

# MPRI 2.30: Proofs of Security Protocols

## 1. The CCSA Approach to Computational Security

---

Adrien Koutsos

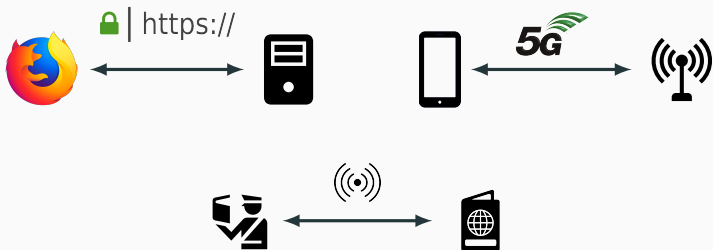
2024/2025

# Introduction

---

## Security Protocols

- **Distributed programs** which aim at providing some security properties.
- Uses **cryptographic primitives**: e.g. encryption.



## Context: Security Properties

There is a large variety of **security properties** that such protocols must provide.



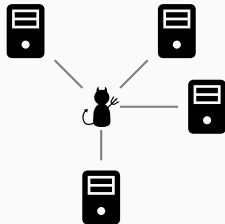
# Context: Attacker Model

Against whom should these properties hold?

- **concretely**, in the **real world**: malicious individuals, corporations, state agencies, ...
- more **abstractly**, one (or many) computers sitting on the network.

## Abstract attacker model

- **Network capabilities**: worst-case scenario: *eavesdrop, block and forge* messages.
- **Computational capabilities**: the adversary's *computational power*.
- **Side-channels capabilities**: observing the agents (e.g. time, power-consumption)  
⇒ not in this lecture.



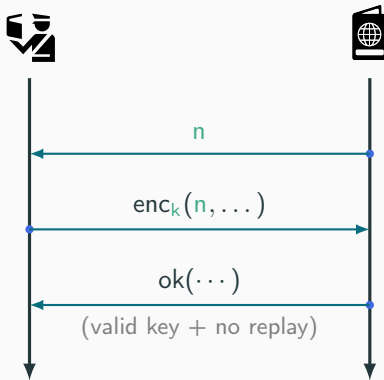
# BAC Protocol (simplified)

The Basic Access Control protocol in **e-passports**:

- uses an RFID tag.
- guard access to information stored.
- should guarantee **data confidentiality** and **user privacy**.

Some security mechanisms:

- **integrity**: obtaining key  $k$  requires **physical access**.
- **no replay**: random nonce  $n$ , old messages cannot be re-used.



# BAC Protocol (simplified)

## Privacy: Unlinkability

No adversary can know whether it interacted with a particular user, **in any context**.

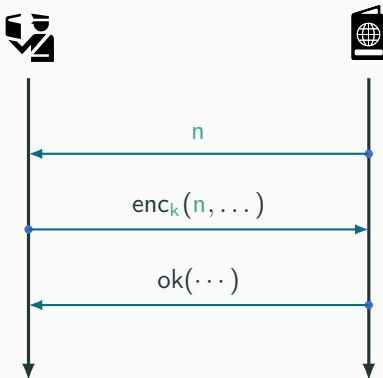
**Example.** For two user sessions:

$$\text{att} \left( \begin{array}{c} \text{[Globe Icon]} \\ \text{[Globe Icon]} \end{array}, \begin{array}{c} \text{[Globe Icon]} \\ \text{[Globe Icon]} \end{array} \right) = \left\{ \begin{array}{l} \begin{array}{c} \text{[Blue Globe Icon]} \\ \text{[Blue Globe Icon]} \end{array}, \begin{array}{c} \text{[Blue Globe Icon]} \\ \text{[Red Globe Icon]} \end{array} ? \\ \begin{array}{c} \text{[Blue Globe Icon]} \\ \text{[Red Globe Icon]} \end{array}, \begin{array}{c} \text{[Blue Globe Icon]} \\ \text{[Red Globe Icon]} \end{array} ? \end{array} \right.$$

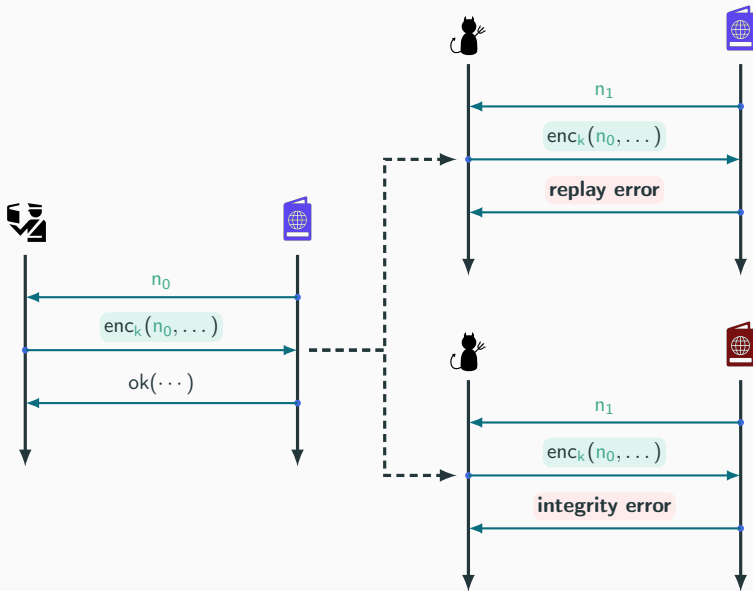
French version of BAC:

- $\neq$  **error messages** for replay and integrity checks.

$\Rightarrow$  **unlinkability attack.**



# BAC Protocol: Privacy Attack





# BAC Protocol: Lessons

Take-away lessons:

- This is a **protocol-level attack**: no issue with cryptography:  
⇒ cryptographic primitives are but an **ingredient**.
- **Innocuous-looking changes** can **break** security:  
⇒ designing security protocols is **hard**.

How to get a **strong confidence** in a protocol's **security guarantees**?

