Introduction to Privacy and Privacy-Enhancing Technologies

Vicenç Torra

October, 2013

Institut d'Investigació en Intel·ligència Artificial (IIIA-CSIC), Bellaterra, Catalonia

A kind of preface

Data Privacy

- Technological point of view
  - Computer science (communications, security, data mining)
  - Statistical disclosure control

Data Privacy: Why?

- Why data privacy?
  - Legislation
  - Companies own interest
  - Scandals

Data Privacy: Why?

- Why data privacy?
  - Legislation
  - Companies own interest
  - Scandals
    - ⋆ AOL scandal

      In order to help the information retrieval community, AOL released a large set of logs. The set was anonymized removing identifiers, but no other action was done. This ended up with not only important damage to AOL users privacy, but also a major damage to AOL itself with several class action suits and complains against the company

Data Privacy: Why? (II)

- Why data privacy?
  And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for 'landscapers in Lilburn, Ga,' several people with the last name Arnold and 'homes sold in shadow lake subdivision gwinnett county georgia.' It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. 'Those are my searches,' she said, after a reporter read part of the list to her. (New York Times, August 9, 2006)

Data Privacy: Why? (III)

- Set of queries of user user No. 4417749
- landscapers in lilburn ga.
- gwinnett county homes sold in lilburn
- homes sold in lilburn ga.
- homes sold in lilburn
- competitive market analysis of homes in lilburn
- homes sold in shadow lake subdivision gwinnett county georgia
- atlanta humane society
- pine straw lilburn delivery
- panasonic vacuum dealer in lilburn ga.
- aameetings in georgia
- pne straw in lilburn ga.
- morgan stanley atlanta ga.
- shadow lake subdivision gwinnett county georgia
- pine straw in lilburn ga.
- morgan staanley dean witter atlanta georgia
- morgan stanley dean witter atlanta georgia
- dean witter atlanta ga.
- single dances in atlanta
- jarrett arnold
- jarrett t. arnold
- jarrett t. arnold eugene oregon
- jack t. arnold
- eugene oregon jarrett arnold
- eugene oregon jaylene arnold
- jaylene and jarrett arnold eugene or.
- jarrett arnold eugene oregoon

Outline

Introduction to Privacy

Introduction to privacy

Parts:

- Definition of Privacy
- Legal Aspects
- Research Areas in Privacy

Introduction to privacy

Definition of privacy:

- A philosophical point of view.
- A technical point of view.

Introduction to privacy

Definition of privacy:

- A philosophical point of view.
  - The nature of privacy (p.2 and 3 (Schoeman, 1984))

Introduction to privacy

Definition of privacy:

- A philosophical point of view.
  - The nature of privacy (p.2 and 3 (Schoeman, 1984))
    - Privacy as a claim, entitlement, or right of an individual to determine what information about himself or herself may be communicated to others.

Introduction to privacy

Definition of privacy:

- A philosophical point of view.
  - The nature of privacy (p.2 and 3 (Schoeman, 1984))
    - ⋆ Privacy as a claim, entitlement, or right of an individual to determine what information about himself or herself may be communicated to others.
    - ⋆ Privacy as the measure of control an individual has over: information about himself; intimacies of personal identity; or who has sensory access to him.

Introduction to privacy

Definition of privacy:

- A philosophical point of view.
  - The nature of privacy (p.2 and 3 (Schoeman, 1984))
    - ⋆ Privacy as a claim, entitlement, or right of an individual to determine what information about himself or herself may be communicated to others.
    - ⋆ Privacy as the measure of control an individual has over: information about himself; intimacies of personal identity; or who has sensory access to him.
    - ⋆ Privacy as a state or condition of limited access to a person

Introduction to privacy

Definition of privacy:

- A philosophical point of view.
  - The nature of privacy (p.2 and 3 (Schoeman, 1984))
    - ⋆ Privacy as a claim, entitlement, or right of an individual to determine what information about himself or herself may be communicated to others.
    - ⋆ Privacy as the measure of control an individual has over: information about himself; intimacies of personal identity; or who has sensory access to him.
    - ⋆ Privacy as a state or condition of limited access to a person
  - Is privacy coherent and distinctive?
  - Is privacy culturally relative?

