Probabilistic Information Flow

Catuscia Palamidessi
INRIA Saclay & Ecole Polytechnique

Based on joint work with these people

Prakash Panangaden

Miguel E. Andrés

Mário S. Alvim

Konstantinos Chatzikokolakis

Plan of the talk

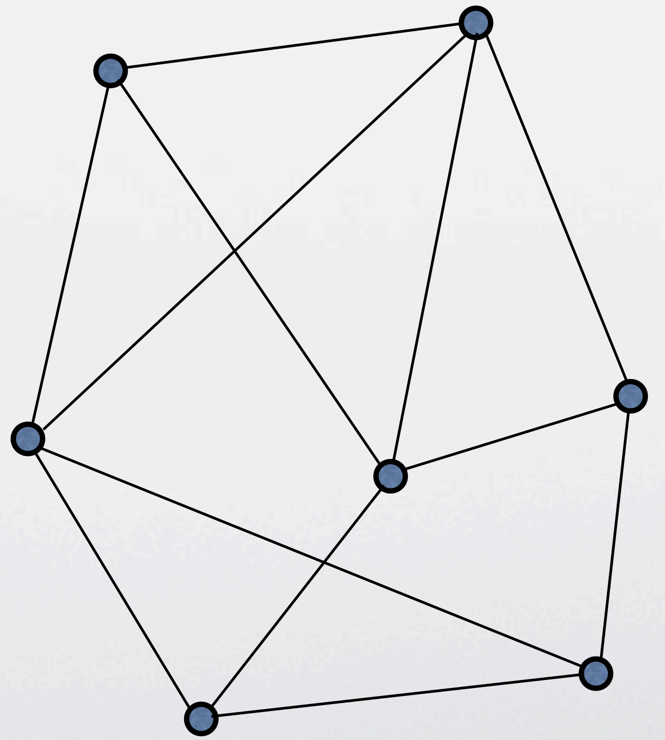
- Information Flow in a probabilistic setting. Examples
- Possibilistic approaches
- Probabilistic approaches
- Information-theoretic approaches
- Approach based on statistical inference and Bayesian risk
- Some relations between the various approaches
- Problems in extending the framework to the interactive case
- Verification

Information hiding

- Information flow (originally): leak of information from high variables to low variables
- Information flow is an aspect of a general problems called information-hiding: ***Prevent an observer from inferring secret information from the information made available to him (observables).***
- Other problems that can be seen as Information-hiding problems: Anonymity, Privacy, Untreaceability, Confidentiality, Secrecy ...
- In particular, the communities of **Information Flow** and of (Theory of) **Anonymity** are converging on the formal approaches
- This talk will be about the common foundations, with particular focus on the **probabilistic aspects**
- Two examples from Anonymity: DC Nets and Crowds

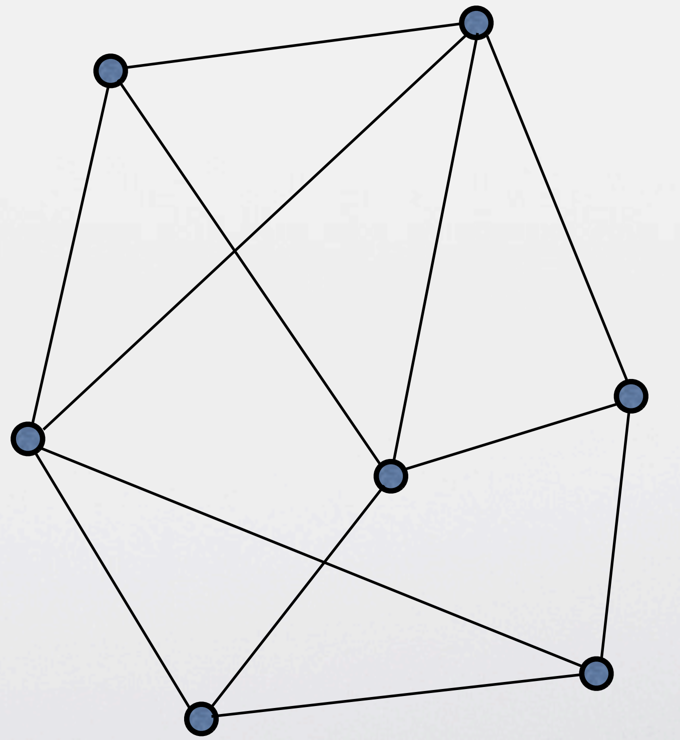
Example: DC Nets (Chaum 88)

- A set of nodes with some communication channels (edges).
- One of the nodes (source) wants to broadcast one bit b of information
- The source must remain anonymous