# High-Confidence Security Guarantees

## Verification

Formal mathematical proof of security protocols:



- Must be **sound**: proof  $\Rightarrow$  property always holds.
- Usually **undecidable**: approaches either **incomplete** or **interactive**.
- **Machine-checked proofs** yield a high degree of confidence.
  - ▶ **general-purpose** tools (e.g. **Coq** and **Lean**).
  - ▶ in security protocol analysis, mostly **dedicated** tools.  
E.g. **CryptoVerif**, **EasyCrypt**, **SQUIRREL**.

## Goal

Design **formal frameworks** allowing for **mechanized verification** of **cryptographic protocols**.

- At the intersection of **cryptography** and **verification**.
- Particular verification challenges:
  - ▶ small or medium-sized programs
  - ▶ complex properties
  - ▶ concurrent and probabilistic programs + arbitrary adversary

# The CCSA Approach to Cryptographic Protocol Verification

The Computationally Complete Symbolic Attacker (**CCSA**) [1] is a framework in the **computational model** for the **verification** of cryptographic protocols.

## Key ingredients

- **Protocol executions** modeled as **pure symbolic terms**.
- A **probabilistic logic**.  
⇒ interpret terms as PTIME-computable bitstring distributions.
- **Reasoning rules** capturing **cryptographic arguments**.
- **Abstract approach**: no probabilities, no security parameter.

# Outline

Introduction

Processes

Terms

Process Semantics

A Motivating Example

Symbolic Protocol Execution

Terms

Symbolic Rules

Semantics of Terms

# Processes

---

# Process: Syntax

## Elementary processes:

$$E ::= \mathbf{in}(c, x) \mid \mathbf{out}(c, t) \mid \nu n \mid \mathbf{if } b \text{ then } E \mid \quad (c \in \mathcal{C}, x, n \in \mathcal{X}) \\ E.E \mid \mathbf{null}$$

where  $\mathcal{C}$  is a set of channel symbols and  $\mathcal{X}$  a set of variables.

## Processes:

$$P_0 ::= E \mid (P_0 \mid P_0) \quad P ::= P_0 \mid \nu n.P \quad (n \in \mathcal{X})$$

We let  $\text{chans}(P)$  be the channels of a process  $P$ .

**Restrictions:** elementary process must use a single channel, and distinct elementary processes must use distinct channels. I.e. if  $P = E_1 \mid \cdots \mid E_n$

$$\text{then } \forall i. |\text{chans}(E_i)| \leq 1 \quad \forall i \neq j, \text{chans}(E_i) \cap \text{chans}(E_j) = \emptyset.$$

## Example of a Protocol

As an example, we consider a simple authentication protocol:

### The Private Authentication (PA) Protocol, v1

$$I = \nu n_I. \quad \text{out}(I, \{\langle pk_I, n_I \rangle\}_{pk_S})$$
$$S = \nu n_S. \text{in}(S, x). \text{out}(S, \{\langle \pi_2(\text{dec}(x, sk_I)), n_S \rangle\}_{pk_I})$$

where  $pk_I \equiv pk(k_I)$  and  $pk_S \equiv pk(k_S)$ .

The full protocol is  $\nu k_I. \nu k_S. (I \mid S)$ .

*Notation:*  $\equiv$  denotes *syntactic equality of terms*.



# Processes

---

Terms

# Terms

We use **terms** to model *protocol messages*, built upon a set of **symbols**  $\mathcal{S}$  which includes:

- **Variables**  $\mathcal{X}$ , used, e.g.  $x$  in  $\text{in}(A, x)$  or  $n$  in  $\nu n$ .
- **Function symbols**  $\mathcal{F}$ , e.g.:

$A, B, \langle \cdot, \cdot \rangle, \pi_1(\cdot), \pi_2(\cdot), \{\cdot\}!, \text{pk}(\cdot), \text{sk}(\cdot),$   
 $\text{if } \cdot \text{ then } \cdot \text{ else } \cdot, \cdot = \cdot, \cdot \wedge \cdot, \cdot \vee \cdot, \cdot \rightarrow \cdot$

We note  $\mathcal{T}(\mathcal{S})$  the set of well-typed (see next slide) terms over symbols  $\mathcal{S}$ . Terms are usually written  $t$ , and boolean terms  $b$ .

## Examples

$\text{pk}(k_A)$

$\{\langle \text{pk}_A, n_A \rangle\}_{\text{pk}_B}$

$\pi_1(n_A)$

# Terms: Types

## Types

Each symbol  $s \in \mathcal{S}$  comes with a type  $\text{type}(s)$  of the form:

$$(\tau_b^1 \star \dots \star \tau_b^n) \rightarrow \tau_b \quad \text{or} \quad \tau_b$$

where  $\tau_b^1, \dots, \tau_b^n, \tau_b$  are all **base types** in  $\mathbb{B}$ .

- We ask that  $\mathbb{B}$  contains at least the `message` and `bool` types.
- We restrict *variables* to base types, i.e.:

$$\forall x \in \mathcal{X}, \text{type}(x) \in \mathbb{B}.$$

- We require that terms are well-typed and of a base type:

$$\vdash t : \tau_b \quad \text{where } \tau_b \in \mathbb{B}.$$

The **interpretation**  $\langle t \rangle_{\mathbb{L}}^{\eta, \sigma}$  of a term  $t$  as a bitstring is parameterized by:

- the **security parameter**  $\eta$ ;
- a **library**  $\mathbb{L}$  which provides the semantics  $\langle \cdot \rangle_{\mathbb{L}}$  of **symbols** in  $\mathcal{F}$  (details on next slides);
- the **valuation**  $\sigma : \mathcal{X} \hookrightarrow \{0, 1\}^*$  maps variables to their values.<sup>1</sup>

We may omit  $\sigma$ ,  $\mathbb{L}$  and  $\eta$  when they are clear from the context.

---

<sup>1</sup> $f : \mathbb{A} \hookrightarrow \mathbb{B}$  denotes a *partial*  $f$  function from  $\mathbb{A}$  to  $\mathbb{B}$ .