Introduction to privacy

Definition of privacy:

- A philosophical point of view.
  - The nature of privacy (p.2 and 3 (Schoeman, 1984))
    - ⋆ Privacy as a claim, entitlement, or right of an individual to determine what information about himself or herself may be communicated to others.
    - ⋆ Privacy as the measure of control an individual has over: information about himself; intimacies of personal identity; or who has sensory access to him.
    - ⋆ Privacy as a state or condition of limited access to a person
  - Is privacy coherent and distinctive?
  - Is privacy culturally relative?
- Schoeman, F. D. (1984) Philosophical dimensions of privacy: an anthology, Cambridge University Press.

Introduction to privacy

Definition of privacy:

- A technical point of view.
  - A few definitions related to privacy will follow.

Introduction to privacy

Definition of privacy:

- A technical point of view.
  - A few definitions related to privacy will follow.
  - Anonymity, Pseudonyms, Disclosure

Introduction to privacy

Legal aspects:

Introduction to privacy

Legal aspects:

- A fundamental right protected at different levels:

Introduction to privacy

Legal aspects:

- A fundamental right protected at different levels:
  - Universal, European, National levels

Introduction to privacy

Legal aspects: Universal

- The Universal Declaration of Human Rights

Introduction to privacy

Legal aspects: Universal

- The Universal Declaration of Human Rights
  - **Article 12.** No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks. (10 December 1948, UN General Assembly)

Introduction to privacy

Legal aspects: European

- European Convention on Human Rights

Introduction to privacy

Legal aspects: European

- European Convention on Human Rights
  - **Article 8** Right to respect for private and family life
    1. Everyone has the right to respect for his private and family life, his home and his correspondence.
    2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

    (Council of Europe, 3 September 1953)

Introduction to privacy

Legal aspects: National level

Introduction to privacy

Legal aspects: National level

- Information at the European Union of National level legislation

Introduction to privacy

Legal aspects: National level

- Information at the European Union of National level legislation
- http://ec.europa.eu/justice/data-protection/index_en.htm

Introduction to privacy

Legal aspects: National level

- Information at the European Union of National level legislation
- http://ec.europa.eu/justice/data-protection/index_en.htm
- xsCountry Reports:
  - http://ec.europa.eu/justice/data-protection/document/studies/inde
  - Some EU countries: Czech Republic, Denmark, France, Germany, Greece, United Kingdom

Introduction to privacy

Legal aspects: National level

- Information at the European Union of National level legislation
- http://ec.europa.eu/justice/data-protection/index_en.htm
- xsCountry Reports:
  - http://ec.europa.eu/justice/data-protection/document/studies/inde
  - Some EU countries: Czech Republic, Denmark, France, Germany, Greece, United Kingdom
  - Some Non-European countries: USA (discusses: Federal level, California, New Jersey), Australia, Hong Kong, India, Japan.

Introduction to privacy

Legal aspects: USA

Introduction to privacy

Legal aspects: USA

- Most relevant laws

Introduction to privacy

Legal aspects: USA

- Most relevant laws
  - Health Insurance Portability and Accountability Act (HIPAA, 1996)
  - Patriot Act (2001)
  - Homeland Security Act (2002)

Introduction to privacy

Research areas:

Introduction to privacy

Research areas:

- Privacy and Security
  Privacy in communications (information security), anonymity

Introduction to privacy

Research areas:

- Privacy and Security
  Privacy in communications (information security), anonymity
- Data mining
  Privacy preserving data mining

Introduction to privacy

Research areas:

- Privacy and Security
  Privacy in communications (information security), anonymity
- Data mining
  Privacy preserving data mining
- Statistics
  Statistical disclosure control

Definitions: Anonymity and Pseudonyms

Anonymity and Pseudonyms

Technical definition:

- Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others. Viewed in terms of the relation of the individual to social participation, privacy is the voluntary and temporary withdrawal of a person from the general society through physical or psychological means, either in a state of solitude or small-group intimacy or, when among larger groups, in a condition of anonymity or reserve (Westin 67[1], p.7; also in Pfitzmann, Hansen, 2010, p.6)

[1] Alan F. Westin: Privacy and Freedom; Atheneum, New York 1967

Anonymity and Pseudonyms

Parts (definitions):

- Identity and Identifiers
- Types of Pseudonyms
- Anonymity

Anonymity and Pseudonyms: Anonymity

Pfitzmann, A., Hansen, M. (2010) A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management.

http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf

Anonymity and Pseudonyms

Definitions. Framework according to the perspective in communications.

