ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

S. Mödersheim
Ch. Sprenger
AS 2009

Formal Methods for Information Security

Exercise Sheet 5

Hand-in date: Nov 2, 2009

Assignment 5.1: Intruder knowledge in abstracted NSL

Consider NSL, the Needham-Schroeder public-key protocol with Lowe's fix (i.e. with second message $\{NA, NB, B\}_{\mathsf{pk}(A)}$). Let us work with the abstraction where Alice creates NA as $na(A, B)$ and Bob creates NB as $nb(B, A)$. The set of nonces in this model is

$$\mathtt{Nonce} = \{ni\} \cup \{na(A, B) \mid A, B \in \mathtt{Agent}\} \cup \{nb(B, A) \mid A, B \in \mathtt{Agent}\}$$

where ni is a nonce that the intruder initially knows.

We consider a typed model where variables of type agent and nonce can only be instantiated with values from the set of agents and nonces, respectively. Let the initial intruder knowledge be $IK_0 = \mathtt{Agent} \cup \{\mathsf{pk}, \mathsf{inv}(\mathsf{pk}(i))\}$, where $\mathtt{Agent} = \{a, b, i\}$ is the set of agents considered.

(a) Give the set of rules that describes all messages that the intruder can derive (in the style of module 6, slide 36). Note that only the nonces that an agent creates should be abstracted, while the nonces that the agent receives are not abstracted.

(b) List the entire intruder knowledge that the intruder can ever obtain—resticted to messages that some honest agent can obtain and submessages thereof. (In other words, ignore any uninteresting terms like $\langle i, \langle i, i \rangle \rangle$ that the intruder can construct.)

(c) Can we infer from this list that the nonces are secrets between the respective A and B?

Solution

(a) Here are the derivation rules for NSL:

$$\frac{}{\{na(A,B),A\}_{\mathsf{pk}(B)} \in \mathcal{DY}_{NSL}(M)}\ Step_1\ (A,B \in \texttt{Agent})$$

$$\frac{\{NA,A\}_{\mathsf{pk}(B)} \in \mathcal{DY}_{NSL}(M)}{\{NA,nb(B,A),B\}_{\mathsf{pk}(A)} \in \mathcal{DY}_{NSL}(M)}\ Step_2$$

$$\frac{\{na(A,B),NB,B\}_{\mathsf{pk}(A)} \in \mathcal{DY}_{NSL}(M)}{\{NB\}_{\mathsf{pk}(B)} \in \mathcal{DY}_{NSL}(M)}\ Step_3$$

$$\frac{t_1 \in \mathcal{DY}_{NSL}(M) \quad \ldots \quad t_n \in \mathcal{DY}_{NSL}(M)}{f(t_1,\ldots,t_n) \in \mathcal{DY}_{NSL}(M)}\ Compose\ (f \in \Sigma_p)$$

$$\frac{\langle m_1,m_2\rangle \in \mathcal{DY}_{NSL}(M)}{m_i \in \mathcal{DY}_{NSL}(M)}\ Proj_i\ (i \in \{1,2\})$$

$$\frac{\{m\}_k \in \mathcal{DY}_{NSL}(M) \quad \mathsf{inv}(k) \in \mathcal{DY}_{NSL}(M)}{m \in \mathcal{DY}_{NSL}(M)}\ Decrypt$$

(b) The initial intruder knowledge is $IK_0 = \texttt{Agent} \cup \{pk, \mathsf{inv}(\mathsf{pk}(i)), ni\}$, where $\texttt{Agent} = \{a,b,i\}$. We will now construct in a series of derivation steps the "interesting" subset $C \subseteq \mathcal{DY}_{NSL}(IK_0)$ from the initial knowledge IK_0 (where "interesting" means protocol messages and their derivable submessages). This yields a monotonically increasing series of intruder knowledges $IK_0 \subseteq IK_1 \subseteq \ldots \subseteq IK_n$ ending in the fixed point $IK_n = C$ (for $n = 10$). Each step i below mentions the new messages in $IK_i \setminus IK_{i-1}$ that are added to the intruder knowledge. We also introduce the abbreviation $Nonces_i = IK_i \cap \texttt{Nonce}$, since we are particularly interested in the nonces we can derive. Initially, we have $Nonces_0 = \{ni\}$, since $ni \in IK_0$

The intruder can then derive the following:

(1) By ($Step_1$): $\{na(A,B),A\}_{\mathsf{pk}(B)} \in IK_1$ for each $A,B \in \texttt{Agent}$.

(2) By ($Decrypt/Proj_1$): $na(A,i) \in IK_2$ for all $A \in \texttt{Agent}$. At this point we have $Nonces_2 = \{ni\} \cup \{na(A,i) \mid A \in \texttt{Agent}\}$.

(3) By ($Step_2$): $\{na(A,B),nb(B,A),B\}_{\mathsf{pk}(A)} \in IK_3$ for $A,B \in \texttt{Agent}$.

(4) By ($Decrypt/Proj_i$): $na(i,B) \in IK_4$ and $nb(B,i) \in IK_4$ for all $B \in \texttt{Agent}$. Now we have $Nonces_4 = \{ni\} \cup \{na(A,i), na(i,A), nb(A,i) \mid A \in \texttt{Agent}\}$.

(5) By ($Step_3$): $\{nb(B,A)\}_{\mathsf{pk}(B)} \in IK_5$ for all $A,B \in \texttt{Agent}$.

