RAPPORT TECHNIQUE EVA

Abstract Interpretation for Secrecy using Patterns

Date	:	November 21, 2002
Authors	:	L. Bozga, Y. Lakhnech, M. Périn
Title	:	Abstract Interpretation for Secrecy using Patterns
Report number / Version	:	5/ 2

TRUSTED LOGIC S.A.
5 rue du Bailliage
78000 Versailles, France
www.trusted-logic.fr

Laboratoire Spécification Vérification
CNRS UMR 8643, ENS Cachan
61, avenue du président-Wilson
94235 Cachan Cedex, France
www.lsv.ens-cachan.fr

Laboratoire Verimag
CNRS UMR 5104,
Univ. Joseph Fourier, INPG
2 av. de Vignate,
38610 Gières, France
www-verimag.imag.fr

Abstract Interpretation for Secrecy using Patterns

L. Bozga, Y. Lakhnech, M. Périn

November 21, 2002

1 Introduction

At the heart of almost every computer security architecture is a set of cryptographic protocols that use cryptography to encrypt and sign data. They are used to exchange confidential data such as pin numbers and passwords, to authentify users or to guarantee anonymity of principals. It is well known that even under the idealized assumption of perfect cryptography, logical flaws in the protocol design may lead to incorrect behavior with undesired consequences. Maybe the most prominent example showing that cryptographic protocols are notoriously difficult to design and test is the Needham-Schroeder protocol for authentification. It has been introduced in 1978 [29]. An attack on this protocol has been found by G. Lowe using the CSP model-checker FDR in 1995 [23]; and this led to a corrected version of the protocol [24]. Consequently there has been a growing interest in developing and applying formal methods for validating cryptographic protocols [25, 13]. Most of this work adopts the so-called Dolev and Yao model of intruders. This model assumes perfect cryptographic primitives and a nondeterministic intruder that has total control of the communication network and capacity to forge new messages. It is known that reachability is undecidable for cryptographic protocols in the general case [18], even when a bound is put on the size of messages [17]. Because of these negative results, from the point of view of verification, the best we can hope for is either to identify decidable sub-classes as in [5, 32, 26] or to develop correct but incomplete verification algorithms as in [28, 22, 20].

In this paper, we present a correct but, in general, incomplete verification algorithm to prove secrecy without putting any assumption on messages nor on the number of sessions. Proving secrecy means proving that secrets, which are pre-defined messages, are not revealed to unauthorized agents. The main contribution of our paper is an original method for proving that a secret is not revealed by a set of rules that model how the initial set of messages known by the intruder evolves. In contrast to almost all existing methods, we do not try to compute or approximate the sets of messages that can be known by the intruder. Our algorithm is rather based on the notion of "the secret being *guarded*, or *kept under the hat* of a safe message". For example, suppose that our secret is the nonce N_B and that the key K_B^{-1} – the inverse of K_B – is not known by the intruder. Then, any message that contains N_B and that is encrypted with K_B is

a guard for N_B. For instance, N_B is guarded in the message $\{\{N_A, N_B\}_{K_B}\}_{K_I}$ by the sub-message $\{N_A, N_B\}_{K_B}$. The idea is then to compute a set of guards that will keep the secret unrevealed in all sent messages and such that the inverses of the keys used in this set are also secrets. The difficulty here is that this set is, in general, infinite. Therefore, we use *terms* to represent sets of guards. For instance the term $\{x, x_s\}_{K_B}$ says that the secret will be guarded in any message $\{m, m'\}_{K_B}$, where the secret is not a sub-message of m but may be a sub-message of m'. The problem is, however, that there might be a rule $\{I, y\}_{K_B} \rightarrow y$ that will send y unencrypted to the intruder if (s)he produces the message $\{I, y\}_{K_B}$. Hence, the term $\{x, x_s\}_{K_B}$ will guard the secret except when x is I. Thus, our abstract domain consists of pairs $(\mathcal{G}, \mathcal{B})$ of terms. Those in $\mathcal{G}$ correspond to good messages that guard the secrets, whereas those in $\mathcal{B}$ denote "bad exceptions", that is, the particular instances of terms in $\mathcal{G}$ that do not guard the secrets. A weakness of terms is, however, that variables appear only at the leafs, and hence, they do not allow to describe, for instance, the set of terms that share a common sub-term. To mitigate this weakness, we introduce *pattern terms*, that is, terms with an interpreted constructor, Sup, where a term $Sup(t)$ is meant for the set of terms that contain t as sub-term. To use these pattern terms in our verification method, we solve a generalized form of the unification problem and introduce a widening operator.

We developed a prototype in Caml that implements this method. We have been able to verify several protocols taken from [10] including, for instance, Needham-Schroeder-Lowe (0.03 sec), Yahalom (12.67 sec), Otway-Rees (0.01 sec), Skeme (0.06 sec) and Kao-Chow (0.78 sec).

Related work

Decidability Dolev, Even and Karp introduced the class of ping-pong protocols and showed its decidability. The restriction put on these protocols are, however, too restrictive and none of the protocols of [10] falls in this class. Recently, Comon, Cortier and Mitchell [12] extended this class allowing pairing and binary encryption while the use of nonces still cannot be expressed in their model. Reachability is decidable for the bounded number of sessions [5, 32, 26] or when nonce creation is not allowed and the size of messages is bounded [17]. These assumtions are in practice not always justified or rather rarely justified.

Security protocols debugging For the general case, model-checking tools have been applied to discover flaws in cryptographic protocols [23, 33, 27, 11]. The tool described in [11] is a model-cheker dedicated to cryptograhic protocols. Most of these methods bound the number of sessions to be considered as well as the size of the messages.

Deductive methods Methods based on induction and theorem proving have been developed (e.g. [30, 8, 15]). These methods are general, i.e., can handle unbounded protocols, but are not automatic with exception of [15]. This work can be seen as providing a general proof strategy for verifying security protocols. The strategy is implemented

on the top of PVS and allows to handle many known protocols. The termination of this strategy is, however, not guaranteed.

Logic programming based methods These methods are based on modeling protocols in Horn Logic, e.g. as Prolog programs, as in [35, 7, 3] and developing suitable proof strategies. The main difficulty in these methods is that termination of the analysis is not guaranteed.

Typing and Abstraction-based methods Type systems and type-checking have also been advocated as a method for verifying security protocols (e.g. [1, 21, 2]). Although, these techniques can handle unbounded protocols they are as far as we know not yet completely automatic. Closest to our work are partial algorithms based on abstract interpretation and tree automata that have been presented in [28, 22, 20]. The main difference is, however, that we do not compute the set of messages that can be known by the intruder but a set of guards as explained above. Our method can handle unbounded protocols fully automatically with the price that it may discover false attacks. Interesting enough is that this does not happen on any of the practical protocols we tried (see Table 1 in Section 6.3). We are actually working on a method that allows to analyse possible attacks.

2 Preliminary

If $n \in \mathbb{N}$ then we denote by $\mathbb{N}_n$ the set $\{1, \cdots, n\}$. Let $\mathcal{X}$ be a countable set of variables and let $\mathcal{F}^i$ be a countable set of function symbols of arity i, for every $i \in \mathbb{N}$. Let $\mathcal{F} = \bigcup_{i \in \mathbb{N}} \mathcal{F}^i$. The set of *terms over* $\mathcal{X}$ *and* $\mathcal{F}$, denoted by $\mathcal{T}(\mathcal{X}, \mathcal{F})$, is the smallest set containing $\mathcal{X}$ and closed under application of the function symbols in $\mathcal{F}$, i.e., $f(t_1, \cdots, t_n)$ is a term in $\mathcal{T}(\mathcal{X}, \mathcal{F})$, if $t_i \in \mathcal{T}(\mathcal{X}, \mathcal{F})$, for $i = 1, \cdots, n$, and $f \in \mathcal{F}^n$. As usual, function symbols of arity 0 are called constant symbols. *Ground terms* are terms with no variables. We denote by $\mathcal{T}(\mathcal{F})$ the set of ground terms over $\mathcal{F}$.

A tree tr is a function from a finite subset of ω^* to $\mathcal{X} \cup \mathcal{F}$ such that $tr(u) \in \mathcal{F}^n$ iff $u \cdot j \in dom(tr)$, for every $j \in \{0, \cdots, n-1\}$. We identify terms with trees by associating to each term t a tree $Tr(t)$ as follows: (1) if x is a variable, then $dom(Tr(x)) = \{\epsilon\}$ and $Tr(x)(\epsilon) = x$, (2) if f is a constant symbol, then $dom(Tr(f)) = \{\epsilon\}$ and $Tr(f)(\epsilon) = f$ and (3) for a term $t = f(t_0, \cdots, t_{n-1})$, $dom(Tr(t)) = \{\epsilon\} \cup \bigcup_{i=0}^{n-1} i \cdot dom(Tr(t_i))$, where $\cdot$ is word concatenation extended to sets, $Tr(t)(\epsilon) = f$ and $Tr(t)(i \cdot u) = Tr(t_i)(u)$. Henceforth, we tacitly identify the term t with $Tr(t)$. The elements of $dom(t)$ are called *positions* in t. We use $\prec$ to denote the prefix relation on ω^*. We write $t(p)$ to denote the symbol at position p in t and $t_{|p}$ to denote the subterm of t at position p, which corresponds to the tree $t_{|p}(x) = t(p \cdot x)$ with $x \in dom(t_{|p})$ iff $x \cdot p \in dom(t)$. We write $t \preceq t'$ (resp. $t \prec t'$) to denote that t is a sub-term (resp. proper sub-term) of t' . Moreover, $t[t'/p]$ denotes the term obtained from t by substituting t' for $t_{|p}$. The set of variables in a term t is defined as usual and is denoted by $var(t)$.

3 Models for cryptographic protocols

In this section, we describe how we model cryptographic protocols and give a precise definition of the properties we want to verify. We begin by describing the messages involved in a protocol model.

3.1 Messages

The set of messages is denoted by $\mathcal{T}(\mathcal{F})$ and contains terms constructed from constant symbols and the function symbols $\mathbf{encr} : \mathcal{T}(\mathcal{F}) \times \mathcal{K} \to \mathcal{T}(\mathcal{F})$ and $\mathbf{pair} : \mathcal{T}(\mathcal{F}) \times \mathcal{T}(\mathcal{F}) \to \mathcal{T}(\mathcal{F})$. Constant symbols are also called atomic messages and are defined as follows:

1. *Principal names* are used to refer to the principals in a protocol. The set of all principals is $\mathcal{P}$.

2. *Nonces* can be thought as randomly generated numbers. As no one can predict their values, they are used to convince for the freshness of a message. We denote by $\mathcal{N}$ the set of nonces.

3. *Keys* are used to encrypt messages. We have three key constructors pbk, pvk and smk of type : $(\mathcal{N} \times \mathcal{P}^*) \cup \mathcal{P}^* \to \mathcal{K}$, where pbk, pvk and smk stand respectively for *public*, *private* and *symmetric* keys.

   The nonce that can appear as the first parameter of a key term is used to model fresh keys. For example, if N is a nonce produce by a server, then $smk(N, A, B)$ is a *fresh symmetric session key* between A and B, whereas $smk(A, B)$ is a session independant symmetric key between A and B. The creation of fresh keys is useful to model session key; this feature is used in the Yahalom's protocol.

   The key $pbk(n, p_1, \cdots, p_r)$ is an inverse of the key $pvk(n, p_1, \cdots, p_r)$ and vice versa; and a key $smk(n, p_1, \cdots, p_r)$ is its self-inverse. If k is a key then we use k^{-1} to denote its inverse. Moreover, we model as usual, we write K_A instead of $pbk(A)$, K_A^{-1} instead of $pvk(A)$, K_{AS} instead of $smk(A, S)$, and K_{AB} with $fresh(K_{AB})$ instead of $smk(N, A, B)$ with $fresh(N)$.

   We denote by $\mathcal{K}$ the set of ground keys, i.e., $\mathcal{K} = \mathcal{T}(\mathcal{N} \cup \mathcal{P} \cup \{pbk, pvk, smk\})$.

For the sake of simpicity we left out the signatures and hash functions but we can easlily handle them in our model. Let $\mathcal{A} = \mathcal{P} \cup \mathcal{N} \cup \mathcal{K}$ and $\mathcal{F} = \mathcal{A} \cup \{\mathbf{encr}, \mathbf{pair}\}$. As usual, we write (m_1, m_2) for $\mathbf{pair}(m_1, m_2)$ and $\{m\}_k$ instead of $\mathbf{encr}(m, k)$. *Message terms* are the elements of $\mathcal{T}(\mathcal{X}, \mathcal{F})$, that is, terms over the atoms $\mathcal{A}$, a set of variables $\mathcal{X}$ and the binary function symbols **encr** and **pair**. Thus, *messages* are ground terms in $\mathcal{T}(\mathcal{X}, \mathcal{F})$.

