Information-hiding Protocols as Opaque Channels

Catuscia Palamidessi

Based on joint work with
Kostas Chatzikokolakis and Prakash Panangaden

Supported by
INRIA/DREI project PRINTEMPS and INRIA/ARC project ProNoBiS

Plan of the talk

- Motivation
- Protocols as channels
- Preliminary notions of Information Theory
- Opacity as converse of channel capacity
- Intended leak of information
- Relation with other notions in literature
- Computing the capacity of the protocol/channel
- Statistical inference and Bayesian risk
- Conclusion and future work

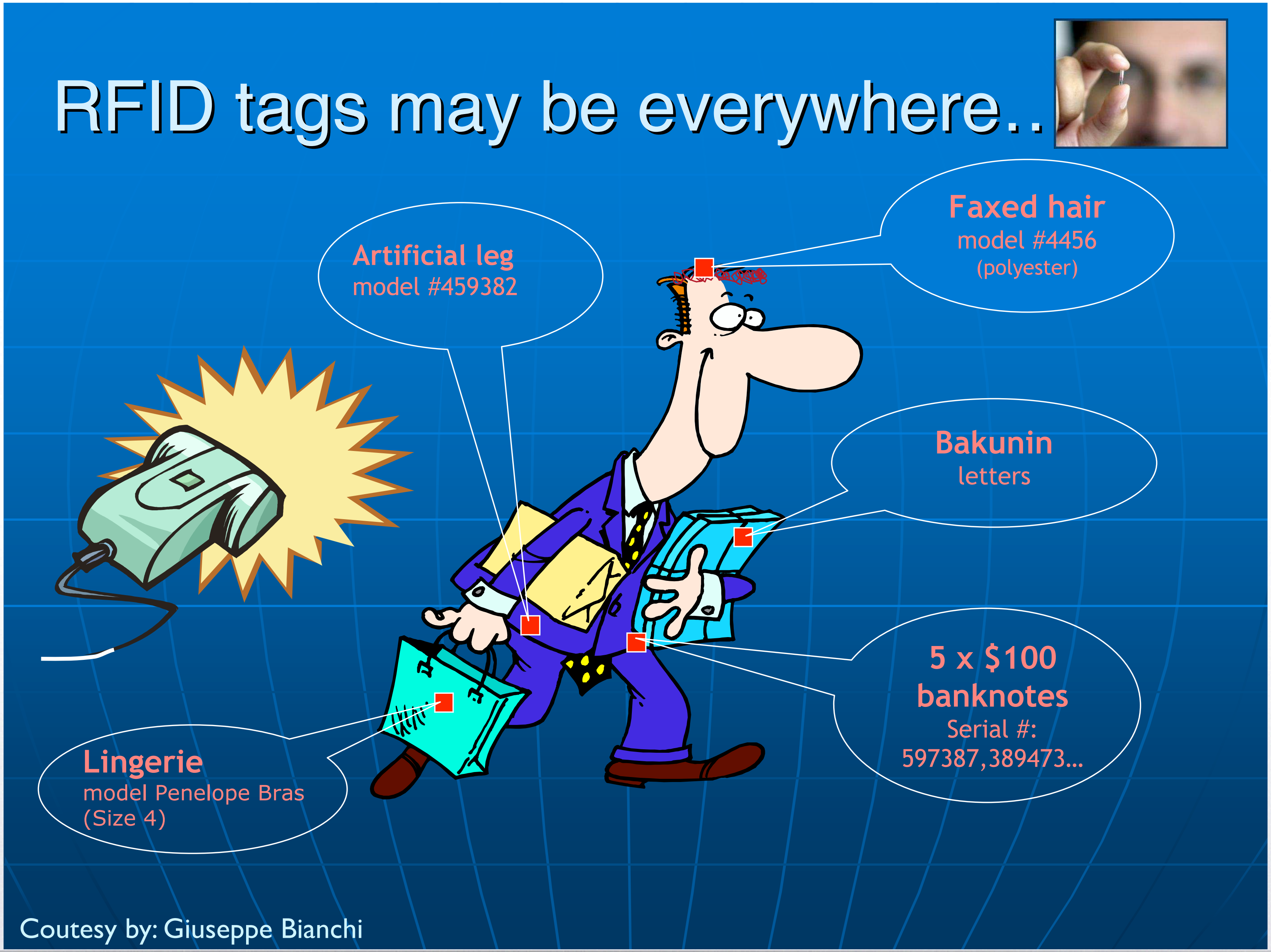
~~Information-hiding~~ Privacy

- 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 in Global/Pervasive Computing

