

# A Higher-Order Indistinguishability Logic for Cryptographic Reasoning

David Baelde, Adrien Koutsos, Joseph Lallemand

Univ Rennes, CNRS & IRISA      Inria Paris

GT MFS Days 2023



# New foundation for the Squirrel prover

Squirrel is a **proof assistant** for verifying cryptographic protocols in the computational model. It is based on the **CCSA approach**.

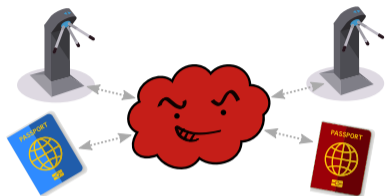


 Gergei Bana & Hubert Comon. *A Computationally Complete Symbolic Attacker for Equivalence Properties*. CCS 2014.

## Outline

- Brief presentation of CCSA **base logic**, and Squirrel's **meta-logic**.
- How a **higher-order CCSA logic** solves several problems.

## Example protocol: Basic Hash



Each tag ( $T_i$ ) owns a secret key  $k_i$ .

Reader ( $R$ ) knows all legitimate keys.

$T_i \rightarrow R : \langle n_T, h(n_T, k_i) \rangle$

$R \rightarrow T_i : \text{ok}$

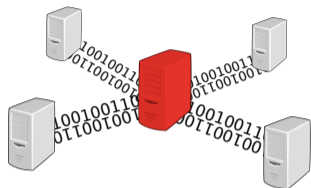
Scenario under consideration:

- Roles  $R, T_1, \dots, T_n$  with arbitrary number of sessions for each role.
- Attacker can intercept messages, inject new messages.

Security properties:

- Readers must accept only legitimate inputs.
- It must not be possible to track tags.

# Cryptographer's model for provable security



Messages = bitstrings

Secrets = random samplings

Participants = PPTIME Turing machines

Rule out unavoidable, unimportant attacks:

- Attacks with **negligible** probability of success (asymptotically smaller than any  $\eta^{-k}$ ).
- Attacks that cannot run in probabilistic polynomial-time.

## Cryptographic assumptions

- Keyed hash function may be collision resistant, unforgeable, pseudo-random.
- Encryptions should ensure various forms of indistinguishability.
- Etc.

# Outline

1 Running example

**2 Base logic**

3 Meta-logic

4 Higher-order logic

5 Conclusion

## Base logic syntax

First-order logic with terms interpreted as **probabilistic computations of bitstrings**.

- Names  $n, r, k \dots$ : special constants represent random samplings.
- Honest function symbols  $f, enc, ok \dots$ : represent primitives, public constants, etc.
- Adversarial function symbols  $att$ ; represent attacker computations.

### Example

In models where  $h$  is an unforgeable hash function,  
 $att(h(true, k))$  and  $h(false, k)$  are unlikely to compute the same bitstring.

## Base logic syntax

First-order logic with terms interpreted as **probabilistic computations of bitstrings**.

- Names  $n, r, k, \dots$ : special constants represent random samplings.
- Honest function symbols  $f, \text{enc}, \text{ok}, \dots$ : represent primitives, public constants, etc.
- Adversarial function symbols  $\text{att}$ ; represent attacker computations.

### Example

In models where  $h$  is an unforgeable hash function,  
 $\text{att}(h(\text{true}, k))$  and  $h(\text{false}, k)$  are unlikely to compute the same bitstring.

Formulas built from a single predicate  $\vec{u} \sim \vec{v}$  expressing indistinguishability.

### Example

$\exists x. (x \sim \text{ok}) \Rightarrow \perp$

## Application to Basic Hash protocol

Simple privacy scenario:  $T_1, T_2$  vs.  $T_1, T'_1$ .

$$\text{Let } u_1, u_2 = \langle n_t, h(n_t, k_1) \rangle, \langle n'_t, h(n'_t, k_2) \rangle$$

$$v_1, v_2 = \langle n_t, h(n_t, k_1) \rangle, \langle n'_t, h(n'_t, k_1) \rangle$$

Privacy expressed as  $u_1, u_2 \sim v_1, v_2$ .