Role terms To describe the transitions that can be performed by a principal in a session of a cryptographic protocol, we introduce *role terms*. Let $\mathcal{X}_N$ be a set of variables that range over nonces with $n, n', \ldots$ as typical variables and $\mathcal{X}_P$ be a set of variables that range over principals with $p, p_1, \ldots$ as typical variables. We assume that $\mathcal{X}_M$, $\mathcal{X}_N$ and $\mathcal{X}_P$ are pairwisely disjoint.

Role terms over $\mathcal{X}$ are terms constructed from variables in $\mathcal{X}_M \sqcup \mathcal{X}_N \sqcup \mathcal{X}_P$ using the binary functions symbols **encr** and **pair** and where constants are not allowed. More, precisely role terms are defined by the following tree grammar:

$$\begin{array}{rcl} \mathit{Key} & ::= & \mathit{pbk}(x_1, \cdots, x_r) \mid \mathit{pvk}(x_1, \cdots, x_r) \mid \mathit{smk}(x_1, \cdots, x_r) \\ & & \mathit{pbk}(n, x_1, \cdots, x_r) \mid \mathit{pvk}(n, x_1, \cdots, x_r) \mid \mathit{smk}(n, x_1, \cdots, x_r) \\ \mathit{RT} & ::= & n \mid p \mid \mathit{Key} \mid x \mid \mathbf{pair}(\mathit{RT}_1, \mathit{RT}_2) \mid \mathbf{encr}(\mathit{RT}, \mathit{Key}) \end{array}$$

where $x_1, \cdots, x_r \in \mathcal{X}_P$, $n \in \mathcal{X}_N$, $p \in \mathcal{X}_P$ and $x \in \mathcal{X}_M$.

3.2 Cryptographic Protocols - Syntax

To describe cryptographic protocols, we need to describe the transitions the principals can perform. In our setting, transitions have the form $t \to t'$, where t and t' are role terms with $\mathit{var}(t') \subseteq \mathit{var}(t)$, t is called the *guard* of the transition and t' its *action*.

Now, a cryptographic protocol is described by a parameterized session description where the parameters are the involved principals, the fresh nonces and used keys. A *session description* is then given by a tuple $(P, \mathit{tran}, \mathit{fresh})$, where

- P is a vector $(p_1, \cdots, p_r)$, $r \geq 1$, of distinct principal variables in $\mathcal{X}_P$,
- *tran* is a function that associates to each principal variable in P a finite list of transitions,
- *fresh* associates to each principal variable p in P a disjoint finite set of nonce variables in $\mathcal{X}_N$.

Example 3.1 The Needham-Schroeder protocol for authentification can be described as follows using the usual informal notation for cryptographic protocols:

$$\begin{array}{rcl} A \to B & : & \{A, N_A\}_{k_B} \\ B \to A & : & \{N_A, N_B\}_{k_A} \\ A \to B & : & \{N_B\}_{k_B} \end{array}$$

Intuitively, A plays the role of the initiator of the session; while B is a responder. In our setting it is described by the session description given in Figure 1, where $P = (p_1, p_2)$, $\mathit{fresh}(p_1) = \{n\}$ and $\mathit{fresh}(p_2) = \{n'\}$. As one can see, our description is much more detailed and elevates many of the ambiguities of the informal description. □

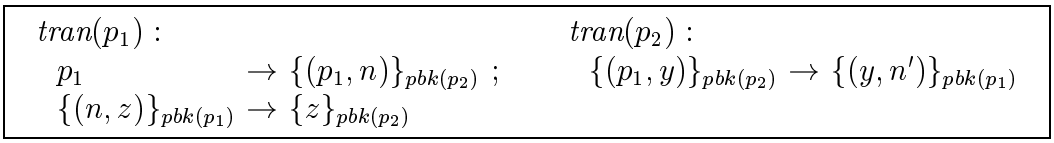

Figure 1: Needham-Schreoder protocol

3.3 The intruder model

In this section, we describe how an intruder can create new messages from already known messages. We use the most commonly used model, introduced by Dolev and Yao [16], which is given by a formal system $\vdash$. The intruder capabilities for intercepting messages and sending (fake) messages are fixed by the operational semantics. Thus, the *derivability of a message* m from a set E of messages, denoted by $E \vdash m$, is described by the following axiom and rules:

- If $m \in E$ then $E \vdash m$.
- If $E \vdash m_1$ and $E \vdash m_2$ then $E \vdash \mathbf{pair}(m_1, m_2)$. This rule is called pairing.
- If $E \vdash m$ and $E \vdash k \in \mathcal{K}$ then $E \vdash \mathbf{encr}(m, k)$. This is called encryption.
- If $E \vdash \mathbf{pair}(m_1, m_2)$ then $E \vdash m_1$ and $E \vdash m_2$. This is called projection.
- If $E \vdash \mathbf{encr}(m, k)$, $E \vdash k'$ and k and k' are inverses then $E \vdash m$. This is called decryption.

Pairing and encryption rules are called *introduction* rules while projection and decryption are called *elimination* rules. As usual, derivations in the system $\vdash$ can be seen as proof trees. Moreover, $\vdash$ can be seen as a natural deduction system. It enjoys the *normalized proof property* [31, 11] that states that if m is derivable from E then there is a proof of m from E where no elimination rule appears in the proof tree above (as a father of) any introduction rule.

It is worth noticing that the intruder can not forge any key term from the knowledge of its subterms, e.g, $A, B \nvdash smk(A, B)$. No rule are provided to the intruder to do so. Consequently, from the intruder point of view, the key terms are atomic keys.

Critical and non-critical positions We now define *critical* and *non-critical* positions in a message. The idea is that there is no way to deduce the key used for encryption from an encrypted message. So, the key position in messages of the form $\mathbf{encr}(m, k)$ is not critical and it is a safe place for a secret. The critical position corresponds to the subterm relation in the strand space model [34, 19].

Formally, given a term t, a position p in t is called *non-critical*, if there is a position q such that $t(q) = \mathbf{encr}$ and $p = q \cdot 2$; otherwise it is called *critical*. We will also use the

notation $s \in_c m$ to denote that s appears in m at a critical position, i.e., there exists $p \in dom(m)$ such that p is critical and $m_{|p} = s$.

For a term t, we use the notation $E \not\vdash t$ to denote that no instantiation of t is derivable from E, that is, for no substitution $\sigma : \mathcal{X} \to \mathcal{T}(\mathcal{F})$, we have $E \vdash \sigma(t)$.

We also use the notation $E \not\vdash^{\in_c} t$ to denote that no message derivable from E contains an instantiation of t at a critical position, that is, for every message m if $E \vdash m$ then $\sigma(t) \not\in_c m$, for any σ. The relation $\not\vdash^{\in_c}$ is naturally extended to set of terms.

3.3.1 Operational semantics

In the rest of this section, let $\mathcal{S} = (P, tran, fresh)$ be a given session description. We want to describe the behavior of the protocol described by $\mathcal{S}$ without any restriction on the numbers of sessions and principals. To do so, we need to define *instantiated transitions* and *instantiated sessions*. We use natural numbers as session identifiers. A session instance is fixed by a pair (i, π), where i is its identifier and π is a vector of principals that instantiate the principal variables $p_1, \cdots, p_r$. Therefore, we introduce the set $\text{Inst} = \mathbb{N} \times \mathcal{P}^r$. As, we impose that the variables in P are distinct, we can use $\pi(p_j)$ to refer the j^{th} principals name in the vector π, i.e., we can identify π with a function.

Session instances We assume that we have for each nonce variable $n \in \mathcal{X}_N$ and each principal $p \in \mathcal{P}$ the **bijective** function $n_p : \omega \times \mathcal{P}^r \to \mathcal{N}(p)$ such that $n_p(i, \pi) \neq m_p(i, \pi)$, if n and m are syntactically different. For short, we write $N_p^{i,\pi}$ instead of $n_p(i, \pi)$. Intuitively, we use $N_p^{i,\pi}$ as the nonce corresponding to the nonce variable n used by p in the session (i, π). In order to produce an instance of the session description we have to chose a session number and a substitution that associates a constant name to each principal variable in P.

Given $(i, \pi) \in \text{Inst}$, we generate a session instance, denoted by $(\mathcal{S})_\pi^i$, by applying the following transformations to all role terms that appear in $\mathcal{S}$:

- we replace each principal variable p by $\pi(p)$,
- each nonce variable $n \in fresh(p)$ by $N_{\pi(p)}^{i,\pi}$.

We denote by t_π^i the message term obtained from t by applying the transformations above. Then, the (i, π)-*instance* of a transition $t \to t'$ is $t_\pi^i \to (t')_\pi^i$. Given $p \in P$, we denote by $tran_\pi^i(p)$ the list of (i, π)-instantiated transitions obtained from $tran(p)$.

Example 3.2 Let $\pi = (A, B)$. Moreover, assume that $n_A(0, \pi) = N_A^0$ and $n'_B(0, \pi) = N_B^0$. Then, the $(0, \pi)$-instance of the Needham-Schroeder protocol contains the transitions given in Figure 2. □

$tran^0_\pi(A)$:		$tran^0_\pi(B)$:	
A	$\rightarrow \{(A, N^0_A)\}_{pbk(B)}$;	$\{(A, y)\}_{pbk(B)}$	$\rightarrow \{(y, N^0_B)\}_{pbk(A)}$
$\{(N^0_A, z)\}_{pbk(A)}$	$\rightarrow \{z\}_{pbk(B)}$		

Figure 2: The transitions of the $(0, \pi)$-instance.

Configurations and transitions In order to define global configurations that may arise during the protocol execution, we need to define the state of each session instance.

The *state* of a session instance is given by a pair (τ, E), the function τ associates for each $p \in P$ a list of (i, π)-*instantiated* transitions and E is a set of messages. We denote by Σ the set of session states.

A *configuration* of the protocol defined by $\mathcal{S}$ is given by a pair (ξ, E), where $dom(\xi)$ is the set of identifiers of the sessions created in the configuration, $\xi(i)$ is the state of session i and E is the set of messages intercepted by the intruder. The operational semantics of a protocol is defined by transitions on configurations. There are two sets of transitions: 1.) transitions that create new sessions and 2.) transition that are induced by transition inside sessions:

$$\frac{i \notin dom(\xi)}{(\xi, E) \rightarrow (\xi[i \mapsto (\pi, tran^i_\pi)], E)}, \text{ where } \pi \in \mathcal{P}^r.$$

$$\frac{(\tau, E) \Rightarrow (\tau', E')}{(\xi, E) \rightarrow (\xi[i \mapsto (\pi, \tau')], E')},$$

where $\xi(i) = (\pi, \tau)$ and $\Rightarrow$ is defined below.

The relation $\Rightarrow$ describes session state changes caused by firing principal transitions. We have $(\tau, E) \Rightarrow (\tau', E')$, if there is $t \rightarrow t'$ which is the first transition in $\tau(p)$ for $p \in P$ and there is a substitution $\rho : \mathcal{X} \rightarrow \mathcal{T}(\mathcal{F})$ such that $E \vdash \rho(t)$ and $E' = E \cup \{\rho(t')\}$.

Example 3.3 Consider again our running example, the Needham-Schroeder protocol, and a session between A and B, identified by 0, with principal A in the last step of the protocol. The state $\xi(0)$ of the session is described by Let the values of the parameters: $\pi(p_1) = A$ and $\pi(p_2) = B$; and by the program counter: $\tau(p_1) = \{(N^0_A, Z)\}_{K_A} \rightarrow \{Z\}_{K_B}$ and $\tau(p_2) = \epsilon$.

Moreover, let $\{N^0_A, N^0_B\}_{K_A} \in E$ then A can fire its last transition which modifies the session state: $(\tau, E) \Rightarrow (\tau', E')$, where $\tau'(p_1) = \tau'(p_2) = \epsilon$ and $E' = E \cup \{\{N^0_B\}_{K_B}\}$.

Clarifying remarks In our model the intruder has the ability to intercept any message sent by a principal and principals have no guarantee about the origin of a message. Thus, the intruder can intercept messages use them to create fake messages and deliver these to the principals. Following Bolignano [9], in our model this is realized by modeling sending of messages as adding messages to the set E and by modeling reception of messages as reading messages deducible from E. Principals use, however, the guards

of the transitions to check the genuineness of received messages. For instance, in the Needham-Schroeder example, the guard $\{(p_1, y)\}_{pbk(p_2)}$ of the transition of principal p_2 means that principal p_2 accepts any and only messages that are sent by p_1 and encrypted by $pbk(p_2)$ of a pair of messages. Consider now the guard of the second transition of p_1, namely $\{(n, z)\}_{pbk(p_1)}$. Here, p_1 refuses (and the execution blocks) if the message to be read is not an encryption by $pbk(p_1)$ of a pair whose first message is the nonce n sent in the first transition.

