Tree Automata for Detecting Attacks on Protocols with Algebraic Cryptographic Primitives

Boichut[1]

INRIA-LANDE
IRISA
Rennes, France

Héam[2] Kouchnarenko[3]

INRIA-CASSIS
LIFC
Besançon, France

Abstract

This paper extends a rewriting approximations-based theoretical framework in which the security problem – secrecy preservation against an active intruder – may be semi-decided through a reachability verification. In a recent paper, we have shown how to semi-decide whether a security protocol using algebraic properties of cryptographic primitives is safe. In this paper, we investigate the dual - insecurity - problem: we explain how to semi-decide whether a protocol using cryptographic primitive algebraic properties is unsafe. The main advantage of our work is that the approximation functions make it possible to automatically verify security protocols with an arbitrary number of sessions. Furthermore, our approach is supported by the tool TA4SP successfully applied for analysing the NSPK-xor protocol and the Diffie-Hellman protocol.

Keywords: Security protocol, algebraic properties, automatic approximation.

1 Introduction

Security protocols are part of systems for which the security problem is in general undecidable. Approximations and abstractions represent a well-suited alternative for verifying them in practice. A lot of investigations have been carried out on this topic [2,12,7,15,17,20,16,19].

An often encountered difficulty is about encoding with non-atomic keys. A non-atomic key is a key established in several steps from several data. This topic comes

[1] Email: boichut@irisa.fr

[2] Email: heamp@lifc.univ-fcomte.fr

[3] Email: kouchna@lifc.univ-fcomte.fr

close to the handling of operators with algebraic properties. On a strongly typed model (model in which the structure of a compound key is clearly specified), most of the developed methods are able to perform a protocol analysis. Unfortunately a secure strongly typed model is not a secure model because of type confusing attacks.

That is why our previous contribution [4] has extended the verification method in [3] in order to verify – without typing – security protocols bringing into play operators with algebraic properties. This improvement has made the computation of sound over-approximations of the intruder knowledge possible. Consequently, the safety, i.e., the secrecy preservation on protocols using algebraic properties of the *exclusive or* (xor) operator or the *exponential* (exp) operator can be established automatically. However, there is a lack of the attack detection, i.e. of showing that a protocol is unsafe.

The main contribution of this paper consists of showing the feasibility of the automatic unsafety verification for protocols when 1) the number of sessions is unbounded, and 2) the cryptographic primitives use algebraic operators properties. We propose sufficient conditions on term rewriting systems (TRSs for short), under which attack detection on such protocols becomes possible.

To illustrate the contributions, experiments on the detection of attacks against protocols with the primitives using xor or exp (xored and exped protocols, for short), are reported.

Structure of the paper The paper is organised as follows. After giving preliminary notions on tree automata and TRSs, we introduce in Section 2 a substitution depending on rules of a TRS, and a notion of compatibility between such substitutions and finite tree automata, both suitable for reachability analysis in rewriting with non left-linear TRSs. In Section 3, we present the extension of [4] dealing with under-approximations. Finally, before concluding, we give in Section 4 a brief overview of related works, and we explain how to apply the obtained new results to analyse xored or exped protocols.

2 Background and Notations

In this section basic notions on finite tree automata, term rewriting systems and approximations are recalled. The reader is referred to [9] for more detail. Moreover, detailled examples are given [5].

2.1 Notations

Given the set $\mathbb{N}$ of natural integers, $\mathbb{N}^*$ denotes the finite strings over $\mathbb{N}$. Let $\mathcal{F}$ be a finite set of symbols with their arities. The set of symbols of $\mathcal{F}$ of arity i is denoted $\mathcal{F}_i$. Let $\mathcal{X}$ be a finite set whose elements are variables. We assume that $\mathcal{X} \cap \mathcal{F} = \emptyset$. A finite ordered tree t over a set of labels $(\mathcal{F}, \mathcal{X})$ is a function from a prefix-closed set $\mathcal{P}\text{os}(t) \subseteq \mathbb{N}^*$ to $\mathcal{F} \cup \mathcal{X}$. A term t over $\mathcal{F} \cup \mathcal{X}$ is a labeled tree whose domain $\mathcal{P}\text{os}(t)$ satisfies the following properties: $\mathcal{P}\text{os}(t)$ is non-empty and prefix closed, for each $p \in \mathcal{P}\text{os}(t)$, if $t(p) \in \mathcal{F}_n$ (with $n \neq 0$), then $\{i \mid p.i \in \mathcal{P}\text{os}(t)\} = \{1, \ldots, n\}$ and, for each $p \in \mathcal{P}\text{os}(t)$, if $t(p) \in \mathcal{X}$ or $t(p) \in \mathcal{F}_0$, then $\{i \mid p.i \in \mathcal{P}\text{os}(t)\} = \emptyset$. Each element of $\mathcal{P}\text{os}(t)$ is called a position of t. For each subset $\mathcal{K}$ of $\mathcal{X} \cup \mathcal{F}$ and each term t we denote by $\mathcal{P}\text{os}_{\mathcal{K}}(t)$ the subset of positions p's of t such that $t(p) \in \mathcal{K}$. Each position p of t such that $t(p) \in \mathcal{F}$, is called a functional position. The set of terms over $(\mathcal{F}, \mathcal{X})$ is denoted $\mathcal{T}(\mathcal{F}, \mathcal{X})$. A ground term is a term t such that $\mathcal{P}\text{os}(t) = \mathcal{P}\text{os}_{\mathcal{F}}(t)$ (i.e. such that $\mathcal{P}\text{os}_{\mathcal{X}}(t) = \emptyset$). The set of ground terms is denoted $\mathcal{T}(\mathcal{F})$. A subterm $t_{|p}$ of $t \in \mathcal{T}(\mathcal{F}, \mathcal{X})$ at position p is defined by: $\mathcal{P}\text{os}(t_{|p}) = \{i \mid p.i \in \mathcal{P}\text{os}(t)\}$ and, For all $j \in \mathcal{P}\text{os}(t_{|p})$, $t_{|p}(j) = t(p.j)$. We denote by $t[s]_p$ the term obtained by replacing in t the subterm $t_{|p}$ by s.

