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
- The Concurrency Theory approach
- 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
- Comparison with other (purely) probabilistic approaches to anonymity

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.

The concurrency approach

- We will focus on the "Concurrency theory" approach, which was first proposed by Schneider and Sidiropoulos, 1996

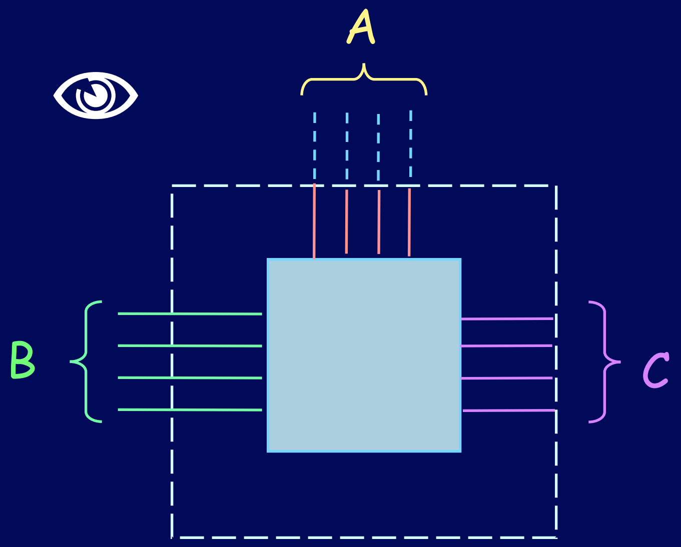
- Agents and protocols are described as processes
- Actions for which we want anonymity of the agent are modeled as consisting of two components:
  - the action itself, a,
  - the identity of the agent performing the action, i

$$a(i)$$

- Anonymous Agents: the agents who want to remain secret
- Anonymous Actions $A = \{ a(i) \mid i \in \text{Anonymous Agents} \}$

The concurrency approach

- Anonymity is defined w.r.t. the following partition on Actions:
  - $A = \{ a(i) \mid i \in$ 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



The definition of nondeterministic anonymity by Schneider and Sidiropoulus:

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

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

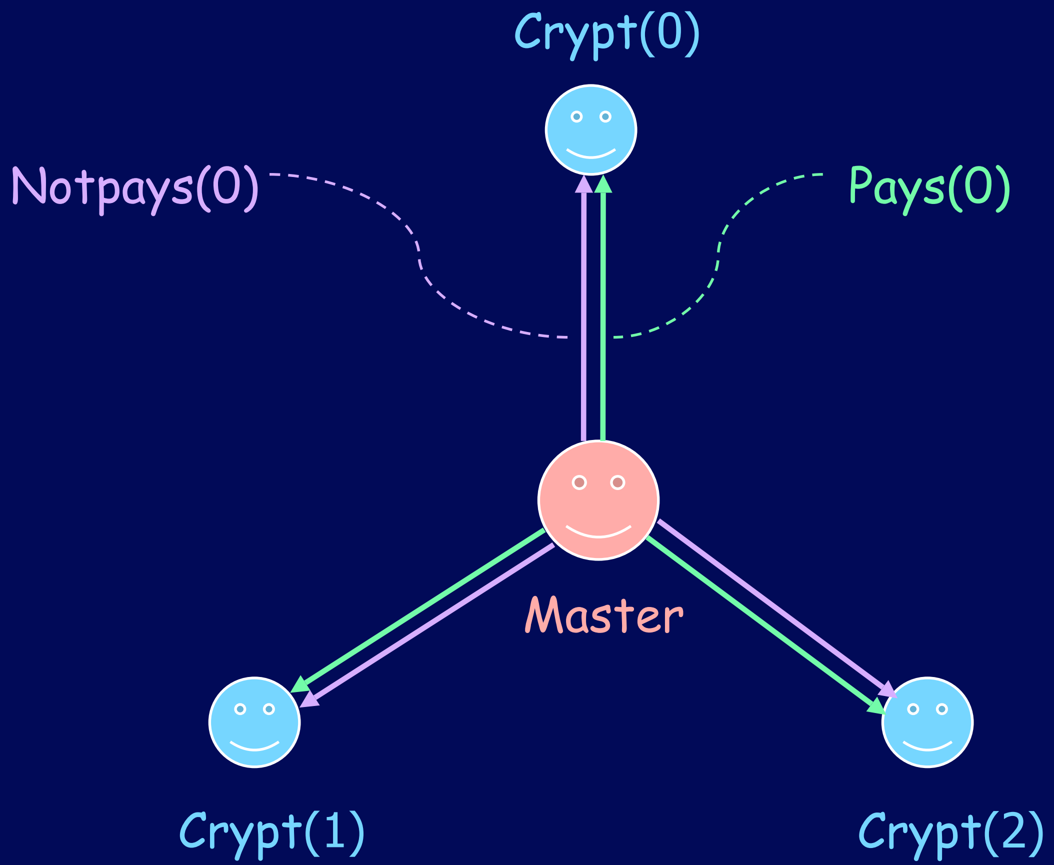
Example: The dining cryptographers

- **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



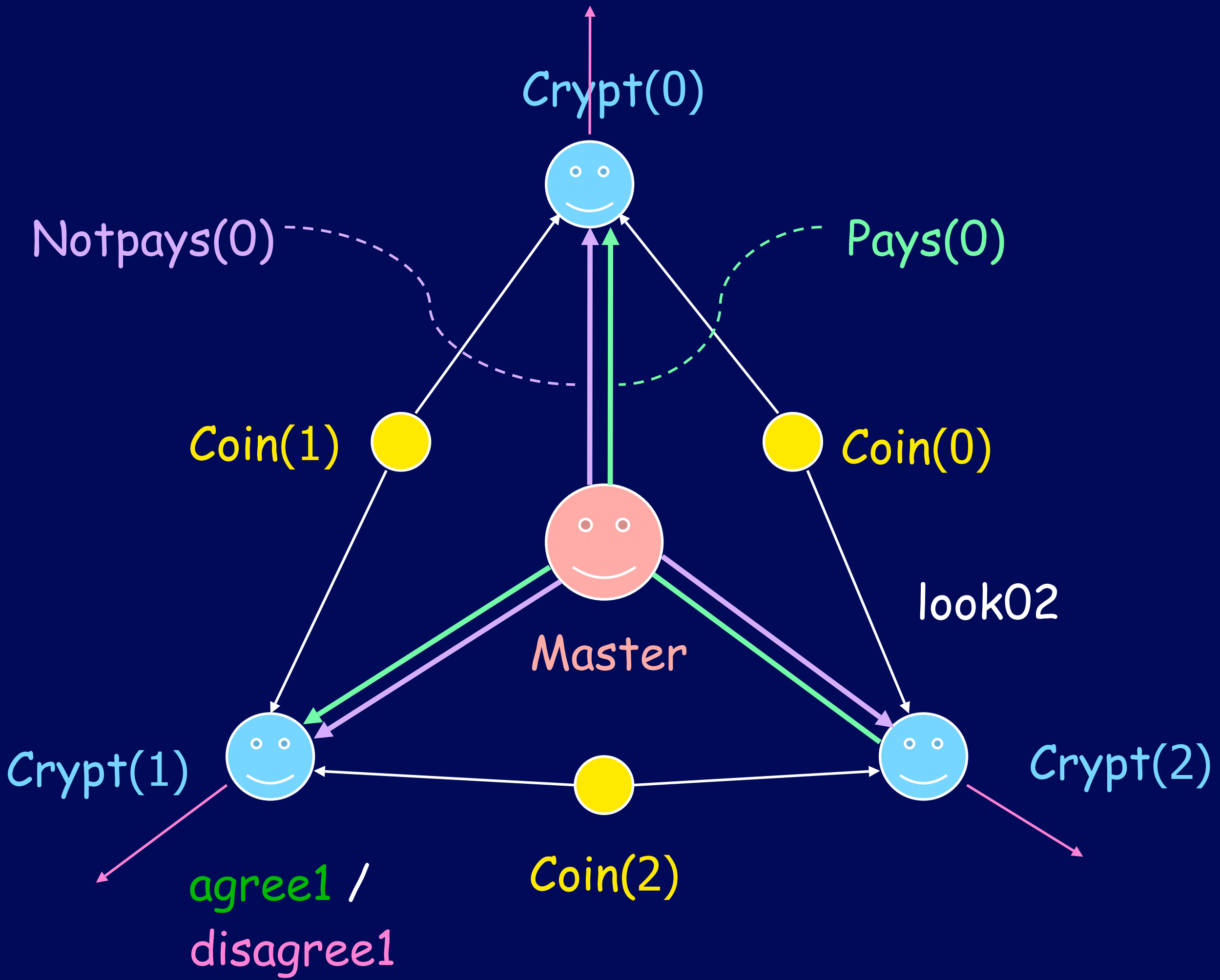
The dining cryptographers
A solution (Chaum 1988)

- 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



Crypt(0)

Notpays(0)

Pays(0)

Coin(1)

Coin(0)

look02

Master

Crypt(1)

Crypt(2)

agree1 / disagree1

Coin(2)