3.4 Secrecy modeling

A *secrecy* goal states that a designated message should not be made public. A secret is public when it is deducible from the set of messages intercepted by the intruder. In our setting, a secret is defined by a role term. For instance, in the Needham-Schroeder example a secret we want to prove is n', the nonce sent by p_2. More precisely, each session instance is associated with a secret we want to prove. Here arises the important question concerning the initial knowledge of the intruder and his ability to profit from the actions of honest participants in parallel and previous sessions. In other words, when proving that the secret associated to session i running between the participants A and B remains unrevealed, we have to take into account that an intruder can profit from a session between A and C to break the protocol. Actually, we cannot even rely on the honesty of C; she can be seen as an intruder's accomplice.

As in the previous section, let $\mathcal{S} = (P, tran, fresh)$ be a given session description.

A *secret template* is given by a role term t. Given $(i, \pi) \in \text{Inst}$ and a secret template t, let $C(E, \pi, i)$ denote the following constraint on E:

$$\text{For every nonce variable } n \in \bigcup_{p \in P} fresh(p),$$

$$E \not\vdash^{\in_c} N^{i,\pi}_{\pi(p)}.$$

Intuitively, this means that the intruder cannot know messages that contain fresh nonces.

Moreover, let $C(E)$ denote the condition:

$$\forall (i, \pi) \in \text{Inst} \cdot C(E, \pi, i).$$

We are now ready to define our secrecy property formally. The protocol described by $\mathcal{S}$ satisfies the secrecy property defined by the secret template t in the initial set E_0 of intruder's messages , denoted by $Secret(\mathcal{S}, t, E_0)$, if for every $(i, \pi) \in \text{Inst}$

$$\text{if } C(E_0),\ (\emptyset, E_0) \rightarrow^* (\xi, E) \text{ and } \xi(i) = (\pi, \tau) \text{ then } E \not\vdash t^i_\pi.$$

The definition of secrecy can be easily extended to a set T of secret templates by: $Secret(\mathcal{S}, T, E_0)$ iff $Secret(\mathcal{S}, t, E_0)$, for all $t \in T$.

4 Finite abstraction of atomic messages and sessions

Also, in this section we fix an arbitrary cryptographic protocol given by a session description $\mathcal{S} = (P, tran, fresh)$ and fix a secret s given by a role term. To prove that s is a secret, we are faced with the following problems:

1. The definition of our verification problem is a reachability problem quantified universally over all $(i, \pi) \in \text{Inst}$.
2. There is no bound on the number of sessions that can be created.
3. There is no bound on the size of the messages that occur during execution of the protocol.

In this section, we present an abstraction that copes with the first two problems. The other problem is handled in the next section. We proceed in two steps. First, we present an abstraction that is parameterized by $(i_0, \pi_0) \in \text{Inst}$, then we argue that the abstract system we obtain does not depend on the choice of (i_0, π_0). The main idea of the abstraction is as follows. Clearly, the behavior of a participant does not depend on its identity. This is simply a consequence of defining protocol sessions in a parameterized manner as we did. It also does not depend on the identifier associated to the session.

Therefore, we fix an arbitrary session where the same principal, say A, plays the (say two) different roles in the protocol. Then, we identify the intruder I with all principals other than A. Moreover, we identify all sessions in which A is not involved. Concerning the other sessions, that is, those where A is involved, we identify:

- all sessions where A plays both roles and which are different from the fixed session,
- all sessions where A plays the first role while the second role is played by an other principal,
- all sessions where A plays the second role while the first role is played by a different principal.

Identifying sessions means also identifying the nonces and keys used in these sessions. This leaves us with a system where we have unbounded number of messages as the size of the messages is not bounded. To summarize, we model a protocol as a set of transitions that can be taken in any order and any number of times. Notice that this abstraction involves considering only one honest principal and one dishonest principal (the intruder). The proof that this abstraction is safe and actually also exact is given in [14].

We now present this idea formally. Let $(i_0, \pi_0) \in \text{Inst}$ be fixed. For a concrete semantic object x, we use the notation $x^{(i_0,\pi_0)}$ to denote its abstraction, and in case (i_0, π_0) is known from the context we use $x^\sharp$.

We start by defining the abstract domains $\mathbb{N}^\sharp = \{\top, \bot\}$ and $\mathcal{P}^\sharp = \{A, I\}$ and the abstractions:

- $i^\sharp = \top$, if $(i, \pi) = (i_0, \pi_0)$ and $i^\sharp = \bot$; otherwise, and

- $p^\sharp = A$, if $p = \pi_0(p_i)$, and $p^\sharp = I$; otherwise.

We extend the abstraction of participants to vectors of participants by taking the abstractions of the components.

For keys, we take the abstract set $\mathcal{K}^\sharp$ that consists of a distinguished key K_I and the keys in $\mathcal{P}(p_1^\sharp, \cdots, p_l^\sharp)$ with $p_1^\sharp, \cdots, p_l^\sharp \in \mathcal{P}^\sharp$ and $p_j^\sharp \neq I$, for all $j \in \mathbb{N}_l$. The abstraction of a key $k(p_1, \cdots, p_n)$ is defined by:

$$k^\sharp(p_1, \cdots, p_n) = \begin{cases} k(p_1^\sharp, \cdots, p_n^\sharp); \text{ if } p_i^\sharp \neq I, \text{ for } i = 1, \cdots, n \\ K_I \qquad\qquad\qquad\quad ; \text{otherwise} \end{cases}$$

Example 4.1 If the number of roles $r = 2$ and $\pi_0 = (A, B)$ then $\mathcal{K}^\sharp$ consits of K_I and $pbk(A)$, $pvk(A)$, $pbk(A, A)$, $pvk(A, A)$, $smk(A, A)$. □

It remains to define the abstraction of nonces. The abstraction of the nonce $N_p^{i,\pi}$, denoted by $(N_p^{i,\pi})^\sharp$, is given by:

- N_I, if $\pi^\sharp = (I, \cdots, I)$,
- $N_{p^\sharp}$, if $i^\sharp = \top$ and $\pi^\sharp = \pi_0$, and
- $N_{p^\sharp}^{\pi^\sharp}$, otherwise.

Example 4.2 For Needham-Schroeder, we have the following set of abstract nonces:

$$\mathcal{N}^\sharp = \{N_I, N_A, N_B, N_A^{A,x}, N_B^{x,A} \mid x \in \{A, I\}\}.$$

Recall that we use N_B instead of n'_A. Siimilarly, for $N_B^{x,A}$ with $x \in \{A, I\}$. □

The abstraction of a message term t, denoted by $t^\sharp$, is obtained as the homomorphic extension of the abstractions on participants, nonces and keys. For a set T of terms, let $T^\sharp = \{t^\sharp \mid t \in T\}$.

The set $\mathcal{T}(\mathcal{F})^\sharp$ of abstract messages is the set of ground terms over $\mathcal{A}^\sharp$ and the constructors **encr** and **pair** as for $\mathcal{T}(\mathcal{F})$. Similarly, we can define the set of abstract terms by allowing variables in $\mathcal{X}$.

We are now ready to define the abstraction of a cryptographic protocol that will be given as a pair (C, R) of constraints of the form $E \not\vdash^{\in_c} m$, where $m \in \mathcal{T}(\mathcal{F})^\sharp$ and a set of abstract transitions. We call (C, R) an *abstract protocol*. The pair (C, R) defines a transition system whose initial states are sets $E \subseteq \mathcal{T}(\mathcal{F})^\sharp$ that satisfy C and where we have $E \rightarrow_R E'$, if there is $t \rightarrow t'$ in R and $\rho : \mathcal{X} \rightarrow \mathcal{T}(\mathcal{F})^\sharp$ such that $E \vdash \rho(m)$ and $E' = E \cup \{\rho(m')\}$.

The abstraction $\mathcal{S}^\sharp$ of the cryptographic protocol defined by $\mathcal{S}$ is defined by:

- $C^\sharp(E) = \{E \not\vdash^{\in_c} m^\sharp \mid E \not\vdash^{\in_c} m \text{ in } C(E, \pi_0, i_0)\}$ and
- the set R of abstract transitions $t_1^\sharp \rightarrow_R t_2^\sharp$ such that $t_1 \rightarrow t_2$ is a transition in some session instance $\mathcal{S}_\pi^i$.

We also call R abstract transitions rules. Let $\mathcal{PT} = (C, R)$ be an abstract protocol and $E_0 \subseteq \mathcal{T}(\mathcal{F})^\sharp$. We say that $\mathcal{PT}$ *preserves the secret* s *in* E_0, denoted by $E_0 \not\vdash_{\mathcal{PT}} s$, if for all $E \subseteq \mathcal{T}(\mathcal{F})^\sharp$, if $C(E_0)$ and $E_0 \rightarrow_R^* E$ then $E \not\vdash s$.

To relate a cryptographic protocol and its abstraction, we need to relate derivation by the intruder on the concrete and abstract messages. We can prove by structural induction on m the following:

Lemma 4.1 *Let E be a set of messages and $E^\sharp = \{m^\sharp \mid m \in E\}$. Then, $E \vdash m$ implies $E^\sharp \vdash m^\sharp$, for any message $m \in \mathcal{T}(\mathcal{F})$.* □

We can also prove the following lemma to relate concrete and abstract term instantiations:

Lemma 4.2 *Let t_1 and t_2 be terms and let $\rho : \mathcal{X} \rightarrow \mathcal{T}(\mathcal{F})$. Then, $\rho(t_1) = \rho(t_2)$ implies $\rho^\sharp(t_1^\sharp) = \rho^\sharp(t_2^\sharp)$, where $\rho^\sharp(X)$ is defined as $\rho(X)^\sharp$.* □

Using Lemma 4.1 and Lemma 4.2, we can prove that $(C^\sharp, R)$ is indeed an abstraction of $\mathcal{S}$ where the abstraction of a configuration (ξ, E) is $E^\sharp$:

Proposition 4.1 *Let (ξ_1, E_1) and (ξ_2, E_2) be concrete configurations. Then,*

$$(\xi_1, E_1) \rightarrow (\xi_2, E_2) \text{ implies } E_1^\sharp \rightarrow_R E_2^\sharp.$$

Moreover, if $C(E)$ is true then also $C^\sharp(E^\sharp)$. □

Exploiting Proposition 4.1 and the fact that $(C^\sharp, R)$ does not depend on (i_0, π_0), that is, we have the same constraints and transitions for all $(i, \pi) \in \text{Inst}$, we can prove:

Theorem 4.1 *The protocol defined by $\mathcal{S}$ satisfies te secrecy property defined by S in E_0, if its abstraction $(C^\sharp, R)$ preserves $S^\sharp$ in $E_0^\sharp$, i.e.,*

$$E_0^\sharp \not\vdash_{(C^\sharp, R)} S^\sharp \text{ implies } Secret(\mathcal{S}, S, E_0).$$

Example 4.3 Let us consider again our running example the Needham-Schroeder protocol. The set R of rules contains for instance the rules of Figure 4.3.

Moreover, the constraint $C^\sharp(E)$ is defined by $E \not\vdash^{\in_c} \{N_A, N_B\}$.

5 Verification based on patterns keeping secrets

Throughout, this section we assume that we are given a protocol $P = (\mathcal{C}, \mathcal{R})$ and a set of secrets defined by a set $\mathcal{S}$ of messages. We present an algorithm that allows to verify that a protocol preserves a set of secrets.

$$\pi = \pi_0 \quad \frac{A}{\{A,N_A\}_{pbk(A)}}; \quad \pi = (A,A) \quad \frac{A}{\{A,N_A^{AA}\}_{pbk(A)}};$$

$$\pi = (A,I) \quad \frac{A}{\{A,N_A^{AI}\}_{pbk(I)}};$$

$$\pi = \pi_0 \quad \frac{\{A,Y\}_{pbk(A)}}{\{Y,N_B\}_{pbk(A)}}; \quad \pi = (A,A) \quad \frac{\{A,Y\}_{pbk(A)}}{\{Y,N_B^{AA}\}_{pbk(A)}};$$

$$\pi = (I,A) \quad \frac{\{I,Y\}_{pbk(A)}}{\{Y,N_B^{IA}\}_{pbk(I)}};$$

$$\pi = \pi_0 \quad \frac{\{N_A,Z\}_{pbk(A)}}{\{Z\}_{pbk(A)}}; \quad \pi = (A,A) \quad \frac{\{N_A^{AA},Z\}_{pbk(A)}}{\{Z\}_{pbk(A)}};$$

$$\pi = (A,I) \quad \frac{\{N_A^{AI},Z\}_{pbk(A)}}{\{Z\}_{pbk(I)}};$$

Figure 3: The abstract rules of NS

5.1 Hat-messages: messages keeping secrets under a hat

If a principal A wants to protect a secret s, then he has to use a key whose inverse is not known by the intruder in order to encrypt every occurence of s in every message sent. The secret s itself need not be directly encrypted. Indeed, it should be enough that it only appears as part of encrypted messages. The basic idea of our method is to compute the set of encrypted messages that protect the secrets. As we will see, encryption with a safe key is not always sufficient to protect a secret in every message, as honest principals following the protocol can unwillingly help the intruder in decrypting the message.