For all sets A and B, we denote by $\Sigma(A, B)$ the set of functions from A to B. If $\sigma \in \Sigma(\mathcal{X}, B)$, then for each term $t \in \mathcal{T}(\mathcal{F}, \mathcal{X})$, we denote by $t\sigma$ the term obtained from t by replacing for each $x \in \mathcal{X}$, the variable x by $\sigma(x)$. A term rewriting system $\mathcal{R}$ over $\mathcal{T}(\mathcal{F}, \mathcal{X})$ is a finite set of pairs (l, r) from $\mathcal{T}(\mathcal{F}, \mathcal{X}) \times \mathcal{T}(\mathcal{F}, \mathcal{X})$, denoted $l \to r$, such that the set of variables occurring in r is included in the set of variables of l. A TRS is left-linear if for each rule $l \to r$, every variable occurs at most once in l. For each ground term t, we denote by $\mathcal{R}(t)$ the set of ground terms $t^{'}$ such that there exist a rule $l \to r$ of $\mathcal{R}$, a function $\mu \in \Sigma(\mathcal{X}, \mathcal{T}(\mathcal{F}))$ and a position p of t satisfying $t_{|p} = l\mu$ and $t^{'} = t[r\mu]_p$. The relation $\{(t, t^{'}) \mid t^{'} \in \mathcal{R}(t)\}$ is classically denoted $\to_{\mathcal{R}}$. If $t \to_{\mathcal{R}} t'$ for $t, t' \in \mathcal{T}(\mathcal{F})$, then t is a *rewriting predecessor* of t' and t' is *rewriting successor* of t. For each set of ground terms B we denote by $\mathcal{R}^*(B)$ the set of ground terms related to an element of B modulo the reflexive-transitive closure of $\to_{\mathcal{R}}$.

A tree automaton $\mathcal{A}$ is a tuple $(\mathcal{Q}, \Delta, F)$, where $\mathcal{Q}$ is the set of states, Δ the set of transitions, and F the set of final states. Transitions are rewriting rules of the form $f(q_1, \ldots, q_k) \to q$, where $f \in \mathcal{F}_k$ and the q_i's are in $\mathcal{Q}$. A term $t \in \mathcal{T}(\mathcal{F})$ is accepted or recognised by $\mathcal{A}$ if there exists $q \in F$ such that $t \to_{\Delta}^{*} q$ (we also write $t \to_{\mathcal{A}}^{*} q$). The set of terms accepted by $\mathcal{A}$ is denoted $\mathcal{L}(\mathcal{A})$. For each state $q \in \mathcal{Q}$, we write $\mathcal{L}(\mathcal{A}, q)$ for the tree language $\mathcal{L}((\mathcal{Q}, \Delta, \{q\}))$. A tree automaton is finite if its set of transitions is finite.

In [4], a new kind of substitution has been introduced. We recall this definition below. Notice that the domain of these substitutions is not the set of variables anymore, but a set of positions. Thus, given a variable, this allows a symbolic representation of its values.

Definition 2.1 Let $\mathcal{R}$ be a term rewriting system, $\mathcal{Q}$ a set of states and $l \to r \in \mathcal{R}$. An $(l \to r)$-substitution is a function from $\mathcal{P}\text{os}_{\mathcal{X}}(l)$ into $\mathcal{Q}$.

