

On the Symbolic Analysis of Low-Level Cryptographic Primitives:  
Modular Exponentiation and the Diffie-Hellman Protocol

**MARZIA BUSCEMI**, University of Pisa  
buscem1@di.unipi.it

**MICHELE BOREALE**, University of Florence  
boreale@dsi.unifi.it

Ottawa, FCS'03

## Security Protocol Analysis

- Formal methods can be used for the verification of security protocols: agents executing the protocols are modelled as concurrent processes, described by means of an appropriate language, i.e., the spi-calculus.
- Model checking and reachability analysis rely on exploration of the space of this system. The sequences of actions (traces) are analysed to check whether they lead to an insecure state.

## The Attacker Model

Formal methods for protocol analysis usually follow the Dolev-Yao attacker model: the (hostile) environment can manipulate any message travelling over the network, synthesize new messages by pairing, encryption, and decryption of previously known messages, etc.

**State Explosion Problem:** protocols typically generate an infinite number of traces, as an agent may expect any of the infinitely many messages the environment can send on the network at any time.

## Symbolic Analysis

- Replace the infinitely many transitions arising from an input action by a single **symbolic** transition;

- Represent the received message as a variable;

- Add **constraints** on this variable as the computation proceeds.

## Bibliography:

- [Amadio, Lugiez00], [Boreale01], [Millen, Shmatikov01], etc.

## Example

Let  $P = a(x) \cdot \text{let } y = \text{dec}_k(x) \text{ in } P'$ .

$$\langle \varepsilon, P \rangle \longleftarrow \langle a \langle x \rangle, \text{let } y = \text{dec}_k(x) \text{ in } P' \rangle$$

$$\longleftarrow \langle a \langle \{x'\}_k \rangle, P'[\{x'\}_k/x][x'/y] \rangle$$

**Note:** The decryption of  $x$  using  $k$  is resolved by unifying  $x$  and  $\{x'\}_k$  for some fresh  $x'$  (**mgv**  $\theta = \{x'/x\}_k$ );  $y$  is replaced by the result of the decryption, i.e.,  $x'$ .

## Problem

The Dolev-Yao model assumes black-box enciphering and perfect cryptography. In some cases, this is an over-simplification.

For instance, the analysis of a protocol based on modular exponentiation should take into account the operations involved in the protocol (exponentiation, product) and their 'relevant' algebraic laws; even operations not explicitly mentioned, provided they are considered feasible, must be accounted for, since an adversary might take advantage of them.

## Diffie-Hellman Protocol

The protocol has two public parameters: a large prime  $p$  and a generator  $\alpha$  for  $Z_p^*$ .  $A$  and  $B$  generates  $n_A$  and  $B$  generates  $n_B \in Z_p^*$ . Next,  $A$  and  $B$  exchange their public values:

$$1. \quad A \rightarrow B : \exp(\alpha, n_A)$$

$$2. \quad B \rightarrow A : \exp(\alpha, n_B)$$

Finally,  $A$  computes  $K = \exp(\exp(\alpha, n_B), n_A) = \exp(\alpha, n_A \times n_B)$ , and  $B$  computes  $K = \exp(\exp(\alpha, n_A), n_B) = \exp(\alpha, n_A \times n_B)$ .

$$3. \quad A \rightarrow B : \{d\}_K$$

## Contribution

We extend a symbolic method for automatic analysis of security protocols to handle modular exponentiation and, in particular, the Diffie-Hellman key exchange.

**Related Work.** Recently, some protocol analysis techniques ([MS03,

CLS03, Her02]) have been developed which take into account an attacker model with some equational theory.

Our model is less precise than others (notably [MS03]), its main limitation being a bound on the number of factors that may appear in any exponent. However, for such restricted model, we offer a decision method.



## Symbolic Framework

Our approach relies on a symbolic framework for protocol analysis in presence of a variety of crypto-functions and low-level operations.