In order to develop this idea formally we need to introduce a few definitions. Recall that critical and non-critical positions as well as the notation $\in_c$ has been introduced in Section 3.3.

Definition 5.1 *We call* hat-term *any term of the form* $\mathbf{encr}(t,k)$. *It is called* hat-message *if it is a ground term. Then, a secret s is protected by a set of hat-messages H in a set of messages M, denoted by $M\langle H\rangle s$, if the following condition is satisfied:*

> *for all messages $m \in M$, for all critical positions p in m with $m_{|p} = s$, there exists a position $q \prec p$ such that $m_{|q} \in H$.* □

Example 5.1 According to our definition, the hat-message $\mathbf{encr}(s,k)$ protects the secret s in the messages $\mathbf{encr}(\mathbf{encr}(s,k),k')$ but it does not protect it neither in the message $\mathbf{encr}(\mathbf{encr}(s,k'),k)$ nor in $\mathbf{pair}(\mathbf{encr}(s,k),s)$. Furthermore, even if the key k is part of the secrets, it does not need to be protected in the previous messages for it never appears in critical position. □

In order to extend the previous definition to a set of secrets, we introduce a function $\mathcal{H} : \mathcal{T}(\mathcal{F}) \to \mathscr{P}(\mathcal{T}(\mathcal{F}))$ which associates each secret with a set of hat-messages. Since

there will be no ambiguity, we overload the notation of Definition 5.1. Let $\mathcal{S}$ be a set of secrets, we define : $M\langle\mathcal{H}\rangle\mathcal{S} = \bigwedge_{s\in\mathcal{S}} M\langle\mathcal{H}(s)\rangle s$.

The notion of a message being protected by a hat-message introduced above does not take into account the capabilities of the intruder to decompose and compose new messages. We have to give particular care to the treatment of composed secrets as they can be obtained either by composition or decomposition. To do so, let us define the weak closure under decomposition of a term. Let T be a set terms. Then, T is called *weakly closed under decomposition*, or just weakly closed for short, if for any term $f(t_1, \cdots, t_n) \in T$ there is some $i \in \mathbb{N}_n$ such that $t_i \in T$. Then, a set T' is called a *weak closure* of T, if it is weakly closed, contains T and no proper sub-set of it satisfies these properties. Now, we are ready to extend the notion of a message being protected by a hat-message taking into account the intruder.

Definition 5.2 *A set $\mathcal{S}$ of messages is called* strongly protected *by $\mathcal{H}$ in M, denoted by $M[\mathcal{H}]\mathcal{S}$, if the following conditions are satisfied:*

1. *$M \vdash m \Rightarrow m\langle\mathcal{H}\rangle\mathcal{S}$, i.e., the secrets in $\mathcal{S}$ must be protected in any message that the intruder can deduce from M.*
2. *$\mathcal{S}$ is weakly closed.*
3. *$\bigcup_{s\in\mathcal{S}} Keys(\mathcal{H}(s))^{-1} \subseteq \mathcal{S}$ where $Keys(H)^{-1} = \{k^{-1} \mid \mathbf{encr}(t,k) \in H\}$, i.e., the inverses of all the keys used in hat-messages are also secrets.* □

Intuitively, Condition (2) ensures that the intruder will always miss at least one part of a secret preventing him from deducing it by composition. Condition (3) ensures that the intruder will not be able to decrypt a secret protected by a hat-message.

The work in [34] introduces the notion of an (co-)ideal induced by a set of messages. The ideal $I(S)$ induced by S is the set of messages that have to be protected in order not to reveal any secret in S. This notion has been also used in [15]. The notion of a message being strongly protected by $\mathcal{H}$ in M is finer in the sense that it tells us why the intruder cannot get a secret. That is, it tells which messages really protect the secret.

Our main motivation in introducing $M[\mathcal{H}]\mathcal{S}$ lays in the following Proposition that states that adding a message m in which the secrets of $\mathcal{S}$ are protected preserves strong protection of $\mathcal{S}$.

Proposition 5.1 *Let $M, \mathcal{S} \subseteq \mathcal{T}(\mathcal{F})$ be sets of messages, m be a message and $\mathcal{H} : \mathcal{T}(\mathcal{F}) \to \mathscr{P}(\mathcal{T}(\mathcal{F}))$. If $M[\mathcal{H}]\mathcal{S}$ and $m\langle\mathcal{H}\rangle\mathcal{S}$ then $M \cup \{m\}[\mathcal{H}]\mathcal{S}$.* □

The proof of this proposition is presented in the appendix. Notice while reading the proof that it does not rely on the fact that keys are atomic. Actually, our method can be extended to cover the case of non atomic keys.

Example 5.2 Consider a composed secret $s = \mathbf{pair}(s_1, s_2)$ and a $\mathcal{H}$ function defined by:

$\mathcal{H}(s_1) = \mathcal{H}(s) = \{\mathbf{encr}(\mathbf{pair}(s_1, s_2), k)\}$
$\mathcal{H}(s_2) = \mathcal{H}(s_1) \cup \{\mathbf{encr}(s_2, k)\}$
$\mathcal{H}(k) = \mathcal{H}(k^{-1}) = \emptyset$

Let $\mathcal{S} = \{k, k^{-1}, s, s_2\}$. Notice that $\mathcal{S}$ is weakly closed. Then, for the set of messages $M = \{\mathbf{encr}(\mathbf{pair}(s_1, s_2), k)\}$ and $m = \mathbf{pair}(s_1, \mathbf{encr}(s_2, k))$, we have $M[\mathcal{H}]\mathcal{S}$ and $m\langle\mathcal{H}\rangle s$. Indeed, k^{-1} does not appear at all; k is never used in critical positions; s and s_2 are protected in M by the hat $\mathbf{encr}(\mathbf{pair}(s_1, s_2), k)$; s_2 is protected in m by the hat $\mathbf{encr}(s_2, k)$; and s does not appear in m. Then, it follows from the proposition that $\mathcal{S}$ is protected by $\mathcal{H}$ in $M \cup \{m\}$, even if s_1 is not kept in m. Besides, notice that there is no hat term able to keep s_1 in m. □

Let now $r = t_1 \to t_2$ be a rule in $\mathcal{R}$ (recall that we have fixed an arbitrary protocol $P = (\mathcal{C}, \mathcal{R})$. We say that the pair $(\mathcal{S}, \mathcal{H})$ composed of a hat function and a set of secrets is *stable w.r.t. a rule* r, if for every subtitution σ, we have $\sigma(t_2)\langle\mathcal{H}\rangle\mathcal{S}$; it is stable w.r.t. a set of rules $\mathcal{R}$, if it stable w.r.t. to each rule in $\mathcal{R}$. Then, using Proposition 5.1, we can prove by induction the following Theorem:

Theorem 5.1 *Let $\mathcal{S}$ be a set of secrets, $\mathcal{H}$ be a hat function. If $(\mathcal{S}, \mathcal{H})$ is stable w.r.t. all rules in $\mathcal{R}$ and $E_0[\mathcal{H}]\mathcal{S}$, for every set of messages E_0 that satisfies $\mathcal{C}$, then $\not\vdash_P \mathcal{S}$, i.e., the secrets in $\mathcal{S}$ are preserved by P.*

5.2 Computing stable secrets and hat-messages

According to Theorem 5.1, given a protocol $P = (\mathcal{C}, \mathcal{R})$ and a set $\mathcal{S}$ of secrets, if we can find a hat function $\mathcal{H}$ and a set $\mathcal{S}'$ of secrets such that:

- the constraint $\mathcal{C}$ on E_0 – the messages initially known by the intruder – imply $E_0[\mathcal{H}]\mathcal{S}'$,
- $\mathcal{S} \subseteq \mathcal{S}'$ and
- $(\mathcal{S}', \mathcal{H})$ is stable w.r.t. $\mathcal{R}$,

we can conclude that the secrets in $\mathcal{S}$ are preserved by $P = (\mathcal{C}, \mathcal{R})$. In this section, we develop an algorithm that computes a stable pair $(\mathcal{S}', \mathcal{H})$. This is done in two steps. First, we develop a semantic version of the algorithm in which we do not consider questions related to representing sets of hat-messages. Then, we define a symbolic representation for hat-messages and develop a symbolic algorithm.

input: $\mathcal{R}$, a weakly closed $\mathcal{S}$ and $\mathcal{H}$
output: $\mathcal{S}'$, $\mathcal{H}'$ such that $(\mathcal{S}', \mathcal{H}')$ is stable w.r.t. $\mathcal{R}$.
$\mathcal{S}' := \mathcal{S}$; $\mathcal{H}' := \mathcal{H}$;
repeat
  — *first, add to the secrets the inverse of the keys used in hat-messages then,*
  — *compute the closure that adds to S' one subpart of each compound secret of $\mathcal{S}'$*
  $\mathcal{S}' := \mathcal{S}' \cup Keys(\mathcal{H}')^{-1}$; $\mathcal{S}' := Closure(\mathcal{S}')$; $\mathcal{S}_c := \mathcal{S}'$; $\mathcal{H}_c := \mathcal{H}'$;
  for each $t_p \to t_c \in \mathcal{R}$
    — *compute all Dangerous Substitutions of rule $t_p \to t_c$ where a secret is*
    — *not kept in the conclusion*
    $DS := \{\sigma : \mathcal{X} \to \mathcal{T}(\mathcal{F}) \mid \neg(\, \sigma(t_c)\langle \mathcal{H}' \rangle \mathcal{S}' \,)\}$;
    — *compute the corresponding Dangerous Premisses*
    $DP := \{\sigma(t_p) \mid \sigma \in DS\}$;
    — *update the secret and hat-messages according to the dangerous premisses:*
    — *case 1 removes the hat-messages that appear in dangerous premisses*
    for each $s \in \mathcal{S}'$, $m \in DP$ **do**
      — *the Bad Hats-messages of s are those which protect a secret*
      — *that can be disclosed by a rule*
      $BHat := \{m_{|q} \mid \exists p.\ q \prec p \wedge m_{|p} = s \wedge m_{|q} \in \mathcal{H}'(s)\}$
      — *select a subset of hat-messages to be removed*
      choose $BHat' \subseteq BHat$ **with** $BHat \neq \emptyset \ \Rightarrow\ BHat' \neq \emptyset$
      — *update the $\mathcal{H}$ function at point s*
      $\mathcal{H}'(s) := \mathcal{H}'(s) \setminus BHat'$;
    od
    — *case 2 adds to the secrets all bad premisses than do not contain any secret*
    $newS := \{m \in DP \mid \forall s \in \mathcal{S}'.\ s \not\in_c m\}$; $\mathcal{S}' := \mathcal{S}' \cup newS$
  od
until $(\mathcal{S}', \mathcal{H}') = (\mathcal{S}_c, \mathcal{H}_c)$

Figure 4: The semantic version of the verification algorithm

5.3 A semantic version of the verification algorithm

In Figure 4, we present an algorithm that computes a pair $(\mathcal{S}, \mathcal{H})$ which is stable w.r.t. the rules of the protocol. It uses the function *Closure* that associates to a set T of terms a weak closure of T. The algorithm takes a set of rules $\mathcal{R}$, a set of secrets $\mathcal{S}$ and a function $\mathcal{H}$ as input. It is a fixpoint computation, starting with $(\mathcal{S}, \mathcal{H})$. If it terminates, it returns an augmented set of secrets $\mathcal{S}'$ and a function $\mathcal{H}'$ such that $\mathcal{H}'(s) \subseteq \mathcal{H}(s)$, for every secret s. We now explain the intuitively the clue point of the algorithm. Let us take a rule $t_p \to t_c$ in $\mathcal{R}$, a substitution $\sigma : \mathcal{X} \to \mathcal{T}(\mathcal{F})$ and a hat-message h such that h protects a secret in $\sigma(t_p)$. If the secret is not protected by any hat-messages in $\sigma(t_c)$ – the conclusion of the instanciated rule, then the hat-message h is not safe indeed and it must be removed from the set of hat-messages. Actually, even when the inverse of the keys used in the hat-messages are unknown by the intruder, the secret can be unwillingly revealed by a principal. Think for instance of a protocol with $\{(y, x)\}_{pbk(A)} \to \{x\}_{pbk(y)}$ as a rule of principal A. On reception of the message $\{(I, Secret)\}_{pbk(A)}$, the principal A will respond $\{Secret\}_{pbk(I)}$ thus unwillingly decrypting the secret for the intruder. So, $\{(I, Secret)\}_{pbk(A)}$ is a particular case where $pbk(A)$ does not protect the secret and it must be removed from the set of hat-messages. Case 2 in the algorithm considers the case where a secret is not protected in the conclusion and the premisse does not contain a secret. In this case, the apparently harmless premise is as compromising as the secret, and hence, is added to the set of secrets.