We then adapt this kind of substitution to the rewriting framework in the following way. Let $l \to r \in \mathcal{R}$ and σ be an $(l \to r)$-substitution. We denote by $l\sigma$ the term of $\mathcal{T}(\mathcal{F}, \mathcal{Q})$ such that $\mathcal{P}\text{os}(l\sigma) = \mathcal{P}\text{os}(l)$, and for each $p \in \mathcal{P}\text{os}(l)$, if $p \in \mathcal{P}\text{os}_{\mathcal{X}}(l)$ then $l\sigma(p) = \sigma(p)$, otherwise $l\sigma(p) = l(p)$. Similarly, we denote by $r\sigma$ the term of $\mathcal{T}(\mathcal{F}, \mathcal{Q})$ defined by: $\mathcal{P}\text{os}(r\sigma) = \mathcal{P}\text{os}(r)$ and, for each $p \in \mathcal{P}\text{os}(r)$, if $p \notin \mathcal{P}\text{os}_{\mathcal{X}}(r)$ then $r\sigma(p) = r(p)$ and $r\sigma(p) = \sigma(p^{'})$ otherwise, where $p^{'} = \min \mathcal{P}\text{os}_{r(p)}(l)$ (positions are lexicographically ordered). For a given tree automaton, a particular class of $(l \to r)$-substitutions can be drawn.

Definition 2.2 Let $\mathcal{A}$ be a finite tree automaton. We say that an $(l \to r)$-

substitution σ is $\mathcal{A}$-compatible if for each $x \in \mathcal{V}\mathrm{ar}(l)$,

$$\bigcap_{p \in \mathcal{P}\mathrm{os}_{\{x\}}(l)} \mathcal{L}(\mathcal{A}, \sigma(p)) \neq \emptyset.$$

Finally, the last notion we introduce is the definition of an approximation function.

Definition 2.3 Let $\mathcal{A}$ be a finite tree automaton. An approximation function (for $\mathcal{A}$) is a function associating with each tuple $(l \to r, \sigma, q)$, where $l \to r \in \mathcal{R}$, σ is an $\mathcal{A}$-compatible $(l \to r)$-substitution and q a state of $\mathcal{A}$, a mapping from $\mathcal{P}\mathrm{os}(r)$ to $\mathcal{Q}$.

This notion is very useful for reachability analysis in rewriting with non left-linear TRSs as shown in the following section.

2.2 Reachability Analysis in Rewriting with non Left-linear TRSs

This section recalls the approximation-based framework we have been developing, and explains our objectives from a formal point of view.

Given a tree automaton $\mathcal{A}$ and a TRS $\mathcal{R}$ (for several classes of automata and TRSs), the tree automata completion [15,14] algorithm computes a tree automaton $\mathcal{A}_k$ such that $\mathcal{L}(\mathcal{A}_k) = \mathcal{R}^*(\mathcal{L}(\mathcal{A}))$ when it is possible (for the classes of TRSs covered by this algorithm see [14]), and such that $\mathcal{L}(\mathcal{A}_k) \supseteq \mathcal{R}^*(\mathcal{L}(\mathcal{A}))$ otherwise.

The tree automata completion works as follows. From $\mathcal{A} = \mathcal{A}_0$ completion builds a sequence $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_k$ of automata such that if $s \in \mathcal{L}(\mathcal{A}_i)$ and $s \rightarrow_{\mathcal{R}} t$ then $t \in \mathcal{L}(\mathcal{A}_{i+1})$. If there is a fix-point automaton $\mathcal{A}_k$ such that $\mathcal{R}^*(\mathcal{L}(\mathcal{A}_k)) = \mathcal{L}(\mathcal{A}_k)$, then one has $\mathcal{L}(\mathcal{A}_k) = \mathcal{R}^*(\mathcal{L}(\mathcal{A}_0))$ (or $\mathcal{L}(\mathcal{A}_k) \supseteq \mathcal{R}^*(\mathcal{L}(\mathcal{A}))$ if $\mathcal{R}$ is not in one class of [14]). In particular, for non left-linear TRSs, the completion is not sound. Indeed, if the completion converges towards a fix-point automaton $\mathcal{A}_k$, $\mathcal{L}(\mathcal{A}_k)$ is not necessarily either $\mathcal{R}^*(\mathcal{L}(\mathcal{A}))$ or a super set of $\mathcal{R}^*(\mathcal{L}(\mathcal{A}))$.

In [4], the completion procedure has been improved so that the method is sound for non left-linear TRSs. This technique is introduced below. As mentioned previously, the completion builds a sequence $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_k$ of tree automata such that the set of terms reachable in one step of rewriting from $\mathcal{L}(\mathcal{A}_i)$ are in $\mathcal{L}(\mathcal{A}_{i+1})$. To build $\mathcal{A}_{i+1}$ from $\mathcal{A}_i$, we achieve a *completion step* which consists of finding *critical pairs* between $\rightarrow_{\mathcal{R}}$ and $\rightarrow_{\mathcal{A}_i}$. Formally, for an approximation function γ, a rule $l \rightarrow r \in \mathcal{R}$ and an $\mathcal{A}_i$-compatible $(l \to r)$-substitution σ, a critical pair is an instance $l\sigma$ of l such that there exists $q \in \mathcal{Q}$ satisfying $l\sigma \rightarrow^*_{\mathcal{A}_i} q$ and $r\sigma \not\rightarrow^*_{\mathcal{A}_i} q$. For every critical pair, such that $l\sigma \rightarrow^*_{\mathcal{A}_i} q$ and $r\sigma \not\rightarrow^*_{\mathcal{A}_i} q$, detected between $\mathcal{R}$ and $\mathcal{A}_i$, $\mathcal{A}_{i+1}$ is built by adding new transitions to $\mathcal{A}_i$, so that it recognizes $r\sigma$ in q, i.e. $r\sigma \rightarrow_{\mathcal{A}_{i+1}} q$.