### Main Idea

- To formalise a protocol by means of a language akin to the applied pi.
- To give a symbolic operational semantics based on **unification** and such that provides finite and effective protocol models.
- To provide a method to carry out trace analysis directly on the symbolic model. Under certain conditions on the primitives the symbolic method is proven **complete**.

## General Framework

- For  $\Sigma$  a signature of function symbols  $f, g, \dots$ ,  $T_\Sigma$  is the algebra of terms given by the grammar:

$$\zeta, \eta ::= n \mid f(\zeta)$$

- A **frame**  $\mathcal{F}$  is a triple  $(\Sigma, \mathcal{M}, \uparrow)$ , where:

–  $\Sigma$  is a signature;

–  $\mathcal{M} \subseteq \mathcal{F}_\Sigma$  is a set of **messages**  $M, N, \dots$ ;

–  $\uparrow \subseteq \mathcal{F}_\Sigma \times \mathcal{M}$  is an **evaluation relation**.

- Given an initial set of messages  $S$ , the set  $\mathcal{H}(S)$  consists of all the expressions inductively built by applying functions of  $\Sigma$  to elements of  $S$ . A message  $M$  is **deducible** from  $S$ ,  $S \vdash M$ , if and only if  $\exists \zeta \in \mathcal{H}(S) : \zeta \uparrow M$ .

## A Frame for Diffie-Hellman Protocol (1/3)

SIGNATURE	$\Sigma = \{ \alpha, \text{unit}, 1, \cdot \}, \{ \cdot \}, \text{dec}(\cdot)(\cdot), \text{exp}(\cdot, \cdot), \text{root}(\cdot, \cdot), \cdot \times \cdot, \text{mult}(\cdot, \cdot), \text{inv}(\cdot), \text{inv}'(\cdot), (\cdot)^{-1} \}$
FACTORS	$f ::= u \mid u^{-1}$
PRODUCTS	$F ::= 1 \mid f_1 \times \dots \times f_k$
KEYS	$K, H ::= f \mid \text{exp}(\alpha, F)$
MESSAGES	$M, N ::= F \mid K \mid \{M\}_K$

## A Frame for Diffie-Hellman Protocol (2/3)

		(UNIT <sub>2</sub> )
	$1 \rightsquigarrow \text{unit}$	
		(UNIT <sub>1</sub> )
	$\zeta \rightsquigarrow \zeta \times \text{unit}$	
		(INV <sub>4</sub> )
	$\text{inv}'(\zeta) \times \zeta \rightsquigarrow \text{unit}$	
		(INV <sub>3</sub> )
	$1 \rightsquigarrow \zeta$	
		(INV <sub>2</sub> )
	$\zeta \rightsquigarrow \text{inv}'(\zeta^{-1})$	
		(INV <sub>1</sub> )
$1 \leq n$	$\text{inv}(\zeta_1 \times \dots \times \zeta_n) \rightsquigarrow \text{inv}'(\zeta_1) \times \dots \times \text{inv}'(\zeta_n)$	
$1 \leq k < n \leq l$	$\text{mult}(\zeta_1 \times \dots \times \zeta_k, \zeta_{k+1} \times \dots \times \zeta_n) \rightsquigarrow \zeta_{i_1} \times \dots \times \zeta_{i_n}$	(MULT)

## A Frame for Diffie-Hellman Protocol (3/3)

(DEC)  $\text{dec}_\eta(\{\zeta\}_\eta) \rightsquigarrow \zeta$

(EXP)  $\text{exp}(\text{exp}(\zeta, \eta), \zeta) \rightsquigarrow \text{exp}(\zeta, \text{mult}(\eta, \zeta))$

(ROOT)  $\text{root}(\text{exp}(\zeta, \eta), \zeta) \rightsquigarrow \text{exp}(\zeta, \text{mult}(\eta, \text{inv}(\zeta)))$

(CTX) 
$$\frac{\zeta \rightsquigarrow \zeta'}{c[\zeta] \rightsquigarrow c[\zeta']}$$

EVALUATION  $\zeta \uparrow \eta \quad \# \quad \zeta \rightsquigarrow_* \eta$

## Discussion

The relation  $\rightsquigarrow$  is terminating but not confluent. In fact, the non-determinism of  $\rightsquigarrow$  is intended to model the commutativity and the associativity of the product operation, as reflected in the rule (MULT).