Anonymity and Pseudonyms

Definitions. Framework according to the perspective in communications.

- senders and recipients communicate (messages) through a communication network

Anonymity and Pseudonyms

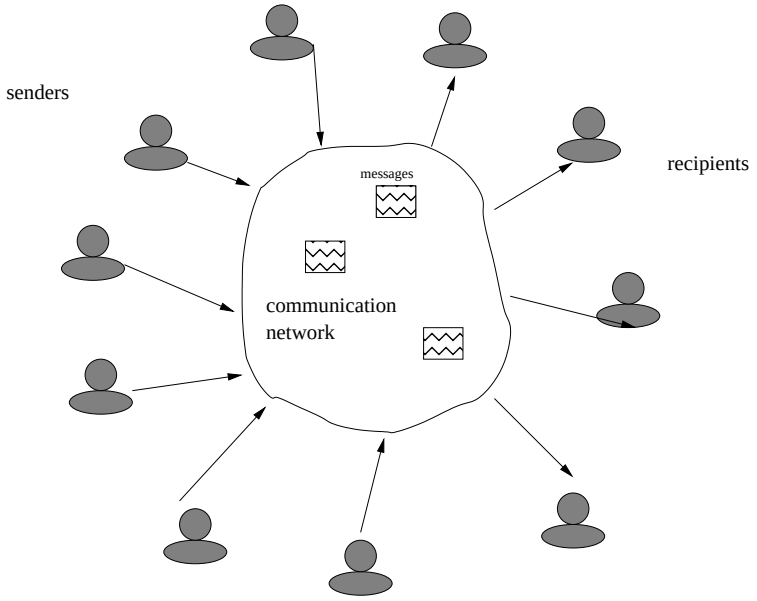
Definitions. Framework according to the perspective in communications.

- senders and recipients communicate (messages) through a communication network
  - senders, also called actors, receivers, also called actees

Anonymity and Pseudonyms

Definitions. Framework according to the perspective in communications.

- senders and recipients communicate (messages) through a communication network
  - senders, also called actors, receivers, also called actees
  - no distinction on whether the senders/receivers are human or not



Anonymity and Pseudonyms

Definitions. Adversary

Anonymity and Pseudonyms

Definitions. Adversary

- Attacker, adversary, intruder

Anonymity and Pseudonyms

Definitions. Adversary

- Attacker, adversary, intruder
  - used to refer the entities working against some protection goal

Anonymity and Pseudonyms

Definitions. Adversary

- Attacker, adversary, intruder
  - used to refer the entities working against some protection goal
  - attacker being of older use

Anonymity and Pseudonyms

Definitions. Adversary

- Attacker, adversary, intruder
  - used to refer the entities working against some protection goal
  - attacker being of older use
  - adversary more recently used in the security research community

Anonymity and Pseudonyms

Definitions. Adversary

- Attacker, adversary, intruder
  - used to refer the entities working against some protection goal
  - attacker being of older use
  - adversary more recently used in the security research community
  - privacy preserving data mining uses the terms adversary and adversarial attacks

Anonymity and Pseudonyms

Definitions. Adversary

- Attacker, adversary, intruder
  - used to refer the entities working against some protection goal
  - attacker being of older use
  - adversary more recently used in the security research community
  - privacy preserving data mining uses the terms adversary and adversarial attacks
  - intruder is the most used term in statistical disclosure control

Anonymity and Pseudonyms

Definitions. Goal of the adversary

Anonymity and Pseudonyms

Definitions. Goal of the adversary

- increase his knowledge on the items of interest (IOI).

Anonymity and Pseudonyms

Definitions. Goal of the adversary

- increase his knowledge on the items of interest (IOI).
  - Knowledge can be described in terms of probabilities on the items of interest
  - The more knowledge, the more the probabilities are closer to the true probabilities

Anonymity and Pseudonyms

Definition. Anonymity

- **Definition.** Anonymity of a subject means that the subject is not identifiable within a set of subjects, the anonymity set.