$$\begin{tikzcd}
l\sigma \arrow[r, "\mathcal{R}"'] \arrow[d, "\mathcal{A}_i"', "*"] & r\sigma \arrow[dl, dashed, bend left, "\mathcal{A}_{i+1}", "*"'] \\
q &
\end{tikzcd}$$

Before giving a definition of a completion step (Def. 2.5), we introduce a normalisation step described in Definition 2.4. Note that the transition $r\sigma \rightarrow q$ is not necessarily a transition of the form $f(q_1, \dots, q_n)\rightarrow q'$ and so has to be normalized first. For example, to normalize a transition of the form $f(g(a), h(q'))\rightarrow q$, we need to find some states q_1, q_2, q_3 and replace the previous transition by a set of normalized transitions: $\{a\rightarrow q_1, g(q_1)\rightarrow q_2, h(q')\rightarrow q_3, f(q_2, q_3)\rightarrow q\}$. The states used in a normalization step do not grow on trees and it is of the approximation function γ concern to deliver them at each completion step. Formally,

Definition 2.4 Let $\mathcal{A} = (\mathcal{Q}_0, \Delta, F_0)$ be a finite tree automaton, γ be an approximation function for $\mathcal{A}$, $l \rightarrow r$ be a rule of $\mathcal{R}$, σ be an $\mathcal{A}$-compatible $(l \rightarrow r)$-substitution, and q be a state of $\mathcal{A}$. We denote by $\mathrm{Norm}_\gamma(l \rightarrow r, \sigma, q)$ the following set of transitions, called *normalization* of $(l \rightarrow r, \sigma, q)$:

$$\begin{array}{ll}\{f(q_1, \dots, q_k)\rightarrow q^{'} & |p \in \mathcal{P}\mathrm{os}_{\mathcal{F}}(r),\ r(p) = f, \\ & q^{'} = q \text{ if } p = \varepsilon \text{ otherwise } q^{'} = \gamma(l \rightarrow r, \sigma, q)(p) \\ & q_i = \gamma(l \rightarrow r, \sigma, q)(p.i) \text{ if } p.i \notin \mathcal{P}\mathrm{os}_{\mathcal{X}}(r), \\ & q_i = \sigma(\min\{p^{'} \in \mathcal{P}\mathrm{os}_{\mathcal{X}}(l) \mid l(p^{'}) = r(p.i)\})\text{otherwise}\}\end{array}$$

The min is computed for the lexical order.

Notice that the set $\{p^{'} \in \mathcal{P}\mathrm{os}_{\mathcal{X}}(l) \mid l(p^{'}) = r(p.i)\}$ used in the above definition is not empty. Indeed, in a TRS, variables occurring in the right-hand side must, by definition, occur in the left-hand side too.

Definition 2.5 Let $\mathcal{R}$ be a TRS. Let $\mathcal{A}_0 = (\mathcal{Q}_0, \Delta_0, F_0)$ be a finite tree automaton and γ an approximation function for $\mathcal{A}_0$. The automaton $\mathcal{C}_\gamma(\mathcal{A}_0) = (\mathcal{Q}_1, \Delta_1, F_1)$ is defined by:

$$\Delta_1 = \Delta_0 \cup \bigcup \mathrm{Norm}_\gamma(l \rightarrow r, \sigma, q)$$

where the union involves all rules $l \rightarrow r \in \mathcal{R}$, all states $q \in \mathcal{Q}_0$, all $\mathcal{A}_0$-compatible $(l \rightarrow r)$-substitutions σ such that $l\sigma\rightarrow^*_{\mathcal{A}_0} q$ and $r\sigma \not\rightarrow^*_{\mathcal{A}_0} q$, $F_1 = F_0$ and $\mathcal{Q}_1 = \mathcal{Q}_0 \cup \mathcal{Q}_2$, where $\mathcal{Q}_2$ denotes the set of states occurring in left/right-hand sides of transitions of Δ_1.

Following theorem was proved in [4].

Theorem 2.6 *Let $(\mathcal{A}_n)$ and (γ_n) be respectively a sequence of finite tree automata and a sequence of approximation functions such that for each integer n, γ_n is an approximation function for $\mathcal{A}_n$ and $\mathcal{A}_{n+1} = \mathcal{C}_{\gamma_n}(\mathcal{A}_n)$. If there exists a positive integer N, such that for every $n \geq N$, $\mathcal{A}_n = \mathcal{A}_N$, then $\mathcal{R}^*(\mathcal{L}(\mathcal{A}_0)) \subseteq \mathcal{L}(\mathcal{A}_N)$.*

From a verification point of view, this technique is very helpful. Indeed, for a system Σ whose transition relation is Δ, one specifies the initial configuration of Σ by a tree language E, and Δ by a TRS $\mathcal{R}$. With a well-suited approximation function γ, an over-approximation of reachable configurations of Σ, denoted $E^\gamma_{\mathcal{R}}$, can be computed. Finally, a set of bad configurations, denoted E_{Bad}, can be encoded with a tree language and if $E^\gamma_{\mathcal{R}} \cap E_{Bad}$ is empty, then no bad configuration is reachable.