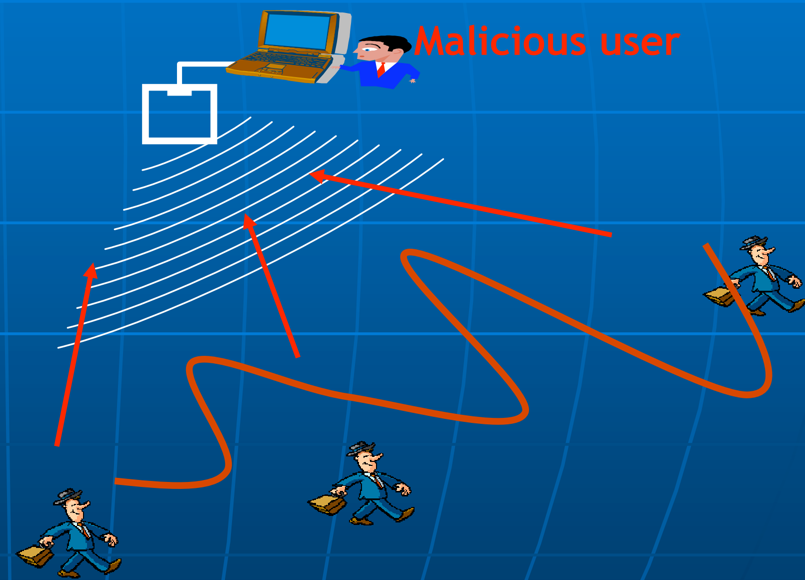
- Issue of privacy protection exacerbated by orders of magnitude:
  - Electronic devices and their continuous interaction with users $\Rightarrow$ possibility to gather and store a huge amount of information
  - Profiling / data mining techniques $\Rightarrow$ precise definition of the individual's preferences
  - Personal information on consumers perceived as asset $\Rightarrow$ often subject matter of commercial transactions
- Result:
  - A tremendous 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..
Artificial leg
model #459382
Faxed hair
model #4456
(polyester)
Bakunin
letters
5 x $100
banknotes
Serial #:
597387,389473...
Lingerie
model Penelope Bras
(Size 4)
Coutesy by: Giuseppe Bianchi

