversary breaks a cryptographic design, often assuming that the attacker has limited resources that are insufficient to solve a mathematical problem. While EasyCrypt excels at quantifying the probability of adversarial success, it lacks support for keeping track of the complexity of adversarial computations. This is a limitation that is common to other tools in computer-aided cryptography, and it means that manual inspection is required to check that the formalized claims refer to probabilistic programs that fall in the correct complexity classes. While this may be acceptable for very simple constructions, for more intricate proofs it may be difficult to interpret what a proved claim means in the cryptographic sense; in particular, existing computer-aided tools cannot fully express the subtleties that arise in compositional approaches such as Universal Composability [18]. This is an important limitation, as compositional approaches are ideally suited for proving security of complex cryptographic designs involving many layers of simpler building blocks. This work overcomes this limitation and showcases the benefits of reasoning about computational complexity in EasyCrypt, through three broad contributions.

*Formal verification of complexity statements.* We define a formal system for specifying and proving complexity claims. Our formal system is based on an expressive module system, which enriches the existing EasyCrypt module system with declarations of memory footprints (specifying what is read and written) and cost (specifying which oracles can be called and how often). This richer module system is the basis for modeling the cost of a program as a tuple. The first component of the tuple represents the intrinsic cost of the program, i.e. its cost in a model where oracle and adversary calls are free. The remaining components of the tuple represent the number of calls to oracles and adversaries. This style of modeling is compatible with cryptographic practice and supports reasoning compositionally about (open) programs.

Our formal system is built on top of the module system and takes the form of a Hoare logic for proving complexity claims that upper bound the cost of expressions and commands. Furthermore, an embedding of the formal system into a higher-order logic provides support for reductionist statements relating adversarial advantage and execution cost, for instance:

$$\forall \mathcal{A}. \exists \mathcal{B}. \text{adv}_{\mathcal{S}}(\mathcal{A}) \leq \text{adv}_{\mathcal{H}}(\mathcal{B}) + \epsilon \wedge \text{cost}(\mathcal{B}) \leq \text{cost}(\mathcal{A}) + \delta$$

where typically  $\epsilon$  and  $\delta$  are polynomial expressions in the number of oracle calls. The above statement says that every adversary  $\mathcal{A}$  can be turned into an adversary  $\mathcal{B}$ , with sensibly equivalent resources, such that the advantage of  $\mathcal{A}$  against a cryptographic scheme  $\mathcal{S}$  is upper bounded by the advantage of  $\mathcal{B}$  against a hardness assumption  $\mathcal{H}$ . Note that the statement is only meaningful because the cost of  $\mathcal{B}$  is conditioned on the cost of  $\mathcal{A}$ , as the advantage of an unbounded adversary is typically large (e.g., it succeeds with probability 1). The ability to prove and instantiate such  $\forall \exists$ -statements is essential for capturing compositional reasoning principles.

We show correctness of our formal system w.r.t. an interpretation of programs. Our interpretation provides the first complete semantics for the EasyCrypt module system, which was previously lacking. This semantics is of independent interest and could be used to prove soundness of the two program logics supported by EasyCrypt: a Relational Hoare Logic [10] and a Union Bound logic [9].

*Implementation in the EasyCrypt proof assistant.* We have implemented our formal system as an extension to the EasyCrypt proof assistant, which provides mechanisms for declaring the cost of operators and for helping users derive the cost of expressions and programs. Our implementation brings several contributions of independent interest, including an improvement of the memory restriction system of EasyCrypt, and a library and automation support to reason about extended

integers that are used for reasoning about cost. For the latter we follow [40] and reduce formulae about extended integers to integer formulae that can be sent to SMT solvers. Another key step is to embed our Hoare logic for cost into the ambient higher-order logic—similar to what is done for the other program logics of EasyCrypt. This allows us to combine judgments from different program logics, and it enhances the expressiveness of the approach. Implementation-wise, this extension required to add or rewrite around 8 kLoC of EasyCrypt. The implementation and examples (including those of the paper as well as classic examples from the EasyCrypt distribution, including Bellare and Rogaway BR93 Encryption, Hashed ElGamal encryption, Cramer-Shoup encryption, and hybrid arguments) are open source [23].

*Case study: Universal Composability.* Universal Composability [17, 19] (UC) is a popular framework for reasoning about cryptographic systems. Its central notion, called UC-emulation, formalizes when a protocol  $\pi_1$  can safely replace a protocol  $\pi_2$ . Informally, UC-emulation imposes that there exists a simulator  $\mathcal{S}$  capable of *fooling* any environment  $\mathcal{Z}$  by presenting to it a view that is fully consistent with an interaction with  $\pi_1$ , while it is in fact interacting with  $\mathcal{S}(\pi_2)$ . This intuition, however, must be formalized with tight control over the capabilities of the environment and the simulator. If this were not the case, the definition would make no sense: existential quantification over unrestricted simulators is too weak (it is crucial for the compositional security semantics that simulators use comparable resources to real-world attackers), whereas universal quantification over unrestricted environments results in a definition that is too strong to be satisfied [17, 18]. Moreover, when writing proofs in the UC setting, it is often necessary to consider the joint resources of a sub-part of a complex system that involves a mixture of concrete probabilistic algorithms and abstract adversarial entities, when they are grouped together to build an attacker for a reductionistic proof. In these cases, it is very hard to determine by inspection whether the constructed adversaries are within the complexity classes for which the underlying computational assumptions are assumed to hold. Therefore, tool support for complexity claims is of particular importance with UC – conversely, UC is a particularly challenging example for complexity claims.

Using our enriched implementation of EasyCrypt, we develop a new *fully mechanized* formalization of UC. In contrast to [21], which chooses to follow closely the classic execution model for UC, our mechanization adopts a more EasyCrypt-friendly approach that is closer to the simplified version of UC proposed by Canetti, Cohen and Lindell in [20]; this is further discussed in Section 6. Our mechanization covers the core notions of UC, the classic composition lemmas, transitivity and composability, which respectively state that UC-emulation (as a binary relation between cryptographic systems) is closed under transitivity and arbitrary adversarial contexts. More importantly, our development captures for the first time the complexity aspects of these general results. As an illustrative application of our approach we revisit the example used in [21], where modular proofs for Diffie-Hellman key exchange and encryption over ideal authenticated channels are composed to construct a UC secure channel.

*Discussion.* The possibility to quantify over adversary using complexity claims introduces conceptual simplifications in layered proofs by i. supporting compositional reasoning and ii. avoiding the use of explicit cost accounting modeling. The downside is that it also introduces some additional burden on users, who now must prove complexity claims. However, we note that our extension does not invalidate existing EasyCrypt developments: complexity claims are optional, existing proofs have been left unchanged, and their type-checking remains as fast as before. Furthermore, it is possible to layer the complexity claims on top of standard EasyCrypt proofs that do not capture the complexity aspects – in effect, this is what we did in our example. We have also provided rudimentary support to automate proofs of complexity claims, and could enhance this support

even further by adopting ideas from cost analysis. We think that the current tool is significantly more usable and scalable than prior versions without support for complexity reasoning.

To make this claim more concrete, let us consider the implications of refactoring an existing EasyCrypt development and extend it to take advantage of cost analysis for both dealing with query counts and to include complexity claims. Removing the layer of modular wrapping that explicitly keeps track of query counts leads to more readable code, and has essentially no impact on the proofs. However, when it comes to complexity claims, new specifications and proof scripts must be added to the development. The new specifications consist of the description of the cost model and the declarations of the types of the various algorithms, which include explicit cost expressions. The additional proof effort consists of applying our logic to prove complexity claims and discharging the relevant side-conditions. As a coarse metric on the additional proof and specification efforts required, we consider the ratio of the number of lines of codes related to the cost analysis over the total number of lines. For the example presented in the next section, this ratio is 117/495. For the Universal Composability example, the ratio is 270/2300 for the concrete protocol and 791/1775 for the general composition theorems. We also note that there is a large body of work on automated complexity analysis, as mentioned in the related work section, which might reduce this overhead.

This paper is an extended version of [6], which was presented at the ACM CCS'21 conference.

## ACKNOWLEDGMENTS

The research leading to these results has received funding the French National Research Agency (ANR) under the project TECAP ANR-17-CE39-0004-01.

This work is financed by National Funds through the FCT - Fundação para a Ciência e a Tecnologia within project PTDC/CCI-INF/31698/2017.

Work by Gilles Barthe was supported by the Office of Naval Research (ONR) under project N00014-15-1-2750.

This work received funding from the France 2030 program managed by the French National Research Agency under grant agreement No. ANR-22-PECY-0006.

## 2 WARM UP EXAMPLE: PKE FROM A ONE-WAY TRAPDOOR PERMUTATION

To illustrate our approach we use a public-key encryption (PKE) scheme proposed by [13] (BR93) that uses a one-way trapdoor permutation  $f$  (used with public key  $pk$ ) and a cryptographic hash function  $H.o$  modeled as a random oracle (RO). The encryption of a message  $m$  by the BR93 scheme is  $(f\ pk\ r\ ||\ (H.o(r) \oplus m))$  where  $||$  is bit-string concatenation and  $\oplus$  is bit-wise xor. We start by a quick presentation of EasyCrypt modules, before delving into the proof of security of this scheme.

Modules are a key ingredient of an EasyCrypt formalization. Roughly, a module is a structure packaging together variable and procedures declarations. Modules can be parameterized by one or more modules — such parameterized modules are called functors. Fig. 1 gives a small example of a functor named BR93 implementing the BR93 encryption function. This module is parameterized by a module  $H$  modeling the random oracle used by the encryption scheme, which it uses in its implementation of the `enc` procedure. Modules types are used to structure module declarations, by declaring the signature of the procedures that must be included in a module. Coming back to our example, RO is a module type stating that  $H$  must declare a procedure named `o` from values of type `rand` to values of type

---

```

module BR93 (H : RO) = {
  proc enc(pk, m) = {
    var r : rand, h : plaintext;

    r  $\stackrel{\$}{\leftarrow}$  drand;
    h  $\leftarrow$  H.o(r);
    return (f pk r || h  $\oplus$  m);
  }
}

```

---

Fig. 1. Example of an EasyCrypt module.

---

```

module type RO = {
  proc o (r:rand) : plaintext compl[intr :  $t_o$ ]
}.

module type Scheme (H: RO) = {
  proc kg() : pkey * skey
  proc enc(pk:pkey, m:plaintext) : ciphertext
  proc dec(sk:skey, c:ciphertext) : plaintext option
}.

module type Adv (H: RO) = {
  proc choose(p:pkey) :
    (plaintext * plaintext) compl[intr :  $t_c$ , H.o :  $k_c$ ]
  proc guess(c:ciphertext) :
    bool compl[intr :  $t_g$ , H.o :  $k_g$ ]
}.
    
```

---

```

module (Inv : INV) (H : RO) (A:Adv) = {
  var qs : rand list
  module QH = {
    proc o(x:rand) = {
      qs ← x::qs;
      r ← H.o(x);
      return r; }
  }
  proc invert(pk:pkey,y:rand) : rand = {
    qs ← [];
    (m0,m1) ← A(QH).choose(pk);
    h ←  $\overset{\$}{\leftarrow}$  dplaintext;
    b ← A(QH).guess(y || h);
    while (qs ≠ []) {
      r ← head qs;
      if (f pk r = y) return r;
      qs ← tail qs; }
  }
}.
    
```

---

Fig. 2. Inverter for trapdoor permutation in EasyCrypt.

plaintext (c.f. Fig. 2) – the type of values outputted by the encryption. Finally, a module can contains sub-modules, allowing for a rich hierarchical presentation.

Intuitively, the RO is used to convert the message into a random input for the trapdoor permutation so as to allow a reduction to the one-wayness property. This proof strategy is used in BR93 and many other schemes, including OAEP [13]. Fig. 2 shows the code of an inverter for the trapdoor permutation that is constructed from an attacker against the encryption scheme.<sup>1</sup> This inverter simulates the single random oracle used by the encryption scheme for the attacker and recovers the pre-image to  $y$  with essentially the same probability as the attacker breaks the encryption scheme.

We first define module types for random oracles RO, schemes Scheme, and adversaries Adv. The module type for random oracles declares a single procedure  $o$  with cost  $\leq t_o$ . The module type for schemes declares three procedures for key generation, encryption, and decryption, and is parameterized by a random oracle H. No cost declaration is necessary. The module type for (chosen-plaintext) adversaries declares two procedures: choose for choosing two plaintexts  $m_0$  and  $m_1$ , and guess for guessing the (uniformly sampled) bit  $b$  given an encryption of  $m_b$ . The cost of these procedures is a pair: the second component is an upper bound on the number of times it can call the random oracle, and the first is an upper bound on its intrinsic cost, i.e. its cost assuming that oracle calls (modeled as functor parameters) have a cost of 0. This style of modeling is routinely used in cryptography and is better suited to reason about open code. This cost model is also more fine-grained than counting the total cost of the procedure including the cost of the oracles, as we have a guarantee on the number of time oracles are called.

Next, we define the inverter  $Inv$  for the one-way trapdoor permutation. It runs the adversary A, keeping track of all the calls that A makes to H in a list  $qs$  (using the sub-module QH), and then searches in the list  $qs$  for a pre-image of  $y$  under  $f$  pk. Search is done through a while loop, which we write in a slightly beautified syntax. This inverter can be used to state the following

<sup>1</sup>We use the following notation:  $\overset{\$}{\leftarrow}$  denotes a random sampling;  $[]$  is the empty list;  $a :: l$  appends a to the list l.

reductionist security theorem relating the advantage and execution cost of an adversary against chosen-plaintext security of the PKE with the advantage of the inverter against one-wayness.

**THEOREM 2.1 (SECURITY OF BR93).** *Let  $t_f$  represent the cost of applying the one-way function  $f$  and  $t_o$  denote the cost of  $H.o$ , i.e. the implementation of a query to a lazily sampled random oracle. Fix the type for adversaries  $\tau_{\mathcal{A}}$  such that:*

$$\text{cost } \mathcal{A}.choose \leq \text{compl}[\text{intr} : t_c, H.o : k_c] \quad \text{and} \quad \text{cost } \mathcal{A}.guess \leq \text{compl}[\text{intr} : t_g, H.o : k_g]$$

and fix  $\tau_I$  such that:

$$\text{cost } I.invert \leq \text{compl}[\text{intr} : (5 + t_f) \cdot (k_c + k_g) + 4 + t_o \cdot (k_c + k_g) + t_c + t_g].$$

Then,  $\forall \mathcal{A} \in \tau_{\mathcal{A}}, \exists I \in \tau_I, \text{adv}_{\text{IND-CPA}}^{\text{BR93}}(\mathcal{A}) \leq \text{adv}_{\text{OW}}^f(I)$ .

Here, IND-CPA refers to the standard notion of ciphertext indistinguishability under chosen-plaintext attacks for PKE, where the adversary is given the public key and asked to guess which of two messages of its choice has been encrypted in a challenge ciphertext; OW refers to the standard one-wayness definition for trapdoor permutations, where the attacker is given the public parameters and the image of a random pre-image, which it must invert. In the former, advantage is the absolute bias of the adversary's boolean output w.r.t. 1/2; in the latter, advantage is the probability of successful inversion.

We prove the statement by providing  $\text{Inv}(\mathcal{A})$  as a witness for the existential quantification, which creates two sub-goals. The first sub-goal establishes the advantage bound, which we prove using relational Hoare logic. The second sub-goal establishes that our inverter satisfies the required cost restrictions, and is proved using our Hoare logic for complexity. We declare the type of  $\text{Inv}$  as:

$$\begin{aligned} \text{cost } \text{Inv}.invert &\leq \text{compl}[\text{intr} : (5 + t_f) \cdot (k_c + k_g) + 4, \\ &H.o = k_c + k_g, \mathcal{A}.choose = 1, \mathcal{A}.guess = 1] \end{aligned}$$

and so we first must establish that  $\text{Inv}$  belongs to this functor type. It is easy to show that  $\text{A.choose}$  and  $\text{A.guess}$  are called exactly once, and that  $\text{H.o}$  is called at most  $k_c + k_g$  times. So we turn to the intrinsic complexity of  $\text{Inv}$ . The key step for this proof is to show that the loop does at most  $k_c + k_g$  iterations. We use the length of  $qs$  as a variant: the length of the list is initially 0, and incremented by 1 by each call to the random oracle, therefore its length at the start of the loop is at most  $k_c + k_g$ . Moreover, the length decreases by 1 at each iteration, so we are done. The remaining reasoning is standard,<sup>2</sup> using the cost of each operator—fixed by choice in this particular example to 1, except for the operator  $f$ . Our modeling of cost enforces useful invariants that simplify reasoning. For instance, proving upper bounds on the execution cost of  $\text{Inv}$  requires proving an upper bound on the number of iterations of the loop, and therefore on the length of  $qs$  upon entering the loop. We derive the complexity statement in the theorem, which shows only the intrinsic cost of  $\text{Inv}$ , by instantiating the complexity type of  $\text{Inv}$  with the cost of its module parameter  $\mathcal{A}$ . This illustrates how our finer-grained notion of cost is useful for compositional reasoning.

*Comparison with EasyCrypt.* Our formalization follows the same pattern as the BR93 formalization from the EasyCrypt library. However, the classic module system of EasyCrypt only tracks read-and-write effects and lacks first-class support for bounding the number of oracle calls and for reasoning about the complexity of programs. To compensate for this first point, classic EasyCrypt proofs use wrappers to explicitly count the number of calls and to return dummy answers when the number of adversarial calls to an oracle exceeds a threshold. The use of wrappers suffices for reasoning

<sup>2</sup>Notice that the condition of the loop is executed at most  $k_c + k_g$  time.

<p>Expressions (<i>distribution expressions are similar</i>):</p> $e ::= v \in \mathcal{V} \quad (\text{variable})$ $  f(e_1, \dots, e_n) \quad (\text{if } f \in \mathcal{F}_E)$ <p>Statements:</p> $s ::= \mathbf{abort} \quad (\text{abort})$ $  \mathbf{skip} \quad (\text{skip})$ $  s_1; s_2 \quad (\text{sequence})$ $  x \leftarrow e \quad (\text{assignment})$ $  x \overset{\$}{\leftarrow} d \quad (\text{sampling})$ $  x \leftarrow \mathbf{call} F(\vec{e}) \quad (\text{proc. call})$ $  \mathbf{if } e \mathbf{ then } s_1 \mathbf{ else } s_2 \quad (\text{cond.})$ $  \mathbf{while } e \mathbf{ do } s \quad (\text{loop})$ <p>Procedure body:</p> $\text{body} ::= \{ \text{var } (\vec{v} : \vec{\tau}); s; \text{return } e \}$	<p>Function paths:</p> $F ::= p.f \quad (\text{proc. lookup})$ <p>Module paths:</p> $p ::= x \quad (\text{mod. ident.})$ $  p.x \quad (\text{mod. comp.})$ $  p(p) \quad (\text{func. app.})$ <p>Module expressions:</p> $m ::= p \quad (\text{mod. path})$ $  \mathbf{struct } st \mathbf{ end} \quad (\text{structure})$ $  \mathbf{func}(x : M) m \quad (\text{functor})$ <p>Module structures:</p> $st ::= d_1; \dots; d_n \quad (n \in \mathbb{N})$ <p>Module declarations:</p> $d ::= \mathbf{module } x = m$ $  \mathbf{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}$
--	---

Fig. 3. Program and module syntax

about adversarial advantage. However, no similar solution can be used for reasoning about the computational cost of adversaries.

Therefore, the BR93 formalization from the EasyCrypt library makes use of the explicit definition of  $\mathcal{I}$ , and users must analyze the complexity of  $\mathcal{I}$  outside the tool. As a result, machine-checked security statements are partial (complexity analysis is missing), cluttered (existential quantification is replaced by explicit witnesses), and compositional reasoning is hard (existential quantification over module types cannot be used meaningfully).

### 3 ENRICHED EASYCRYPT MODULE SYSTEM

We present a formalization of our extended module system for EasyCrypt. It is based on EasyCrypt current imperative probabilistic programming language and module system, which we enrich to track the read-and-write memory footprint and complexity cost of module components through *module restrictions*. These module restrictions are checked through a type system: memory footprint type-checking is fully automatic, while type-checking a complexity restriction generates a proof obligation that is discharged to the user — using the cost Hoare logic we present later, in Section 4.

#### 3.1 Syntax of Programs and Modules

The syntax of our language and module system is (quite) standard and summarized in Fig. 3. We describe it in more detail below. We assume given a set of operators  $\mathcal{F}_E$  and a set of distribution operators  $\mathcal{F}_D$ . For any  $g \in \mathcal{F}_E \cup \mathcal{F}_D$ , we assume given its type:  $\text{type}(g) = \tau_1 \times \dots \times \tau_n \rightarrow \tau$  where  $\tau_1, \dots, \tau_n, \tau \in \mathbb{B}$  with  $\mathbb{B}$  the set of base types. We require that `bool` is a base type, and otherwise leave  $\mathbb{B}$  unspecified.

We consider well-typed arity-respecting expressions built from  $\mathcal{F}_E$  and variables in  $\mathcal{V}$ . Similarly, distribution expressions  $d$  are built upon  $\mathcal{F}_D$  and  $\mathcal{V}$ . For any expression  $e$ , we let  $\text{vars}(e)$  be the set of variables appearing in  $e$  (idem for distribution expression).



Signature structures (for any  $n \in \mathbb{N}$ ):

$$S ::= D_1; \dots; D_n$$

Module signature declarations:

$$D ::= \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r \mid \text{module } x : M$$

Module signatures:

$$M ::= \text{sig } S \text{ restr } \theta \text{ end} \mid \text{func}(x : M) M'$$

**Module restrictions:**

$$\theta ::= \epsilon \mid \theta, (f : \lambda) \qquad \lambda ::= \top \mid \lambda_m \wedge \lambda_c$$

**Memory restrictions** (for any  $l \in \mathbb{N}$ ):

$$\lambda_m ::= +\text{all mem} \setminus \{v_1, \dots, v_l\} \mid \{v_1, \dots, v_l\}$$

**Complexity restrictions** (for any  $l, k, k_1, \dots, k_l \in \mathbb{N}$ ):

$$\lambda_c ::= \top \mid \text{compl}[\text{intr} : k, x_1.f_1 : k_1, \dots, x_l.f_l : k_l]$$

Fig. 4. Module signatures and restrictions

We assume a simple language for program statements. A statement  $s$  can be an abort, a skip, a statement sequence  $s_1; s_2$ , an assignment  $x \leftarrow e$  of an expression to a variable, a random sampling  $x \xleftarrow{\$} d$  from a distribution expression, a conditional, a while loop, or a procedure call  $x \leftarrow \text{call } F(\vec{e})$ .

*The module system.* In a procedure call,  $F$  is a function path of the form  $p.f$  where  $f$  is the procedure name and  $p$  is a module path. Basically, when calling  $p.f$ , the module system will resolve  $p$  to a module structure, which must declare the procedure  $f$  (this will be guaranteed by our type system). Formally, a module structure  $st$  is a list of module declarations, and a module declaration  $d$  is either a procedure (with typed arguments, and a body which comprises a list of local variables with their types  $\vec{v} : \vec{\tau}$ , a statement  $s$  and a return expression  $e$ ) or a sub-module declaration.

The component  $c$  of a module  $x$  can be accessed through the module path expression  $x.c$ . Since a module can contain sub-modules, we can have nested accesses, as in  $x. \dots .z.c$ . Hence, a module path is either a module identifier, a component access of another module path  $p$ , or a functor application. Finally, a module expression  $m$  is either a module path, a module structure or a functor.

### 3.2 Module Signatures and Restrictions

The novel part of our system is the use of module restrictions in module signatures. Objects related to module restriction are **highlighted in red** throughout this paper (this is only here to improve readability, not to convey additional information). The syntax of module signatures and restrictions is given in Fig. 4. A module structure signature  $S$  is a list of module signature declarations, which are procedure signatures or sub-module signatures. Then, a module signature  $M$  is either a functor signature, or a structure signature with a module restriction  $\theta$  attached.

*Module restrictions.* A module restriction restricts the effects of a module's procedures. We are interested in two types of effects. First, we characterize the memory footprint (i.e. global variables which are read or written to) of a module's procedures through *memory restrictions*. Second, we upper bound the execution cost of a procedure, and the number of calls a functor's procedure can make to the functor's parameters, through *complexity restrictions*.

Restrictions are useful for compositional reasoning, as they allow stating and verifying properties of a module's procedures at declaration time. In the case of an abstract module (i.e. a module whose code is unknown), restrictions allow to constrain, through the type system, its possible instantiations. This is a key idea of our approach, which we exploit to prove complexity properties of cryptographic reductions.

For example, we give in Fig. 5 EasyCrypt code corresponding to an adversary against a hardware security module. In this scenario the goal of the adversary is to recover the secret key stored in the module `Hsm`. The example uses two types of restrictions. The module-level restriction `{+all mem, -Hsm}` states that such an adversary can access all the memory, except for the memory used by the module `Hsm`. The procedure-level restriction `[intr : k0, H.enc : k]` attached to `guess`, states that `guess` execution time is at most `k0` (excluding calls to `H.enc`), and that `guess` can make at most `k` queries to the procedure `H.enc`.

Formally, a module restriction is a list of pairs comprising a procedure identifier  $f$  and a procedure restriction  $\lambda$ , and a procedure restriction  $\lambda$  is either  $\top$  (no restriction), or the conjunction of a memory restriction  $\lambda_m$  and a complexity restriction  $\lambda_c$ :

*Memory.* A memory restriction  $\lambda_m$ , attached to a procedure  $f$ , restricts the variables that  $f$  can access *directly*. We allow for positive memory restrictions  $\{v_1, \dots, v_l\}$ , which states that  $f$  can only access the variables  $v_1, \dots, v_l$ ; and negative memory restrictions `+all mem \ {v1, ..., vl}`, which states that  $f$  can access any global variables except the variables  $v_1, \dots, v_l$ .

Note that  $\lambda_m$  only restricts  $f$ 's *direct* memory accesses: this excludes the memory accessed by the procedure oracles (which are modeled as functor's parameters). This is crucial, as otherwise, an adversary that is not allowed to access some oracle's memory (a standard assumption in security proofs) would not be allowed to call this oracle. E.g., the adversary of Fig. 5 can call the oracle `H.enc` (which can be instantiated by `Hsm`), even though it cannot access directly `Hsm`'s memory.

*Complexity.* A complexity restriction  $\lambda_c$  attached to a procedure  $f$  restricts its execution time and the number of calls that  $f$  can make to its parameters: it is either  $\top$ , i.e. no restriction; or the restriction `compl[intr : k, x1.f1 : k1, ..., xl.fl : kl]`, which states that: i) its execution time (excluding calls to the parameters) must be at most  $k$ ; ii)  $f$  can call, for every  $i$ , the parameter's procedure  $x_i.f_i$  at most  $k_i$  times. We require that all parameters' procedures appear in the restriction. This can be done w.l.o.g. by assuming that any missing entry is zero (which is exactly what is done in our EasyCrypt implementation).

### 3.3 Typing Enriched Module Restrictions

We now present the core rules of our module type system, which are summarized in Fig. 6 and Fig. 8. The rest of the rules are postponed to Appendix A. For clarity of presentation, our module type system requires module paths to always be long modules paths, from the root of the program to the sub-module called (we give a simple example in Fig. 7). This allows to have a simpler module resolution mechanism, by removing any scoping issues. This is done without loss of

---

```

module type HSM = {
  proc enc (x:msg) : cipher.
module Hsm : HSM = {
  proc enc (x:msg) : cipher = { . . . }.
module type Adv (H : HSM) {+all mem, -Hsm} = {
  proc guess () : skey compl[intr : k0, H.enc : k].
    
```

---

Fig. 5. Example of adversary with restrictions.

---

```

module A = {
module B = { . . . }
module C = {
  module E = A.B    (* Valid full path *)
  module F = B      (* Invalid path *)
}
    
```

---

Fig. 7. Example: valid and invalid paths.

<b>Module path typing</b> $\Gamma \vdash p : M$ .				
<b>NAME</b> $\frac{\Gamma(p) = \_ : M}{\Gamma \vdash p : M}$	<b>COMPNT</b> $\frac{\Gamma \vdash p : \text{sig } S_1; \text{ module } x : M; S_2 \text{ restr } \theta \text{ end}}{\Gamma \vdash p.x : M}$	<b>FUNCAPP</b> $\frac{\Gamma \vdash p : \text{func}(x : M') M \quad \Gamma \vdash p' : M'}{\Gamma \vdash p(p') : M[x \mapsto \text{mem}_\Gamma(p')]}$		
<b>Module expression typing</b> $\Gamma \vdash_p m : M$ .				
<i>We omit the rules <math>\Gamma \vdash M</math> to check that a module signature <math>M</math> is well-formed.</i>				
<b>ALIAS</b> $\frac{\Gamma \vdash p_a : M}{\Gamma \vdash_p p_a : M}$	<b>STRUCT</b> $\frac{\Gamma \vdash_{p,\theta} \text{st} : S}{\Gamma \vdash_p \text{struct st end} : \text{sig } S \text{ restr } \theta \text{ end}}$	<b>FUNC</b> $\frac{\Gamma \vdash M_0 \quad \Gamma(x) \not\leq_{\text{undef}}}{\Gamma, \text{ module } x = \text{abs}_{\text{param}} : M_0 \vdash_{p(x)} m : M} \quad \Gamma \vdash_p \text{func}(x : M_0) m : \text{func}(x : M_0) M$	<b>SUB</b> $\frac{\Gamma \vdash_p m : M_0 \quad \Gamma \vdash M_0 < : M}{\Gamma \vdash_p m : M}$	
<b>Module structure typing</b> $\Gamma \vdash_{p,\theta} \text{st} : S$ .				
<b>PROCDECL</b> $\frac{\Gamma_f \vdash s \quad \text{body} = \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s; \text{return } r \} \quad \vec{v}, \vec{v}_1 \text{ fresh in } \Gamma \quad \Gamma_f = \Gamma, \text{ var } \vec{v} : \vec{\tau}, \text{ var } \vec{v}_1 : \vec{\tau}_1 \quad \Gamma(p.f) \not\leq_{\text{undef}} \quad \Gamma, \text{ proc } p.f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body} \vdash_{p,\theta} \text{st} : S}{\Gamma \vdash_{p,\theta} (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}; \text{st}) : (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r; S)}$				
<b>MODDECL</b> $\frac{\Gamma \vdash_{p,x} m : M \quad \Gamma(p.x) \not\leq_{\text{undef}} \quad \Gamma, \text{ module } p.x = m : M \vdash_{p,\theta} \text{st} : S}{\Gamma \vdash_{p,\theta} (\text{module } x = m; \text{st}) : (\text{module } x : M; S)}$			<b>STRUCTEMP</b> $\frac{}{\Gamma \vdash_{p,\theta} \epsilon : \epsilon}$	
<b>Environments typing</b> $\vdash \mathcal{E}$				
<b>ENVEMP</b> $\frac{}{\vdash \epsilon}$	<b>ENVSEQ</b> $\frac{\vdash \mathcal{E} \quad \mathcal{E} \vdash \delta}{\vdash \mathcal{E}, \delta}$	<b>ENVVAR</b> $\frac{\mathcal{E}(v) \not\leq_{\text{undef}}}{\mathcal{E} \vdash \text{var } v : \tau}$	<b>ENVMOD</b> $\frac{\mathcal{E} \vdash_x m : M \quad \mathcal{E}(x) \not\leq_{\text{undef}}}{\mathcal{E} \vdash (\text{module } x = m : M)}$	<b>ENVABS</b> $\frac{\mathcal{E} \vdash M_l \quad \mathcal{E}(x) \not\leq_{\text{undef}}}{\mathcal{E} \vdash (\text{module } x = \text{abs}_K : M)}$

Fig. 6. Core typing rules.

generality: in practice, one can always replace short module paths with long module paths when parsing a program.

A typing environment  $\Gamma$  is a list of typing declarations. A typing declaration, denoted  $\delta$ , is either a variable, module, abstract module or procedure declaration, with a type.

$$\delta ::= \text{var } v : \tau \mid \text{module } p = m : M \mid \text{module } x = \text{abs}_K : M \mid \text{proc } p.f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}$$

$$K ::= \text{open} \mid \text{param} \quad \Gamma ::= \epsilon \mid \Gamma, \delta$$

Note that module and procedure declarations can be rooted at an arbitrary path  $p$ .