## Function symbols.

For a function symbols  $f \in \mathcal{F}$ , we simply apply  $\langle f \rangle_{\perp}$ :

$$\langle f(t_1, \dots, t_n) \rangle_{\perp}^{\eta, \sigma} \stackrel{\text{def}}{=} \langle f \rangle_{\perp}(1^{\eta}, \langle t_1 \rangle_{\perp}^{\eta, \sigma}, \dots, \langle t_n \rangle_{\perp}^{\eta, \sigma})$$

**Restriction:**  $\langle f \rangle_{\perp}$  must be **poly-time** computable and **deterministic**.

## Variables.

The interpretation  $\langle x \rangle_{\perp}^{\eta, \sigma}$  of a **variable**  $x \in \mathcal{X}$  is given by the valuation  $\sigma$ :

$$\langle x \rangle_{\perp}^{\eta, \sigma} \stackrel{\text{def}}{=} \sigma(x)$$

## Terms: Builtins

We **force** the interpretation of some **function symbols**.

- $\text{if } \cdot \text{ then } \cdot \text{ else } \cdot$  is interpreted as **branching**:

$$\llbracket \text{if } \cdot \text{ then } \cdot \text{ else } \cdot \rrbracket_{\mathbb{M}}(b, v_1, v_2) \stackrel{\text{def}}{=} \begin{cases} v_1 & \text{if } b = 1 \\ v_2 & \text{otherwise} \end{cases}$$

- $\cdot = \cdot$  is interpreted as an **equality** test:

$$\llbracket \cdot = \cdot \rrbracket_{\mathbb{M}}(v_1, v_2) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } v_1 = v_2 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, we force the interpretations of  $\wedge, \vee, \rightarrow, \text{true}, \text{false}$ .

# Processes

---

## Process Semantics

## Process: Concrete Configuration

A **concrete configuration** is a tuple  $(\phi, \sigma, P)$  representing a **partially executed process** where:

- the **concrete frame**  $\phi$  is the sequence bitstrings  $w_1, \dots, w_n \in \{0, 1\}^*$  outputted since the protocol execution started.
- the **valuation**  $\sigma$ .
- the **process**  $P$  is the process that remains to be executed.

A concrete configuration records the **current state** of an execution.

**Initial configuration:**  $(\epsilon, [\text{st} \mapsto 0], P)$

- $\text{st}$  is a special variable used to store the **adversarial state**.



# Finite Distributions

Processes are **probabilistic**:

⇒ The semantics of a process is a **distribution** of **configurations**.

**Discrete distributions.**

A **discrete distribution** over a set  $\mathcal{S}$  is a **formal sum**  $\sum_{i \in I} a_i \cdot s_i$  where:

$$I \text{ is countable} \quad \sum_{i \in I} a_i = 1 \quad a_i \geq 0 \text{ for any } i \in I$$

## Examples

- $\frac{1}{3} \cdot \text{"0"} + \frac{2}{3} \cdot \text{"42"}$
- $\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1$  (unbiased coin flip).

## Process: Concrete Semantics

A **trace**  $\text{tr}$  is a sequence of **observable actions**  $\alpha_1, \dots, \alpha_n$ . The trace  $\text{tr}$  represents a protocol/adversary **interaction scenario**.

$$\alpha ::= \text{in}(c) \mid \text{out}(c) \quad \text{tr} ::= \epsilon \mid \alpha \mid \text{tr}, \text{tr} \quad (\text{where } c \in \mathcal{C})$$

The **concrete semantics** is given by the relation (see next slides):

$$(\phi, \sigma, P) \xrightarrow[\mathbb{L}, \mathcal{A}, \eta]{\text{tr}} \sum_{i \in I} a_i \cdot (\phi_i, \sigma_i, P_i)$$

**Adversary.**  $\mathcal{A}$  is a **stateful**, **probabilistic** and **poly-time** program.

**Notations.** We omit  $\mathcal{A}, \mathbb{L}, \eta$  when they are fixed or clear from context. Further, we write  $\Rightarrow$  instead of  $\xrightarrow{\epsilon}$ .

## Structural rules:

$$\frac{}{(\phi, \sigma, \mathbf{null} \mid P) \Rightarrow (\phi, \sigma, P)} \quad \frac{}{(\phi, \sigma, P) \Rightarrow (\phi, \sigma, P') \text{ when } P \approx_{AC} P'}$$

where  $\approx_{AC}$  is the small congruence relation over processes which contains:

- **commutativity**  $P_0 \mid P_1 \approx_{AC} P_1 \mid P_0$ ;
- **associativity**  $(P_0 \mid P_1) \mid P_2 \approx_{AC} P_0 \mid (P_1 \mid P_2)$ ;
- **$\alpha$ -renaming**  $\nu n_0. P \approx_{AC} \nu n_1. P[n_0 \mapsto n_1]$   
(same for inputs  $\mathbf{in}(c, x). P$ ).

# Process: Concrete Semantics

## Branching rules:

$$\frac{\langle b \rangle_{\perp}^{\eta, \sigma} = \text{true}}{(\phi, \sigma, \text{if } b \text{ then } P \mid P') \Rightarrow (\phi, \sigma, P \mid P')}$$

$$\frac{\langle b \rangle_{\perp}^{\eta, \sigma} = \text{false}}{(\phi, \sigma, \text{if } b \text{ then } P \mid P') \Rightarrow (\phi, \sigma, P')}$$

## Output rules:

$$\frac{\phi' = (\phi, \langle t \rangle_{\perp}^{\eta, \sigma})}{(\phi, \sigma, (\text{out}(c, t); P) \mid P') \xrightarrow{\text{out}(c)} (\phi', \sigma, P \mid P')}$$

$$\frac{\begin{array}{l} \phi' = (\phi, \text{error}) \\ c \notin \text{chans}(P) \text{ or } P = (\text{in}(c, x). P_0) \end{array}}{(\phi, \sigma, P) \xrightarrow{\text{out}(c)} (\phi', \sigma, P)}$$

