

Confidentiality analyses correct wrt. computational semantics

Peeter Laud

peeter.l@ut.ee

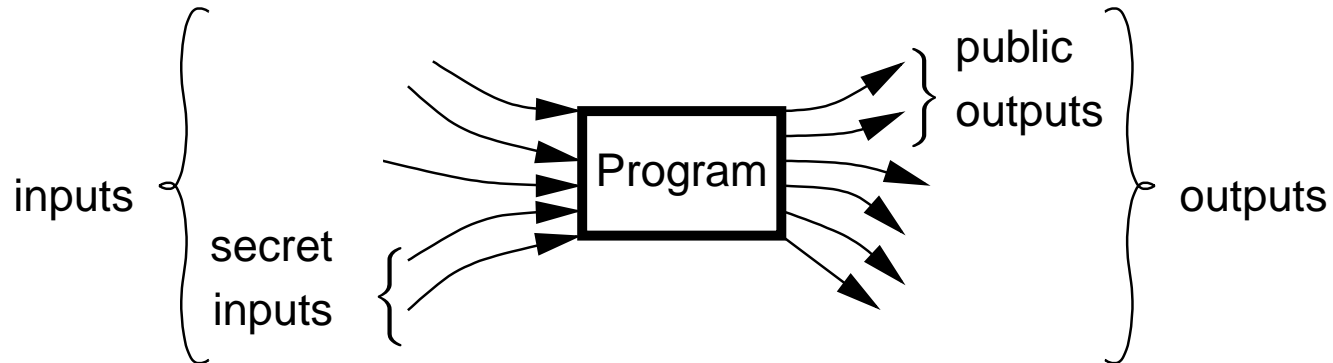
Tartu Ülikool

Cybernetica AS

Overview

- Computationally secure information flow.
- A program analysis, correct wrt. above.
- Confidentiality in cryptographic protocols.
- A very simple analysis.
- Using the def. of secure encryption.

Problem statement



Inputs come from a known source, i.e. the distribution of inputs is known.

- Public outputs should be independent of secret inputs.
- We want tools checking that.
- The input to these tools is the program text.
 - ... possibly also the description of input distribution.

Programming language — syntax

The WHILE-language (simple imperative language).

$$\begin{array}{l} P ::= x := o(x_1, \dots, x_k) \\ | \textit{skip} \\ | P_1; P_2 \\ | \textit{if } b \textit{ then } P_1 \textit{ else } P_2 \\ | \textit{while } b \textit{ do } P' \end{array}$$

$b, x, x_1, \dots, x_k \in \mathbf{Var.}$ $o \in \mathbf{Op.}$ $\textit{enc}, \textit{gen} \in \mathbf{Op.}$

Programming language — semantics

Denotational semantics: $\llbracket P \rrbracket : \text{State} \rightarrow \text{State}_\perp$.

$\text{State} = \text{Var} \rightarrow \text{Val}$.

State_\perp has an extra element \perp , denoting nontermination.

For each $o \in \text{Op}$ with arity k , a function $\llbracket o \rrbracket : \text{Val}^k \rightarrow \text{Val}$ is given.

Semantics is defined inductively over program structure.

This is the traditional setup...

Cryptographic considerations

- Security definitions in theoretical cryptography require
 - primitives with probabilistic functionality;
 - the **security parameter**.
- Also, all values are bit-strings.

Therefore:

- $\llbracket P \rrbracket = \{\llbracket P \rrbracket_n\}_{n \in \mathbb{N}}$;
- $\llbracket P \rrbracket_n : \mathbf{State}_n \rightarrow \mathcal{D}(\mathbf{State}_{n \perp})$;
- $\mathbf{State}_n = \mathbf{Var} \rightarrow \mathbf{Val}_n$;
- $\mathbf{Val}_n = \{0, 1\}^*$.

Also, $\llbracket o \rrbracket = \{\llbracket o \rrbracket_n\}_{n \in \mathbb{N}}$, $\llbracket o \rrbracket_n : \mathbf{Val}_n^k \rightarrow \mathcal{D}(\mathbf{Val}_n)$.

Computationally secure information flow

A program has CSIF, if its public outputs are computationally independent from its secret inputs.

- Secret inputs — initial values of variables in $\text{Var}_S \subseteq \text{Var}$.
- Public outputs — final values of variables in $\text{Var}_P \subseteq \text{Var}$.

