The intruder deduction problem in presence of the XOR theory

For simplicity of presentation we only consider messages built from constants using pairing and symmetric encryption.

Recall the following rules we previously considered for the intruder deduction problem.

$$\text{(M)} \qquad \frac{}{T \vdash m}\ m \in T$$

$$\text{(P)} \qquad \frac{T \vdash m_1 \quad T \vdash m_2}{T \vdash \langle m_1, m_2 \rangle}$$

$$\text{(E)} \qquad \frac{T \vdash m_1 \quad T \vdash m_2}{T \vdash \{m_1\}_{m_2}}$$

$$\text{(UL)} \quad \frac{T \vdash \langle m_1, m_2 \rangle}{T \vdash m_1} \qquad \text{(UR)} \quad \frac{T \vdash \langle m_1, m_2 \rangle}{T \vdash m_2}$$

$$\text{(D)} \quad \frac{T \vdash \{m_1\}_{m_2} \quad T \vdash m_2}{T \vdash m_1}\, m_2 \text{ is symmetric}$$

We now add the new rule for XOR.

$$\text{(X)} \quad \frac{T \vdash m_1 \ldots T \vdash m_k}{T \vdash m_1 + \ldots + m_k}$$

For any t let $t\downarrow$ be its normal form obtained by applying repeatedly the cancellation rules and removing the 0 symbol. We assume that in the above rules every term is implicitly normalized after applying each rule.

Define $sub(T)$ to be the least set such that

- If $t \in T$ then $t \in sub(T)$
- If $\langle u, v \rangle \in sub(T)$ then $u, v \in sub(T)$
- If $\{u\}_v \in sub(T)$ then $u, v \in sub(T)$
- If $u_1 + \ldots + u_n \in sub(T)$ and u_i are not headed with $+$ then $u_1, \ldots, u_n \in sub(T)$

Define a minimal derivation using the intruder deduction rules to be the one of smallest size (the number of applications of the rules).

Observation 1 If a minimal derivation δ has an analysis rule at the end, i.e. it is of one of the following forms:

$$\dfrac{\dfrac{\delta_1 \ldots \delta_n}{T \vdash \langle u, v\rangle}}{T \vdash u} \qquad \dfrac{\dfrac{\delta_1 \ldots \delta_n}{T \vdash \langle v, u\rangle}}{T \vdash u} \qquad \dfrac{\dfrac{\delta_1 \ldots \delta_n}{T \vdash \{u\}_v} \quad \begin{matrix}\vdots \\ T \vdash v\end{matrix}}{T \vdash u}$$

Then $\langle u, v\rangle \in sub(T)$ (resp. $\langle v, u\rangle \in sub(T)$, resp. $\{u\}_v \in sub(T)$)

We consider δ to be of the first form. The other cases are similar.

We claim that for any subtree

$$\delta' = \frac{\delta'_1 \ldots \delta'_m}{T \vdash w}$$

occurring in one of $\delta'_1, \ldots, \delta'_m$, if $\langle u, v \rangle$ is a subterm of w and $w \notin T$ then for at least one i,

the root of δ'_i is labeled with $T \vdash w'$ and $\langle u, v \rangle$ is a subterm of w'.

To show this consider the last rule used in δ'. If it is (M), (UL), (UR) or (D) then the claim is obvious.

If the last rule used is (X), i.e. $w = (u_1 + \ldots + u_n) \downarrow$, then since $\langle u, v \rangle$ is a subterm of w, it is also a subterm of one of the u_i.

If the last rule used is (P) then $w = \langle w_1, w_2 \rangle \neq \langle u, v \rangle$ because δ is minimal. Hence $\langle u, v \rangle$ is a subterm of w_1 or w_2. The case where the last rule is (E) is similar.

This proves the claim.

It follows that at least one leaf node of the derivation δ is labeled with $T \vdash w$ with $w \in T$ and $\langle u, v \rangle$ a subterm of w. Hence $\langle u, v \rangle \in sub(T)$. This proves Observation 1.

Observation 2 If there is a minimal derivation δ of $T \vdash u$ then it only contains nodes of the form $T \vdash v$ with $v \in sub(T \cup \{u\})$.

We do induction on size of δ. If the last rule in δ is (M), (P) or (E) then the argument is as in the case without equations.

If the last rule used is (UL), (UR) or (D) then the result follows from Observation 1.