# Process: Concrete Semantics

## New rule:

$$\frac{n \notin \text{dom}(\sigma)}{(\phi, \sigma, (\nu n; P) \mid P') \Rightarrow \sum_{|w|=\eta} \frac{1}{2^\eta} \cdot (\phi, \sigma[n \mapsto w], P \mid P')}$$

## Input rule:

$$\frac{x \notin \text{dom}(\sigma) \quad \mathcal{A}(1^\eta, \phi, \sigma[\text{st}]) = \sum_i a_i \cdot (w_i, s_i)}{(\phi, \sigma, (\text{in}(c, x); P) \mid P') \xrightarrow{\text{in}(c)} \sum_i a_i \cdot (\phi', \sigma[x \mapsto w_i, \text{st} \mapsto s_i], P \mid P')}$$

$$\frac{c \notin \text{chans}(P) \text{ or } P = (\text{out}(c, t). P_0) \quad \mathcal{A}(1^\eta, \phi, \sigma[\text{st}]) = \sum_i a_i \cdot (w_i, s_i)}{(\phi, \sigma, P) \xrightarrow{\text{in}(c)} \sum_i a_i \cdot (\phi', \sigma[\text{st} \mapsto s_i], P)}$$

**Remark:**  $\text{dom}(f) \stackrel{\text{def}}{=} \{x \mid f(x) \text{ defined}\}$  is the domain of the partial function  $f$ .

## Transitivity rule:

$$\frac{(\phi, \sigma, P) \xrightarrow{\text{tr}_0} \sum_i a_i \cdot (\phi_i, \sigma_i, P_i) \quad \forall i. (\phi_i, \sigma_i, P_i) \xrightarrow{\text{tr}_1} \sum_j b_{i,j} \cdot (\phi_{i,j}, \sigma_{i,j}, P_{i,j})}{(\phi, \sigma, P) \xrightarrow{\text{tr}_0, \text{tr}_1} \sum_{i,j} a_i \cdot b_{i,j} \cdot (\phi_{i,j}, \sigma_{i,j}, P_{i,j})}$$

## Process: Concrete Semantics

The relation  $\Rightarrow$  is **non-deterministic**.

Still, thanks to the **restrictions** on processes, different non-deterministic choices yield **identical distributions** on frames:

$$\begin{aligned} & (\phi, \sigma, P) \xRightarrow{\text{tr}} \sum_i a_i \cdot (\phi_0^i, \sigma_0^i, P_0^i) \\ \wedge & (\phi, \sigma, P) \xRightarrow{\text{tr}} \sum_j b_j \cdot (\phi_1^j, \sigma_1^j, P_1^j) \\ \Rightarrow & \sum_i a_i \cdot \phi_0^i = \sum_j b_j \cdot \phi_1^j \end{aligned}$$

Thus, given a process  $P$  and a trace  $\text{tr}$ , if:

$$(\epsilon, [\text{st} \mapsto 0], P) \xRightarrow[\mathbb{L}, \mathcal{A}, \eta]{\text{tr}} \sum_i a_i \cdot (\phi_i, \sigma_i, P_i)$$

then the **distribution**  $\sum_i a_i \cdot \phi_i$  is the **concrete execution** of  $P$  over  $\text{tr}$ , denoted  $\text{exec}_{\mathbb{L}, \mathcal{A}}^\eta(P, \text{tr})$ .

## A Motivating Example

---



# Toy Protocol

We consider a toy protocol as a motivating example:

$A : \nu n. \quad \text{out}(A, \{n\}_k). \text{out}(A, n)$   
 $B : \text{in}(B, x). \text{if } \text{dec}(x, k) \neq \perp \text{ then}$   
 $\quad \text{in}(B, y). \text{out}(B, \text{dec}(x, k) = y)$

## Security property.

*B outputs false if A does not send its second message.*

## Assumptions. (informally)

We assume that the *symmetric encryption*  $\{\cdot\}_k$  is a secure AE:

- **Integrity:** if  $\text{dec}(m, k) \neq \perp$  then  $m$  is an **honest encryption**.
- **Confidentiality:**  $\{m_0\}_k$  and  $\{m_1\}_k$  are **indistinguishable**.  
(when  $m_0$  and  $m_1$  are of the same length.)

## Security of our Toy Protocol

**A** :  $\nu n.$     **out**(A,  $\{n\}_k$ ). **out**(A,  $n$ )  
**B** : **in**(B,  $x$ ). if  $\text{dec}(x, k) \neq \perp$  then  
          **in**(B,  $y$ ). **out**(B,  $\text{dec}(x, k) = y$ )

When executing the trace

**out**(A), **in**(B), **in**(B), **out**(B),

the adversary sees the messages:

$\{n\}_k$ , if  $(\text{dec}(\text{att}_1(\Phi_1), k) \neq \perp)$  then  $(\text{dec}(\text{att}_1(\Phi_1), k) = \text{att}_2(\Phi_1))$

where  $\Phi_1 = \{n\}_k$ , and **att**<sub>1</sub>, **att**<sub>2</sub> represents, resp., the first and second inputs to **B** chosen by the adversary.