Let $D_n \in \mathcal{D}(\text{State}_n)$ be the distribution of input states for security parameter n . Computational independence means:

$$\{(s_n | \mathbf{Var}_S, t_n | \mathbf{Var}_P) : s_n \leftarrow D_n, t_n \leftarrow \llbracket \mathbf{P} \rrbracket_n(s_n)\} \approx \{(s_n | \mathbf{Var}_S, t'_n | \mathbf{Var}_P) : s_n, s'_n \leftarrow D_n, t'_n \leftarrow \llbracket \mathbf{P} \rrbracket_n(s'_n)\}$$

Programs running in polynomial time

This def. is good for programs running in expected polynomial time.

If a program leaks information only after exponentially long time, then the previous definition still considers it insecure.

- Let P^ℓ be a program that makes at most $\ell(n)$ steps of P .
If P has not stopped, then P^ℓ stops in a special state \perp .
(ℓ — a polynomial)
- P^ℓ can be expressed in the WHILE-language.
 - The rewrite of P to P^ℓ is quite simple.

P is secure $:\iff \forall \ell : P^\ell$ is secure.

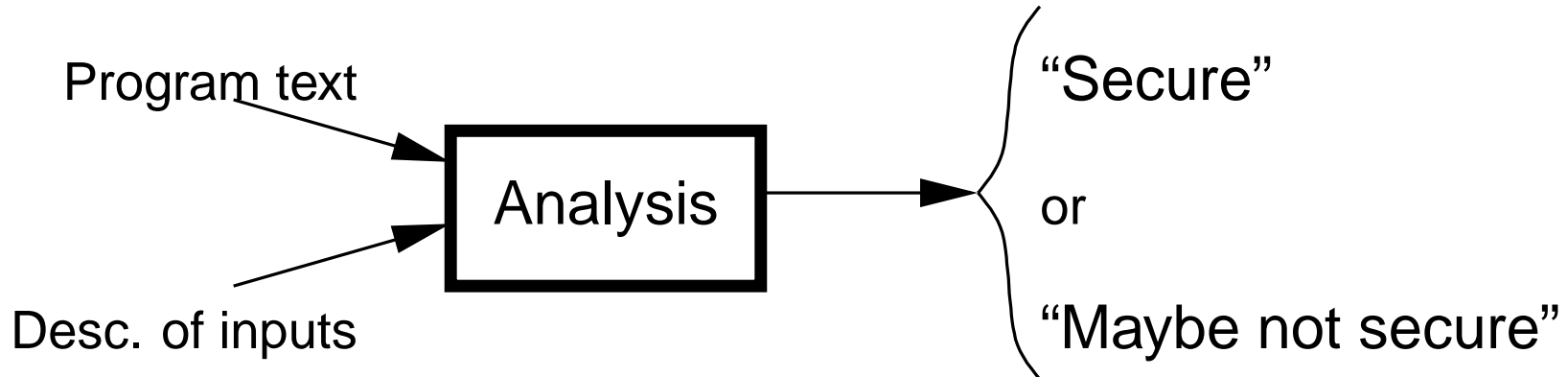
Timing-insensitive def.

- Definition on previous slide is timing-sensitive.
 - This is good.
- Sometimes we do not want timing sensitivity.
 - Good timing-sensitive analyses are hard to construct.
 - Timing issues seem to be orthogonal to computational issues.

P is secure $:\iff \exists l_0 \forall l \geq l_0 : P^l$ is secure.

- To analyse P , we analyse P^l .
 - ...but the number of executed steps is only checked at loop heads.

Program analysis's approach



- Having secure information flow is uncomputable in general.
- Description of inputs — whatever is known about D .
 - ... and expressible in the domain of the analysis.

Domain of the analysis

- Analysis maps the description of the input distribution to the description of the output distribution.
- Description of $D = \{D_n\}_{n \in \mathbb{N}}$ is $(\mathcal{X}, \mathcal{K}) \in \mathcal{P}(\mathcal{P}(\mathbf{Var}) \times \mathcal{P}(\mathbf{Var})) \times \mathcal{P}(\mathbf{Var})$.
 - $(X, Y) \in \mathcal{X}$, if X and Y are independent in D .
 - $k \in \mathcal{K}$, if (the value of) k is distributed like a key.
- Assume the program does not change the variables in \mathbf{Var}_S .
- If $(\mathbf{Var}_S, \mathbf{Var}_P) \in \mathcal{X}_{\text{output}}$, then the program has secure information flow.
- The analysis is defined inductively over the program structure.