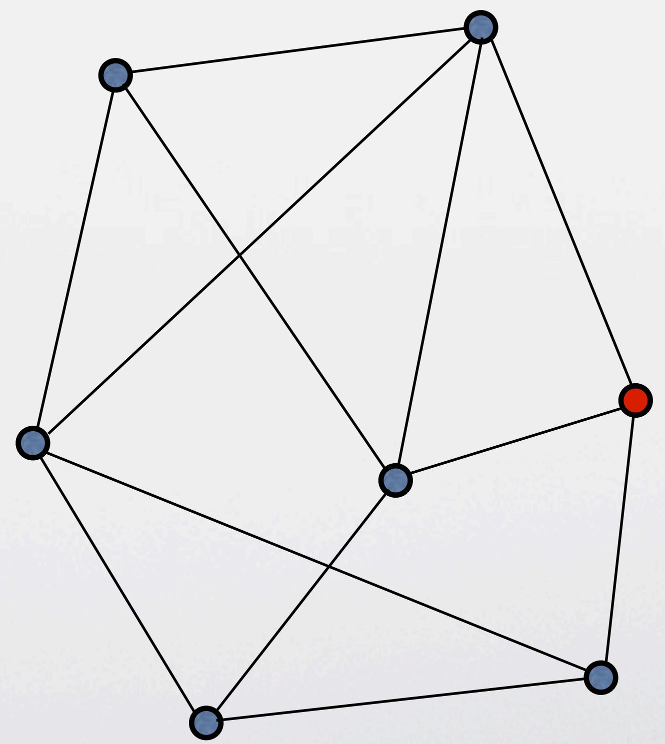
A possible solution

- Associate to each edge a fair coin
- Toss the coins
- Each node computes the binary sum of the incident edges. The source adds b. They all broadcast their results

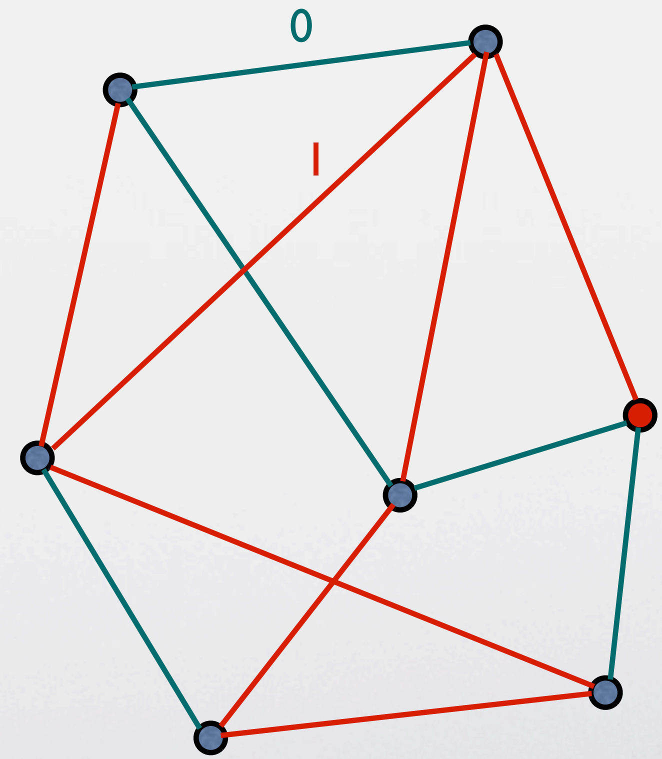


A possible solution

- Associate to each edge a fair coin
- Toss the coins
- Each node computes the binary sum of the incident edges. The source adds b. They all broadcast their results



