Probabilistic and Nondeterministic Aspects of Anonymity

Catuscia Palamidessi, INRIA & LIX

Based on joint work with:

Mohit Bhargava, IIT New Delhi

Plan of the talk

- The concept of Anonymity
- Example: the Dining Cryptographers
- The Nondeterministic approach of Schneider and Sidiropoulos
- Motivations for considering the probabilistic aspects
- The hierarchy of Reiter and Rubin:
  - Strong Anonymity and weaker notions
- Formalization of Strong Probabilistic Anonymity
- Conclusion and future work

The concept of anonymity

- **Goal:**
  - To ensure that the identity of the agent performing a certain action remains secret.
- Examples of situations in which anonymity may be desirable:
  - Electronic elections
  - File sharing
  - Donations
  - ...
- Some systems:
  - Crowds [Reiter and Rubin,1998],
    - anonymous communication (anonymity of the sender)
  - Onion Routing [Syverson, Goldschlag and Reed, 1997]
    - anonymous communication
  - Freenet [Clarke et al. 2001]
    - anonymous information storage and retrieval

Formal approaches to Anonymity

- Concurrency Theory (CSP)
  - Schneider and Sidiropoulos, 1996
- Epistemic Logic
  - Sylverson and Stubblebine, 1999
  - Halpern and O'Neil, 2004
- Function views
  - Hughes and Shmatikov, 2004

- *All these approaches are either purely nondeterministic or purely probabilistic*
- **However, most anonymity protocols, including Crowds, Onion Routing, and Freenet, have both:**
  - **probabilistic aspects:** randomized primitives.
  - **nondeterministic aspects:** users, scheduler, other unknown factors.