## Application to Basic Hash protocol

Simple privacy scenario:  $T_1, T_2$  vs.  $T_1, T'_1$ .

$$\text{Let } u_1, u_2 = \langle n_t, h(n_t, k_1) \rangle, \langle n'_t, h(n'_t, k_2) \rangle$$

$$v_1, v_2 = \langle n_t, h(n_t, k_1) \rangle, \langle n'_t, h(n'_t, k_1) \rangle$$

Privacy expressed as  $u_1, u_2 \sim v_1, v_2$ .

Authentication for  $T_1, T_2$  expressed as  $\phi \sim \text{true}$  with  $\phi$  as follows with  $\text{input}_R = \mathbf{att}(u_1, u_2)$ :

$$(\text{snd}(\text{input}_R) \stackrel{\bullet}{=} h(\text{fst}(\text{input}_R), k_1) \dot{\vee}$$

$$\text{snd}(\text{input}_R) \stackrel{\bullet}{=} h(\text{fst}(\text{input}_R), k_2))$$

$$\Rightarrow (\text{input}_R \stackrel{\bullet}{=} u_1 \dot{\vee} \text{input}_R \stackrel{\bullet}{=} u_2)$$

We assume some builtin connectives with their expected semantics:  $- \stackrel{\bullet}{=} -, - \dot{\vee} -, - \Rightarrow \dots$

## Semantics of terms

### Definition (Semantics in a model $\mathcal{M}$ )

For a term  $t$ ,  $\llbracket t \rrbracket_{\mathcal{M}}$  is a **deterministic polynomial time Turing machine** taking as inputs:

- the security parameter  $\eta$ ;
- two infinite randomness tapes  $\rho = (\rho_h, \rho_a)$  for **honest** and **attacker** samplings.

Random samplings can be tracked syntactically:

- Honest randomness only accessed by names.
- Attacker randomness only accessed by **att** <sub>$i$</sub>  symbols.
- Variables can access anything!

### Example

$$\Pr[ \llbracket n \dot{=} t \rrbracket_{\mathcal{M}}(\eta, \rho) = \text{true} ] = 2^{-\eta} \text{ if } t \text{ closed, } n \notin t$$

## Semantics of formulas

### Definition (Computational indistinguishability)

$\mathcal{M} \models \vec{u} \sim \vec{v}$  when, for any polynomial time Turing machine  $\mathcal{D}$ ,

$$| \Pr[\mathcal{D}(\llbracket \vec{u} \rrbracket_{\mathcal{M}}(\eta, \rho_h, \rho_a), \eta, \rho_a)] - \Pr[\mathcal{D}(\llbracket \vec{v} \rrbracket_{\mathcal{M}}(\eta, \rho_h, \rho_a), \eta, \rho_a)] | \text{ is negligible.}$$

The rest is as usual in first-order logic.

## Semantics of formulas

### Definition (Computational indistinguishability)

$\mathcal{M} \models \vec{u} \sim \vec{v}$  when, for any polynomial time Turing machine  $\mathcal{D}$ ,

$$| \Pr[\mathcal{D}(\llbracket \vec{u} \rrbracket_{\mathcal{M}}(\eta, \rho_h, \rho_a), \eta, \rho_a)] - \Pr[\mathcal{D}(\llbracket \vec{v} \rrbracket_{\mathcal{M}}(\eta, \rho_h, \rho_a), \eta, \rho_a)] | \text{ is negligible.}$$

The rest is as usual in first-order logic.

### Example

- $u \sim \text{true}$  means that  $u$  is true with overwhelming probability.

## Semantics of formulas

### Definition (Computational indistinguishability)

$\mathcal{M} \models \vec{u} \sim \vec{v}$  when, for any polynomial time Turing machine  $\mathcal{D}$ ,

$$| \Pr[\mathcal{D}([\vec{u}]_{\mathcal{M}}(\eta, \rho_h, \rho_a), \eta, \rho_a)] - \Pr[\mathcal{D}([\vec{v}]_{\mathcal{M}}(\eta, \rho_h, \rho_a), \eta, \rho_a)] | \text{ is negligible.}$$

The rest is as usual in first-order logic.

### Example

- $u \sim \text{true}$  means that  $u$  is true with overwhelming probability.
- $\text{true} \sim (t \neq n)$  is valid when  $t$  closed and does not contain  $n$ .