# Security of our Toy Protocol

Our toy protocol is **secure** if the following **indistinguishability** holds:

$$\{n\}_k, \text{ if } (\text{dec}(\mathbf{att}_1(\Phi_1), k) \neq \perp) \text{ then } (\text{dec}(\mathbf{att}_1(\Phi_1), k) = \mathbf{att}_2(\Phi_1)) \\ \approx \{n\}_k, \text{ if } (\text{dec}(\mathbf{att}_1(\Phi_1), k) \neq \perp) \text{ then false}$$

## Informal Security Proof

Using the **integrity** of the encryption, we have:

$$(\text{dec}(\mathbf{att}_1(\Phi_1), k) \neq \perp) \Leftrightarrow (\mathbf{att}_1(\Phi_1) = \{n\}_k).$$

Thus:

$$\begin{aligned} & \{n\}_k, \text{ if } (\text{dec}(\mathbf{att}_1(\Phi_1), k) \neq \perp) \text{ then } (\text{dec}(\mathbf{att}_1(\Phi_1), k) = \mathbf{att}_2(\Phi_1)) \\ \approx & \{n\}_k, \text{ if } (\mathbf{att}_1(\Phi_1) = \{n\}_k) \quad \text{then } (\text{dec}(\mathbf{att}_1(\Phi_1), k) = \mathbf{att}_2(\Phi_1)) \\ \approx & \{n\}_k, \text{ if } (\mathbf{att}_1(\Phi_1) = \{n\}_k) \quad \text{then } (n = \mathbf{att}_2(\Phi_1)) \end{aligned}$$

## Informal Security Proof

Using the **integrity** of the encryption, we have:

$$(\text{dec}(\mathbf{att}_1(\Phi_1), k) \neq \perp) \Leftrightarrow (\mathbf{att}_1(\Phi_1) = \{n\}_k).$$

Thus:

$$\begin{aligned} & \{n\}_k, \text{ if } (\text{dec}(\mathbf{att}_1(\Phi_1), k) \neq \perp) \text{ then } (\text{dec}(\mathbf{att}_1(\Phi_1), k) = \mathbf{att}_2(\Phi_1)) \\ \approx & \{n\}_k, \text{ if } (\mathbf{att}_1(\Phi_1) = \{n\}_k) \quad \text{then } (\text{dec}(\mathbf{att}_1(\Phi_1), k) = \mathbf{att}_2(\Phi_1)) \\ \approx & \{n\}_k, \text{ if } (\mathbf{att}_1(\Phi_1) = \{n\}_k) \quad \text{then } (n = \mathbf{att}_2(\Phi_1)) \end{aligned}$$

Using the **confidentiality** of the encryption, we can replace  $\{n\}_k$  by  $\{0^\eta\}_k$  (assuming  $\text{len}(n) = \eta$ ). Thus:

$$\begin{aligned} & \{n\}_k, \text{ if } (\mathbf{att}_1(\Phi_1) = \{n\}_k) \quad \text{then } (n = \mathbf{att}_2(\Phi_1)) \\ \approx & \{0^\eta\}_k, \text{ if } (\mathbf{att}_1(\{0^\eta\}_k) = \{0^\eta\}_k) \text{ then } (n = \mathbf{att}_2(\{0^\eta\}_k)) \end{aligned}$$

# Informal Security Proof

Using the **integrity** of the encryption, we have:

$$(\text{dec}(\mathbf{att}_1(\Phi_1), k) \neq \perp) \Leftrightarrow (\mathbf{att}_1(\Phi_1) = \{n\}_k).$$

Thus:

$$\begin{aligned} & \{n\}_k, \text{ if } (\text{dec}(\mathbf{att}_1(\Phi_1), k) \neq \perp) \text{ then } (\text{dec}(\mathbf{att}_1(\Phi_1), k) = \mathbf{att}_2(\Phi_1)) \\ & \approx \{n\}_k, \text{ if } (\mathbf{att}_1(\Phi_1) = \{n\}_k) \quad \text{then } (\text{dec}(\mathbf{att}_1(\Phi_1), k) = \mathbf{att}_2(\Phi_1)) \\ & \approx \{n\}_k, \text{ if } (\mathbf{att}_1(\Phi_1) = \{n\}_k) \quad \text{then } (n = \mathbf{att}_2(\Phi_1)) \end{aligned}$$

Using the **confidentiality** of the encryption, we can replace  $\{n\}_k$  by  $\{0^\eta\}_k$  (assuming  $\text{len}(n) = \eta$ ). Thus:

$$\begin{aligned} & \{n\}_k, \text{ if } (\mathbf{att}_1(\Phi_1) = \{n\}_k) \quad \text{then } (n = \mathbf{att}_2(\Phi_1)) \\ & \approx \{0^\eta\}_k, \text{ if } (\mathbf{att}_1(\{0^\eta\}_k) = \{0^\eta\}_k) \text{ then } (n = \mathbf{att}_2(\{0^\eta\}_k)) \end{aligned}$$

By **probabilistic independence**:  $n \neq \mathbf{att}_2(\{0^\eta\}_k)$  (overwhelmingly), yielding:

$$\{0^\eta\}_k, \text{ if } (\mathbf{att}_1(\{0^\eta\}_k) = \{0^\eta\}_k) \text{ then false}$$

We conclude by a similar proof in the other direction.

## Verification of Cryptographic Protocols

$$\forall \mathcal{A} \in \mathcal{C}. (\mathcal{A} \parallel P) \models \Phi$$

To build a verification framework from what we just did, we need to:

- represent the **adversary/protocol interaction**  $(\mathcal{A} \parallel P)$  as **symbolic terms**;
- express the **security property**  $\Phi$  using a **logical formula**;
- capture the **cryptographic arguments**  $\models$  as **reasoning rules**.

# Symbolic Protocol Execution

---



# Symbolic Protocol Execution

**Goal:** obtain **symbolic representations** of protocol executions that:

- **faithfully** model the protocol semantics;
- are amenable to **formal reasoning**.

**How?** Use the same **techniques** as in our motivating example.

- **1. Explicit probabilistic dependencies:**  
randomness is sampled **eagerly** and not **lazily**.
- **2. Pure encoding of the adversary:**  
no adversarial state.

