Formal approaches to Information-hiding

- An overview -

Catuscia Palamidessi
INRIA Futurs & Ecole Polytechnique

based on joint work with

Kostas Chatzikokolakis and Prakash Panangaden
Ecole Polytechnique Mc Gill University

Plan of the talk

- Motivation
- Protocols for information-hiding
- Possibilistic approaches
- Probabilistic approaches
- Information-theoretic approaches
- Approach based on statistical inference and Bayesian risk
- Some relations between the various approaches
- Verification

Information-hiding: Privacy

- Ability of an individual or group to stop information about themselves from becoming known to people other than those they choose to give the information to [Wikipedia]
  - **Protection of private data** (credit card number, personal info etc.)
  - **Anonymity**: protection of identity
  - **Unlinkability**: protection of link between information and user
  - **Unobservability**: impossibility to determine what the user is doing

More precise definition @ www.freehaven.net/anonbib/cache/terminology.pdf

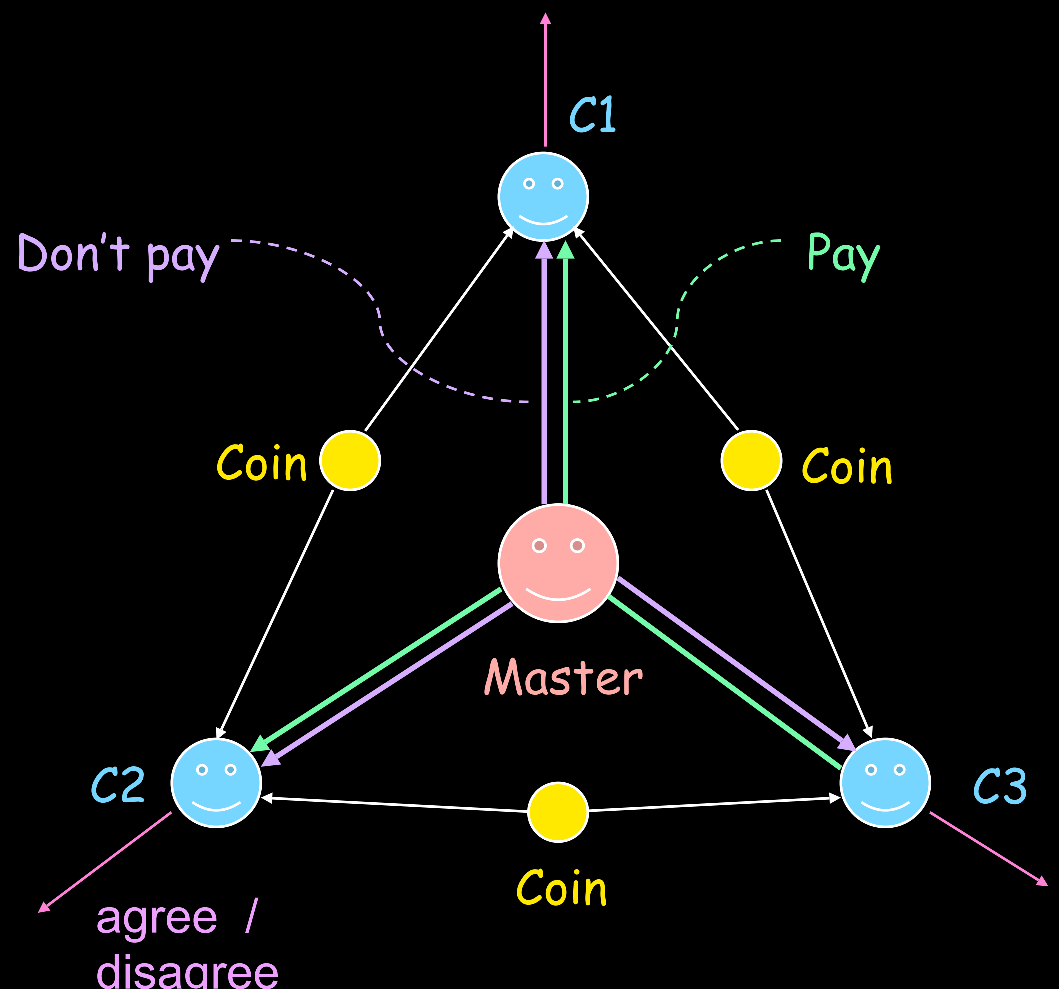
Privacy issues in the modern world

- Issue of privacy protection exacerbated by orders of magnitude:
  - Electronic devices and their continuous interaction with users
    ⇒ possibility to gather and store a huge amount of information
  - Profiling / data mining techniques
    ⇒ precise definition of the individual's preferences
  - Personal information on consumers perceived as asset
    ⇒ often subject matter of commercial transactions