Example: analysing assignments

Consider the program $x := o(x_1, \dots, x_k)$.

If $(X \cup \{x_1, \dots, x_k\}, Y) \in \mathcal{X}_{\text{input}}$

then $(X \cup \{x_1, \dots, x_k, x\}, Y) \in \mathcal{X}_{\text{output}}$.

Analysing encryptions — problems

Let k be distributed like a key in D_{input} .

- Consider the program $l := k + 1$.

Then $\{l\}$ is not independent of $\{k\}$ in D_{output} .

- Consider the program $x := \text{Enc}(k, y)$.

Then $\{x\}$ is not independent of $\{k\}$ in D_{output} .

- To check whether x and k come from the same or from different samples of D_{output} , try to decrypt x with k .

These two cases should be distinguished as l is usable for decryption but x is not.

Encrypting black boxes

- Let $\mathbf{k} \in \text{Var}$. Let S_n be a program state.
- $S_n([\mathbf{k}]_{\mathcal{E}})$ denotes a black box that encrypts with \mathbf{k} . I.e.
 - $S_n([\mathbf{k}]_{\mathcal{E}})$ has an input tape and an output tape;
 - When a bit-string w is written on the its tape,

$$\llbracket \mathcal{E}nc \rrbracket_n(S_n(\mathbf{k}), w)$$

is invoked and the result written to the output tape.

- Indistinguishability can be defined for distributions over black boxes.
 - Independence can be defined, too.
- Security of $(\llbracket \mathcal{G}en \rrbracket, \llbracket \mathcal{E}nc \rrbracket)$ is defined as the indistinguishability of certain black boxes.

Security of encryption

- $(\mathcal{G}, \mathcal{E})$ is **secure against CPA**, iff

$$\{\boxed{\mathcal{E}_k(\cdot)} : k \leftarrow \mathcal{G}\} \approx \{\boxed{\mathcal{E}_k(\mathbf{0})} : k \leftarrow \mathcal{G}\}$$

- $(\mathcal{G}, \mathcal{E})$ is **which-key concealing**, iff

$$\{(\boxed{\mathcal{E}_k(\cdot)}, \boxed{\mathcal{E}_{k'}(\cdot)}) : k, k' \leftarrow \mathcal{G}\} \approx \{(\boxed{\mathcal{E}_k(\cdot)}, \boxed{\mathcal{E}_k(\cdot)}) : k \leftarrow \mathcal{G}\}$$

$(\llbracket \mathcal{G}en \rrbracket, \llbracket \mathcal{E}nc \rrbracket)$ must satisfy both.

Modified domain of the analysis

- Let $\widetilde{\mathbf{Var}} = \mathbf{Var} \uplus \{[\mathbf{x}]_{\varepsilon} : \mathbf{x} \in \mathbf{Var}\}$.
- Description of a distribution D is

$$(\mathcal{X}, \mathcal{K}) \in \mathcal{P}(\mathcal{P}(\widetilde{\mathbf{Var}}) \times \mathcal{P}(\widetilde{\mathbf{Var}})) \times \mathcal{P}(\mathbf{Var}) .$$

- $(X, Y) \in \mathcal{X}$ if X and Y are independent in D .
- $\mathbf{k} \in \mathcal{K}$, if the distribution of $[\mathbf{k}]_{\varepsilon}$ according to D is indistinguishable from $\boxed{[[\mathcal{E}nc]]_{\mathbf{k}}(\cdot)}$.

Analysing encryptions

Consider the program $x := \mathcal{E}nc(k, y)$.

If $(X, Y) \in \mathcal{X}_{\text{input}}$

and $k \in \mathcal{K}_{\text{input}}$

and $(\{[k]_{\mathcal{E}}\}, X \cup Y \cup \{y\}) \in \mathcal{X}_{\text{input}}$

then $(X \cup \{x\}, Y) \in \mathcal{X}_{\text{output}}$.

Generally $(\{[k]_{\mathcal{E}}\}, \{[k]_{\mathcal{E}}\}) \in \mathcal{X}_{\text{input}}$, **hence** $(\{x\}, \{[k]_{\mathcal{E}}\}) \in \mathcal{X}_{\text{output}}$.