… and at stake

Personal data Gathering

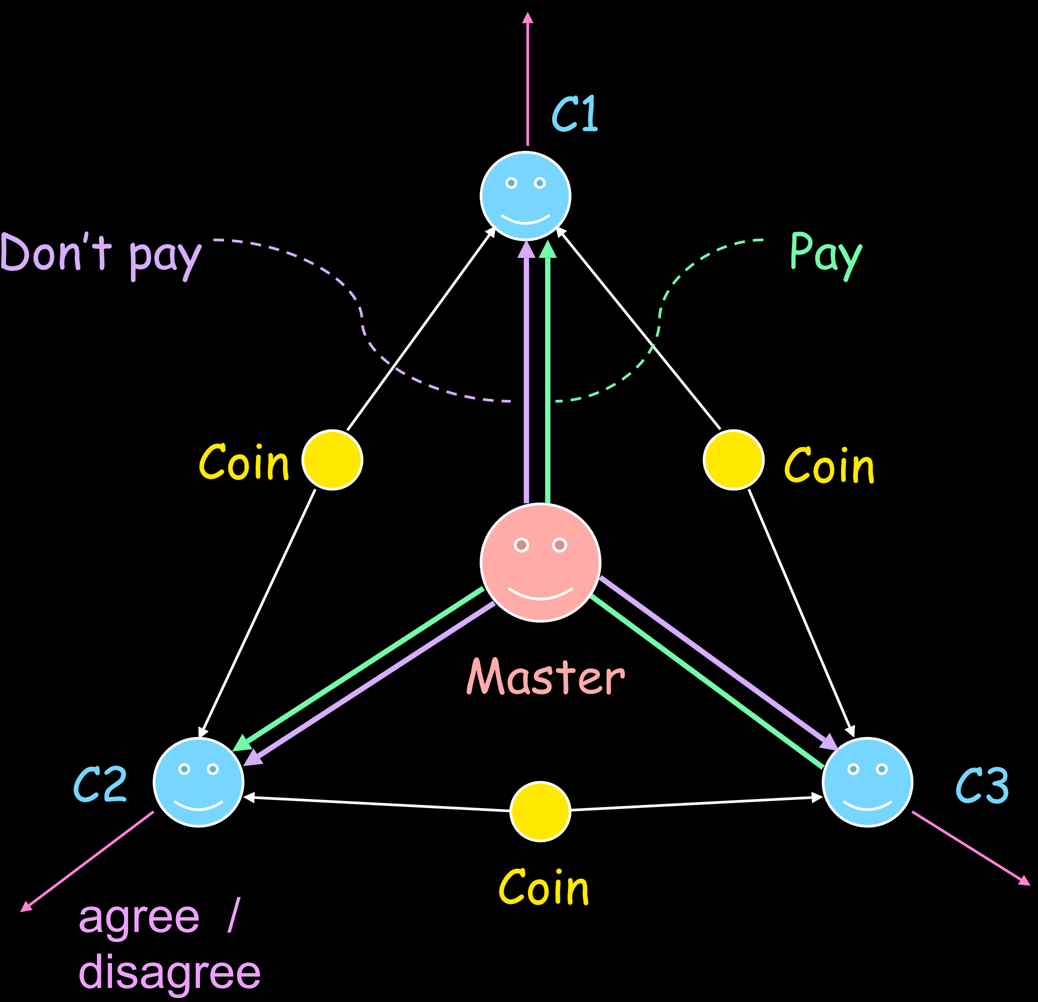
Tracking



Can't kill the RFID tag when dealing with tagged cash...