- Result:
  - An enormous amount of information on the individual is gathered, processed, exchanged, used
  - The individual often has not consented to this processing
  - In the worst scenario, he is not even aware of it

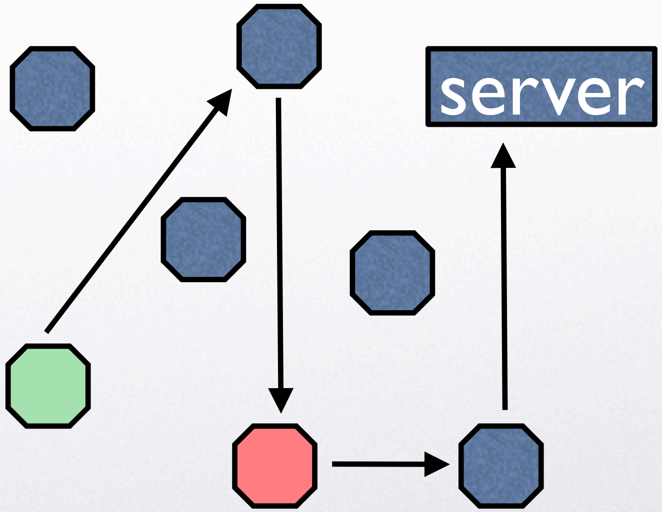

RFID tags may be everywhere..
Faxed hair
model #4456
(polyester)
Artificial leg
model #459382
Bakunin
letters
5 x $100
banknotes
Serial #:
597387,389473...
Lingerie
model Penelope Bras
(Size 4)
Coutesy by: Giuseppe Bianchi