Anonymity and Pseudonyms

Definition. Anonymity

- **Definition.** Anonymity of a subject means that the subject is not identifiable within a set of subjects, the anonymity set.
  - not identifiable: the subject is not distinguishable from the other subjects within the anonymity set.
  - quantification of the anonymity is implicit.

Anonymity and Pseudonyms

Definition. On the quantification of Anonymity

Anonymity and Pseudonyms

Definition. On the quantification of Anonymity

- Different sizes of the anonymity set imply different levels of anonymity.

Anonymity and Pseudonyms

Definition. On the quantification of Anonymity

- Different sizes of the anonymity set imply different levels of anonymity.
  Example. s_1 and s_2 different levels of anonymity: anonymity set of s_1 has only three subjects and the one of s_2 has ten subjects.

Anonymity and Pseudonyms

Definition. On the quantification of Anonymity

- Different sizes of the anonymity set imply different levels of anonymity.
  Example. s_1 and s_2 different levels of anonymity: anonymity set of s_1 has only three subjects and the one of s_2 has ten subjects.
- With anonymity sets of the same cardinality, different levels of identification.

Anonymity and Pseudonyms

Definition. On the quantification of Anonymity

- Different sizes of the anonymity set imply different levels of anonymity.
  Example. s_1 and s_2 different levels of anonymity: anonymity set of s_1 has only three subjects and the one of s_2 has ten subjects.
- With anonymity sets of the same cardinality, different levels of identification.
  Example. $AS(s_1) = \{s_{11}, s_{12}, s_{13}\}$ be the anonymity set of s_1; then, it is better with respect to anonymity the distribution $(1/3, 1/3, 1/3)$ than the distribution $(0.9, 0.05, 0.05)$

Anonymity and Pseudonyms

Definition. Anonymity (making explicit the level of identification)

Anonymity and Pseudonyms

Definition. Anonymity (making explicit the level of identification)

- **Definition.** From an adversary's persective, anonymity of a subject s means that the adversary cannot achieve certain level of identification for the subject s within the anonymity set.

Anonymity and Pseudonyms: Unlinkability

Anonymity and Pseudonyms

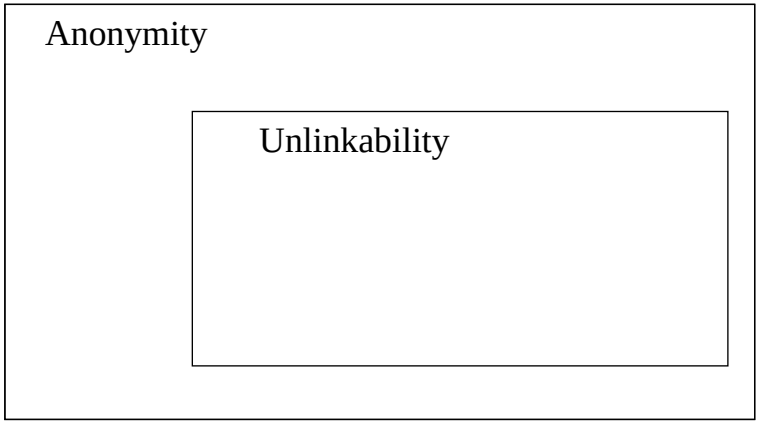
Definition. Unlinkability

- **Definition.** Unlinkability of two or more items of interest (IOIs, e.g. subjects, messages, actions, ...) from an attacker's perspective means that within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether these IOIs are related or not.

Anonymity and Pseudonyms

Discussion of unlinkability

- (Pfitzmann, Hansen, 2010) point out that unlinkability is a sufficient condition of anonymity but not a necessary condition. That is, unlinkability implies anonymity.



Anonymity and Pseudonyms

Discussion of unlinkability

- (Pfitzmann, Hansen, 2010) point out that unlinkability is a sufficient condition of anonymity but not a necessary condition. That is, unlinkability implies anonymity.
- Case of linkability and anonymity not compromised.

Anonymity and Pseudonyms

Discussion of unlinkability

- (Pfitzmann, Hansen, 2010) point out that unlinkability is a sufficient condition of anonymity but not a necessary condition. That is, unlinkability implies anonymity.
- Case of linkability and anonymity not compromised.
  Example. An attacker can link all messages of a transaction, due to timing, nevertheless if all of them are encrypted so that no information can be obtained about the subjects in the transactions, then anonymity is not compromised.