In particular, in [4], we have used this technique for verifying security protocols bringing into play the `xor` operator ($\oplus$). Note that the nilpotence property of $\oplus$ is specified with a non left-linear rule, i.e., $x \oplus x\rightarrow 0$. The tree languages specify the intruder knowledge and the configurations of the network. The TRS specifies the protocol and the intruder abilities for decoding, coding, depairing messages. Thus, if a secret term t does not belong to an over-approximation of the knowledge that the intruder might have, then t is actually secret.

3 Under-Approximations for non Left-linear TRSs

The over-approximation results in [4] do not provide a way to prove that a particular term is reachable: the method is not complete. This section adapts the means and extends the results in [4] to under-approximations computations. In the security protocol framework, computing under-approximations allows an under-estimation of the intruder knowledge, and thus secrecy flaws detection. Indeed, if a secret datum is in the intruder knowledge under-estimation, then the intruder actually knows this secret. The main idea (and problem) behind the under-approximations is that one wants the languages of computed tree automata to be in the set of terms reachable by rewriting . Having some conditions on the TRS makes it possible to control the completion, and proving that a term is actually reachable is then possible.

We define here γ to be an injective approximation function from $\mathcal{R} \times (\mathbb{N}^* \mapsto \mathcal{Q}) \times \mathbb{N}^* \times \mathcal{Q}$ into $\mathcal{Q}$. Theorem 3.2 shows that with such an approximation function, an under-approximation of the set of reachable terms is possible. Before, Lemma 3.1 presents an intermediary result useful for proving Theorem 3.2: this result reveals some features of terms recognised by $\mathcal{C}_\gamma(\mathcal{A})$ for which there exists a rewriting predecessor recognised by $\mathcal{A}$. In the following, we introduce the notation $NLV(t)$ which for a term t of $\mathcal{T}(\mathcal{F},\mathcal{X})$, denotes the set of non-linear variables of t, i.e., the set of variables occurring at least twice within t.

Lemma 3.1 *Let $\mathcal{R}$ be a right-linear TRS for which $NLV(l) \cap \mathcal{V}\mathrm{ar}(r) = \emptyset$ for all $l \to r \in \mathcal{R}$. Let $\mathcal{A}$ be the current tree automaton and $\mathcal{C}_\gamma(\mathcal{A})$ be the tree automaton obtained after one completion step with $\mathcal{R}$ and γ. If there exist a ground term t over $\mathcal{F}$, a state q of $\mathcal{A}$ and a function τ from $\mathcal{P}\mathrm{os}(t)$ to Q such that $t \in \mathcal{L}(\mathcal{C}_\gamma(\mathcal{A}), q)$, $t \notin \mathcal{L}(\mathcal{A}, q)$ and τ satisfies the following conditions: (i) $\tau(\varepsilon) = q$; (ii) for all $p \in \mathcal{P}\mathrm{os}(t)$, $t_{|p} \in \mathcal{L}(\mathcal{C}_\gamma(\mathcal{A}), \tau(p))$ and, (iii) for all $p \in \mathcal{P}\mathrm{os}(t) \setminus \{\varepsilon\}$, if $\tau(p)$ is a state of $\mathcal{A}$, then $t_{|p} \in \mathcal{L}(\mathcal{A}, \tau(p))$. Then there exists $t_0 \in \mathcal{T}(\mathcal{F})$ such that $t_0 \in \mathcal{L}(\mathcal{A}, q)$ and $t_0 \to_{\mathcal{R}} t$.*

The proof of Lemma 3.1 is in [5]. The following result shows that each term of the language $\mathcal{C}_\gamma(\mathcal{A}_0)$ is reachable by rewriting from $\mathcal{A}_0$ and using $\mathcal{R}$.

Theorem 3.2 *Let $\mathcal{A}_0 = (\mathcal{Q}_0, \Delta_0, F_0)$ be a finite tree automaton. Let $\mathcal{R}$ be a right-linear TRS. Given the approximation function γ defined at the beginning of Section 3, if for all $l \to r \in \mathcal{R}$, $\mathcal{V}\mathrm{ar}(r) \cap NLV(l) = \emptyset$ then $\mathcal{L}(\mathcal{C}_\gamma(\mathcal{A}_0)) \subseteq \mathcal{R}^*(\mathcal{L}(\mathcal{A}_0))$.*

The proof of Theorem 3.1 can be found in [5]. Let $\mathcal{C}_\gamma^{(n)}(\mathcal{A}_0)$ be the tree automaton obtained after n completion steps performed from $\mathcal{A}_0$ by using the TRS $\mathcal{R}$ and the approximation function γ. Finally, Proposition 3.3 shows that the approximation function γ provides a sound under-approximation of reachable terms (see [5]).