Example: the dining cryptographers



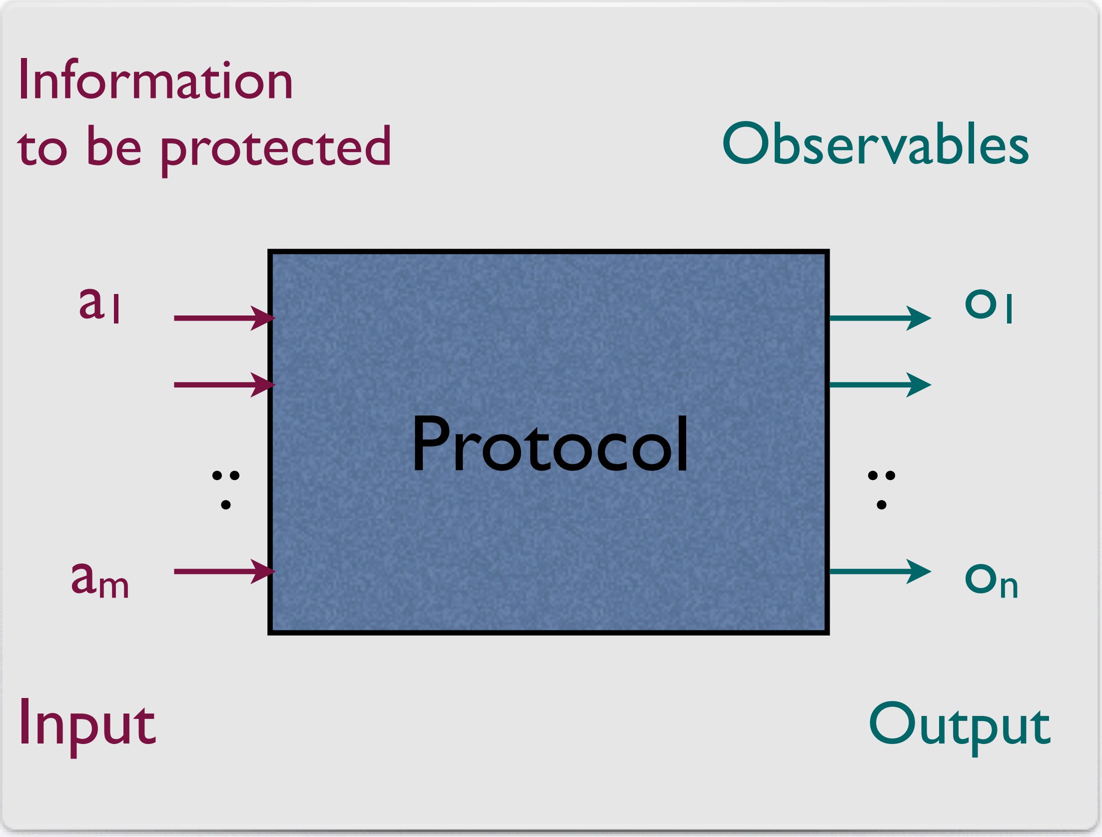
Crowds

- A crowd is a group of n nodes
- The initiator selects randomly a node (called forwarder) and forwards the request to it
- A forwarder:
  - With prob. p_f selects randomly a new node and forwards the request to him
  - With prob. $1-p_f$ sends the request to the server