The following proposition 5.2 summarizes the properties of the algorithm:

Proposition 5.2 *If the algorithm of Figure 4 applied to $(\mathcal{R}, \mathcal{S}, \mathcal{H})$ terminates, it returns $\mathcal{S}'$ and $\mathcal{H}'$ which satisfy the following conditions:*

1. *$(\mathcal{S}', \mathcal{H}')$ is stable w.r.t. $\mathcal{R}$,*
2. *$\mathcal{S} \subseteq \mathcal{S}'$ and*
3. *$Keys(\mathcal{H}')^{-1} \subseteq \mathcal{S}'$.* □

Using Proposition 5.2 and Theorem 5.1, we can prove the following:

Corollary 5.1 *If the algorithm of Figure 4 terminates with $(\mathcal{S}', \mathcal{H}')$ as result and each set of messages E_0 that satisfies $\mathcal{C}(E_0)$ also satisfies $E_0[\mathcal{H}']\mathcal{S}'$, we can conclude $\not\vdash_P \mathcal{S}'$, and hence, $\not\vdash_P \mathcal{S}$.* □

5.4 A symbolic representation of hat-messages: Pattern Terms

To develop an effective version of our semantic algorithm, we need to represent (potentially infinite) sets of hat-messages. To do so, we introduce a symbolic representation that consists in a pair $(\mathcal{G}, \mathcal{B})$ of sets of terms over variables in $\mathcal{X} \cup \mathcal{X}_s$. Here, $\mathcal{X}_s$ denotes a new disjoint set of variables. We use $x_s, y_s, \cdots$ as typical variables in $\mathcal{X}_s$. Roughly speaking, a hat-message for a secret is an instance of a term in $\mathcal{G}$ that is not an instance of any term in $\mathcal{B}$. Therefore, we call the terms in $\mathcal{G}$ *good patterns* and those in $\mathcal{B}$ *bad*

patterns. In good patterns, we use the variables in $\mathcal{X}_s$ to mark the positions in which a term containing a secret is allowed to appear. While in bad patterns, they are used to mark positions where it is not guaranteed that the secret is not revealed. For instance, if $\mathcal{G}$ contains the term $\{x_s\}_K$ and $\mathcal{B}$ contains the term $\{(A, x_s)\}_K$ then the message $\{(B, \mathit{Secret})\}_K$ will be in the set represented by $(\mathcal{G}, \mathcal{B})$, while the message $\{(A, \mathit{Secret})\}_K$ will not. Let us now define this symbolic representation formally.

To do so, we introduce *pattern terms* defined by the following BNF:

$$pt \ ::= \ N \mid P \mid K \mid x \mid x_s \mid \mathbf{pair}(pt_1, pt_2) \mid \mathbf{encr}(pt, K) \mid \mathit{Sup}\,(pt)$$

where $N \in \mathcal{N}$, $P \in \mathcal{P}$, $K \in \mathcal{K}$, $x \in \mathcal{X}$, and $x_s \in \mathcal{X}_s$. The set of pattern terms is denoted by $\mathcal{PT}(\mathcal{X} \cup \mathcal{X}_s, \mathcal{F})$. Notice that every term in $\mathcal{T}(\mathcal{X} \cup \mathcal{X}_s, \mathcal{F})$ is also a pattern term in $\mathcal{PT}(\mathcal{X} \cup \mathcal{X}_s, \mathcal{F})$. The difference between the two is that patterns terms make use of the special *Sup* function symbol.

Intuitively, as can be seen from the following definition, $\mathit{Sup}\,(t)$ represents all terms containing the term t as a sub-term. For instance, the terms A, $\mathbf{pair}(x, A)$, $\mathbf{encr}(A, K)$, $\cdots$ all belong to $[\![\mathit{Sup}\,(A)]\!]$.

Definition 5.3 *Given a pattern term* pt*, let* $[\![\mathrm{pt}]\!]$ *be defined as follows:*

$$\begin{array}{l}
[\![\mathrm{pt}]\!] \ = \ \{\mathrm{pt}\} \quad \textit{if pt is a constant or a variable} \\
[\![\mathbf{pair}(\mathrm{pt}_1, \mathrm{pt}_2)]\!] \ = \ \{\mathbf{pair}(t_1, t_2) \mid t_1 \in [\![\mathrm{pt}_1]\!], t_2 \in [\![\mathrm{pt}_2]\!]\} \\
[\![\mathbf{encr}(\mathrm{pt}_1, k)]\!] \ = \ \{\mathbf{encr}(t_1, k) \mid t_1 \in [\![\mathrm{pt}_1]\!]\} \\
[\![\mathit{Sup}\,(\mathrm{pt})]\!] \ = \ \{t \mid \ \textit{there is a position } p \textit{ in } t \textit{ s.t. } t_{|p} \in [\![\mathrm{pt}]\!]\}
\end{array}$$

□

To define the concretization of pattern terms we need to introduce *positive* and *negative substitutions*. As usual a substitution is a mapping $\sigma : \mathcal{X} \cup \mathcal{X}_s \to \mathcal{T}(\mathcal{X} \cup \mathcal{X}_s, \mathcal{F})$. A ground substitution is a mapping $\sigma : \mathcal{X} \cup \mathcal{X}_s \to \mathcal{T}(\mathcal{F})$. A substitution is called *positive*, if $\sigma(x)_{|p} \notin \mathcal{X}_s \cup \mathcal{S}$, for every $x \in \mathit{dom}(\sigma) \cap \mathcal{X}$ and critical position p in $\sigma(x)$. It is *negative*, if there is a variable $x_s \in \mathcal{X}_s \cap \mathit{dom}(\sigma)$ such that $\sigma(x_s)_{|p} \in \mathcal{X}_s \cup \mathcal{S}$, for some critical position p in $\sigma(x_s)$. Then, let $\gamma^+_{\mathcal{S}}(pt) = \{\sigma(t) \mid t \in [\![pt]\!], \sigma \text{ is ground and positive}\}$ and $\gamma^-_{\mathcal{S}}(pt) = \{\sigma(t) \mid t \in [\![pt]\!], \sigma \text{ is ground and negative}\}$. As usual, we can extend $\gamma^+_{\mathcal{S}}$ and $\gamma^-_{\mathcal{S}}$ pointwisely to sets of pattern terms. We use $\gamma^-_{\mathcal{S}}$ for the concretization of bad patterns and $\gamma^+_{\mathcal{S}}$ for the concretization of good patterns. We represent pairs $(\mathcal{S}, \mathcal{H})$ by triples $\langle \mathcal{S}, \mathcal{G}, \mathcal{B} \rangle$ of finite sets of secrets, good and bad patterns. Good patterns correspond to "maybe safe" hat terms and bad patterns define "really unsafe" hat terms. The concretization of $\langle \mathcal{S}, \mathcal{G}, \mathcal{B} \rangle$ is the pair $(\mathcal{S}, \mathcal{H})$, where $\mathcal{H}$ associates to every $s \in \mathcal{S}$ the same set $\gamma^+_{\mathcal{S}}(\mathcal{G}) \setminus \gamma^-_{\mathcal{S}}(\mathcal{B})$ of hat-messages. More formally, we have the following:

Definition 5.4 *A* symbolic representation *SR is triple* $\langle \mathcal{S}, \mathcal{G}, \mathcal{B} \rangle$*, where*

- $\mathcal{S}$ *is a finite set of terms that represents the secrets,*
- $\mathcal{G}$ *is a finite set of terms that represent the good hat messages, and*

- $\mathcal{B}$ *is a finite set of pattern terms that represent the bad hat messages*

A symbolic representation $\langle \mathcal{S}, \mathcal{G}, \mathcal{B} \rangle$ *defines the pair* $(\mathcal{S}, \mathcal{H})$ *where* $\mathcal{H}(s) = \gamma^{+}_{\mathcal{S}}(\mathcal{G}) \setminus \gamma^{-}_{\mathcal{S}}(\mathcal{B})$, *for each* $s \in \mathcal{S}$. □

Example 5.3 Let $g = \mathbf{encr}(x_s, K)$ and $SR = \langle \{N, K^{-1}\}, \{g\}, \emptyset \rangle$. Then, the concretization of SR is $(\{N, K^{-1}\}, [N \mapsto H, K^{-1} \mapsto H])$, where H consists of all ground instances of g that contains N or K^{-1}. Thus, N is protected in $\mathbf{encr}(\ \mathbf{encr}(\mathbf{encr}(N, K''), K),\ K')$, for the hat message $\mathbf{encr}(\mathbf{encr}(N, K''), K)$ belongs to $\mathcal{H}(N) = H$ as an instance of g with the substitution $[x_s \mapsto \mathbf{encr}(N, K'')]$. □

6 A symbolic verification algorithm

The symbolic algorithm is obtained from the algorithm of Figure 4 by replacing each operation by a corresponding symbolic one that operates on $\langle \mathcal{S}, \mathcal{G}, \mathcal{B} \rangle$. For the sake of presentation, we first explain the symbolic algorithm in the particular case where $\mathcal{B}$ consists of terms rather than pattern terms, i.e., *Sup* does not occur in any term in $\mathcal{B}$. We will explain later how it extends to pattern terms and what are the difficulties to solve.

6.1 The algorithm on terms

Before presenting the algorithm we need to introduce the following definitions. Let t and t' be terms. A substitution $\sigma : \mathcal{X} \cup \mathcal{X}_s \to \mathcal{T}(\mathcal{X} \cup \mathcal{X}_s, \mathcal{F})$ is called *a positive matcher* of t' on t, if it is a positive substitution with $\sigma(t') = t$. Given a set $\mathcal{G}$ of terms, we say that t matches positively with $\mathcal{G}$, if there is a term $t' \in \mathcal{G}$ and a positive matcher of t' on t. Moreover, a substitution $\sigma : \mathcal{X} \cup \mathcal{X}_s \to \mathcal{T}(\mathcal{X} \cup \mathcal{X}_s, \mathcal{F})$ is called *a negative unifier* of t' and t, if it is a negative substitution with $\sigma(t') = \sigma(t)$.

The symbolic algorithm takes as input a set of rules $\mathcal{R}$, a set of secrets $\mathcal{S}$, a set of good terms $\mathcal{G}$ and an empty set of bad terms $\mathcal{B} = \emptyset$. It computes new bad terms and secrets until the concretization of $\langle \mathcal{S}, \mathcal{G}, \mathcal{B} \rangle$ becomes stable w.r.t. all rules in $\mathcal{R}$. Let us now sketch its main steps:

1. The set $\mathcal{S}$ of secrets becomes $\mathcal{S} \cup \mathit{Keys}(\mathcal{G})^{-1}$, where $\mathit{Keys}(\mathcal{G})^{-1}$ is the set of keys of the form k^{-1} such that $\mathbf{encr}(t, k)$ is in $\mathcal{G}$.

2. Given a rule $t_p \to t_c$ in $\mathcal{R}$, we have to consider all possible occurrences of a secret in the conclusion t_c. This is done by replacing variables in t_c by secret variables. In other words, we consider all rules in $\mathcal{R}' = \{\sigma(t_p) \to \sigma(t_c) \mid t_p \to t_c \in \mathcal{R},\ \sigma : \mathit{var}(t_c) \to \mathcal{X}_s\}$.

3. Given a rule $t_p \to t_c$ in $\mathcal{R}'$, the algorithm computes the finite set of dangerous substitution DS as follows. A substitution $\sigma : \mathcal{X} \cup \mathcal{X}_s \to \mathcal{T}(\mathcal{X} \cup \mathcal{X}_s, \mathcal{F})$ is *dangerous* if there is a critical position p in t_c such that $t_{c|p} \in \mathcal{X}_s \cup \mathcal{S}$ and for every position

$q \prec p$, when $t_{c|q}$ matches positively with $\mathcal{G}$, it also unifies negatively by σ with a term in $\mathcal{B}$. Then, $DS := \{\sigma : \mathcal{X} \cup \mathcal{X}_s \to \mathcal{T}(\mathcal{X} \cup \mathcal{X}_s, \mathcal{F}) \mid \sigma \textit{ is dangerous}\}$. The paragraph below illustrates the computation of dangerous substitutions.