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). It can be described using a process calculus which allows to express both probabilistic and nondeterministic choices.
  - There are many proposals in literature. We will use the probabilistic asynchronous π-calculus [Herescu & Palamidessi, 2000]

- 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

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

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

The fully nondeterministic variant

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

Nondeterministic choice

```
$Crypt_i = m_i(x) \,.\, c_{i,i}(y) \,.\, c_{i,i\oplus 1}(z) \,.$
        if $x = \mathsf{p}$
            then $\overline{pay}_i$ . if $y = z$
                then $\overline{out}_i\, disagree$
                else $\overline{out}_i\, agree$
            else if $y = z$
                then $\overline{out}_i\, agree$
                else $\overline{out}_i\, disagree$
```

Anonymous actions

Observables

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

Nondeterministic choice

Probabilistic choice

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

The dining cryptographers - Nondeterministic anonymity

- Schneider and Sidiropoulos considered the nondeterministic version of the Dining Cryptographers (expressed in CSP). They proved automatically using FDR that the D.C. satisfy their (nondeterministic) definition of anonymity, which is:

For any permutation ρ of the anonymous actions

$$\rho(\mathit{Traces}(P)) = \mathit{Traces}(P)$$

Where

P is the fully nondeterministic variant of the system,

Traces are the traces on

- Anonymous actions: the pay(i)'s,
- Observables: the out_i agrees and out_i disagrees

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.

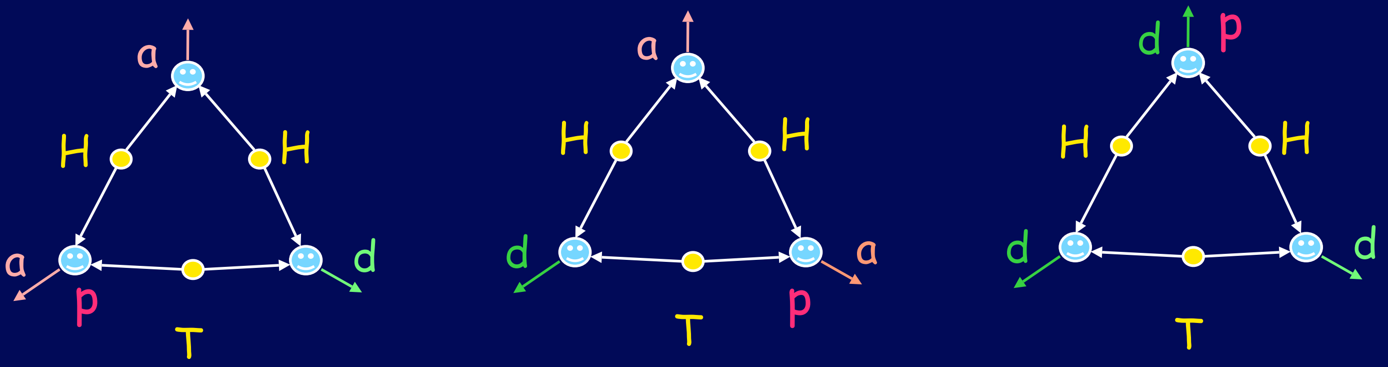
For instance: The (informal) hierarchy of Reiter and Rubin

- **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).

Towards the formalization of Strong Probabilistic Anonymity

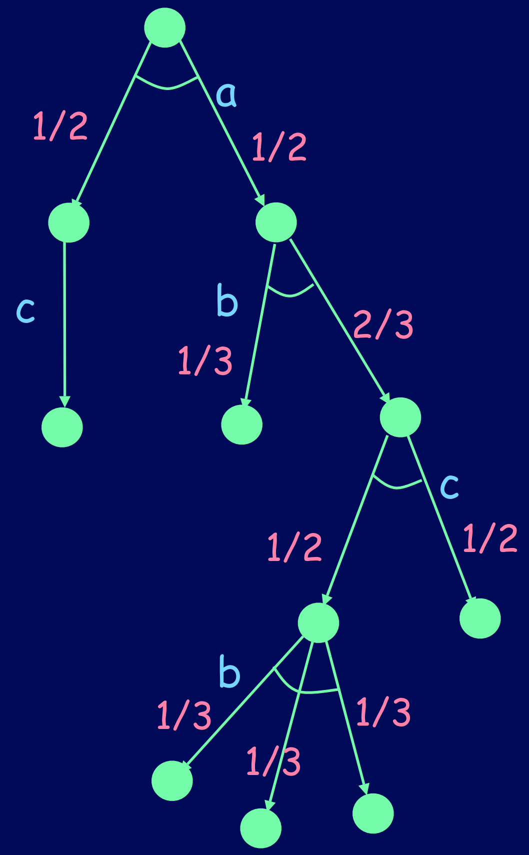
- The rest of this talk is dedicated to formalizing the notion of “beyond suspicion” (strong probabilistic anonymity)

