Formal methods for software security

Thomas Jensen, INRIA

Forum "Méthodes formelles"
Toulouse, 31 January 2017
Formal methods for software security
Formal methods for software security

- Certification
- Availability
- Cryptographic protocols
- Security types
- Reference monitors
- Integrity
- Refinement
- Computational model
- Side channels
- Security types
- Enforcement
- Secure OS
- SEL4
- Coq
- Confidentiality
- Information flow
- Non-interference
- B
- Secure programming
- Reference monitors
- Enforcement
Formal methods for software security

1. Basic concepts
2. Cryptographic protocols
3. Secure OS
4. Certification of software
5. Information flow
Basic security concepts

Confidentiality

• my secrets will not be disclosed … at least not more than I'm willing to accept.

Integrity

• my data and decisions are not influenced by intruders.

Availability

• software and services are there when I need them.
Basic security concepts

Confidentiality

• my secrets will not be disclosed … at least not more than I'm willing to accept.

Integrity

• my data and decisions are not influenced by intruders.

Availability

• software and services are there when I need them.

Security ≠ Safety
Basic security concepts

Confidentiality

• my secrets will not be disclosed … at least not more than I'm willing to accept.

Integrity

• my data and decisions are not influenced by intruders.

Availability

• software and services are there when I need them.

Security ≠ Safety

… but they are strongly related
Attacker model

Security is open-ended!

The question

Is my software secure?

must be complemented by an **attacker model**, stating the threats we are up against.

Specify the attackers

- observational power (output, network messages, time, …)
- actions (code insertion, message injection, …)
- access to machine (physical, through network, …)
Enforcement mechanisms

Certification of applications
  • Common Criteria
  • Formal methods for reaching upper levels.

Security-enhancing software development
  • secure programming guidelines
  • secure compilation.

Static code analysis
  • eg, Java's byte code verifier, information flow analysis.

Reference monitors and run-time analysis
Cryptographic protocols
Models of cryptographic protocols

Symbolic models

- specified as a series of exchanges of messages
- assuming perfect cryptography

Example: two agents A, B

1. \( A \leftrightarrow B : \{N_A, A\}_{K_B} \)
2. \( B \leftrightarrow A : \{N_A, N_B, B\}_{K_A} \)
3. \( A \leftrightarrow B : \{N_B\}_{K_B} \)

Attackers may

- intercept and re-send messages
- encrypt and decrypt messages (with available keys)
Verification

Model

- state = current message + state of A, B, and attacker
- rewriting rules defining protocol and attacker

\[ \{ (msg)_{\text{key}}, \text{key} \} \rightarrow \{ \text{msg}, (msg)_{\text{key}}, \text{key} \} \]

Security properties

- secrecy ("no state where attacker has the secret")
- authentification, re-play, ...
- specific properties ("key may not be used on stored content", "vote has been counted")
Tools

A variety of mature tools

- AVISPA, Tamarin, ProVerif, Timbuk, …

based on

- term and multi-set rewriting, Horn clauses, …

Interfaces for writing and animating protocols

- eg as Message Sequence Charts (SPAN).
Computational models

A model closer to reality:

- Messages: bit strings,
- Crypto primitives: functions on bit strings,
- Attacker: any probabilistic poly-time Turing machine.

Properties proved for all traces except for a set of traces of negligible probability.

Secrecy: attacker can distinguish secret from random number with only infinitesimal probability.

Proofs by refinement of models.

See eg. the cryptooverif tool
Implementations of crypto protocols
Implementations of crypto protocols

Security concerns with implementations of protocols and basic operations of cryptography.
Implementations of crypto protocols

Security concerns with implementations of protocols and basic operations of cryptography.

Implementations of cryptographic primitives are prone to side channel attacks:

- leaking secrets via timing or energy consumption,
- a challenge for implementors
Implementations of crypto protocols

Security concerns with implementations of protocols and basic operations of cryptography.

Implementations of cryptographic primitives are prone to side channel attacks:

• leaking secrets via timing or energy consumption,
• a challenge for implementors

Implementations of entire protocols are prone to programming errors:

• see the Verified TLS project for building a formally verified implementation of TLS.
Secure operating systems
Security and OS
Organized Sharing of resources between processes

- using the same memory
- communicating via IPC

and still guarantee isolation properties.
Security and OS

Organized Sharing of resources between processes

- using the same memory
- communicating via IPC

and still guarantee **isolation properties**.