(6) By ($Decrypt/Proj_i$): $nb(i,A) \in IK_6$ for all $A \in \texttt{Agent}$. At this point we have

$$Nonces_6 = \{ni\} \cup \{na(A,i), na(i,A), nb(A,i), nb(i,A) \mid A \in \texttt{Agent}\}$$

(7) (Composition) of messages matching the premise of ($Step_2$): $\{N,A\}_{\mathsf{pk}(B)} \in IK_7$ for all $A,B \in \texttt{Agent}$ and $N \in Nonces_6$.

(8) By $(Step_2)$: $\{N, nb(B,A), B\}_{\mathsf{pk}(A)} \in IK_8$ for $A, B \in \texttt{Agent}$ and $N \in Nonces_6$.

(9) (Composition) of messages matching the premise of rule $Step_3$, i.e., for all $A, B \in \texttt{Agent}$ and all $N \in Nonces_6$, we have $\{na(A,B), N, B\}_{\mathsf{pk}(A)} \in IK_9$.

(10) By $(Step_3)$: $\{N\}_{\mathsf{pk}(B)} \in IK_{10}$ for $B \in \texttt{Agent}$ and $N \in Nonces_6$.

Here, we have reached the fixed point $C = IK_{10}$, since no new protocol messages can be constructed. Note in particular, that step (6) onwards decryption and projection would not produce any new nonces. Hence, we have $C \cap \texttt{Nonce} = Nonces_6$.

(c) We have in the fixedpoint C neither $na(A,B)$ nor $nb(B,A)$ such that both $A \neq i$ and $B \neq i$. Consider the secret-events that are generated in any trace:

- $signal(secret, A, B, na(A,B))$ where $A \neq i$ (because the intruder does not generate such events) — in this case either $B = i$ or the intruder does not know $na(A,B)$.
- Similarly for $signal(secret, B, A, nb(B,A))$.

Thus, no secrecy goal is violated on any trace.

Assignment 5.2: Authentication in abstracted NSL

Consider again the abstracted NSL protocol of the previous assignment and the following "attack" trace:

$$\begin{array}{ll}
a \to b: & \{na(a,b), a\}_{\mathsf{pk}(b)} \\
b \to a: & \{na(a,b), nb(b,a), b\}_{\mathsf{pk}(a)} \\
a \to b: & \{nb(b,a)\}_{\mathsf{pk}(a)} \\
i(a) \to b: & \{ni, a\}_{\mathsf{pk}(b)} \\
b \to i(a): & \{ni, nb(b,a), b\}_{\mathsf{pk}(a)} \\
i(a) \to b: & \{nb(b,a)\}_{\mathsf{pk}(a)}
\end{array}$$

(a) Which authentication/agreement goal is violated by this trace? (Insert the corresponding signals into the role descriptions.)

(b) Does this attack trace have a counter-part in the concrete protocol model (without the abstraction)? Explain.

(c) (**For experts:**) What can we do about it?

Solution

(a) The goal $B\ weakly\ authenticates\ A\ on\ NA$ (non-injective agreement).[1] The respective events in the trace are then:

$$\begin{array}{ll} a \to b: & \{na(a,b), a\}_{\mathsf{pk}(b)} \\ & signal(running_A(a, b, na(a, b))) \\ b \to a: & \{na(a,b), nb(b,a), b\}_{\mathsf{pk}(a)} \\ a \to b: & \{nb(b,a)\}_{\mathsf{pk}(a)} \\ & signal(commit_B(b, a, na(a, b))) \\ i(a) \to b: & \{ni, a\}_{\mathsf{pk}(b)} \\ b \to i(a): & \{ni, nb(b,a), b\}_{\mathsf{pk}(a)} \\ i(a) \to b: & \{nb(b,a)\}_{\mathsf{pk}(a)} \\ & signal(commit_B(b, a, \mathbf{ni})) \end{array}$$

This trace is an attack: the intruder convinces b that his own nonce ni came from a. The attack works since the "fresh nonce" of b is the same as in b's first protocol run and therefore the intruder can replay a's response from the first run.

Technically, the trace constitutes an attack, since the second $commit_B$ signal is not matched by a corresponding $running_A$ signal.

(b) This attack does not work in the concrete model because b will generate a fresh nonce in the second step and will therefore not accept the nonce from a previous session in step 3; the new nonce cannot be recovered by the intruder because he does not have $\mathsf{inv}(\mathsf{pk}(a))$.

Note that this proves only that this particular attack trace does not work in the original model, there may be other attacks on weak authentication. The proof that weak authentication holds is in fact done as sketched in the following.

(c) One may consider a finer abstraction, i.e. partition the set of all nonces into a larger number of equivalence classes. A simple way to do it is to leave $NA \mapsto na(A, B)$ as is, but to refine the abstraction of NB into $nb(B, A, NA)$ i.e. making the new nonce a function of the agent names and the nonce that B has received. The only problem is that this leads to an infinite set of equivalence classes and thus to termination problems. A finite alternative is to consider $NB \mapsto nb(B, A, bool)$ where $bool = 1$ if the incoming nonce that B received was $na(A, B)$ and 0 otherwise.

[1] Of course injective agreement is thus also violated, but as said in the lecture, the injective variant of agreement cannot hold in this model, anyway.