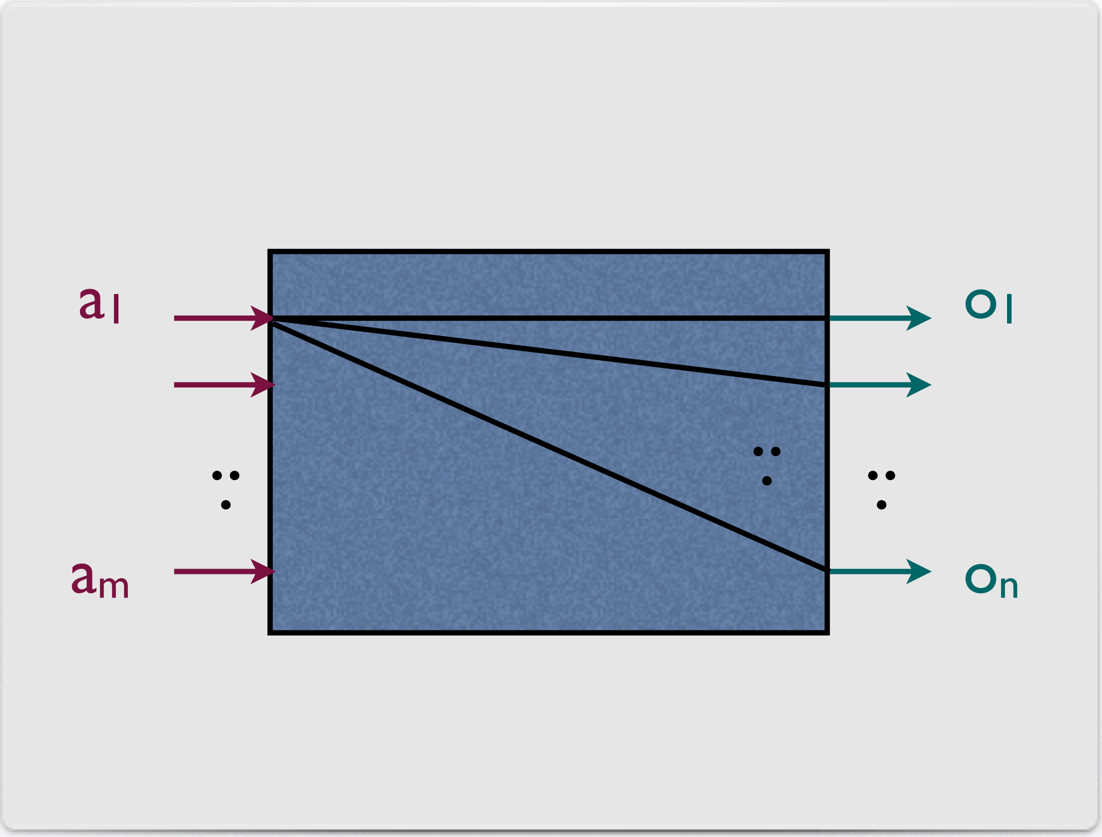


Common features of information-hiding protocols

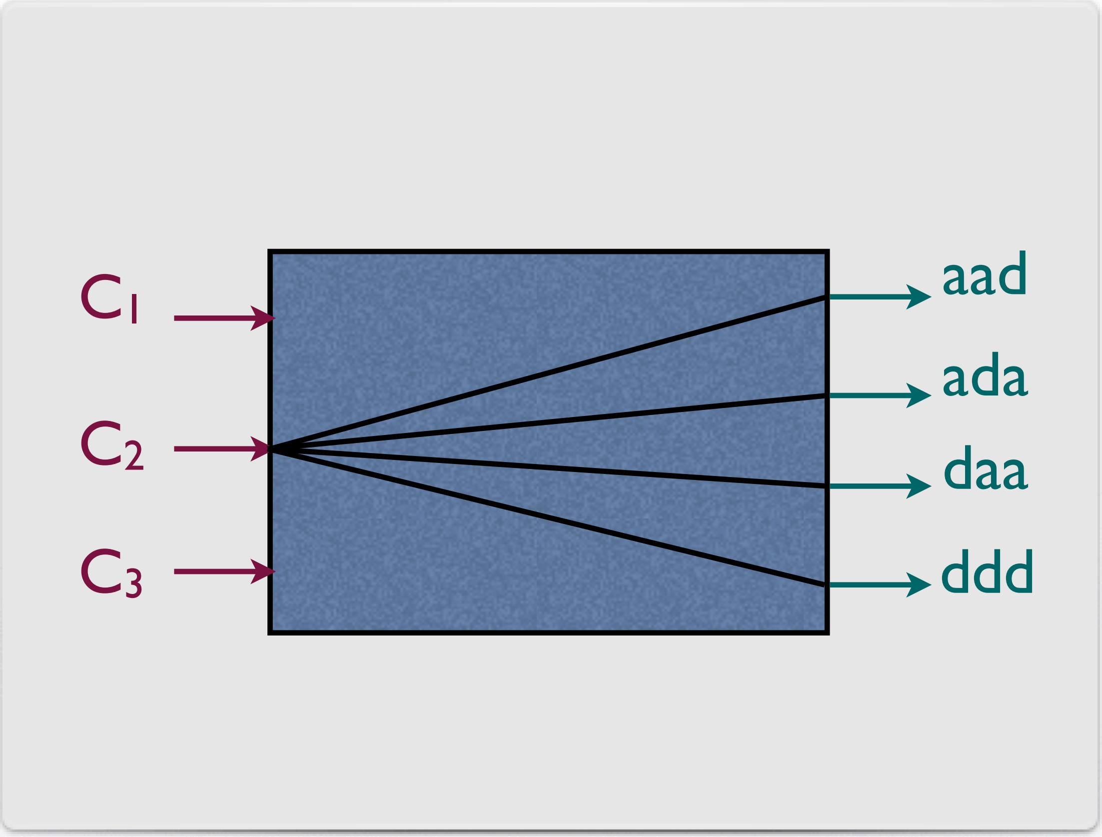
- There is information that we want to keep hidden
  - the user who pays in D.C.
  - the user who initiates the request in Crowds
- There is information that is revealed (observables)
  - agree/disagree in D.C.
  - the users who forward messages to a corrupted user in Crowds
- Protocols often use randomization to hide the link between hidden and observable information
  - coin tossing in D.C.
  - random forwarding to another user in Crowds