An abstract module declaration  $\text{module } x = \text{abs}_K : M$  states that  $x$  is a module with signature  $M$  whose code is unknown. This is used either for open code, or to represent a functor parameter at typing time. Open modules and parameters are treated differently by the type system: a memory restriction ignores the memory footprint of a functor parameter; and a complexity restriction restricts the number of calls that can be made by parameters' procedures. Therefore, we annotate an abstract module with its kind, which can be **open** or **param**. Finally, module and procedure declarations come with the absolute path from the root of the program to the parent module where the declaration is made (variable and abstract modules are always declared at top-level).

For example, the entry  $(\text{module } p.x = m : M)$  means that there is a sub-module  $m$  named  $x$  and with type  $M$  declared at path  $p$ . As usual we require that typing environments do not contain two declarations with the same path. This allows to see a typing environment  $\Gamma$  as a partial function from variable names  $v$ , module paths  $p$  or procedure paths  $p.f$  to (base, module, abstract modules or procedure) values and their types, defined as follows:

$$\Gamma(v) = \tau \quad (\text{if } \Gamma = (\Gamma_1; \text{var } v : \tau; \Gamma_2))$$

$$\begin{aligned}
 \Gamma(p) &= m : M && (\text{if } \Gamma = (\Gamma_1; \text{ module } p = m : M; \Gamma_2)) \\
 \Gamma(x) &= \text{abs}_K x : M && (\text{if } \Gamma = (\Gamma_1; \text{ module } x = \text{abs}_K : M; \Gamma_2)) \\
 \Gamma(p.f) &= \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body} && (\text{if } \Gamma = (\Gamma_1; \text{ proc } p.f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}; \Gamma_2))
 \end{aligned}$$

and  $\Gamma(z) = \text{undef}$  otherwise. Also, we write  $\Gamma(z) \not\leq_{\text{undef}}$  when  $\Gamma(z') = \text{undef}$ . for any prefix  $z'$  of  $z$ .<sup>3</sup>

*Abstract modules.* Abstract modules representing open code (i.e. with kind **open**) are restricted to low-order signatures:

$$\begin{aligned}
 M_l &::= \text{sig } S_l \text{ restr } \theta \text{ end} \mid \text{func}(x : \text{sig } S_l \text{ restr } \theta \text{ end}) M_l \\
 S_l &::= D_{l1}; \dots; D_{ln} && D_l ::= \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r
 \end{aligned}$$

Basically, we only allow module structures, or functors whose parameters are module structures. This restriction is motivated by the fact that further generality is not necessary for cryptographic proofs (adversaries and simulations usually return base values, not procedures); and, more importantly, this restriction allows the abstract call rule of our instrumented Hoare logic **ABS** presented in Fig. 12 to remain tractable.

For any  $M_l$ , we let  $\text{procs}(M_l) = \{f_1, \dots, f_n\}$  be the set of procedure names declared in  $M_l$ .

*Environments.* The semantics of programs, presented later in Section 5, is parametrized by an environment  $\mathcal{E}$ . Essentially, a environment is a typing environment that do not contain abstract module declarations of kind, and which contains only top-level module declaration (i.e. with a module path of the form  $x$ ).

$$\mathcal{E} ::= \epsilon \mid \mathcal{E}, \text{ var } v : \tau \mid \mathcal{E}, \text{ module } x = m : M \mid \mathcal{E}, \text{ module } x = \text{abs}_{\text{open}} : M_l$$

For any  $\mathcal{E}$ , we let  $\text{abs}(\mathcal{E}) = \{x_1, \dots, x_n\}$  be the set of abstract module names declared in  $\mathcal{E}$ .

*Typing module paths.* The typing judgment  $\Gamma \vdash p : M$  states that the module path  $p$  refers to a module with type  $M$ . Its typing rules, which are given in Fig. 6, are standard [32], except for the functor application typing rule **FUNCAPP**:

$$\text{FUNCAPP} \quad \frac{\Gamma \vdash p : \text{func}(x : M') M \quad \Gamma \vdash p' : M'}{\Gamma \vdash p(p') : M[x \mapsto \text{mem}_\Gamma(p')]}$$

A key point here is that we need to substitute  $x$  in the module signature. The substitution function is standard (see Fig. 24), except for module restrictions, which are modified as follows:

- a memory restriction restricts the variables that a procedure can access *directly* – however, memory accesses done through functor parameters are purposely not restricted. Hence, when we instantiate a functor parameter  $x$  by a module path  $p'$ , we must add its memory footprint, which is  $\text{mem}_\Gamma(p')$ . This is handled when substituting  $x$  in a memory restriction:

$$\lambda_m[x \mapsto \text{mem}_\Gamma(p')] = \lambda_m \sqcup \text{mem}_\Gamma(p')$$

- a complexity restriction gives upper bounds on a procedure execution time, and on the number of calls it can make to its functors' parameters. When we instantiate a functor, we remove a functor parameter, and therefore remove the corresponding entries in the

<sup>3</sup>Meaning that the (variable, module or procedure) path  $z$  is not declared by  $\Gamma$ , even through a sub-module or functor application.

**Restriction checking**  $\Gamma \vdash \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s; \text{return } e \} \triangleright \theta$ .

$$\begin{array}{c}
 \text{RESTRCHECK} \\
 \frac{\Gamma \vdash \text{body} \triangleright \lambda_m}{\Gamma \vdash \text{body} \triangleright \lambda_c} \\
 \Gamma \vdash \text{body} \triangleright \lambda_m \wedge \lambda_c
 \end{array}
 \qquad
 \begin{array}{c}
 \text{RESTRMEM} \\
 \frac{\Gamma \vdash s \triangleright \lambda_m \quad \Gamma \vdash e \triangleright \lambda_m}{\Gamma \vdash \{ \_ ; s; \text{return } e \} \triangleright \lambda_m}
 \end{array}
 \qquad
 \begin{array}{c}
 \text{RESTRMEMS} \\
 \frac{\text{mem}_\Gamma(s) \sqsubseteq \lambda_m}{\Gamma \vdash s \triangleright \lambda_m}
 \end{array}
 \qquad
 \begin{array}{c}
 \text{RESTRMEME} \\
 \frac{\text{vars}(e) \sqsubseteq \lambda_m}{\Gamma \vdash e \triangleright \lambda_m}
 \end{array}$$
  

$$\begin{array}{c}
 \text{RESTRCOMPLTOP} \\
 \hline
 \Gamma \vdash \text{body} \triangleright \top
 \end{array}
 \qquad
 \begin{array}{c}
 \text{RESTRCOMPL} \\
 \frac{\mathcal{E} \vdash \{ \top \} s \{ \psi \mid t \} \quad \vdash \{ \psi \} r \leq t_r \quad (t + t_r \cdot \mathbb{1}_{\text{conc}}) \leq_{\text{compl}} \lambda_c}{\mathcal{E} \vdash \{ \_ ; s; \text{return } r \} \triangleright \lambda_c}
 \end{array}$$

**Notes:** the relation  $\sqsubseteq$  checks the inclusion of a memory restriction into another, and is defined in Fig. 22. Also,  $\text{mem}_\Gamma(s)$  computes an over-approximation of a instruction's memory footprint, and is defined in Fig. 23.

Fig. 8. Restriction checking rules.

complexity restrictions.

$$\begin{aligned}
 \text{compl}[\text{intr} : k, y_1.f_1 : k_1, \dots, y_l.f_l : k_l][x \mapsto \_] = \\
 \text{compl}[\text{intr} : k, (y_1.f_1 : k_1)[x \mapsto \_], \dots, (y_l.f_l : k_l)[x \mapsto \_]] \\
 \text{where } (y.f : k)[x \mapsto \_] = \begin{cases} \epsilon & \text{if } y = x \\ y.f : k & \text{otherwise} \end{cases}
 \end{aligned}$$

Also, note that when substituting  $x$  into  $p$  in  $p.y$ , we do not substitute the module component identifier  $y$  (essentially, only top-level module names are substituted). Similarly, when we substitute  $x$  into  $p$  in a module declaration (module  $y = m$ ), we ignore  $y$ .

*Other typing rules.* The typing judgment for module expressions  $\Gamma \vdash_p m : M$  states that the module expression  $m$ , declared at path  $p$ , has type  $M$ . Functor are typed by the **FUNC** rule. Note that the functor body is typed in an extended typing environment, where the module parameter  $x$  has been declared as an abstract module with kind **param**.

The typing judgment for module structures  $\Gamma \vdash_{p,\theta} st : S$  is annotated by both the module path of the structure being typed, and the module restriction  $\theta$  that the structure must verify. Remark that when we type a procedure using **PROCDECL**, we check that the procedure  $f$  body satisfies the module restriction  $\theta[f]$  by requiring that the restriction checking judgment  $\Gamma \vdash \text{body} \triangleright \theta[f]$  holds.

*Restrictions.* The restriction checking rules are given in Fig. 8. The **RESTRMEM** rule checks that a procedure body  $\{ \_ ; s; \text{return } e \}$  (where  $s$  is the procedure's instructions, and  $e$  the returned expression) verifies a *memory* restriction through a fully automatic syntactic check done in the auxiliary rules **RESTRMEMS** and **RESTRMEME**.

$$\begin{array}{c}
 \text{RESTRMEM} \\
 \frac{\Gamma \vdash s \triangleright \lambda_m \quad \Gamma \vdash e \triangleright \lambda_m}{\Gamma \vdash \{ \_ ; s; \text{return } e \} \triangleright \lambda_m}
 \end{array}
 \qquad
 \begin{array}{c}
 \text{RESTRMEMS} \\
 \frac{\text{mem}_\Gamma(s) \sqsubseteq \lambda_m}{\Gamma \vdash s \triangleright \lambda_m}
 \end{array}
 \qquad
 \begin{array}{c}
 \text{RESTRMEME} \\
 \frac{\text{vars}(e) \sqsubseteq \lambda_m}{\Gamma \vdash e \triangleright \lambda_m}
 \end{array}$$

This syntactic check uses  $\text{mem}_\mathcal{E}(s)$  and  $\text{vars}(e)$ , which are sound over-approximations of an instruction and expression memory footprint (the approximation is not complete, e.g. it will include memory accesses done by unreachable code).

The **RESTRCOMPL** rule checks that an instruction verifies some *complexity* restriction. The rule generates proof obligations in a Hoare logic for cost. These proof obligations are discharged

**Module path resolution**  $\text{res}_\Gamma(p)$  **to module expression**

$$\begin{aligned} \text{res}_\Gamma(p) &= \text{res}_\Gamma(\bar{m}) && (\text{if } \Gamma(p) = \bar{m} : \_) \\ \text{res}_\Gamma(p.x) &= \text{res}_\Gamma(m) && (\text{if } \text{res}_\Gamma(p) = \text{struct } st_1; \text{ module } x = m : M; st_2 \text{ end}) \\ \text{res}_\Gamma(p(p')) &= \text{res}_\Gamma(m_0[x \mapsto p']) && (\text{if } \text{res}_\Gamma(p) = \text{func}(x : M) m_0) \\ \text{res}_\Gamma(p(p')) &= (\text{abs}_{\mathcal{K}} x)(\bar{p}_0, p') && (\text{if } \text{res}_\Gamma(p) = (\text{abs}_{\mathcal{K}} x)(\bar{p}_0)) \end{aligned}$$

**Module expression resolution**  $\text{res}_\Gamma(\bar{m})$

$$\begin{aligned} \text{res}_\Gamma(\text{struct } st \text{ end}) &= \text{struct } st \text{ end} \\ \text{res}_\Gamma(\text{func}(x : M) m) &= \text{func}(x : M) m \\ \text{res}_\Gamma((\text{abs}_{\mathcal{K}} x)(\bar{p})) &= (\text{abs}_{\mathcal{K}} x)(\bar{p}) \end{aligned}$$

**Module procedure resolution**  $f\text{-res}_\Gamma(m.f)$

(note that this includes resolution for function paths  $f\text{-res}_\Gamma(p.f)$ )

$$\begin{aligned} f\text{-res}_\Gamma(p.f) &= (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}) && (\text{if } \Gamma(p.f) = (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body})) \\ f\text{-res}_\Gamma(m.f) &= (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}) && (\text{if } \text{res}_\Gamma(m) = \text{struct } st_1; \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}; st_2 \text{ end}) \\ f\text{-res}_\Gamma(m.f) &= (\text{abs}_{\mathcal{K}} x)(\bar{p}).f && (\text{if } \text{res}_\Gamma(m) = (\text{abs}_{\mathcal{K}} x)(\bar{p})) \end{aligned}$$

Fig. 9. Resolution functions for paths, module expressions and module procedure.

interactively using the proof system we present later, in Section 4.

$$\frac{\text{RESTRCOMPL} \quad \mathcal{E} \vdash \{\top\} s \{\psi \mid t\} \quad \vdash \{\psi\} r \leq t_r \quad (t + t_r \cdot \mathbb{1}_{\text{conc}}) \leq_{\text{compl}} \lambda_c}{\mathcal{E} \vdash \{\_ ; s ; \text{return } r\} \triangleright \lambda_c}$$

Here, the proof obligation  $\mathcal{E} \vdash \{\top\} s \{\psi \mid t\}$  states that the execution of  $s$  in any memory has a complexity upper bounded by  $t$ , and that the post-condition  $\psi$  holds after  $s$ 's execution. The proof obligation  $\vdash \{\psi\} r \leq t_r$  upper-bounds the cost of evaluating the return expression  $r$ . Finally, the rule checks that the sum of  $t$  and  $t_r$  is compatible with the complexity restriction  $\lambda_c$  through the premise  $(t + t_r \cdot \mathbb{1}_{\text{conc}}) \leq_{\text{compl}} \lambda_c$ . We leave the precise definition of  $\leq_{\text{compl}}$  to Section 4 (see Fig. 13). Intuitively,  $t$  is a record of entries of the form  $(x.f \mapsto l_f)$ , each stating that the abstract module  $x$ 's procedure  $f$  has been called at most  $l_c$  times, plus a special entry  $(\text{conc} \mapsto l_c)$  stating that  $s$  execution time, excluding abstract calls, is at most  $l_c$ . Then,  $t_0 \leq_{\text{compl}} \lambda_c$  checks that  $t_0[x.f] \leq \lambda_c[x.f]$  for every functor parameter  $x.f$ , and that  $\lambda_c[\text{intr}]$  upper-bounds everything else in  $t_0$ .

*Remark 3.1.* Note that the complexity checking rule **RESTRCOMPL** is *not* extended to typing environment, because the cost Hoare judgment  $\mathcal{E} \vdash \{\top\} s \{\psi \mid t\}$  is not defined for typing environment. While we could probably extend **RESTRCOMPL** to allow typing in a *typing environment*  $\Gamma$ , this would complicate a lot the soundness proof of our logic. Indeed, as it stands, we do not need to show closure of Hoare logic derivations under substitution of a module parameter  $x$  of type  $\text{abs}_{\text{param}} : M$  by a concrete module  $m$  of the same type  $M$  (because an environment  $\mathcal{E}$  cannot contain a declaration of an abstract module of kind **param**, only of open modules of kind **open**, which are never substituted, only instantiated). Instead, we only need to show closure under such substitution for *typing judgment* (not Hoare logic derivations), which simplifies the proof.

### 3.4 Module Resolution

We present the semantics of our module system, which we use to give the semantics of our programming language in Section 5. Our module system semantics is given by a module resolution mechanism, which describes how module expression  $m$  are evaluated in a typing environment  $\Gamma$ .

*Extended module resolution.* Because a module expression  $m$  is evaluated in a typing environment  $\Gamma$  that can contain abstract modules (representing open code or functor parameters), the resolved module  $\text{res}_\Gamma(m)$  may not be a module expression according to the syntactic categories defined in Fig. 21. We let extended module expressions be the elements of the form:

$$\tilde{m} ::= m \mid \text{abs}_K x$$

Note that it would not make much sense to extend the syntax of module expressions to allow them to contain abstract modules, as abstract modules of kind **param** are reserved to the type system; and **open** modules must be introduced at the logical level (in the ambient higher-order logic).

*Module resolution.* The resolution function  $\text{res}_\Gamma(\_)$  evaluates a module path, in  $\Gamma$ , into a (resolved) extended module expression, which can be a module structure, a functor, or an (potentially applied) abstract module. Mostly,  $\text{res}_\Gamma(\_)$  take care of functor application through the rules:

$$\begin{aligned} \text{res}_\Gamma(p(p')) &= \text{res}_\Gamma(m_0[x \mapsto p']) && (\text{if } \text{res}_\Gamma(p) = \text{func}(x : M) m_0) \\ \text{res}_\Gamma(p(p')) &= (\text{abs}_K x)(\vec{p}_0, p') && (\text{if } \text{res}_\Gamma(p) = (\text{abs}_K x)(\vec{p}_0)) \end{aligned}$$

(the full definition is in Fig. 9). In the concrete functor case, we must substitute the module identifier  $x$  into a path  $p'$  in a module expression  $m_0$ .

*Example 3.1.* Consider a typing environment  $\Gamma$ , and the path  $x.y(z)(v)(w)$ , which must be read as  $((x.y)(z))(v)(w)$ . Then, assuming that  $\Gamma(z) = \text{abs}_{\text{open}} z$ ,  $\Gamma(v) = m_v$ ,  $\Gamma(w) = \text{abs}_{\text{param}} w$  and:

$$\Gamma(x) = \text{struct module } y = \text{func}(u : \_) u \text{ end}$$

where  $m_v$  is some module expression, then  $\text{res}_\Gamma(x.y(z)(v)(w)) = (\text{abs}_{\text{open}} z)(v, w)$ .

We define the module procedure resolution function  $f\text{-res}_\Gamma(m.f)$ . A resolved module procedure  $f\text{-res}_\mathcal{E}(m.f)$  is: i) either a concrete procedure declaration ( $\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}$ ); ii) or the procedure component  $f$  of a resolved (potentially applied) abstract module  $(\text{abs}_K x)(\vec{p}).f$ .

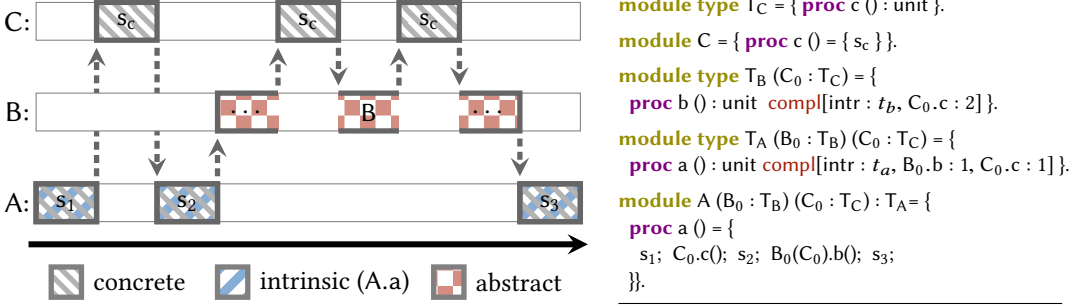
## 4 COMPLEXITY REASONING IN EASYCRYPT

We now present our Hoare logic for cost, which allows to formally prove complexity upper-bounds of programs. This logic manipulates judgment of the form  $\mathcal{E} \vdash \{\phi\} s \{\psi \mid t\}$ , where  $s$  is a statement,  $\phi, \psi$  are assertions, and  $t$  is a cost. We leave the assertion language unspecified, and only require that the models of an assertion formula  $\phi$  are memories, and write  $v \in \phi$  whenever  $v$  satisfies  $\phi$ .

Essentially, the judgment  $\mathcal{E} \vdash \{\phi\} s \{\psi \mid t\}$  states that  $s$  is a program well-typed in the environment  $\mathcal{E}$  (e.g. this means that  $s$  can only call concrete or abstract procedures declared in  $\mathcal{E}$ ), and that: i) the execution of the program  $s$  on any initial memory  $v_i$  satisfying the precondition  $\phi$  (i.e.  $v_i \in \phi$ ) terminates in time at most  $t$ ; and ii), the final memory  $v_f$  obtained by executing  $s$  starting from  $v_i$  satisfies the post-condition  $\psi$  (i.e.  $v_f \in \psi$ ).

### 4.1 Cost Judgment

A key point of our Hoare logic for cost is that it allows to split the cost of a program  $s$  between its concrete and abstract costs, i.e. between the time spent in concrete code, and the time spent in abstract procedures. To reflect this separation between concrete and abstract cost, a cost  $t$  is a record of entries mapping each abstract procedure  $x.f$  to the number of times this procedure was



Judgment  $\mathcal{E} \vdash \{\top\} A(B, C).a \{\top \mid [\text{conc} \mapsto t_{\text{conc}}, B.b \mapsto 1]\}$  where  $\mathcal{E} = (\text{module } B = \text{abs}_{\text{open}} : T_B)$ .

Fig. 10. Graphical representation of the different cost measurements.

called, and mapping a special element `conc` to the concrete execution time (i.e. excluding abstract procedure calls). Since the set of available abstract procedures (and consequently the number of entries in the cost  $t$ ) depends on the current environment  $\mathcal{E}$ , we parameterize the notion of cost by the environment  $\mathcal{E}$  considered:

*Definition 4.1.* A  $\mathcal{E}$ -cost is an element of the form:

$$t ::= [\text{conc} \mapsto k, x_1.f_1 \mapsto k_1, \dots, x_l.f_l \mapsto k_l]$$

where  $\mathcal{E}$  is an environment,  $k, k_1, \dots, k_l$  are integers, and the  $x_i.f_i$  are all the abstract procedures declared in  $\mathcal{E}$ .

*Example 4.1.* Consider  $\mathcal{E}$  with two abstract modules  $x$  and  $y$ :

$$\mathcal{E} = (\text{module } x = \text{abs}_{\text{open}} : \text{sig} (\text{proc } f \_ ) \text{ restr } \_ \text{ end}; \\ (\text{module } y = \text{abs}_{\text{open}} : \text{sig} (\text{proc } h \_ ) \text{ restr } \_ \text{ end})$$

Then  $[\text{conc} \mapsto 10; x.f \mapsto 0; y.h \mapsto 3]$  represents a concrete cost of 10, at most three calls to  $y.h$ , and none to  $x.f$ .

*Definition 4.2.* A *cost judgment* for a statement is an element of the form  $\mathcal{E} \vdash \{\phi\} s \{\psi \mid t\}$  where  $\mathcal{E}$  must be well-typed,  $s$  must be well-typed in  $\mathcal{E}$  and  $t$  must be an  $\mathcal{E}$ -cost. We define similarly a cost judgment for a procedure  $\mathcal{E} \vdash \{\phi\} F \{\psi \mid t\}$ .

In Fig. 10, we give a graphical representation of a cost judgment for the procedure  $A(B, C).a$ , where  $A$  and  $C$  are concrete modules, and  $B$  is an abstract functor with access to  $C$  as a parameter. Then, intuitively, the cost judgment:

$$\mathcal{E} \vdash \{\top\} A(B, C).a \{\top \mid [\text{conc} \mapsto t_{\text{conc}}, B.b \mapsto 1]\}$$

is valid whenever  $t_{\text{conc}}$  upper-bounds the concrete cost (in hatched gray ) which is the sum of: i) the intrinsic cost of  $A.a$ , which is the cost of  $A.a$  without counting parameter calls, represented in hatched blue in the figure, and must be at most  $t_a$  as stated in  $T_A$ 's restriction; and ii) the sum of the cost of the three calls to  $C.c$ .

The cost of the execution of the abstract procedure  $B.b$  (in hatched red ) , which excludes the two calls  $B.b$  makes to  $C.c$ , are accounted for by the entry  $(B.b \mapsto 1)$  in the cost judgment. Note that it is crucial that this excludes the cost of the two calls to  $C.c$ , which are already counted in the concrete cost  $t_{\text{conc}}$  .



$$\begin{array}{c}
\text{SKIP} \\
\hline
\mathcal{E} \vdash \{\phi\} \text{ skip } \{\phi \mid 0\} \\
\\
\text{WEAK} \\
\hline
\mathcal{E} \vdash \{\phi'\} s \{\psi' \mid t'\} \quad \phi \Rightarrow \phi' \quad \psi' \Rightarrow \psi \quad t' \leq t \\
\hline
\mathcal{E} \vdash \{\phi\} s \{\psi \mid t\} \\
\\
\text{FALSE} \\
\hline
\mathcal{E} \vdash \{\perp\} s \{\psi \mid t\} \\
\\
\text{ASSIGN} \\
\hline
\vdash \{\phi\} e \leq t_e \\
\hline
\mathcal{E} \vdash \{\phi \wedge \psi[x \leftarrow e]\} x \leftarrow e \{\psi \mid t_e\} \\
\\
\text{RAND} \\
\hline
\vdash \{\phi_0\} d \leq t \\
\phi = (\phi_0 \wedge \forall v \in \text{dom}(d). \psi[x \leftarrow v]) \\
\hline
\mathcal{E} \vdash \{\phi\} x \stackrel{\$}{\leftarrow} d \{\psi \mid t\} \\
\\
\text{SEQ} \\
\hline
\mathcal{E} \vdash \{\phi\} s_1 \{\phi' \mid t_1\} \\
\mathcal{E} \vdash \{\phi'\} s_2 \{\psi \mid t_2\} \\
\hline
\mathcal{E} \vdash \{\phi\} s_1; s_2 \{\psi \mid t_1 + t_2\} \\
\\
\text{IF} \\
\hline
\mathcal{E} \vdash \{\phi \wedge e\} s_1 \{\psi \mid t\} \\
\mathcal{E} \vdash \{\phi \wedge \neg e\} s_2 \{\psi \mid t\} \quad \vdash \{\phi\} e \leq t_e \\
\hline
\mathcal{E} \vdash \{\phi\} \text{ if } e \text{ then } s_1 \text{ else } s_2 \{\psi \mid t + t_e\} \\
\\
\text{WHILE} \\
\hline
I \wedge e \Rightarrow c \leq N \\
\forall k, \mathcal{E} \vdash \{I \wedge e \wedge c = k\} s \{I \wedge k < c \mid t(k)\} \quad \forall k \leq N, \vdash \{I \wedge e \wedge c = k\} e \leq t_e(k) \quad \vdash \{I \wedge \neg e\} e \leq t_e(N+1) \\
\hline
\mathcal{E} \vdash \{I \wedge 0 \leq c\} \text{ while } e \text{ do } s \{I \wedge \neg e \mid \sum_{i=0}^N t(i) + \sum_{i=0}^{N+1} t_e(i)\} \\
\\
\text{CALL} \\
\hline
\text{args}_{\mathcal{E}}(F) = \vec{v} \quad \vdash \{\phi[\vec{v} \leftarrow \vec{e}]\} \vec{e} \leq t_e \\
\mathcal{E} \vdash \{\phi\} F \{\psi[x \leftarrow \text{ret}] \mid t\} \\
\hline
\mathcal{E} \vdash \{\phi[\vec{v} \leftarrow \vec{e}]\} x \leftarrow \text{call } F(\vec{e}) \{\psi \mid t_e + t\} \\
\\
\text{CONC} \\
\hline
f\text{-res}_{\mathcal{E}}(F) = (\text{proc } f(\vec{v} : \vec{r}) \rightarrow \tau_r = \{\_ ; s; \text{return } r \}) \\
\mathcal{E} \vdash \{\phi\} s \{\psi[\text{ret} \leftarrow r] \mid t\} \quad \vdash \{\psi\} r \leq t_{\text{ret}} \\
\hline
\mathcal{E} \vdash \{\phi\} F \{\psi \mid t + t_{\text{ret}}\}
\end{array}$$

**Convention:** ret cannot appear in programs (i.e.  $\text{ret} \notin \mathcal{V}$ ).

Fig. 11. Basic rules for cost judgment.

*Expression cost.* We have a second kind of judgment  $\vdash \{\phi\} e \leq t_e$ , which states that the cost of evaluating  $e$  in any memory satisfying  $\phi$  is at most  $t_e$ , where  $t_e$  is an *integer*, not a  $\mathcal{E}$ -cost (indeed, an expression cost is always fully concrete, as expressions do not contain procedure calls). We do not provide a complete set of rules for such judgments, as this depends on low-level implementation details and choices, such as data-type representation and libraries implementations. In practice, we give rules for some built-ins, a way for the user to add new rules, and an automatic rewriting mechanism which automatically prove such judgments from the user rules in most cases.

## 4.2 Hoare Logic for Cost Judgment

We present our Hoare logic for cost, which allows to prove cost judgments of programs. Our logic has one rule for each possible program construct (assignment, loop,...), plus some structural rules (e.g. weakening). We start by describing the rules for the basic program constructs (the rules can be found in Fig. 11).

Basically, our cost judgment are standard Hoare logic judgment with the additional cost information, and both aspects must be handled by the rules of our logic.

In some cases, these can be handled separately. E.g. the rule:

$$\begin{array}{c}
\text{IF} \\
\hline
\vdash \{\phi\} e \leq t_e \quad \mathcal{E} \vdash \{\phi \wedge e\} s_1 \{\psi \mid t\} \quad \mathcal{E} \vdash \{\phi \wedge \neg e\} s_2 \{\psi \mid t\} \\
\hline
\mathcal{E} \vdash \{\phi\} \text{ if } e \text{ then } s_1 \text{ else } s_2 \{\psi \mid t + t_e\}
\end{array}$$

state that if: i) the evaluation of the condition  $e$  takes time at most  $t_e$ ; ii) the execution of the then branch program  $s_1$ , assuming pre-condition  $\phi \wedge e$ , guarantees the post-condition  $\psi$  and takes time at most  $t$ ; iii) and the execution of the else branch, assuming the pre-condition  $\phi \wedge \neg e$ , guarantees the same post-condition  $\psi$ , and also takes time at most  $t$ ; then the full conditional statement **if**  $e$  **then**  $s_1$  **else**  $s_2$ , assuming pre-condition  $\phi$ , guarantees the post-condition  $\psi$  in time at most

$$\begin{array}{c}
 \text{Abs} \\
 \frac{
 \begin{array}{l}
 \text{f-res}_{\mathcal{E}}(F) = (\text{abs}_{\text{open}} \ x)(\vec{p}).f \quad \mathcal{E}(x) = \text{abs}_{\text{open}} \ x : (\text{func}(\vec{y} \ _) \ \text{sig} \ \_ \ \text{restr} \ \theta \ \text{end}) \\
 \theta[f] = \lambda_m \wedge \lambda_c \quad \lambda_c = \text{compl}[\text{intr} : K, z_{j_1}.f_1 : K_1, \dots, z_{j_l}.f_l : K_l] \quad \text{FV}(I) \cap \lambda_m = \emptyset \\
 \vec{k} \ \text{fresh in } I \quad \forall i, \forall \vec{k} \leq (K_1, \dots, K_l), \vec{k}[i] < K_i \rightarrow \mathcal{E} \vdash \{I \ \vec{k}\} \vec{p}[j_i].f_i \ \{I(\vec{k} + \mathbb{1}_i) \mid t_i \ k\}
 \end{array}
 }{
 \mathcal{E} \vdash \{I \ \vec{0}\} \ F \ \{\exists \vec{k}, I \ \vec{k} \wedge \vec{0} \leq \vec{k} \leq (K_1, \dots, K_l) \mid T_{\text{abs}}\}
 } \\
 \text{where } T_{\text{abs}} = \{x.f \mapsto 1; (G \mapsto \sum_{i=1}^l \sum_{k=0}^{K_i-1} (t_i \ k)[G])_{G \neq x.f}\}
 \end{array}$$

**Conventions:**  $\vec{y}$  can be empty (this corresponds to the non-functor case).

Fig. 12. Abstract call rule for cost judgment.

$t + t_e$ . Note that we use the same cost upper-bound  $t$  for both branches: essentially,  $t$  can be chosen to be the maximum of the execution times of both branches.

We designed rules for all basic constructs of the logic. E.g. the assignment rule **ASSIGN** lets the user provide a dedicated pre-condition  $\phi$  used to upper-bound the cost of evaluating  $e^4$ ; and the weakening rule **WEAK** is the standard Hoare logic weakening rule, with an additional premise  $t' \leq t$ .

Other rules are more involved, and require the user to show simultaneously invariants of the memory state of the program and cost upper-bounds. This is the case of the abstract call rule, and of the instantiation rule.