### Model Restrictions:

1. a fixed upper bound ( $l$ ) on the number of factors;
2. product and inverse operations cannot be applied to exponentials and to encrypted terms;
3. exponentiation starts from the basis  $\alpha$ , and exponents can only be products.

**Note.** Starting from a term obeying the above conditions, an attacker is capable of 'deducing' all AC variants of the message represented by that term.

## Process Syntax

$A, B ::= \mathbf{0}$	(null)
$a(x).A$	(input)
$\bar{a}\langle \zeta \rangle.A$	(output)
$\text{let } y = \zeta \text{ in } A$	(evaluation)
$[\zeta = \eta]A$	(matching)
$A \parallel B$	(parallel composition)
$(\text{new } a)A$	(restriction)

## Diffie-Hellman Specification

$$\begin{aligned} A &= (\text{new } n_A) \text{ a1} \langle \text{exp}(\alpha, n_A) \rangle \cdot \text{a2}(x) \cdot \text{let } z = \text{exp}(x, n_A) \text{ in } \overline{\text{a3}} \langle \{d\}^z \rangle \cdot 0 \\ B &= (\text{new } n_B) \text{ b1}(y) \cdot \overline{\text{b2}} \langle \text{exp}(\alpha, n_B) \rangle \cdot \text{let } w = \text{exp}(y, n_B) \text{ in } \text{b3}(t) \cdot \\ &\quad \text{let } t' = \text{dec}_w(t) \text{ in } B' \\ P &= A \parallel B. \end{aligned}$$



## Concrete Operational Semantics

- An **action** is a term of the form  $a\langle M \rangle$  (**input** action) or  $\bar{a}\langle M \rangle$  (**output** action).
  - A **trace** is a sequence of actions such that any input message can be synthesized from the knowledge the environment has previously acquired. Formally, it is a closed string  $s$  such that
$$\forall s_1, s_2, a\langle M \rangle, \quad \text{if } s = s_1 \cdot a\langle M \rangle \cdot s_2 \text{ then } s_1 \vdash M.$$
  - A **configuration**  $\langle s, P \rangle$  is a pair consisting of a trace  $s$  and a process  $P$ .
- Note:**  $s \vdash M$  abbreviates  $\text{msg}(s) \vdash M$ .

## Transition System

Basic reduction rules:

$$\text{(INP)} \quad \langle s, a(x).P \rangle \longrightarrow \langle s \cdot a \langle M \rangle, P[M/x] \rangle \quad s \vdash M, M \text{ closed}$$

$$\text{(OUT)} \quad \langle s, \bar{a} \langle \zeta \rangle . P \rangle \longrightarrow \langle s \cdot \bar{a} \langle M \rangle, P \rangle \quad \zeta \uparrow M, M \text{ closed}$$

**Note:** Rule (INP) makes the transition relation **infinite-branching**, since  $M$  ranges over the infinite set of deducible messages.

## Transition System (Cont.)

Other rules:

$$\begin{array}{l}
 \langle s, \text{let } y = \zeta \text{ in } P \rangle \longleftarrow \langle s, P[M/y] \rangle \quad \zeta \uparrow M, M \text{ closed} \quad (\text{LET}) \\
 \langle s, P \rangle \longleftarrow \langle s, P[\eta] \rangle \quad \zeta \uparrow M, \eta \uparrow N, M = N \quad (\text{MATCH}) \\
 \frac{\langle s, P \rangle \longleftarrow \langle s', P' \rangle \quad \langle \tilde{O} \parallel P \rangle \longleftarrow \langle \tilde{O} \parallel P' \rangle}{\langle s, P \rangle \longleftarrow \langle s', P' \rangle} \quad (\text{PAR})
 \end{array}$$

