From CryptoVerif Specifications to Computationally Secure Implementations of Protocols

Bruno Blanchet
joint work with David Cadé

INRIA Paris
Bruno.Blanchet@inria.fr

Year 2023-24
# Protocol verification

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Symbolic</th>
<th>Computational</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDR, AVISPA, ProVerif, Scyther, Tamarin, ...</td>
<td>CryptoVerif, CertiCrypt, EasyCrypt, ...</td>
<td></td>
</tr>
<tr>
<td>Implementations</td>
<td>FS2PV, F7, Spi2Java, C2PV, ...</td>
<td>FS2CV, Computational F7, C2CV, ...</td>
</tr>
</tbody>
</table>
## Protocol verification

<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Computational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifications</td>
<td>FDR, AVISPA, ProVerif, Scyther, Tamarin, …</td>
<td>CryptoVerif, CertiCrypt, EasyCrypt, …</td>
</tr>
<tr>
<td>Implementations</td>
<td>FS2PV, F7, Spi2Java, C2PV, …</td>
<td>FS2CV, Computational F7, C2CV, <em>our work</em>, …</td>
</tr>
</tbody>
</table>
Our approach

Generate protocol implementations from specifications.

- Specification proved secure in the computational model by CryptoVerif.
- Specification translated into an OCaml implementation by our compiler.

Goal: proved implementations of cryptographic protocols.

Remark: FS2CV does the translation in the other direction!
Overview of our approach

- CryptoVerif specification
- Our Compiler
- Protocol Code
- Network Code
- OCaml Compiler
- Implementation
- Proof in the computational model

Caption: Tool Input Result
Choice of the target language

- Why OCaml?
  - Memory safe. Easier to show that the network code does not access the protocol memory.
  - Clean semantics.
  - Crypto library available.

- Writing a compiler into another language would not be difficult.
  Proving the security of the generated protocol may be more difficult.
The CryptoVerif specification language: terms

CryptoVerif represents protocols and games in a process calculus.

\[ M, N ::= \text{terms} \]
\[ x \quad \text{variable} \]
\[ f(M_1, \ldots, M_m) \quad \text{function application} \]

Function symbols \( f \) correspond to functions computable by polynomial-time deterministic Turing machines.
The CryptoVerif specification language: processes

\[ Q ::= \]

\[ 0 \quad \text{oracle definitions} \]

\[ Q \mid Q' \quad \text{nil} \]

\[ \text{foreach } i \leq n \text{ do } Q \quad \text{parallel composition} \]

\[ O(x_1 : T_1, \ldots, x_k : T_k) := P \quad \text{replication } n \text{ times} \]

\[ P ::= \]

\[ \text{return}(M_1, \ldots, M_k); Q \quad \text{oracle declaration} \]

\[ \text{yield} \quad \text{return} \]

\[ x \leftarrow T; P \quad \text{end} \]

\[ x : T = M \text{ in } P \quad \text{random number} \]

\[ \text{if } M \text{ then } P \text{ else } P' \quad \text{assignment} \]

\[ \text{insert } Tbl(M_1, \ldots, M_k); P \quad \text{conditional} \]

\[ \text{get } Tbl(x_1 : T_1, \ldots, x_k : T_k) \text{ suchthat } M \text{ in } P \text{ else } P' \quad \text{insert in table} \]

\[ \text{get from table} \]
A $\rightarrow$ B : enc($r$, \(K_{ab}\))

process \(O_{\text{start}}() := \quad K_{ab} \xleftarrow{R} \text{key}; \quad \text{return}(); \)

( foreach \(i_1 \leq N\) do run \(\text{processA}(K_{ab})\) |
foreach \(i_2 \leq N\) do run \(\text{processB}(K_{ab})\) )

- The process generates \(K_{ab}\).
- This symmetric key will not be known by the opponent.
- Only after the key has been generated, we can call at most \(N\) times \(\text{processA}\) and at most \(N\) times \(\text{processB}\).
Example

\[ A \rightarrow B : \text{enc}(r, Kab) \]

\begin{align*}
\text{let } & \text{processA}(Kab) = \\
& O_A() := r \xleftarrow{R} \text{nonce}; s \xleftarrow{R} \text{seed}; \\
& \text{return(\text{enc(\text{nonceToBitstring}(r), Kab, s))}.}
\end{align*}

\begin{align*}
\text{let } & \text{processB}(Kab) = \\
& O_B(m : \text{bitstring}) := \\
& \text{let injbot(\text{nonceToBitstring}(r' : \text{nonce})) = \text{dec}(m, Kab) in} \\
& \text{return()}. \\
\end{align*}