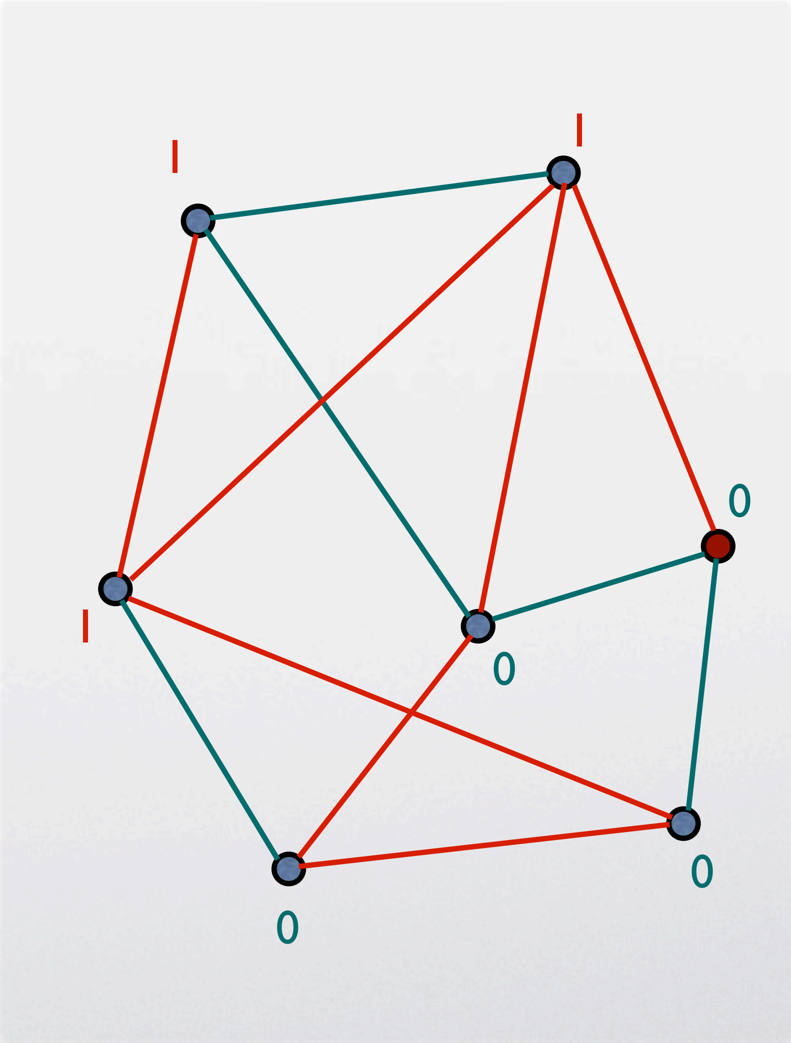
A possible solution



- Associate to each edge a fair coin
- Toss the coins
- Each node computes the binary sum of the incident edges. The source adds b. They all broadcast their results

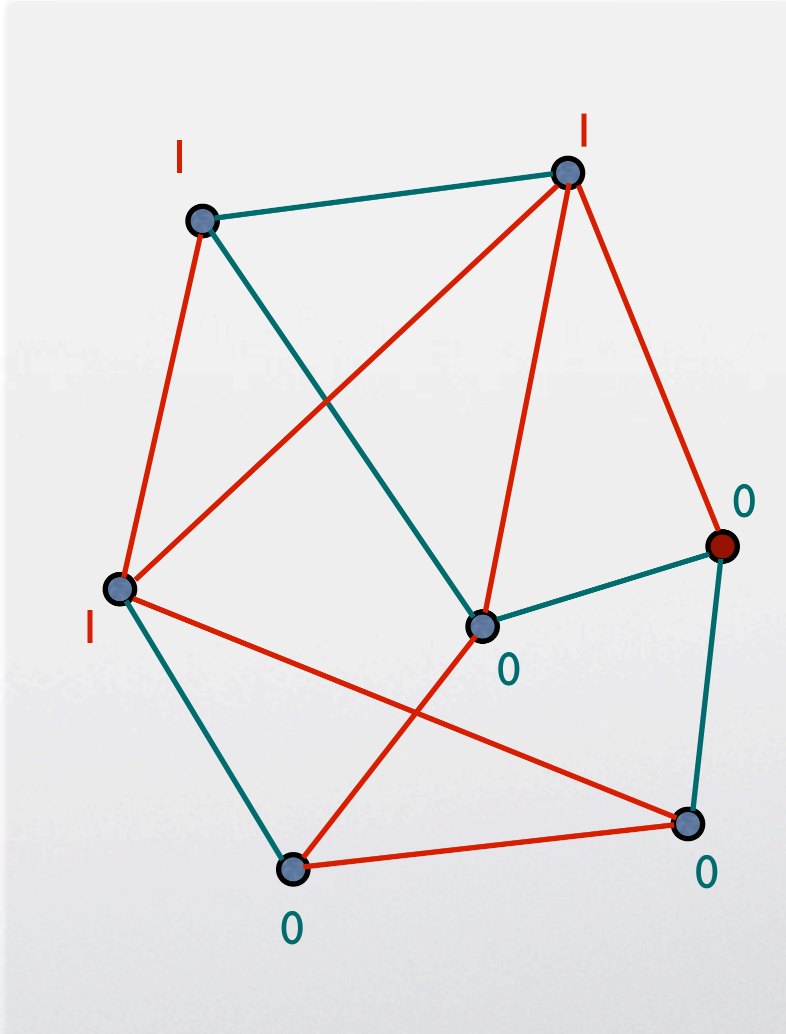
Correctness

- Associate to each edge a fair coin
- Toss the coins
- Each node computes the binary sum of the incident edges. The source adds b. They all broadcast their results
- The total binary sum is computed
- **Correctness**: The total binary sum equals b