### 4.3 Abstract Call Rule

The abstract call rule allows to upper-bound the cost of a call to an abstract procedure  $F$ . To ease the presentation, we first present a version of the rule for usual Hoare judgment without costs, and explain how to add costs after.

$$\begin{array}{c}
 \text{ABS-PARTIAL} \\
 \frac{
 \begin{array}{l}
 \text{f-res}_{\mathcal{E}}(F) = (\text{abs}_{\text{open}} \ x)(\vec{p}).f \quad \mathcal{E}(x) = \text{abs}_{\text{open}} \ x : (\text{func}(\vec{y} \ _) \ \text{sig} \ \_ \ \text{restr} \ \theta \ \text{end}) \\
 \theta[f] = \lambda_m \wedge \_ \quad \text{FV}(I) \cap \lambda_m = \emptyset \quad \forall p_0 \in \vec{p}, \forall g \in \text{procs}_{\mathcal{E}}(p_0), \mathcal{E} \vdash \{I\} \ p_0.g \ \{I\}
 \end{array}
 }{
 \mathcal{E} \vdash \{I\} \ F \ \{I\}
 }
 \end{array}$$

First, the function path  $F$  is resolved to  $(\text{abs}_{\text{open}} \ x)(\vec{p}).f$ , i.e. a call to the procedure  $f$  of an abstract functor  $x$  applied to the modules  $\vec{p}$  (the case where  $x$  is not a functor is handled by taking  $\vec{p} = \epsilon$ ). Then,  $x$ 's module type is lookup in  $\mathcal{E}$ , and we retrieve the module restriction  $\theta$  attached to it. The rule allows to prove that some formula  $I$  is an invariant of the abstract call, by showing two things.

First, we show that  $I$  is an invariant of  $x.f$ , excluding calls to the functor parameters. This is done by checking that  $x.f$  cannot access the variables used in  $I$ , using its memory restriction  $\lambda_m$  (looked-up by the premise  $\theta[f] = \lambda_m \wedge \_$ ) and requiring that  $\text{FV}(I) \cap \lambda_m = \emptyset$ .

Then, we prove that  $I$  is an invariant of  $x.f$ 's calls to functor parameters. This is guaranteed by requiring that for every functor parameter  $p_0 \in \vec{p}$ , for any of  $p_0$ 's procedure  $g \in \text{procs}_{\mathcal{E}}(p_0)$ , the judgment  $\mathcal{E} \vdash \{I\} \ p_0.g \ \{I\}$  is valid.

*Abstract Call Rule.* We now present our **Abs** rule for *cost judgments*, which is given in Fig. 12. Essentially, the cost of the call to  $x(\vec{p}).f$  is decomposed between:

- the intrinsic cost of  $x.f$  excluding the cost of the calls to  $x$ 's functor parameters. This is accounted for by the entry  $(x.f \mapsto 1)$  in the final cost  $T_{\text{abs}}$ .
- the cost of the calls to  $x.f$  functor parameters, which are enumerated in the restriction:

$$\lambda_c = \text{compl}[\text{intr} : K, z_{j_1}.f_1 : K_1, \dots, z_{j_l}.f_l : K_l]$$

<sup>4</sup>If the rule forced to take  $\phi = \psi[x \leftarrow e]$ , then it would not be complete, as prior information on the value on  $x$  (e.g. coming from a previous assignment to  $x$ ) is erased, which may prevent us from proving a precise upper-bound on  $\vdash \{\phi\} \ e \leq t_e$ .

## INSTANTIATION

$$\frac{\begin{array}{l} M_1 = \text{func}(\vec{\gamma} : \vec{M}) \text{ sig } S_1 \text{ restr } \theta \text{ end} \quad \mathcal{E} \vdash_x m : \text{erase}_{\text{compl}}(M_1) \\ \vec{z} \text{ fresh in } \mathcal{E} \quad \forall f \in \text{procs}(S_1), (\mathcal{E}, \text{module } \vec{z} : \text{abs}_{\text{open}} \vec{M} \vdash \{\top\} m(\vec{z}).f \{\top \mid t_f\}) \\ \forall f \in \text{procs}(S_1), t_f \leq_{\text{compl}} \theta[f] \quad \mathcal{E}, \text{module } x = \text{abs}_{\text{open}} : M_1 \vdash \{\phi\} s \{\psi \mid t_s\} \end{array}}{\mathcal{E}, \text{module } x = m : M_1 \vdash \{\phi\} s \{\psi \mid T_{\text{ins}}\}}$$

where:

$$\begin{aligned} T_{\text{ins}} &= \{G \mapsto t_s[G] + \sum_{f \in \text{procs}(S_1)} t_s[x.f] \cdot t_f[G]\} \\ t_f \leq_{\text{compl}} \theta[f] &= \forall z_0 \in \vec{z}, \forall g \in \text{procs}(\vec{M}[z_0]), t_f[z_0.g] \leq \theta[f][z_0.g] \wedge \\ &\quad t_f[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}) \\ h \in \text{procs}_{\mathcal{E}}(A)}} t_f[A.h] \cdot \text{intr}_{\mathcal{E}}(A.h) \leq \theta[f][\text{intr}] \end{aligned}$$

**Conventions:**  $\text{intr}_{\mathcal{E}}(A.h)$  is the  $\text{intr}$  field in the complexity restriction of the abstract module procedure  $A.h$  in  $\mathcal{E}$ .

Fig. 13. Instantiation rule for cost judgment.

We require, for every  $i$ , a bound on the cost of the  $k$ -th call to the functor argument  $z_{j_i}$  procedure's  $f_i$ , where  $k$  can range anywhere between 0 and the maximum number of calls  $x.f$  can make to  $z_{j_i}$ , which is  $K_i$ . The cost of the  $k$ -th call to  $z_{j_i}.f_i$  is bounded by  $(t_i k)$  where  $k = \vec{k}[i]$  and:

$$\mathcal{E} \vdash \{I \vec{k}\} \vec{p}[j_i].f_i \{I (\vec{k} + \mathbb{1}_i) \mid t_i k\}$$

To improve precision, we let the invariant  $I$  depend on the number of calls to the functor parameters through the integer vector  $\vec{k}$ . After calling  $\vec{p}[j_i].f_i$ , we update  $\vec{k}$  by adding one to its  $i$ -th entry ( $\mathbb{1}_i$  is the vector where the  $i$ -th entry is one and all other entries are zero).

The final cost  $T_{\text{abs}}$  (except for  $x.f$ ) is obtained by taking the sum, over all functor parameters and number of calls to this functor parameter, of the cost of each call.

#### 4.4 Instantiation Rule

The **INSTANTIATION** rule, given in Fig. 13, allows to instantiate an abstract module  $x$  by a concrete module  $m$ . Assume that we can upper-bound the cost of a statement  $s$  by  $t_s$ , when  $x$  is abstract:

$$\mathcal{E}, \text{module } x = \text{abs}_{\text{open}} : M_1 \vdash \{\phi\} s \{\psi \mid t_s\}$$

Then we can instantiate  $x$  by a concrete module  $m$  as long as  $m$  complies with the module signature  $M_1$ , which is checked through two conditions.

First, we check that  $m$  has the correct module type, except for complexity restrictions, through the premise  $\mathcal{E} \vdash_x m : \text{erase}_{\text{compl}}(M_1)$

Then, we check that  $m$  satisfies the complexity restriction  $\theta$  in  $M_1$ , by requiring that for any procedure  $f$  of  $x$ :

$$\mathcal{E}, \text{module } \vec{z} : \text{abs}_{\text{open}} \vec{M} \vdash \{\top\} m(\vec{z}).f \{\top \mid t_f\}$$

where  $t_f$  must respect  $\theta[f]$ , which is guaranteed by  $t_f \leq_{\text{compl}} \theta[f]$ , which does two checks:

- first, it ensures that the number of calls to any functor parameter  $z_0$  of  $x$  done by  $m.f$  is upper-bounded by  $\theta[f][z_0]$ .
- then, it verifies that the bound of  $x$ 's intrinsic cost  $\theta[f][\text{intr}]$  upper-bounds the cost of the execution of  $m.f$ , excluding functor parameter calls, through the condition:

$$t_f[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}) \\ h \in \text{procs}_{\mathcal{E}}(A)}} t_f[A.h] \cdot \text{intr}_{\mathcal{E}}(A.h) \leq \theta[f][\text{intr}]$$

where  $\text{intr}_{\mathcal{E}}(A.h)$  is the upper-bound on  $A.h$  intrinsic cost declared in  $\mathcal{E}$  (if  $A.h$  declares no intrinsic bound in  $\mathcal{E}$ , then  $\text{intr}_{\mathcal{E}}(A.h)$  is undefined (hence  $A.h$  execution time can be arbitrarily large), and the **INSTANTIATION** rule cannot be applied). In other words, the concrete execution time  $t_f[\text{conc}]$  of  $x.f$ , plus the abstract execution time of  $x.f$  (excluding functor parameters, already accounted for), must be bounded by  $\theta[f][\text{intr}]$ .

The final cost  $T_{\text{ins}}$  (in Fig. 13) is the sum of the cost  $t_s$  of  $s$  (which excludes the cost of  $x$ 's procedures), plus the sum, for any procedure  $f$  of  $x$ , of the number of times  $s$  called  $x.f$  (which is  $t_s[x.f]$ ), times the cost of  $x.f$  (which is  $t_f$ ).

## 5 INSTRUMENTED SEMANTICS

We now define the denotational semantics of our programming language and cost judgments. We first quickly introduce the main aspects of our semantics below, before defining it formally in the rest of the section. We use this semantics to state our main soundness theorem. The proof of our main theorem is postponed to Appendix D.

*Program semantics.* The semantics  $\llbracket s \rrbracket_v^{\mathcal{E}, \rho}$  of our language depends on the initial memory  $v$ , the environment  $\mathcal{E}$ , and on the interpretation  $\rho$  of  $\mathcal{E}$ 's abstract modules. Essentially,  $\llbracket s \rrbracket_v^{\mathcal{E}, \rho}$  is a discrete distribution over  $\mathcal{M} \times \mathbb{N}$ , where the integer component is the cost of evaluating  $s$  in  $(\mathcal{E}, \rho)$ , starting from the memory  $v$ . Then, the  $\mathcal{E}$ -cost of an instruction  $s$  under memory  $v$  and interpretation of  $\mathcal{E}$ 's abstract modules  $\rho$ , denoted by  $\text{cost}_v^{\mathcal{E}, \rho}(s) \in \mathbb{N} \cup \{+\infty\}$ , is the maximum execution cost in any final memory, defined as:

$$\text{cost}_v^{\mathcal{E}, \rho}(s) = \inf \{c' \mid \Pr((\_, c) \leftarrow \llbracket s \rrbracket_v^{\mathcal{E}, \rho}; c \leq c') = 1\} \quad (1)$$

*Judgments semantics.* Basically, the judgment  $\mathcal{E} \vdash \{\phi\} s \{\psi \mid t\}$  states that: i) the memory  $v$  obtained after executing  $s$  in an initial memory  $v \in \phi$  must satisfy  $\psi$ ; ii) the complexity of the instruction  $s$  is upper-bounded by the complexity of the concrete code in  $s$ , plus the sum over all abstract oracles  $A.f$  of the number of calls to  $A.f$  times the intrinsic complexity of  $A.f$ . Formally:

$$\text{cost}_v^{\mathcal{E}, \rho}(s) \leq t[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}) \\ f \in \text{procs}(\mathcal{E}(A))}} t[A.f] \cdot \text{compl}_{A.f}^{\mathcal{E}, \rho}$$

where  $\text{compl}_{A.f}^{\mathcal{E}, \rho}$  is the intrinsic complexity of the procedure  $A.f$ , i.e. its complexity excluding calls to  $A$ 's functor parameters.

### 5.1 Semantics

We now present the semantics of our programs. For any set  $A$ , we denote by  $\mathbb{D}(A)$  the set of discrete sub-distributions over  $A$  — i.e. the set of function  $\mu : A \rightarrow [0, 1]$  with discrete support s.t.  $\mu$  is summable and  $|\mu| = \sum_x \mu(x) \leq 1$ . For  $x \in A$ , the *Dirac distribution at  $x$*  is written  $\mathbb{1}_x^A$  or  $\mathbb{1}_x$  when  $A$  is clear from the context. If  $\mu \in \mathbb{D}(A)$  and  $\mu' \in A \rightarrow \mathbb{D}(B)$ , the expected distribution of  $\mu'$  in  $\mathbb{D}(B)$  when ranging over  $\mu$ , written  $\mathbb{E}_{x \sim \mu}[\mu'(x)]$  or  $\mathbb{E}_{\mu}[\mu']$ , is defined as  $\mathbb{E}_{\mu}[\mu'] = b \in B \mapsto \sum_{a \in A} \mu(a) \mu'(a)(b)$ . For  $\mu' \in \mathbb{D}(A)$  and  $f : A \rightarrow B$ , the marginal of  $\mu'$  w.r.t.  $f$ , written  $f^{\#}(\mu') \in \mathbb{D}(B)$ , is defined as  $f^{\#}(\mu') = b \mapsto \sum_{a \in A \mid f(a)=b} \mu'(a)$ . We write  $\pi_1^{\#}$  (resp.  $\pi_2^{\#}$ ) for resp. the first and second marginal — i.e. when  $f$  is resp. the first and second projection. For any base type  $\tau \in \mathbb{B}$ , we assume an interpretation domain  $\mathbb{V}_{\tau}$ . We let  $\mathbb{V}$  be the set of all possible values  $\cup_{\tau \in \mathbb{B}} \mathbb{V}_{\tau}$ . A memory  $v \in \mathcal{M}$  is a function from  $\mathcal{V}$  to  $\mathbb{V}$ . We write  $v[x]$  for  $v(x)$ . For  $v \in \mathcal{M}$  and  $v \in \mathbb{V}$ , we write  $v[x \leftarrow v]$  for the memory that maps  $x$  to  $v$  and  $y$  to  $v[y]$  for  $y \neq x$ .

*Expressions semantics.* For any operator  $f \in \mathcal{F}_{\mathcal{E}}$  with type  $\tau_1 \times \dots \times \tau_n \rightarrow \tau$ , we assume given its semantics  $(f) : \mathbb{V}_{\tau_1} \times \dots \times \mathbb{V}_{\tau_n} \mapsto \mathbb{V}_{\tau}$ , and the cost of its evaluation  $c_{\mathcal{E}}(f, \cdot) : \mathbb{V}_{\tau_1} \times \dots \times \mathbb{V}_{\tau_n} \mapsto \mathbb{N}$ .

$$\begin{aligned}
\llbracket \mathbf{skip} \rrbracket_v^{\mathcal{E}, \rho} &= \mathbb{1}_{(v, 0)} \\
\llbracket \mathbf{abort} \rrbracket_v^{\mathcal{E}, \rho} &= 0 \\
\llbracket s_1; s_2 \rrbracket_v^{\mathcal{E}, \rho} &= \mathbb{E}_{(v', c') \sim \llbracket s_1 \rrbracket_v^{\mathcal{E}, \rho}} [\llbracket s_2 \rrbracket_{v'}^{\mathcal{E}, \rho} \oplus c'] \\
\llbracket x \leftarrow e \rrbracket_v^{\mathcal{E}, \rho} &= \mathbb{1}_{(v[x \leftarrow \langle e \rangle_v], c_E(e, v))} \\
\llbracket x \xleftarrow{\$} d \rrbracket_v^{\mathcal{E}, \rho} &= \mathbb{E}_{v \sim \langle d \rangle_v} [\mathbb{1}_{(v[x \leftarrow v], c_D(d, v))}] \\
\llbracket \mathbf{if } e \mathbf{ then } s_1 \mathbf{ else } s_2 \rrbracket_v^{\mathcal{E}, \rho} &= \begin{cases} \llbracket s_1 \rrbracket_v^{\mathcal{E}, \rho} \oplus c_E(e, v) & \text{if } \langle e \rangle_v \neq 0 \\ \llbracket s_2 \rrbracket_v^{\mathcal{E}, \rho} \oplus c_E(e, v) & \text{otherwise} \end{cases} \\
\llbracket \mathbf{while } e \mathbf{ do } s \rrbracket_v^{\mathcal{E}, \rho} &= \lim_{n \rightarrow \infty} \llbracket \mathbf{loop}_n^{e, s} \rrbracket_v^{\mathcal{E}, \rho} \\
\text{where } \mathbf{loop}_{n+1}^{e, s} &= \mathbf{if } e \mathbf{ then } (s; \mathbf{loop}_n^{e, s}) \mathbf{ else skip} \\
\text{and } \mathbf{loop}_0^{e, s} &= \mathbf{if } e \mathbf{ then abort else skip}
\end{aligned}$$

Moreover, if  $f\text{-res}_{\mathcal{E}}(m.f) = \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \_ ; s ; \text{return } r \}$ :

$$\begin{aligned}
\llbracket x \leftarrow \mathbf{call } m.f(\vec{e}) \rrbracket_v^{\mathcal{E}, \rho} &= \text{let } v_0 = v[\vec{v} \leftarrow \langle \vec{e} \rangle_v] \text{ in} \\
&\quad \mathbb{E}_{(v', c') \sim \llbracket s \rrbracket_{v_0}^{\mathcal{E}, \rho}} [\mathbb{1}_{v'[x \leftarrow \langle r \rangle_{v'}], c' + c_E(\vec{e}, v) + c_E(r, v')}]
\end{aligned}$$

And if  $f\text{-res}_{\mathcal{E}}(m.f) = (\mathbf{abs}_{\text{open}} x)(\vec{p}).f$ :

$$\llbracket x \leftarrow \mathbf{call } m.f(\vec{e}) \rrbracket_v^{\mathcal{E}, \rho} = \llbracket x \leftarrow \mathbf{call } \rho(x)(\vec{p}).f(\vec{e}) \rrbracket_v^{\mathcal{E}, \rho}$$

Fig. 14.  $(\mathcal{E}, \rho)$ -denotational semantics  $\llbracket \_ \rrbracket_v^{\mathcal{E}, \rho}$ .

The semantics  $\langle e \rangle_v : \mathcal{M} \rightarrow \mathbb{V}$  of a well-typed expression  $e$  in a memory  $v$  is defined inductively by:

$$\langle e \rangle_v = \begin{cases} v(x) & \text{if } e = x \in \mathcal{V} \\ \langle f \rangle(\langle e_1 \rangle_v, \dots, \langle e_n \rangle_v) & \text{if } e = f(e_1, \dots, e_n) \end{cases}$$

And the cost of the evaluation of a well-typed expression  $c_E(e, \cdot) : \mathcal{M} \mapsto \mathbb{N}$  is defined by:

$$c_E(e, v) = \begin{cases} 1 & \text{if } e = x \in \mathcal{V} \\ c_f + \sum_{1 \leq i \leq n} c_E(e_i, v) & \text{if } e = f(e_1, \dots, e_n) \\ \text{and } c_f = c_E(f, \langle e_1 \rangle_v, \dots, \langle e_n \rangle_v) & \end{cases}$$

For technical reasons, we assume that there exists one operator with a non-zero cost.<sup>5</sup>

For any distribution operator  $d \in \mathcal{F}_D$  with type  $\tau_1 \times \dots \times \tau_n \rightarrow \tau$ , we assume given its semantics  $\langle d \rangle : \mathbb{V}_{\tau_1} \times \dots \times \mathbb{V}_{\tau_n} \mapsto \mathbb{D}(\mathbb{V}_{\tau})$ , and the cost of its evaluation  $c_D(d, \cdot) : \mathbb{V}_{\tau_1} \times \dots \times \mathbb{V}_{\tau_n} \mapsto \mathbb{N}$ . We define similarly  $\langle d \rangle_v : \mathcal{M} \rightarrow \mathbb{D}(\mathbb{V})$  and  $c_D(d, \cdot) : \mathcal{M} \mapsto \mathbb{N}$ .

*Environment and  $\mathcal{E}$ -pre-interpretation.* To give the semantics of a program in an environment  $\mathcal{E}$ , we need an interpretation of  $\mathcal{E}$ 's abstract modules. A  $\mathcal{E}$ -pre-interpretation is a function  $\rho$  from  $\mathcal{E}$ 's abstract modules to module expressions, with the correct types, *except for complexity restrictions*. We will specify what it means for a module expression to verify a complexity restriction later, after having defined the semantics of our language.

<sup>5</sup>Some of our lemmas do not hold if all programs have a cost of zero.

*Definition 5.1.* Let  $\text{erase}_{\text{compl}}(M)$  be the module signature  $M$  where every complexity restriction  $\lambda_c$  has been erased, by replacing it by  $\top$ . Then  $\rho$  is a  $\mathcal{E}$ -pre-interpretation if and only if for every  $x$  such that  $\mathcal{E} = \mathcal{E}_1$ ; module  $x = \text{abs}_{\text{open}} : M_1; \mathcal{E}_2$ , we have  $\mathcal{E}_1 \vdash_{\epsilon} \rho(x) : \text{erase}_{\text{compl}}(M_1)$ .

Note that we type  $\rho(x)$  in  $\mathcal{E}_1$ , which lets the interpretation of  $x$  use any module or abstract module declared before  $x$  in  $\mathcal{E}$ .

*Programs semantics.* If  $\mu \in \mathbb{D}(\mathcal{M} \times \mathbb{N})$  and  $n \in \mathbb{N}$ , we write  $\mu \oplus n$  for the distribution  $f^{\#}(\mu)$  where  $f : (m, c) \mapsto (m, c + n)$ . Let  $\mathcal{E}$  be a well-typed environment, and  $s$  be a well-typed instruction in  $\mathcal{E}$ , i.e. such that  $\mathcal{E} \vdash s$ . The  $\mathcal{E}$ -denotational semantics of an instruction  $s$  under the memory  $v$  and  $\mathcal{E}$ -pre-interpretation  $\rho$ , written  $\llbracket s \rrbracket_v^{\mathcal{E}, \rho} \in \mathbb{D}(\mathcal{M} \times \mathbb{N})$ , is defined in Fig. 14.

We give the semantics for an extended syntax, which allows procedure calls to be of the form  $x \leftarrow \text{call } m.f(\vec{e})$  where  $m$  is a module expression. Note that this subsumes the syntax of statements, since a module expression  $m$  can be a module path  $p$ . This allows to concisely define the semantics of a call to an abstract procedure  $(\text{abs}_{\text{open}} x)(\vec{p}).f$  as the semantics of a call to  $\rho(x)(\vec{p}).f$ .

The  $\mathcal{E}$ -cost of an instruction  $s$  under memory  $v$  and  $\mathcal{E}$ -pre-interpretation  $\rho$ , denoted by  $\text{cost}_v^{\mathcal{E}, \rho}(s) \in \mathbb{N} \cup \{+\infty\}$ , is defined as:

$$\text{cost}_v^{\mathcal{E}, \rho}(s) = \sup(\text{supp}(\pi_2^{\#}(\llbracket s \rrbracket_v^{\mathcal{E}, \rho})))$$

where  $\text{supp}$  is the support of a distribution (this definition is equivalent to the one given in Equ. 1).

## 5.2 Cost Judgement Semantics and Soundness of our Proof System

To define the semantics of our cost judgments, we need two additional complexity measures: the number of calls a program execution makes to some abstract procedure, and the intrinsic cost of a program execution (i.e. the cost of the program without the cost of parameters calls). For space reasons, these additional complexity measures are defined in Appendix B.

*Soundness.* We now have all the tools to define the semantics of our expression and program cost judgments.

*Definition 5.2.* the judgment  $\vdash \{\phi\} e \leq t_e$  stands for:

$$\forall v \in \phi, c_E(e, v) \leq t_e$$

*Definition 5.3.* The judgment  $\mathcal{E} \vdash \{\phi\} s \{\psi \mid t\}$  means that for any  $\mathcal{E}$ -interpretation  $\rho$  and  $v \in \phi$ :

$$\text{supp}(\pi_1^{\#}(\llbracket s \rrbracket_v^{\mathcal{E}, \rho})) \subseteq \psi \wedge \text{cost}_v^{\mathcal{E}, \rho}(s) \leq t[\text{conc}] + \sum_{A \in \text{abs}(\mathcal{E})} f \in \text{procs}(\mathcal{E}(A)) t[A.f] \cdot \text{compl}_{A.f}^{\mathcal{E}, \rho}$$

where a  $\mathcal{E}$ -interpretation is a  $\mathcal{E}$ -pre-interpretation mapping abstract modules to concrete module satisfying the required memory and complexity restrictions, and  $\text{compl}_{A.f}^{\mathcal{E}, \rho}$  is the intrinsic complexity of the procedure  $A.f$ , i.e. its complexity excluding calls to  $A$ 's functor parameters (the detailed definition is postponed to Appendix B).

Basically, the complexity of the instruction  $s$  is upper-bounded by the complexity of the concrete code in  $s$ , plus the sum over all abstract oracles  $A.f$  of the number of calls to  $A.f$  times the intrinsic complexity of  $A.f$ .

We are now ready to state our main theorem showing the soundness of our Hoare logic for cost.

**THEOREM 5.1.** *The proof rules in Figures 12, 11 and 13 are sound.*

The proof can be found in Appendix D

## 6 EXAMPLE: UNIVERSAL COMPOSABILITY

UC security guarantees that a protocol  $\pi_1$  can safely replace a protocol  $\pi_2$  while preserving both the functionality and the security of the overall system. The most common application of this framework is to set  $\pi_2$  to be an idealized protocol that assumes a trusted-third-party (TTP) to which parties delegate the computation; the specification of the TTP is called an *ideal functionality*  $\mathcal{F}$ . An ideal functionality  $\mathcal{F}$  defines what protocol  $\pi_1$  should achieve both in terms of correctness and security to securely replace the TTP. Moreover,  $\mathcal{F}$  can be used as an ideal sub-component when designing higher-level protocols, which then can be instantiated with protocol  $\pi_1$  to obtain a fully concrete real-world protocol.

The UC framework defines an execution model where protocol participants, attackers and contexts are modeled as Interactive Turing Machines (ITM). The model was carefully designed to give a good balance between expressive power—e.g., one can capture complex interactions in distributed protocols involving multiple parties in a variety of communication models, various forms of corruption, etc.—and a tailored (and relatively simple) resource analysis mechanism that permits keeping track of the computing resources available to both honest and malicious parties.

The model is described in detail in [17, 18]. However, most UC proofs found in the literature refer only to a common understanding of the semantics of the execution model and a set of high-level restrictions that are inherent to the model. These include the allowed interactions between different machines, the order in which machines are activated, predefined sequences of events, etc. More fine-grained descriptions of the execution model are sometimes introduced locally in proofs, when they are needed to deal with more subtle points or technicalities that can only be clarified at the cost of extra details. This stands in contrast with typical game-based proofs for simpler cryptographic primitives [10], where security proofs are given in great detail. This is one of the reasons why, while there has been impressive progress in machine-checking game-based proofs [5], we are only now giving the first steps in formalizing proofs in the UC setting [21, 26, 33]. Another reason is that the ITM model for communication is difficult to express in procedure-based semantics offered by tools that target game-based proofs.

To overcome these difficulties, we propose a new approach to machine-checking UC proofs that shares many features of the simplified version of UC proposed by Canetti, Cohen and Lindell in [20]. As in [20], we statically fix the machines/modules in the execution model and we allow an adversarial entity to control which module gets to be executed next, rather than allowing machines to pass control between them more freely as in the original UC execution model. The crucial difference to the ITM execution model is that the above interactions are procedure-based, which means that whenever the environment passes control to the protocol, the internal protocol structure will follow a procedure call tree that guarantees (excluding the possibility of non-terminating code) that control returns to the environment.<sup>6</sup> As in [20], we lose some expressiveness, but we do not go as far as hard-wiring a specific communications model for protocols based on authenticated channels; instead, we leave it to the protocol designer to specify the communications model by using an appropriate module structure. We recover the authenticated communications model of [20] by explicitly defining a hybrid real-world, in which concrete modules for ideal authenticated channels are available to the communicating parties. We discuss the trade-offs associated with our approach more in depth at the end of this section, drawing a parallel to the work in [21].

<sup>6</sup>Intuitively, the UC model expresses a single line of execution using a token-passing mechanism that allows one machine to *transfer* computational resources to another, and even to create new machines. In our setting, resource analysis is much simpler. All modules representing honest and adversarial entities are fixed from the start and the cost model is concrete: all adversarial entities have a resource usage type, which means they are known to execute a maximum number of operations and perform a bounded number of procedure calls. Hence the resources used by any subset of modules in our formalizations can be stated as expressions over these type parameters.

<pre> <b>module type</b> IO = {   <b>proc</b> inputs (i:inputs) : unit   <b>proc</b> outputs(o:ask_outputs) : outputs option }.  <b>module type</b> BACKDOORS = {   <b>proc</b> step (m:step) : unit   <b>proc</b> backdoor (m:ask_backdoor) : backdoor option }.         </pre>	<pre> <b>module type</b> E_INTERFACE = {   <b>include</b> IO   <b>include</b> BACKDOORS }.  <b>module type</b> PROTOCOL = {   <b>proc</b> init() : unit   <b>include</b> E_INTERFACE }.         </pre>
--	--

Fig. 15. PROTOCOL type in EasyCrypt.

### 6.1 Mechanized Formalization in EasyCrypt

We propose a natural simplification of the UC execution model that is based on EasyCrypt modules and show that this opens the way for a lightweight formalization of UC proofs. This formalization has been conducted in our extension of EasyCrypt (the proofs of the lemmas and theorems of this section are fully mechanized).

*Protocols and Functionalities as EasyCrypt modules.* The basic component in our UC execution model is a module of type PROTOCOL given in Fig. 15. Inhabitants of this type represent a full real-world configuration—a distributed protocol executed by a fixed number of parties—or an ideal-world configuration—an ideal functionality executing a protocol as a trusted-third party. The type of a protocol has a fixed interface, but it is parametric on the types of values exchanged via this interface. The fixed interface is divided into three parts: i) `init` allows modeling some global protocol setup; ii) `IO` captures the interaction of a higher level protocol using this protocol as a sub-component; and iii) `BACKDOORS` captures the interaction of an adversary with the protocol during its execution.

When we define real-world protocols, a module of type PROTOCOL will be constructed from sub-modules that emulate the various parties and the communications channels between them. In this case, BACKDOORS models adversarial power in this communication model. For ideal-world protocols, a PROTOCOL is typically a flat description of the ideal computation in a single module; here BACKDOORS models unavoidable leakage (e.g., the length of secret inputs or the states of parties in an interactive protocol) and external influence over the operation of the trusted-third party (e.g., blocking the computation to model a possible denial of service attack).<sup>7</sup>

*Execution Model.* The real- and ideal-world configurations are composed by a statically determined set of modules, which communicate with each-other using a set of hardwired interfaces. The execution model is defined by an experiment in which an external environment interacts with the protocol via its IO and BACKDOORS interfaces until, eventually, it outputs a boolean value (Fig. 16). The IO interface allows the environment to pass an input to the protocol using `inputs` or to retrieve an output produced by the protocol using `outputs`. For example in the real-world, the environment can use these procedures to give input to or obtain an output from one of the sub-modules that represent the computing parties involved in the protocol. The BACKDOORS interface allows the environment to read some message that may be produced by the protocol using `backdoor` or make one of the protocol sub-components (parties) advance in its execution using `step` to deliver a message.

<sup>7</sup>Ideal-world backdoors are used to weaken the security requirements and are usually tailored to bring the security definition down to a level that can be met by real-world protocols. Note that the definition of meaningful ideal functionalities is a crucial aspect of UC security theory; here we just provide a mechanism that permits formalizing such definitions in EasyCrypt.



<pre> <b>module</b> UC_emul (E:ENV) (P:PROTOCOL) = {   <b>proc</b> main() = {     <b>var</b> b;     P.init(); b ← E(P).distinguish(); <b>return</b> b;   } </pre>	<pre> <b>module</b> CompS(F:IDEAL.PROTOCOL, S:SIMULATOR)   : PROTOCOL = {   <b>proc</b> init() = { F.init(); S(F).init(); }   <b>include</b> F [ inputs, outputs ]   <b>include</b> S(F) [ step, backdoor ] } </pre>
---	--

Fig. 16. Execution model for real/ideal worlds (top) and composition of functionality with a simulator (bottom).