Proposition 3.3 *If $\mathcal{R}$ is right-linear and for all $l \to r \in \mathcal{R}$, $NLV(l) \cap \mathcal{V}\mathrm{ar}(r) = \emptyset$*

then for all $n \leq 0$*,* $\mathcal{L}(\mathcal{C}_\gamma^{(n)}(\mathcal{A}_0)) \subseteq \mathcal{R}^*(\mathcal{L}(\mathcal{A}_0))$*,* $\mathcal{L}(\mathcal{C}_\gamma^{(n)}(\mathcal{A}_0)) \subseteq \mathcal{L}(\mathcal{C}_\gamma^{(n+1)}(\mathcal{A}_0))$ *and* $\bigcup_{n\geq 0} \mathcal{L}(\mathcal{C}_\gamma^{(n)}(\mathcal{A}_0)) = \mathcal{R}^*(\mathcal{L}(\mathcal{A}_0))$*.*

At this point, we have developed theoretical frameworks which lead either to over-approximations of the set of reachable terms in general, or to its under-approximations under additional conditions on TRSs. The obtained results allow us to apply the approximation-based methods to system verification as presented in the next section.

4 Experiments and Related Works

With the extension brought for the under-approximations computation, we are now able to detect whether a protocol using algebraic properties of cryptographic primitives is flawed or not. Actually, while the protocol is flawed in the rewriting model for any number of sessions, in general, it turns out to be a real attack against the protocol when an attack concerns a long term secret between two honnest agents. We consider a long term secret such as a secret which is never revealed, even in future sessions. In this section, we present some experimental results obtained on two protocols, well-known to be flawed, which are NSPK-xor and the key establishment à la Diffie-Helmann protocol. The technique presented in this paper has been implemented in the tool `TA4SP`.

4.1 TA4SP for Attack Detection

This section details two protocols, well-known to be flawed, which are NSPK-xor and the key establishment à la Diffie-Helmann protocol in its simplest form. The notations used are the following: `X -> Y: Z` specifies that `X` sends the message `Z` to `Y`, `X.Y` is the concatenation of data `X` and `Y`, and $\{\mathtt{X}\}_{\mathtt{Y}}$ (or `{X}_Y`) is the encoding of the message `X` by the message `Y`. Moreover, data `Na`, `Nb`, `ni(Na)` and `ni(Nb)` with `i` being an integer, are fresh random numbers, also called a *nonces*. Finally, the last concept to know concerns the keys, which can be public, private or symmetric. To a public key `Pka` is associated a private key `Prka`. A message encoded by one can be decoded by the other: $\{\{\mathtt{M}\}_{\mathtt{Pka}}\}_{\mathtt{Prka}} = \{\{\mathtt{M}\}_{\mathtt{Prka}}\}_{\mathtt{Pka}} = \mathtt{M}$. A symmetric key `K` can decode a message encoded by itself: $\{\{\mathtt{M}\}_{\mathtt{K}}\}_{\mathtt{K}} = \mathtt{M}$.

The NSPK-xor Protocol is composed of three steps so that each participant can authenticate the other. First, the agent A sends the message $\{Na.A\}_{K_B}$ to the agent B. Second, B sends $\{Nb.Na \oplus B\}_{K_A}$ to A. Finally, A sends $\{Nb\}_{K_B}$ to B as a confirmation. Using `TA4SP`, we obtain in 71.03 seconds that the protocol does not preserve the secrecy of the data Nb against an intruder. Thanks to the `AVISPA` toolset, one can use one of three other tools (in this case CL-AtSe [21]) for exhibiting the following attack trace.

```
1. i -> (a,6):  start
2. (a,6) -> i:  {n9(Na).a}_ki

3. i -> (a,3):  start
4. (a,3) -> i:  {n1(Na).a}_kb

5. i -> (b,4):  {xor(i,xor(b,n9(Na))).a}_kb
6. (b,4) -> i:  {n5(Nb).xor(i,n9(Na))}_ka
```

```
7. i -> (a,6):  {n5(Nb).xor(i,n9(Na))}_ka
8. (a,6) -> i:  {n5(Nb)}_ki
```