If we have a program $1 := k + 1$, then $(\{1\}, \{[k]_{\mathcal{E}}\}) \notin \mathcal{X}_{\text{output}}$.

On security def. of encryptions

- In the definition a system is considered, consisting of
 - the adversary,
 - the encrypting black box,
 - ...
- The key is inside the black box.
 - I.e. the usage of the key is quite restricted.
- Programming language puts no restrictions on the usage of the variable containing the key.
- Requirement $(\{[k]_{\mathcal{E}}\}, X \cup Y \cup \{y\}) \in \mathcal{X}_{\text{input}}$ gives the necessary restrictions.

Analysing key generations

Consider the program $k := \mathit{Gen}()$.

if $(X, Y) \in \mathcal{X}_{\text{input}}$

then $k \in \mathcal{K}_{\text{output}}$

and $(X \cup \{[k]_{\mathcal{E}}\}, Y \cup \{[k]_{\mathcal{E}}\}) \in \mathcal{X}_{\text{output}}$.

Analysing if-then-else

Consider the program *if* b *then* P_1 *else* P_2 .

- Let $\{x_1, \dots, x_k\} = \text{Var}_{\text{asgn}} \subseteq \text{Var}$ be the set of variables assigned to in P_1 and P_2 .

- Let $\text{Var}' = \text{Var} \cup \{N, x_1^{\text{true}}, \dots, x_k^{\text{true}}, x_1^{\text{false}}, \dots, x_k^{\text{false}}\}$

- Program at right has the same functionality.

- P_1^{true} is P_1 , where each x_i is replaced with x_i^{true} .

- Similarly for P_2^{false} .

$N := b$

$x_1^{\text{true}} := x_1$

$x_1^{\text{false}} := x_1$

...

$x_k^{\text{true}} := x_k$

$x_k^{\text{false}} := x_k$

P_1^{true}

P_2^{false}

$x_1 := N ? x_1^{\text{true}} : x_1^{\text{false}}$

...

$x_k := N ? x_k^{\text{true}} : x_k^{\text{false}}$

Analysing ? :

Consider the program $x := b ? y : z$. Let $y, z \in \mathcal{K}_{\text{input}}$.

- If $(X, Y) \in \mathcal{X}_{\text{input}}$ and $(\{[y]_{\varepsilon}\}, \{[z]_{\varepsilon}\}, X \cup Y \cup \{b\}) \in \mathcal{X}_{\text{input}}$ then $(X \cup \{[x]_{\varepsilon}\}, Y \cup \{[x]_{\varepsilon}\}) \in \mathcal{X}_{\text{output}}$.
- If $(\{[y]_{\varepsilon}, [z]_{\varepsilon}\}, \{b\}) \in \mathcal{X}_{\text{input}}$ then $x \in \mathcal{K}_{\text{output}}$.

$(X_1, \dots, X_k) \in \mathcal{X}$ means

$$(X_1, X_2) \in \mathcal{X}$$

$$(X_1 \cup X_2, X_3) \in \mathcal{X}$$

.....

$$(X_1 \cup \dots \cup X_{k-1}, X_k) \in \mathcal{X}$$

Analysing loops

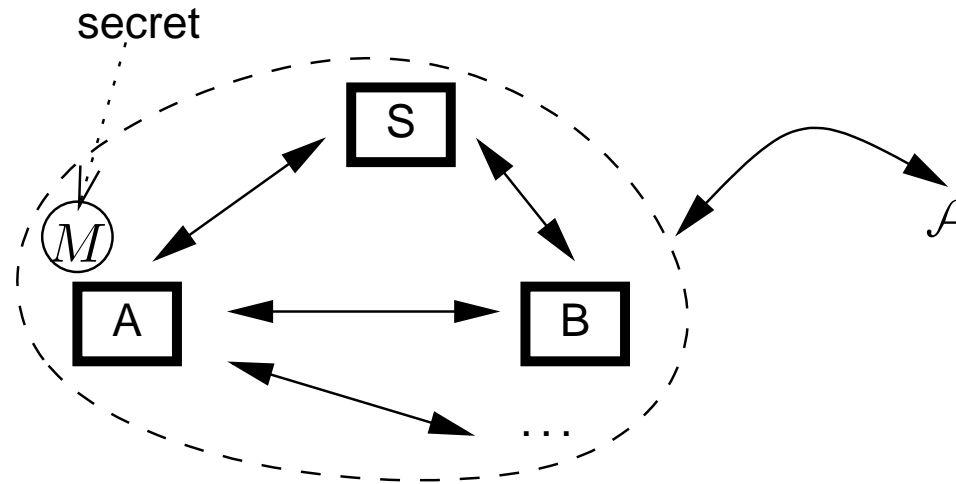
Consider the program *while* b *do* P .