4. The set of *bad premises* is given by $BP = \{\sigma(t_p) \mid \sigma \in DS\}$. The new bad terms are the subterms of a bad premise $t \in BP$ that were supposed to protect a secret in t as they match positively with $\mathcal{G}$. So, bad terms are particular instances of terms in $\mathcal{G}$ that must be removed. Formally,

   $$BHat = \{t_{|q} \mid t \in BP,\ \exists p.\ q \prec p\ \wedge\ t_{|p} \in \mathcal{X}_s \cup \mathcal{S}\ \wedge\ t_{|q} \textit{ matches positively with } \mathcal{G}\}$$

   Notice that we use the same set of hat-terms for all the secrets, so each occurrence of $\mathcal{H}'$ and $\mathcal{H}'(s)$ is now replaced by the set of bad terms $\mathcal{B}'$ and the restriction of $\mathcal{H}'(s)$ is replaced by: $\mathcal{B}' := \mathcal{B}' \cup BHat$.

5. Finally, $newS$ is replaced by $newS := \{t \in DP \mid var(t) \cap \mathcal{X}_s = \emptyset,\ \forall s \in \mathcal{S}'.\ s \notin_c t\}$.

Computation of dangerous substitutions

We present the algorithm that computes the dangerous substitutions induced by a rule $t_p \to t_c$, a set of good terms $\mathcal{G}$ and a set of bad terms $\mathcal{B}$. We apply the following algorithm to each secret (a message in $\mathcal{S}$ or a variable in $\mathcal{X}_s$) that appears at critical position in the conclusion of the rule.

Let p be a position of a secret in t_c. Let $\mathcal{PP}$ be the set of positions p_i above p such that a good term of $\mathcal{G}$ matches positively with $t_{c|p_i}$. In the sequel, the found positive matchers are of no use; we only retain the protecting positions $p_1 \prec p_2 \prec \ldots \prec p_n (\prec p)$. We define below the function Φ that computes all the negative unifiers that cancel each protecting position by a bad term. Formally, the dangerous substitutions are the negative unifiers σ that satisfy: $\bigwedge_{p_i \in \mathcal{PP}} \sigma(t_c)_{|p_i} = \sigma(b_i)$ with $b_i \in \mathcal{B}$.

Initially, Φ is called with the set $\mathcal{PP}$ of protecting positions and a set of substitutions DS containing only the empty substitution: $DS = \{[\,]\}$. Then, it takes in turn each protecting position and if it is possible, it completes the substitutions of DS in order to cancel the current position by a bad term.

$$\Phi(t_c,\ \mathcal{B},\ \mathcal{PP},\ DS) = \begin{cases} \{\sigma \in DS \mid \sigma \textit{ is a negative unifier}\} & \text{if } \mathcal{PP} = \emptyset \\ \Phi(t_c,\ \mathcal{B},\ \mathcal{PP} \setminus \{p_i\},\ \bigcup\limits_{\sigma_j \in DS} \{\sigma_j \cup \sigma_1^{i,j}, \cdots, \sigma_j \cup \sigma_{n_{i,j}}^{i,j}\}\ ) & \text{for a position } p_i \text{ chosen in } \mathcal{PP} \end{cases}$$

where the $\sigma_k^{i,j}$ in the fourth argument are the negative unifiers resulting of the unification of $\sigma_j(t_c)_{|p_i}$ with some bad terms of $\mathcal{B}$.

The same algorithm is used in the case where $\mathcal{G}$ and $\mathcal{B}$ are sets of terms or sets of patterns terms. We only need to adapt the unification algorithm. In the case of terms, we use the standard *most general unifier*; and for patterns terms, we define a unification algorithm presented in Section 6.2.

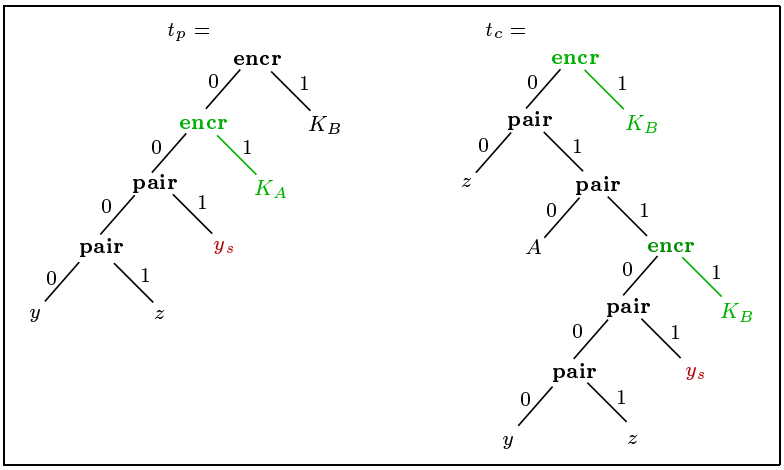

Figure 5: Illustration of computing dangerous substitutions.

Example 6.1 We illustrates the computation of dangerous substitutions on the sets of terms

$$\mathcal{G} = \{\ \mathbf{encr}(x_s, K_B),\ \mathbf{encr}(x_s, K_A)\ \}$$
$$\mathcal{B} = \{\ \mathbf{encr}(\mathbf{pair}(I, x'_s), K_B),\ \mathbf{encr}(\mathbf{pair}(\mathbf{pair}(A, x''), x''_s), K_B)\ \}$$

and a rule $t_p \to t_c$ given in Figure 5.

We consider the conclusion of the rule. The first step consists in looking for all the critical positions in the conclusion where a secret or a secret variable appears. We find one variable $y_s \in \mathcal{X}_s$ at position 0.1.1.0.1 in the term t_c. For each critical position, we look for the positions above it that match with a good term in $\mathcal{G}$. For y_s, we found exaclty two protecting positions: the term $\mathbf{encr}(x_s, K_B)$ matches positively at position $p_1 = \epsilon$ and also at position $p_2 = 0.1.1$. Then, the function Φ looks for all substitutions that negatively unify some bad terms with the terms at the protecting positions p_1 and p_2. Starting with position $p_1 = \epsilon$, it unifies $t_{c|\epsilon}$ with the bad term $\mathbf{encr}(\mathbf{pair}(I, x'_s), K_B)$ for the negative unifier $\sigma' = [z = I, x'_s = t_{c|0.1}]$. This cancels the top most protection. Then, the function Φ attempts to complete the substitution σ' so that it also cancels the protection at position $p_2 = 0.1.1$. To do so, it tries to unify the term $\sigma'(t_c)_{|p_2} = \mathbf{encr}(\mathbf{pair}(\mathbf{pair}(y, I), y_s), K_B)$ with a bad term and succeeds with the bad term $\mathbf{encr}(\mathbf{pair}(\mathbf{pair}(A, x''), x''_s), K_B)$ and the negative unifier $\sigma'' = [y = A, x'' = I, y_s = x''_s]$. The two unifiers are then composed and restricted to the domain $var(t_c)$ resulting in the substitution $\sigma = (\sigma' \cup \sigma'')_{/var(t_c)} = [y = A, z = I]$. Pursuing the computation does not provide other substitutions and finally Φ returns the set of dangerous substitutions $\{\sigma\}$. We now look at the premise of the rule to compute the new bad terms induced by σ. The position of the secret variable y_s in t_p is protected by the good term $\mathbf{encr}(x_s, K_A)$ that matches positively on $t_{p|0}$. However, the dangerous substitution σ tells that this protection will not work in case where y is A and z is I. Consequently, we refine the set of hat messages by removing this particular case. In our symbolic representation, this comes out to add $\sigma(t_{p|0}) = \mathbf{encr}(\ \mathbf{pair}(\mathbf{pair}(A, I), y_s),\ K_A)$

to the set of bad terms. □

6.2 Dealing with pattern terms

First of all, it is worth to mention that pattern terms are more expressive than terms, that is, there are sets of messages that can be described as pattern terms but not as terms. This is for instance the case for the set of messages that contain the constant A as sub-message. In fact, introducing the interpreted function symbol *Sup* corresponds to adding the sub-term relation to a logic on terms. Moreover, it is not difficult to exhibit examples of protocols where one needs the expressive power of pattern terms to represent the bad hat-messages.

Unification and matching are the key operation in the 3^{rd} anf 4^{th} steps of the symbolic algorithm. The problem we need to solve for obtaining our symbolic algorithm is, however, *not* unification of pattern terms. The problem we need to solve is the following: Given two pattern terms u and t, determine the set $\mathcal{U}(u,t)$ of substitutions σ such that there exist terms $u' \in [\![u]\!]$ and $t' \in [\![t]\!]$ such that $\sigma(u') = \sigma(t')$. More precisely, we want to characterize the set of most general unifiers that unify some terms in $[\![u]\!]$ and $[\![t]\!]$. Actually, the problem we need to solve for our symbolic algorithm is a simpler one where at least one of the pattern terms u and t is simply a term, that is, without occurence of *Sup* in it [1]. We prefer, however, to present a solution for the general case. We will do this in a general setting.

Let us consider a finite set $\mathcal{F}$ of function symbols such that $Sup \notin \mathcal{F}$ and let $\mathcal{X}$ be a countable set of variables (see Section 2 for the notations). The set of pattern terms induced by $\mathcal{F}$ and $\mathcal{X}$, denoted by $\mathcal{PT}(\mathcal{F},\mathcal{X})$ is defined by the following BNF:

$$t ::= x \mid f(t_1,\cdots,t_n) \mid Sup\,(t)$$

where x is a variable in $\mathcal{X}$ and f is a function symbol of arity $n \geq 0$. Let $\mathcal{F}^{(i)}$ denote the function symbols in $\mathcal{F}$ of arity i. As usual, function symbols of arity 0, i.e. elements of $\mathcal{F}^{(0)}$, are called constants. The meaning $[\![t]\!]$ of a pattern term t is a set of terms in $\mathcal{T}(\mathcal{X},\mathcal{F})$ defined in the same way as in Definition 5.3.

Definition 6.1 *Given two pattern terms u and t, a substitution $\sigma : \mathcal{X} \to \mathcal{PT}(\mathcal{X},\mathcal{F})$ is a called a* maximal general unifier *for u and t, if the following conditions are satisfied:*

1. *it is a most general unifier for some terms $u' \in [\![u]\!]$ and $t' \in [\![t]\!]$ and*
2. *for every substitution σ' that unifies terms in $[\![u]\!]$ and $[\![t]\!]$, σ' is not more general than σ, that is, for no substitution ρ, we have $\sigma = \rho\sigma'$.*

We denote by $\mathcal{U}(u,t)$ the set of maximal general unifiers for u and t. □

In general there will be more than one maximal general unifier for u and t even modulo renaming. The definition of $\mathcal{U}$ can be extended in the usual way –as for unification–

[1] Indeed, we need to unify the conclusion of a rule, which is a term, with a bad pattern.

to sets $\{(u_i, t_i) \mid i \in [1, n]\}$ of pairs of pattern terms. In the sequel, we prefer to write $u_i = t_i$ instead of (u_i, t_i) as our algorithm essentially consists in manipulating some kind of equations.

In this section, we want to develop an algorithm that given $E = \{u_i = t_i \mid i \in [1, n]\}$ determines $\mathcal{U}(E)$. From now on, we will call such a set E a *generalized equational problem*, written *gep* for short. It turns out that an extension of the set of transformations that solve the usual unification problem (cf. [6]) will give the solution.

An equation $u = t$ is called simple, if the argument of any occurence of *Sup* in u and t is either a constant or a variable. A gep E is called simple, if all its equations are simple. It is not difficult to see that every gep E can be transformed into an equivalent simple gep E'. Equivalent here means that $\mathcal{U}(E')$ restricted to the variables in E is the same as $\mathcal{U}(E)$.

Example 6.2 *Consider the following gep* $E = \{Sup\,(f(b, Sup\,(g(x)))) = f(b, g(Sup\,(a)))\}$. *Then, E is equivalent to E' that consists of the following equations:*

$$Sup\,(x_0) = f(b, g(Sup\,(a))),\ x_0 = f(b, Sup\,(x_1)),\ x_1 = g(x).$$

□

We first recall in Figure 6 the usual rules for solving unification.

Delete	$\{t = t\} \cup E$	$\Rightarrow$	E
Decompe	$\{f(u_1, \cdots, u_n) = f(t_1, \cdots, t_n)\} \cup E$	$\Rightarrow$	$\{u_i = t_i \mid i \in [0, n]\} \cup E$
Orient	$\{t = x\} \cup E$	$\Rightarrow$	$\{x = t\} \cup E$
Eliminate	$\{x = t\} \cup E$	$\Rightarrow$	$\{x = t\} \cup E[t/x]$
Clash	$\{f(u_1, \cdots, u_n) = g(t_1, \cdots, t_m)\} \cup E$	$\Rightarrow$	$\mathcal{U} = \emptyset$
Occurs-Check	$\{x = t\} \cup E$	$\Rightarrow$	$\mathcal{U} = \emptyset$, if x is not t and $x \in var(t)$

Figure 6: Rules of solving unification

Now to deal with *Sup*, we add new rules. We attract the reader's attention to the fact that the first rule (Splitting) transforms a simple gep E into a set of simple geps. Indeed, it yields a new gep for each sub-term of $f(t_1, \cdots, t_n)$. This is not the case for the usual unification rules. The rules for *Sup* are presented in Figure 7.