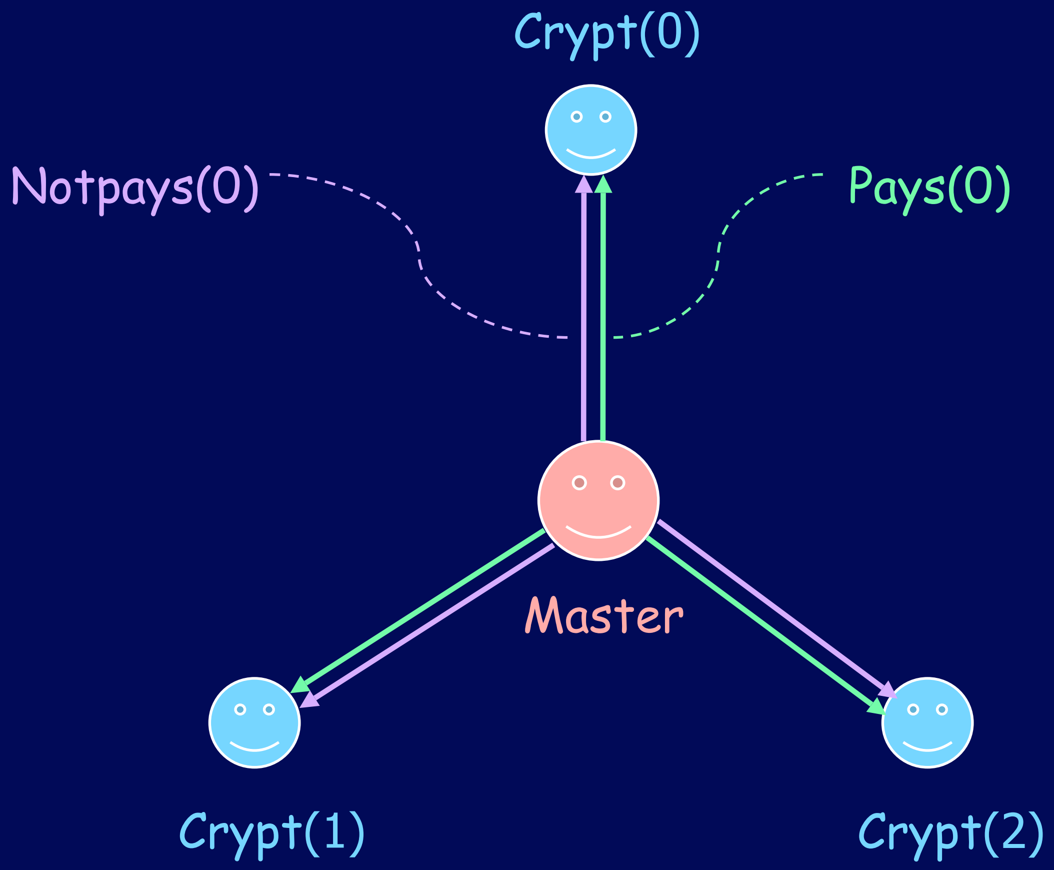
Example: The dining cryptographers

- **Problem formulated originally by David Chaum, 1988**

- **The Problem:**
  - Three cryptographers share a meal
  - The meal is paid either by the organization (master) or by one of them. The master decides who pays
  - Each of the cryptographers is informed by the master whether or not he is has to pay

- **GOAL:**
  - The cryptographers would like to make known whether the meal is being paid by the master or by one of them, but without knowing who among them, if any, is paying. They cannot involve the master

The dining cryptographers



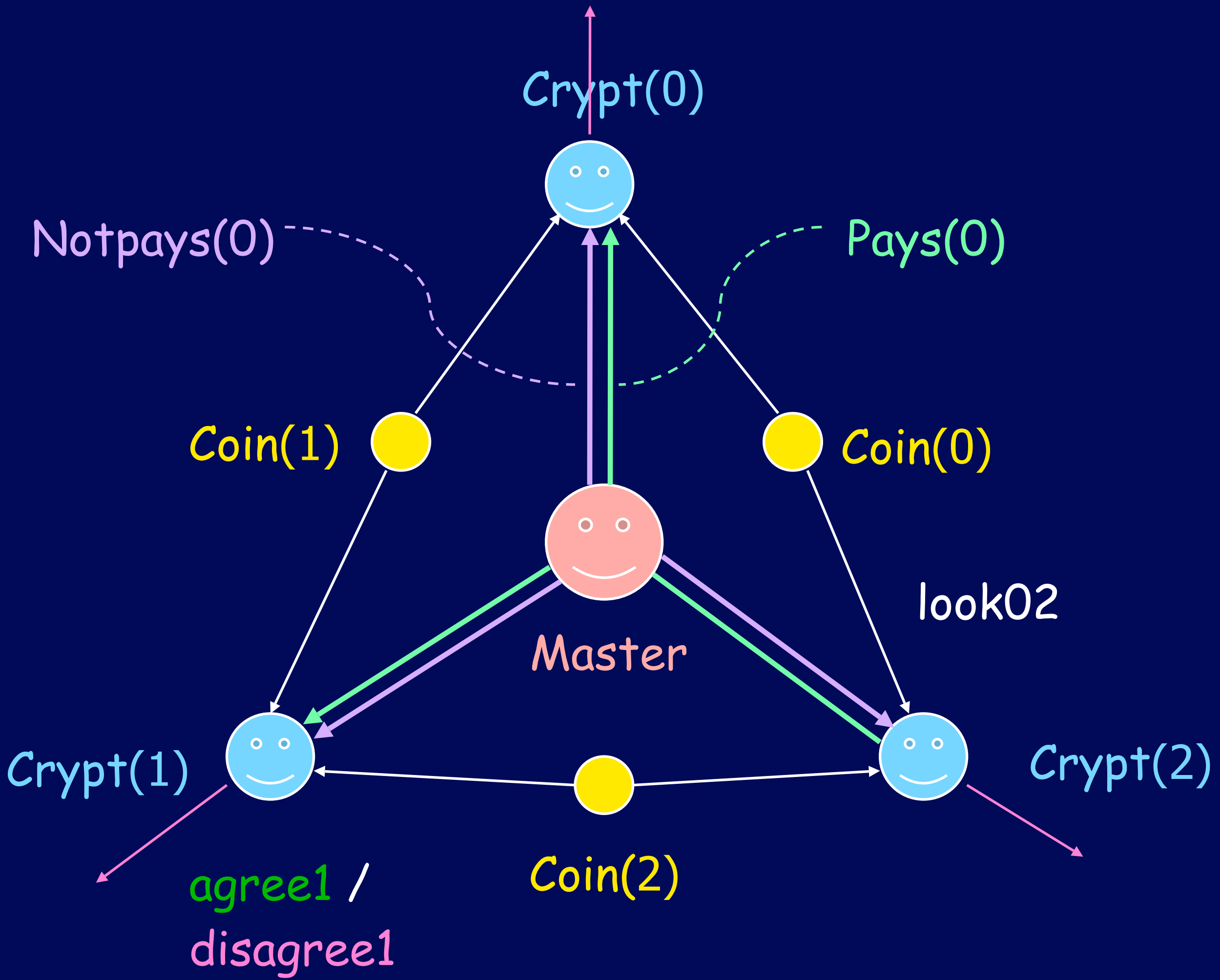
The dining cryptographers A solution (Chaum 1988)