- \textit{processA} sends the encryption of \( r \) under \( Kab \) (probabilistic encryption)
- \textit{processB} decrypts the received message
Example — summary

\[
\text{let } \text{processA}(Kab) = \\
O_A() := r \xleftarrow{R} \text{nonce}; s \xleftarrow{R} \text{seed}; \\
\text{return}(\text{enc(\text{nonceToBitstring}(r), Kab, s)}).
\]

\[
\text{let } \text{processB}(Kab) = \\
O_B(m : \text{bitstring}) := \\
\text{let injbot(\text{nonceToBitstring}(r' : \text{nonce})) = dec(m, Kab) in} \\
\text{return()}. \\
\]

\[
\text{process } O_{\text{start}}() := Kab \xleftarrow{R} \text{key}; \text{return();} \\
(\text{foreach } i1 \leq N \text{ do run processA}(Kab) | \\
\text{foreach } i2 \leq N \text{ do run processB}(Kab))
\]
Annotations: Separation in multiple programs

let \textit{processA}(Kab) =
\[ \textit{pA}\{(O_A() := r \overset{R}{\leftarrow} \text{nonce}; s \overset{R}{\leftarrow} \text{seed}; \}
\text{return(\text{enc(nonceToBitstring(r), Kab, s))})\}. \]

let \textit{processB}(Kab) =
\[ \textit{pB}\{(O_B(m : \text{bitstring}) := \}
\text{let injbot(nonceToBitstring(r' : \text{nonce})) = \text{dec}(m, Kab) in \}
\text{return()}\}. \]

process \textit{keygen} [Kab > fileKab] \{O_{\text{start}}() := Kab \overset{R}{\leftarrow} \text{key}; \text{return()}\};
(\text{foreach } i_1 \leq N \text{ do run processA(Kab) |}
\text{foreach } i_2 \leq N \text{ do run processB(Kab))}
let \( \text{processA}(Kab) = \)

\[
\begin{align*}
    & \text{pA}\{ O_A() := r \xleftarrow{R} \text{nonce}; s \xleftarrow{R} \text{seed}; \\
    & \text{return}(\text{enc(\text{nonceToBitstring}(r), Kab, s)})\}
\end{align*}
\]

let \( \text{processB}(Kab) = \)

\[
\begin{align*}
    & \text{pB}\{ O_B(m : \text{bitstring}) := \\
    & \text{let injbot(\text{nonceToBitstring}(r' : \text{nonce})) = \text{dec}(m, Kab) in} \\
    & \text{return()})\}
\end{align*}
\]

process \( \text{keygen} [Kab > fileKab] \{ O_{\text{start}}() := Kab \xleftarrow{R} \text{key}; \text{return()}\}; \)

( foreach \( i1 \leq N \) do run \( \text{processA}(Kab) \) |

foreach \( i2 \leq N \) do run \( \text{processB}(Kab) \))
Annotations: types and functions

- OCaml type representing a CryptoVerif type:
  ```ocaml```
  ```
  implementation type keyseed = 128.  (bitstring of 128 bits)
  implementation type host = "string" [serial = "id", "id"].
  ```
- OCaml function representing a function in the protocol specification:
  ```ocaml```
  ```
  implementation fun enc = "sym_enc".
  implementation fun injbot = "injbot" [inverse = "injbot_inv"].
  ```
  - In the CryptoVerif specification, there are assumptions about these functions.
    - Functional assumptions: \( \text{dec(enc}(m, k, s), k) = \text{injbot}(m) \).
    - Security assumptions: encryption is IND-CPA and INT-CTXT.
  - These assumptions must be manually verified.
get / insert handle tables of keys:

- insert \( \text{keytbl}(h, k) \)
  inserts element \( h, k \) in the table \( \text{keytbl} \).
- get \( \text{keytbl}(h', k') \) such that \( h' = h \) in \( P \) else \( P' \)
  stores in \( h', k' \) an element of table \( \text{keytbl} \) such that \( h' = h \),
  i.e., stores in \( k' \) the key of \( h \), and runs \( P \).
  Runs \( P' \) when no such element exists.