## Security Properties: Specification

Properties are expressed in terms of **correspondence assertions**:

- $\alpha \rightarrow \beta$  in  $s$  if action  $\alpha$  always occurs prior to action  $\beta$  in  $s$ ;
- $s \models \alpha \rightarrow \beta$  if, for each ground substitution  $\rho$ ,  $\alpha\rho \rightarrow \beta\rho$  in  $s$ .
- $C \models \alpha \rightarrow \beta$  if, for all traces  $s$  generated by  $C$ ,  $s \models \alpha \rightarrow \beta$ .

Correspondence assertions are suited for security properties, such as secrecy and authentication.

$$\langle \varepsilon, P \parallel g(x).\mathbf{0} \rangle \models \perp \rightarrow \langle p \rangle g$$

The property that the protocol  $P$  in the above example should not leak the datum  $d$  can be expressed also by saying that the adversary will never be capable of synthesising  $d$ , without prior knowledge of it.

This can be formalised by considering an 'absurd' action  $\perp$  (that is nowhere used in agent expression) and by extending  $P$  with a 'guardian' process:  $g(x).\mathbf{0}$ . The **security** property to check is then:

## An Example: Security

## Symbolic Semantics

We equip the frame for modular exponentiation with a **symbolic** evaluation relation  $(\downarrow^s)$ , which is in agreement with its concrete counterpart  $(\downarrow)$ . Intuitively,  $\zeta \downarrow^s \eta$  means that  $\zeta$  evaluates to  $\eta$  under all instances  $\rho$  of  $\theta$ . The relation  $\downarrow^s$  is presented in the next slide.

A **symbolic trace**  $\sigma$  is a trace that may contain variables.

A **symbolic configuration**  $\langle \sigma, A \rangle_s$  is a pair composed by a symbolic trace  $\sigma$  and an agent  $A$ .

## Symbolic Evaluation Relation (1/2)

		(UNIT <sup>s</sup> )
	$\text{unit} \times \zeta \xrightarrow{s} \zeta$	
	$\text{unit} \times \zeta \xrightarrow{s} \zeta$	(UNIT <sup>s</sup> )
	$\text{inv}'(\zeta) \times \eta \xrightarrow{s} \text{unit}$	(INV4 <sup>s</sup> )
	$\text{inv}'(\zeta) \xrightarrow{s} \zeta^{-1}$	(INV3 <sup>s</sup> )
	$\text{inv}'(\zeta) \xrightarrow{s} \zeta$	(INV2 <sup>s</sup> )
	$\text{inv}'(\zeta) \times \dots \times \text{inv}'(\zeta) \xrightarrow{s} \theta$	(INV1 <sup>s</sup> )
$\left. \begin{array}{l} \text{inv}'(\zeta) \times \dots \times \text{inv}'(\zeta) \xrightarrow{s} \theta \\ 1 \leq n \leq l, \end{array} \right\}$	$\theta$	(MULT <sup>s</sup> )
$\left. \begin{array}{l} \text{inv}'(\zeta) \times \dots \times \text{inv}'(\zeta) \xrightarrow{s} \theta \\ 1 \leq k < n \leq l, \end{array} \right\}$	$\text{mult}(\zeta_1, \zeta_n) \xrightarrow{s} \theta$	
$\left. \begin{array}{l} \text{inv}'(\zeta) \times \dots \times \text{inv}'(\zeta) \xrightarrow{s} \theta \\ 1 \leq n \leq l, \end{array} \right\}$	$\text{mult}(\zeta_1, \zeta_n) \xrightarrow{s} \theta$	