Its analysis is the repeated application of the analysis of

if b *then* P *else skip*

It stabilises due to finiteness of the domain and monotonicity of the analysis.

Active adversaries — problem statement



M remains confidential if for all adversaries \mathcal{A} , the adversary's experience is independent of M .

Language for protocols

- A party is a sequence of statements. Statements are:

$k := \mathit{Gen}$	$x := \mathit{random}$
$x := (y_1, \dots, y_m)$	$y := \pi_i^m(x)$
$x := \mathit{encr}_k(y)$	$y := \mathit{decr}_k(x)$
$\mathit{send } x$	$x := \mathit{receive}$
$\mathit{check}(x = y)$	

- Protocol is a set of parties.
- Some additional statements (generation of long-term keys) are done at the very beginning of execution.
- Each variable may occur at LHS at most once.

Semantics

Protocol runs in parallel with the adversary.

- Adversary takes care of message forwarding.
- If something goes wrong during the execution of a party, then this party becomes stuck.
 - $\text{check}(x = y)$ returns false;
 - operand types do not match the operator;
 - a message does not decrypt.
- Parties execute one statement at a time, the adversary does the scheduling.
 - When a party gets stuck, the adversary is not notified immediately.

Adversary's experience

- Adversary learns the values of the variables x , where $\text{send } x$ is a statement in some party.
- No timing information is available, because the adversary schedules.
- Therefore there is again a set of public variables Var_P , whose values make up the entire experience.

$$\text{Var}_P = \{x \mid \text{some party contains send } x\}$$

Denning-style analysis

- Suppose a statement $x := O(x_1, \dots, x_m)$ occurs in some party.
 - x_1, \dots, x_m are all variables occurring in RHS.
 - O can be any operation — tupling, projection, decryption, encryption.
- There is **information flow** from x_i to x .
 - Denote $x_i \Rightarrow x$.
- Protocol is **insecure**, if $M \xRightarrow{*} x$ for some $x \in \text{Var}_P$.
 - Otherwise it is **secure**.

An extremely conservative analysis.

Security against CCA

Encryption system $(\mathcal{G}, \mathcal{E}, \mathcal{D})$ is **secure against CCA**, if

$$\{(\boxed{\mathcal{E}_k(\cdot)}, \boxed{\mathcal{D}_k(\cdot)}) : k \leftarrow \mathcal{G}\}$$

is indistinguishable from

$$\{(\boxed{\mathcal{E}_k(\mathbf{0})}, \boxed{\mathcal{D}_k(\cdot)}) : k \leftarrow \mathcal{G}\}$$

by all adversaries that do not give the output of the left black box to the right black box.

Main idea

- Replace statements $x := \text{encr}_k(y)$ with statements $x := \text{encr}_k(Z)$, where $\llbracket Z \rrbracket = 0$.
 - Z is a new variable.
- This makes the information flow relation \Rightarrow sparser.
- The replacement is valid only when certain conditions are satisfied.
 - Valid \equiv does not change the adversary's experience.

Conditions for replacing

When replacing the statement $x := \text{encr}_k(y) \dots$

- We must know exactly, where else the key k is used.
 - The same key may occur under different names.
 - To find it out, we symbolically execute the protocol.
- When computing the values of the variables in Var_P , the key k may only be used to encrypt and decrypt.
- We may not decrypt the ciphertexts created with key k .
 - We achieve this with a program transformation.

Symbolic execution of protocols

We assign a term to each variable. The terms T are

$$\underline{const}(\mathbf{x}) \quad \underline{tuple}^m(T_1, \dots, T_m)$$

$$\underline{secret}(\mathbf{M}) \quad \underline{\pi}_i^m(T)$$

$$\underline{key}(\mathbf{x}) \quad \underline{encr}(l, T_k, T_y)$$

$$\underline{key}(l) \quad \underline{decr}(T_k, T_y)$$

$$\underline{random}(l) \quad \underline{received}(l)$$

$$\underline{stuck}$$

- l — statement label.
- $\underline{\dots}(\mathbf{x})$ is assigned to the variable \mathbf{x} that is initialised before the run of the protocol.
- There are some obvious simplification rules.