Anonymity and Pseudonyms

Discussion of unlinkability: examples of anonymity in communication

- Sender anonymity: of a subject means that to this potentially sending subject, each message is unlinkable

Anonymity and Pseudonyms

Discussion of unlinkability: examples of anonymity in communication

- Sender anonymity: of a subject means that to this potentially sending subject, each message is unlinkable
- Recipient anonymity: of a subject means that to this potentially sending subject, each message is unlinkable

Anonymity and Pseudonyms

Discussion of unlinkability: examples of anonymity in communication

- Sender anonymity: of a subject means that to this potentially sending subject, each message is unlinkable
- Recipient anonymity: of a subject means that to this potentially sending subject, each message is unlinkable
- Relationship anonymity: of a pair of subjects, the potentially sending subject and the potentially receiving subject, means that to this potentially communicating pair of subjects, each message is unlinkable

Anonymity and Pseudonyms: Disclosure

Anonymity and Pseudonyms

Definition. Disclosure (concept used within SDC and PPDM communities)

Anonymity and Pseudonyms

Definition. Disclosure (concept used within SDC and PPDM communities)

- **Definition.** Disclosure takes place when the attacker takes advantage of the observation of available data to improve his knowledge on some confidential information about an item of interest.

Anonymity and Pseudonyms

Definition. Disclosure (concept used within SDC and PPDM communities)

- **Definition.** Disclosure takes place when the attacker takes advantage of the observation of available data to improve his knowledge on some confidential information about an item of interest.
  - Defined in terms of the additional confidential information or knowledge that an adversary can acquire from observing the system.

Anonymity and Pseudonyms

Definition. Disclosure (concept used within SDC and PPDM communities)

- **Definition.** Disclosure takes place when the attacker takes advantage of the observation of available data to improve his knowledge on some confidential information about an item of interest.
  - Defined in terms of the additional confidential information or knowledge that an adversary can acquire from observing the system.
  - In SDC and PPDM, typically observes a protected file or database. Then, disclosure is about improving the knowledge of a particular subject in the database.

Anonymity and Pseudonyms

Disclosure. Two types of disclosure

Anonymity and Pseudonyms

Disclosure. Two types of disclosure

- **Identity disclosure.** When the adversary can link a respondent with a particular record in the protected data set.

Anonymity and Pseudonyms

Disclosure. Two types of disclosure

- **Identity disclosure.** When the adversary can link a respondent with a particular record in the protected data set.
  - Only respondents whose data have been published can be identified
  - It relates to the concept of linkability discussed above
  - Identity disclosure is also known as entity disclosure

Anonymity and Pseudonyms

Disclosure. Two types of disclosure

- **Identity disclosure.** When the adversary can link a respondent with a particular record in the protected data set.
  - Only respondents whose data have been published can be identified
  - It relates to the concept of linkability discussed above
  - Identity disclosure is also known as entity disclosure
- **Attribute disclosure.** When the adversary can learn something new about an attribute of a respondent, even when no relationship can be established between the individual and the data.

Anonymity and Pseudonyms

Disclosure. Two types of disclosure

- **Identity disclosure.** When the adversary can link a respondent with a particular record in the protected data set.
  - Only respondents whose data have been published can be identified
  - It relates to the concept of linkability discussed above
  - Identity disclosure is also known as entity disclosure
- **Attribute disclosure.** When the adversary can learn something new about an attribute of a respondent, even when no relationship can be established between the individual and the data.
  - The intruder increases his accuracy on an attribute of the respondent

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- Identity disclosure neither stronger nor weaker than attribute disclosure.

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- Identity disclosure neither stronger nor weaker than attribute disclosure.
  $\rightarrow$ Most common: identity disclosure implies attribute disclosure.
  (E.g. we identify one record using a subset of the attributes)

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- Identity disclosure neither stronger nor weaker than attribute disclosure.
  $\rightarrow$ Most common: identity disclosure implies attribute disclosure.
  (E.g. we identify one record using a subset of the attributes)
- We may have
  - (1) Neither identity nor attribute disclosure
  - (2) Identity disclosure and attribute disclosure
  - (3) No identity disclosure but attribute disclosure
  - (4) Identity disclosure but not attribute disclosure
    ((4) is against the implication above)

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (2) Identity disclosure and attribute disclosure

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (2) Identity disclosure and attribute disclosure
  - **Re-identification and attribute disclosure.** Establishing a link with the records gives information.