At steps `1.` and `2.` of the attack, the agent `a` initiates a session with the intruder by sending the message `{n9(Na).a}_ki` to the intruder where `n9(Na)` is a nonce generated by `a` and `ki` is the public key of the intruder. At steps `3.` and `4.`, the agent `a` initiates a session with the agent `b`. The intruder composes at step `5.` the message `xor(i,xor(b,n9(Na))).a` and sends it to `b` after having encoded it with the public key of the agent `b`. The agent `b` deduces at step `6.` that this message comes from the agent `a` thanks to the identity occurring in the received message. Moreover, `b` considers the message xor(i,xor(b,n9(Na)))' as the nonce generated by `a`. Consequently, `b` performs the second step of the protocol. At step `6.` of the attack trace, `b` composes `n5(Nb).xor(b,xor(i,xor(b,n9(Na))))` which is equivalent to `n5(Nb).xor(i,n9(Na))` after considering the algebraic properties of $\oplus$ (xor operator). Then, he sends it to `a` after having encoded it with the public key of `a`. The agent `b` declares also the nonce `n5(Nb)` as a secret shared between himself and the agent `a`. But, according to the point of view of the agent `a`, the message `{n5(Nb).xor(i,n9(Na))}_ka` should come from `i` (the intruder) because `n5(Nb)` identifies the agent `i` for `a`. According to his deduction, the agent `a` sends `{n5(Nb)}_ki` to the intruder. Finally, the latter can deduce `n5(Nb)` which is a secret supposed to be shared between `b` and `a`.

The Diffie-Helmann Protocol is a key establishment protocol between two agents A and B. The simplest version of this protocol is composed of three steps. At step 1, A generates the nonce Na and computes $exp(G, Na)$ (standing for G^{Na}) where G is a number known by every agents. Thus A sends the message $exp(G, Na)$ to the agent B. At step 2, the agent B generates also a number Nb and computes on the one hand $exp(G, Nb)$ and on the other hand $K = exp(X, Nb)$ where X is the message received i.e. $exp(G, Na)$. The former is sent to A and the latter stands for the symmetric key shared between A and B. As soon as B receives the message $exp(G, Nb)$ from A, (s)he then computes $exp(exp(G, Nb), Na)$ and thus considers it as the symmetric key shared with A. Indeed, according to the algebraic properties of the exponentiation, $K = exp(exp(G, Na), Nb) = exp(exp(G, Nb), Na)$. Finally, the message $\{secret\}_K$ is sent by A to B in which *secret* is a datum initially known uniquely by A and B. Using `TA4SP` this protocol has been shown as being flawed in 24.73 seconds. For this protocol, a MIM (*Man in the Middle*) attack is known and is detailed below with the attack trace outputted with the `AVISPA` tool-set.

```
1. i -> (a,3):  start
2. (a,3) -> i:  exp(g,n1(Na))

3. i -> (b,4):  g
4. (b,4) -> i:  exp(g,n5(Nb))

5. i -> (a,3):  g
6. (a,3) -> i:  {secab}_(exp(g,n1(Na)))

7. i -> (b,4):  {secab}_(exp(g,n5(Nb)))
8. (b,4) -> i:  ()
```

Roughly, the intruder establishes two keys: `exp(exp(g,n1(Na)),g)` with `a` at steps `2` and `5` and `exp(exp(g,n5(Nb)),g)` with `b` at steps `3` and `4`. At step `6`, the agent a sends the secret data to b with the key unfortunately shared with the intruder. The intruder then extracts the secret data and forwards it to b with the other key. Finally, b is persuaded that this message comes from a.

4.2 Related Work

In [18] it has been shown that using equational tree automata under associativity and/or commutativity is relevant for security problems of cryptographic protocols with an equational property. For protocols modeled by associative-commutative TRSs, the authors announce the possibility for the analysis to be done automatically thanks to the tool ACTAS manipulating associative-commutative tree automata and using approximation algorithms. However, the engine has still room to be modified and optimised to support an automated verification.

In [11], the authors study the IBM 4758 CCA (Common Cryptographic Architecture) API which has been shown as flawed in [6]. In response to this flaw, IBM then has proposed three recommendations designed to prevent it. The formalisation of these recommendations leads Cortier et al. to draw up a particular class of security protocols using the operator $\oplus$ for which the secrecy problem is decidable with an unbounded number of sessions. They have then shown that any one of the three recommendations is sufficient to secure the API against a Dolev-Yao intruder [13].

In the recent survey [10], the authors give an overview of the existing methods in formal approaches to analyse cryptographic protocols. In the same work, a list of some relevant algebraic properties of cryptographic operators is established, and for each of them, the authors provide examples of protocols or attacks using these properties. This survey lists two drawbacks with the recent results aiming at the analysis of protocols with algebraic properties. First, in most of the papers a particular decision procedure is proposed for a particular property. Second, the authors emphasise the fact that the results remain theoretical, and very few implementations automatically verify protocols with algebraic properties.

5 Conclusion

The main purpose of this paper is to show that the symbolic approximation-based approach we have been developing is well-adapted for detecting attacks on protocols using algebraic properties while considering an unbounded number of sessions. Indeed, the automatically generated symbolic under-approximation function enables us 1) an automated normalisation of transitions, and 2) an automated completion procedure within the set of reachable terms.

With this extension our approximation-based framework proposes verification methods using either over-approximations of the set of reachable terms in general, or its under-approximations under additional conditions on TRSs. The contributions of the paper have been integrated into the push-button tool TA4SP [1] successfully applied for analysing the NSPK-xor protocol and the Diffie-Hellman protocol. Let us remark that TA4SP is used for protocols specified in the standard High Level Protocol Specification Language (HLPSL) [8]. This language is known to be suitable for industrial users.