Protocols as channels



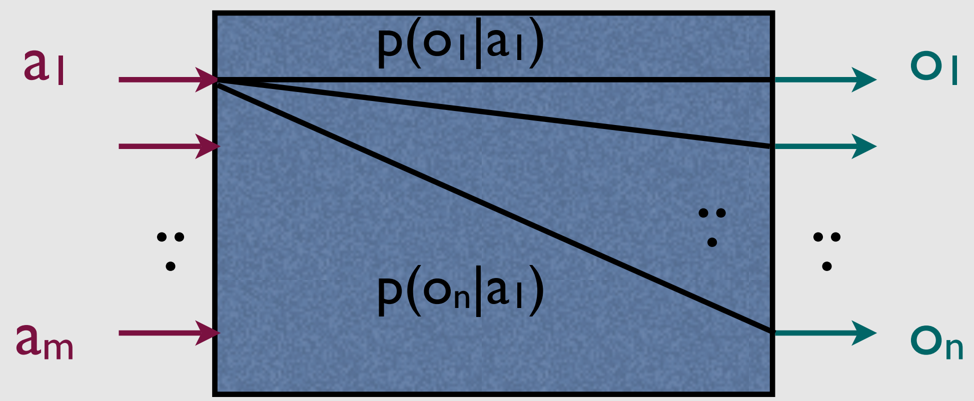
Protocols as noisy channels



Example: The protocol of the dining cryptographers

Assumptions

- We consider probabilistic protocols
  - 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)
- We assume that the protocol at each session receives exactly one input and produces exactly one observable
- We want to define the degree of protection independently from the input's distribution, i.e. the users of the protocol



The conditional probabilities

	o_1	$\cdots$	o_n
a_1	$p(o_1 \mid a_1)$	$\cdots$	$p(o_n \mid a_1)$
$\vdots$	$\vdots$		
a_m	$p(o_1 \mid a_m)$		$p(o_n \mid a_m)$

The channel is completely characterized by the array of conditional probabilities

Possibilistic approaches

- [Schneider and Sidiropoulus], [Halpern and O'Neill]
- Key idea: Replace the random choices by nondeterministic choices
- Common principle: A protocol provides protection iff:
  For every pair of hidden events a, a′, P[a] is "equivalent" to P[a′]
- Criticism: Too weak!

Probabilistic approaches

Notions of total protection in literature

In the following, a, a′ are hidden events, o is an observable

1. [Halpern and O'Neill - like] for all a, a': $p(a|o) = p(a'|o)$
2. [Chaum], [Halpern and O'Neill]: for all a, o: $p(a|o) = p(a)$
3. [Bhargava and Palamidessi]: for all a, a', o: $p(o|a) = p(o|a')$

- Criticism to (1): it depends on the input's distribution rather than on the features of the protocol and it is too strong because it is equivalent equivalent to requiring $p(a) = p(a')$ for all a, a'
- (2) and (3) are equivalent
- These notions are 0-1. 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) does not abstract from the input distribution.
  (3) seems the best to us, but it is controversial

Statistical Inference

- A natural definition of the degree of protection: the ‘probability of error’ (i.e. the probability of guessing wrong) when we try to infer the hidden information from the observables

Statistical inference

- $\boldsymbol{o} = o_1, o_2, \ldots, o_n$: a sequence of n observations
- f : the function used by the adversary to infer the input from a sequence of observations
- Error region of f for input a: $E_f(a) = \{\boldsymbol{o} \in \mathcal{O}^n \mid f(\boldsymbol{o}) \neq a\}$
- Probability of error for input a: $\eta(a) = \sum_{\boldsymbol{o} \in E_f(a)} p(\boldsymbol{o}|a)$
- Probability of error for f :

$$P_{f_n} = \sum_{a \in A} p(a)\eta(a)$$

MAP decision functions

- *MAP: Maximum Aposteriory Probability*
- Applicable when the input's distribution is known.
  Use Bayes theorem:

$$p(a \mid o) = (p(o \mid a)\, p(a)) / p(o)$$

- f is a MAP decision function if $f(o) = a$ implies

$$p(o \mid a)\, p(a) \;>=\; p(o \mid a')\, p(a') \qquad \text{for all } a, a' \text{ and } o$$