## Symbolic Evaluation Relation (2/2)

$$\text{(DEC}_s\text{)} \quad \text{dec}_\eta(\zeta) \stackrel{\theta}{\rightsquigarrow}_s x_1 \theta \quad \theta = \text{mgu}(\zeta = \{x_1\}x_2, \eta = x_2)$$

$$\text{(EXP1}_s\text{)} \quad \text{exp}(x, \zeta) \stackrel{\theta}{\rightsquigarrow}_s \text{exp}(\alpha, \text{mult}(x_1, \zeta)) \quad \theta = [\text{exp}(\alpha, x_1)]/x]$$

$$\text{(EXP2}_s\text{)} \quad \text{exp}(\text{exp}(\xi, \eta), \zeta) \stackrel{\theta}{\rightsquigarrow}_s \text{exp}(\xi, \text{mult}(\eta, \zeta))$$

$$\text{(ROOT1}_s\text{)} \quad \text{root}(x, \zeta) \stackrel{\theta}{\rightsquigarrow}_s \text{exp}(\alpha, \text{mult}(x_1, \text{inv}(\zeta))) \quad \theta = [\text{exp}(\alpha, x_1)]/x]$$

$$\text{(ROOT2}_s\text{)} \quad \text{root}(\text{exp}(\xi, \eta), \zeta) \stackrel{\theta}{\rightsquigarrow}_s \text{exp}(\xi, \text{mult}(\eta, \text{inv}(\zeta)))$$

$$\text{(CTX}_s\text{)} \quad \frac{\zeta \stackrel{\theta}{\rightsquigarrow}_s \zeta'}{C[\zeta] \stackrel{\theta}{\rightsquigarrow}_s C[\zeta']}$$

SYMBOLIC EVALUATION  $\zeta \uparrow_\theta \eta$  iff  $\zeta \stackrel{\theta}{\rightsquigarrow}_s \theta_1 \dots \theta_n$  and  $\theta = \theta_1 \dots \theta_n$





## Symbolic Semantics: Example

Let  $P = \bar{a}\langle k \rangle \cdot a(x) \cdot \text{let } z = \text{root}(x, k) \text{ in } P'$ . Then,

$$\langle \varepsilon, P \rangle_s \xrightarrow{s} \langle \bar{a}\langle k \rangle \cdot a(x), \theta \rangle_s \xrightarrow{s} \langle \bar{a}\langle k \rangle \cdot a(x), \theta, P'\theta\theta' \rangle_s,$$

with  $\theta = [\text{exp}(\alpha, x_1)/x] [x/(I_{x_1}) \text{ fresh}]$  and  $\theta' = [\text{exp}(\alpha, x_1 \times k_{-1})/z]$ .

## Symbolic vs. Concrete Semantics

**Theorem (soundness and completeness).** Let  $C$  be an initial configuration and  $s$  a trace. Then  $C$  generates  $s$  if and only if there is  $\sigma$  s.t.  $C$  symbolically generates  $\sigma$  and  $s$  is an instance of  $\sigma$ .

By this result, the task of checking correspondence assertions is reduced to the analysis of the **finately many** symbolic traces the initial configuration can generate.

## Trace Consistency

Many symbolic traces are **inconsistent**, i.e., they cannot be instantiated to any concrete trace. For example,  $\langle \{x\}^k \cdot \bar{a}\langle x \rangle$  is inconsistent.

**Proposition** Let  $\sigma$  be a symbolic trace. Then there exists a finite set of

traces **Refinement**( $\sigma$ ), which are instances of  $\sigma$  and have the following

property: for any  $s$ ,  $s$  is a solution of  $\sigma$  if and only if  $s$  is a solution of some

$\sigma' \in \text{Refinement}(\sigma)$ .

Roughly, the set **Refinement**( $\sigma$ ) is computed by repeatedly unifying each

input message in  $\sigma$  to terms that can be synthesized out of previous

messages in  $\sigma$ .