Courtesy by: Giuseppe Bianchi

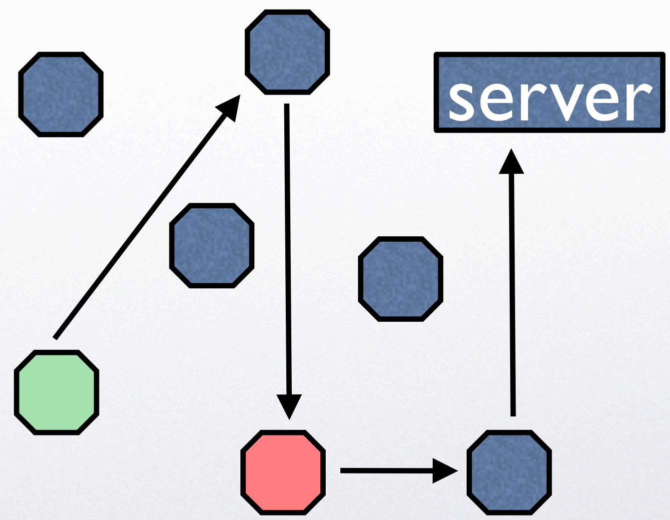
Example: the dining cryptographers



$$
\begin{aligned}
Master &= \textstyle\sum_{i=0}^{2} \tau \,.\, \overline{m}_i \mathsf{p} \,.\, \overline{m}_{i\oplus 1} \mathsf{n} \,.\, \overline{m}_{i\oplus 2} \mathsf{n} \,.\, 0 \\
&\quad + \;\; \tau . \overline{m}_0 \mathsf{n} \,.\, \overline{m}_1 \mathsf{n} \,.\, \overline{m}_2 \mathsf{n} \,.\, 0 \\
Crypt_i &= m_i(x) \,.\, c_{i,i}(y) \,.\, c_{i,i\oplus 1}(z) \,. \\
&\qquad \text{if } x = \mathsf{p} \\
&\qquad\quad \text{then } \overline{pay}_i \,.\, \text{if } y = z \\
&\qquad\qquad \text{then } \overline{out_i}\, disagree \\
&\qquad\qquad \text{else } \overline{out_i}\, agree \\
&\qquad\quad \text{else if } y = z \\
&\qquad\qquad \text{then } \overline{out_i}\, agree \\
&\qquad\qquad \text{else } \overline{out_i}\, disagree \\
Coin_i &= p_h \tau \,.\, Head_i + p_t \tau \,.\, Tail_i \\
Head_i &= \overline{c}_{i,i}\, head \,.\, \overline{c}_{i\ominus 1,i}\, head \,.\, 0 \\
Tail_i &= \overline{c}_{i,i}\, tail \,.\, \overline{c}_{i\ominus 1,i}\, tail \,.\, 0 \\
DCP &= (\nu \vec{m})(Master \\
&\qquad \mid \;\; (\nu \vec{c})(\Pi_{i=0}^{2} Crypt_i \;\mid\; \Pi_{i=0}^{2} Coin_i) \,)
\end{aligned}
$$

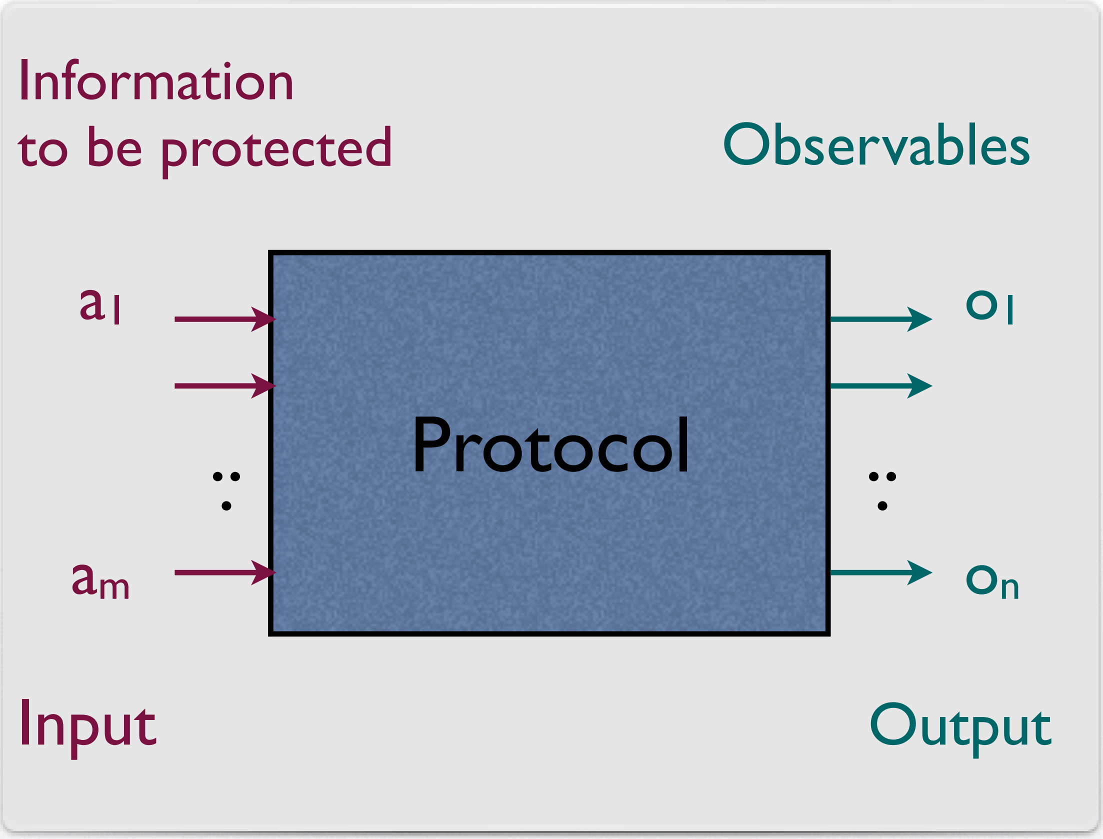
Crowds

- 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. $1\text{-}p_f$ selects randomly a new node and forwards the request to him
  - With prob. p_f sends the request to the server