- Tables are stored in files:
  implementation table \( \text{keytbl} = "\text{filekeytbl}" \).
Treatment of tables in CryptoVerif

For proving the protocol, CryptoVerif encodes tables as arrays:

- The variables are considered as arrays with one cell for each copy of the definition.
  - Useful for remembering all values taken by the variable.
- \( \text{foreach } i \leq n \text{ do } \ldots \text{insert } keytbl(h, k) \)
  becomes
  \( \text{foreach } i \leq n \text{ do } \ldots \text{let } keytbl_1[i] = h \text{ in let } keytbl_2[i] = k \text{ in} \)
- \( \text{get } keytbl(h', k') \text{ suchthat } h' = h \text{ in } P \text{ else } P' \)
  becomes
  \( \text{find } u \leq n \text{ suchthat defined}(keytbl_1[u], keytbl_2[u]) \land keytbl_1[u] = h \text{ then let } h' = keytbl_1[u] \text{ in let } k' = keytbl_2[u] \text{ in } P \text{ else } P' \)
For proving the protocol, CryptoVerif encodes tables as arrays:

- The variables are considered as arrays with one cell for each copy of the definition.
  - Useful for remembering all values taken by the variable.
- \[\text{foreach } i \leq n \text{ do } \ldots \text{ insert } \text{keytbl}(h, k)\]
  becomes
  \[\text{foreach } i \leq n \text{ do } \ldots \text{ let } \text{keytbl}_1[i] = h \text{ in let } \text{keytbl}_2[i] = k \text{ in}\]
- \[\text{get } \text{keytbl}(h', k') \text{ suchthat } h' = h \text{ in } P \text{ else } P'\]
  becomes
  \[\text{find } u \leq n \text{ suchthat defined}(\text{keytbl}_1[u], \text{keytbl}_2[u]) \land \text{keytbl}_1[u] = h \text{ then let } h' = \text{keytbl}_1[u] \text{ in let } k' = \text{keytbl}_2[u] \text{ in } P \text{ else } P'\]
- Generalized to several insertions by looking up in the variables defined at each insertion.
Treatment of tables in CryptoVerif

For proving the protocol, CryptoVerif encodes tables as arrays:

- The variables are considered as arrays with one cell for each copy of the definition.
  - Useful for remembering all values taken by the variable.
- \texttt{foreach } \texttt{i \leq n do ... insert keytbl(h, k)}
  
  becomes
  \texttt{foreach } \texttt{i \leq n do ... let keytbl}_1[\texttt{i}] = h \texttt{ in let keytbl}_2[\texttt{i}] = k \texttt{ in}

- \texttt{get keytbl(h', k')} suchthat \texttt{h'} = h \texttt{ in P else P'}
  
  becomes
  \texttt{find } u \leq n \texttt{ suchthat defined(keytbl}_1[u], keytbl}_2[u]) \land keytbl}_1[u] = h \texttt{ then let h'} = keytbl}_1[u] \texttt{ in let k'} = keytbl}_2[u] \texttt{ in P else P'}

- Generalized to several insertions by looking up in the variables defined at each insertion.

Avoiding arrays is more intuitive and simplifies the compilation.
Compilation to OCaml

For each program, the compiler generates an OCaml module where it defines a function for each oracle.

- A function $\text{init} : \text{unit} \to \tau$ returns the tuple of functions representing the oracles available at the beginning of the program.
  - $\text{init}$ may also read variables from files when needed.
- Each oracle declaration $Q$ is represented by a function that
  - takes as argument the arguments of the oracle $Q$
  - and returns
    - the tuple of functions representing oracles that follow $Q$,
    - the result returned at the end of $Q$. 
Compilation to OCaml: example

let \( \text{processA}(Kab) = pA\{ O_A() := r \overset{R}{\leftarrow} \text{nonce}; s \overset{R}{\leftarrow} \text{seed}; \)
\[
\text{return}(\text{enc}(\text{nonceToBitstring}(r, Kab, s))) \}\.