# Semantics of formulas

## Definition (Computational indistinguishability)

$\mathcal{M} \models \vec{u} \sim \vec{v}$  when, for any polynomial time Turing machine  $\mathcal{D}$ ,

$$| \Pr[\mathcal{D}([\vec{u}]_{\mathcal{M}}(\eta, \rho_h, \rho_a), \eta, \rho_a)] - \Pr[\mathcal{D}([\vec{v}]_{\mathcal{M}}(\eta, \rho_h, \rho_a), \eta, \rho_a)] | \text{ is negligible.}$$

The rest is as usual in first-order logic.

## Example

- $u \sim \text{true}$  means that  $u$  is true with overwhelming probability.
- $\text{true} \sim (t \neq n)$  is valid when  $t$  closed and does not contain  $n$ .
- $\forall \vec{x}, \vec{y}, \vec{x}', \vec{y}'. (\vec{x}, \vec{y} \sim \vec{x}', \vec{y}') \Rightarrow (\vec{x}, f(\vec{y}) \sim \vec{x}', f(\vec{y}'))$  is valid for any function symbol  $f$ .

# Outline

1 Running example

2 Base logic

**3 Meta-logic**

4 Higher-order logic

5 Conclusion

## Basic Hash protocol in the meta-logic

Simple privacy scenario:  $T_1, T_2$  vs.  $T_1, T'_1$ .

$$\begin{aligned} \text{Let } u_1, u_2 &= \langle n_t, h(n_t, k_1) \rangle, \langle n'_t, h(n'_t, \boxed{k_2}) \rangle \\ v_1, v_2 &= \langle n_t, h(n_t, k_1) \rangle, \langle n'_t, h(n'_t, \boxed{k_1}) \rangle \end{aligned}$$

Privacy expressed as  $u_1, u_2 \sim v_1, v_2$ .



## Basic Hash protocol in the meta-logic

Privacy expressed as  $\text{frame}@_{\tau} \sim_{\mathcal{P}, \mathcal{P}'} \text{frame}@_{\tau}$ .

- Timestamp  $\tau$  refers to an arbitrary point in any arbitrary execution trace.
- Macro  $\text{frame}@_{\tau}$  stands for the sequence of all protocol outputs before  $\tau$ .
- Protocol  $\mathcal{P}$  = multiple tags playing multiple sessions.
- Protocol  $\mathcal{P}'$  = multiple tags playing single sessions.

## Basic Hash protocol in the meta-logic

Privacy expressed as  $\text{frame}@_\tau \sim_{\mathcal{P}, \mathcal{P}'} \text{frame}@_\tau$ .

- Timestamp  $\tau$  refers to an arbitrary point in any arbitrary execution trace.
- Macro  $\text{frame}@_\tau$  stands for the sequence of all protocol outputs before  $\tau$ .
- Protocol  $\mathcal{P}$  = multiple tags playing multiple sessions.
- Protocol  $\mathcal{P}'$  = multiple tags playing single sessions.

Authentication for  $T_1, T_2$  expressed as  $\phi \sim \text{true}$  with  $\phi$  as follows with  $\text{input}_R = \text{att}(u_1, u_2)$ :

$$\begin{aligned} & (\text{snd}(\text{input}_R) \doteq \text{h}(\text{fst}(\text{input}_R), k_1) \dot{\vee} \\ & \quad \text{snd}(\text{input}_R) \doteq \text{h}(\text{fst}(\text{input}_R), k_2)) \\ \Rightarrow & (\text{input}_R \doteq u_1 \dot{\vee} \text{input}_R \doteq u_2) \end{aligned}$$

## Basic Hash protocol in the meta-logic

Privacy expressed as  $\text{frame}@_{\tau} \sim_{\mathcal{P}, \mathcal{P}'} \text{frame}@_{\tau}$ .

- Timestamp  $\tau$  refers to an arbitrary point in any arbitrary execution trace.
- Macro  $\text{frame}@_{\tau}$  stands for the sequence of all protocol outputs before  $\tau$ .
- Protocol  $\mathcal{P}$  = multiple tags playing multiple sessions.
- Protocol  $\mathcal{P}'$  = multiple tags playing single sessions.