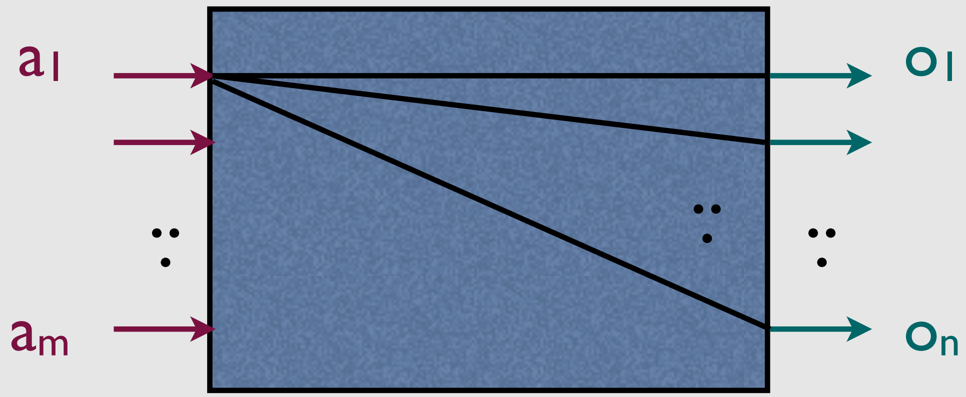


Common features of information-hiding protocols

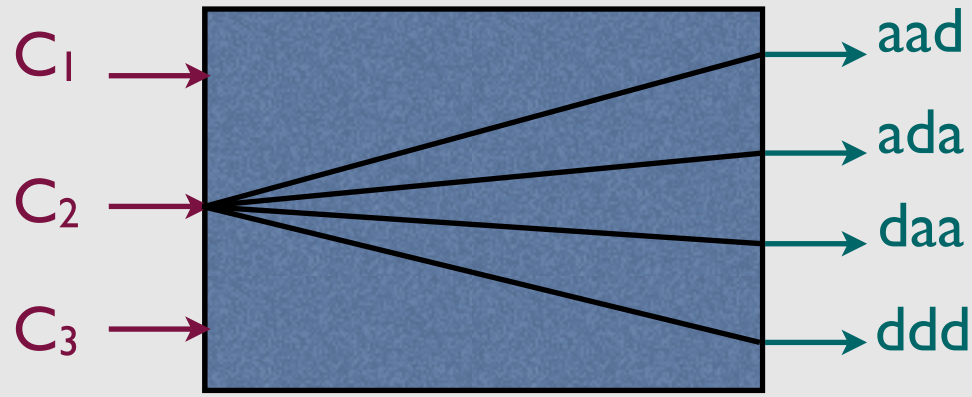
- There is information that we want to keep hidden
  - the user who pays in D.C.
  - the user who initiates the request in Crowds
- There is information that is revealed
  - agree/disagree in D.C.
  - the users who forward messages to a corrupted user in Crowds
- Protocols often use randomization to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