We describe now the typical sequence of events in a real-world execution; the ideal-world will become clear when we describe the notion of UC emulation below. When the adversarial environment uses the IO interface to pass input to a computing party, this may trigger the computing party to perform some computations and, in turn, provide inputs to other sub-modules included in the protocol description; in most cases this will correspond to sending a message using an idealized communications channel represented by an ideal functionality.<sup>8</sup> Our convention is that inputs calls do not allow obtaining information back (the return type is unit). This means that any outputs produced by parties need to be *pulled* by the environment with separate calls to outputs. Similarly, when the environment asks a party for an output, the party may perform some computation and call the outputs interface of a hybrid ideal functionality (e.g., to see if a message has been delivered) before returning the output to the environment.

The BACKDOORS interface follows these conventions closely. The backdoor method allows the environment to retrieve leakage that may be available for it to collect (e.g., the public part of a party's state, or a buffered message in an authenticated channel). The step procedure allows the environment to pass control to any module inside the protocol; this is important to make sure that the environment always has full control of the liveness of the execution model and can schedule the execution of the various processes at will whenever there are several possible lines of execution.

*UC emulation.* The central notion to Universal Composability is called UC-emulation, which is a relation between two protocols  $\pi_1$  and  $\pi_2$ : if  $\pi_1$  UC-emulates  $\pi_2$  with small advantage  $\epsilon$  then  $\pi_1$  can replace  $\pi_2$  in any context (within a complexity class).

*Definition 6.1 (UC emulation).* Protocol  $\pi_1$  UC emulates  $\pi_2$  under complexity restrictions  $c_{\text{sim}}$  and  $c_{\text{env}}$  and advantage bound  $\epsilon$  if

$$\exists S \in \tau_{\text{sim}}^{\pi_1, \pi_2, c_{\text{sim}}}, \forall Z \in \tau_{\text{env}}^{\pi_1, \pi_2, S, c_{\text{env}}}, |\Pr[\mathcal{Z}(\pi_1) : \top] - \Pr[\mathcal{Z}(\langle \pi_2 \parallel S(\pi_2) \rangle) : \top]| \leq \epsilon$$

We write this as  $\text{Adv}_{c_{\text{sim}}, c_{\text{env}}}^{\text{uc}}(\pi_1, \pi_2) \leq \epsilon$ .

The first probability term corresponds to the event that the environment returns true in the real-world execution model described above, i.e., in game UC\_emul parameterized with  $\text{ENV} = \mathcal{Z}$  and  $P = \pi_1$ . The second probability term corresponds to the equivalent event in the ideal-world (or reference) execution model where, as shown in Fig. 17 (right),  $\pi_2$  is typically an ideal functionality; this corresponds to game UC\_emul parameterized with  $\text{ENV} = \mathcal{Z}$  and a protocol  $P$  that results from attaching  $S$  to the BACKDOORS interface of  $\pi_2$ . We denote this ideal-world  $P$  by  $\langle \pi_2 \parallel S(\pi_2) \rangle$ , corresponding to the EasyCrypt functor CompS also shown in Fig. 16.

UC-emulation imposes that a simulator  $S$  is capable to *fool* any environment by presenting a view that is fully consistent with the real-world, while learning only what the BACKDOORS interface of  $\pi_2$  allows. If such a simulator exists, then clearly  $\pi_2$  cannot be worse than  $\pi_1$  in the information it

<sup>8</sup>Real-world settings using ideal functionalities as sub-components are called *hybrid*.

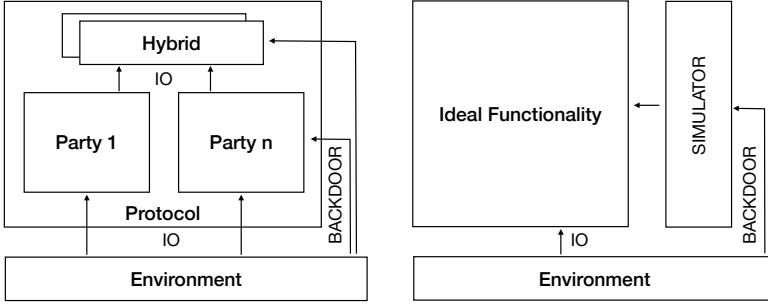


Fig. 17. Module restrictions. Arrows indicate ability to make procedure calls via the interface specified as a label; all other cross-boundary memory access is disallowed.

reveals to the environment via its BACKDOORS interface.<sup>9</sup> Our UC-emulation definition quantifies over simulators and environments using types that give a full characterization of their use of resources, including the ability to access memory, number and types of procedure calls and intrinsic computational costs. The memory access restrictions are depicted in Fig. 17, and they can be easily matched to the standard restrictions in the UC framework. Not shown are the cost restrictions, which give explicit bounds for the resources used by various parts of the execution model; these are crucial for obtaining, not only a meaningful definition, but also for obtaining meaningful reductions to computational assumptions, as will be seen below.

Let us examine the types of  $\mathcal{Z}$  and  $\mathcal{S}$  in more detail. We first note that the definition of emulation is parametric in the resource restrictions  $c_{sim}$  and  $c_{env}$ . Clearly the IO interface of  $\pi_2$  must match the type of the IO interface of  $\pi_1$ , which is consistent with the goal that  $\pi_1$  can replace  $\pi_2$  in any context, and this is enforced by our type system. This need not be the case for the BACKDOORS interface and, in fact, if  $\pi_2$  is an ideal functionality, the BACKDOORS interface in the ideal world is of a different nature altogether than the one in the real world: it specifies leakage and adversarial control that are unavoidable even when the functionality is executed by a trusted third-party on behalf of the parties. The type of the simulator  $\mathcal{S}$  is given by  $\tau_{sim}^{\pi_1, \pi_2, c_{sim}}$ , which defines the type of modules that has access to the BACKDOORS interface of  $\pi_2$ , exposes the BACKDOORS interface of  $\pi_1$  and is restricted memory-wise to exclude the memory of  $\pi_2$  and resource-wise by  $c_{sim}$ . Note that, if  $\mathcal{S}$  could *look inside* the ideal functionality, then it would know all the information that is also given to the real-world protocol: a trivial simulator would always exist and the definition would be meaningless because all protocols would be secure. The type of the environment is given by  $\tau_{env}^{\pi_1, \pi_2, \mathcal{S}, c_{env}}$ , the type of modules that have oracle access to the IO and BACKDOORS interfaces of  $\pi_1$ , and are restricted memory-wise to exclude the memories of  $\pi_1$ ,  $\pi_2$  and  $\mathcal{S}$ , and resource-wise by  $c_{env}$ . In this case, if the environment could *look inside*  $\pi_1$ ,  $\pi_2$  or  $\mathcal{S}$  it could directly detect with which world it is interacting, and no protocol would be secure. For concreteness, the cost restriction on

<sup>9</sup>The emulation notions in [17, 18] quantify over a restricted class of *balanced* environments. Intuitively, such environments must be *fair* to the simulator in that polynomial-time execution in the size of its inputs is comparable to the execution time of the real-world adversary. Without this restriction, the definition would require the existence of a simulator that uses much less resources than the real-world attacker, which makes the definition too strong. Balanced environments guarantee that the resources given to the simulator match those given to the real-world adversary; moreover, the dummy adversary is formally explicit in the real-world to enable this resource accounting. In our setting we deal with this issue differently: the EasyCrypt resource model is concrete, which means that one can explicitly state in the security definition which resources are used by the simulator and assess what this means in terms of protocol security. We refer the interested reader to [17, Section 4.4] for a discussion of quantitative UC definitions such as the one we adopt. For this reason, as we show below, we also do not need to keep the dummy adversary explicitly in the real world.

the type of the environment imposed by  $c_{\text{env}}$  is of the form:

$$c_{\text{env}} := \text{compl}[\text{intr} : c_1, \pi.\text{inputs} : c_2, \pi.\text{outputs} : c_3, \pi.\text{backdoor} : c_4, \pi.\text{step} : c_5]$$

where type refinements can set  $c_i$  to depend on the types of other modules in the context.

## 6.2 Warm-up: Transitivity of UC emulation

It is easy to show that UC-emulation is a transitive relation: if  $\pi_1$  UC-emulates  $\pi_2$  and this, in turn, UC-emulates  $\pi_3$ , then  $\pi_1$  UC-emulates  $\pi_3$ . When stating this lemma in EasyCrypt we move the existential quantifications over the simulators in the hypotheses to global universal quantifications; this logically equivalent formulation allows us to refer to the memory of these simulators when quantifying over all adversarial environments in the consequence: we quantify only over those that cannot *look inside* the simulators that are assumed to exist by hypothesis, which is a natural (and necessary) restriction. In other examples we use the same approach. The lemma is stated in EasyCrypt as follows (we adapt the  $\text{Adv}_{\cdot}^{\text{uc}, \mathcal{S}}(\cdot, \cdot)$  notation by indicating the universally quantified simulator  $\mathcal{S}$  in superscript).

LEMMA 6.1 (TRANSITIVITY). *For all  $\epsilon_{1,2}, \epsilon_{2,3} \in \mathbb{R}^+$ , all protocols  $\pi_1, \pi_2$  and  $\pi_3$  s.t. the IO interfaces of all three protocols are of the same type, all cost restrictions  $c_{\text{sim}(1,2)}, c_{\text{sim}(2,3)}$  and all simulators  $\mathcal{S}_{1,2} \in \tau_{\text{sim}}^{\pi_1, \pi_2, c_{\text{sim}(1,2)}}$ ,  $\mathcal{S}_{2,3} \in \tau_{\text{sim}}^{\pi_2, \pi_3, c_{\text{sim}(2,3)}}$ , we have that:*

$$\begin{aligned} \text{Adv}_{c_{\text{sim}(1,2)}, \hat{c}_{\text{env}(1,2)}}^{\text{uc}, \mathcal{S}_{1,2}}(\pi_1, \pi_2) \leq \epsilon_{1,2} &\Rightarrow \text{Adv}_{c_{\text{sim}(2,3)}, \hat{c}_{\text{env}(2,3)}}^{\text{uc}, \mathcal{S}_{2,3}}(\pi_2, \pi_3) \leq \epsilon_{2,3} \\ &\Rightarrow \text{Adv}_{\hat{c}_{\text{sim}(1,3)}, c_{\text{env}(1,3)}}^{\text{uc}}(\pi_1, \pi_3) \leq \epsilon_{1,2} + \epsilon_{2,3} \end{aligned}$$

where  $\hat{c}_{\text{sim}(1,3)}$  corresponds to the cost of sequentially composing  $\mathcal{S}_{1,2}$  and  $\mathcal{S}_{2,3}$ ,  $\hat{c}_{\text{env}(2,3)}$  must allow for an adversarial environment that results from converting a distinguisher between  $\pi_1$  and  $\pi_3$  in  $c_{\text{env}(1,3)}$  and composing it with  $\mathcal{S}_{1,2}$ , and  $\hat{c}_{\text{env}(1,2)} = c_{\text{env}(1,3)}$ .

In the statement of the lemma we use notation  $\hat{c}$  to denote the fact that these cost restrictions are fixed as a function of the costs of other algorithms: intuitively, the cost of the environment in the consequence is free and it constrains the costs of environments in the hypotheses; then, if for some cost restrictions  $c_{\text{sim}(1,2)}$  and  $c_{\text{sim}(2,3)}$  the hypotheses hold, these in turn fix the cost of the simulator we give as a witness. This pattern is observable in the remaining examples we give in this section.

From the proof, we get a witness simulator  $\mathcal{S}_{1,3} = \text{SeqS}(\mathcal{S}_{2,3}, \mathcal{S}_{1,2})$  that results from plugging together the two simulators implied by the assumptions: intuitively,  $\mathcal{S}_{2,3}$  is able to interact with  $\pi_3$  and emulate the BACKDOORS of  $\pi_2$ , and this is sufficient to enable  $\mathcal{S}_{1,2}$  to emulate the BACKDOORS interface of  $\pi_1$ , as required. Technically, the proof shows first that one can break down  $\mathcal{S}_{1,3}$  and put  $\pi_2$  in the place of  $\text{CompS}(\pi_3, \mathcal{S}_{2,3})$ . To show this, we aggregate  $\mathcal{S}_{1,2}$  into the environment to construct a new environment that would break  $\pi_2$  if such a modification was noticeable, contradicting the second hypothesis. The proof then follows by applying the first hypothesis. Note that this proof strategy is visible in the resources used by  $\mathcal{S}_{1,3}$ , since they are those required to run the composed module  $\text{SeqS}(\mathcal{S}_{2,3}, \mathcal{S}_{1,2})$ . Moreover, the quantification over the resources of the environments in the second hypothesis must accommodate an environment that *absorbs* simulator  $\mathcal{S}_{1,2}$  and runs it internally.

## 6.3 The dummy adversary in UC

The standard notion of UC emulation [17, 18] enriches the real-world with an explicit adversary  $\mathcal{A}$  representing an attacker that has access to the real-world BACKDOORS interface and colludes with the environment to break the protocol. In this case, the real- and ideal- world execution models

become structurally identical, in that the environment interacts with the BACKDOORS interface via adversarial entities in both worlds.<sup>10</sup>

Consider the functor in Fig. 18, which extends any real-world protocol with abstract adversary  $\mathcal{A}$  (A in EasyCrypt notation) at its BACKDOORS interface. The type of  $\mathcal{A}$  is parametric in the BACKDOORS offered by the protocol in our basic execution model, and it fixes the type of the BACKDOORS interface in the extended execution model NONDUMMY.PROTOCOL. This means that when we quantify over such adversaries, we quantify also over the potential forms of environment-to-adversary information exchange. The following theorem shows that we do not lose generality by working with an (implicit) dummy adversary in our execution model.

---

```

module type ADV(B : BACKDOORS) = {
  include NONDUMMY.BACKDOORS
}

module A_PROTOCOL(A : ADV, P : PROTOCOL)
  : NONDUMMY.PROTOCOL = {
  proc init() : unit = { P.init(); }
  include P [inputs, outputs]
  include A(P) [step,backdoor]
}
    
```

---

Fig. 18. Real-world protocol with adversary.

**THEOREM 6.2 (DUMMY ADVERSARY).** *UC emulation is equivalent to UC emulation with an explicit real-world adversary. More precisely:*

- *Emulation with an implicit dummy adversary implies emulation with an explicit arbitrary adversary: For all  $\epsilon \in \mathbb{R}^+$ , all protocols  $\pi_1$  and  $\pi_2$  with IO interfaces of the same type, all complexity restrictions  $c_{\text{sim}}, c_{\text{env}}$  and all simulators  $\mathcal{S} \in \tau_{\text{sim}}^{\pi_1, \pi_2, c_{\text{sim}}}$ , we have*

$$\text{Adv}_{c_{\text{sim}}, c_{\text{env}}}^{\text{uc}, \mathcal{S}}(\pi_1, \pi_2) \leq \epsilon \implies \forall \mathcal{A} \in \tau_{\text{adv}}, \text{Adv}_{\hat{c}_{\text{sim}}, c_{\text{env}}}^{\text{uc}}(\langle \pi_1 \parallel \mathcal{A}(\pi_1) \rangle, \pi_2) \leq \epsilon$$

where  $\hat{c}_{\text{sim}}$  allows for a simulator  $\mathcal{S}'$  that combines adversary  $\mathcal{A}$  and simulator  $\mathcal{S}$ .

- *Emulation with an implicit dummy adversary is implied by emulation with an explicit arbitrary adversary: For all  $\epsilon \in \mathbb{R}^+$ , all protocols  $\pi_1$  and  $\pi_2$  with IO interfaces of the same type, all complexity restrictions  $c_{\text{sim}}, c_{\text{env}}$  and all simulator memory spaces  $\mathcal{M}$ , we have*

$$\forall \mathcal{A} \in \tau_{\text{adv}}, \text{Adv}_{c_{\text{sim}}, c_{\text{env}}}^{\text{uc}, \mathcal{M}}(\langle \pi_1 \parallel \mathcal{A}(\pi_1) \rangle, \pi_2) \leq \epsilon \implies \text{Adv}_{c_{\text{sim}}, c_{\text{env}}}^{\text{uc}, \mathcal{M}}(\pi_1, \pi_2) \leq \epsilon$$

where  $\tau_{\text{adv}}$  accommodates the dummy adversary.

Our proof gives a simulator  $\mathcal{S}'$  for the first part of the theorem that joins together simulator  $\mathcal{S}$  and adversary  $\mathcal{A}$ : intuitively the new simulator uses the existing one to fool the (non-dummy) real-world adversary into thinking it is interacting with the real-world protocol and, in this way, it can offer the expected BACKDOORS view generated by  $\mathcal{A}$  to the environment. The resources used by  $\mathcal{S}'$  are those required to run the composition of  $\mathcal{S}$  and  $\mathcal{A}$ . The proof of the second part of the theorem is more interesting: we construct an explicit dummy adversary and use this to instantiate the hypothesis and obtain a simulator for this adversary, which we then show must also work when the dummy adversary is only implicit: this second step is an equivalence proof showing that, if the simulator matches the explicit dummy adversary which just passes information along, then it is also good when the environment is calling the protocols' BACKDOORS interface directly. The resulting simulator is therefore guaranteed to belong to the same cost-annotated type over which we quantify existentially in the hypothesis.

We note a technicality in the second part of the theorem: since the hypothesis quantifies over adversaries before quantifying existentially over simulators, we cannot use the approach adopted

<sup>10</sup>For this reason the simulator is often called an *ideal world adversary*; we do not adopt this terminology here to avoid confusion. The order of the quantifiers in the emulation definition is crucial for its compositional properties: it requires that, for all adversaries  $\mathcal{A}$ , there exists a simulator  $\mathcal{S}$  such that, for all environments  $\mathcal{Z}$ , the real- and ideal- worlds are indistinguishable. We now show that the same result holds in our setting.

in the transitivity proof and in the first part of the theorem, where we use global universal quantifications over hypothesized simulators. Instead, we quantify globally over a memory space  $\mathcal{M}$ , restrict simulators in the hypothesis to only use  $\mathcal{M}$ , and prevent other algorithms to interfere with this memory space where appropriate (we abuse notation by indicating  $\mathcal{M}$  in  $\text{Adv}^{\text{uc}}$  to denote this).

## 6.4 Universal Composability

The fundamental theorem of Universal Composability is stated in our EasyCrypt formalization as follows.

**THEOREM 6.3 (UNIVERSAL COMPOSABILITY).** *For all  $\epsilon_\rho, \epsilon_\pi \in \mathbb{R}^+$ , all ideal functionalities  $f, \mathcal{F}$ , all protocols  $\rho(f)$  and  $\pi$ , such that the IO interfaces of  $\pi$  and  $f$  (resp.  $\rho$  and  $\mathcal{F}$ ) are of the same type, all cost restrictions  $c_{\text{sim}(\rho)}$ ,  $c_{\text{sim}(\pi)}$ , and all simulators  $S_\rho \in \tau_{\text{sim}}^{\rho(f), \mathcal{F}, c_{\text{sim}(\rho)}}$  and  $S_\pi \in \tau_{\text{sim}}^{\pi, f, c_{\text{sim}(\pi)}}$ , we have:*

$$\text{Adv}_{c_{\text{sim}(\pi)}, \hat{c}_{\text{env}(\pi)}}^{\text{uc}, S_\pi}(\pi, f) \leq \epsilon_\pi \Rightarrow \text{Adv}_{c_{\text{sim}(\rho)}, \hat{c}_{\text{env}(\rho)}}^{\text{uc}, S_\rho}(\rho(f), \mathcal{F}) \leq \epsilon_\rho \Rightarrow \text{Adv}_{\hat{c}_{\text{sim}}, c_{\text{env}}}^{\text{uc}}(\rho(\pi), \mathcal{F}) \leq \epsilon_\rho + \epsilon_\pi$$

where  $\hat{c}_{\text{env}(\pi)}$  accommodates an environment that internally uses  $c_{\text{env}}$  resources and additionally runs  $\rho$ ,  $\hat{c}_{\text{sim}}$  corresponds to the cost of composing  $S_\pi$  and  $S_\rho$ ,  $\hat{c}_{\text{env}(\rho)}$  allows for an adversarial environment built by composing  $S_\pi$  with an environment in  $c_{\text{env}}$ .

This theorem establishes that any protocol  $\rho(f)$  that UC-emulates a functionality  $\mathcal{F}$  when relying on an ideal sub-component  $f$  offers the same level of security when it is instantiated with a protocol  $\pi$  that UC-emulates  $f$ . The proof first shows that the simulator  $S_\pi$  that exists by hypothesis can be converted into a simulator that justifies that  $\rho(\pi)$  UC-emulates  $\rho(f)$ : intuitively this new simulator uses  $S_\pi$  when interacting with the backdoors of  $f$  and just passes along the environment's interactions with the backdoors of  $\rho$ . This part of the proof combines any successful environment  $\mathcal{Z}$  against the composed protocol into a successful environment that absorbs  $\rho$  and breaks  $\pi$ . This justifies the cost restriction on  $c_{\text{env}}$ . Then, we know by hypothesis that  $\rho(f)$  UC emulates  $\mathcal{F}$ , and the result follows by applying the transitivity lemma, which also explains the remaining cost restrictions.

*Example: Composing key exchange with encryption.* We conclude this section with an example of the use of our framework and general lemmas stated above for concrete protocols. Consider the code snippets in Fig. 19. On the left we show the inner structure of a two-party protocol formalization (Diffie-Hellman) when one assumes an ideal sub-component (in this case a bi-directional ideal authenticated channel F2Auth exposing IO interface  $\text{Pi.REAL.IO}$ ). The full real-world configuration is obtained by applying a functor  $\text{CompRF}$  that composes this protocol with F2Auth and exposes the backdoors of both DHKE and F2Auth in a combined BACKDOORS interface. The IO interface to this real-world protocol is simply the input/output interface for both parties; parties take as input a role (initiator/responder) and the identities of parties involved in the protocol (type unit pkg); they output a session key when the protocol completes.

The Initiator code is shown in Fig. 20. On initialization it samples its ephemeral key pair and resets the derived key. When the environment provides input, which includes the identities of the parties that will take part in the key exchange, the ephemeral public key is transmitted via one of the ideal authenticated channels. The party then returns control to the environment (note that delivering a message to the authenticated channel does not pass control to the authenticated channel). When the environment calls `step`, the initiator checks the incoming ideal channel to see if it received a message. At any point the environment can check the initiator output using `output`. The backdoor interface provides no information, since all communications go through the authenticated channels. The responder code is symmetric.

<pre> <b>module</b> (DHKE : RHO) (F2Auth: Pi.REAL.IO) = {   <b>module</b> Initiator = { ... }   <b>module</b> Responder = { ... }    <b>proc</b> init() : unit = {     Initiator.init(); Responder.init(); }    <b>proc</b> inputs(r : role, p : unit pkg)   : unit = {     <b>if</b> (r = I) { Initiator.inputs(p); }     <b>else</b> { Responder.inputs(p); } }    <b>proc</b> outputs(r : role)   : group option = { ... }    <b>proc</b> step(r : role) : unit = { ... }    <b>proc</b> backdoor(r : role) : unit option = {     <b>var</b> rr;     <b>if</b> (r = I) { rr ← Initiator.backdoor(); }     <b>else</b> { rr ← Responder.backdoor(); }     <b>return</b> rr; } } </pre>	<pre> <b>module</b> FAuth : PROTOCOL = {   <b>var</b> st : state    <b>proc</b> init() : unit = {     st ← init_st; }    <b>proc</b> inputs(r : role, p : msg pkg)   : unit = {     st ← set_msg st r p; }    <b>proc</b> outputs(r : role, p : unit pkg)   : msg option = {     <b>return</b> get_msg st r p; }    <b>proc</b> step() : unit = {     st ← unblock st; }    <b>proc</b> backdoor()   : leakage option = {     <b>return</b> leak st; } } </pre>	<pre> <b>module</b> FKE : PROTOCOL = {   <b>var</b> st : state    <b>proc</b> init() : unit = {     k ←<sup>\$</sup> gen; st ← init k; }    <b>proc</b> inputs(r : role, p : unit pkg)   : unit = {     st ← party_start st r p; }    <b>proc</b> outputs(r : role)   : key option = {     <b>return</b> party_output st r; }    <b>proc</b> step() : unit = {     st ← unblock st; }    <b>proc</b> backdoor()   : leakage option = {     <b>return</b> leak st; } } </pre>
--	---	--

Fig. 19. Examples of real-world (left) and ideal-world protocols (middle and right). Left: structure of a Diffie-Hellman protocol relying on FAuth for authenticated communication (one shot each way). Middle: ideal functionality for one-shot authenticated channel FAuth. Right: ideal functionality for key exchange.

```

module Initiator = {
  proc init() : unit = { st ← IInit; _x ←$ FDistr.dt; _X ← g^_x; _K ← None; }

  proc inputs(_p : unit pkg) : unit = {
    if (st = IInit) { p ← _p; Auth.inputs(Left (I, (snd p, rcv p, _X))); st ← ISent; }}

  proc outputs() : group option = { return _K; }

  proc step() : unit = {
    if (st = ISent) {
      _Y ← Auth.outputs(Right (R, (rcv p, snd p, ()))));
      if (_Y ≠ None) { _K ← Some (oget (getr (oget _Y)) ^ _x); st ← IDone; }
    }}

  proc backdoor() : unit option = { return None; }}

```

Fig. 20. Diffie-Hellman Initiator.

In the middle code-snippet of Fig. 19 we give an example ideal functionality for a simple one-shot unidirectional authenticated channel; one party provides input with the party identities and the message to transmit (type msg pkg), and the other party can obtain the message if it calls outputs with matching identities (type unit pkg.) The attacker can use the backdoor procedure to observe the state of the channel, including the transmitted message and the party identities and it can use the step procedure to control when the message is delivered (the unlock operator changes the state so that, if a message is buffered, then it is made available at the output procedure) to the receiving party (get\_message is checking for identity consistency, which models authentication).

The example starts with a proof that the Diffie-Hellman protocol on the left of Fig. 19 UC-emulates the ideal functionality for key exchange shown on the right of Fig. 19 in a hybrid-real world where the parties have access to authenticated channels. The FKE functionality runs internally a state machine that waits for both parties to provide input, and allows an adversary/simulator interacting with its BACKDOORS interface to control when the different parties obtain a fresh shared secret key. This result is stated as follows; note the accounting of resources spent by the combined Diffie-Hellman attacker, making it explicit that the DDH assumption must be valid for such an attacker.

LEMMA 6.4 (SECURITY OF DHKE). *Fix  $c_{\text{ddh}} \in \mathbb{R}^+$  and let  $\epsilon_{\text{DDH}}$  be the maximum advantage of any DDH attacker against the group over which we implement DHKE. Then, we have that*

$$\text{Adv}_{c_{\text{sim}}(\text{DHKE}), c_{\text{env}}(\text{DHKE})}^{\text{uc}}(\text{DHKE}(\text{F2Auth}), \text{FKE}) \leq \epsilon_{\text{DDH}}$$

where  $c_{\text{sim}}(\text{DHKE})$  is the cost of a concrete simulator  $\mathcal{S}_{\text{DHKE}}$  that just samples random group elements as the protocol messages and mimics the states of the real-world parties and  $\text{F2Auth}$ ;  $c_{\text{env}}(\text{DHKE})$  must be such that  $c_{\text{ddh}}$  accommodates the cost of an adversary that runs internally the entire UC emulation experiment (including the environment) and interpolates between the real and ideal worlds, depending on the external DDH challenge.

The second result shows that the ideal functionality for key exchange can be combined with one-time-pad encryption to transform a one-shot authenticated channel into a one-shot secure channel that also guarantees confidentiality. Formally:

LEMMA 6.5 (SECURITY OF OTP). *Fix any  $c_{\text{env}}(\text{OTP})$ . Then we have*

$$\text{Adv}_{c_{\text{sim}}(\text{OTP}), c_{\text{env}}(\text{OTP})}^{\text{uc}}(\text{OTP}(\text{FKE}, \text{FAuth}), \text{FSC}) = 0$$

where  $c_{\text{sim}}(\text{OTP})$  is the cost of a concrete simulator  $\mathcal{S}_{\text{OTP}}$  that just samples a random string in place of the ciphertext and mimics the states of the real-world parties, FKE and  $\text{FAuth}$ .

Here,  $\text{FSC}$  represents the secure channel ideal functionality, which operates exactly as  $\text{FAuth}$ , but does not leak the transmitted message; leakage includes only information on the state of the channel. The protocol runs in a hybrid world where it has access to both FKE and  $\text{FAuth}$ , uses the former to obtain a shared key between the two parties, and then transmits the one-time-padded message using  $\text{FAuth}$ . We apply our Universal Composability theorem to derive that FKE can be replaced by the DHKE protocol, resulting in a protocol that still UC-emulates the secure channel functionality. The final theorem is stated as follows.

THEOREM 6.6 (SECURITY OF OTP COMPOSED WITH DHKE). *Fix  $c_{\text{ddh}} \in \mathbb{R}^+$  and let  $\epsilon_{\text{DDH}}$  be the maximum advantage of any DDH attacker against the group over which we implement DHKE. Then*

$$\text{Adv}_{c_{\text{sim}}, c_{\text{env}}}^{\text{uc}}(\text{OTP}(\text{DHKE}, \text{FAuth}), \text{FSC}) \leq \epsilon_{\text{DDH}}$$

where  $c_{\text{env}}$  is constrained so that  $c_{\text{env}}(\text{DHKE})$  accommodates an environment that internally uses  $c_{\text{env}}$  resources and additionally runs OTP, and  $c_{\text{sim}}$  corresponds to the cost of composing  $\mathcal{S}_{\text{OTP}}$  and  $\mathcal{S}_{\text{DHKE}}$ .

The crucial application of the complexity restrictions is visible in the attacker against the DDH assumption, which now has a more complex structure that results from the application of the composition theorem: for this application of composition to be meaningful, it is crucial that the global environment is computationally bounded (even though the OTP protocol is information-theoretically secure) as a function of  $c_{\text{ddh}}$ , as otherwise the reduction to DDH would be meaningless. Indeed, the class of DDH attackers must allow for the extra resources required to run a simulation of OTP protocol in the reduction. Note also that the execution time of the global simulator is given

by  $\mathcal{S}_{\text{OTP}}$  and  $\mathcal{S}_{\text{DHKE}}$ , which are very efficient; hence the UC emulation result has a small simulation overhead [18, 19].

For the proof we used an auxiliary lemma, which is a specialization of the Universal Composability theorem for the case where the hybrid functionality is the parallel composition of two ideal functionalities and we apply the Universal Composability theorem to instantiate only one of them.

*Our formalization vs EasyUC.* Our Diffie-Hellman example is an alternative formalization of the example given by Canetti, Stoughton and Varia [21] for the EasyUC framework. We borrow it because, as in [21], it is a good toy example with which to validate and demonstrate our formalization. This example is also convenient to show that the approach in this paper and EasyUC in effect complement each other. An important design goal of EasyUC is to follow the UC execution model as closely as possible; this allows a more direct translation of protocols and ideal functionalities.

In contrast, our goal is to take advantage of the EasyCrypt machinery to reduce proof effort and development size: our development (including complexity) takes 2300 lines of code and it includes general UC theorems that can be reused in future work; this compares to 18K lines of code for EasyUC.<sup>11</sup> The downside of our approach is the impact in the way one specifies protocols and ideal functionalities: message passing corresponds to procedure calls, and these must adhere to the EasyCrypt tree-based procedure call semantics. For example, we do not allow an execution environment where a party communicates with an ideal functionality arbitrarily without relying on the environment for scheduling; one could of course formalize a message passing mechanism on top of EasyCrypt as in [21] to allow for this, but this would then fall out of the scope of our general composition theorems. Moreover, it would lead to larger developments and increased proof effort, which would defeat our original purpose.

In short, one can think of the EasyUC approach as a front-end for cryptographers, and our approach as a convenient back-end for conducting the machine-checked proofs. We leave it as an interesting direction for future work to develop a sound translation between these two approaches to modeling UC for a representative class of protocols such as those considered in [20]. Another interesting direction for future work is to identify UC security proofs that cannot be naturally expressed using our approach to formalizing UC and to investigate how it can be extended to deal with these examples.

## 7 RELATED WORK

*Cost analysis.* There is a very large body of work that uses program logics for cost analysis of imperative programs. [36] uses Hoare logic for proving upper bounds on execution time of deterministic programs. In the probabilistic setting, [30] uses a pre-expectation calculus inspired from Kozen [31] and Morgan, McIver and Seidel [34] to compute upper bounds on the expected cost of probabilistic programs. Building on prior work in automation of amortized analysis, [35] develop a compositional and automated logic for proving upper bounds on expected time. Recently, [28] show how the pre-expectations calculus can also be used for proving lower bounds on expected time. While most of the cryptography literature considers worst-case execution time, expected-time arises naturally in some applications; see e.g. [24] for a discussion of expected polynomial-time in cryptography. It would be interesting to develop an adaptation of our system for expected-time complexity, based on the approach developed in [30]. In addition, there is a long line of work on automating cost analysis, both for deterministic and for probabilistic programs, see e.g. [1, 15, 25]. These techniques could be helpful to alleviate users efforts, and connecting with tools that support them is an important direction for future work.