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

- We use a formalism expressing both probabilistic and nondeterministic behavior
  - the probabilistic asynchronous pi-calculus [Herescu & Palamidessi, 2000]
  - semantics based on the Probabilistic Automata [Segala & Lynch 1994]

- Users-independence: even in case of probabilistic users, the definition should be independent from the probability distribution of the users

Fully probabilistic automata

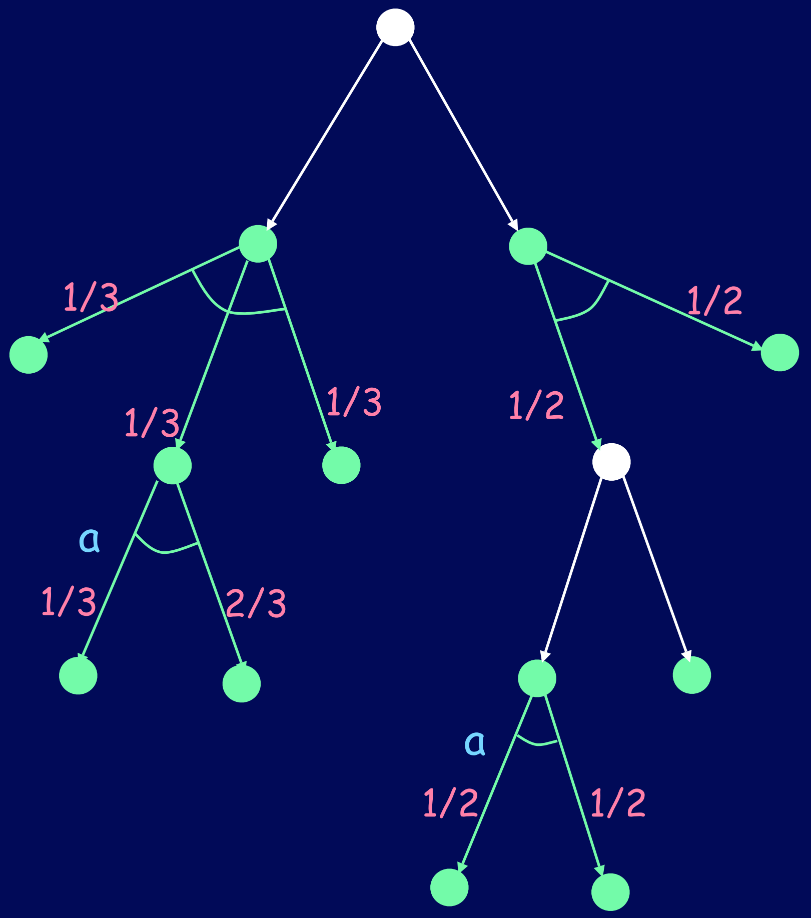


- **Observable actions:** a, b, c
- **Execution:** a path from the root to a leaf
- **Probability of an execution:** the product of the probabilities on the edges

- **Event:** a set of executions
- **Probability of an event:** the sum of the probabilities of the executions

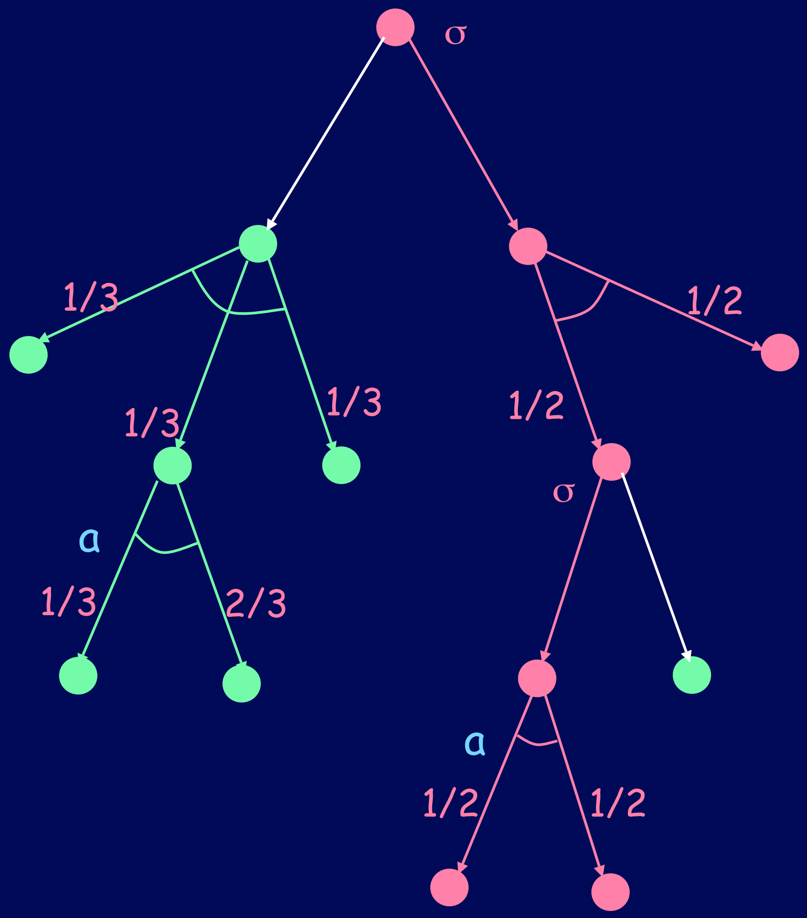
- Examples:
  - The event c has probability
    - p(c) = 1/2 + 1/6 = 2/3
  - The event ab has probability
    - p(ab) = 1/6 + 1/18 = 2/9