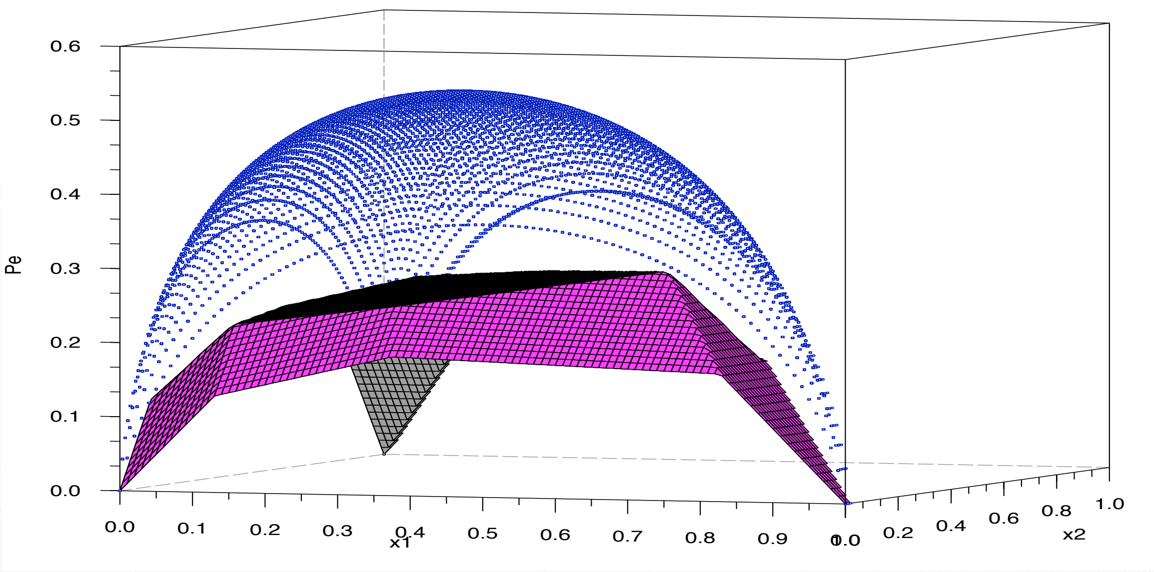
- **Proposition 1:** the MAP decision functions minimize the probability of error (which in this case is called Bayes risk)

Relation with the probabilistic notion of strong anonymity

- Proposition 2:
  the Bayes risk is maximum iff Capacity = 0
  (i.e) iff for all a, a', o, $p(o|a) = p(o|a')$

Bayesian Risk and Information Theory

- Object of study since decades
- Philosophical and practical motivations



- Relation with Conditional Entropy   H(A|O)
- Bounds by Rény ’66, Hellman-Raviv ’70, Santhi-Vardy ‘06
- Tighter bound obtained by studying the ‘corner points’

Independence from the input distribution

- Under a certain condition Cond, for large sequences of observations the input distribution becomes negligible:

- **Proposition 3:** Under Cond, any MAP decision function f can be approximated by a function g s.t.
  $g(o) = a$ implies p(o | a) > p(o | a') for all a, a' and o
  g is called a ML (Maximum Likelihood) function

- "approximated" means that the more observations we make, the smaller is the difference in the probability of error for f and for g

and ... guess what!

- The condition Cond for ML to approximate MAP is the negation (almost) of our old friend:

  Cond: for all a, a′, there exists o: $p(o|a) \neq p(o|a')$

- **Proposition 4:** If Cond holds, then the probability of error under MAP (and ML) converges to 0

- **Proposition 5:** If Cond does not hold, then the probability of error does not converge to 0 under *any* decision function. (Provided that a, a’ have positive probab.)

How to compute the matrix of the channel associated to a protocol

- Express the protocol in your favorite formalism
- Establish the hidden events (inputs) and the observable events (outputs)
- The matrix of the channel (i.e. the conditional probabilities) is completely determined by the protocol and can be computed either by hand or by model checking
- The capacity is completely determined by the matrix and can be approximated by using the Arimoto-Blahut algorithm. In some particular cases is given by a formula

Example: D.C. in the probabilistic asynchronous π-calculus

$$\boxed{\begin{aligned} Master &= \textstyle\sum_{i=0}^{2} \tau \,.\, \overline{m}_i \mathsf{p} \,.\, \overline{m}_{i\oplus 1}\mathsf{n} \,.\, \overline{m}_{i\oplus 2}\mathsf{n} \,.\, 0 \\ &\quad + \;\; \tau . \overline{m}_0 \mathsf{n} \,.\, \overline{m}_1 \mathsf{n} \,.\, \overline{m}_2 \mathsf{n} \,.\, 0 \end{aligned}}$$