Suppose the last rule is (X) to obtain $u = (u_1 + \ldots + u_n) \downarrow$ from derivations δ_i of $T \vdash u_i$. It suffices to show that each $u_i \in sub(T \cup \{u\})$. Wlog the last rule in δ_i is not (X). If u_i is not headed with $+$ then $u_i \in sub(\{u\})$. Otherwise the last rule in δ_i is (M), (UL), (UR) or (D). In the first case, the result is easy and in the other cases we apply Observation 1.

Hence to check whether a message m can be obtained from a set T of messages by the above rules, we proceed as in the non-equational case. We keep on generating more and more messages of $sub(T \cup \{m\})$ which can be generated from the existing messages using one of the rules.

It remains to show how to check that a term t can be obtained from a set of terms $t_1 + \ldots + t_n$ using the rule (X).

Each t_i is of the form $c_{i,1}u_1 + \ldots + c_{i,k}u_k$ and t is of the form $c_1u_1 + \ldots + c_ku_k$, where u_i are pairwise distinct and are not headed with $+$, and $c_i \in \{0, 1\}$.

Hence we need to check whether there are some $x_1, \ldots, x_n \in \{0, 1\}$ such that

$$x_1 c_{1,1} \oplus \ldots \oplus x_n c_{n,1} = c_1$$
$$\ldots$$
$$x_1 c_{1,k} \oplus \ldots \oplus x_n c_{n,k} = c_k$$

where $\oplus$ is the xor operation on the set $\{0, 1\}$.

This can be checked in polynomial time, e.g. using Gaussian elimination.

Hence the intruder deduction problem can be solved in polynomial time in presence of the exclusive-or operation.

Secrecy analysis for protocols with bounded number of sessions

In case of bounded number of sessions we earlier obtained an NP algorithm for deciding non-secrecy by assuming that the size of messages remains bounded.

We now remove this restriction, and show that the problem is still in NP.

For simplicity consider again only pairing and symmetric encryption operators.

We need to consider intruder deduction problems over terms involving variables.

We follow a presentation due to H. Comon-Lundh.

We will deal with constraints of the form

$$T_1 \models u_1 \wedge \ldots \wedge T_n \models u_n$$

where T_i are sets of terms containing variables and u_i are terms containing variables.

A solution of such a constraint is a substitution σ which maps every variable occurring in the constraint to some ground term (containing no variables).

σ is a solution of the above constraint if for every i we have

$$T_i\sigma \vdash u_i\sigma$$

The symbol $\vdash$ is the one considered before for intruder deduction on ground terms. The symbol $\models$ is used in the constraints involving non-ground terms.

Given a certain number of sessions of a protocol, as in the case of bounded message size, we first guess a suitable interleaving of the steps of the protocol. This gives us a sequence of the form

$$r_1 \Rightarrow s_1, \ldots, r_n \Rightarrow s_n$$

where r_i, s_i are terms involving variables. Intuitively r_i is the message received in a certain step and s_i is the message sent in that step.

We need to check whether such a sequence of steps is feasible, for certain values of the variables in the protocol steps.

This amounts to deciding whether the intruder can construct at each step the required message which should be received by some agent. At each step, the sent message is added to the knowledge of the intruder.

We represent this problem as a constraint. Let T_0 be the initial knowledge and s the secret message. We have the following constraint

$$
\begin{aligned}
T_0 &\models r_1 \\
T_0, s_1 &\models r_2 \\
T_0, s_1, s_2 &\models r_3 \\
\dots &\models \\
T_0, s_1, \dots, s_{n-1} &\models r_n \\
T_0, s_1, \dots, s_n &\models s
\end{aligned}
$$

We have the following well-formedness assumptions on the constraint

$$C = T_1 \models u_1 \wedge \ldots \wedge T_n \models u_n$$

- The set $\{T_1, \ldots, T_n\}$ is totally ordered wrt the subset relation.
- For every $T \models u$ occurring in the constraint and every variable $x \in Var(T)$, the set

  $$T_x = min\{T' \mid T' \models v \in C, x \in Var(v)\}$$

  exists and $T_x \subsetneq T$.

(Var denotes the set of variables occurring in a term.)

These assumptions are true for the constraint constructed from a realistic protocol, because every variable is introduced first in a received message.