(Simplified) Probabilistic Automata



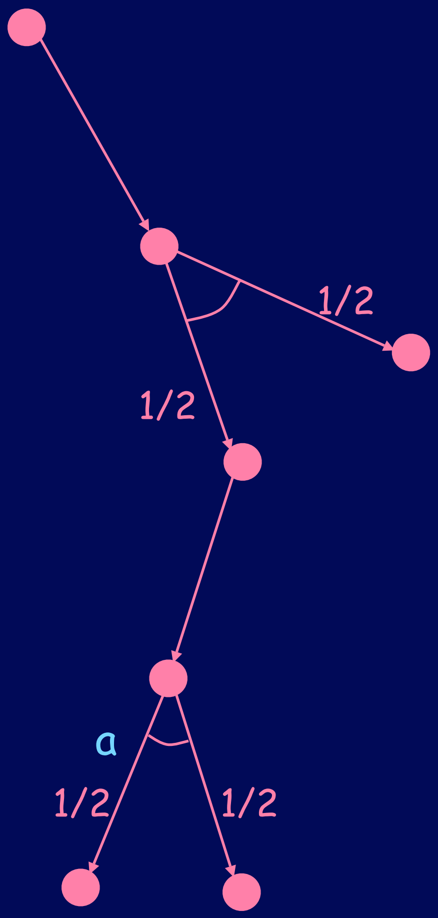
- White nodes: nondeterministic
  Green nodes: probabilistic
- Scheduler: a function that associates to each nondeterministic nodes a node among its successors
- Etree(σ): the fully probabilistic automaton obtained by pruning the tree from the choices not selected by σ
- p_σ (o) = the probability of the event o under σ

(Simplified) Probabilistic Automata



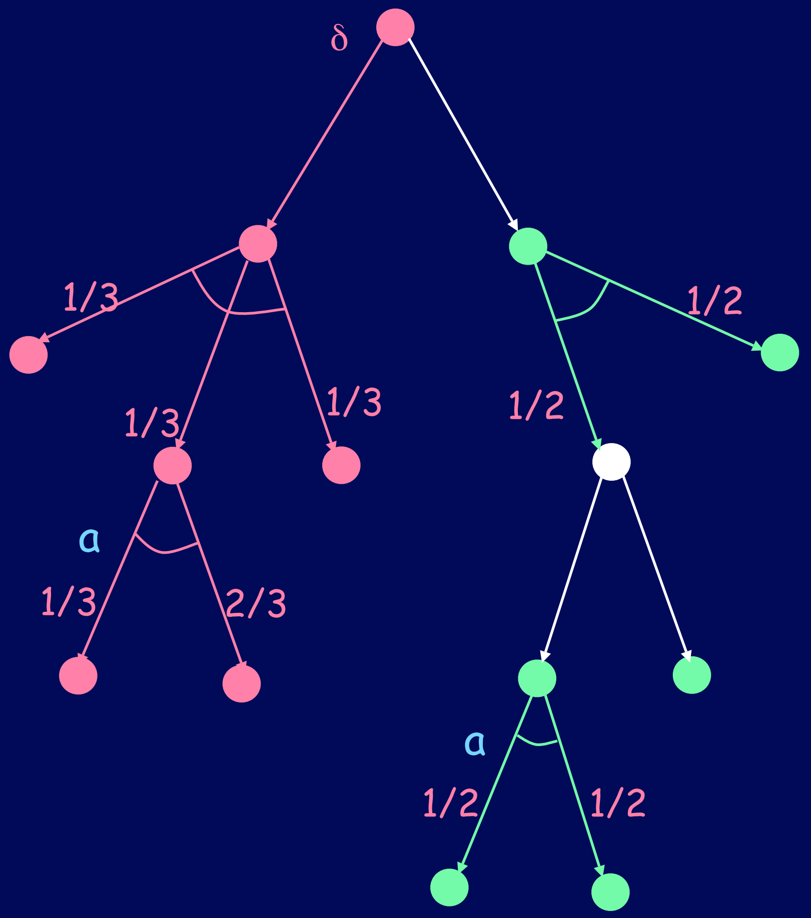
- White nodes: nondeterministic
  Green nodes: probabilistic
- Scheduler: a function that associates to each nondeterministic nodes a node among its successors
- Etree(σ): the fully probabilistic automaton obtained by pruning the tree from the choices not selected by σ
- p_σ (o) = the probability of the event o under σ