<sup>11</sup>The count excludes general purpose libraries, but we should note that the exact numbers are not important, as the size of a development varies significantly with style of coding and the use of automation.



There is also a large body of work on cost analysis of higher-order languages, both in the deterministic and probabilistic settings. Some of these works specifically focus on cryptography; e.g., [4] develop a type system that accomodates non-heriditarily polynomial-time computations (i.e. computations that contain non-polynomial-time subroutines), and use their type system to analyze the complexity of the reduction from hardcore predicates to one-way functions. Many of these works are based on (linear) type systems [22, 29], but other works use a fine-grained program logic to support value-dependent cost analysis [38]. These approaches could be used in conjunction with our type-and-effect module system to support compositional reductionist proofs for a higher-order language.

*Computer-aided cryptography.* CryptoVerif [14] is an automated tool for computational security proofs. CryptoVerif uses approximate equivalences to find (or check) cryptographic reductions, and keeps track of the complexity of adversaries. Most other tools for computational security proofs, including CertiCrypt [10], Foundational Cryptography Framework [37], and CryptHOL [12], share their foundations and overall approach with EasyCrypt. However, these tools offer limited support for complexity reasoning and they do not support the use of modules for defining cryptographic schemes and notions. This is not a fundamental limitation, since these tools are embedded in a general-purpose proof assistant. However, extending these tools to achieve similar effects as our type-and-effect module system and program logic for complexity would represent a significant endeavor.

Our module system is inspired from EasyCrypt [8, 11]. However, the EasyCrypt module system lacks complexity restrictions, which hampers the use of compositional approaches. Beyond EasyCrypt, several other tools and approaches use structures similar to modules for formalizing cryptographic schemes and their security. Computational Indistinguishability Logic (CIL) [7] rely on oracle systems, which are very closely related to our modules. Interestingly, the main judgment of CIL establishes the approximate equivalence of two oracle systems, and is explicitly quantified by the resources of an adversary. State-separating proofs [16] pursue similar goals, using a notion of package. Packages have the expressivity of modules, but additionally support private functions. Our modules can emulate private functions using restrictions. At present, there is no tool support for state-separating proofs. [39] introduces the notion of interface, which is similar to module, for formalizing cryptography.

## 8 CONCLUSION

We have developed an extension of the EasyCrypt proof assistant to support reasoning complexity claims. The extension captures reductionist statements that faithfully match the cryptographic literature and supports compositional reasoning. As a main example, we have shown how to formalize key results from Universal Composability, a long-standing goal of computer-aided cryptography.

## REFERENCES

- [1] Elvira Albert, Puri Arenas, Samir Genaim, Miguel Gómez-Zamalloa, German Puebla, D. Ramírez, G. Román, and Damiano Zanardini. 2009. Termination and Cost Analysis with COSTA and its User Interfaces. *Electr. Notes Theor. Comput. Sci.* 258, 1 (2009), 109–121.
- [2] José Bacelar Almeida, Manuel Barbosa, Gilles Barthe, Matthew Campagna, Ernie Cohen, Benjamin Grégoire, Vitor Pereira, Bernardo Portela, Pierre-Yves Strub, and Serdar Tasiran. 2019. A Machine-Checked Proof of Security for AWS Key Management Service. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019, London, UK, November 11-15, 2019*, Lorenzo Cavallaro, Johannes Kinder, XiaoFeng Wang, and Jonathan Katz (Eds.). ACM, 63–78. <https://doi.org/10.1145/3319535.3354228>
- [3] José Bacelar Almeida, Cecile Baritel-Ruet, Manuel Barbosa, Gilles Barthe, François Dupressoir, Benjamin Grégoire, Vincent Laporte, Tiago Oliveira, Alley Stoughton, and Pierre-Yves Strub. 2019. Machine-Checked Proofs for Cryptographic Standards: Indifferentiability of Sponge and Secure High-Assurance Implementations of SHA-3. In *Proceedings*

- of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019, London, UK, November 11-15, 2019, Lorenzo Cavallaro, Johannes Kinder, XiaoFeng Wang, and Jonathan Katz (Eds.). ACM, 1607–1622. <https://doi.org/10.1145/3319535.3363211>
- [4] Patrick Baillot, Gilles Barthe, and Ugo Dal Lago. 2015. Implicit Computational Complexity of Subrecursive Definitions and Applications to Cryptographic Proofs. In *Logic for Programming, Artificial Intelligence, and Reasoning - 20th International Conference, LPAR-20 2015, Suva, Fiji, November 24-28, 2015, Proceedings (Lecture Notes in Computer Science, Vol. 9450)*, Martin Davis, Ansgar Fehnker, Annabelle McIver, and Andrei Voronkov (Eds.). Springer, 203–218. [https://doi.org/10.1007/978-3-662-48899-7\\_15](https://doi.org/10.1007/978-3-662-48899-7_15)
- [5] Manuel Barbosa, Gilles Barthe, Karthikeyan Bhargavan, Bruno Blanchet, Cas Cremers, Kevin Liao, and Bryan Parno. 2021. SoK: Computer-Aided Cryptography. In *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, Los Alamitos, CA, USA, 777–795. <https://doi.org/10.1109/SP40001.2021.00008>
- [6] Manuel Barbosa, Gilles Barthe, Benjamin Grégoire, Adrien Koutsos, and Pierre-Yves Strub. 2021. Mechanized Proofs of Adversarial Complexity and Application to Universal Composability. In *CCS*. ACM, 2541–2563.
- [7] Gilles Barthe, Marion Daubignard, Bruce M. Kapron, and Yassine Lakhnech. 2010. Computational indistinguishability logic. In *Proceedings of the 17th ACM Conference on Computer and Communications Security, CCS 2010, Chicago, Illinois, USA, October 4-8, 2010*, Ehab Al-Shaer, Angelos D. Keromytis, and Vitaly Shmatikov (Eds.). ACM, 375–386. <https://doi.org/10.1145/1866307.1866350>
- [8] Gilles Barthe, François Dupressoir, Benjamin Grégoire, César Kunz, Benedikt Schmidt, and Pierre-Yves Strub. 2013. EasyCrypt: A Tutorial. In *Foundations of Security Analysis and Design VII - FOSAD 2012/2013 Tutorial Lectures (Lecture Notes in Computer Science, Vol. 8604)*, Alessandro Aldini, Javier López, and Fabio Martinelli (Eds.). Springer, 146–166. [https://doi.org/10.1007/978-3-319-10082-1\\_6](https://doi.org/10.1007/978-3-319-10082-1_6)
- [9] Gilles Barthe, Marco Gaboardi, Benjamin Grégoire, Justin Hsu, and Pierre-Yves Strub. 2016. A Program Logic for Union Bounds. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy (LIPIcs, Vol. 55)*, Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 107:1–107:15. <https://doi.org/10.4230/LIPIcs.ICALP.2016.107>
- [10] Gilles Barthe, Benjamin Grégoire, and Santiago Zanella Béguelin. 2009. Formal certification of code-based cryptographic proofs. In *Proceedings of the 36th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2009, Savannah, GA, USA, January 21-23, 2009*, Zhong Shao and Benjamin C. Pierce (Eds.). ACM, 90–101. <https://doi.org/10.1145/1480881.1480894>
- [11] Gilles Barthe, Benjamin Grégoire, Sylvain Heraud, and Santiago Zanella Béguelin. 2011. Computer-Aided Security Proofs for the Working Cryptographer. In *Advances in Cryptology - CRYPTO 2011 - 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011. Proceedings (Lecture Notes in Computer Science, Vol. 6841)*, Phillip Rogaway (Ed.). Springer, 71–90. [https://doi.org/10.1007/978-3-642-22792-9\\_5](https://doi.org/10.1007/978-3-642-22792-9_5)
- [12] David A. Basin, Andreas Lochbihler, and S. Reza Sefidgar. 2020. CryptHOL: Game-Based Proofs in Higher-Order Logic. *J. Cryptology* 33, 2 (2020), 494–566. <https://doi.org/10.1007/s00145-019-09341-z>
- [13] Mihir Bellare and Phillip Rogaway. 1993. Random Oracles are Practical: A Paradigm for Designing Efficient Protocols. In *CCS '93, Proceedings of the 1st ACM Conference on Computer and Communications Security, Fairfax, Virginia, USA, November 3-5, 1993*, Dorothy E. Denning, Raymond Pyle, Ravi Ganesan, Ravi S. Sandhu, and Victoria Ashby (Eds.). ACM, 62–73. <https://doi.org/10.1145/168588.168596>
- [14] Bruno Blanchet. 2006. A Computationally Sound Mechanized Prover for Security Protocols. In *2006 IEEE Symposium on Security and Privacy (S&P 2006), 21-24 May 2006, Berkeley, California, USA*. IEEE Computer Society, 140–154. <https://doi.org/10.1109/SP.2006.1>
- [15] Marc Brockschmidt, Fabian Emmes, Stephan Falke, Carsten Fuhs, and Jürgen Giesl. 2014. Alternating Runtime and Size Complexity Analysis of Integer Programs. In *Tools and Alg. for the Constr. and Anal. of Systems - 20th Int. Conf. (TACAS'14)*. 140–155.
- [16] Chris Brzuska, Antoine Delignat-Lavaud, Cédric Fournet, Konrad Kohbrok, and Markulf Kohlweiss. 2018. State Separation for Code-Based Game-Playing Proofs. In *Advances in Cryptology - ASIACRYPT 2018 - 24th International Conference on the Theory and Application of Cryptology and Information Security, Brisbane, QLD, Australia, December 2-6, 2018, Proceedings, Part III (Lecture Notes in Computer Science, Vol. 11274)*, Thomas Peyrin and Steven D. Galbraith (Eds.). Springer, 222–249. [https://doi.org/10.1007/978-3-030-03332-3\\_9](https://doi.org/10.1007/978-3-030-03332-3_9)
- [17] Ran Canetti. 2000. Universally Composable Security: A New Paradigm for Cryptographic Protocols. Cryptology ePrint Archive, Report 2000/067. <https://eprint.iacr.org/2000/067>.
- [18] Ran Canetti. 2001. Universally Composable Security: A New Paradigm for Cryptographic Protocols. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*. IEEE Computer Society, 136–145. <https://doi.org/10.1109/SFCS.2001.959888>
- [19] Ran Canetti. 2001. Universally composable security: a new paradigm for cryptographic protocols. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. 136–145.

- [20] Ran Canetti, Asaf Cohen, and Yehuda Lindell. 2015. A Simpler Variant of Universally Composable Security for Standard Multiparty Computation. In *Advances in Cryptology – CRYPTO 2015*, Rosario Gennaro and Matthew Robshaw (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 3–22.
- [21] Ran Canetti, Alley Stoughton, and Mayank Varia. 2019. EasyUC: Using EasyCrypt to Mechanize Proofs of Universally Composable Security. In *32nd IEEE Computer Security Foundations Symposium, CSF 2019, Hoboken, NJ, USA, June 25-28, 2019*. IEEE, 167–183. <https://doi.org/10.1109/CSF.2019.00019>
- [22] Ugo Dal Lago and Marco Gaboardi. 2011. Linear Dependent Types and Relative Completeness. In *Proceedings of the 2011 IEEE 26th Annual Symposium on Logic in Computer Science (LICS '11)*. 133–142.
- [23] The EasyCrypt development team. 2021. Source code of our EasyCrypt. <https://github.com/EasyCrypt/easycrypt>.
- [24] Oded Goldreich. 2010. On Expected Probabilistic Polynomial-Time Adversaries: A Suggestion for Restricted Definitions and Their Benefits. *J. Cryptology* 23, 1 (2010), 1–36. <https://doi.org/10.1007/s00145-009-9050-5>
- [25] Sumit Gulwani, Krishna K. Mehra, and Trishul Chilimbi. 2009. SPEED: Precise and Efficient Static Estimation of Program Computational Complexity. In *Proceedings of the 36th Annual Symposium on Principles of Programming Languages (Savannah, GA, USA) (POPL '09)*. 127–139.
- [26] Helene Haagh, Aleksandr Karbyshev, Sabine Oechsner, Bas Spitters, and Pierre-Yves Strub. 2018. Computer-Aided Proofs for Multiparty Computation with Active Security. In *31st IEEE Computer Security Foundations Symposium, CSF 2018, Oxford, United Kingdom, July 9-12, 2018*. IEEE Computer Society, 119–131. <https://doi.org/10.1109/CSF.2018.00016>
- [27] Shai Halevi. 2005. A plausible approach to computer-aided cryptographic proofs. *IACR Cryptol. ePrint Arch.* 2005 (2005), 181. <http://eprint.iacr.org/2005/181>
- [28] Marcel Hark, Benjamin Lucien Kaminski, Jürgen Giesl, and Joost-Pieter Katoen. 2020. Aiming low is harder: induction for lower bounds in probabilistic program verification. *Proc. ACM Program. Lang.* 4, POPL (2020), 37:1–37:28. <https://doi.org/10.1145/3371105>
- [29] Jan Hoffmann. 2011. *Types with Potential: Polynomial Resource Bounds via Automatic Amortized Analysis*. Ph.D. Dissertation. Ludwig-Maximilians-Universität München.
- [30] Benjamin Lucien Kaminski, Joost-Pieter Katoen, Christoph Matheja, and Federico Olmedo. 2016. Weakest Precondition Reasoning for Expected Run-Times of Probabilistic Programs. In *Programming Languages and Systems - 25th European Symposium on Programming, ESOP 2016, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2016, Eindhoven, The Netherlands, April 2-8, 2016, Proceedings (Lecture Notes in Computer Science, Vol. 9632)*, Peter Thiemann (Ed.). Springer, 364–389. [https://doi.org/10.1007/978-3-662-49498-1\\_15](https://doi.org/10.1007/978-3-662-49498-1_15)
- [31] Dexter Kozen. 1985. A Probabilistic PDL. *J. Comput. Syst. Sci.* 30, 2 (1985), 162–178. [https://doi.org/10.1016/0022-0000\(85\)90012-1](https://doi.org/10.1016/0022-0000(85)90012-1)
- [32] Xavier Leroy. 1995. Applicative Functors and Fully Transparent Higher-Order Modules. In *POPL*. ACM Press, 142–153.
- [33] Kevin Liao, Matthew A. Hammer, and Andrew Miller. 2019. ILC: a calculus for composable, computational cryptography. In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2019, Phoenix, AZ, USA, June 22-26, 2019*, Kathryn S. McKinley and Kathleen Fisher (Eds.). ACM, 640–654. <https://doi.org/10.1145/3314221.3314607>
- [34] Carroll Morgan, Annabelle McIver, and Karen Seidel. 1996. Probabilistic Predicate Transformers. *ACM Trans. Program. Lang. Syst.* 18, 3 (1996), 325–353. <https://doi.org/10.1145/229542.229547>
- [35] Van Chan Ngo, Quentin Carbonneaux, and Jan Hoffmann. 2018. Bounded expectations: resource analysis for probabilistic programs. In *Proceedings of the 39th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2018, Philadelphia, PA, USA, June 18-22, 2018*, Jeffrey S. Foster and Dan Grossman (Eds.). ACM, 496–512.
- [36] Hanne Riis Nielson. 1987. A Hoare-Like Proof System for Analysing the Computation Time of Programs. *Sci. Comput. Program.* 9, 2 (1987), 107–136. [https://doi.org/10.1016/0167-6423\(87\)90029-3](https://doi.org/10.1016/0167-6423(87)90029-3)
- [37] Adam Petcher and Greg Morrisett. 2015. A Mechanized Proof of Security for Searchable Symmetric Encryption. In *IEEE 28th Computer Security Foundations Symposium, CSF 2015, Verona, Italy, 13-17 July, 2015*, Cédric Fournet, Michael W. Hicks, and Luca Viganò (Eds.). IEEE Computer Society, 481–494.
- [38] Ivan Radicek, Gilles Barthe, Marco Gaboardi, Deepak Garg, and Florian Zuleger. 2018. Monadic refinements for relational cost analysis. *Proc. ACM Program. Lang.* 2, POPL (2018), 36:1–36:32. <https://doi.org/10.1145/3158124>
- [39] Mike Rosulek. 2020. *The Joy of Cryptography*. <https://joyofcryptography.com>
- [40] Asankhaya Sharma, Shengyi Wang, Andreea Costea, Aquinas Hobor, and Wei-Ngan Chin. 2015. Certified Reasoning with Infinity. In *FM 2015: Formal Methods - 20th International Symposium, Oslo, Norway, June 24-26, 2015, Proceedings (Lecture Notes in Computer Science, Vol. 9109)*, Nikolaj Bjørner and Frank S. de Boer (Eds.). Springer, 496–513. [https://doi.org/10.1007/978-3-319-19249-9\\_31](https://doi.org/10.1007/978-3-319-19249-9_31)

**Appendix Outline.** We quickly outline the structure of the appendix. We present the remaining typing rules for our programming language and module system in Appendix A. In Appendix B, we give the definition of the additional complexity measures used to define our cost judgements: the

number of calls a program execution makes to some abstract procedure, and the intrinsic cost of a program execution (i.e. the cost of the program without the cost of parameters calls). Appendix C proves that that our type system and module resolution mechanism have the *subject reduction* property: essentially, evaluating a module expression preserves its type. This is a crucial property, as it proves that module restrictions are preserved by the evaluation of a module using the module resolution mechanism. Finally, we present the full set of rules of our Hoare logic for cost and prove their soundness in Appendix D.

## A ADDITIONAL TYPING RULES

To help the reader, we give a summary of all the syntactic categories of our programming language and module system in Fig. 21.

### A.1 Additional Typing Rules

The memory restriction union  $\sqcup$ , intersection  $\sqcap$  and the memory restriction subset  $\sqsubseteq$  operations are defined in Fig. 22. In Fig. 23, we present our sub-typing rules, our typing rules for statements and expressions, and the definition of the function  $\text{mem}_\Gamma(p)$  which computes the memory footprint of  $p$  in  $\Gamma$ . Note that we need two different rules to type function paths: **T-PROC1** does a lookup of the procedure as a component of an already typed module; and **T-PROC2** does a lookup of the procedure in the typing environment, in case the procedure is declared in one of the parent modules of the current sub-module being typed (consequently, these modules are not yet fully typed).

*Example A.1.* Consider the modules and procedures given in Fig. 25. When typing  $h$ , the typing environment contains one module declaration and one procedure declaration:

$$\Gamma = (\text{module } A.B = \_ : \text{struct proc } f() \rightarrow \text{unit end}; \\ (\text{proc } A.C.g() \rightarrow \text{unit} = \dots))$$

Here, the call to  $f$  in  $h$  is typed using the **T-PROC1** rule, while the call to  $g$  is typed using **T-PROC2**.

## B ADDITIONAL INSTRUMENTED SEMANTICS

We give the definition of the additional complexity measures used to define our cost judgements, using two auxiliary instrumented semantics:  $\gamma.g \llbracket s \rrbracket_v^{\mathcal{E}, \rho}$  counts the number of times  $s$  calls an abstract procedure  $\gamma.g$ ; and  $^i \llbracket s \rrbracket_v^{\mathcal{E}, \rho}$  measures the intrinsic cost of an instruction (i.e. without counting the cost of function calls in a functor parameters).

*Function call counting.* The function call counting semantics  $\gamma.g \llbracket s \rrbracket_v^{\mathcal{E}, \rho}$ , given in in Fig. 26, evaluates the instruction  $s$  under the memory  $v$  and  $\mathcal{E}$ -pre-interpretation  $\rho$ , counting the number of calls to the abstract procedure  $\gamma.g$ .

The maximum number of calls of an instruction  $s$  or module procedure  $m.f$  to  $\gamma.g$  in  $(\mathcal{E}, \rho)$  is:

$$\begin{aligned} \# \text{calls}_{\gamma.g, v}^{\mathcal{E}, \rho}(s) &= \sup(\text{supp}(\pi_2^\#(\gamma.g \llbracket s \rrbracket_v^{\mathcal{E}, \rho}))) && \text{(in memory } v) \\ \# \text{calls}_{\gamma.g}^{\mathcal{E}, \rho}(s) &= \max_{v \in \mathcal{M}}(\# \text{calls}_{\gamma.g, v}^{\mathcal{E}, \rho}(s)) && \text{(in any memory)} \end{aligned}$$

---

```

module A = {
module B = { proc f () : unit = ... }
module C = {
proc g () : unit = ...
proc h () : unit = {
  A.B.f();
  A.C.g();
}}}.
    
```

---

Fig. 25. Example of procedures typing.

Expressions (*distribution expressions are similar*):

$e ::= v \in \mathcal{V}$  (variable)  
 $| f(e_1, \dots, e_n)$  (if  $f|_n \in \mathcal{F}_E$ )

Statements:

$s ::= \mathbf{abort}$  (abort)  
 $| \mathbf{skip}$  (skip)  
 $| s_1; s_2$  (sequence)  
 $| x \leftarrow e$  (assignment)  
 $| x \stackrel{\$}{\leftarrow} d$  (sampling)  
 $| x \leftarrow \mathbf{call} F(\vec{e})$  (procedure call)  
 $| \mathbf{if} e \mathbf{then} s_1 \mathbf{else} s_2$  (conditional)  
 $| \mathbf{while} e \mathbf{do} s$  (loop)

Procedure body:

body  $::= \{ \text{var } (\vec{v} : \vec{\tau}); s; \text{return } e \}$

Module paths:

$p ::= x$  (mod. ident.)  
 $| p.x$  (mod. comp.)  
 $| p(p)$  (func. app.)

Function paths:

$F ::= p.f$  (proc. lookup)

Module expressions:

$m ::= p$  (mod. path)  
 $| \mathbf{struct} \text{ st } \mathbf{end}$  (structure)  
 $| \mathbf{func}(x : M)$  (functor)

Module structures:

$\text{st} ::= d_1; \dots; d_n$  (for any  $n \in \mathbb{N}$ )

Module declarations:

$d ::= \mathbf{proc} f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}$  (proc.)  
 $| \mathbf{module} x = m$  (module)

Signature structures:

$S ::= D_1; \dots; D_n$  (for any  $n \in \mathbb{N}$ )

Module signature declarations:

$D ::= \mathbf{proc} f(\vec{v} : \vec{\tau}) \rightarrow \tau_r$  (procedure decl.)  
 $| \mathbf{module} x : M$  (mod. decl.)

Module signatures:

$M ::= \mathbf{sig} S \mathbf{restr} \theta \mathbf{end}$  (restr. sig. struct.)  
 $| \mathbf{func}(x : M) M'$  (functor)

Typing environment declarations:

$\delta ::= \mathbf{var} v : \tau$  (variable decl.)  
 $| \mathbf{module} p = m : M$  (module decl.)  
 $| \mathbf{module} x = \mathbf{abs}_K : M$  (abs. mod.)  
 $| \mathbf{proc} p.f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}$  (proc. decl.)

Abstract module kind:

$K ::= \mathbf{open}$  (open module)  
 $| \mathbf{param}$  (module parameter)

Typing Environment:

$\Gamma ::= \epsilon \mid \Gamma, \delta$

Module restrictions:

$\theta ::= \epsilon \mid \theta, (f : \lambda)$   
 $\lambda ::= \top \mid \lambda_m \wedge \lambda_c$

Memory restrictions:

$\lambda_m ::= +\text{all mem} \setminus \{v_1, \dots, v_l\}$  (for any  $l \in \mathbb{N}$ )  
 $| \{v_1, \dots, v_l\}$  (for any  $l \in \mathbb{N}$ )

Complexity restrictions:

$\lambda_c ::= \top \mid \mathbf{compl}[\text{intr} : k, x_1.f_1 : k_1, \dots, x_l.f_l : k_l]$   
 (for any  $l, k, k_1, \dots, k_l \in \mathbb{N}$ )

Low-order module signatures:

$M_l ::= \mathbf{sig} S_l \mathbf{restr} \theta \mathbf{end}$  (restr. sig. struct.)  
 $| \mathbf{func}(x : \mathbf{sig} S_l \mathbf{restr} \theta \mathbf{end}) M_l$  (low-order func.)

Low-order signature structures:

$S_l ::= D_{l1}; \dots; D_{ln}$  (for any  $n \in \mathbb{N}$ )

Low-order module signature declarations:

$D_l ::= \mathbf{proc} f(\vec{v} : \vec{\tau}) \rightarrow \tau_r$  (procedure decl.)

Extended module expressions:

$\tilde{m} ::= m \mid \mathbf{abs}_K x$

Fig. 21. Syntax of statements and modules.

$$\begin{aligned}
 & \text{Memory restriction union } \sqcup \\
 (+\text{all mem} \setminus \{v_1, \dots, v_n\}) \sqcup (+\text{all mem} \setminus \{v'_1, \dots, v'_m\}) &= +\text{all mem} \setminus (\{v_1, \dots, v_n\} \cap \{v'_1, \dots, v'_m\}) \\
 \{v_1, \dots, v_n\} \sqcup \{v'_1, \dots, v'_m\} &= \{v_1, \dots, v_n\} \cup \{v'_1, \dots, v'_m\} \\
 (+\text{all mem} \setminus \{v_1, \dots, v_n\}) \sqcup \{v'_1, \dots, v'_m\} &= +\text{all mem} \setminus (\{v_1, \dots, v_n\} \setminus \{v'_1, \dots, v'_m\}) \\
 & \text{Memory restriction intersection } \sqcap \\
 (+\text{all mem} \setminus \{v_1, \dots, v_n\}) \sqcap (+\text{all mem} \setminus \{v'_1, \dots, v'_m\}) &= +\text{all mem} \setminus (\{v_1, \dots, v_n\} \cup \{v'_1, \dots, v'_m\}) \\
 \{v_1, \dots, v_n\} \sqcap \{v'_1, \dots, v'_m\} &= \{v_1, \dots, v_n\} \cap \{v'_1, \dots, v'_m\} \\
 (+\text{all mem} \setminus \{v_1, \dots, v_n\}) \sqcap \{v'_1, \dots, v'_m\} &= \{v'_1, \dots, v'_m\} \setminus \{v_1, \dots, v_n\} \\
 & \text{Memory restriction subset } \sqsubseteq \\
 (+\text{all mem} \setminus \{v_1, \dots, v_n\}) \sqsubseteq (+\text{all mem} \setminus \{v'_1, \dots, v'_m\}) &= \{v'_1, \dots, v'_m\} \sqsubseteq \{v_1, \dots, v_n\} \\
 \{v_1, \dots, v_n\} \sqsubseteq \{v'_1, \dots, v'_m\} &= \{v_1, \dots, v_n\} \sqsubseteq \{v'_1, \dots, v'_m\} \\
 (+\text{all mem} \setminus \{v_1, \dots, v_n\}) \sqsubseteq \{v'_1, \dots, v'_m\} &= \perp \\
 \{v_1, \dots, v_n\} \sqsubseteq (+\text{all mem} \setminus \{v'_1, \dots, v'_m\}) &= \{v_1, \dots, v_n\} \cap \{v'_1, \dots, v'_m\} = \emptyset
 \end{aligned}$$

Fig. 22. Memory restriction operations and type erasure functions.

$$\#\text{calls}_{\mathcal{S}_{y.g}}^{\mathcal{E}, \rho}(m.f) = \begin{cases} \#\text{calls}_{\mathcal{S}_{y.g}}^{\mathcal{E}, \rho}(s) & \text{when } \text{f-res}_{\mathcal{E}}(m.f) = (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \_ ; s ; \_ \}) \\ \#\text{calls}_{\mathcal{S}_{y.g}}^{\mathcal{E}, \rho}(\rho(x)(\vec{p}).f) \nu & \text{when } \text{f-res}_{\mathcal{E}}(m.f) = (\text{abs}_x \text{ open})(\vec{p}).f \end{cases}$$

Note that when  $\text{f-res}_{\mathcal{E}}(m.f) = (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \_ ; s ; \_ \})$ , we ignore the return expression, since expression cannot contain procedure calls (only operator applications).

*Intrinsic cost.* The  $(\mathcal{E}, \rho)$ -denotational semantics of an instruction  $s$  with intrinsic cost under memory  $\nu$  and parameters  $\vec{x}$ , written  $^i \llbracket s \rrbracket_{\nu}^{\mathcal{E}, \rho, \vec{x}}$  is the cost of the execution of  $s$  under  $\nu$  in  $\rho$ , without counting the costs of function calls to the parameters  $\vec{x}$ . Formally,  $^i \llbracket \_ \rrbracket_{\nu}^{\mathcal{E}, \rho, \vec{x}}$  is defined exactly like  $\llbracket \_ \rrbracket_{\nu}^{\mathcal{E}, \rho, \vec{x}}$  in Fig. 14, except for the abstract procedure call case, which is replaced by:

$$^i \llbracket x \leftarrow \text{call } m.f(\vec{e}) \rrbracket_{\nu}^{\mathcal{E}, \rho, \vec{x}} = \begin{cases} \llbracket \mathbb{E}_{(v', c') \sim i} \llbracket x \leftarrow \text{call } \rho(x)(\vec{p}).f(\vec{e}) \rrbracket_{\nu}^{\mathcal{E}, \rho, \vec{x}} \llbracket (v', c_{\mathbb{E}}(\vec{e}, \nu)) \rrbracket & \text{if } z \in \vec{x} \\ ^i \llbracket x \leftarrow \text{call } \rho(x)(\vec{p}).f(\vec{e}) \rrbracket_{\nu}^{\mathcal{E}, \rho, \vec{x}} & \text{if } z \notin \vec{x} \end{cases}$$

where  $\text{f-res}_{\mathcal{E}}(m.f) = (\text{abs}_{\text{open}} z)(\vec{p}).f$ .