Authentication expressed as meta-logic formula:

$$\forall i. \text{snd}(\text{input}@_{\tau}) = \text{h}(\text{fst}(\text{input}@_{\tau}), \text{k}(i)) \Rightarrow \exists j. \text{T}(i, j) < \tau \wedge \text{input}@_{\tau} = \text{output}@_{\text{T}(i, j)}$$

- Macro  $\text{input}@_{\tau}$  defined as  $\text{att}(\text{frame}@_{\tau})$ . Macro  $\text{output}@_{\tau}$  defined depending on  $\tau$ .
- Indices  $i, j$  are elements of an arbitrary finite set.
- Total order  $<$  on timestamps corresponds to execution order.

## Meta-logic overview

We are now reasoning **over all traces** (all **trace models**  $\mathbb{T}$ )  
and all implementations of functions (all computational models  $\mathcal{M}$ ).

Meta-logic term $t$	$\xrightarrow{\mathbb{T}}$	base logic term $(t)_{\mathcal{P}}^{\mathbb{T}}$	$\xrightarrow{\mathcal{M}}$	PPTM returning bitstring
Local meta-logic formula $\phi$	$\xrightarrow{\mathbb{T}}$	base logic term $(\phi)_{\mathcal{P}}^{\mathbb{T}}$	$\xrightarrow{\mathcal{M}}$	PPTM returning boolean
Global meta-logic formula $\Phi$	$\xrightarrow{\mathbb{T}}$	base logic formula $(\Phi)^{\mathbb{T}}$	$\xrightarrow{\mathcal{M}}$	SAT/UNSAT

### Important points

- Local formula quantifications only over indices and timestamps:  
translated to finite disjunctions and conjunctions once  $\mathbb{T}$  fixed.
- Translation of macros depends on  $\mathcal{P}$ : definition of logic tied to a class of protocols.

## Proving meta-logic formulas

Custom proof system:

- hides probabilistic reasoning when proving local formulas;
- most axioms are liftings of base logic axioms.

## Proving meta-logic formulas

Custom proof system:

- hides probabilistic reasoning when proving local formulas;
- most axioms are liftings of base logic axioms.

### Freshness in base logic

$(t \dot{\neq} n) \sim \text{true}$  is valid for any closed term  $t$  that doesn't contain  $n$ .

### Freshness in meta-logic

Local meta-logic formula  $\phi \Rightarrow t \neq n(i)$  is valid if  
for any  $\mathbb{T}$  such that  $(\phi)^{\mathbb{T}} \neq \text{false}$ ,  $(n(i))^{\mathbb{T}}$  does not occur in  $(t)^{\mathbb{T}}$ .

### Example (Basic Hash)

$\forall i, j, \tau. \tau < T(i, j) \Rightarrow n_t(i, j) \neq \text{fst}(\text{input}@_{\tau})$

# Outline

- 1 Running example
- 2 Base logic
- 3 Meta-logic
- 4 Higher-order logic**
- 5 Conclusion

## From machines to random variables

Terms of the old base logic:  
probabilistic polynomial-time machines.

$$\llbracket t \rrbracket_{\mathcal{M}}(\eta, \rho) \in \{0, 1\}^*$$

Terms of the new logic:  
 *$\eta$ -indexed families of random variables.*

$$\llbracket t \rrbracket_{\mathcal{M}}(\eta, \rho) \in \llbracket \tau \rrbracket_{\mathcal{M}}$$



## From machines to random variables

Terms of the old base logic:  
probabilistic polynomial-time machines.

$$\llbracket t \rrbracket_{\mathcal{M}}(\eta, \rho) \in \{0, 1\}^*$$

Terms of the new logic:  
 *$\eta$ -indexed families of random variables.*

$$\llbracket t \rrbracket_{\mathcal{M}}(\eta, \rho) \in \llbracket \tau \rrbracket_{\mathcal{M}}$$

### Non-PTIME function symbols

E.g. talking about discrete logarithm is useful in specifications and proofs.

# From machines to random variables

Terms of the old base logic:  
probabilistic polynomial-time machines.

$$\llbracket t \rrbracket_{\mathcal{M}}(\eta, \rho) \in \{0, 1\}^*$$

Terms of the new logic:  
 $\eta$ -indexed families of random variables.

$$\llbracket t \rrbracket_{\mathcal{M}}(\eta, \rho) \in \llbracket \tau \rrbracket_{\mathcal{M}}$$

## Non-PTIME function symbols

E.g. talking about discrete logarithm is useful in specifications and proofs.