Nondeterministic choice

$$\begin{aligned} Crypt_i &= m_i(x) \,.\, c_{i,i}(y) \,.\, c_{i,i\oplus 1}(z) \,. \\ &\qquad \text{if } x = \mathsf{p} \\ &\qquad\quad \text{then } \overline{pay}_i \,.\, \text{if } y = z \\ &\qquad\qquad \text{then } \overline{out}_i\, disagree \\ &\qquad\qquad \text{else } \overline{out}_i\, agree \\ &\qquad\quad \text{else if } y = z \\ &\qquad\qquad \text{then } \overline{out}_i\, agree \\ &\qquad\qquad \text{else } \overline{out}_i\, disagree \end{aligned}$$

Anonymous actions

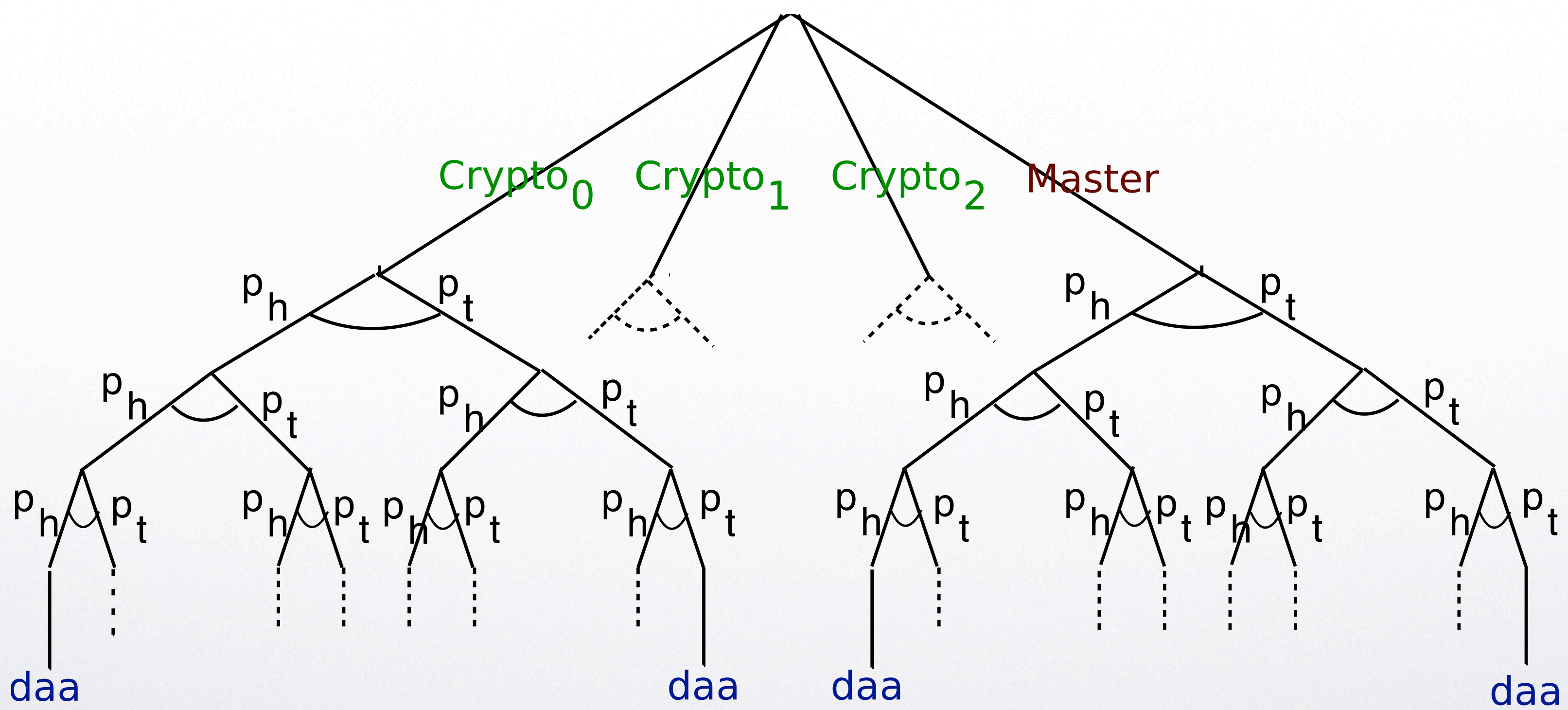
Observables

$$\boxed{Coin_i = p_h \tau \,.\, Head_i + p_t \tau \,.\, Tail_i}$$

Probabilistic choice

$$\begin{aligned} Head_i &= \overline{c}_{i,i} head \,.\, \overline{c}_{i\ominus 1,i} head \,.\, 0 \\ Tail_i &= \overline{c}_{i,i} tail \,.\, \overline{c}_{i\ominus 1,i} tail \,.\, 0 \\ DCP &= (\nu \vec{m})(Master \\ &\qquad \mid \;\; (\nu \vec{c})(\Pi_{i=0}^{2} Crypt_i \;\mid\; \Pi_{i=0}^{2} Coin_i)\;) \end{aligned}$$

Probabilistic automaton associated to the probabilistic π program for the D.C.