Anonymity

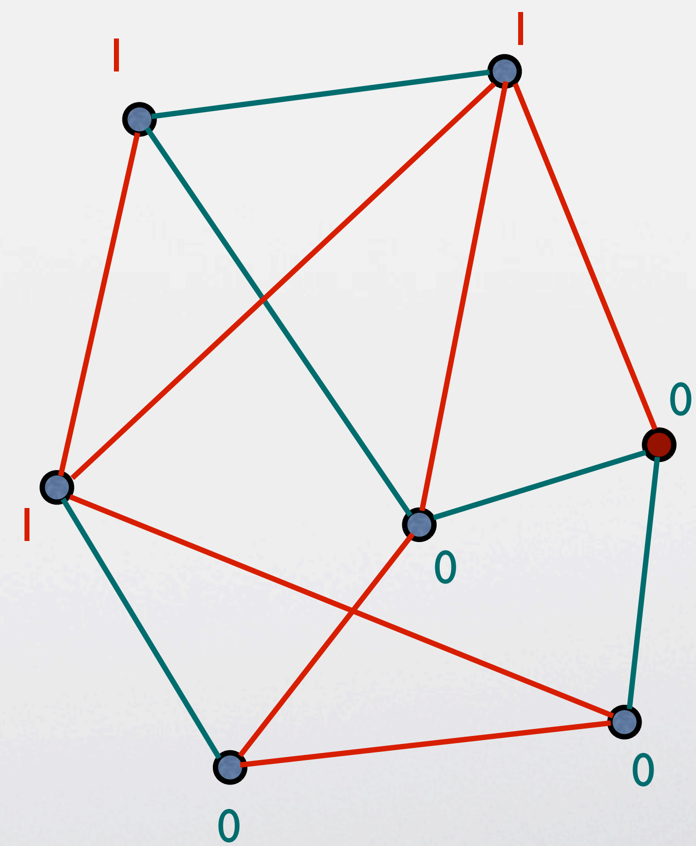
- How should anonymity be formulated ?



Strong anonymity

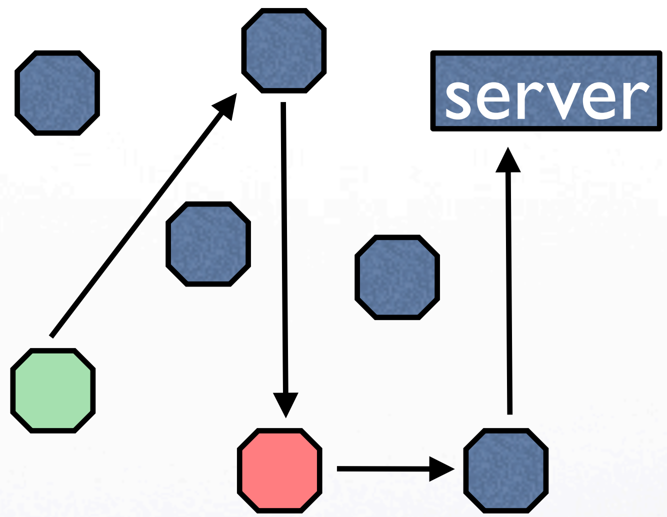
- **Strong anonymity:**
  If the graph is connected and the coins are fair, then for an external observer, the *a posteriori* probability that a certain node is the source is equal to its *a priori* probability

- Question: what about the internal nodes?



Example: Crowds

- Problem: A user (initiator) wants to send a message anonymously to a server.
- Crowds: A group of n users who agree to participate in the protocol.
- The initiator selects randomly anotehr user (forwarder) and forwards the request to it
- A forwarder:
  - With prob. p_f selects randomly another forwarder and
    forwards the request to him
  - With prob. 1-p_f sends the request to the server