Protocols as channels



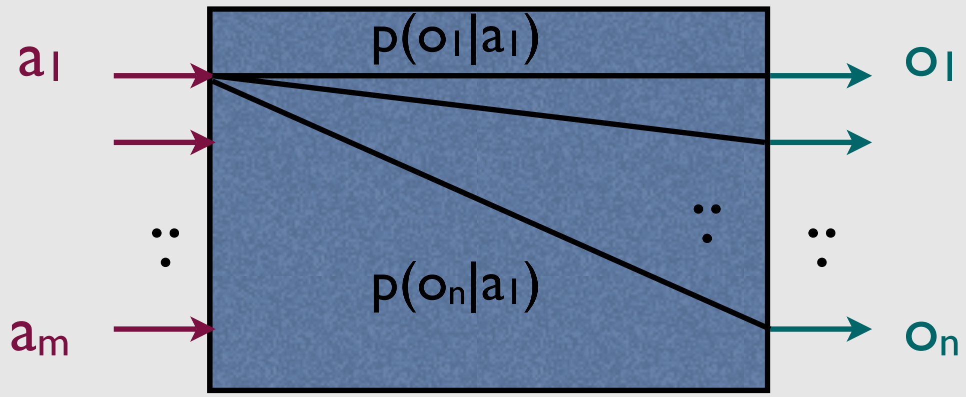
Protocols as noisy channels



The protocol of the dining cryptographers

Protocols as noisy channels

- We consider a probabilistic approach
  - Inputs: elements of a random variable A
  - Outputs: elements of a random variable O
  - For each input a_i, the probability that we obtain an observable o_j is given by $p(o_j \mid a_i)$
- We assume that the protocol receives exactly one input at each session
- We want to define the degree of protection independently from the input's distribution, i.e. the users



The conditional probabilities

	o_1	...	o_n
a_1	$p(o_1 \mid a_1)$	...	$p(o_n \mid a_1)$
$\vdots$	$\vdots$		
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

Preliminaries of Information Theory

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

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

Opacity

- Necessity to give a quantitative measure of the degree of protection provided by a protocol
- We define Opacity as the converse of the Capacity of the channel:

$$C = \max_{p(a)} I(A;O)$$

- Note that this definition is independent from the distribution on the inputs, as desired

Relative privacy

- Some information about A may be revealed intentionally
- Example: elections

- We model the revealed information with a third random variable R

R = number of users who voted for c

Relative privacy

- We use the notion of conditional mutual information

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

- And define the conditional capacity similarly

$$C_R = \max_{p(a)} I(A;O|R)$$

Partitions: a special case of relative privacy

- We say that R partitions $\mathcal{X}$ iff $p(r|x)$ is either 0 or 1 for every r, x
- Examples: elections, group anonymity

Theorem

If R partitions $\mathcal{A}$ and $\mathcal{O}$ then the transition matrix of the protocol is of the form

$$\begin{array}{c|cccc} & \mathcal{O}_1 & \mathcal{O}_2 & \dots & \mathcal{O}_l \\ \hline \mathcal{A}_1 & M_1 & 0 & \dots & 0 \\ \mathcal{A}_2 & 0 & M_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_l & 0 & 0 & \dots & M_l \end{array}$$

and

$$C_R \leq d \quad \Leftrightarrow \quad C_i \leq d, \forall i \in 1..l$$

where C_i is the capacity of matrix M_i.

Relation with existing notions

Strong probabilistic anonymity

$p(a) = p(a|o) \quad \forall a, o$ — [Chaum, 88], aka "conditional anonymity" [Halpern and O'Neill, 03].

$p(o|a_i) = p(o|a_j) \quad \forall o, i, j$ — [Bhargava and Palamidessi, 05]

Proposition

An anonymity protocol satisfies strong probabilistic anonymity iff $C = 0$.

Example: Dining cryptographers

	100	010	001	111
a_1	1/4	1/4	1/4	1/4
a_2	1/4	1/4	1/4	1/4
a_3	1/4	1/4	1/4	1/4

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

- Express the protocol in your favorite formalism
- Establish the anonymous 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