## Arbitrary quantifiers

Easily defined since terms need not be computable:

$$\llbracket \forall x : \alpha. \phi \rrbracket_{\mathcal{M}}^{\sigma}(\eta, \rho) = \text{true} \quad \text{when} \quad \llbracket \phi \rrbracket_{\mathcal{M}}^{\sigma, x \mapsto a}(\eta, \rho) = \text{true} \quad \text{for all } a \in \llbracket \alpha \rrbracket_{\mathcal{M}}.$$

## Going higher-order

Quantifications allow to express reasoning steps internally:

$$\forall x, y : \text{message}. \quad \text{fst } \langle x, y \rangle = x$$

## Going higher-order

Terms are arbitrary typed  $\lambda$ -terms: functions, functions taking functions taking ...

Not for fun, but because it provides simple, uniform foundations.

Quantifications allow to express reasoning steps internally:

$\forall x, y : \text{message}. \quad \text{fst } \langle x, y \rangle = x$

$\forall f : \alpha \rightarrow \beta, c : \text{bool}, x, y : \alpha. \quad (\text{if } c \text{ then } f \ x \ \text{else } f \ y) = (f \ (\text{if } c \ \text{then } x \ \text{else } y))$

## Going higher-order

Terms are arbitrary typed  $\lambda$ -terms: functions, functions taking functions taking ...

Not for fun, but because it provides simple, uniform foundations.

Quantifications allow to express reasoning steps internally:

$\forall x, y : \text{message.}$

$\text{fst } \langle x, y \rangle = x$

$\forall f : \alpha \rightarrow \beta, c : \text{bool}, x, y : \alpha.$

$(\text{if } c \text{ then } f \ x \ \text{else } f \ y) = (f \ (\text{if } c \ \text{then } x \ \text{else } y))$

$\forall p : \text{timestamp} \rightarrow \text{bool.}$

$(\forall \tau. (\forall \tau'. \tau' < \tau \Rightarrow p \ \tau') \Rightarrow p \ \tau) \Rightarrow \forall \tau. p \ \tau$

$\forall p : \alpha \rightarrow \text{bool.}$

$(\text{not } (\exists a : \alpha. p \ a)) = (\forall a : \alpha. \text{not } (p \ a))$

## Generalizing indistinguishability

Global formulas are still first-order formulas, but over higher-order terms.

Arbitrary predicates can be considered:

- $\vec{u} \sim \vec{v}$ : distinguisher still polynomial time.
- $\text{adv}(t)$ :  $t$  can be computed in polynomial time without access to  $\rho_h$ .
- ...

Higher-order indistinguishability allows to decompose old axioms:

$$\forall \vec{x}, \vec{y}, \vec{x}', \vec{y}'. \quad \text{adv}(\mathbf{f}) \Rightarrow (\vec{x}, \vec{y} \sim \vec{x}', \vec{y}') \Rightarrow (\vec{x}, \mathbf{f}(\vec{y}) \sim \vec{x}', \mathbf{f}(\vec{y}'))$$

## Generalizing indistinguishability

Global formulas are still first-order formulas, but over higher-order terms.

Arbitrary predicates can be considered:

- $\vec{u} \sim \vec{v}$ : distinguisher still polynomial time, takes functions (oracles) as inputs.
- $\text{adv}(t)$ :  $t$  can be computed in polynomial time without access to  $\rho_h$ .
- ...

Higher-order indistinguishability allows to decompose old axioms:

$$\forall \vec{x}, \vec{y}, \vec{x}', \vec{y}'. \quad \text{adv}(f) \Rightarrow (\vec{x}, \vec{y} \sim \vec{x}', \vec{y}') \Rightarrow (\vec{x}, f(\vec{y}) \sim \vec{x}', f(\vec{y}'))$$

$$\forall f, x, y, f', x', y'. \quad (x, f, y \sim x', f', y') \Rightarrow (x, f, y \sim x', f', y')$$

$$\forall f, x, y, x', y'. \quad \text{adv}(f) \Rightarrow (x, y \sim x', y') \Rightarrow (x, f, y \sim x', f, y')$$

## Recursive definitions

Equip our  $\lambda$ -calculus with recursive definitions to recover the macros of the old meta-logic.