![Diagram showing processes and OS layers]
Security and OS

Organized Sharing of resources between processes

- using the same memory
- communicating via IPC

and still guarantee isolation properties.

Large, complex software - long history of security alerts.
The SEL4 project

Project run at NICTA 2004-2014.

Formal verification of Liedtke's L4 micro-kernel.

- small code base (9 K Loc),
- threads, memory management, IPC, interrupts, capability-based AC,
- running on ARM,
- verified using the Isabelle/HOL theorem prover.

Prove:

- Functional correctness (and a lot of safety properties)
- Non-interference
SEL4: proof structure

Proof by refinement

Abstract model → HOL

Executable model → Haskell

C implementation → C

Binary kernel → HOL4 binary spec
SEL4: proof structure

Proof by refinement

Abstract model  ←  HOL
   □

Executable model  ←  Haskell
   □

C implementation  ←  C
   □

Binary kernel  ←  HOL4 binary spec

On the "Abstract model", build
  • access control model,
  • integrity and confidentiality proof
SEL4: proof structure

Proof by refinement

Abstract model \rightarrow HOL

Executable model \rightarrow Haskell

C implementation \rightarrow C

Binary kernel \rightarrow HOL4 binary spec

On the "Abstract model", build

- access control model,
- integrity and confidentiality proof

200 000 lines of Isabelle/HOL proof

25 person-years
Prove & Run's ProvenCore

SEL4 uses Isabelle/HOL and Haskell

- higher-order logic and lazy functional programming is still not main-stream development tools.

Prove & Run has developed a formally verified microkernel ProvenCore

- refinement proof method
- isolation properties.

using their SMART development framework:

- functional, executable specification
- closer to programmer's intuition
- equipped with a dedicated prover
Certification of Java Card applications
Java Card certification
Java Card certification

Java Card

- reduced dialect of Java for bank cards and SIM,
- no dynamic loading, reflection, floating points, threads,…
- "resource-constrained" programming practice.
Java Card certification

Java Card

- reduced dialect of Java for bank cards and SIM,
- no dynamic loading, reflection, floating points, threads,…
- "resource-constrained" programming practice.

Industrial context:

- Applications developed by third-parties and put on an app store.
- Must be certified according to industry norms (eg, AFSCM* norms for NFC applications).
- Need "light-weight" certification techniques.

*Association Française du Sans Contact Mobile
AFSCM norms/guidelines

Enforce good programming practice and resource usage

- catch exceptions, call methods with valid args,
- no recursion and almost no dynamic allocation,
- don't call method xxx.

Avoid exceptions due to

- null pointers, array indexing, class casts,
- illegal applet interaction through the firewall.
The Java Card analyser
The Java Card analyser

A combination of numeric and points-to analysis

- tailored to the application domain,
- take advantage of imposed restrictions,
- precise (flow-sensitive, inter-proc, trace partitioning).
The Java Card analyser

A combination of numeric and points-to analysis

- tailored to the application domain,
- take advantage of imposed restrictions,
- precise (flow-sensitive, inter-proc, trace partitioning).

Major challenge: modelling the Java Card API.
The Java Card analyser

A combination of numeric and points-to analysis

• tailored to the application domain,
• take advantage of imposed restrictions,
• precise (flow-sensitive, inter-proc, trace partitioning).

Major challenge: modelling the Java Card API.

Outcome: an abstract model of execution states

• mined by queries formalising the AFSCM norms.
The Java Card analyser

A combination of numeric and points-to analysis

- tailored to the application domain,
- take advantage of imposed restrictions,
- precise (flow-sensitive, inter-proc, trace partitioning).

Major challenge: modelling the Java Card API.

Outcome: an abstract model of execution states

- mined by queries formalising the AFSCM norms.
Information flow analysis
Back to confidentiality

Classify data as either

- private/secret/confidential
- public

A basic security policy:

"Confidential data should not become public"
Breaking confidentiality

```c
int secret s;    // s ∈ \{0,1\}
int public p;
```
Breaking confidentiality

```c
int secret s;    // s ∈ \{0,1\}
int public p;

p := s;          Direct flow```

\( \)
Breaking confidentiality