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

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

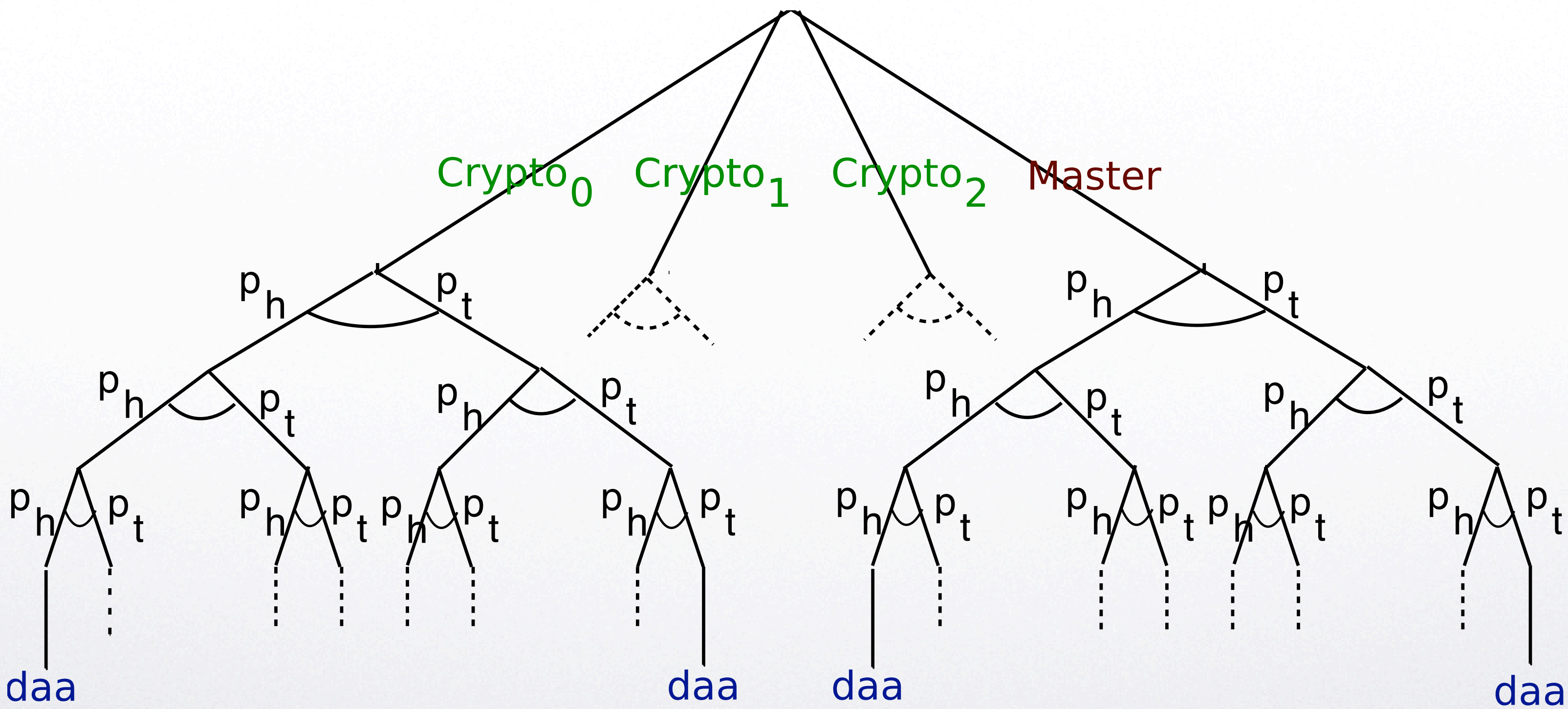
Observables

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

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

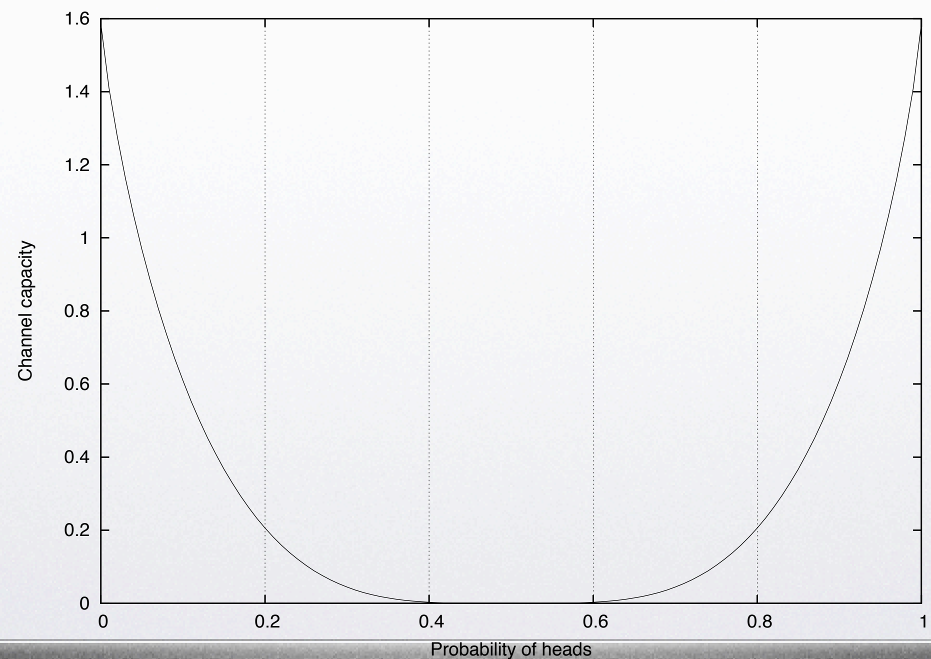
	daa	ada	aad	ddd	aaa	dda	dad	add
c_1	0.37	0.21	0.21	0.21	0	0	0	0
c_2	0.21	0.37	0.21	0.21	0	0	0	0
c_3	0.21	0.21	0.37	0.21	0	0	0	0
m	0	0	0	0	0.37	0.21	0.21	0.21