# Symbolic Execution: Explicit Probabilistic Dependencies

## 1. Explicit probabilistic dependencies.

### Key idea.

*Sample randomness before-hand (eagerly), and retrieve it as needed.*

Concretely, we rewrite a process  $P$  by **moving randomness early**:

$$\begin{aligned} P = E_1 \mid \dots \mid E_l &\Leftrightarrow (\nu \vec{n}_1. E'_1) \mid \dots \mid (\nu \vec{n}_l. E'_l) \\ &\Leftrightarrow \nu \vec{n}_1, \dots, \vec{n}_l. (E'_1 \mid \dots \mid E'_l) \end{aligned}$$

where  $(\nu \vec{n}_i. E'_i)$  is  $E_i$  with all random samplings moved at the beginning.  
and names are distincts:

$$\forall i \neq j. \vec{n}_i \cap \vec{n}_j = \emptyset.$$

# Symbolic Execution: Explicit Probabilistic Dependencies

More precisely,  $E_i \rightsquigarrow^! (\nu \vec{n}_i. E'_i)$  where  $\rightsquigarrow$  is defined by the rules:

$$(E_0. \nu n. E_1) \rightsquigarrow \nu n. (E_0. E_1) \quad (E_0 \mid \nu n. E_1) \rightsquigarrow \nu n. (E_0 \mid E_1)$$

$$\text{if } b \text{ then } \nu n. E \rightsquigarrow \nu n. \text{if } b \text{ then } E$$

where  $n \notin \text{fv}(E_0)$  and  $n \notin \text{fv}(b)$ .

💡 *Recall that process are taken modulo  $\alpha$ -renaming.*

**Notations.**  $E_0 \rightsquigarrow^! E_1$  iff.  $E_0 \rightsquigarrow^* E_1$  and  $E_1 \not\rightsquigarrow E'$  for all  $E'$ .

$\text{fv}(E)$  and  $\text{fv}(t)$  are the **free variables** of, resp.,  $E$  and  $t$ .

## 2. Pure encoding of the adversary.

We need a **deterministic** and **stateless** representation  $\mathcal{A}_p$  of the adversary  $\mathcal{A}$ .

- **deterministic**: sample the adversary's randomness  $\rho_a$  **eagerly** and pass it as an **explicit argument**.
- **stateless**:  $\mathcal{A}_p$  recomputes  $\mathcal{A}$ 's state each time it is called.  
💡 *This exploits determinism, and requires us to provide the full history each time we call  $\mathcal{A}_p$ .*

# Symbolic Executions: Pure Encoding of the Adversary

Informally, we want that:

$$\begin{aligned} & \mathcal{A}(w_1) \quad , \quad \dots \quad , \quad \mathcal{A}(w_n) \\ \stackrel{=}{\text{distr.}} & \mathcal{A}_p(w_1, \rho_a) \quad , \quad \dots \quad , \quad \mathcal{A}_p(w_1, \dots, w_n, \rho_a) \end{aligned}$$

where  $\rho_a$  is a *long enough* sequence of bits sampled **independently uniformly at random**.

We call  $\mathcal{A}_p$  a **pure representation** of  $\mathcal{A}$ .

We do not detail it, but  $\mathcal{A}_p$  can be systematically built from  $\mathcal{A}$ .

# Symbolic Protocol Execution

---

Terms

# Terms

To define our **symbolic execution rules**, we need to **extend** our set of **term symbols**  $\mathcal{S} = \mathcal{N} \uplus \mathcal{X} \uplus \mathcal{F} \uplus \mathcal{G}$ :

- **Variables**  $\mathcal{X}$ .
- **Function symbols**  $\mathcal{F}$ .
- **Names**  $\mathcal{N}$ .
- **Adversarial function symbols**  $\mathcal{G}$ , of any arity.

## Changes.

- Names are no longer represented by variables in  $\mathcal{X}$ , but by special symbols in  $\mathcal{N}$  with a tailored semantics (presented later).  
(For the sake of simplicity, we ask that all names are of type **message**.)
- Adversarial function symbols  $\mathcal{G}$  are used to represent calls to  $\mathcal{A}_p$ .

# Adversarial function symbols

More precisely, if:

- there has already been  $n$  **outputs**, represented by the terms  $t_1, \dots, t_n$ ;
- and we are doing the  $j$ -th **input** since the protocol started;

then the **input bitstring** is **represented** by:

$$\mathbf{att}_j(t_1, \dots, t_n)$$

where  $\mathbf{att}_j \in \mathcal{G}$  is an **adversarial** function symbol of arity  $n$ .

💡  $j$  allows to have different values for consecutive inputs.



# Symbolic Protocol Execution

---

## Symbolic Rules

# Symbolic Executions

We describe a **systematic method** to compute, given a **process**  $P$  and a **trace**  $\text{tr}$  of **observable actions**, the **terms** representing the **outputted messages** during the execution of  $P$  over  $\text{tr}$ .

This is the **symbolic execution** of  $P$  over  $\text{tr}$ .

We deal with the protocol **randomness** and **adversarial inputs** using the two techniques we just saw.

## Symbolic configuration

A **symbolic configuration** is a tuple  $(\Phi; \lambda; j; \Pi_1, \dots, \Pi_l)$  where:

- $\Phi$  is a sequence of terms (in  $\mathcal{T}(\mathcal{S})$ ).
- $\lambda$  is a finite sequence of mappings  $(x \mapsto t)$  where  $t$  is a term.
- $j \in \mathbb{N}$ .
- for every  $i$ ,  $\Pi_i = (P_i, b_i)$  where  $P_i$  is a protocol and  $b_i$  is a boolean term.

## Symbolic Configuration: Intuition

In a **symbolic configuration**  $(\Phi; \lambda; j; \Pi_1, \dots, \Pi_l)$ :

- $\Phi$  is the **frame**, i.e. the sequence of terms outputted since the execution started.
- $\lambda$  **records inputs**, it maps input variable to their corresponding term.
- $j$  **counts the number of inputs** since the execution started.
- $(P, b)$  **represent the protocol**  $P$  if  $b$  is true (and is **null** otherwise).  
Using this interpretation,  $\Pi_1, \dots, \Pi_l$  is the **current process**.