- We insert a coin between each pair of cryptographers and we toss it.

- The result of each coin-tossing is visible to the adjacent cryptographers, and only to them.

- Each cryptographer examines the two adjacent coins
  - If he is not paying, he announces "agree" if the results are the same, and "disagree" otherwise.
  - If he is paying, he says the opposite

The dining cryptographers



The Dining Cryptographers
Properties of the solution

Proposition 1: if the number of “disagree” is even, then the master is paying. Otherwise, one of them is paying.

Proposition 2 (Anonymity): In the latter case, if the coins are fair (i.e. they give Head and Tail with the same probability) then an external observer (and the non paying cryptographers) will not be able to deduce who is paying

Description of the D.C. using a process calculus

- The D.C. is naturally both nondeterministic (the master) and probabilistic (the coins).

- Special cases:
  - The fully nondeterministic approximation, where coins are nondeterministic [Schneider and Sidiropoulus, 1996]
  - The fully probabilistic variant, where the master is probabilistic
    - with a uniform distribution, or
    - with an arbitrary distribution

- In order to describe the anonymous systems, and to formalize the property of anonymity, we use a process calculus which allows to express both probabilistic and nondeterministic choices.
  - There are many proposals in literature. We use the probabilistic asynchronous π-calculus [Herescu & Palamidessi, 2000]

D.C. in the probabilistic asynchronous π-calculus

$$\boxed{\begin{aligned} \mathit{Master} &= \textstyle\sum_{i=0}^{2} \tau \,.\, \overline{m_i}\mathsf{p} \,.\, \overline{m_{i\oplus 1}}\mathsf{n} \,.\, \overline{m_{i\oplus 2}}\mathsf{n} \,.\, 0 \\ &\quad + \;\; \tau.\overline{m_0}\mathsf{n} \,.\, \overline{m_1}\mathsf{n} \,.\, \overline{m_2}\mathsf{n} \,.\, 0 \end{aligned}}$$

Nondeterministic choice

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

Anonymous actions

Observables

$$\boxed{\mathit{Coin}_i = p_h \tau \,.\, \mathit{Head}_i + p_t \tau \,.\, \mathit{Tail}_i}$$

Probabilistic choice

$$\begin{aligned} \mathit{Head}_i &= \overline{c}_{i,i}\mathit{head} \,.\, \overline{c}_{i\ominus 1,i}\mathit{head} \,.\, 0 \\ \mathit{Tail}_i &= \overline{c}_{i,i}\mathit{tail} \,.\, \overline{c}_{i\ominus 1,i}\mathit{tail} \,.\, 0 \\ \mathit{DCP} &= (\nu \vec{m})(\mathit{Master} \\ &\qquad \mid \;\; (\nu \vec{c})(\mathit{\Pi}_{i=0}^{2}\mathit{Crypt}_i \;\mid\; \mathit{\Pi}_{i=0}^{2}\mathit{Coin}_i)\;) \end{aligned}$$

The fully nondeterministic variant

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

Nondeterministic choice

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

Anonymous actions

Observables

$$\mathit{Coin}_i = p_h \tau \,.\, \mathit{Head}_i + p_t \tau \,.\, \mathit{Tail}_i$$