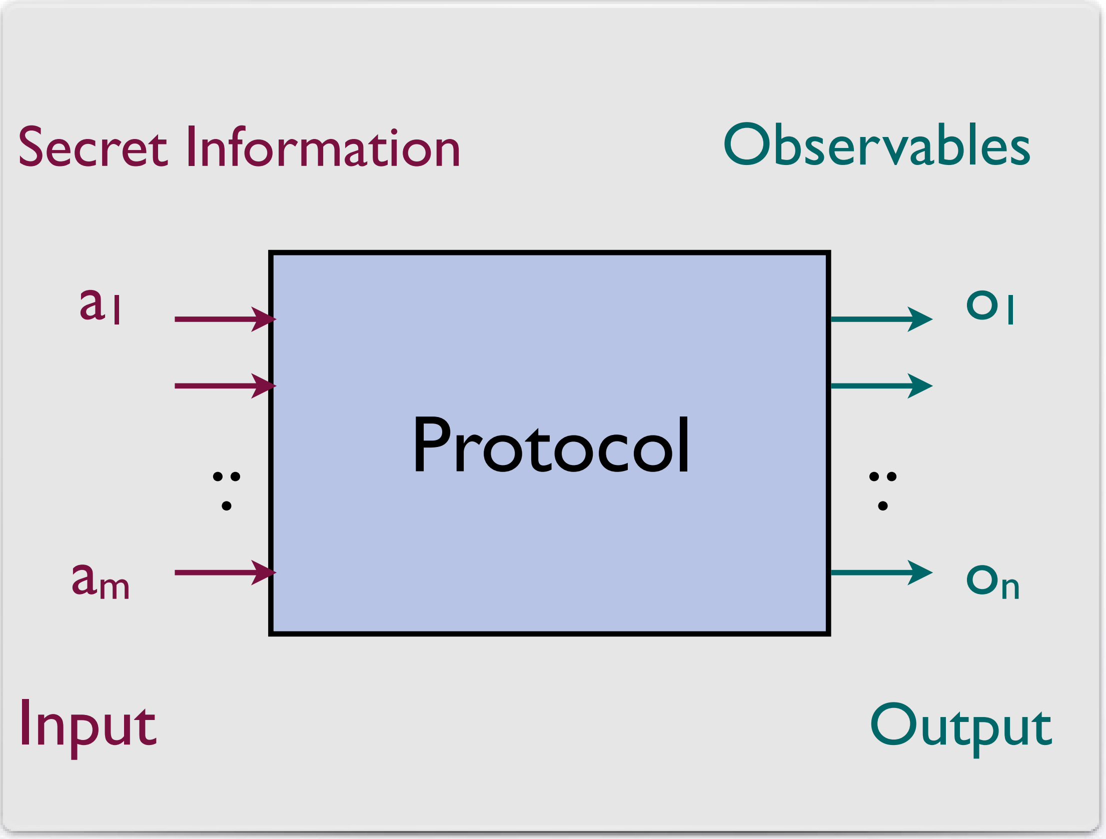
Probable innocence: under certain conditions, the attacker who intercepts the message from x cannot attribute more than 0.5 probability to x to be the initiator

Common features in information hiding

- There is information that we want to keep secret
  - the source in DC Nets
  - the initiator in Crowds
- There is information that is revealed (observables)
  - agree/disagree in DC Nets
  - the users who forward messages to a corrupted user in Crowds
- The value of the secret information may be chosen probabilistically. Furthermore, protocols may use randomization to hide the link between hidden and observable information
  - coin tossing in DC Nets
  - random forwarding to another user in Crowds

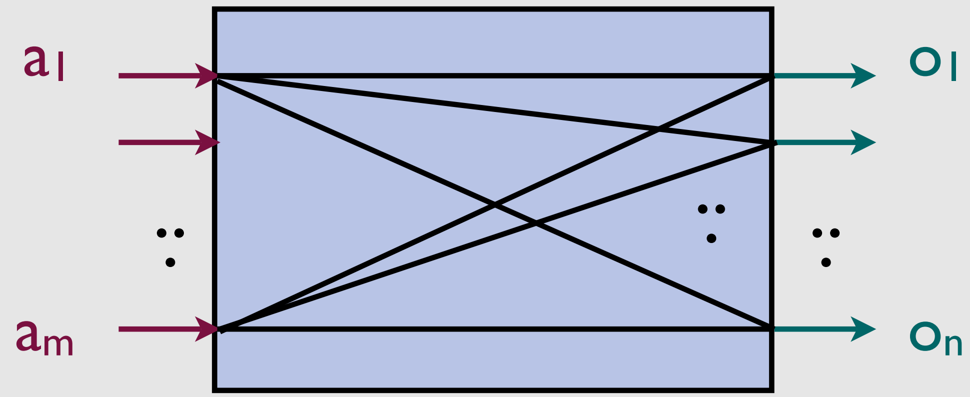
Assumptions

- For the moment we consider the non-interactive case: Each activation of the system receives exactly one input and produces exactly one output
  - Inputs: elements of a random variable A
  - Outputs: elements of a random variable O
  - For each input a, the probability that we obtain an observable o is given by p(o | a)