Examples of channel matrices

- Dining cryptographers, while varying the probability p of the coins to give heads

  - p = 0.5

	daa	ada	aad	ddd	aaa	dda	dad	add
c_1	1/4	1/4	1/4	1/4	0	0	0	0
c_2	1/4	1/4	1/4	1/4	0	0	0	0
c_3	1/4	1/4	1/4	1/4	0	0	0	0
m	0	0	0	0	1/4	1/4	1/4	1/4

  - p = 0.7

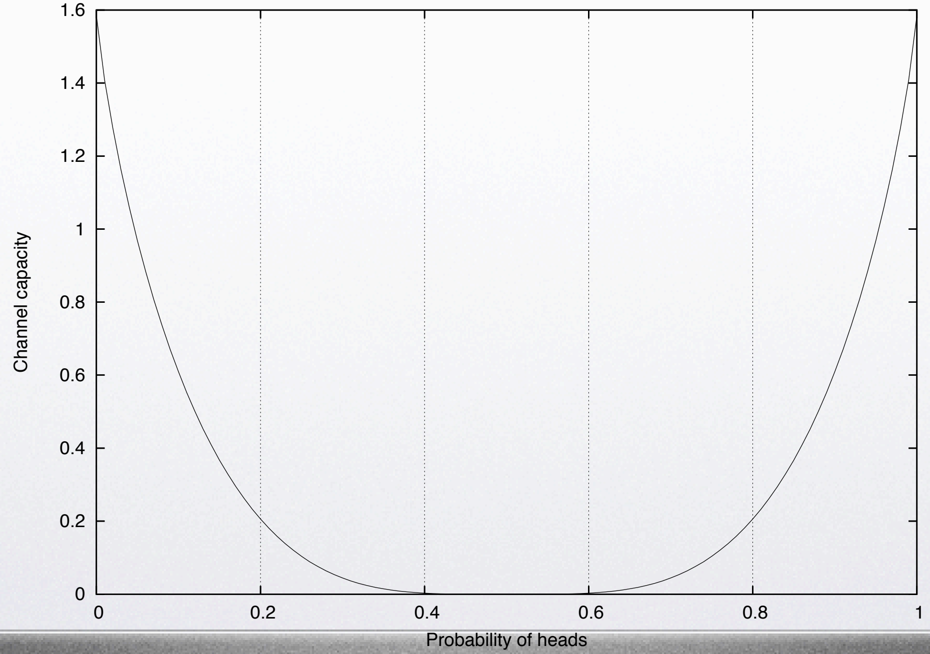
	daa	ada	aad	ddd	aaa	dda	dad	add
c_1	0.37	0.21	0.21	0.21	0	0	0	0
c_2	0.21	0.37	0.21	0.21	0	0	0	0
c_3	0.21	0.21	0.37	0.21	0	0	0	0
m	0	0	0	0	0.37	0.21	0.21	0.21

Computing the capacity from the matrix

- General case: using the Arimoto-Blahut algorithm
  - Approximates the capacity to a given precision
- In particular cases we can exploit the protocol's symmetries
  - Symmetric channel: all rows and all columns are permutations of each other
  - In a symmetric channel: $C = \log|\mathcal{O}| - H(\mathbf{r})$
  - Can be extended to weaker notions of symmetry

Test-case: dining cryptographers

- Fair coins: the protocol is strongly anonymous (C=0)
- Totally biased coins: the payer can be always identified (maximum capacity C = log 3)



Thank you !