(Simplified) Probabilistic Automata



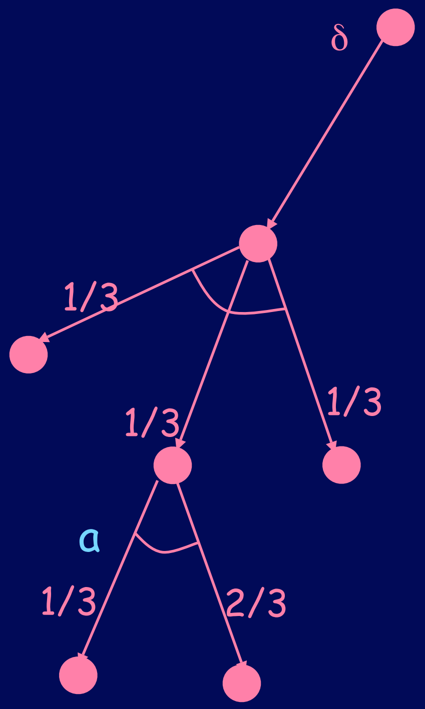
- White nodes: nondeterministic
  Green nodes: probabilistic
- Scheduler: a function that associates to each nondeterministic nodes a node among its successors
- Etree(σ): the fully probabilistic automaton obtained by pruning the tree from the choices not selected by σ
- p_{σ} (o) = the probability of the event o under σ
- p_{σ} (a) = 1/4

(Simplified) Probabilistic Automata



- White nodes: nondeterministic
  Green nodes: probabilistic
- Scheduler: a function that associates to each nondeterministic nodes a node among its successors
- Etree(σ): the fully probabilistic automaton obtained by pruning the tree from the choices not selected by σ
- p_σ (o) = the probability of the event o under σ

(Simplified) Probabilistic Automata



- White nodes: nondeterministic
  Green nodes: probabilistic
- Scheduler: a function that associates to each nondeterministic nodes a node among its successors

- Etree(σ): the fully probabilistic automaton obtained by pruning the tree from the choices not selected by σ
- p_σ (o) = the probability of the event o under σ
- p_δ (a) = 1/9

Towards the 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

Towards the formalization of Strong Probabilistic Anonymity

- Case of probabilistic users:
  - Consider a fully probabilistic version of the D.C. where the master makes its choices with uniform probability distribution. The property which has been proved by Chaum for the case of fair coins is the following:

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

  Namely: the observation of o does not add anything to the knowledge of the probability of a(i), except that the action a has been performed.

- This is similar to **Conditional Anonymity** by Halpern & O'Neill

- Problems with the above notion:
  - In general it may depend on the probability distribution of the users
  - Not applicable for nondeterministic users

Towards the formalization of S.P.A.

- 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 exaustive and mutually exclusive hypotheses $h_1,\dots,h_n$, and an event o, what is the evidence, given o, that h_1 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?

Towards the formalization of S.P.A.