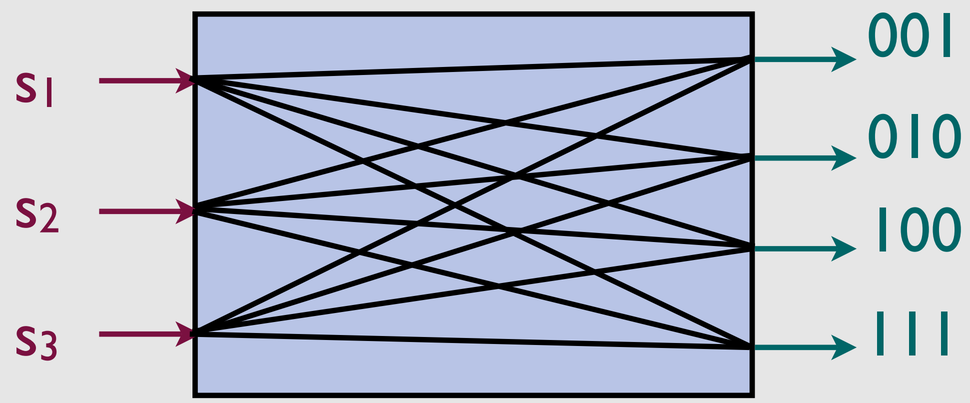
General framework:

Protocols as Information-Theoretic channels

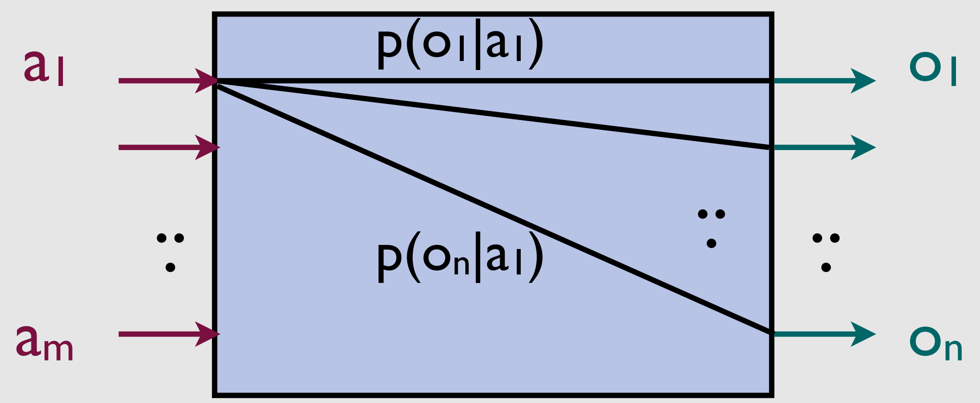


Protocols are noisy channels. Each run has 1 input and 1 output, but:

- an input can generate different outputs (according to a prob. distr.)
- an output can be generated by different inputs



Example: DC Nets with 3 nodes, when b=1



The conditional probabilities

	o_1	...	o_n
a_1	$p(o_1 \mid a_1)$	...	$p(o_n \mid a_1)$
⋮	⋮		
a_m	$p(o_1 \mid a_m)$		$p(o_n \mid a_m)$

A channel is characterized by its matrix:
the array of conditional probabilities

Possibilistic approaches

- Schneider and Sidiropoulus, and many others ...
- Key idea: Replace the random choices by nondeterministic choices
- Common principle: A system P has no leakage iff: For every pair of secret values a, a′, P[a] “is equivalent” to P[a′]
- Criticisms:
  - Too weak: it collapses uniform distrib and non-zero distrib
  - It assumes that the scheduler “helps”

Problem of the scheduler: Consider the following system

$$S \stackrel{\text{def}}{=} (c, out)(A \parallel H_1 \parallel H_2 \parallel Corr),$$

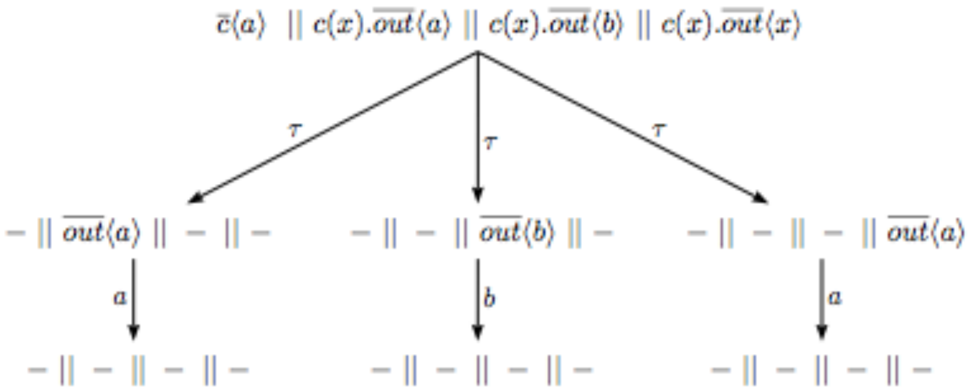
$$A \stackrel{\text{def}}{=} \bar{c}\langle sec \rangle, \quad H_1 \stackrel{\text{def}}{=} c(s).\overline{out}\langle a \rangle, \quad H_2 \stackrel{\text{def}}{=} c(s).\overline{out}\langle b \rangle, \quad Corr \stackrel{\text{def}}{=} c(s).\overline{out}\langle s \rangle$$

- Intuitively, the system is not secure. However *S*[*a*/*sec*] and *S*[*b*/*sec*] are bisimilar
- The problem is that nondeterminism in concurrency is meant as underspecification
- Standard implementation refinement preserves properties expressed on individual paths, but no-leakage is expressed as a global property.