Remark that both semantics coincide on their first component. Indeed, for any  $\mathcal{E}$ -pre-interpretation  $\rho$ :

$$\forall \nu s. \pi_1^{\#}(\llbracket s \rrbracket_{\nu}^{\mathcal{E}, \rho}) = \pi_1^{\#}(^i \llbracket s \rrbracket_{\nu}^{\mathcal{E}, \rho, \vec{x}})$$

The  $(\mathcal{E}, \rho)$ -intrinsic cost  $\text{i-cost}_{\nu}^{\mathcal{E}, \rho, \vec{x}}(\_) \in \mathbb{N} \cup \{+\infty\}$  of an instruction  $s$  is:

$$\text{i-cost}_{\nu}^{\mathcal{E}, \rho, \vec{x}}(s) = \text{sup}(\text{supp}(\pi_2^{\#}(^i \llbracket s \rrbracket_{\nu}^{\mathcal{E}, \rho, \vec{x}})))$$

The intrinsic cost of a procedure  $m.f$ , with parameters  $\vec{x}$ , is:

- If  $\text{f-res}_{\mathcal{E}}(m.f) = (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \_ ; s ; \text{return } r \})$  then:

$$\text{i-cost}_{\nu}^{\mathcal{E}, \rho, \vec{x}}(m.f) = \text{i-cost}_{\nu}^{\mathcal{E}, \rho, \vec{x}}(s) + c_{\mathbb{E}}(r, \nu)$$

- If  $\text{f-res}_{\mathcal{E}}(m.f) = (\text{abs}_{\text{open}} x)(\vec{p}).f$  then:

$$\text{i-cost}_{\nu}^{\mathcal{E}, \rho, \vec{x}}(m.f) = \text{i-cost}_{\nu}^{\mathcal{E}, \rho, \vec{x}}(\rho(x)(\vec{p}).f)$$

**Module signature and structure sub-typing**  $\vdash M_1 <: M_2$  and  $\vdash S_1 <: S_2$ .*We omit the reflexivity and transitivity rules.*

<p><b>SUBSIG</b></p> $\frac{\vdash S_1 <: S_2 \quad \vdash \theta_1 <: \theta_2}{\vdash \text{sig } S_1 \text{ restr } \theta_1 \text{ end} <: \text{sig } S_2 \text{ restr } \theta_2 \text{ end}}$	<p><b>SUBFUNC</b></p> $\frac{\vdash M'_0 <: M_0 \quad \vdash M <: M'}{\vdash \text{func}(x : M_0) M <: \text{func}(x : M'_0) M'}$
<p><b>SUBSTRUCT</b></p> $\frac{\forall i \in \{1; \dots; n\}, \vdash D_i <: D'_i}{\vdash D_1; \dots; D_n <: D'_1; \dots; D'_n}$	<p><b>SUBMODDECL</b></p> $\frac{\vdash M_1 <: M_2}{\vdash \text{module } x : M_1 <: \text{module } x : M_2}$

**Statements and function paths typing**  $\Gamma \vdash s$  and  $\Gamma \vdash F : \_$ .

<p><b>T-ABORT</b></p> $\frac{}{\Gamma \vdash \text{abort}}$	<p><b>T-SKIP</b></p> $\frac{}{\Gamma \vdash \text{skip}}$	<p><b>T-SEQ</b></p> $\frac{\Gamma \vdash s_1 \quad \Gamma \vdash s_2}{\Gamma \vdash s_1; s_2}$	<p><b>T-ASSIGN</b></p> $\frac{\Gamma \vdash x : \tau \quad \Gamma \vdash e : \tau}{\Gamma \vdash x \leftarrow e}$	<p><b>T-RAND</b></p> $\frac{\Gamma \vdash x : \tau \quad \Gamma \vdash d : \tau}{\Gamma \vdash x \xleftarrow{\$} d}$
<p><b>T-CALL</b></p> $\frac{\Gamma \vdash F : \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r \quad \Gamma \vdash x : \tau_r \quad \Gamma \vdash \vec{e} : \vec{\tau}}{\Gamma \vdash x \leftarrow \text{call } F(\vec{e})}$	<p><b>T-PROC1</b></p> $\frac{\Gamma \vdash p : \text{sig } (S_1; \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r; S_2) \text{ restr } \theta \text{ end}}{\Gamma \vdash p.f : (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r)}$			
<p><b>T-PROC2</b></p> $\frac{\Gamma(p.f) = (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r \rightarrow \_)}{\Gamma \vdash p.f : (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r)}$	<p><b>T-IF</b></p> $\frac{\Gamma \vdash e : \text{bool} \quad \Gamma \vdash s_1 \quad \Gamma \vdash s_2}{\Gamma \vdash \text{if } e \text{ then } s_1 \text{ else } s_2}$	<p><b>T-WHILE</b></p> $\frac{\Gamma \vdash e : \text{bool} \quad \Gamma \vdash s}{\Gamma \vdash \text{while } e \text{ do } s}$		

**Expressions typing**  $\Gamma \vdash e : \tau$ .

<p><b>EXPRAPP</b></p> $\frac{\text{type}(f) = \tau_1 \times \dots \times \tau_n \rightarrow \tau \quad \forall i \in \{1; \dots; n\}, \Gamma \vdash e_i : \tau_i}{\Gamma \vdash f(e_1, \dots, e_n) : \tau}$	<p><b>EXPRVAR</b></p> $\frac{\Gamma(v) = \tau}{\Gamma \vdash v : \tau}$
---	---

**Restriction entailment**  $\vdash \theta <: \theta'$ .*We omit the transitivity and reflexivity rules for  $\vdash \theta <: \theta'$ .*

<p><b><math>\sqsubseteq</math>-PROC</b></p> $\frac{\forall f \in \text{dom}(\theta, \theta'), \vdash \theta[f] <: \theta'[f]}{\vdash \theta <: \theta'}$	<p><b><math>\sqsubseteq</math>-SPLIT</b></p> $\frac{\vdash \lambda_m <: \lambda_m' \quad \vdash \lambda_c <: \lambda_c'}{\vdash \lambda_m \wedge \lambda_c <: \lambda_m' \wedge \lambda_c'}$	<p><b><math>\sqsubseteq</math>-TOP</b></p> $\frac{}{\vdash \lambda <: \top}$	<p><b><math>\sqsubseteq</math>-MEM</b></p> $\frac{\lambda_m \sqsubseteq \lambda_m'}{\vdash \lambda_m <: \lambda_m'}$
<p><b><math>\sqsubseteq</math>-MEMTOP</b></p> $\frac{}{\vdash \lambda_c <: \top}$	<p><b><math>\sqsubseteq</math>-COMPL</b></p> $\frac{k \leq k' \quad \forall i, k_i \leq k'_i}{\vdash \text{compl}[\text{intr} : k, x_1.f_1 : k_1, \dots, x_n.f_n : k_n] <: \text{compl}[\text{intr} : k', x_1.f_1 : k'_1, \dots, x_n.f_n : k'_n]}$		

**Memory restriction.**

<p><math>\text{mem}_\Gamma(\text{abort}) = \emptyset</math></p> <p><math>\text{mem}_\Gamma(x \leftarrow e) = \{x\} \sqcup \text{vars}(e)</math></p> <p><math>\text{mem}_\Gamma(s_1; s_2) = \text{mem}_\Gamma(s_1) \sqcup \text{mem}_\Gamma(s_2)</math></p>	<p><math>\text{mem}_\Gamma(\text{skip}) = \emptyset</math></p> <p><math>\text{mem}_\Gamma(x \xleftarrow{\\$} d) = \{x\}</math></p> <p><math>\text{mem}_\Gamma(\text{while } e \text{ do } s) = \text{vars}(e) \sqcup \text{mem}_\Gamma(s)</math></p>
--	--

$$\text{mem}_\Gamma(\text{if } e \text{ then } s_1 \text{ else } s_2) = \text{vars}(e) \sqcup \text{mem}_\Gamma(s_1) \sqcup \text{mem}_\Gamma(s_2)$$

$$\text{mem}_\Gamma(x \leftarrow \text{call } p.f(\vec{e})) = \{x\} \sqcup \text{mem}_\Gamma(p.f) \sqcup \text{vars}(\vec{e})$$

$$\text{mem}_\Gamma(p) = \sqcup_{f \in \text{procs}_\Gamma(p)} \text{mem}_\Gamma(p.f)$$

When  $f\text{-res}_\Gamma(p.f) = (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s; \text{return } r \})$ :

$$\text{mem}_\Gamma(p.f) = (\text{mem}_\Gamma(s) \cup \text{vars}(r)) \setminus \{ \vec{v}; \vec{v}_1 \}$$

When  $f\text{-res}_\Gamma(p.f) = (\text{abs}_K x)(\vec{p}_0).f$ ,  $K = \text{open}$  and  $\Gamma(x) = \text{abs}_K : \text{func } \_ \text{ sig } \_ \text{ restr } \theta \text{ end}$ :

$$\text{mem}_\Gamma(p.f) = \theta[f] \sqcup \text{mem}_\Gamma(\vec{p}_0)$$

When  $f\text{-res}_\Gamma(p.f) = (\text{abs}_{\text{param}} x)(\vec{p}_0).f$ :

$$\text{mem}_\Gamma(p.f) = \text{mem}_\Gamma(\vec{p}_0)$$

Fig. 23. Additional typing rules and operations.

**Substitution in module signatures and declarations**

$$\begin{aligned}
 (\text{func}(y : M_0)) M_1[x \mapsto \lambda_m] &= \begin{cases} \text{func}(y : M_0[x \mapsto \lambda_m]) (M_1[x \mapsto \lambda_m]) & \text{when } y \neq x \\ \text{func}(y : M_0) M_1 & \text{otherwise} \end{cases} \\
 (\text{sig } S \text{ restr } \theta \text{ end})[x \mapsto \lambda_m] &= \text{sig } (S[x \mapsto \lambda_m]) \text{ restr } (\theta[x \mapsto \lambda_m]) \text{ end} \\
 (D_1; \dots; D_n)[x \mapsto \lambda_m] &= D_1[x \mapsto \lambda_m]; \dots; D_n[x \mapsto \lambda_m] \\
 \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r[x \mapsto \lambda_m] &= \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r \\
 \text{module } y : M[x \mapsto \lambda_m] &= \text{module } y : (M[x \mapsto \lambda_m])
 \end{aligned}$$

**Substitution in module restriction**

$$\begin{aligned}
 ((f_1 : \lambda_1); \dots; (f_n : \lambda_n))[x \mapsto \lambda_m] &= (f_1 : \lambda_1[x \mapsto \lambda_m]); \dots; (f_n : \lambda_n[x \mapsto \lambda_m]) \\
 (\lambda_m^0 \wedge \lambda_c)[x \mapsto \lambda_m] &= \lambda_m^0[x \mapsto \lambda_m] \wedge \lambda_c[x \mapsto \lambda_m] \\
 \top[x \mapsto \lambda_m] &= \top
 \end{aligned}$$

**Substitution in memory restriction**

$$\lambda_m^0[x \mapsto \lambda_m] = \lambda_m^0 \sqcup \lambda_m$$

**Substitution in complexity restriction**

$$\text{compl}[\text{intr} : k, y_1.f_1 : k_1, \dots, y_l.f_l : k_l][x \mapsto \lambda_m] = \text{compl}[\text{intr} : k, (y_1.f_1 : k_1)[x \mapsto \lambda_m], \dots, (y_l.f_l : k_l)[x \mapsto \lambda_m]]$$

$$\text{where } (y.f : k)[x \mapsto \lambda_m] = \begin{cases} \epsilon & \text{if } y = x \\ y.f : k & \text{otherwise} \end{cases}$$

**Substitution in module paths**

$$\begin{aligned}
 y[x \mapsto p] &= \begin{cases} p & \text{if } y = x \\ y & \text{otherwise} \end{cases} & (p'.y)[x \mapsto p] &= (p'[x \mapsto p]).y \\
 & & (p'(p''))[x \mapsto p] &= (p'[x \mapsto p])(p''[x \mapsto p])
 \end{aligned}$$

**Substitution in module expressions**

$$(\text{func}(y : M) m)[x \mapsto p] = \text{func}(y : M) (m[x \mapsto p]) \quad (\text{struct st end})[x \mapsto p] = \text{struct st}[x \mapsto p] \text{ end}$$

**Substitution in module structures, declarations and procedure body**

$$\begin{aligned}
 (d_1; \dots; d_n)[x \mapsto p] &= d_1[x \mapsto p]; \dots; d_n[x \mapsto p] \\
 (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body})[x \mapsto p] &= \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = (\text{body}[x \mapsto p]) \\
 (\text{module } y = m)[x \mapsto p] &= \text{module } y = (m[x \mapsto p]) \\
 \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s; \text{return } e \}[x \mapsto p] &= \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s[x \mapsto p]; \text{return } e \}
 \end{aligned}$$

**Substitution in statements**

$$\begin{aligned}
 \text{abort}[x \mapsto p] &= \text{abort} & \text{skip}[x \mapsto p] &= \text{skip} \\
 x \leftarrow e[x \mapsto p] &= x \leftarrow e & x \stackrel{\$}{\leftarrow} d[x \mapsto p] &= x \stackrel{\$}{\leftarrow} d \\
 (s_1; s_2)[x \mapsto p] &= s_1[x \mapsto p]; s_2[x \mapsto p] & x \leftarrow \text{call } p'.f(\vec{e})[x \mapsto p] &= x \leftarrow \text{call } (p'[x \mapsto p]).f(\vec{e}) \\
 \left( \begin{array}{l} \text{if } e \text{ then } s_1 \\ \text{else } s_2 \end{array} \right) [x \mapsto p] &= \text{if } e \text{ then } (s_1[x \mapsto p]) \\ & \quad \text{else } (s_2[x \mapsto p]) & (\text{while } e \text{ do } s)[x \mapsto p] &= \text{while } e \text{ do } s[x \mapsto p]
 \end{aligned}$$

Fig. 24. Substitution functions.

And the intrinsic cost *in any memory* of an instruction  $s$  or a module procedure  $m.f$  is:

$$\begin{aligned}
 \text{i-cost}^{\mathcal{E}, \rho, \vec{z}}(s) &= \max_{v \in \mathcal{M}} \text{i-cost}_v^{\mathcal{E}, \rho, \vec{z}}(s) \\
 \text{i-cost}^{\mathcal{E}, \rho, \vec{z}}(m.f) &= \max_{v \in \mathcal{M}} \text{i-cost}_v^{\mathcal{E}, \rho, \vec{z}}(m.f)
 \end{aligned}$$

*Interpretations.* We now define when a pre-interpretation is an interpretation.



$$\begin{aligned}
\gamma.g \llbracket \mathbf{skip} \rrbracket_v^{\mathcal{E}, \rho} &= \mathbb{1}_{(v, 0)} \\
\gamma.g \llbracket \mathbf{abort} \rrbracket_v^{\mathcal{E}, \rho} &= \mathbf{0} \\
\gamma.g \llbracket s_1; s_2 \rrbracket_v^{\mathcal{E}, \rho} &= \mathbb{E}_{(v', c') \sim \gamma.g \llbracket s_1 \rrbracket_v^{\mathcal{E}, \rho}} [\gamma.g \llbracket s_2 \rrbracket_{v'}^{\mathcal{E}, \rho} \oplus c'] \\
\gamma.g \llbracket x \leftarrow e \rrbracket_v^{\mathcal{E}, \rho} &= \mathbb{1}_{(v[x \leftarrow (e)], 0)} \\
\gamma.g \llbracket x \xleftarrow{\$} d \rrbracket_v^{\mathcal{E}, \rho} &= \mathbb{E}_{v \sim (d)}_v [\mathbb{1}_{(v[x \leftarrow v], 0)}] \\
\gamma.g \llbracket x \leftarrow \mathbf{call} \ m.f(\vec{e}) \rrbracket_v^{\mathcal{E}, \rho} &= \text{let } v_0 = v[\vec{v} \leftarrow (\vec{e})]_v \text{ in} \\
&\quad \mathbb{E}_{(v', c') \sim \gamma.g \llbracket s \rrbracket_{v_0}^{\mathcal{E}, \rho}} [\mathbb{1}_{(v'[x \leftarrow (r)], v', c')}] \\
&\quad \text{(if } f\text{-res}_{\mathcal{E}}(m.f) = \text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \_ ; s ; \text{return } r \} \text{)} \\
\gamma.g \llbracket x \leftarrow \mathbf{call} \ m.f(\vec{e}) \rrbracket_v^{\mathcal{E}, \rho} &= \gamma.g \llbracket x \leftarrow \mathbf{call} \ \rho(x)(\vec{p}).f(\vec{e}) \rrbracket_v^{\mathcal{E}, \rho} \oplus \mathbf{1} \quad \text{(if } f\text{-res}_{\mathcal{E}}(m.f) = (\mathbf{abs}_{\text{open}} \ x)(\vec{p}).f \text{ and } x.f = \gamma.g \text{)} \\
\gamma.g \llbracket x \leftarrow \mathbf{call} \ m.f(\vec{e}) \rrbracket_v^{\mathcal{E}, \rho} &= \gamma.g \llbracket x \leftarrow \mathbf{call} \ \rho(x)(\vec{p}).f(\vec{e}) \rrbracket_v^{\mathcal{E}, \rho} \quad \text{(if } f\text{-res}_{\mathcal{E}}(m.f) = (\mathbf{abs}_{\text{open}} \ x)(\vec{p}).f \text{ and } x.f \neq \gamma.g \text{)} \\
\gamma.g \llbracket \mathbf{if } e \text{ then } s_1 \text{ else } s_2 \rrbracket_v^{\mathcal{E}, \rho} &= \begin{cases} \gamma.g \llbracket s_1 \rrbracket_v^{\mathcal{E}, \rho} \oplus \mathbf{0} & \text{if } (e)_v \neq \mathbf{0} \\ \gamma.g \llbracket s_2 \rrbracket_v^{\mathcal{E}, \rho} \oplus \mathbf{0} & \text{otherwise} \end{cases} \\
\gamma.g \llbracket \mathbf{while } e \text{ do } s \rrbracket_v^{\mathcal{E}, \rho} &= \lim_{n \rightarrow \infty} \gamma.g \llbracket \mathbf{loop}_n^{e, s} \rrbracket_v^{\mathcal{E}, \rho}
\end{aligned}$$

Fig. 26. Function call counting semantics  $\gamma.g \llbracket \_ \rrbracket_v^{\mathcal{E}, \rho}$ .

*Definition B.1.* Let  $\mathcal{E}$  be a well-typed environment. A  $\mathcal{E}$ -pre-interpretation  $\rho$  is an  $\mathcal{E}$ -interpretation if for every module identifier  $x$  such that  $\mathcal{E} = \mathcal{E}_1$ ; module  $x = \mathbf{abs}_{\text{open}} : M_1; \mathcal{E}_2$  where:

$$M_1 = \text{func}(\vec{z} : \vec{M}) \text{ sig } S_1 \text{ restr } \theta \text{ end}$$

and for every procedure  $f \in \text{procs}(S)$ , for every valuation  $\vec{m}$  of the functor's parameters such that, for every  $1 \leq i \leq |\vec{z}|$ , if we let  $z_i = \vec{z}[i]$ ,  $m_i = \vec{m}[i]$  and  $M_i = \vec{M}[i] = \text{sig } \_ \text{ restr } \lambda_c^i \wedge \_ \text{ end}$ ,<sup>12</sup> if:

$$\mathcal{E} \vdash (\text{module } z_i = m_i : \text{erase}_{\text{compl}}(M_i))$$

and

$$\forall g \in \text{procs}(M_i), \text{cost}_v^{\mathcal{E}}(m_i.g) \leq \lambda_c^i$$

(with the convention that  $j \leq \top$  for any integer  $j$ ) then the execution of  $f$  in any memory verifies the complexity restriction in  $\theta[f]$ . Formally, let  $\mathcal{E}' = \mathcal{E}$ ; module  $\vec{z} = \mathbf{abs}_{\text{open}} : \vec{M}$  and  $\rho' = \rho, (\vec{z} : \vec{m})$ , and:

$$\theta[f] = \_ \wedge \lambda_c = \mathbf{compl}[\text{intr} : k, \gamma_1.f_1 : k_1, \dots, \gamma_l.f_l : k_l]$$

Then for every  $1 \leq j \leq l$ ,

$$\#\text{calls}_{\gamma_j.f_j}^{\mathcal{E}', \rho'}(x(\vec{z}).f) \leq k_j \quad \text{and} \quad \text{i-cost}^{\mathcal{E}', \rho', \vec{z}}(x(\vec{z}).f) \leq k$$

*Intrinsic cost of a functor.* Finally, the  $(\mathcal{E}, \rho)$ -intrinsic complexity of a functor procedure  $x.f$ , denoted by  $\text{compl}_{x.f}^{\mathcal{E}, \rho} \in \mathbb{N} \cup \{+\infty\}$ , is the maximal intrinsic cost of  $x.f$ 's body over all possible memories and instantiation of  $x$ 's functor parameters. Let  $\mathcal{E}(x) = \mathbf{abs}_{\text{open}}(\text{func}(\vec{z} : \vec{M}) \text{ sig } \_ \text{end})$  and  $\mathcal{E}' = (\mathcal{E}; \text{module } \vec{z} = \mathbf{abs}_{\text{open}} : \vec{M})$ . Also, let  $\mathcal{I}$  be the set  $\mathcal{E}'$ -interpretation  $\rho'$  extending  $\rho$ . Then:

$$\text{compl}_{x.f}^{\mathcal{E}, \rho} = \sup_{\rho' \in \mathcal{I}} \text{i-cost}^{\mathcal{E}', \rho', \vec{z}}(x(\vec{z}).f)$$

<sup>12</sup>Indeed, since  $M_1$  is a low-order signature,  $M_i$  must be a module structure signature.

## C SUBJECT REDUCTION PROPERTY FOR MODULE RESOLUTION

To prove the soundness of our logic, we are going to need to show that our module resolution mechanism has the subject reduction property, i.e. that evaluating a module expression preserves its type. This is a crucial property that allows to show that memory and complexity restrictions are preserved by the evaluation of a module using the module resolution mechanism. That is, the module expression resulting from the evaluation of another module expression has the same memory footprint, and satisfies the same complexity restrictions.

Unfortunately, this does not directly hold, for a technical reason we describe later (Appendix C.1). We circumvent this problem by defining an alternative notion of typing judgement (Appendix C.2), for which we show that subject reduction does hold (Appendix C.4). This alternative notion of typing judgements for module expressions will be sufficient to prove the soundness of our logic later, in Appendix D.

### C.1 Resolution and sub-module declarations

The following example shows why subject reduction does not hold for our *initial* notion of typing judgements, because of sub-modules declarations.

*Example C.1.* Consider a well-typed typing environment  $\vdash \Gamma$ , and a module path  $p$  where:

$$\Gamma \vdash p : \text{sig } S_1; \text{ module } x : M; \_ \text{ restr } \_ \text{ end}$$

We are going to assume that some kind of subject reduction property holds for  $p$ . More precisely, we assume that:

$$f\text{-res}_\Gamma(p) = (\text{struct } st_1; \text{ module } x = m; \_ \text{ end})$$

and that we have a derivation:

$$\Gamma \vdash_p (\text{struct } st_1; \text{ module } x = m; \_ \text{ end}) : (\text{sig } S_1; \text{ module } x : M; \_ \text{ restr } \_ \text{ end})$$

Then, we know that  $p.x$  resolves to  $m$ , i.e.  $f\text{-res}_\Gamma(p) = m$ . But we do not have:

$$\Gamma \vdash_{p.x} m : M$$

The problem is that the sub-module  $m$  may use sub-modules declared in  $st_1$ . Consequently, it is not well-typed in  $\Gamma$ , but in an extended typing environment, where the sub-module declarations in  $st_1$  (which have types  $S_1$ ) have been added to  $\Gamma$ . For example, we can have:

$$st_1 = (\text{module } z = m_0) \qquad m = z$$

Therefore, we cannot state a subject reduction property for the module resolution function w.r.t. the typing judgment  $\Gamma \vdash_p m : M$ .

*Alternative typing judgements for module expressions.* Instead, we introduce another typing judgment, noted  $\Gamma \Vdash m : M$ , which is similar to the typing judgment  $\Gamma \vdash_p m : M$  of Fig. 6, but is used to type a module expression in an environment which has already been typed, while  $\Gamma \vdash_p m : M$  is used to type a module declaration in an environment where some modules have not yet been fully typed. We postpone its definition to Appendix C.2. Using this alternative typing judgment notion, we can state the subject reduction property we want

**LEMMA C.1 (SUBJECT REDUCTION).** *If  $\Gamma \Vdash \mathcal{E}$  and  $\Gamma \Vdash m : M$  then  $\Gamma \Vdash \text{res}_\mathcal{E}(m) : M$  whenever  $\text{res}_\mathcal{E}(m)$  is well-defined.*

### Module expression typing $\Gamma \Vdash \tilde{m} : M$ .

$$\begin{array}{c}
\text{D-ALIAS} \\
\frac{\Gamma \vdash p_a : M}{\Gamma \Vdash p_a : M} \\
\\
\text{D-FUNC} \\
\frac{\Gamma \vdash M_0 \quad \Gamma(x) \not\equiv_{\text{undef}} \quad \Gamma, \text{module } x = \text{absparam} : M_0 \Vdash m : M}{\Gamma \Vdash \text{func}(x : M_0) m : \text{func}(x : M_0) M} \\
\\
\text{D-STRUCT} \\
\frac{\Gamma \Vdash_{\theta} \text{st} : S}{\Gamma \Vdash \text{struct st end} : \text{sig } S \text{ restr } \theta \text{ end}} \\
\\
\text{D-SUB} \\
\frac{\Gamma \Vdash \tilde{m} : M_0 \quad \vdash M_0 <: M}{\Gamma \Vdash \tilde{m} : M} \\
\\
\text{D-ABSEMP} \\
\frac{\Gamma \Vdash x : M_1}{\Gamma \Vdash \text{abs}_{\mathcal{K}} x : M_1} \\
\\
\text{D-ABS} \\
\frac{\Gamma \Vdash (\text{abs}_{\mathcal{K}} x)(\vec{p}) : \text{func}(y : M') M \quad \Gamma \Vdash p' : M'}{\Gamma \Vdash (\text{abs}_{\mathcal{K}} x)(\vec{p}, p') : M[y \mapsto \text{mem}_{\Gamma}(p')]} \\
\\
\text{D-COMPNT} \\
\frac{\Gamma \Vdash \tilde{m} : \text{sig } S_1; \text{ module } x : M; S_2 \text{ restr } \theta \text{ end}}{\Gamma \Vdash \tilde{m}.x : M} \\
\\
\text{D-FUNCAPP} \\
\frac{\Gamma \Vdash \tilde{m} : \text{func}(x : M') M \quad \Gamma \vdash p' : M'}{\Gamma \Vdash \tilde{m}(p') : M[x \mapsto \text{mem}_{\Gamma}(p')]}
\end{array}$$

### Module structure typing $\Gamma \Vdash_{\theta} \text{st} : S$ .

$$\begin{array}{c}
\text{D-PROCDECL} \\
\frac{\text{body} = \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s; \text{return } r \} \quad \vec{v}, \vec{v}_1 \text{ fresh in } \Gamma \quad \Gamma_f = \Gamma, \text{var } \vec{v} : \vec{\tau}, \text{var } \vec{v}_1 : \vec{\tau}_1 \\
\Gamma_f \vdash s \quad \Gamma_f \vdash r : \tau_r \quad \Gamma \vdash \text{body} \triangleright_{\theta}[f] \quad \Gamma \Vdash_{\theta} \text{st} : S}{\Gamma \Vdash_{\theta} (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{body}; \text{st}) : (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r; S)} \\
\\
\text{D-MODDECL} \\
\frac{\Gamma \Vdash m : M \quad \Gamma \Vdash_{\theta} \text{st} : S}{\Gamma \Vdash_{\theta} (\text{module } x = m; \text{st}) : (\text{module } x : M; S)} \\
\\
\text{D-STRUCTEMP} \\
\frac{}{\Gamma \Vdash_{\theta} \epsilon : \epsilon}
\end{array}$$

### Environments typing $\Gamma \Vdash \mathcal{E}$

$$\begin{array}{c}
\text{D-ENVSEQ} \\
\frac{\Gamma \Vdash \Gamma_1 \quad \Gamma \Vdash \Gamma_2}{\Gamma \Vdash (\Gamma_1; \Gamma_2)} \\
\\
\text{D-ENVEMP} \\
\frac{}{\Gamma \Vdash \epsilon} \\
\\
\text{D-ENVVAR} \\
\frac{\Gamma(v) = (\text{var } v : \tau)}{\Gamma \Vdash (\text{var } v : \tau)} \\
\\
\text{D-ENVMOD} \\
\frac{\Gamma \Vdash m : M}{\Gamma \Vdash (\text{module } x = m : M)} \\
\\
\text{D-ENVABS} \\
\frac{\Gamma \vdash M_1 \quad \Gamma(x) = \text{abs}_{\mathcal{K}} x : M_1}{\Gamma \Vdash (\text{module } x = \text{abs}_{\mathcal{K}} x : M_1)}
\end{array}$$

Fig. 27. Typing rules in typed environments.

*Proof outline.* The proof is essentially the proof that simply typed  $\lambda$ -calculus has the subject reduction properties, with some consequent additional work to handle module restrictions.

The rest of this section is organized as follows: we define the typing rules for  $\Gamma \Vdash m : M$  in Appendix C.2, and prove the link between  $\Gamma \Vdash m : M$  and  $\Gamma \vdash_p m : M$ ; we present an alternative, more computational, way of defining the module resolution procedure in Appendix C.3; finally, we prove that the module resolution procedure has the subject reduction property in Appendix C.4.

## C.2 Typing in Typed Environments

During the resolution of a module expression  $m$ , we may have to resolve applied module paths of the form  $p(p')$ . This is done by first evaluating  $p$  into a functor or an abstract module, and then

performing the application. Consequently, we can have intermediate expressions of the form  $m(p)$ . Similarly, when resolving a module component access  $p.x$ , we first resolve  $p$  into a module structure and then access its component  $x$ . This yields intermediate expressions of the form  $m.x$ .

*Example C.2.* We recall Example 3.1 which we presented in Section 3. We recall that  $\Gamma$  is a typing environment  $\Gamma$  such that:

$$\begin{aligned} \Gamma(x) &= \text{struct module } y = \text{func}(u : \_) \text{ u end} & \Gamma(z) &= \text{abs}_{\text{open}} z \\ \Gamma(v) &= m_v & \Gamma(w) &= \text{abs}_{\text{param}} w \end{aligned}$$

where  $m_v$  is some module expression. The resolution of the module expression  $x.y(z)(v)(w)$  in  $\Gamma$  is as follows:

$$\begin{aligned} &x.y(z)(v)(w) \\ &\Rightarrow_{\text{res. step}} (\text{struct module } y = \text{func}(u : \_) \text{ u end}).y(z)(v)(w) \\ &\Rightarrow_{\text{res. step}} (\text{func}(u : \_) \text{ u})(z)(v)(w) \\ &\Rightarrow_{\text{res. step}} z(v)(w) \\ &\Rightarrow_{\text{res. step}} (\text{abs}_{\text{open}} z)(v)(w) \\ &\Rightarrow_{\text{res. step}} ((\text{abs}_{\text{open}} z)(v))(w) \\ &\Rightarrow_{\text{res. step}} (\text{abs}_{\text{open}} z)(v, w) \end{aligned}$$

Therefore, to prove subject reduction for the module resolution procedure, we consider an extended syntax for modules (and add the corresponding typing rules).

*Definition C.1.* A *partially resolved* module expression  $\tilde{m}$  is an element of the form:

$$\tilde{m} ::= \tilde{m} \mid \tilde{m}(p) \mid \tilde{m}.x$$

It is enough to allow only top-level applications of module expressions to module paths and module component accesses.

We give the rules for typing in already typed environment in Fig. 27. Roughly, these rules are the module expression, module structure and module environment typing rules given in Fig. 6, with the following changes:

- when typing a module declaration of a module structure, the typing environment  $\Gamma$  is not extended with the inferred type, as it must already be present in  $\Gamma$ . Consequently, we do not need to keep track of the current path of the sub-module expression being typed, which simplifies the rules.
- we added a rule for abstract modules, for the application of a module *expression* to a module path, and for component access. This allows to type the intermediate terms that appear during the module resolution procedure.
- when typing an environment, the typing environment is not extended with the inferred types.

Before going further, we give the definition of well-formed typing environments. Essentially, a typing environment  $\Gamma$  is well-formed if it contains no duplicate declarations, and it contains only well-formed module signature.

*Definition C.2.* A typing environment  $\Gamma$  is well-formed iff:

- whenever  $\Gamma = (\Gamma_0; \text{Decl}_z; \Gamma_1)$  then  $(\Gamma_0; \Gamma_1)(z) \not\stackrel{!}{\downarrow} \text{undef}$ , where:

$$\text{Decl}_z \in \{ \text{module } z = m : M; \text{ var } z : \tau; \text{ proc } z(\vec{v} : \vec{\tau}) \rightarrow \tau_r \}$$

- whenever  $\Gamma = \Gamma_0$ ; module  $x = m : M$ ;  $\Gamma_1$  then  $M$  is well-formed in  $\Gamma_0$ , i.e.  $\Gamma_0 \vdash M$ .

It is straightforward to check that if  $\Gamma$  is well-typed, then  $\Gamma$  is well-formed.

**PROPOSITION C.2.** *If  $\vdash \Gamma$  then  $\Gamma$  is well-formed.*

**PROOF.** This is by induction over  $\vdash \Gamma$ . □

We recall that given a module path  $p$ ,  $\Gamma(p)$  returns the typing declaration corresponding to  $p$  in  $\Gamma$ , if it exists. Note that this does not descend in sub-module declarations to retrieve the type of  $p$ . For example, assume that:

$$\Gamma = (\text{module } p = \text{struct module } y = m \text{ end} : (\text{sig module } y : M \text{ restr } \_ \text{ end}))$$

then  $\Gamma(p) = \_ : (\text{sig module } y : M \text{ restr } \_ \text{ end})$ , but  $\Gamma(p.y)$  is not defined. We introduce a new notion of lookup in a typing environment,  $\Gamma[p]$ , which descends in sub-module declarations. It also descends below functor definitions, in the case where the functor argument is a module identifier. Continuing the example above, we have  $\Gamma[p.y] = m : M$ . Formally:

**Definition C.3.** For every well-formed typing environment  $\Gamma$  and module path  $p$ , we let:

$$\begin{aligned} \Gamma[p] &= m : M && (\text{if } \Gamma(p) = m : M) \\ \Gamma[p(x)] &= m : M && (\text{if } \Gamma[p] = \text{func}(x : M_0) m : \text{func}(x : M_0) M) \end{aligned}$$

where the functor rule is module alpha-renaming. Moreover, if:

$$\Gamma[p] = (\text{struct } \_ ; \text{ module } x = m ; \_ \text{ end}) : (\text{sig } (\_ ; \text{ module } x : M ; \_) \text{ restr } \_ \text{ end})$$

then:

$$\Gamma[p.x] = m : M$$

The following proposition allows to replace a typing environment  $\Gamma_0$  by  $\Gamma_1$ , as long as they coincide w.r.t.  $\_ [p]$  on every  $p$  such that  $\Gamma_0[p]$  is well-defined.