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (2) Identity disclosure and attribute disclosure
  - **Re-identification and attribute disclosure.** Establishing a link with the records gives information.
  - **Example.** Adversary has the following information $(HYU, Tarragona, 58)$

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (2) Identity disclosure and attribute disclosure
  - **Re-identification and attribute disclosure.** Establishing a link with the records gives information.
  - **Example.** Adversary has the following information $(HYU, Tarragona, 58)$

Respondent	City	Age	Illness
ABD	Barcelona	30	Cancer
COL	Barcelona	30	Cancer
GHE	Tarragona	60	AIDS
CIO	Tarragona	60	AIDS
HYU	Tarragona	58	Heart attack

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (3) No identity disclosure but attribute disclosure

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (3) No identity disclosure but attribute disclosure
  - **Attribute disclosure without re-identification.** This is the case of some k-anonymous data. All k-anonymous records have the same value for a confidential attribute, we have attribute disclosure for this attribute. This issue motivated the definition of l-diversity.

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (3) No identity disclosure but attribute disclosure
  - **Attribute disclosure without re-identification.** This is the case of some k-anonymous data. All k-anonymous records have the same value for a confidential attribute, we have attribute disclosure for this attribute. This issue motivated the definition of l-diversity.
  - **Example.** Adversary has the following information $(ABD, Barcelona, 30)$

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (3) No identity disclosure but attribute disclosure
  - **Attribute disclosure without re-identification.** This is the case of some k-anonymous data. All k-anonymous records have the same value for a confidential attribute, we have attribute disclosure for this attribute. This issue motivated the definition of l-diversity.
  - **Example.** Adversary has the following information $(ABD, Barcelona, 30)$

Respondent	City	age	illness
ABD	Barcelona	30	Cancer
COL	Barcelona	30	Cancer
GHE	Tarragona	60	AIDS
CIO	Tarragona	60	AIDS

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (4) Identity disclosure but not attribute disclosure (case against implication)

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- (4) Identity disclosure but not attribute disclosure (case against implication)
  - **Re-identification without attribute disclosure.** E.g. when all attributes are needed for re-identification, or when the attributes not used for re-identification do not cause disclosure.
  - **Example.** Adversary has the following information $(HYU, Tarragona, 60, Heartattack)$

Anonymity and Pseudonyms

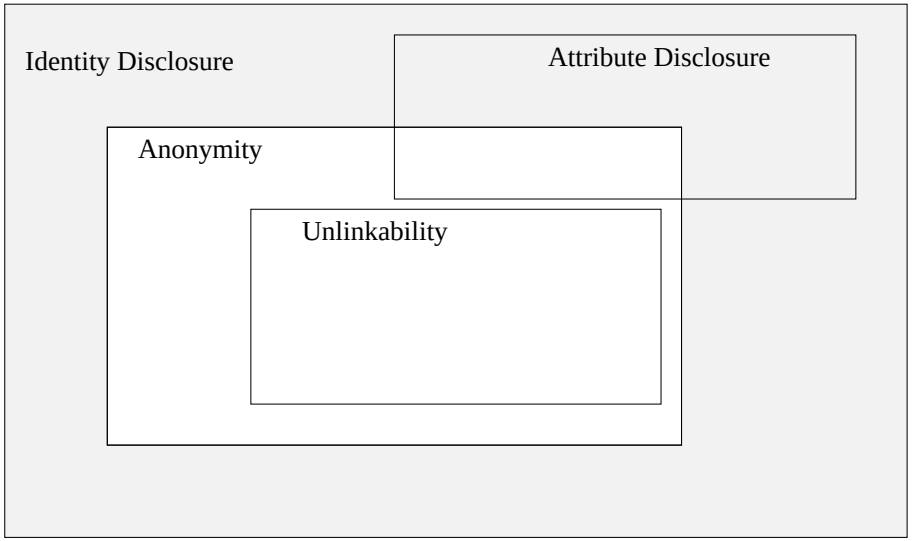
Disclosure. Attribute disclosure vs. identity disclosure

- (4) Identity disclosure but not attribute disclosure (case against implication)
  - **Re-identification without attribute disclosure.** E.g. when all attributes are needed for re-identification, or when the attributes not used for re-identification do not cause disclosure.
  - **Example.** Adversary has the following information $(HYU, Tarragona, 60, Heartattack)$

Respondent	City	age	illness
ABD	Barcelona	30	Cancer
COL	Barcelona	30	Cancer
GHE	Tarragona	60	AIDS
CIO	Tarragona	60	AIDS
HYU	Tarragona	60	Heart attack

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure



Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- Identity disclosure can be understood as a binary property. That is, an individual is re-identified or not.