**Initial configuration:**  $(\epsilon; \emptyset; 0; (P, \top))$

# Symbolic Execution: Branching and Input Rules

## Rule for protocol branching:

$$\begin{aligned} & (\Phi; \lambda; j; (\text{if } b_0 \text{ then } P, b_1), \Pi_1, \dots, \Pi_l) \\ \hookrightarrow & \quad (\Phi; \lambda; j; (P, b_0 \wedge b_1), \Pi_1, \dots, \Pi_l) \end{aligned}$$

## Rules for inputs:

$$\begin{aligned} & (\Phi; \lambda; j; (\mathbf{in}(c, x).P, b), \Pi_1, \dots, \Pi_l) \\ \stackrel{\mathbf{in}(c)}{\hookrightarrow} & (\Phi; \lambda[x \mapsto \mathbf{att}_j(\Phi)]; j + 1; (P, b), \Pi_1, \dots, \Pi_l) \end{aligned} \quad (x \notin \text{dom}(\lambda))$$

$$(\Phi; \lambda; j; \Pi_1, \dots, \Pi_l) \stackrel{\mathbf{in}(c)}{\hookrightarrow} (\Phi; \lambda; j + 1; \Pi_1, \dots, \Pi_l) \quad (\dagger)$$

where  $(\dagger)$ :  $c \notin \text{chans}(\Pi_1, \dots, \Pi_l)$  or  $\exists i$  s.t.  $\Pi_i$  starts with  $\mathbf{out}(c, \cdot)$ .

# Symbolic Execution: Output Rule

## Rules for outputs:

$$\begin{array}{c} (\Phi; \lambda; j; (\mathbf{out}(c, t).P, b), \Pi_1, \dots, \Pi_l) \\ \xrightarrow{\mathbf{out}(c)} (\Phi, (\text{if } b \text{ then } t)\lambda; \lambda; j; (P, b), \Pi_1, \dots, \Pi_l) \end{array}$$

$$(\Phi; \lambda; j; \Pi_1, \dots, \Pi_l) \xrightarrow{\mathbf{out}(c)} (\Phi, \text{error}; \lambda; j; \Pi_1, \dots, \Pi_l) \quad (\ddagger)$$

where  $(\ddagger)$ :  $c \notin \text{chans}(\Pi_1, \dots, \Pi_l)$  or  $\exists i$  s.t.  $\Pi_i$  starts with  $\mathbf{out}(c, \cdot)$ .

💡 *The input and output rules make sense because we restrict ourselves to elementary processes with distinct channels.*

# Symbolic Execution

Given a process  $P$  (without  $\nu$ ) and a trace  $\text{tr}$ , if:

$$(\epsilon; \emptyset; 0; (P, \top)) \xrightarrow{\text{tr}} (\Phi; \_; \_; \_)$$

then  $\Phi$  is the **symbolic execution** of  $P$  over  $\text{tr}$ , denoted  $\text{s-exec}(P, \text{tr})$ .

## Handling the $\nu$ construct.

If  $P$  contains  $\nu$ , we compute  $P_0$  s.t.  $P \rightsquigarrow^! \nu \vec{n}. P_0$  with  $\vec{n} \in \mathcal{N}$ , and then symbolically execute  $P_0$ .

# Symbolic Execution: Soundness

## Claim (informal).

The symbolic execution is **sound** w.r.t. the concrete semantics.

More precisely, for every library  $\mathbb{L}$ , adversary  $\mathcal{A}$ , and security parameter  $\eta$ :

$$\text{exec}_{\mathbb{L}, \mathcal{A}}^{\eta}(P, \text{tr}) =_{\text{distr.}} \llbracket \text{s-exec}(P, \text{tr}) \rrbracket_{\mathbb{L}[\text{att} \mapsto \mathcal{A}_p]}^{\eta, \rho}$$

where:

- $\rho$  is sampled uniformly at random among bitstrings of sufficient length.
- $\mathcal{A}_p$  is a pure representation of  $\mathcal{A}$ .

**Remark:**  $\llbracket t \rrbracket_{\mathbb{M}}^{\eta, \rho}$  is the semantics of the terms coming from the symbolic execution, which we define in the next section.



## Symbolic Execution: Exercises

### Exercise

What are all the **possible symbolic executions** of the following protocols?

$\text{in}(c, x). \text{out}(c, t)$

$\text{out}(A, t_1) \mid \text{in}(B, x). \text{out}(B, t_2)$

if  $b$  then  $\text{out}(c, t_1)$  else  $\text{out}(c, t_2)$

if  $b$  then  $\text{out}(A, t_1)$  else  $\text{out}(B, t_2)$

### Exercise

Extend the **symbolic** algorithm with a rule allowing to handle processes with let bindings.

Could the same thing be done for mutable, inter-process, state?

# Semantics of Terms

---

# Semantics of Terms

We showed how to represent **protocol execution**, on some fixed trace of observables  $\text{tr}$ , as a **sequence of terms**.

Intuitively, the terms corresponds to **PTIME-computable bitstring distributions**.

## Example

If  $\langle \_ , \_ \rangle$  is the concatenation, and samplings are done uniformly at random among bitstrings of length  $\eta \in \mathbb{N}$ , then:

$$\nu n_0, \nu n_1, \mathbf{out}(c, \langle n_0, \langle 00, n_1 \rangle \rangle) \quad \text{yields} \quad \langle n_0, \langle 00, n_1 \rangle \rangle$$

which represent a distribution over bitstrings of length  $2 \cdot \eta + 2$ , where all bits are sampled uniformly and independently, except for the bits at positions  $\eta$  and  $\eta + 1$ , which are always 0.

# Semantics of Terms

We interpret  $t \in \mathcal{T}(\mathcal{S})$  as a **Probabilistic Polynomial-time Turing machine** (PPTM), with:

- a **working tape** (also used as input tape);
- two **read-only tapes**  $\rho = (\rho_a, \rho_h)$  for adversary and honest randomness.

We let  $\mathcal{D}$  be the set of such machines.

💡 *The machine must be polynomial in the size of its input on the working tape only.*

## Terms Interpretation

The **interpretation**  $\llbracket t \rrbracket_{\mathbb{M}} \in \mathcal{D}$  of a term  $t$  is parameterized by a **model**  $\mathbb{M}$  which provides:

- the set of random tapes  $\mathbb{T}_{\mathbb{M},\eta} = \mathbb{T}_{\mathbb{M},\eta}^a \times \mathbb{T}_{\mathbb{M},\eta}^h$ , where  $\mathbb{T}_{\mathbb{M},\eta}^a$  and  $\mathbb{T}_{\mathbb{M},\eta}^h$  are **finite** same-length set of bit-strings.