**PROPOSITION C.3.** *For every well-formed environments  $\Gamma_0$  and  $\Gamma_1$ , if  $\Gamma_0[p] = \Gamma_1[p]$  for every  $p$  such that  $\Gamma_0[p]$  is well-defined, then for any  $m$  and  $M$ , if  $\Gamma_0 \Vdash m : M$  then  $\Gamma_1 \Vdash m : M$*

**PROOF.** This is immediate by induction over the typing derivations  $\Gamma_0 \Vdash m : M$ . □

**Example C.3.** The following two typing environments  $\Gamma_0$  and  $\Gamma_1$  verify the proposition hypothesis:

$$\Gamma_0 = (\text{module } p.x = m : M ; \text{ module } p.y = m' : M')$$

and:

$$\Gamma_1 = \left( \text{module } p = (\text{struct module } x = m ; \text{ module } y = m' \text{ end}) \right. \\ \left. : (\text{sig module } x : M ; \text{ module } y : M' \text{ restr } \_ \text{ end}) \right)$$

We can strengthen the context.

**PROPOSITION C.4.** *For any well-formed environment  $\Gamma$ , if  $(\Gamma, \Gamma_1)$  is well-formed then  $\Gamma ; \Gamma_1 \Vdash m : M$  whenever  $\Gamma \Vdash m : M$ .*

**PROOF.** This is straightforward by induction over the typing derivation  $\Gamma \Vdash m : M$ . □

We can replay a typing derivation if we have a finer type for a module declaration in the context.

PROPOSITION C.5. For any well-formed environment  $\Gamma$ , if

$$\Gamma, \text{ module } p = m : M, \Gamma_1 \Vdash m_0 : M_0$$

then for any  $\vdash M' <: M$ , we have

$$\Gamma, \text{ module } p = m : M', \Gamma_1 \Vdash m_0 : M_0$$

PROOF. We show this by induction over the typing derivation  $\Gamma, \text{ module } p : M, \Gamma_1 \Vdash m : M_0$ , by replacing any application of the **D-ALIAS** typing rule on a path starting by  $p$ , i.e. of the form  $p(p_0, \dots, p_n).p'$ , by an application of **D-ALIAS** on  $p$ , followed by **D-SUB** and by the applications of **D-FUNCAAPP** and **D-COMPNT** to replay the path typing derivation of  $\Gamma \vdash p(p_0, \dots, p_n).p' : M$  as a module expression typing derivation of  $\Gamma \Vdash p(p_0, \dots, p_n).p' : M$ .  $\square$

LEMMA C.6. For every well-formed typing environment  $\Gamma$  such that  $\Gamma(p) \not\downarrow_{\text{undef}}$ :

- if  $\Gamma \vdash_p m : M$  then there exists  $M'$  well-formed in  $\Gamma$  such that  $\vdash M' <: M$  and:

$$\Gamma; \text{ module } p = m : M' \Vdash m : M' \quad (2)$$

- if  $\Gamma \vdash_{p, \theta} (d_1, \dots, d_n) : (D_1, \dots, D_n)$  then there exists  $D'_1, \dots, D'_n$  well-formed such that  $\vdash D'_i <: D_i$  for every  $i$ , and:

$$\Gamma; \delta_1, \dots, \delta_n \Vdash_{\theta} (d_1, \dots, d_n) : (D_1, \dots, D_n)$$

where for every  $i$ :

$$\delta_i = \text{ module } p.x = m : M$$

if  $D'_i = (\text{ module } x : M)$  and  $d_i = (\text{ module } x = m)$ , and

$$\delta_i = \text{ proc } p.f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{ body}$$

if  $d_i = (\text{ proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{ body})$ .

PROOF. The proof is by induction over the typing derivations in hypothesis.

- the **ALIAS** and **STRUCTEMP** cases are immediate. Note that we use the hypothesis  $\Gamma(p) \not\downarrow_{\text{undef}}$  in the **ALIAS** case, to ensure that the path lookup in  $\Gamma$  is still well-defined.
- we show the **STRUCT** case by applying the induction hypothesis, and using Proposition C.3.
- for **MODDECL** we have  $st = (\text{ module } x = m_0; st_0)$ , and  $S = (\text{ module } x : M_0; S_0)$ , and a derivation of the form:

$$\frac{\text{ModDECL} \quad \Gamma \vdash_{p,x} m_0 : M_0 \quad \Gamma(p.x) \not\downarrow_{\text{undef}} \quad \Gamma, \text{ module } p.x = m_0 : M_0 \vdash_{p, \theta} st_0 : S_0}{\Gamma \vdash_{p, \theta} (\text{ module } x = M_0; st_0) : (\text{ module } x : M_0; S_0)}$$

We know that  $st_0 = (\text{ struct } d_1, \dots, d_n \text{ end})$  and:

$$S_0 = (\text{ sig } D_1, \dots, D_n \text{ restr } \_ \text{ end})$$

By applying the induction hypothesis twice, we know that there are derivations of:

$$\Gamma, \text{ module } p.x = m_0 : M'_0 \Vdash m_0 : M'_0 \quad (3)$$

$$\Gamma, \text{ module } p.x = m_0 : M_0, \Gamma'' \Vdash_{\theta} (d_1, \dots, d_n) : (D'_1, \dots, D'_n) \quad (4)$$

where  $\vdash M'_0 <: M_0$ , and  $\Gamma'' = (D'_1, \dots, D'_n)$  where for every  $i$ ,  $\vdash D'_i <: D_i$  and:

$$D''_i = \begin{cases} \text{ module } p.y = m : M \\ \quad (\text{ if } D'_i = (\text{ module } y : M) \text{ and } d_i = (\text{ module } y = m)) \\ \text{ proc } p.f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{ body} \\ \quad (\text{ if } d_i = (\text{ proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \text{ body})) \end{cases}$$

By Proposition C.4, we deduce from Eq. (3) that there is a derivation of:

$$\Gamma, \text{module } p.x = m_0 : M'_0, \Gamma'' \Vdash m_0 : M'_0$$

By Proposition C.5, we deduce from Eq. (4) that there is a derivation of:

$$\Gamma, \text{module } p.x = m_0 : M'_0, \Gamma'' \Vdash_{\theta} (d_1, \dots, d_n) : (D'_1, \dots, D'_n)$$

We conclude by applying the **D-MODDECL** typing rule.

- **PROCDECL** is similar to the **MODDECL** case. We omit the details.
- for **SUB**, we have:

$$\frac{\text{SUB} \quad \Gamma \vdash_p m : M_0 \quad \vdash M_0 <: M}{\Gamma \vdash_p m : M}$$

By induction hypothesis, we have a derivation of  $\Gamma, \text{module } p : M'_0 \Vdash m : M'_0$  where  $\vdash M'_0 <: M_0$ . Using the sub-typing transitivity rule, we have that  $\vdash M'_0 <: M$ . This concludes this case.

- for **FUNC** we have  $m = \text{func}(x : M_0) m'$  and  $M = \text{func}(x : M_0) M'$ , and the derivation:

$$\frac{\text{FUNC} \quad \Gamma \vdash M_0 \quad \Gamma(x) \not\leq_{\text{undef}} \quad \Gamma, \text{module } x = \text{abs}_{\text{param}} : M_0 \vdash_{p(x)} m' : M'}{\Gamma \vdash_p \text{func}(x : M_0) m' : \text{func}(x : M_0) M'}$$

By induction hypothesis, we have  $M''$  such that  $\vdash M'' <: M'$  and:

$$\Gamma, \text{module } x = \text{abs}_{\text{param}} : M_0, \text{module } p(x) = m' : M'' \Vdash m' : M''$$

is derivable. We re-order the declarations in the context, to obtain a derivation of:

$$\Gamma, \text{module } p(x) = m' : M'', \text{module } x = \text{abs}_{\text{param}} : M_0 \Vdash m' : M''$$

(note that the new typing environment is still well-formed). By applying the **D-FUNC** rule, there is a derivation of:

$$\Gamma, \text{module } p(x) = m' : M'' \Vdash \text{func}(x : M_0) m' : \text{func}(x : M_0) M''$$

By Proposition C.3, there is a derivation of:

$$\Gamma, \text{module } p = (\text{func}(x : M_0) m') : (\text{func}(x : M_0) M'') \Vdash \text{func}(x : M_0) m' : \text{func}(x : M_0) M'' \quad \square$$

We extend sub-typing judgments to typing environments, by requiring that  $\vdash \Gamma_0 <: \Gamma_1$  whenever  $\Gamma_0$  and  $\Gamma_1$  are of the same length, and any declaration in  $\Gamma_0$  is a sub-type of the corresponding declaration in  $\Gamma_1$ .

**LEMMA C.7.** *If  $\vdash \Gamma$  then there exists  $\vdash \Gamma' <: \Gamma$  such that  $\Gamma' \Vdash \Gamma$ .*

**PROOF.** We use Lemma C.6 to replace the **ENVMOD** rules by **D-ENVMOD** rules. We omit the details.  $\square$

### C.3 Module Resolution as a Rewrite Relation

We define in Fig. 28 the relation  $\rightarrow_{\Gamma}$  on partially resolved module expressions. This relation is exactly the evaluation strategy used by the module resolution procedure defined in Fig. 9.

We let  $\tilde{m} \downarrow_{\Gamma}$  be the normal form of  $\tilde{m}$  w.r.t. the reflexive and transitive closure of  $\rightarrow_{\Gamma}$ , if it exists.

We now state a property of  $\rightarrow_{\Gamma}$  in the case where  $\Gamma$  is an *environment*, i.e. top-level module declarations are not module aliases.

**PROPOSITION C.8.** *The relation  $\rightarrow_{\Gamma}$  is deterministic on well-formed module expression: for every well-formed  $\tilde{m}$ , if  $\tilde{m} \rightarrow_{\Gamma} \tilde{m}_0$  and  $\tilde{m} \rightarrow_{\Gamma} \tilde{m}_1$  then  $\tilde{m}_0 = \tilde{m}_1$ .*

*Moreover, for any well-formed  $\mathcal{E}$  and  $\tilde{m}$ , if  $\text{res}_{\mathcal{E}}(\tilde{m})$  is well-defined then  $\text{res}_{\mathcal{E}}(\tilde{m}) = \tilde{m} \downarrow_{\mathcal{E}}$ .*

$p \rightarrow_{\Gamma} m$	if $\Gamma(p) = m : \_$
$m.x \rightarrow_{\Gamma} m'$	if $m = \text{struct } st_1; \text{ module } x = m'; st_2 \text{ end}$
$m.x \rightarrow_{\Gamma} m'.x$	if $m \rightarrow_{\Gamma} m'$
$m(p) \rightarrow_{\Gamma} m_0[x \mapsto p]$	if $m = \text{func}(x : M) m_0$ and $x$ is fresh in $\mathcal{E}$
$\tilde{m}(p) \rightarrow_{\Gamma} (\text{abs}_{\kappa} x)(\tilde{p}_0, p)$	if $\tilde{m} = (\text{abs}_{\kappa} x)(\tilde{p}_0)$
$m(p) \rightarrow_{\Gamma} m'(p)$	if $m \rightarrow_{\Gamma} m'$

Fig. 28. The relation  $\rightarrow_{\Gamma}$ .

PROOF. The fact that  $\rightarrow_{\Gamma}$  is deterministic is straightforward from the definition of  $\rightarrow_{\Gamma}$  in Fig. 28: indeed, we only need to observe that well-formed module expressions cannot have two module declarations with the same name.

Moreover, for any  $\tilde{m}_1$  and  $\tilde{m}_2$ , if  $\tilde{m}_1 \rightarrow_{\Gamma} \tilde{m}_2$  and  $\tilde{m}_1$  is well-formed then so is  $\tilde{m}_2$ . Hence if  $\rightarrow_{\Gamma}$  terminates on a well-formed  $\tilde{m}$  then  $\tilde{m} \downarrow_{\Gamma}$  exists.

Let  $\mathcal{E}$  be a well-typed environment. We show the second point by induction over the length of the computation of  $\text{res}_{\mathcal{E}}(\tilde{m})$ . In the environment lookup case, we use the fact that  $\mathcal{E}(p)$  is either a module structure, functor or an abstract module, and is therefore in  $\rightarrow_{\mathcal{E}}$ -normal form. We omit the rest of the proof.  $\square$

#### C.4 Subject Reduction

Before proving that our system has the subject reduction property, we state the following substitution lemmas.

LEMMA C.9 (SUBSTITUTION 1). *We have the following substitution properties:*

- if  $\vdash \theta_1 <: \theta_2$  then  $\vdash \theta_1[x \mapsto \lambda_m] <: \theta_2[x \mapsto \lambda_m]$ .
- if  $\vdash \lambda_1 <: \lambda_2$  then  $\vdash \lambda_1[x \mapsto \lambda_m] <: \lambda_2[x \mapsto \lambda_m]$ .
- if  $\vdash M_0 <: M'_0$  then  $\vdash M_0[x \mapsto \lambda_m] <: M'_0[x \mapsto \lambda_m]$ .

PROOF. The properties above are shown by induction over their respective typing derivation.  $\square$

LEMMA C.10 (SUBSTITUTION 2). *For every:*

$$\Gamma_1; \text{ module } x = \text{abs}_{\text{param}} : M; \Gamma_2$$

*which is a well-formed typing environment, if:*

$$\Gamma_1; \text{ module } x = \text{abs}_{\text{param}} : M; \Gamma_2 \vdash e : \tau$$

*then:*

$$\Gamma_1; \Gamma_2[x \mapsto \lambda_m] \vdash e : \tau$$

PROOF. The proof is immediate by induction over the typing derivation, since module types in the typing environment are not used when typing expressions.  $\square$

PROPOSITION C.11. *For any well-formed typing environment  $\Gamma$ , if  $\Gamma \vdash p : M$  and  $\text{mem}_{\Gamma}(p) = \lambda_m$  then:*

$$\text{mem}_{\Gamma}(F[x \mapsto p]) \subseteq (\text{mem}_{\Gamma, \text{module } x = \text{abs}_{\text{param}} : M(F)})(x \mapsto \lambda_m)$$

LEMMA C.12 (SUBSTITUTION 3). *For every well-formed environment  $\Gamma$  of the form:*

$$\Gamma_1; \text{ module } x = \text{abs}_{\text{param}} : M; \Gamma_2$$

*for every module path  $p$  such that  $\Gamma_1 \vdash p : M$ , if  $\text{mem}_{\Gamma_1}(p) = \lambda_m$  then:*



- if  $\Gamma \Vdash m : M_0$  then:

$$\Gamma_1; \Gamma_2[x \mapsto \lambda_m] \Vdash m[x \mapsto p] : M_0[x \mapsto \lambda_m]$$

- if  $\Gamma \vdash \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s; \text{return } e \} \triangleright \lambda$  then:

$$\Gamma_1; \Gamma_2[x \mapsto \lambda_m] \vdash \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s[x \mapsto p]; \text{return } e \} \triangleright \lambda[x \mapsto \lambda_m]$$

PROOF. We prove the two properties above simultaneously, by induction over the corresponding typing derivations. The proof is straightforward (we omit the details).

Note that we do not need to show closure under substitution of module parameters in *Hoare* derivations, only in typing derivations (see Remark 3.1).  $\square$

LEMMA C.13. *If  $\Gamma \Vdash \mathcal{E}$  and  $\Gamma \Vdash m : M$  then if  $m \rightarrow_{\Gamma} m'$  then  $\Gamma \Vdash m' : M$ .*

PROOF. W.l.o.g., we assume that the typing derivations  $\Gamma \Vdash \mathcal{E}$  and  $\Gamma \Vdash m : M$  never apply the **D-SUB** typing rules twice in a row (using the transitivity rule for sub-typing judgments). To simplify derivations, we also assume that **D-SUB** is applied once between each typing rule application (using the sub-typing reflexivity rule if necessary).

We do a case analysis on the reduction  $m \rightarrow_{\Gamma} m'$ .

- if  $x \rightarrow_{\Gamma} \mathcal{E}(x)$ . We check that we must have a derivation of the following form:

$$\frac{\text{D-ALIAS} \frac{\Gamma(x) = \_ : M_0}{\Gamma \Vdash x : M_0} \quad \vdash M_0 <: M}{\text{D-SUB} \quad \Gamma \Vdash x : M}$$

Using the fact that  $\Gamma \Vdash \mathcal{E}$ , we know that  $\mathcal{E} = (\mathcal{E}_0; \text{module } x = m : M_0; \mathcal{E}_1)$  and  $\Gamma \Vdash m : M_0$ . We conclude by applying **D-SUB**.

- if  $m.x \rightarrow_{\Gamma} m'$  where  $m = (\text{struct } st_1; \text{module } x = m'; st_2 \text{ end})$ . We check that we must have a derivation of the following form:

$$\frac{\text{D-COMPNT} \frac{\Gamma \Vdash m : \text{sig } S_1; \text{module } x : M_0; S_2 \text{ restr } \theta \text{ end}}{\Gamma \Vdash m.x : M_0} \quad \vdash M_0 <: M}{\text{D-SUB} \quad \Gamma \Vdash m.x : M}$$

and:

$$\frac{\text{D-STRUCT+} \dots \frac{\dots \quad \Gamma \Vdash m' : M' \quad \vdash M' <: M'_0}{\Gamma \Vdash m : \text{sig } S'_1; \text{module } x : M'_0; S'_2 \text{ restr } \theta' \text{ end}} \quad \vdash M'_0 <: M_0 \quad \dots}{\text{D-SUB} \quad \Gamma \Vdash m : \text{sig } S_1; \text{module } x : M_0; S_2 \text{ restr } \theta \text{ end}}$$

Hence we have a derivation of  $\Gamma \Vdash m' : M'$ . We conclude by sub-typing, using the fact that we can derive  $\vdash M' <: M$ .

- if  $m.x \rightarrow_{\Gamma} m'.x$  where  $m \rightarrow_{\Gamma} m'$ , we know that our derivation is of the form:

$$\frac{\text{D-COMPNT} \frac{\Gamma \Vdash m : \text{sig } S_1; \text{module } x : M_0; S_2 \text{ restr } \theta \text{ end}}{\Gamma \Vdash m.x : M_0} \quad \vdash M_0 <: M}{\text{D-SUB} \quad \Gamma \Vdash m.x : M}$$

By induction hypothesis, we have a derivation of:

$$\Gamma \Vdash m' : \text{sig } S_1; \text{module } x : M_0; S_2 \text{ restr } \theta \text{ end}$$

We conclude immediately using **D-COMPNT** and **D-SUB**.

- the case where  $m(p) \rightarrow_{\Gamma} m'(p)$  with  $m \rightarrow_{\Gamma} m'$  is the same.

- if  $m(p) \rightarrow_{\Gamma} m_0[x \mapsto p]$  where  $m = \text{func}(x : M_p) m_0$ , we have a derivation of the form:

$$\text{D-FUNCAPP} \frac{\Gamma \Vdash m : \text{func}(x : M'_p) M' \quad \Gamma \vdash p : M'_p}{\Gamma \Vdash m(p) : M' [x \mapsto \lambda_m]} \vdash M' [x \mapsto \lambda_m] <: M$$

$$\text{D-SUB} \frac{\Gamma \Vdash m(p) : M}{\Gamma \Vdash m(p) : M}$$

where  $\lambda_m = \text{mem}_{\Gamma}(p')$ . Since  $m = \text{func}(x : M_p) m_0$ , we must have a derivation of the form:

$$\text{D-FUNC} \frac{\Gamma \vdash M_p \quad \Gamma(x) \not\downarrow_{\text{undef}} \quad \Gamma, \text{module } x = \text{abs}_{\text{param}} : M_p \Vdash m_0 : M''}{\Gamma \Vdash \text{func}(x : M_p) m_0 : \text{func}(x : M_p) M''} \vdash M'_p <: M_p \quad \vdash M'' <: M'$$

$$\text{D-SUB} \frac{\Gamma \Vdash \text{func}(x : M_p) m_0 : \text{func}(x : M_p) M''}{\Gamma \Vdash \text{func}(x : M_p) m_0 : \text{func}(x : M'_p) M'}$$

Since  $\Gamma \vdash p : M'_p$  and  $\vdash M'_p <: M_p$ , we know that we have a derivation of  $\Gamma \vdash p : M_p$ . Since:

$$\Gamma, \text{module } x = \text{abs}_{\text{param}} : M_p \Vdash m_0 : M''$$

is derivable, we apply Lemma C.12 to get a derivation of:

$$\Gamma \Vdash m_0[x \mapsto p] : M'' [x \mapsto \lambda_m]$$

Using Lemma C.9, we know that since  $\vdash M'' <: M'$ , we have a derivation of:

$$\Gamma \Vdash m_0[x \mapsto p] : M' [x \mapsto \lambda_m]$$

- the case where  $\bar{m}(p) \rightarrow_{\Gamma} (\text{abs}_K x)(\vec{p}_0, p)$  and  $\bar{m} = (\text{abs}_K x)(\vec{p}_0)$  is immediate.  $\square$

We recall and prove Lemma C.1:

LEMMA (SUBJECT REDUCTION). *If  $\Gamma \Vdash \mathcal{E}$  and  $\Gamma \Vdash m : M$  then  $\Gamma \Vdash \text{res}_{\mathcal{E}}(m) : M$  whenever  $\text{res}_{\mathcal{E}}(m)$  is well-defined.*

PROOF. From Proposition C.8, we know that  $\text{res}_{\mathcal{E}}(m) = m \downarrow_{\mathcal{E}}$ . From Lemma C.13, we know that the type of the module expression is preserved. This concludes this proof.  $\square$

## D SOUNDNESS OF OUR HOARE LOGIC FOR COST

We now prove the soundness of our Hoare logic rules. We recall Theorem 5.1.

THEOREM. *The proof rules in Figures 12, 11 and 13 are sound.*

We focus on the two most complicated rules of our system, **ABS** (Section D.1) and **INSTANTIATION** (Section D.2). The soundness of the remaining rules is straightforward to show (we omit the proof).

First, some notation. We extend the program semantics  $\llbracket \_ \rrbracket_v^{\mathcal{E}, \rho}$  to function paths, by letting  $\llbracket F \rrbracket_v^{\mathcal{E}, \rho}$  be the execution of  $F$ 's body without the procedure's arguments evaluation phase:

$$\llbracket F \rrbracket_v^{\mathcal{E}, \rho} = \mathbb{E}_{(v', c') \sim \llbracket s \rrbracket_v^{\mathcal{E}, \rho}} \left[ \mathbb{1}_{v' [x \leftarrow (r)_{v'}, c' + c_{\mathcal{E}}(r, v')]} \right]$$

$$\text{(if } f\text{-res}_{\mathcal{E}}(F) = (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \_ ; \text{return } r \}) \text{)}$$

$$\llbracket F \rrbracket_v^{\mathcal{E}, \rho} = \llbracket \rho(x)(\vec{p}) . f \rrbracket_v^{\mathcal{E}, \rho} \quad \text{(if } f\text{-res}_{\mathcal{E}}(F) = (\text{abs}_{\text{open}} x)(\vec{p}) . f \text{)}$$

The cost of  $F$  is defined as expected, by taking the maximum over all memories of the support of the second marginal of  $\llbracket F \rrbracket_v^{\mathcal{E}, \rho}$ :

$$\text{cost}_v^{\mathcal{E}, \rho}(F) = \sup(\text{supp}(\pi_2^{\#}(\llbracket F \rrbracket_v^{\mathcal{E}, \rho})))$$

Finally, cost judgment  $\mathcal{E} \vdash \{\phi\} F \{\psi \mid t\}$  on function path have the same semantics than cost judgment on statements, i.e. for every well-typed environment  $\mathcal{E}$ , for every  $\mathcal{E}$ -interpretation  $\rho$  and  $v \in \phi$ :

$$\text{supp}(\pi_1^\#(\llbracket F \rrbracket_v^{\mathcal{E}, \rho})) \subseteq \psi \quad \wedge \quad \text{cost}_v^{\mathcal{E}, \rho}(F) \leq t[\text{conc}] + \sum_{A \in \text{abs}(\mathcal{E})} t[A.f] \cdot \text{compl}_{A.f}^{\mathcal{E}, \rho} \quad (5)$$

### D.1 Abstract Call Rule Soundness

We now prove that the abstract call rule **ABS** in Fig. 12 is sound.

First, remark that since the rule **ABS** uses a different upper-bound  $t_{H.g}(k)$  on the cost of the  $k$ -th call to the oracle  $H.g$ , and since the fact that we are in the  $k$ -th call is characterized by the invariant  $I$ , we must prove both properties (the invariant on the memories, and the upper-bound on the complexity of the call) simultaneously. The proofs is by induction on the size in  $\mathcal{E}$  of the procedure  $\rho(x).f$  called, where procedure calls are inlined. Formally, for every environment  $\mathcal{E}$  and  $\mathcal{E}$ -pre-interpretation  $\rho$ , we define:

$$\begin{aligned} \# \text{size}_\rho^{\mathcal{E}}(\mathbf{abort}) &= 1 \\ \# \text{size}_\rho^{\mathcal{E}}(\mathbf{skip}) &= 1 \\ \# \text{size}_\rho^{\mathcal{E}}(s_1; s_2) &= 1 + \# \text{size}_\rho^{\mathcal{E}}(s_1) + \# \text{size}_\rho^{\mathcal{E}}(s_2) \\ \# \text{size}_\rho^{\mathcal{E}}(x \leftarrow e) &= 1 = 1 \\ \# \text{size}_\rho^{\mathcal{E}}(x \xleftarrow{\$} d) &= 1 = 1 \\ \# \text{size}_\rho^{\mathcal{E}}(x \leftarrow \mathbf{call} F(\vec{e})) &= 1 + \# \text{size}_\rho^{\mathcal{E}}(s) \quad (\text{if } f\text{-res}_{\mathcal{E}}(F) = \text{proc } f \_ = \{ \_ ; s ; \_ \}) \\ \# \text{size}_\rho^{\mathcal{E}}(x \leftarrow \mathbf{call} F(\vec{e})) &= 1 + \# \text{size}_\rho^{\mathcal{E}}(x \leftarrow \mathbf{call} \rho(x)(\vec{p}).f(\vec{e})) \\ &\quad (\text{if } f\text{-res}_{\mathcal{E}}(F) = (\mathbf{abs}_{\text{open}} x)(\vec{p}).f) \\ \# \text{size}_\rho^{\mathcal{E}}(\mathbf{if } e \mathbf{ then } s_1 \mathbf{ else } s_2) &= 1 + \# \text{size}_\rho^{\mathcal{E}}(s_1) + \# \text{size}_\rho^{\mathcal{E}}(s_2) \\ \# \text{size}_\rho^{\mathcal{E}}(\mathbf{while } e \mathbf{ do } s) &= 1 + \# \text{size}_\rho^{\mathcal{E}}(s) \end{aligned}$$

Note that we do not care about the size of the expressions appearing in the statement.

We now define and prove our (generalized) induction property, which we need to prove the soundness of **ABS** after.

**LEMMA D.1.** *Let  $I$  a formula,  $\mathcal{E}$  a well-typed environment,  $\rho$  an  $\mathcal{E}$ -interpretation and  $\vec{x}$  be functor parameters with module types  $\mathbf{abs}_{\text{open}} \vec{M}$  which are module structure signatures (i.e. not functors).*

*For every statement  $s$  well-typed in  $\mathcal{E}$  with additional functor parameters  $\mathbf{abs}_{\text{param}} \vec{x} : \vec{M}$ , and satisfying a memory restriction  $\lambda_m^s$ :*

$$\Gamma \vdash s \quad \Gamma \vdash s \triangleright \lambda_m^s \quad \text{where} \quad \Gamma = \mathcal{E}, \text{ module } \vec{x} = \mathbf{abs}_{\text{param}} : \vec{M}$$

*Then for every valuation  $\vec{p}$  of the parameters  $\vec{x}$  well-typed in  $\mathcal{E}$  (i.e.  $\mathcal{E} \vdash \vec{p} : \vec{M}$ ), if we let  $\mathcal{E}' = \mathcal{E}, \text{ module } \vec{x} = \mathbf{abs}_{\text{open}} : \vec{M}$ , and  $\vec{O}$  be an enumeration of the parameters  $\vec{x}$ 's procedures, i.e. of:*

$$\{y.h \mid y \in \vec{x} \wedge h \in \text{procs}(\vec{M}[y])\}$$

*Then if:*

- *the memory of the statement  $s$  is independent of  $I$ , except for calls to the parameters  $\vec{x}$ , i.e.:*

$$\lambda_m^s \cap \text{FV}(I) = \emptyset$$

- for every parameters' procedure  $H.g \in \text{procs}_\Gamma(\vec{x})$ , the  $k$ -th call to  $H.g$  (when  $H$  is instantiated by  $\vec{p}[H]$ ) preserves the invariant and has a cost upper-bounded by  $t_{H.g}[k]$  (for  $k \leq \lambda_c[H.g]$ ):

$$\forall \vec{k} \leq \lambda_c[\vec{O}], \vec{k}[H.g] < \lambda_c[H.g] \rightarrow \mathcal{E} \vdash \{I \vec{k}\} \vec{p}[H].f \{I (\vec{k} + \mathbb{1}_{H.g}) \mid t_{H.g}(\vec{k}[H.g])\} \quad (6)$$

- the number of calls from  $s$  to the parameters' procedure is upper-bounded by  $\lambda_c[\vec{O}] - \vec{k}$  (for memory satisfying the invariant  $I \vec{k}$ ):

$$\forall \vec{k} \leq \lambda_c[\vec{O}], \forall v \in I \vec{k}, \forall H.g \in \text{procs}_{\mathcal{E}}(\vec{x}), \#\text{calls}_{H.g,v}^{\mathcal{E}',\rho'}(s) \leq \lambda_c[H.g] - \vec{k}[H.g] \quad (7)$$

For every  $H.g \in \text{procs}_{\mathcal{E}}(\vec{x})$  and  $j < \lambda_c[H.g]$ , we define:

$$\text{call-cost}_{H.g}^{\mathcal{E},\rho}(j) = t_{H.g}(j)[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}) \\ f \in \text{procs}_{\mathcal{E}}(A)}} t_{H.g}(j)[A.f] \cdot \text{compl}_{A.f}^{\mathcal{E},\rho}$$

Let  $\rho' = \rho, (\vec{x} : \vec{p})$ . We have that for every  $\vec{k} \leq \lambda_c[\vec{O}]$  and  $v \in I \vec{k}$ :

$$\text{supp}(\pi_1^\#(\llbracket s \rrbracket_v^{\mathcal{E}',\rho'})) \subseteq I (\vec{k} + \#\text{calls}_{O,v}^{\mathcal{E}',\rho'}(s)) \quad (8)$$

and:

$$\text{cost}_v^{\mathcal{E}',\rho'}(s) \leq \text{i-cost}_v^{\mathcal{E}',\rho',\vec{x}}(s) + \sum_{H.g \in \text{procs}_{\mathcal{E}}(\vec{x})} \sum_{j=\vec{k}[H.g]}^{\vec{k}[H.g] + \#\text{calls}_{H.g,v}^{\mathcal{E}',\rho'}(s) - 1} \text{call-cost}_{H.g}^{\mathcal{E},\rho}(j) \quad (9)$$

That is, the cost of  $s$  executed in  $\mathcal{E}', \rho'$  and memory  $v$  (satisfying the invariant  $I \vec{k}$ ) is upper-bounded by the intrinsic cost of  $s$  with functors parameters  $\vec{x}$ , of the sum over functor parameters' procedures  $H.g \in \text{procs}_{\mathcal{E}}(\vec{x})$  of the sum over all calls to  $H.g$  (which ranges from  $\vec{k}[H.g]$  to  $\#\text{calls}_{H.g,v}^{\mathcal{E}',\rho'}(s) - 1$ ) of the concrete cost of (the  $j$ -th call to)  $H.g$  (upper-bounded by  $t_{H.g}(j)[\text{conc}]$ ), plus the sum, over all abstract procedures  $A.f$  (in the original environment  $\mathcal{E}$ ), of the number of times the  $j$ -th call to  $H.g$  called  $A.f$  (upper-bounded by  $t_{H.g}(j)[A.f]$ ) times the maximal cost of  $A.f$  (which is  $\text{compl}_{A.f}^{\mathcal{E},\rho}$ ).