Example 6.3 *Reconsidering Example 6.2, our algorithm yields* $\{x = Sup\,(a)\}$. *It can be easily verified that indeed* $\mathcal{U}(E) = \{x = Sup\,(a)\}$.

Termination of the algorithm can be proved using lexicographic ordering and the ranking function that maps a gep E to (m_1, m_2, m_3), where:

- m_1 is the number of variables for which there is no equation of the form $x = t$,
- m_2 is the size of E, i.e. $\Sigma_{u=t \in E}(|u| + |t|)$ and

Splitting-var	$\{Sup\,(x) = f(t_1, \cdots, t_n)\} \cup E$	$\Rightarrow$	$\{x = t'\} \cup E$, for $t' \leq f(t_1, \cdots, t_n)$
Splitting-const	$\{Sup\,(a) = f(t_1, \cdots, t_n)\} \cup E$	$\Rightarrow$	$\{a = t'\} \cup E$, for $t' < f(t_1, \cdots, t_n)$
Simplification	$\{t' = Sup\,(t), t' = t\} \cup E$	$\Rightarrow$	$\{t' = t\} \cup E$
Sup-var-const-2	$\{Sup\,(x) = Sup\,(a)\} \cup E$	$\Rightarrow$	E, if $\bigcup_{i\geq 2} \mathcal{F}^{(i)} \neq \emptyset$
Sup-var-const-1	$\{Sup\,(x) = Sup\,(a)\} \cup E$	$\Rightarrow$	$\{x = Sup\,(a)\} \cup E$, if $\bigcup_{i\geq 2} \mathcal{F}^{(i)} = \emptyset$
Sup-const-2	$\{Sup\,(a) = Sup\,(b)\} \cup E$	$\Rightarrow$	E, if $\bigcup_{i\geq 2} \mathcal{F}^{(i)} \neq \emptyset$
Sup-const-1	$\{Sup\,(a) = Sup\,(b)\} \cup E$	$\Rightarrow$	$\{a = b\} \cup E$, if $\bigcup_{i\geq 2} \mathcal{F}^{(i)} = \emptyset$
Sup-var-2	$\{Sup\,(x) = Sup\,(y)\} \cup E$	$\Rightarrow$	E, if $\bigcup_{i\geq 2} \mathcal{F}^{(i)} \neq \emptyset$
Sup-var-1	$\{Sup\,(x) = Sup\,(y)\} \cup E$	$\Rightarrow$	$\{x = Sup\,(y)\} \cup E$, $\{Sup\,(x) = y\} \cup E$, if $\bigcup_{i\geq 2} \mathcal{F}^{(i)} = \emptyset$

Figure 7: Rules for *Sup*

- m_3 is the number of equations $t = x$ in E.

To prove soundness of the algorithm, we prove for each rule $E \Rightarrow E_1, \cdots, E_n$ that we have $\mathcal{U}(E) = \bigcup_{1\leq i\leq n} \mathcal{U}(E_i)$. This proof is presented in the full paper.

6.3 On the termination of the symbolic algorithm

In this section, we present a technique that makes a depth-first implementation of the symbolic verification algorithm always terminate, at a price of safe approximation of the results. In fact, our prototype implementation of our verification algorithm, named HERMES, terminates with precise results on all pratical examples of protocols we tried. That is, the results did not show any false attack (see Table 1).

A sequence $(t_i)_{i\geq 0}$ of pattern terms is called *increasing at a sequence* $(p_i)_{i\geq 0}$ of positions, if the following conditions are satisfied for every $i \geq 0$:

1. $p_i \in dom(t_i)$ and $p_i \preceq p_{i+1}$,
2. $t_0[z/p_0] = t_i[z/p_0]$, where z is fresh variable.
3. $t_{i|p_i} = t_{0|p_0}$.

Let us consider an example to clarify these definitions.

Example 6.4 *Consider the following rule from the session (A, A) of Needham-Schroeder-Lowe protocol presented in Section 6.4:*

$$r = \frac{\{(A, (N_1^{AA}, y))\}_{K_A}}{\{y\}_{K_A}}.$$

Consider the sequence $(\{\theta^i(I, x)\}_{K_A})_{i\geq 0}$, where $\theta(z) = (A, (N_1^{AA}, z))$. The first three terms of the sequence are $\{\theta^0(I, x)\}_{K_A} = \{(I, x)\}_{K_A}$, $\{\theta^1(I, x)\}_{K_A} = \{(A, (N_1^{AA}, (I, x)))\}_{K_A}$

and $\{\theta^2(I,x)\}_{K_A} = \{(A,(N_1^{AA},(A,(N_1^{AA},(I,x)))))\}_{K_A}$. *The whole sequence can be obtained by iteratively computing the bad patterns induced by the rule* r *starting from the bad term* $\{(I,x)\}_{K_A}$. *Thus, a naive application of our symbolic algorithm will not terminate. On the other hand, this sequence is increasing at* $(p_i = 0(11)^i)_{i\geq 0}$. *Indeed,* $\{\theta^i(I,x)\}_{K_A}[z/p_0] = \{z\}_{K_A}$ *and* $(\{\theta^i(I,x)\}_{K_A})_{|p_i} = (I,x)$, *for every* $i \geq 0$. *We will see now how this fact can be exploited to make the algorithm to converge.* □

The clue of our technique for enforcing termination of the symbolic algorithm is expressed by the following proposition:

Proposition 6.1 *Let* $(t_i)_{i\geq 0}$ *be increasing at* $(p_i)_{i\geq 0}$. *Then,*

$$\bigcup_{i\geq 0} [\![t_i]\!] \subseteq \bigcup_{i<j} [\![t_i]\!] \cup [\![t_j[Sup\,(t_{j\,|p_j})/p_j]]\!], \text{ for every } j \geq 0.$$

Example 6.5 *Consider again our Example 6.4. Then, if we choose* $j = 1$, *we obtain a set consisting of the two pattern terms* $\{(I,x)\}_{K_A}$ *and* $\{(A,(N_1^{AA},\, Sup\,(I,x)\,))\}_{K_A}$ *which approximates the whole sequence* $(\{\theta^i(I,x)\}_{K_A})_{i\geq 0}$.

Protocol Name	Result	Time (sec)
Yahalom	OK	12.67
Needham-Schroeder Public Key	Attack	0.04
Needham-Schroeder Public Key (with a key server)	Attack	0.90
Needham-Schroeder-Lowe	OK	0.03
Otway-Rees	OK[1]	0.01
Denny Sacco Key Distribution with Public Key	Attack	0.02
Wide Mouthed Frog (modified)	OK	0.04
Kao-Chow	OK	0.78
Neumann-Stubblebine	OK[1]	0.04
Needham-Schroeder Symmetric Key	Attack	0.08
ISO Symmetric Key One-Pass Unilateral Authentication	Attack	0.01
ISO Symmetric Key Two-Pass Unilateral Authentication	OK	0.01
Andrew Secure RPC	Attack	0.03

Table 1: The results provided by HERMES, our prototype for verifying secrecy properties, running on a Pentium III 600Mhz PC under Linux 2.2.19.

[1]There is a known attack of the untyped version of the protocol. This attack relies on the misuse of a message as an encryption key. Discovering this type attack automatically requires to deal with non-atomic keys. This is not yet implemented in HERMES.

6.4 Needham-Schroeder-Lowe Protocol

The corrected version of the Needham-Schroeder protocol is also called Needham-Schroeder-Lowe as it is G. Lowe who found the attack and corrected the protocol. The difference with the initial version is in the second transition of principal B:

$$\begin{array}{lcl} A \to B & : & \{A, N_1\}_{K_B} \\ B \to A & : & \{B, N_1, N_2\}_{K_A} \\ A \to B & : & \{N_2\}_{K_B} \end{array}$$

We run our verification algorithm with $\mathcal{S} = \{N_2, K_A^{-1}\}$, the empty set of bad patterns and the set $\mathcal{G} = \{\{x_s\}_{K_A}\}$ of good patterns. The algorithm terminates with the set of secrets unchanged and the set $\mathcal{B}$ of bad patterns given in Figure 8. As the initial

$$\{I, x_s\}_{K_A}, \quad \{A, (N_1^{AA}, Sup\,(I, x_s))\}_{K_A}, \quad \{A, (N_1, Sup\,(I, x_s))\}_{K_A}.$$

Figure 8: The bad pattern-terms for the Needham-Schroeder-Lowe protocol

constraints are $E \not\vdash^{\in_c} \{N_1, N_2, K_A^{-1}\}$, that is, none of the messages in $\{N_1, N_2, K_A^{-1}\}$ is contained at a critical position in a message derivable from E, it is easy to prove that we have $E[\gamma_S^+(\mathcal{G}) \setminus \gamma_S^-(\mathcal{B})]\,\mathcal{S}$. Hence, we can conclude that the Needham-Schroeder-Lowe protocol preserves the secret N_2. Concerning, the uncorrected version of Example 3.1, during computation of new secrets and bad patterns, we arrive at a situation where we have to add $\{A, N_1^{AI}\}_{K_I}$ as a secret. As this message contains neither a fresh nonce nor a secret, we stop the computation and follow it back to try construct an attack. This way, we obtain the attack known as "man in the middle". □

7 Conclusion

In this paper, we presented a method based on abstract interpretation for verifying secrecy properties of cryptographic protocols in a general model. Our method deals with unbounded number of sessions, unbounded number of principals, unbounded message depth and unbounded creation of fresh nonces. However, in contrast to the work in [5, 32, 26], where the session number is bounded, our method is not complete. Indeed, the problem is in its most general form undecidable even when pairing is not allowed as shown in [4]. The main contribution of the paper is a verification algorithm that consists of computing an inductive invariant using patterns as symbolic representation. Our method can already deal with models in which we distinguish between long term and short term keys and which contain variables ranging over keys. The idea here is that short term keys can be revealed to the intruder when a session has terminated. This is not the case for long term keys. This allows a more faithful modeling of some protocols.

An version of our tool together with the examples of Table 1 is available at the url `http://www-verimag.imag.fr/~lbozga/hermes/hermes.php`.

References

[1] Martín Abadi. Secrecy by typing in security protocols. In *Theoretical Aspects of Computer Software*, volume 1281 of *Lecture Notes in Computer Science*, pages 611–638. Springer-Verlag, 1997.

[2] Martín Abadi and Bruno Blanchet. Secrecy Types for Asymmetric Communication. In F. Honsell and M. Miculan, editors, *Foundations of Software Science and Computation Structures (FoSSaCS 2001)*, volume 2030 of *Lecture Notes in Computer Science*, pages 25–41, Genova, Italy, April 2001. Springer-Verlag.

[3] Martín Abadi and Bruno Blanchet. Analyzing Security Protocols with Secrecy Types and Logic Programs. In *29th Annual ACM SIGPLAN - SIGACT Symposium on Principles of Programming Languages (POPL 2002)*, pages 33–44, Portland, Oregon, January 2002. ACM Press.

[4] Roberto M. Amadio and Witold Charatonik. On name generation and set-based analysis in dolev-yao model. Technical Report 4379, INRIA, 2002.

[5] Roberto M. Amadio and Denis Lugiez. On the reachability problem in cryptographic protocols. In *International Conference on Concurrency Theory*, volume 1877, pages 380–394, 2000.

[6] Franz Baader and Tobias Nipkow. *Term Rewriting and All That.* Cambridge University Press, 1998.

[7] Bruno Blanchet. Abstracting Cryptographic Protocols by Prolog Rules (invited talk). In Patrick Cousot, editor, *8th International Static Analysis Symposium (SAS'2001)*, volume 2126 of *Lecture Notes on Computer Science*, pages 433–436, Paris, France, July 2001. Springer Verlag.

[8] D. Bolignano. Integrating proof-based and model-checking techniques for the formal verification of cryptographic protocols. *Lecture Notes in Computer Science*, 1427:77–87, 1998.

[9] Dominique Bolignano. An approach to the formal verification of cryptographic protocols. In *ACM Conference on Computer and Communications Security*, pages 106–118, 1996.

[10] J. Clark and J. Joacob. A survey on authentification protocol. http://www.cs.york.ac.uk/~jac/papers/drareviewps.ps, 1997.

[11] E.M. Clarke, S. Jha, and W. Marrero. Using state space exploration and a natural deduction style message derivation engine to verify security protocols. In *IFIP Working Conference on Programming Concepts and Methods (PROCOMET)*, 1998.