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)



Privacy and Statistical Inference

- Opacity as converse of Capacity.
  Ok, it seems 'reasonable'.
  But is it the most natural notion?

- An uncontroversially natural notion is be the 'probability of error' of an adversary trying to infer the hidden information (input) from the observables (output)

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

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

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

Independence from the input distribution

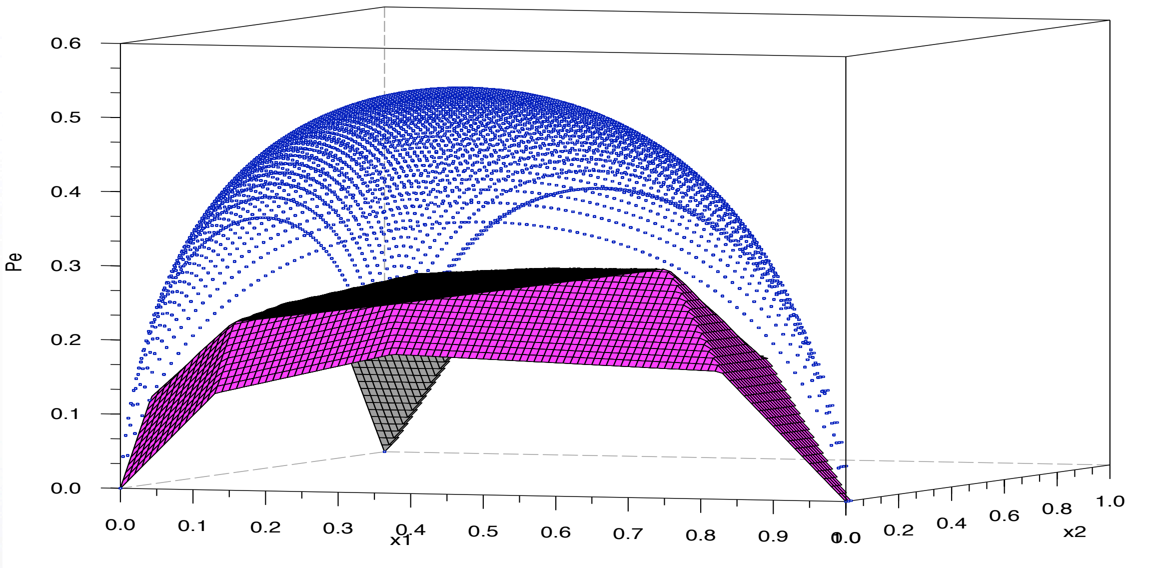
- Under certain conditions, for large sequences of observations the input distribution becomes negligible:
- **Proposition:** A MAP decision function f can be approximated by a function g such that $g(o) = a$ implies

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

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

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’

What about the relation between the Probability of error and Capacity ?

- $p(a|o)$ vs $H(A|O)$

- $p(a|o) / p(a)$ vs $H(A|O) - H(A)$?

Future work

- Explore more in depth the relation between the capability of inferring info about the input and the capacity, or other quantitative notions depending on the channel's matrix.
- Inference of the input distribution without the power of forcing the input to remain the same through the observations
- Characterizations of other (weaker) notions of privacy which are easy to model check, in the sense that they do not require to analyze the capacity as a function of the input distribution
- Develop a logic for efficient model checking

Thank you !