- There are various definitions of evidence in literature. We adopt the following ones:

- **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

Strong probabilistic anonymity

- Assumption: The a(i) 's form a partition of event a
  Namely: there is at most one agent that performs a (at most one culprit)

- Definition of S.P.A. for probabilistic users

$$\boxed{\forall\, i, j, o\,.\ \text{ if }\ p(a(i)) > 0\ \text{ and } p(a(j)) > 0\quad \text{then}\quad p(o \mid a(i)\,) = p(o \mid a(j))} \quad (1)$$

- **Properties:**
  - (1) is satisfied by the D. C. with fair probabilistic coins and probabilistic users
  - (1) does not depend on the probability distribution of the a(i)'s
  - If $\forall$ o, either o $\Rightarrow$ a or o $\Rightarrow$ not a, then (1) is equivalent to

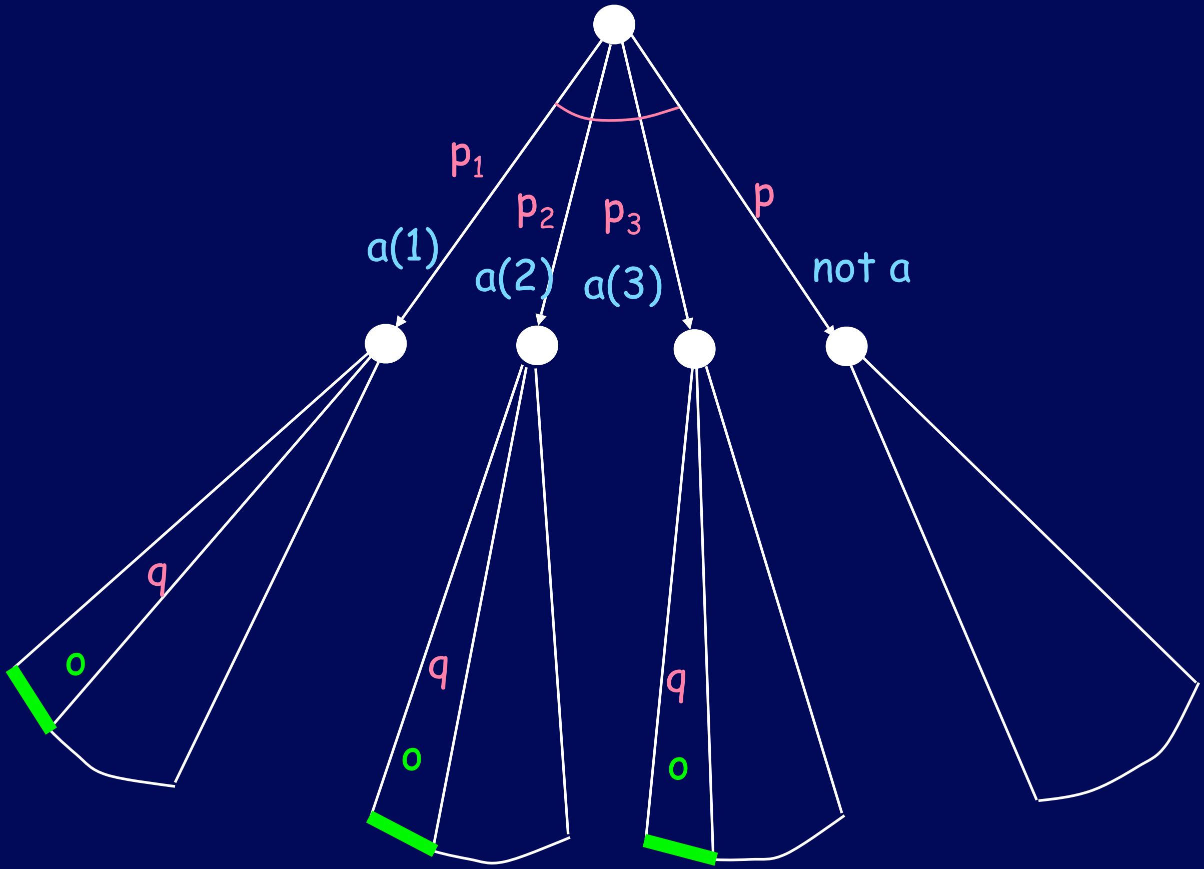
$$\boxed{\forall\, i, o\,.\ \text{ if }\ o \Rightarrow a\quad \text{then}\quad p(a(i) \mid o) = p(a(i) \mid a)} \quad (2)$$

- Definition of P.S.A. for nondeterministic users

$$\boxed{\begin{array}{c} \forall \text{ schedulers } \sigma, \delta\quad \forall\, i, j, o, \\ \text{if }\ \sigma\ \text{ selects } a(i),\ \text{and }\ \delta\ \text{ selects } a(j),\ \text{ then }\ p_\sigma(o) = p_\delta(o) \end{array}} \quad (3)$$