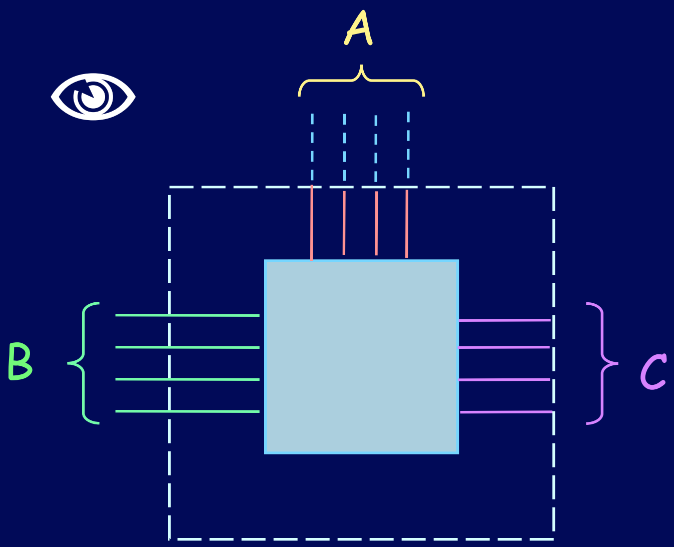
Nondeterministic choice

Probabilistic choice

$$\begin{aligned} \mathit{Head}_i &= \overline{c}_{i,i} \mathit{head} \,.\, \overline{c}_{i\ominus 1,i} \mathit{head} \,.\, 0 \\ \mathit{Tail}_i &= \overline{c}_{i,i} \mathit{tail} \,.\, \overline{c}_{i\ominus 1,i} \mathit{tail} \,.\, 0 \\ \mathit{DCP} &= (\nu \vec{m})(\mathit{Master} \\ &\qquad \mid \; (\nu \vec{c})(\mathit{\Pi}_{i=0}^{2} \mathit{Crypt}_i \mid \mathit{\Pi}_{i=0}^{2} \mathit{Coin}_i) \;) \end{aligned}$$

The purely nondeterministic approach by Schneider and Sidiropoulus

- Anonymity is defined w.r.t. the following partition on Actions:
  - A = { $a(i)$ | $i \in$ Anonymous Agents } : the anonymous actions
  - B = the actions that are visible to the observers
  - C = Actions – (B U A) : The actions we want to hide



Consider the traces on B U A.

Definition: The system P is anonymous if its set of traces is invariant w.r.t. any permutation ρ of the actions in A, namely

$$\rho(\text{Traces}(P)) = \text{Traces}(P) \quad \text{for any } \rho$$

The nondeterministic version of the D.C. satisfies this property

Treating the probabilistic aspects faithfully
Motivations

1. An observer may deduce probabilistic info about the system by making statistical observations
   - This possible leakage of probabilistic info is not captured by the nondeterministic formulation

2. With a probabilistic formulation one can distinguish different levels of strength.

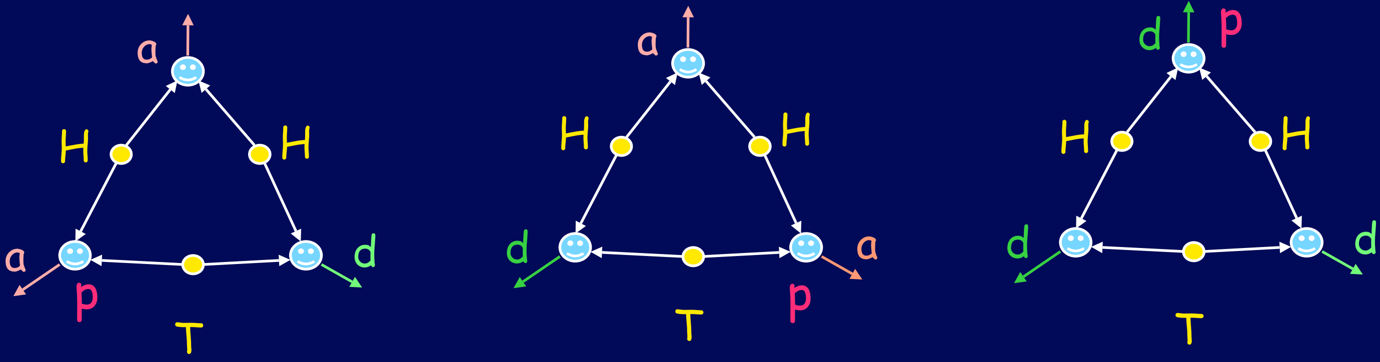
For instance: The (informal) hierarchy of Reiter and Rubin

- **Beyond suspicion**: To the observer, the culprit is not more likely (to be the culprit) than any other agent
- **Probable innocence**: the culprit is less likely than all the other agents together
- **Possible innocence**: the observer cannot be sure that the culprit is indeed the culprit

- The nondeterministic approach corresponds to the lowest level of the hierarchy

Leakage of probabilistic information

- Example. Suppose that in the DC with probabilistic coins we observe with high frequency only the following results



These are 3 of the 4 possible configurations when the payer is a cryptographer

We can deduce that the coins are biased, and how

Therefore we can probabilistically guess who is the payer

This breach in anonymity is not detected by the nondeterministic approach (as long as the fourth configuration is possible).