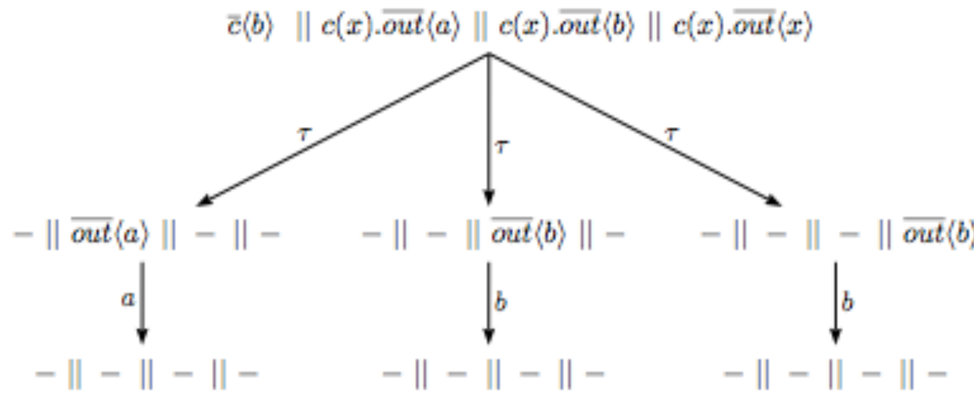
$$S \stackrel{\text{def}}{=} (c, out)(A \parallel H_1 \parallel H_2 \parallel Corr),$$

$$A \stackrel{\text{def}}{=} \bar{c}\langle sec \rangle, \quad H_1 \stackrel{\text{def}}{=} c(s).\overline{out}\langle a \rangle, \quad H_2 \stackrel{\text{def}}{=} c(s).\overline{out}\langle b \rangle, \quad Corr \stackrel{\text{def}}{=} c(s).\overline{out}\langle s \rangle$$

S[*a*/*sec*]



S[*b*/*sec*]



Probabilistic approaches

(1) [Halpern and O'Neill - like] for all a, a': $p(a|o) = p(a'|o)$

(2) [Chaum]: for all a, o: $p(a|o) = p(a)$

(3) [Bhargava and Palamidessi]: for all a, a′, o: $p(o|a) = p(o|a')$

- From standard probability theory we can easily derive that (2) and (3) are equivalent.
- (3) has the following advantages:
  - It does not require to know the a priori distribution p(a)
  - It is independent from the priori distribution and even from its existence

Probabilistic approaches

(1) [Halpern and O'Neill - like] for all a, a': $p(a|o) = p(a'|o)$

(2) [Chaum]: for all a, o: $p(a|o) = p(a)$

(3) [Bhargava and Palamidessi]: for all a, a′, o: $p(o|a) = p(o|a')$

- (1) is equivalent to (2) + the condition $p(a) = p(a')$ for all a, a′
- Uniform probability on the secrets is too strong

Probabilistic approaches

(1) [Halpern and O'Neill - like] for all a, a': $p(a|o) = p(a'|o)$

(2) [Chaum]: for all a, o: $p(a|o) = p(a)$

(3) [Bhargava and Palamidessi]: for all a, a′, o: $p(o|a) = p(o|a')$

- (1) is equivalent to (2) + the condition $p(a) = p(a')$ for all a, a′
- Uniform probability on the secrets is too strong

- But actually all these notions are too strong in practice. We would like a notion that quantifies the *degree of protection*

Information-theoretic approaches

- The entropy $H(A)$ measures the uncertainty about the hidden events:

$$H(A) = -\sum_{a \in \mathcal{A}} p(a) \log p(a)$$

- The conditional entropy $H(A|O)$ measures the uncertainty about A after we know the value of O (after the execution of the protocol).
- The mutual information $I(A;O)$ measures how much uncertainty about A we lose by observing O:

$$I(A;O) = H(A) - H(A|O)$$

Information-theoretic approaches

Various definitions of protection / information leakage

1. Entropy on the hidden information H(A) [Diaz et al.]
2. Mutual information I(A;O) [Malacaria et al.] [Zhu et al.]
3. Capacity $C = \max_{p(a)} I(A;O)$ [Moscowitz et al.] [CPP]