## Refinement

**Definition (Refinement)** ( $\succ$ ) is the least binary relation over symbolic traces given by the following rules, assuming that (1)  $\sigma'$  is the longest prefix of  $\sigma$  which is a trace, (2)  $\sigma = \sigma' \cdot a \langle M \rangle \cdot \sigma''$ , for some  $\sigma''$ , (3)  $N, N' \notin \mathcal{V} \cup \mathcal{V}'$ , (4)  $\theta \neq \varepsilon$ .

$$\frac{N' \in \mathbf{b}(\sigma') \quad M = C[N] \quad \theta = \text{mgu}(N, N')}{\sigma \succ \sigma\theta\theta_0} \text{ (REF1)}$$

$$\frac{\sigma \succ \sigma[x/x]}{x \in \mathbf{v}(M)} \text{ (REF2)}$$

where  $\theta_0 = [x/\hat{x} \mid x \in \mathbf{v}(\sigma) \text{ and } |(\sigma\theta) \setminus \hat{x}| > |\sigma \setminus \hat{x}|]$ .

For any symbolic trace  $\sigma$ ,

$$\text{Refinement}(\sigma) = \{ \sigma' \mid \sigma \succ \sigma' \text{ and } \sigma' \text{ is a trace} \}.$$

## Verification Method

$\overline{M(C, \alpha \mapsto \beta)}$

1. compute  $\text{Mod}_C = \{\sigma \mid C \not\models \sigma\}$ ;
2. **foreach**  $\sigma \in \text{Mod}_C$  **do**
3. **foreach** action  $\gamma$  in  $\sigma$  **do**  
if  $\exists \theta = \text{mgu}(\beta, \gamma)$  **and**  $\exists \sigma' = (\sigma\theta') \in \text{SF}(\sigma)$  s.t.  
 $\alpha\theta'$  does not occur prior to  $\beta\theta'$  in  $\sigma'$
6. **then return**(No,  $\sigma'$ );
7. **return**(Yes);

**Theorem.**

- if  $M(C, \alpha \mapsto \beta)$  returns **Yes** then  $C \models \alpha \mapsto \beta$ .
- if  $M(C, \alpha \mapsto \beta)$  returns **No**,  $\sigma$  then  $C \not\models \alpha \mapsto \beta$ .

## Diffie-Hellman Protocol Analysis

1. The symbolic model  $\text{Mod}^{CDH}$  is computed (in practice, symbolic traces are generated and checked 'on-the-fly').

2. The following symbolic trace  $\sigma$  is considered:

$$s = \underline{a1} \langle \text{exp}(\alpha, n_A) \rangle \cdot \underline{a2} \langle \text{exp}(\alpha, x_0) \rangle \cdot \underline{a3} \langle \{d\}^k \rangle \cdot g \langle t \rangle$$

**3,4.** Action  $\gamma = g \langle t \rangle$  is found that unifies with  $\beta = g \langle d \rangle$ , via  $\theta = [d/t]$ .

**5.** The set **Refinement**( $\sigma\theta$ ) =  $\{\sigma'\}$  is computed where  $\sigma' = \sigma\theta\theta'$ , and  $\theta' = [x_0/x_0]$ . By the above proposition,  $\sigma'$  is a consistent trace.

**6.** Action  $\perp$  does not appear in  $\sigma'$ , hence,

**7.** (No,  $\sigma'$ ) is returned.

An attack to the protocol, i.e. a ground trace  $s$  such that the  $\langle \varepsilon, S \rangle \searrow s$  and  $s \neq \text{Secret}(d) = \perp \mapsto g \langle d \rangle$ , is then retrieved as  $s = \sigma [n_1/x_0]$ .

## Conclusion and Future Work

We have presented a model for the analysis of protocols built around shared-key encryption and modular exponentiation. The model is less precise than others. However, for such restricted model, we offer a decision method.

We believe the method is effective in practice, since the symbolic model is compact and the refinement procedure is invoked only on demand. We are in process of integrating our technique into the STA analysis tool.

<http://www.dsi.unifi.it/~boreale/tool.html>

Our technical development has been confined to multiplication and exponentiation, but we are confident that the approach can be extended to other low-level cryptographic primitives (RSA encryption).