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- Identity disclosure can be understood as a binary property. That is, an individual is re-identified or not.
- Attribute disclosure is not, as it is not uncommon to consider that any release of data would lead to attribute disclosure to some extent. Otherwise, the utility of the protected data might be too low. See e.g.

Anonymity and Pseudonyms

Disclosure. Attribute disclosure vs. identity disclosure

- Identity disclosure can be understood as a binary property. That is, an individual is re-identified or not.
- Attribute disclosure is not, as it is not uncommon to consider that any release of data would lead to attribute disclosure to some extent. Otherwise, the utility of the protected data might be too low. See e.g.

  Quote. At the other extreme, any improvement in our knowledge about an individual could be considered an intrusion. The latter is particularly likely to cause a problem for data mining, as the goal is to improve our knowledge. (p. 7 of (Vaidya, Clifton, Zhu 2006))

Anonymity and Pseudonyms

Disclosure. Attribute disclosure, identity disclosure, and ...

Anonymity and Pseudonyms

Disclosure. Attribute disclosure, identity disclosure, and ...

- Inferential disclosure.

Anonymity and Pseudonyms

Disclosure. Attribute disclosure, identity disclosure, and ...

- Inferential disclosure.
  We do not distinguish it from attribute disclosure

Anonymity and Pseudonyms

Disclosure. Attribute disclosure, identity disclosure, and ...

- Inferential disclosure.
  We do not distinguish it from attribute disclosure
- Difference ID / AD is whether the information is obtained directly from the protected data set, or it is inferred from this data set.

Anonymity and Pseudonyms

Disclosure. Attribute disclosure, identity disclosure, and ...

- Inferential disclosure.
  We do not distinguish it from attribute disclosure
- Difference ID / AD is whether the information is obtained directly from the protected data set, or it is inferred from this data set.
- **Example Inferential Disclosure** The data may show a high correlation between income and purchase price of a home. As the purchase price of a home is typically public information, a third party might use this information to infer the income of a data subject.

Anonymity and Pseudonyms

Disclosure. disclosure vs. anonymity

- Anonymity does not necessarily prevent disclosure

Anonymity and Pseudonyms

Disclosure. disclosure vs. anonymity

- Anonymity does not necessarily prevent disclosure
  - **Example.** If we have that all objects in the anonymity set satisfy a property p, the attacker can infer this property p for the subject of interest. So, there is disclosure but still anonymity.

Anonymity and Pseudonyms

Disclosure. disclosure vs. anonymity

- Anonymity does not necessarily prevent disclosure
  - **Example.** If we have that all objects in the anonymity set satisfy a property p, the attacker can infer this property p for the subject of interest. So, there is disclosure but still anonymity.
  - **Similar.** *no identity disclosure but attribute disclosure*

Anonymity and Pseudonyms

Disclosure. disclosure vs. anonymity

- Anonymity does not necessarily prevent disclosure
  - **Example.** If we have that all objects in the anonymity set satisfy a property p, the attacker can infer this property p for the subject of interest. So, there is disclosure but still anonymity.
  - **Similar.** *no identity disclosure but attribute disclosure*

Disclosure. Disclosure and anonymity

- anonymity cannot increase, disclosure never decreases

Anonymity and Pseudonyms: Undetectability and Unobservability

Anonymity and Pseudonyms

Definition. Undetectability

Anonymity and Pseudonyms

Definition. Undetectability

- **Definition.** Undetectability of an item of interest (IOI) from an attacker's perspective means that the attacker cannot sufficiently distinguish whether it exists or not.

Anonymity and Pseudonyms

Definition. Undetectability

- **Definition.** Undetectability of an item of interest (IOI) from an attacker's perspective means that the attacker cannot sufficiently distinguish whether it exists or not.
  - **Example.** if the attacker's interest is on messages, we have undetectability when he cannot distinguish the messages in the system from random noise.