The generated module \( PA \) has the following interface:

```ocaml
open Base
open Crypto

type type_processA = unit -> (unit * string)
val init : unit -> type_processA
```
Compilation to OCaml: replication

- When a process is **under replication**, it is compiled into an ordinary function:
  
  ```ocaml
  fun [args] -> [body]
  ```

- When a process is **not under replication**, it is compiled into a function that can be called only once:
  
  ```ocaml
  let token = ref true in
  fun [args] ->
    if (!token) then
      begin
        token := false;
        [body]
      end
    else raise Bad_call
  ```
Compilation to OCaml: terms and body (1)

<table>
<thead>
<tr>
<th>CryptoVerif</th>
<th>OCaml</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>$[M]$</td>
</tr>
<tr>
<td>$x$</td>
<td>$[x]$</td>
</tr>
<tr>
<td>$f(M_1, \ldots, M_n)$</td>
<td>$[f] [M_1] \ldots [M_n]$</td>
</tr>
<tr>
<td>$P$</td>
<td>$[P]$</td>
</tr>
<tr>
<td>$x \leftarrow^R T; P$</td>
<td>let $[x] = \text{rand}_T()$ in $[P]$</td>
</tr>
<tr>
<td>let $x = M$ in $P$</td>
<td>let $[x] = [M]$ in $[P]$</td>
</tr>
<tr>
<td>if $M$ then $P$ else $P'$</td>
<td>if $[M]$ then $[P]$ else $[P']$</td>
</tr>
<tr>
<td>yield</td>
<td>raise Match.fail</td>
</tr>
<tr>
<td>return($M$); $Q$</td>
<td>($[Q], [M]$)</td>
</tr>
</tbody>
</table>

When a variable needs to be written to a file, it is written just after its definition.
Compilation to OCaml: terms and body (2)

\[ \text{insert } Tbl(M_1, \ldots, M_n); P \]
compiled into
\[ \text{insert\_in\_table } [Tbl] [[serial_{T_1}] [M_1]; \ldots; [serial_{T_n}] [M_n]]; [P] \]

\[ \text{get } Tbl(x_1 : T_1, \ldots, x_n : T_n) \text{ such that } M \text{ in } P \text{ else } P' \]
compiled into
\[
\begin{align*}
\text{let } l &= \text{get\_from\_table } [Tbl] \\
&\quad \text{(function } [[x_1]']; \ldots; [[x_n]'] \text{ ->} \\
&\quad \qquad \text{let } [[x_1]] = \text{exc\_bad\_file } [Tbl] \text{ (} [[\text{deserial}_{T_1}] [x_1]'] \text{ in } \\
&\quad \qquad \qquad \text{let } [[x_n]] = \text{exc\_bad\_file } [Tbl] \text{ (} [[\text{deserial}_{T_n}] [x_n]'] \text{ in } \\
&\quad \qquad \qquad \text{if } [[M]] \text{ then } ([[x_1]], \ldots, [[x_n]]) \text{ else raise Match\_fail} \\
&\quad \qquad \text{| } _{->} \text{ raise (Bad\_file } [Tbl]])
\end{align*}
\]

in
\[ \text{if } l = [] \text{ then } [P'] \text{ else} \]
\[ \text{let } ([x_1], \ldots, [x_n]) = \text{rand\_list } l \text{ in } [P] \]
Assumptions on the network code:
- No unsafe OCaml functions (such as Obj.magic).
- No mutation of values received from or passed to generated functions.
- No fork after obtaining and before calling a process that can be called only once.
Assumptions

- **Assumptions on the network code:**
  - No unsafe OCaml functions (such as Obj.magic).
  - No mutation of values received from or passed to generated functions.
  - No fork after obtaining and before calling a process that can be called only once.

- **Assumptions on program execution:**
  - Programs are executed in the order specified in the CryptoVerif process.
  - Several programs that insert data in the same table are not run concurrently.
Assumptions

- Assumptions on the network code:
  - No unsafe OCaml functions (such as `Obj.magic`).
  - No mutation of values received from or passed to generated functions.
  - No fork after obtaining and before calling a process that can be called only once.

- Assumptions on program execution:
  - Programs are executed in the order specified in the CryptoVerif process.
  - Several programs that insert data in the same table are not run concurrently.