[12] H. Comon, V. Cortier, and J. Mitchell. Tree automata with one memory, set constraints, and ping-pong protocols. In *Automata, Languages and Programming, 28th International Colloquium, ICALP 2001, Crete, Greece, July 8-12, 2001, Proceedings*, volume 2076 of *Lecture Notes in Computer Science*. Springer, 2001.

[13] H. Comon and V. Shmatikov. Is it possible to decide whether a cryptographic protocol is secure or not? *Journal of Telecommunications and Information Technology.*, 2002.

[14] H. Comon-Lundh and V. Cortier. Security properties: Two agents are sufficient. Technical report, Laboratoire Spécification et Vérification (LSV), ENS de Cachan, August 2002.

[15] V. Cortier, J. Millen, and H. Rueß. Proving secrecy is easy enough. In *Proc. 14th IEEE Computer Security Foundations Workshop (CSFW'01), Cape Breton, Nova Scotia, Canada, June 2001*, pages 97–110. IEEE Comp. Soc. Press, 2001.

[16] D. Dolev and A. C. Yao. On the security of public key protocols. *IEEE Transactions on Information Theory*, 29(2):198–208, 1983.

[17] N. Durgin, P. Lincoln, J. Mitchell, and A. Scedrov. Undecidability of bounded security protocols. In N. Heintze and Proceedings E. Clarke, editors, editors, *Workshop on Formal Methods and Security Protocols — FMSP, Trento, Italy, July 1999.*, 1999.

[18] S. Even and O. Goldreich. On the security of multi-party ping pong protocols. Technical report, Israel Institute of Technology, 1983.

[19] F.J.T. Fábrega, J.C. Herzog, and J.D. Guttman. Strand Spaces: Why is a Security Protocol Correct ? In *Proceedings of the 1998 Conference on Security and Privacy (S&P'98)*, pages 160–171. IEEE, May 1998.

[20] T. Genet and F. Klay. Rewriting for Cryptographic Protocol Verification. In *Proceedings 17th International Conference on Automated Deduction*, volume 1831 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 2000.

[21] A. Gordon and A. Jeffrey. Authenticity by typing for security protocols. In *14th IEEE Computer Security Foundations Workshop (CSFW '01)*, pages 145–159, Washington - Brussels - Tokyo, June 2001. IEEE.

[22] Jean Goubault-Larrecq. A method for automatic cryptographic protocol verification. In Dominique Méry Beverly Sanders, editor, *Fifth International Workshop*

on Formal Methods for Parallel Programming: Theory and Applications (FMPPTA 2000), number 1800 in Lecture Notes in Computer Science. Springer-Verlag, 2000.

[23] G. Lowe. An attack on the Needham-Schroeder public-key authentification protocol. *Information Processing Letters*, 56(3):131–133, November 1995.

[24] G. Lowe. Breaking and fixing the needham-schroeder public-key protocol using fdr. In *TACAS*, number 1055 in Lecture Notes in Computer Science, pages 147–166, 1996.

[25] C. Meadows. Invariant generation techniques in cryptographic protocol analysis. In *PCSFW: Proceedings of The 13th Computer Security Foundations Workshop*. IEEE Computer Society Press, 2000.

[26] J. Millen and V. Shmatikov. Constraint solving for bounded-process cryptographic protocol analysis. In *roc. 8th ACM Conference on Computer and Communications Security (CCS '01)*, pages 166–175, 2001.

[27] J. C. Mitchell, M. Mitchell, and U. Stern. Automated analysis of cryptographic protocols using mur. In *Proceedings of the 1997 Conference on Security and Privacy (S&P-97)*, pages 141–153, Los Alamitos, May 4–7 1997. IEEE Press.

[28] David Monniaux. Abstracting Cryptographic Protocols with Tree Automata. In *Static Analysis Symposium (SAS'99)*, volume 1694 of *Lecture Notes in Computer Science*, pages 149–163. Springer-Verlag, September 1999.

[29] R.M. Needham and M.D. Schroeder. Using encryption for authentication in large networks of computers. *cacm*, 21(12):993–999, 1978.

[30] L. Paulson. Proving properties of security protocols by induction. In *10th IEEE Computer Security Foundations Workshop (CSFW '97)*, pages 70–83, Washington - Brussels - Tokyo, June 1997. IEEE.

[31] D. Prawitz. *Natural Deduction: A Proof-Theoretical Study*. Almquivst & Wikseel, 1965.

[32] M. Rusinowitch and M. Turuani. Protocol insecurity with finite number of sessions is np-complete. In *14th IEEE Computer Security Foundations Workshop (2001), pp. 174–190.*, 2001.

[33] S. Schneider. Verifying authentication protocols with CSP. In *10th IEEE Computer Security Foundations Workshop (CSFW '97)*, pages 3–17, Washington - Brussels - Tokyo, June 1997. IEEE.

[34] J. Thayer, J. Herzog, and J. Guttman. Honest Ideals on Strand Spaces. In *Proceedings of the 11th IEEE Computer Security Foundations Workshop (CSFW '98)*, pages 66–78. IEEE, June 1998.

[35] Christoph Weidenbach. Towards an Automatic Analysis of Security Protocols in First-Order Logic. In Harald Ganzinger, editor, *16th International Conference on Automated Deduction (CADE-16)*, volume 1632 of *Lecture Notes in Artificial Intelligence*, pages 314–328, Trento, Italy, July 1999. Springer-Verlag.

A Proof of the main proposition

Let us first recall the proposition.

Proposition 5.1 *Let $M, S \subseteq \mathcal{T}(\mathcal{F})$ be sets of messages, m be a message and $\mathcal{H} : \mathcal{T}(\mathcal{F}) \to \mathscr{P}(\mathcal{T}(\mathcal{F}))$. If $M[\mathcal{H}]S$ and $m\langle\mathcal{H}\rangle S$ then $M \cup \{m\}[\mathcal{H}]S$.*

Proof Before tackling the proof, we introduce the following definition:

We say that m' is a *derivation-minimal counter-example*, if the following conditions are satisfied:

1. $M, m \vdash m'$,
2. $\neg m'\langle\mathcal{H}\rangle S$ and
3. there is a derivation for $M, m \vdash m'$ which does not contain any strict sub-derivation $M, m \vdash m''$ of a message m'' with $\neg m''\langle\mathcal{H}\rangle S$.

Assume that $M \cup \{m\}[\mathcal{H}]S$ does not hold. Then, there exists a derivation-minimal counter-example m'. The existence of m' can be proved as follows. Take a derivation of $M, m \vdash m'$ and let N_0 be its size. If m' is not a derivation-minimal counter-example then there must exist a sub-derivation $M, m \vdash m'_1$ with $\neg m'_1\langle\mathcal{H}\rangle S$. Clearly, the size N_1 of the derivation tree of m'_1 is strictly smaller than N_0. Repeated application of the same argument must lead to a derivation-minimal counter-example as there are no strictly decreasing chains in $\mathbb{N}$.

Thus, let us come back to our derivation-minimal counter-example m'. We derive a contradiction by case analysis on the last derivation step in $M, m \vdash m'$.

1. $m' \in M \cup \{m\}$. This, contradicts the assumptions $M[\mathcal{H}]S$ and $m\langle\mathcal{H}\rangle S$.
2. Case of encryption: Thus, $m' = \{s\}_k$, $M \cup \{m\} \vdash s$ and $M \cup \{m\} \vdash k$. Since m' is a derivation-minimal counter-example, we have $m' \in S$. Hence, $s \in S$ or $k \in S$ which contradicts the derivation-minimality of m'.
3. Case of pairing: Similar to the previous case.
4. Case of projection: This also contradicts the derivation-minimality assumption.
5. Case of decryption: Then, $M \cup \{m\} \vdash \{m'\}_{k'}$ and $M \cup \{m\} \vdash k'^{-1}$. If $k'^{-1} \notin S$, then $\neg\{m'\}_{k'}\langle\mathcal{H}\rangle S$ which contradicts the derivation-minimality of m'. If $k'^{-1} \in S$ then $M \cup \{m\} \vdash k'^{-1}$ contradicts the derivation-minimality of m'.

□

B Example: The Yahalom Protocol

The aim of the Yahalom protocol (cf. [10] and see Figure 9) is to establish a secret symmetric shared key k_{AB} between two participants A and B using a trusted server S.

$$
\begin{array}{lcl}
A \to B & : & A, N_A \\
B \to S & : & B, \{A, N_A, N_B\}_{k_{BS}} \\
S \to A & : & \{B, k_{AB}, N_A, N_B\}_{k_{AS}}, \{A, k_{AB}\}_{k_{BS}} \\
A \to B & : & \{A, k_{AB}\}_{k_{BS}}, \{N_B\}_{k_{AB}}
\end{array}
$$

Figure 9: The Yahalom protocol

The protocol assumes that A and B already share secure keys k_{AS} respectively k_{BS} with the server S.

The Yahalom protocol can be represented in our setting as follows:

$P = \{p_1, p_2, p_3\}$ with $fresh(p_1) = \{n\}$,$fresh(p_2) = \{n'\}$ and $fresh(p_3) = \{n''\}$. The transitions are described by:

$$
\begin{array}{lcl}
tran(p_1): & & \\
p_1 & \to & p_1, n \\
\{p_2, smk(m, X_1, Y_1), n, Z_1\}_{smk(p_1,p_3)}, W_1 & \to & W_1, \{Z_1\}_{smk(m,X_1,Y_1)} \\
tran(p_2): & & \\
X_2, Y_2 & \to & p_2, \{X_2, Y_2, n'\}_{smk(X_2,p_3)} \\
tran(p_3): & & \\
X_3, \{Y_3, Z_3, W_3\}_{smk(X_3,p_3)} & \to & \{X_3, smk(n'', Y_3, X_3), Z_3, W_3\}_{smk(Y_3,p_3)}, \\
 & & \{Y_3, smk(n'', Y_3, X_3)\}_{smk(X_3,p_3)}
\end{array}
$$

The abstraction of Section 4 yields the following abstract sets:

$$
\begin{array}{l}
P^\sharp = \{A, I\}, \\
K^\sharp = \{K_I, smk(A, A), smk(N_S, A, A), smk(N_S^{AAA}, A, A)\}, \\
N^\sharp = \{N_A, N_A^{AAA}, N_A^{AIA}, N_B, N_B^{AAA}, N_B^{IAA}, N_I\}\}
\end{array}
$$

For the sake of simplicity we write K_{AB} instead of $smk(N_S, A, A)$ respectively K_{AB}^{AAA} instead of $smk(N_S^{AA}, A, A)$. We only present some typical abstract rules in R:

$$\pi = (A, I, A) \quad \frac{I, Y_2}{A, \{I, Y_2, N_A^{IAA}\}_{smk(A,A)}}; \quad \pi = (A, A, A) \quad \frac{A, Y_2}{A, \{A, Y_2, N_A^{AAA}\}_{smk(A,A)}};$$

$$\pi = (X_3, Y_3, A), X_3, Y_3 \in P^\sharp \quad \frac{X_3, \{Y_3, Z_3, W_3\}_{smk(X_3,A)}}{\{X_3, K_{Y_3X_3}^{Y_3X_3A}, Z_3, W_3\}_{smk(Y_3,A)}, \{Y_3, K_{Y_3X_3}^{Y_3X_3A}\}_{smk(X_3,A)}};$$

$$\pi = (A, A, A) \quad \frac{\{A, K_I, N_B^{AAA}, Z_1\}_{smk(A,A)}, W_1}{W_1, \{Z_1\}_{K_I}}; \quad \pi = (A, I, A) \quad \frac{\{I, K, N_B^{AAA}, Z_1\}_{smk(A,A)}, W_1}{W_1, \{Z_1\}_K}, K \in K^\sharp$$

We run our verification algorithm on the set of abstract rules R, the set of secrets $S = \{smk(A, A), N_B, K_{AB}\}$, the empty set of bad patterns, and the set of good patterns $G = \{\{X_S\}_{smk(A,A)}, \{X_S\}_{K_{AB}}\}$.

The algorithm terminates with the set of secrets unchanged and the set of bad patters consisting of:

$\{Sup\,(A, K_I, N_A^{AAA}, X_S)\}_{smk(A,A)}$, $\{Sup\,(I, K_I, N_A^{AIA}, X_S)\}_{smk(A,A)}$, $\{Sup\,(A, K_I, N_A, X_S)\}_{smk(A,A)}$,
$\{Sup\,(I, K_I, N_B^{AIA}, X_S)\}_{smk(A,A)}$, $\{Sup\,(I, X, X_S)\}_{smk(A,A)}$, $\{Sup\,(I, X_S, X)\}_{smk(A,A)}$.

□