- Note that C = 0 iff for all a, a', o, p(o|a) = p(o|a')
- (1) has noting to do with the protocol.
  (2) and (3) are the most commonly accepted ones

Statistical Inference approach

- A natural definition of vulnerability: the "probability of guessing the right value" in one try

- Leakage = A priori vulnerability − A posteriori vulnerability

- A priori vulnerability: max p(a)

- A posteriori vulnerability: weighted average of the max p(a|o) (converse of Bayes risk)

Statistical vs Information Theoretic approach

- **Good news:**

Capacity = 0
iff
A Priori Vulnerability = A Posteriori Vulnerability
iff
p(o|a) = p(o|a') for all a, a',o

- **Bad news (Smith'09):** in general there is not a good match between the IT approach (based on Shannon entropy) and the approach based on the probability of error (difference between a priori and a posteriori vulnerability)

Mismatch btw IT and probability of error

- Example due to Smith'09. Consider A = random number in $[0, 2^{32}-1]$ with uniform a priori

1. O = if (A mod 8) == 0 then A else 0
2. O = A && 37

- These two programs have almost the same Muntual Information. (The one for (2) is slightly higher.)
- However, the a posteriori vulnerability of (1) is much higher than the one of (2):
  - For (1) the a posteriori vulnerability is about 1/8.
  - For (2) the a posteriori vulnerability is $1/2^{27}$

Statistical vs Information Theoretic approach

- **Good news:**

  Capacity = 0
  iff
  A Priori Vulnerability = A Posteriori Vulnerability
  iff
  p(o|a) = p(o|a') for all a, a',o

- **Bad news (Smith'09):** in general there is not a good match between the IT approach (based on Shannon entropy) and the approach based on the probability of error (difference between a priori and a posteriori vulnerability)

- For geometrical distributions Shannon entropy is closely related to the "guessing entropy", defined as the average number of tries necessary to guess the right value. (However, the guessing entropy can be misleading from the security p.o.v.)

- Smith'09: Mutual Information in terms of **Rényi's min entropy** corresponds to the difference between the a posteriori and the a priori vulnerability (in log)

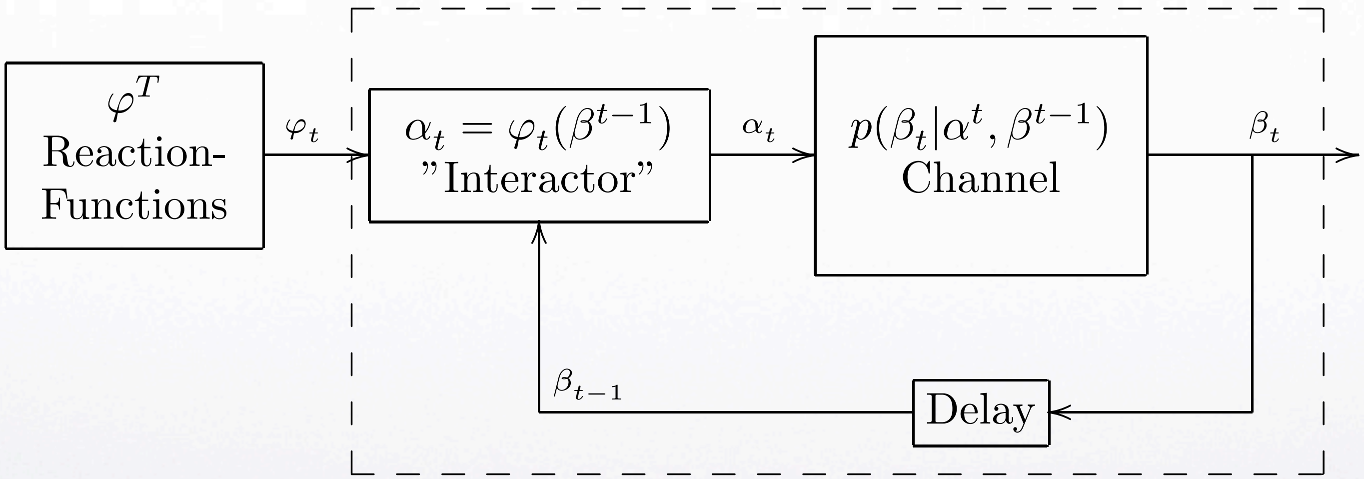
Interactive case

- Secrets and observables may alternate during the execution
- Example: Ebay-like system
- Problems in defining the channel matrix

Interactive case

- Channels with memory and feedback
- Directed mutual information
- Directed capacity
- Open problem: generalize the approach based on the Bayes risk

Interactive case



Thank you !