```plaintext
int secret s; // s ∈ {0,1}
int public p;

p := s;            Direct flow

if s == 1 then
    p := 1
else
    p := 0
```

Indirect flow
Non-interference

Confidentiality can be formalised as non-interference:

Changes in secret values should not be publicly observable
Non-interference

Confidentiality can be formalised as non-interference:

Changes in secret values should not be publicly observable
Non-interference

Confidentiality can be formalised as **non-interference**:

**Changes in secret values should not be publicly observable**
Non-interference

Confidentiality can be formalised as non-interference:

Changes in secret values should not be publicly observable

∀s₁, s₂, s₁′, s₂′, s₁ ~ s₂ ∧ (P, s₁) ↓ s₁′ ∧ (P, s₂) ↓ s₂′ → s₁′ ~ s₂′
Dynamic enforcement

Add a security level ("taint") to all data and variables

Security levels evolve due to assignments
Dynamic enforcement

Add a security level ("taint") to all data and variables

Security levels evolve due to assignments

\[ p := s; \quad \text{// direct flow} \]
Dynamic enforcement

Add a security level ("taint") to all data and variables

Security levels evolve due to assignments

\[ p := s; \quad // \text{direct flow} \]
Dynamic enforcement

Add a security level ("taint") to all data and variables

Security levels evolve due to assignments

\[ p := s; \quad \text{\// direct flow} \]

and when we assign under secret control:

\[
\text{if } s == 1 \text{ then} \\
p := 1
\]
Dynamic enforcement

Add a security level ("taint") to all data and variables

Security levels evolve due to assignments

\[ p := s; \quad \text{// direct flow} \]

and when we assign under secret control:

\[
\text{if } s == 1 \text{ then } \\
p := 1
\]
Secure?

Not enough to enforce confidentiality!

```c
int secret s;  // s ∈ {0,1}
int public p, q;

p := 0; q := 1;
if s == 0 then
    q := 0;
if q == 1 then
    p := 1;
```
Secure?

Not enough to enforce confidentiality!

```c
int secret s;   // s ∈ \{0,1\}
int public p,q;

s=0
p := 0; q := 1;
if s == 0 then
    q := 0;
if q == 1 then
    p := 1;
skip
p=0
```

Secure?

Not enough to enforce confidentiality!

```
int secret s; // s ∈ {0,1}
int public p,q;

p := 0; q := 1;
if s == 0 then
    q := 0;
if q == 1 then
    p := 1;
```

<table>
<thead>
<tr>
<th>s</th>
<th>p</th>
<th>q</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

```
Secure?

Not enough to enforce confidentiality!

\[
\begin{align*}
\text{int } & \text{ secret } s; \quad \text{// } s \in \{0,1\} \\
\text{int } & \text{ public } p, q;
\end{align*}
\]

\[
\begin{align*}
p & := 0; \quad q := 1; & \quad s=0 & \quad s=1 \\
\text{if } s & == 0 \text{ then} & \\
& \quad q := 0; & \quad p=0, q=1 & \quad p=0, q=1 \\
& \quad \text{skip} & \quad p=0, q=0 & \quad \text{skip} \\
& \quad \text{skip} & \quad p=1, q=1 & \quad p=1 \\
p & := 1; & \quad p=0 & \quad p=1
\end{align*}
\]

Need the "no-sensitive-upgrade" principle
Static information flow control

Information flow types:

\[ T,T_x,T_{pc} \in \{\text{public} \sqsubseteq \text{secret}\} \]

Typing rules:

\[ \vdash e : T \quad T \sqsubseteq T_x \quad T_{pc} \sqsubseteq T_x \quad T_{pc} \vdash x := e \]

\[ \vdash e : T \quad T_{pc} \sqcup T \vdash S_i \quad i = 1,2 \quad T_{pc} \vdash \text{if } e \text{ then } S_1 \text{ else } S_2 \]

Well-typed programs are non-interferent
Declassification and side channels

How to declassify confidential data:

• what and when to declassify?
• how much to declassify (passwd, statistics) ?

Information leaks due to other channels

• timing
• energy consumption

Challenge: analysis tools to check constant-time properties of (well-crafted) cryptographic computations.
Coda
Many more topics

Malware detection
  • analysis of (obfuscated) binaries.

Access control
  • formal models and enforcement.

Attack trees.

Web security
  • secure web programming with JavaScript et al.

Privacy
  • differential privacy (theory vs. practice),
  • software in coherence with legislation (EU GDPR).
Formal methods for software security

- Formal methods can improve the security of software.
- Come with solid foundations and mature tools.
- More and more industrial applications.
- Technology is becoming main-stream.
Formal methods for software security

- Formal methods can improve the security of software.
- Come with solid foundations and mature tools.
- More and more industrial applications.
- Technology is becoming main-stream.

Thank you