Formalization of Strong Probabilistic Anonymity

- The rest of this talk is dedicated to formalizing the notion of "beyond suspicion" (strong probabilistic anonymity)

- We want a notion which captures the probabilistic aspects of the protocol, and in which the choices of the users may be either probabilistic or nondeterministic

- **Users-independence:** in case the choices of the users are probabilistic, the definition should be independent from their probability distribution

- Note: in the D.C. example, the choices of the users are represented by the master

Formalization of Strong Probabilistic Anonymity: Notation

- Conditional probability: $p(x \mid y) = p(x \text{ and } y) / p(y)$

- Events:
  - $a(i)$: user i has performed anonymous action a
  - $a = \bigcup_i a(i)$: anonymous action a has been performed
  - $o = b_1 \ldots b_n$: observable actions $b_1, \ldots , b_n$ have been performed

Formalization of S.P.A.: the notion of evidence

- We propose to interpret the notion of "being likely to be the culprit" (in the informal definition of Reiter and Rubin) in terms of the notion of **evidence**

- Notion of evidence:
  - Given a set of exhaustive and mutually exclusive hypotheses $h_1,\dots,h_n$, and an event o, what is the evidence, given o, that h_i holds ?
  - Example: given a coin which is totally biased (p(H) = 1) or fair (p(H) = p(T) = 1/2), and given the event H, what's the evidence that the coin is fair?

The notion of evidence

- **Probabilistic case** - the hypotheses are chosen probabilistically

$$evidence(h_i, o) = p(o|h_i)$$

- **Nondeterministic case**

$$evidence(h_i, o) = \frac{p_{h_i}(o)}{\sum_j p_{h_j}(o)}$$

- Note that the nondeterministic case corresponds to the probabilistic case with uniform distribution

- Relation between evidence and statistics

Strong probabilistic anonymity general definition

- We will say that a system is strongly anonymous iff

For every observable o, for every users i and j,
the evidence that i is the culprit, given o,
is the same as
the evidence that j is the culprit

Strong probabilistic anonymity for probabilistic users

- The definition corresponds to

$$\forall\, i, j, o. \quad p(o \mid a(i)\,) = p(o \mid a(j))$$

- **Properties:**
  - it is satisfied by the D. C. with fair probabilistic coins and probabilistic users
  - it does not depend on the probability distribution of the a(i)'s
  - If $\forall$ o, either $o \Rightarrow a$ or $o \Rightarrow \text{not } a$, then it is equivalent to

$$\forall\, i, o\,.\ \text{ if } \ o \Rightarrow a \ \text{ then } \ p(a(i) \mid o) = p(a(i) \mid a) \qquad (2)$$

known as **conditional anonymity**

Strong probabilistic anonymity for nondeterministic users

- The definition can be equivalently rewritten as

$$\forall\, i, j, o. \quad p_i\,(o) = p_j\,(o)$$

- it is satisfied by the D. C. with fair probabilistic coins and nondeterministic users

Conclusion

Definition of Strong Probabilistic Anonymity for the case of single culprit

- Probabilistic users:
  - independence from probability of users
  - equivalent to conditional anonymity
- Nondeterministic users:
  - naturally corresponds to the definition in the probabilistic case

Future work

- Generalization to the case of multiple culprits
  - Example of application: anonymous elections
  - Note that in case of multiple culprits, in general
    - neither our notion (1), nor conditional anonymity (2), are user-independent
    - (1) and (2) are not equivalent
- Extend the study to weaker notions of probabilistic anonymity
  - Applications to other (real) anonymity protocols
- Extend the study to other notions of information-hiding
- Definition of a suitable logic
  - quantitative aspects
  - a form of implication corresponding to conditional probability
- Automatic verification (model checking)

Thank you !