Constraint solving rules

1. $C \wedge T \vDash u \rightsquigarrow C$
2. $C \wedge T \vDash u \rightsquigarrow_\sigma C\sigma \wedge T\sigma \vDash u\sigma$ if $\sigma = mgu(t, u), t \in Sub(T), t \neq u, t, u$ not variables.
3. $C \wedge T \vDash u \rightsquigarrow_\sigma C\sigma \wedge T\sigma \vDash u\sigma$ if $\sigma = mgu(t_1, t_2), t_1, t_2 \in Sub(T), t_1 \neq t_2, t_1, t_2$ not variables
4. $C \wedge T \vDash \{u\}_v \rightsquigarrow C \wedge T \vDash u \wedge T \vDash v$
5. $C \wedge T \vDash \langle u, v\rangle \rightsquigarrow C \wedge T \vDash u \wedge T \vDash v$
6. $C \wedge T \vDash u \rightsquigarrow \perp$ if $T = \emptyset$ or $Var(T \cup \{u\}) = \emptyset$ and $T \nvDash u$.

$\perp$ denotes an unsatisfiable constraint.

The substitutions σ in $\rightsquigarrow_\sigma$ are to remember the assignments to variables used at various steps for solving the constraints. If the subscript σ is absent, it denotes the identity substitution.

Show: The rules transform a well-formed constraint into a well-formed constraint.

Correctness: If $C_1 \rightsquigarrow_\sigma C_2$ and θ is a solution of C_2 then $\sigma\theta$ is a solution of C_1.

Hence if the new constraint has some solution then the old constraint also has some solution.

Termination: The simplification rules terminate.

Define $\mid C \mid$ to be the sum of the sizes of the right hand sides occurring in the constraints.

We consider the pair $(\mid Vars(C) \mid, \mid C \mid)$ to be the measure of a clause, with lexicographic ordering.

Then the application of the rules make the measure strictly smaller.

In fact only polynomially long sequences of simplification steps are possible.

Finally it remains to show completeness of these set of rules. That is, every constraint which has a solution can be simplified using these rules.

As for the intruder deduction problem we define a notion of simple derivations.

Given sets $T_1 \subseteq \ldots \subseteq T_k$ which occur in the constraint, a derivation δ of $T_i \vdash u$ as left-minimal if for all $j \leq i$ such that $T_j \vdash u$ is derivable, the leaves of δ are of the form $T_i \vdash v$ with $v \in T_j$.

A derivation δ of $T_i \vdash u$ is simple if

1. No branch contains the same node twice.
2. All subproofs are left-minimal.
3. If the last rule applied is a composition, then all nodes of the proof are of the form $T_i \vdash v$ with $v \in Sub(T_i)$.
4. If the last rule applied is a decomposition then all nodes of the proof are of the form $T_i \vdash v$ with $v \in Sub(T_i \cup \{u\})$.

Step 1: If $T_i \vdash u$ has a derivation then it has a simple derivation.

We define T_i to the minimal unresolved left hand side in a constraint C if for all $T_j \subsetneq T_i$ such that $T_j \vDash u \in C$, u is a variable.

In this case we define $T_i' = T_i \cup \{x \mid T_j \vDash x \in C, T_i \subsetneq T_i\}$.

Step 2: Let σ be a solution of a constraint C and T_i a minimal unresolved left hand side of C. If $T_i\sigma \vdash u$ has a simple derivation whose last rule is a decomposition or an axiom, then there exists some $t \in Sub(t_i)$ which is not a variable, such that $u = t\sigma$.

Step 3: Let C be a constraint, σ a solution of C, and σ a minimal unresolved left hand side. If T_i does not contain two distinct unifiable terms, if $T_i\sigma \vdash u\sigma$ has a derivation, if $u \in Sub(T_i)$ and if u is not a variable, then $T_i' \vdash u$ has a derivation.

Define C to be in resolved form if all the right hand sides are variables and all left hand sides are non-empty.

Every resolved form has a solution: assign to every variable a term from the least left hand side (which must be ground).

Step 4: Let σ be a solution of C which is not in resolved form. Then for some θ, τ, C' we have $C \rightsquigarrow_\theta C', \sigma = \theta\tau$ and τ is a solution of C'.

Hence to detect an attack, it suffices to guess a sequence of simplification steps leading to a solved form.

Conclusion: Checking if a protocol with bounded number of sessions has an attack is in NP.