PROOF. We prove this induction over  $\#\text{size}_\rho^{\mathcal{E}}(s)$ . We do a case analysis on  $s$ :

- if  $s$  is **abort** or **skip**, this is immediate.
- if  $s = s_1; s_2$ . Let  $\vec{k} \leq \lambda_c[\vec{O}]$  and  $v \in I \vec{k}$ . We know that:

$$\text{supp}(\pi_1^\#(\llbracket s_1; s_2 \rrbracket_v^{\mathcal{E}',\rho'})) = \bigcup_{v' \in \text{supp}(\pi_1^\#(\llbracket s_1 \rrbracket_v^{\mathcal{E}',\rho'}))} \text{supp}(\pi_1^\#(\llbracket s_2 \rrbracket_{v'}^{\mathcal{E}',\rho'}))$$

Let  $\vec{k}_1 = \vec{k} + \#\text{calls}_{O,v}^{\mathcal{E}',\rho'}(s_1)$ . By induction hypothesis applied on  $s_1$ ,

$$\text{supp}(\pi_1^\#(\llbracket s_1 \rrbracket_v^{\mathcal{E}',\rho'})) \subseteq I \vec{k}_1 \quad (10)$$

Let  $v' \in \text{supp}(\pi_1^\#(\llbracket s_1 \rrbracket_v^{\mathcal{E}',\rho'}))$ , we know that  $\vec{k}_1 \leq \lambda_c[\vec{O}]$ . Hence, by induction hypothesis on  $s_2$ , we deduce that:

$$\text{supp}(\pi_1^\#(\llbracket s_2 \rrbracket_{v'}^{\mathcal{E}',\rho'})) \subseteq I \vec{k}_2$$

where  $\vec{k}_2(v') = \vec{k} + \#\text{calls}_{O,v}^{\mathcal{E}',\rho'}(s_1) + \#\text{calls}_{O,v'}^{\mathcal{E}',\rho'}(s_2)$ . Since:

$$\#\text{calls}_{O,v}^{\mathcal{E}',\rho'}(s_1) + \#\text{calls}_{O,v'}^{\mathcal{E}',\rho'}(s_2) \leq \#\text{calls}_{O,v}^{\mathcal{E}',\rho'}(s_1; s_2) \quad (11)$$

we deduce:

$$\text{supp}(\pi_1^\#(\llbracket s_2 \rrbracket_{v'}^{\mathcal{E}', \rho'})) \subseteq I(\vec{k} + \# \text{calls}_{\vec{O}, v}^{\mathcal{E}', \rho'}(s_1; s_2))$$

which concludes the proofs of the first point. It remains to prove that the complexity of  $s_1; s_2$  is upper-bounded by the wanted quantity. First, we have:

$$\text{cost}_v^{\mathcal{E}', \rho'}(s_1; s_2) = \text{cost}_v^{\mathcal{E}', \rho'}(s_1) + \max_{v' \in \text{supp}(\pi_1^\#(\llbracket s_1 \rrbracket_{v'}^{\mathcal{E}', \rho'}))} \text{cost}_{v'}^{\mathcal{E}', \rho'}(s_2)$$

By applying the induction hypothesis on  $s_1$  and  $s_2$  with, respectively,  $\vec{k}$  and  $\vec{k}_1$ , we get:

$$\text{cost}_v^{\mathcal{E}', \rho'}(s_1) \leq \text{i-cost}_v^{\mathcal{E}', \rho', \vec{x}}(s_1) + \sum_{H.g \in \text{procs}_{\mathcal{E}}(\vec{x})} \sum_{j=\vec{k}[H.g]}^{\vec{k}_1[H.g]-1} \text{call-cost}_{H.g}^{\mathcal{E}, \rho}(j)$$

and:

$$\text{cost}_{v'}^{\mathcal{E}', \rho'}(s_2) \leq \text{i-cost}_{v'}^{\mathcal{E}', \rho', \vec{x}}(s_2) + \sum_{H.g \in \text{procs}_{\mathcal{E}}(\vec{x})} \sum_{j=\vec{k}_1[H.g]}^{\vec{k}_2(v')[H.g]-1} \text{call-cost}_{H.g}^{\mathcal{E}, \rho}(j)$$

Since  $\text{i-cost}_v^{\mathcal{E}', \rho'}(s_1; s_2)$  is equal to:

$$\text{i-cost}_v^{\mathcal{E}', \rho'}(s_1) + \max_{v' \in \text{supp}(\pi_1^\#(\llbracket s_1 \rrbracket_{v'}^{\mathcal{E}', \rho'}))} \text{i-cost}_{v'}^{\mathcal{E}', \rho'}(s_2)$$

and using Equ. (11), we deduce that:

$$\text{cost}_v^{\mathcal{E}', \rho'}(s_1) + \text{cost}_{v'}^{\mathcal{E}', \rho'}(s_2) \leq \text{i-cost}_v^{\mathcal{E}', \rho'}(s_1; s_2) + \sum_{H.g \in \text{procs}_{\mathcal{E}}(\vec{x})} \sum_{j=\vec{k}[H.g]}^{\vec{k}[H.g] + \# \text{calls}_{\vec{O}, v}^{\mathcal{E}', \rho'}(s_1; s_2) - 1} \text{call-cost}_{H.g}^{\mathcal{E}, \rho}(j)$$

We conclude by taking the max over  $v'$  and using Equ. 10.

- if  $s = x \leftarrow e$ . Let  $\vec{k} \leq \lambda_{\vec{O}}(\vec{O})$  and  $v \in I \vec{k}$ . Since  $\text{mem}_{\Gamma}(s) \cap \text{FV}(I) = \emptyset$ , we know that  $x \notin \text{FV}(I)$ . Hence the invariant is preserved, i.e.:

$$\forall v' \in \text{supp}(\pi_1^\#(\llbracket s \rrbracket_{v'}^{\mathcal{E}', \rho'})), v' \in I \vec{k}$$

Since  $\# \text{calls}_{\vec{O}, v}^{\mathcal{E}', \rho'}(s) = 0$ , this proves the first point. Moreover:

$$\text{cost}_v^{\mathcal{E}', \rho'}(x \leftarrow e) = \text{i-cost}_v^{\mathcal{E}', \rho', \vec{x}}(x \leftarrow e)$$

which concludes the proof.

- the random assignment  $x \stackrel{\$}{\leftarrow} d$ , conditional **if**  $e$  **then**  $s_1$  **else**  $s_2$  and while loop **while**  $e$  **do**  $s$  cases are similar. We omit the details.
- if  $s = x \leftarrow \text{call } F(\vec{e})$  and  $\text{f-res}_{\mathcal{E}}(F) = (\text{proc } f(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s'; \text{return } e \})$ , then we proceed as in the previous case, applying the induction hypothesis on  $s'$ . First, we remark that  $s'$  is smaller than  $s$ , since:

$$\# \text{size}_{\rho}^{\mathcal{E}}(s') < \# \text{size}_{\rho}^{\mathcal{E}}(x \leftarrow \text{call } F(\vec{e}))$$

It only remains to check that the induction hypothesis's hypotheses hold. The last two hypotheses are straightforward to show. It only remains to prove that  $\Gamma \vdash s' \triangleright \lambda_m^s$  is derivable. Since  $\Gamma \vdash s \triangleright \lambda_m^s$ , we know that  $\text{mem}_{\Gamma}(s) \subseteq \lambda_m^s$ , where:

$$\text{mem}_{\Gamma}(s) = \{x\} \sqcup \text{mem}_{\Gamma}(s') \sqcup \text{vars}(\vec{e})$$

Which concludes this case.

- idem if  $s = x \leftarrow \mathbf{call} F(\vec{e})$  and  $f\text{-res}_{\mathcal{E}}(F) = (\mathbf{abs}_{\text{open}} H)(\_) . f$  with  $H \notin \vec{x}$ .
- if  $s = x \leftarrow \mathbf{call} F(\vec{e})$  and  $f\text{-res}_{\mathcal{E}}(F) = (\mathbf{abs}_{\text{open}} H)(\vec{p}') . g$  with  $H \in \vec{x}$ , then we use the hypothesis that procedures of the interpretation of the modules parameters  $\vec{x}$  preserves the invariant. First, since  $\vec{x}$  have types  $\vec{M}$  and  $\vec{M}$  are module structure signatures, we know that  $\vec{p}'$  is empty (from the fact that  $s$  is well-typed in  $\Gamma$ ). Hence  $f\text{-res}_{\mathcal{E}}(F) = \mathbf{abs}_{\text{open}} H . g$ .

Let  $\vec{k} \leq \lambda_c[\vec{O}]$  and  $v \in I \vec{k}$ . First, using the hypothesis in Eq. (7), we know that:

$$\vec{k}[H.g] + \#\text{calls}_{H.g,v}^{\mathcal{E}',\rho'}(x \leftarrow \mathbf{call} F(\vec{e})) \leq \lambda_c[H.g]$$

Hence  $\vec{k}[H.g] < \lambda_c[H.g]$ . By applying Equ. (6) on  $H.g$ , we know that:

$$\mathcal{E} \vdash \{I \vec{k}\} \vec{p}[H].g \{I(\vec{k} + \mathbb{1}_{H.g}) \mid t_{H.g}(k)\} \quad (12)$$

Since  $\rho$  is an  $\mathcal{E}$ -interpretation, and since  $\vec{x}$  have type  $\vec{M}$  where  $\vec{M}$  are module structure signatures, we can check that  $\rho'$  is an  $\mathcal{E}'$ -interpretation. Hence, from Equ. (12) and the semantics of the cost judgment given in Equ. (5) we get that:

$$\text{supp}(\pi_1^\#(\llbracket \vec{p}[H].g \rrbracket_v^{\mathcal{E},\rho})) \subseteq I(\vec{k} + \mathbb{1}_{H.g}) \quad (13)$$

and:

$$\begin{aligned} & \text{cost}_v^{\mathcal{E},\rho}(\vec{p}[H].g) \\ & \leq t_{H.g}(\vec{k}[H.g])[\text{conc}] \\ & \quad + \sum_{\substack{A \in \text{abs}(\mathcal{E}) \\ f \in \text{procs}_{\mathcal{E}}(A)}} t_{H.g}(\vec{k}[H.g])[A.f] \cdot \text{compl}_{A.f}^{\mathcal{E},\rho} \\ & \leq \text{call-cost}_{H.g}^{\mathcal{E},\rho}(\vec{k}[H.g]) \end{aligned} \quad (14)$$

Observe that since  $\rho' = \rho, (\vec{x} : \vec{p})$ , we have:

$$\llbracket x \leftarrow \mathbf{call} F(\vec{e}) \rrbracket_v^{\mathcal{E}',\rho'} = \llbracket x \leftarrow \mathbf{call} \vec{p}[H].g(\vec{e}) \rrbracket_v^{\mathcal{E}',\rho'}$$

Hence Equ. (8) follows directly from Equ. (13). It remains to show Equ. (9). First, note that:

$$\text{cost}_v^{\mathcal{E}',\rho'}(x \leftarrow \mathbf{call} F(\vec{e})) = i\text{-cost}_v^{\mathcal{E}',\rho',\vec{x}}(x \leftarrow \mathbf{call} F(\vec{e})) + \text{cost}_v^{\mathcal{E}',\rho'}(\vec{p}[H].g)$$

Since  $\vec{p}$  is well-typed in  $\mathcal{E}$ , we know that  $\text{cost}_v^{\mathcal{E}',\rho'}(\vec{p}[H].g) = \text{cost}_v^{\mathcal{E},\rho}(\vec{p}[H].g)$ . Hence:

$$\text{cost}_v^{\mathcal{E}',\rho'}(x \leftarrow \mathbf{call} F(\vec{e})) = i\text{-cost}_v^{\mathcal{E}',\rho',\vec{x}}(x \leftarrow \mathbf{call} F(\vec{e})) + \text{cost}_v^{\mathcal{E},\rho}(\vec{p}[H].g)$$

We conclude the proof of Equ. (9) using the inequality above and Equ. (14).  $\square$

LEMMA D.2. *The rule ABS given in Fig. 12 is sound.*

PROOF. We just apply Lemma D.1 on  $\rho(x).f$ . The first two hypotheses of the lemma hold thanks to the premises of the ABS rule, and using the fact our module system has the subject reduction property (Lemma C.1). The third hypothesis follows from the fact that  $\rho$  is an  $\mathcal{E}$ -interpretation.  $\square$

## D.2 Instantiation Rule Soundness

We prove the following technical lemma, which allows to extend an  $\mathcal{E}$ -interpretation  $\rho$  into an  $(\mathcal{E}, \text{module } x = \text{abs}_{\text{open}} : M_1)$ -interpretation  $\rho' = \rho, (x \mapsto m)$ . This is possible whenever:

- i)  $m$  has type  $\text{erase}_{\text{compl}}(M_1)$ ;
- ii) and we can show that  $m$  verifies  $M_1$ 's complexity restriction by proving that:

$$\forall f \in \text{procs}(M_1), \quad \mathcal{E}, \text{module } \vec{x} = \text{abs}_{\text{open}} : \vec{M} \vdash \{\top\} m(\vec{x}).f \{\top \mid t_f\}$$

where  $\vec{x}$  are  $M_1$ 's functor parameters, and  $\vec{M}$  their types.

LEMMA D.3. *Let  $\mathcal{E}$  be a well-typed environment,  $M_1$  be low-order module signature s.t.:*

$$M_1 = \text{func}(\vec{x} : \vec{M}) \text{ sig\_restr } \theta \text{ end}$$

and  $m$  be a module expression s.t.  $\mathcal{E} \vdash_x m : \text{erase}_{\text{compl}}(M_1)$ . Let  $\mathcal{E}_a = \mathcal{E}, \text{module } x = \text{abs}_{\text{open}} : M_1$  and  $\rho$  be an  $\mathcal{E}$ -interpretation. If, for every  $f \in \text{procs}(M_1)$ , we have:

$$\mathcal{E}, \text{module } \vec{x} = \text{abs}_{\text{open}} : \vec{M} \vdash \{\top\} m(\vec{x}).f \{\top \mid t_f\} \wedge t_f \leq_{\text{compl}} \theta[f] \quad (15)$$

then  $\rho_a = \rho, (x \mapsto m)$  is an  $\mathcal{E}_a$ -interpretation. Moreover, for any  $f \in \text{procs}(M_1)$ :

$$\text{compl}_{x.f}^{\mathcal{E}_a, \rho_a} \leq t_f[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}) \\ h \in \text{procs}_{\mathcal{E}}(A)}} t_f[A.h] \cdot \text{compl}_{A.h}^{\mathcal{E}, \rho} \quad (16)$$

PROOF. Let  $\vec{m}$  be an evaluation of  $x$ 's parameters s.t. for every  $1 \leq i \leq |\vec{x}|$ , if we let  $x_i = \vec{x}[i]$ ,  $m_i = \vec{m}[i]$  and  $M_i = \vec{M}[i] = \text{sig\_restr } \lambda_c^i \wedge \_ \text{end}$ :

$$\mathcal{E} \vdash (\text{module } x_i = m_i : \text{erase}_{\text{compl}}(M_i))$$

and:

$$\forall g \in \text{procs}(M_i), \text{cost}_v^{\mathcal{E}}(m_i.g) \leq \lambda_c^i$$

Let  $\mathcal{E}'_a = \mathcal{E}_a; \text{module } \vec{x} = \text{abs}_{\text{open}} : \vec{M}$  and  $\rho'_a = \rho_a, (\vec{x} : \vec{m})$ . Also, let  $y \in \vec{x}$  and  $g \in \text{procs}(\vec{M}[y])$ . To prove that  $\rho_a$  is an  $\mathcal{E}_a$ -interpretation, we only need to prove that:

$$\#\text{calls}_{y.g}^{\mathcal{E}'_a, \rho'_a}(x(\vec{x}).f) \leq \theta[f][y.g] \quad (17)$$

and that:

$$i\text{-cost}^{\mathcal{E}'_a, \rho'_a, \vec{x}}(x(\vec{x}).f) \leq \theta[f][\text{intr}] \quad (18)$$

*Restriction to module structures.* First, note that w.l.o.g. we can assume that  $\vec{m}$  are all module structure (i.e. of the form  $\text{struct\_end}$ ). Basically, we show that we can build another interpretation  $\rho''_a$  extending  $\rho_a$  which satisfies the same hypotheses than  $\rho'_a$ , and such that for every  $z \in \vec{x}$ ,  $\rho''_a(z)$  is a module structure such that, for every  $g \in \text{procs}(\vec{M}[z])$ :

$$\#\text{calls}_{z.g}^{\mathcal{E}'_a, \rho'_a}(x(\vec{x}).f) = \#\text{calls}_{z.g}^{\mathcal{E}'_a, \rho''_a}(x(\vec{x}).f)$$

and:

$$i\text{-cost}^{\mathcal{E}'_a, \rho'_a, \vec{x}}(x(\vec{x}).f) = i\text{-cost}^{\mathcal{E}'_a, \rho''_a, \vec{x}}(x(\vec{x}).f)$$

Indeed, assume that there exists some  $y \in \vec{x}$  such that  $\vec{m}[y]$  is not a module structure in  $\rho'_a$ . If  $\vec{m}[y]$  is a module path  $p$  in  $\rho'_a$ , then we resolve it in  $\mathcal{E}, \rho$  (which is always possible, since  $\vec{m}[y]$  is well-typed in  $\mathcal{E}$  and  $\rho$  is an  $\mathcal{E}$ -interpretation) until we get a module structure  $\text{struct st end}$ , and replace  $y$  by  $\text{struct st end}$  in  $\rho'_a$ . Finally,  $\vec{m}[y]$  cannot be a functor (by typing hypothesis) in  $\rho'_a$ . We repeat the steps above until  $(\rho'_a)[\vec{x}]$  are all module structures.

*Proof of Equ. (17).* Since  $\rho'_a(x) = m$  and  $m$  is well-typed in  $\mathcal{E}$ , and since the module expressions  $\vec{m}$  are well-typed in  $\mathcal{E}$ , we can removed  $x$  from the environment and the interpretation while keeping the semantics unchanged. That is, we have:

$$\#calls_{y.g}^{\mathcal{E}', \rho'_a}(x(\vec{x}).f) = \#calls_{y.g}^{\mathcal{E}', \rho'}(m(\vec{x}).f) \quad (19)$$

where:

$$\mathcal{E}' = \mathcal{E}; \text{ module } \vec{x} = \text{abs}_{\text{open}} : \vec{M} \quad \text{and} \quad \rho' = \rho, (\vec{x} : \vec{m})$$

Since  $\mathcal{E}$  is well-typed, and  $\vec{m}$  have types  $\vec{M}$  in  $\mathcal{E}$ , and since  $\vec{M}$  is not a functor type, we can check that  $\mathcal{E}'$  is well-typed, and  $\rho'$  is an  $\mathcal{E}'$ -interpretation. Using Equ. (15), we get:

$$\forall v, \text{cost}_v^{\mathcal{E}', \rho'}(m(\vec{x}).f) \leq t_f[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}') \\ h \in \text{procs}_{\mathcal{E}'}(A)}} t_f[A.h] \cdot \text{compl}_{A.h}^{\mathcal{E}', \rho'} \quad (20)$$

Let  $N > 0$  be an non-zero positive integer. We are going to change the interpretation of  $y$  in  $\rho'$  by adding some code doing nothing and taking time  $N$ . Let  $\text{st}_y$  be such that  $\rho'(y) = \text{struct } \text{st}_y \text{ end}$ . By typing hypothesis, we know that  $\text{st}_y$  is of the shape:

$$\text{st}_y = \text{st}_1; \text{ proc } g(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); s_g; \text{return } e_g \}; \text{st}_2$$

Then, we let  $\text{st}_y^N$  be the module structure:

$$\text{st}_1; \text{ proc } g(\vec{v} : \vec{\tau}) \rightarrow \tau_r = \{ \text{var } (\vec{v}_1 : \vec{\tau}_1); (\text{tic}^N; s_g); \text{return } e_g \}; \text{st}_2$$

and let  $\rho'_N$  be the interpretation with the same domain as  $\rho'$  s.t.:

$$\forall w \in \text{dom}(\rho'), \rho'_N(w) = \begin{cases} \rho'(w) & \text{if } w \neq y \\ \text{struct } \text{st}_y^N \text{ end} & \text{if } w = y \end{cases}$$

Let  $v$  be some arbitrary memory. Since  $\rho'$  is an  $\mathcal{E}'$ -interpretation, then so is  $\rho'_N$ . Using Equ. (15), we get:

$$\text{cost}_v^{\mathcal{E}', \rho'_N}(m(\vec{x}).f) \leq t_f[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}') \\ h \in \text{procs}_{\mathcal{E}'}(A)}} t_f[A.h] \cdot \text{compl}_{A.h}^{\mathcal{E}', \rho'_N} \quad (21)$$

Moreover, we can easily check that:

$$\#calls_{y.g}^{\mathcal{E}', \rho'}(m(\vec{x}).f) = \#calls_{y.g}^{\mathcal{E}', \rho'_N}(m(\vec{x}).f)$$

and:

$$\text{cost}_v^{\mathcal{E}', \rho'_N}(m(\vec{x}).f) = \text{cost}_v^{\mathcal{E}', \rho'}(m(\vec{x}).f) + N \cdot \#calls_{y.g}^{\mathcal{E}', \rho'}(m(\vec{x}).f)$$

From Equ. (21), we have:

$$\text{cost}_v^{\mathcal{E}', \rho'}(m(\vec{x}).f) + N \cdot \#calls_{y.g}^{\mathcal{E}', \rho'}(m(\vec{x}).f) \leq t_f[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}') \\ h \in \text{procs}_{\mathcal{E}'}(A)}} t_f[A.h] \cdot \text{compl}_{A.h}^{\mathcal{E}', \rho'_N}$$



Using the inequality above, and Equ. (20), we have:

$$\begin{aligned}
& N \cdot \#\text{calls}_{y.g}^{\mathcal{E}', \rho'}(m(\vec{x}).f) \\
& \leq \sum_{\substack{A \in \text{abs}(\mathcal{E}') \\ h \in \text{procs}_{\mathcal{E}'}(A)}} t_f[A.h] \cdot (\text{compl}_{A.h}^{\mathcal{E}', \rho'_N} - \text{compl}_{A.h}^{\mathcal{E}', \rho'}) \\
& \leq t_f[y.g] \cdot (\text{compl}_{y.g}^{\mathcal{E}', \rho'_N} - \text{compl}_{y.g}^{\mathcal{E}', \rho'}) \quad (\text{Since } \forall z \neq y, \rho'_N(z) = \rho'(z)) \\
& \leq t_f[y.g] \cdot N
\end{aligned}$$

We deduce that  $\#\text{calls}_{y.g}^{\mathcal{E}', \rho'}(m(\vec{x}).f) \leq t_f[y.g]$ . From Equ. (15), we get that:

$$\#\text{calls}_{y.g}^{\mathcal{E}', \rho'}(m(\vec{x}).f) \leq \theta[f][y.g]$$

This, together with Equ. (19), proves Equ. (17)

Finally, to prove that Equ. (18) and Equ. 16 hold we do the exactly the same reasoning, this time by adding to the interpretation of  $x$  some code doing nothing and taking time  $N$ .  $\square$

We also prove the following weakening lemma for the intrinsic complexity of a procedure.

LEMMA D.4. *For every well-typed environment  $\mathcal{E}$ , if:*

$$\mathcal{E} = \mathcal{E}_1; \mathcal{E}_2 \quad \text{where} \quad \mathcal{E}_1 = \mathcal{E}_0; \text{module } x = \text{abs}_{\text{open}} : M_1$$

and  $\mathcal{E}_2$  contains only module declarations. Then for every  $f \in \text{procs}(M_1)$  and  $\mathcal{E}$ -interpretation  $\rho$ , we have  $\text{compl}_{x.f}^{\mathcal{E}, \rho} = \text{compl}_{x.f}^{\mathcal{E}_1, \rho_1}$  where  $\rho_1$  is the restriction of  $\rho$  to  $\mathcal{E}_1$ 's abstract modules.

PROOF SKETCH. Assume  $M_1 = (\text{func}(\vec{x} : \vec{M}) \text{ sig } \_ \text{ end})$ . Let  $\mathcal{E}' = (\mathcal{E}, \text{module } \vec{x} = \text{abs}_{\text{open}} : \vec{M})$  and  $\mathcal{E}'_1 = (\mathcal{E}_1, \text{module } \vec{x} = \text{abs}_{\text{open}} : \vec{M})$ , we prove that  $\text{compl}_{x.f}^{\mathcal{E}, \rho} \leq \text{compl}_{x.f}^{\mathcal{E}_1, \rho_1}$  and  $\text{compl}_{x.f}^{\mathcal{E}, \rho} \geq \text{compl}_{x.f}^{\mathcal{E}_1, \rho_1}$ .

The latter inequality is straightforward to show, since any  $\mathcal{E}'_1$ -interpretation  $\rho_1$  can be extended into an  $\mathcal{E}'$ -interpretation  $\rho$  that leaves the intrinsic cost of  $x$  unchanged, i.e. such that for any  $v$ :

$$\text{i-cost}_v^{\mathcal{E}', \rho, \vec{x}}(x(\vec{x}).f) = \text{i-cost}_v^{\mathcal{E}'_1, \rho_1, \vec{x}}(x(\vec{x}).f)$$

To prove the former inequality, we show that any  $\mathcal{E}'$ -interpretation  $\rho$  can be transformed into an  $\mathcal{E}'_1$ -interpretation  $\rho'$  such that Equ. (D.2) holds, by inlining all modules of  $\mathcal{E}_2$  in  $\rho$ . We omit the details.  $\square$

LEMMA D.5. *The rule INSTANTIATION given in Fig. 13 is sound.*

PROOF. We consider an instance of the rule INSTANTIATION. We want to prove that:

$$\mathcal{E}, \text{module } x = m : M_1 \vdash \{\phi\} \text{ s } \{\psi \mid T_{\text{ins}}\}$$

Let  $\mathcal{E}_c = (\mathcal{E}, \text{module } x = m : M_1)$  and  $\mathcal{E}_a = (\mathcal{E}, \text{module } x = \text{abs}_{\text{open}} : M_1)$ . We know that:

$$M_1 = \text{func}(\vec{y} : \vec{M}) \text{ sig } S_1 \text{ restr } \theta \text{ end} \quad \mathcal{E} \vdash_x m : \text{erase}_{\text{compl}}(M_1)$$

Let  $\rho$  be an  $\mathcal{E}_c$ -interpretation and  $v \in \phi$ , we need to show that for every memory  $\nu$ :

$$\text{supp}(\pi_1^\#(\llbracket s \rrbracket_v^{\mathcal{E}_c, \rho})) \subseteq \psi \quad (22)$$

$$\text{cost}_v^{\mathcal{E}_c, \rho}(s) \leq T_{\text{ins}}[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}_c) \\ g \in \text{procs}_{\mathcal{E}_c}(A)}} T_{\text{ins}}[A.g] \cdot \text{compl}_{A.g}^{\mathcal{E}_c, \rho} \quad (23)$$

We know that  $M_1 = \text{func}(\vec{\gamma} : \vec{M}) \text{ sig } S_1 \text{ restr } \theta$  end and:

$$\mathcal{E}_a \vdash \{\phi\} \text{ s } \{\psi \mid t_s\} \quad (24)$$

Using Lemma D.3, we know that  $\rho_a = \rho, (x \mapsto m)$  is an  $\mathcal{E}_a$ -interpretation. Hence, using Equ. (24), we deduce that:

$$\text{supp}(\pi_1^\#(\llbracket s \rrbracket_v^{\mathcal{E}_a, \rho_a})) \subseteq \psi \quad (25)$$

$$\text{cost}_v^{\mathcal{E}_a, \rho_a}(s) \leq t_s[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}_a) \\ g \in \text{procs}_{\mathcal{E}_a}(A)}} t_s[A.g] \cdot \text{compl}_{A.g}^{\mathcal{E}_a, \rho_a} \quad (26)$$

Using the fact that  $\rho_a(x) = m$ , we can check (by induction over  $\# \text{size}_{\mathcal{E}_a}^{\rho_a}(s)$ ) that:

$$\llbracket s \rrbracket_v^{\mathcal{E}_a, \rho_a} = \llbracket s \rrbracket_v^{\mathcal{E}_c, \rho} \quad \text{and} \quad \text{cost}_v^{\mathcal{E}_a, \rho_a}(s) = \text{cost}_v^{\mathcal{E}_c, \rho}(s)$$

From the left equality above and Equ. (25), we know that Equ. (22) holds. It remains to prove Equ. (23).

From the right equality above and Equ. (26):

$$\begin{aligned} \text{cost}_v^{\mathcal{E}_c, \rho}(s) &\leq t_s[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}_a) \\ g \in \text{procs}_{\mathcal{E}_a}(A)}} t_s[A.g] \cdot \text{compl}_{A.g}^{\mathcal{E}_a, \rho_a} \\ &\leq t_s[\text{conc}] + \sum_{f \in \text{procs}(S_1)} t_s[x.f] \cdot \text{compl}_{x.f}^{\mathcal{E}_a, \rho_a} \\ &\quad + \sum_{\substack{A \in \text{abs}(\mathcal{E}_c) \\ g \in \text{procs}_{\mathcal{E}_c}(A)}} t_s[A.g] \cdot \text{compl}_{A.g}^{\mathcal{E}_a, \rho_a} \end{aligned}$$

We replace  $\text{compl}_{A.g}^{\mathcal{E}_a, \rho_a}$  by  $\text{compl}_{A.g}^{\mathcal{E}_c, \rho}$  for every  $A \in \text{abs}(\mathcal{E}_c)$  and  $g \in \text{procs}_{\mathcal{E}_c}(A)$  using Lemma D.4:

$$\begin{aligned} \text{cost}_v^{\mathcal{E}_c, \rho}(s) &\leq t_s[\text{conc}] + \sum_{f \in \text{procs}(S_1)} t_s[x.f] \cdot \text{compl}_{x.f}^{\mathcal{E}_a, \rho_a} \\ &\quad + \sum_{\substack{A \in \text{abs}(\mathcal{E}_c) \\ g \in \text{procs}_{\mathcal{E}_c}(A)}} t_s[A.g] \cdot \text{compl}_{A.g}^{\mathcal{E}_c, \rho} \end{aligned} \quad (27)$$

Using Lemma D.3, we upper-bound  $\text{compl}_{x.f}^{\mathcal{E}_a, \rho_a}$  for every  $f \in \text{procs}(S_1)$ :

$$\text{compl}_{x.f}^{\mathcal{E}_a, \rho_a} \leq t_f[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}) \\ g \in \text{procs}_{\mathcal{E}}(A)}} t_f[A.g] \cdot \text{compl}_{A.g}^{\mathcal{E}, \rho}$$

We check that the quantities above are identical when evaluated in  $\mathcal{E}_c$ , hence:

$$\text{compl}_{x.f}^{\mathcal{E}_a, \rho_a} \leq t_f[\text{conc}] + \sum_{\substack{A \in \text{abs}(\mathcal{E}_c) \\ g \in \text{procs}_{\mathcal{E}_c}(A)}} t_f[A.g] \cdot \text{compl}_{A.g}^{\mathcal{E}_c, \rho}$$

Hence, re-organizing the terms in the sum in Equ. (27):

$$\begin{aligned} \text{cost}_v^{\mathcal{E}_c, \rho}(s) &\leq t_s[\text{conc}] + \sum_{f \in \text{procs}(S_1)} t_s[x.f] \cdot t_f[\text{conc}] + \\ &\quad \sum_{A \in \text{abs}(\mathcal{E}_c) \text{ } g \in \text{procs}_{\mathcal{E}_c}(A)} (t_s[A.g] + \sum_{f \in \text{procs}(S_1)} t_s[x.f] \cdot t_f[A.g]) \cdot \text{compl}_{A.g}^{\mathcal{E}_c, \rho} \end{aligned}$$

Which, by definition of  $T_{\text{ins}}$ , is exactly Equ. (23).  $\square$