We equip it with the **uniform** probability measure.

( $\mathbb{T}_{\mathbb{M},\eta}^a$  for the adversary,  $\mathbb{T}_{\mathbb{M},\eta}^h$  for honest functions)

- the semantics  $(\cdot)_{\mathbb{M}}$  of **symbols** in  $\mathcal{S}$  (details on next slides).

(This extends the interpretation  $(\cdot)_{\mathbb{L}}$  of symbols by a library  $\mathbb{L}$ .)

We may omit  $\mathbb{M}$  when it is clear from context.

We define the machine  $\llbracket t \rrbracket_{\mathbb{M}} \in \mathcal{D}$ , by defining its behavior  $\llbracket t \rrbracket_{\mathbb{M}}^{\eta,\rho}$  for every  $\eta \in \mathbb{N}$  and pairs of random tapes  $\rho = (\rho_a, \rho_h) \in \mathbb{T}_{\mathbb{M},\eta}$ .

## Terms Interpretation: Function Symbols

Function symbols interpretations is just **composition**.

For **function symbols** in  $f \in \mathcal{F}$ , we simply apply  $\langle f \rangle_{\mathbb{M}}$ :

$$\llbracket f(t_1, \dots, t_n) \rrbracket_{\mathbb{M}}^{\eta, \rho} \stackrel{\text{def}}{=} \langle f \rangle_{\mathbb{M}}(1^\eta, \llbracket t_1 \rrbracket_{\mathbb{M}}^{\eta, \rho}, \dots, \llbracket t_n \rrbracket_{\mathbb{M}}^{\eta, \rho})$$

**Adversarial function symbols**  $g \in \mathcal{G}$  also have access to  $\rho_a$ :

$$\llbracket g(t_1, \dots, t_n) \rrbracket_{\mathbb{M}}^{\eta, \rho} \stackrel{\text{def}}{=} \langle g \rangle_{\mathbb{M}}(1^\eta, \llbracket t_1 \rrbracket_{\mathbb{M}}^{\eta, \rho}, \dots, \llbracket t_n \rrbracket_{\mathbb{M}}^{\eta, \rho}, \rho_a)$$

**Restrictions.**  $\langle f \rangle_{\mathbb{M}}$  and  $\langle g \rangle_{\mathbb{M}}$  are:

- PTIME-computable;
- **deterministic** (all randomness must come explicitly, from  $\rho$ ).

## Terms Interpretation: Variables and Names

The interpretation  $\langle x \rangle_{\mathbb{M}}$  of a **variable**  $x \in \mathcal{X}$  is an **arbitrary machine** in  $\mathcal{D}$ . Then:

$$\llbracket x \rrbracket_{\mathbb{M}}^{\eta, \rho} \stackrel{\text{def}}{=} \langle x \rangle_{\mathbb{M}}(1^\eta, \rho).$$

**Names**  $n \in \mathcal{G}$  are interpreted as **uniform random samplings** among bitstrings of length  $\eta$ , extracted from  $\rho_h$ :

$$\llbracket n \rrbracket_{\mathbb{M}}^{\eta, \rho} \stackrel{\text{def}}{=} \langle n \rangle_{\mathbb{M}}(1^\eta, \rho_h)$$

For every pair of different names  $n_0, n_1$ , we require that  $\langle n_0 \rangle_{\mathbb{M}}$  and  $\langle n_1 \rangle_{\mathbb{M}}$  extracts disjoint parts of  $\rho_h$ .

💡 Hence different names are **independent random samplings**.

# Terms Interpretation: Names

## Examples

- If  $(n, n_0 : \text{message})$  then:

$$\llbracket n \rrbracket^\eta =_{\text{distr.}} \text{sample } w \text{ in } \{0, 1\}^\eta$$

$$\llbracket (n, n_0) \rrbracket^\eta =_{\text{distr.}} \begin{array}{l} \text{sample } w \text{ in } \{0, 1\}^\eta \\ \text{sample } w' \text{ in } \{0, 1\}^\eta \text{ **independently**} \\ \text{build } (w, w') \end{array}$$

$$\llbracket (n, n) \rrbracket^\eta =_{\text{distr.}} \begin{array}{l} \text{sample } w \text{ in } \{0, 1\}^\eta \\ \text{build } (w, w) \end{array}$$

Indeed:

$$\llbracket (n, n) \rrbracket^{\eta, \rho} = (\llbracket n \rrbracket^{\eta, \rho}, \llbracket n \rrbracket^{\eta, \rho}) = (w, w) \quad (\text{where } w = \langle n \rangle(1^\eta, \rho_h))$$



## Terms Interpretation: Modeling and Randomness

≠ in how randomness is sampled:

- **In the “real-world”**, the adversary  $\mathcal{A}$  samples randomness on-the-fly, as needed.  
⇒ possibly  $P(\eta)$  random bits, where  $P$  is the (polynomial) running-time of  $\mathcal{A}$ .
- **In the logic**, we restrict  $\mathbb{T}_{\mathbb{M},\eta} = \mathbb{T}_{\mathbb{M},\eta}^a \times \mathbb{T}_{\mathbb{M},\eta}^h$  to be finite and fixed by  $\mathbb{M}$ .  
⇒ all randomness sampled eagerly according to  $\mathbb{M}$ , independently of the adversary  $\mathcal{A}$ .

This ≠ of behaviors is not an issue, i.e. the logic can **soundly model** real-world adversaries:

- Indeed, for any adversary  $\mathcal{A}$ , there exists a model  $\mathbb{M}$  with enough randomness.

- [1] G. Bana and H. Comon-Lundh.  
**A computationally complete symbolic attacker for equivalence properties.**  
In *CCS*, pages 609–620. ACM, 2014.