Symbolic execution of *Check-s*

- There are some rules telling us, when the bit-strings corresponding to two terms are certainly different.
- For $\text{check}(x = y)$, we check whether terms assigned to x and y are certainly different.
 - If yes, the protocol party is stuck.
 - If no, then we replace the more complex term with the simpler one everywhere.
 - Complexity is the same as size.
 - But: the terms containing subterms *received*(l) are the most complex.

We consider the key corresponding to *key*(l) to be used exactly where the subterm *key*(l) occurs.

Replacing decryptions

Let k be used for encryption at statements

$$x_1 := \text{encr}_{k_1}(y_1), \quad \dots, \quad x_m := \text{encr}_{k_m}(y_m)$$

Replace $z := \text{decr}_k(w)$ by

$z := \text{case } w \text{ of}$

$x_1 \rightarrow y_1$

.....

$x_m \rightarrow y_m$

$\text{else} \rightarrow \text{decr}_k(w)$

No change to
adversary's
view

Semantics of *case-constructs*

- z is assigned the first y_i , where x_i matches w .
- If this y_i has not been defined yet, then the protocol party gets stuck.
 - This never happens in our transformed protocols.
- A yet undefined x_i never matches.

Ciphertext integrity

An encryption system $(\mathcal{G}, \mathcal{E}, \mathcal{D})$ has **ciphertext integrity**, if:

No PPT algorithm \mathcal{A} with access to oracles $\boxed{\mathcal{E}_k(\cdot)}$ and $\boxed{\mathcal{D}_k(\cdot)}$ can submit to $\boxed{\mathcal{D}_k(\cdot)}$ a bit-string y , such that

- $\mathcal{D}_k(y)$ exists, i.e. y is a valid ciphertext;
- y was not an output of $\boxed{\mathcal{E}_k(\cdot)}$.

i.e. we need no *else*-clause.

- If nothing matches in a *case*-statement, then the protocol party gets stuck.

See [Bellare and Namprempre, ASIACRYPT 2000] for constructions of encryption primitives.

The replacement — wrap-up

- Do the symbolic execution.
- Choose a key $\underline{key}(x)$ or $\underline{key}(l)$, such that
 - In terms assigned to $y \in \text{Var}_P$, this $\underline{key}(\dots)$ occurs only as the key in en-/decryptions.
- Replace the decryption statements $z := \text{decr}_k(y)$, where the term assigned to k is this $\underline{key}(\dots)$.
 - Replace them with *case*-statements.
- Replace the encryption statements $x := \text{encl}_k(y)$, where the term assigned to k is this $\underline{key}(\dots)$.
 - Replace them with $x := \text{encl}_k(Z)$.

Getting rid of *case*-statements

$$z := \text{case } w \text{ of } x_1 \rightarrow y_1 \cdots x_m \rightarrow y_m,$$

where

$$x_1 := \text{encl}_{k_1}(y_1), \quad \dots, \quad x_m := \text{encl}_{k_m}(y_m)$$

is replaced by

$$\begin{aligned} & \text{wait}(s) \\ & \text{check}(w = x_i) \\ & z := y_i \end{aligned}$$

and $\text{signal}(s)$ is added after $x_i := \text{encl}_{k_i}(y_i)$.

i is chosen **nondeterministically** (we get m new protocols).
 s is a new **semaphore**.

Executing $\text{wait}(s)$ before $\text{signal}(s)$ gets stuck.

Handling *wait-s* and *signal-s*

- in the next round, the symbolic execution must proceed in an order consistent with *wait-s* and *signal-s*.
 - We may have to do simultaneous symbolic execution of the parties.
- If there are cyclic dependencies, then the statements in and after the cycle are stuck.

Conclusions

- Cryptographic effects can be faithfully abstracted away.
- Resulting analyses are not overwhelmingly complex.

Future work

- Track the keys in the first analysis presented.
- Do not track the keys in an analysis with active adversaries.
 - Assume that keys are never sent out.
- More expressive language for the second analysis.
- More cryptographic primitives.
 - Public key encryption, digital signatures, . . .
- Other security properties. (Integrity)
- Different security definitions for cryptographic primitives.
 - Encryption as a PRP . . .
- One-way functions.
 - New confidentiality definition is necessary.