### Example

$\text{frame}_{\mathcal{P}}@_{\mathcal{T}} \stackrel{\text{def}}{=} \text{if } \tau = \text{init} \text{ then empty else } \langle \text{output}_{\mathcal{P}}@_{\mathcal{T}}, \text{frame}_{\mathcal{P}}@(\text{pred } \tau) \rangle$

$\text{output}_{\mathcal{P}}@_{\mathcal{T}} \stackrel{\text{def}}{=} \text{match } \tau \text{ with } T(i, j) \mapsto \langle n_t(i, j), h(n_t(i, j), k(i)) \rangle \mid - \mapsto \text{default}$

### Benefits

- Same term can mix macros for different protocols.
- Same logic can deal with different classes of protocols, different attacker models. . .



## Advanced axioms

Axioms cannot be derived as liftings of base logic axioms.

- Recover (adaptations of) previous axioms, justifying them from first principles.
- Opportunity to generalize!

### Freshness in higher-order logic

$\phi \Rightarrow t \neq (n \ i)$  if,

for any  $\mathcal{M}, \eta, \rho$  such that  $\llbracket \phi \rrbracket_{\mathcal{M}}(\eta, \rho) = \text{true}$ ,  $\llbracket t \rrbracket_{\mathcal{M}}(\eta, \rho)$  does not sample  $n$  at  $\llbracket i \rrbracket_{\mathcal{M}}(\eta, \rho)$ .

## Advanced axioms

Axioms cannot be derived as liftings of base logic axioms.

- Recover (adaptations of) previous axioms, justifying them from first principles.
- Opportunity to generalize!

### Freshness in higher-order logic

$\phi \Rightarrow t \neq (n \ i)$  if,

for any  $\mathcal{M}, \eta, \rho$  such that  $\llbracket \phi \rrbracket_{\mathcal{M}}(\eta, \rho) = \text{true}$ ,  $\llbracket t \rrbracket_{\mathcal{M}}(\eta, \rho)$  does not sample  $n$  at  $\llbracket i \rrbracket_{\mathcal{M}}(\eta, \rho)$ .

The same can be done for cryptographic axioms, e.g. IND-CCA.

### Example

Assume  $A$  inputs  $x$  and sends back  $n$  only if  $x = \text{ok}$ .

We are now able to prove  $\text{input}@A \neq \text{ok} \Rightarrow n \neq \text{att}(\text{output}@A)$ .

## Relating global and local quantifiers

### Theorem

For any local formula  $\phi$  we have:

$$\mathcal{M}, \sigma \models \forall(x : \tau).[ \phi ] \quad \text{iff} \quad \mathcal{M}, \sigma \models [\dot{\forall} (x : \tau). \phi]$$

- On the left: for any random variable over  $[[\tau]]$ ,  $\phi$  holds with overwhelming probability.
- On the right: it is overwhelmingly true that  $\phi$  holds for any value in  $[[\tau]]$ .
- Right  $\Rightarrow$  left immediate, other direction requires carefulness with probability theory.

## Relating global and local quantifiers

### Theorem

For any local formula  $\phi$  we have:

$$\mathcal{M}, \sigma \models \forall(x : \tau).[\phi] \quad \text{iff} \quad \mathcal{M}, \sigma \models [\forall (x : \tau).\phi]$$

- On the left: for any random variable over  $[[\tau]]$ ,  $\phi$  holds with overwhelming probability.
- On the right: it is overwhelmingly true that  $\phi$  holds for any value in  $[[\tau]]$ .
- Right  $\Rightarrow$  left immediate, other direction requires carefulness with probability theory.

Same result holds for existential quantifier.

Refined version available when types and random variables are constant, which is the case when we mirror meta-logic reasoning over timestamps and indices.

# Conclusion

## Theory

- General foundation for Squirrel.
- Conceptually simpler: no more meta-logic, not tied to fixed notion of protocol.
- Complete proof system hiding much of the complexity.

## Practice

- Current version of Squirrel implements fragment of new higher-order logic.
- All past case studies carry over.
- New: examples involving corruptions, generic hybrid argument.

Learn more on our website: tech report, tutorials and interactive examples.

<https://squirrel-prover.github.io/>