Future development concerns implementation optimisation. We intend to investigate further algebraic properties that can be handled in practice. In this direction, we project to develop a theoretical framework in order to compute under-approximations without the right-linearity condition required Theorem 3.2. This may for example provide an approximation-based approach for detecting attacks on security protocols with cryptographic primitives using the homomorphism property [10].

References

[1] A. Armando, D. Basin, Y. Boichut, Y. Chevalier, L. Compagna, J. Cuellar, P. Hankes Drielsma, P.-C. Héam, O. Kouchnarenko, J. Mantovani, S. Mödersheim, D. von Oheimb, M. Rusinowitch, J. Santos Santiago, M. Turuani, L. Viganò, and L. Vigneron. The AVISPA Tool for the automated validation of internet security protocols and applications. In *CAV 2005, Proceedings*, volume 3576 of *Lecture Notes in Computer Science*, pages 281–285. Springer, 2005.

[2] B. Blanchet. An efficient cryptographic protocol verifier based on prolog rules. In *CSFW 2001*, pages 82–96. IEEE Computer Society Press, 2001.

[3] Y. Boichut, P.-C. Héam, and O. Kouchnarenko. Automatic Verification of Security Protocols Using Approximations. Research Report RR-5727, INRIA-Lorraine - CASSIS Project, October 2005.

[4] Y. Boichut, P.-C. Héam, and O. Kouchnarenko. Handling algebraic properties in automatic analysis of security protocols. In *ICTAC-06*, volume 4281 of *Lecture Notes in Computer Science*, pages 153–167. Springer Berlin/Heidelberg, 2006.

[5] Y. Boichut, P.-C. Héam, and O. Kouchnarenko. Tree automata for detecting attacks on protocols with algebraic cryptographic primitives. Available at *http://www.irisa.fr/lande/boichut/publications/*, 2007.

[6] M. Bond. Attacks on cryptoprocessor transaction sets. In *CHES 2001, Proceedings*, Lecture Notes on Computer Science, pages 220–234. Springer Verlag, 2001.

[7] L. Bozga, Y. Lakhnech, and M. Perin. Pattern-based abstraction for verifying secrecy in protocols. In *TACAS 2003*, volume 2619 of *Lecture Notes in Computer Science*. Springer-Verlag, 2003.

[8] Y. Chevalier, L. Compagna, J. Cuellar, P. Hankes Drielsma, J. Mantovani, S. Mödersheim, and L. Vigneron. A high level protocol specification language for industrial security-sensitive protocols. In *SAPS 2004*, 2004.

[9] H. Comon, M. Dauchet, R. Gilleron, F. Jacquemard, D. Lugiez, S. Tison, and M. Tommasi. Tree automata techniques and applications, 2002.

[10] V. Cortier, S. Delaune, and P. Lafourcade. A survey of algebraic properties used in cryptographic protocols. *Journal of Computer Security*, 14:1–43, 2006.

[11] V. Cortier, G. Keighren, and G. Steel. Automatic analysis of the security of xor-based key management schemes. In *TACAS 2007*, 2007. To be published in Lecture Notes on Computer Science.

[12] V. Cortier, J. K. Millen, and H. Rueß. Proving secrecy is easy enough. In *14th IEEE Computer Security Foundations Workshop, CSFW 2001, Proceedings*, pages 97–110. IEEE, June 2001.

[13] D. Dolev and A. Yao. On the Security of Public-Key Protocols. *IEEE Transactions on Information Theory*, 2(29), 1983.

[14] G. Feuillade, T. Genet, and V. VietTriemTong. Reachability analysis over term rewriting systems. *Journal of Automated Reasonning*, 33 (3-4), 2004.

[15] Th. Genet and F. Klay. Rewriting for Cryptographic Protocol Verification. In *CADE 2000, Proceedings*, volume 1831 of *Lecture Notes in Computer Science*, pages 271–290. Springer-Verlag, 2000.

[16] C. Meadows. The NRL protocol analyser: An overview. *Journal of Logic Programming*, 1994.

[17] D. Monniaux. Abstracting cryptographic protocols with tree automata. In *SAS 1999*, volume 1694 of *Lecture Notes in Computer Science*. Springer-Verlag, 1999.

[18] H. Ohsaki and T. Takai. Actas: A system design for associative and commutative tree automata theory. In *RULE'2004*, volume 124, Aachen, Germany, June 2004.

[19] L. C. Paulson. The inductive approach to verifying cryptographic protocols. *Journal of Computer Security*, 6(1):85–128, 1998.

[20] D. Song. Athena: A new efficient automatic checker for security protocol analysis. In *CSFW 1999, Proceedings*, pages 192–202. IEEE Computer Society Press, 1999.

[21] M. Turuani. The cl-atse protocol analyser. In *RTA 2006*, volume 4098 of *Lecture Notes in Computer Science*, pages 277–286. Springer-Verlag, 2006.