- Other:
  - Types that represent CryptoVerif data are not recursive.
  - The files used by generated code are not read/written by other code.
Application: SSH

- Secure SHell: an important protocol

SSH Transport Layer

- Key exchange

  enc&MAC tunnel

Authentication of the client

Connection various applications

SSH v. 2.0
SSH Transport Layer Protocol: key exchange

Client $C$

| $id_C = \text{SSH-2.0-version}_C$ |
| $id_S = \text{SSH-2.0-version}_S$ |
| $\text{KEXINIT}, \text{cookie}_C, \text{algos}_C$ |
| $\text{KEXINIT}, \text{cookie}_S, \text{algos}_S$ |

$\begin{align*}
x &\xleftarrow{\text{R}} [2, q - 1], e = g^x \\
K &= f^x \quad \text{(KEYDH_REPLY, pk}_S, f, \text{sign}(H, sk_S))
\end{align*}$

Server $S$

$\begin{align*}
y &\xleftarrow{\text{R}} [1, q - 1], f = g^y \\
K &= e^y
\end{align*}$

$pk_S, \text{sign}(H, sk_S)$ ok?

$\begin{align*}
\text{NEWKEYS} &\rightarrow \text{NEWKEYS}
\end{align*}$

$\text{algos} = \text{diffie-hellman-group14-sha1, ssh-rsa, aes128-cbc, hmac-sha1}$

$H = \text{SHA1}(id_C, id_S, \text{cookie}_C, \text{algos}_C, \text{cookie}_S, \text{algos}_S, pk_S, e, f, K)$
SSH Transport Layer Protocol: packet protocol

\[
\text{sessionid} = H
\]
\[
IV_C = \text{SHA1}(K, H, "A'', \text{sessionid})
\]
\[
IV_S = \text{SHA1}(K, H, "B'', \text{sessionid})
\]
\[
K_{\text{enc,C}} = \text{SHA1}(K, H, "C'', \text{sessionid})
\]
\[
K_{\text{enc,S}} = \text{SHA1}(K, H, "D'', \text{sessionid})
\]
\[
K_{\text{MAC,C}} = \text{SHA1}(K, H, "E'', \text{sessionid})
\]
\[
K_{\text{MAC,S}} = \text{SHA1}(K, H, "F'', \text{sessionid})
\]

\[
\text{packet} = \text{packet}\_\text{length}||\text{padding}\_\text{length}||\text{payload}||\text{padding}
\]

Client \( \text{C} \quad \overset{\text{enc}(K_{\text{enc,C}}, \text{packet}, IV_C), \text{MAC}(K_{\text{MAC,C}}, \text{sequence}\_\text{number}_C||\text{packet})}{\longrightarrow} \quad \text{Server} \quad \text{S} \)

\[
\text{enc}(K_{\text{enc,S}}, \text{packet}, IV_S), \text{MAC}(K_{\text{MAC,S}}, \text{sequence}\_\text{number}_S||\text{packet})
\]
CryptoVerif proof

- Modeled the **SSH Transport Layer Protocol** in CryptoVerif.
- Proved the **authentication of the server** to the client
  - Automatic by CryptoVerif
- The **authentication of the client** to the server requires the authentication protocol.
- **Secrecy of the key** requires extensions of CryptoVerif.
- **Secrecy of messages** sent over the tunnel cannot be proved:
  - Length of the packet leaked,
  - CBC mode with chained IVs.
Generated implementation

- Manually written **cryptographic primitives**.
  - based on CryptoKit.
- Manually written **network code**:
  - Key generators,
  - Client,
  - Server.

They call the code generated from the CryptoVerif model.

- Format respected at the bit level.
  - Interact with other SSH implementations (OpenSSH).
- Some features omitted:
  - Key re-exchange
  - IGNORE, DISCONNECT messages
Demo

- ssh.ocv
- Prove by CryptoVerif
- Compile: key generation, client, server
- Run
Conclusion

- CryptoVerif specifications
  - proved secure in the computational model by CryptoVerif,
  - translated into OCaml implementations.

- Our approach favors the methodology:
  1. Write a formal specification;
  2. Prove it;
  3. Then, build an implementation.

- We proved the soundness of the compiler.
  - specification secure $\Rightarrow$ implementation secure

- Future work: extend the specification language, with loops, mutable variables, . . .
  - extensions of CryptoVerif and of the compiler