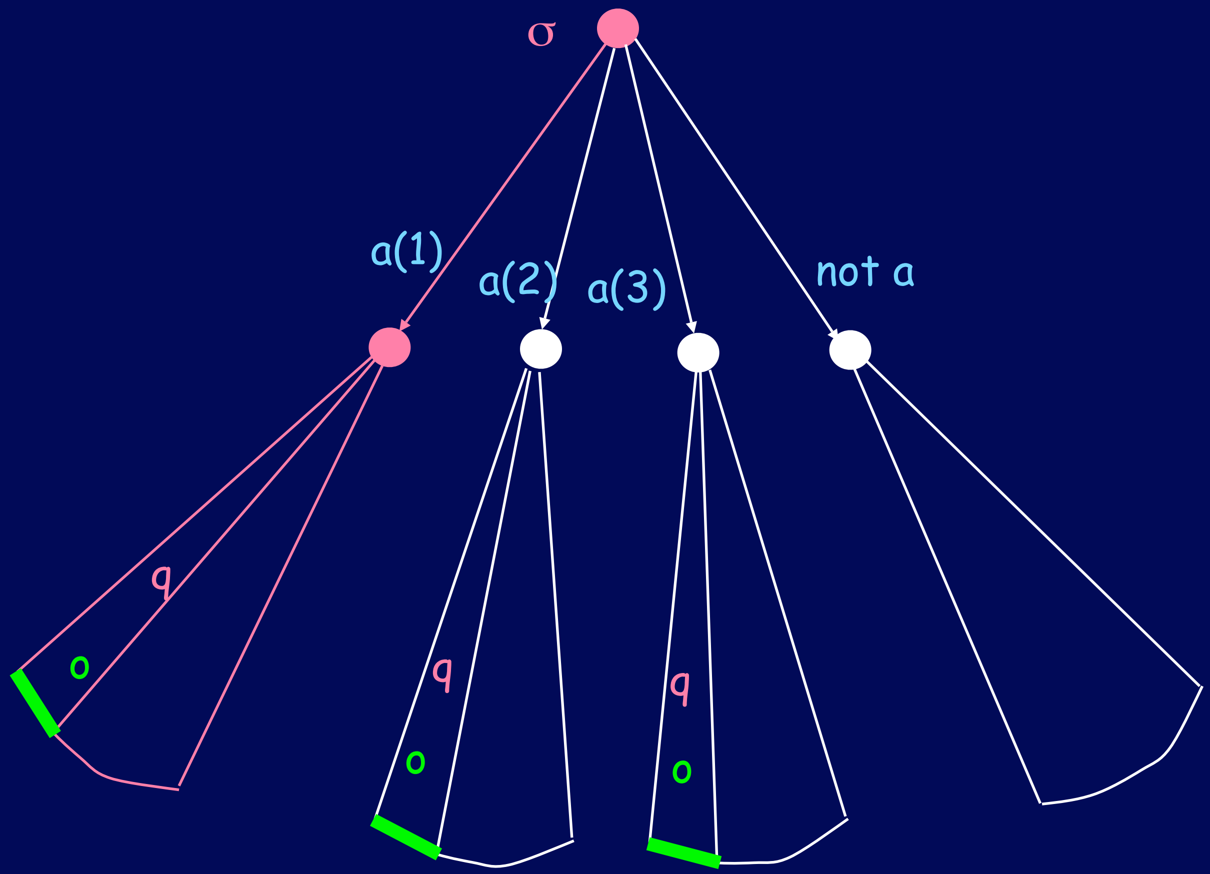
- Property: (3) is satisfied by the D. C. with fair probabilistic coins and nondeterministic users

Independence from the probability distribution of the a(i)'s



$$p(o \mid a(i)) \;=\; p(o \text{ and } a(i)) \,/\, p(a(i) \;=\; q\; p_i \,/ p_i \;=\; q$$
$$=$$
$$p(o \mid a(j))$$

Nondeterministic users



$p_\sigma(o) \;=\; q$

$=$

$p_\delta(o)$

Comparison with other (purely) probabilistic approaches

Halpern & O'Neill proposed the following formalization of Reiter and Rubin 's notions

- **Beyond suspicion:** the culprit is not more likely than any other agent

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

- **Probable innocence:** the culprit is less likely than all the other agents together

  $\forall\, i, j, o.\quad p(a(i) \mid o) < 1/2$

- **Possible innocence:** we are not sure that the culprit is indeed the culprit

  $\forall\, i, j, o\ .\ p(a(i) \mid o) < 1$

Problems:

- These notions do not have an equivalent for nondeterministic users
- They depend on the probability distribution of the users
- This "beyond suspicion" does not hold (in general) for the Dining Crypt with fair coins
- Rubin proved "probable innocence" for Crowds, but this definition of probable innocence does not hold (in general) for Crowds

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 the notion of secrecy
- Definition of a suitable logic
    - quantitative aspects
    - a form of implication corresponding to conditional probability
- Automatic verification (model checking)

Thank you !