Anonymity and Pseudonyms

Discussion. Tools for undetectability

Anonymity and Pseudonyms

Discussion. Tools for undetectability

- Steganography gives tools to embed undetectable messages in other physical or digital objects for their transmission.
  In computer science messages are often embedded in images but also in other objects as databases.
- **Note.**
  - Undetectability not directly related to anonymity / privacy.

Anonymity and Pseudonyms

Discussion. Tools for undetectability

- Steganography gives tools to embed undetectable messages in other physical or digital objects for their transmission.
  In computer science messages are often embedded in images but also in other objects as databases.
- **Note.**
  - Undetectability not directly related to anonymity / privacy.
  - But undetected messages will not raise the interest of the attacker.

Anonymity and Pseudonyms

Discussion. Tools for undetectability

- Steganography gives tools to embed undetectable messages in other physical or digital objects for their transmission.
  In computer science messages are often embedded in images but also in other objects as databases.
- **Note.**
  - Undetectability not directly related to anonymity / privacy.
  - But undetected messages will not raise the interest of the attacker.
  - Counter attack steganographic systems: Steganalysis. (Westfeld, Pfitzman, 2000) describe visual and statistical attacks for steganographic images.

Anonymity and Pseudonyms

Definition. Unobservability

Anonymity and Pseudonyms

Definition. Unobservability

- **Definition.** Unobservability of an item of interest (IOI) means
  - undetectability of the IOI against all subjects uninvolved in it and
  - anonymity of the subject(s) involved in the IOI even against the other subject(s) involved in that IOI.

Anonymity and Pseudonyms

Definition. Unobservability

- **Definition.** Unobservability of an item of interest (IOI) means
  - undetectability of the IOI against all subjects uninvolved in it and
  - anonymity of the subject(s) involved in the IOI even against the other subject(s) involved in that IOI.

An implication.

Unobservability pressumes undetectability and, if items detected, anonymity

So, unobservability implies anonymity and undetectability

Anonymity and Pseudonyms: Pseudonyms and identity

Anonymity and Pseudonyms

Pseudonyms. Previous to computers and internet, e.g. to avoid the disclosure of sensitive information, as gender.

Anonymity and Pseudonyms

Pseudonyms. Previous to computers and internet, e.g. to avoid the disclosure of sensitive information, as gender.

- Amandine Aurore Lucile Dupin (Paris, 1804 - Nohant-Vic, 1876): George Sand
- Caterina Albert (L'Escala, 1869 - L'Escala, 1966): Víctor Català.

Anonymity and Pseudonyms

Pseudonyms. Previous to computers and internet, e.g. to avoid the disclosure of sensitive information, as gender.

- Amandine Aurore Lucile Dupin (Paris, 1804 - Nohant-Vic, 1876): George Sand
- Caterina Albert (L'Escala, 1869 - L'Escala, 1966): Víctor Català.

Definition. A pseudonym is an identifier of a subject other than one of the subject's real names.

Anonymity and Pseudonyms

Pseudonyms. Previous to computers and internet, e.g. to avoid the disclosure of sensitive information, as gender.

- Amandine Aurore Lucile Dupin (Paris, 1804 - Nohant-Vic, 1876): George Sand
- Caterina Albert (L'Escala, 1869 - L'Escala, 1966): Víctor Català.

Definition. A pseudonym is an identifier of a subject other than one of the subject's real names.

- The holder of the pseudonym corresponds to the subject to which the pseudonym refers to.

Anonymity and Pseudonyms

Pseudonyms. Previous to computers and internet, e.g. to avoid the disclosure of sensitive information, as gender.

- Amandine Aurore Lucile Dupin (Paris, 1804 - Nohant-Vic, 1876): George Sand
- Caterina Albert (L'Escala, 1869 - L'Escala, 1966): Víctor Català.

Definition. A pseudonym is an identifier of a subject other than one of the subject's real names.

- The holder of the pseudonym corresponds to the subject to which the pseudonym refers to.
- Several pseudonyms may correspond to the same subject

Anonymity and Pseudonyms

Pseudonyms. Previous to computers and internet, e.g. to avoid the disclosure of sensitive information, as gender.

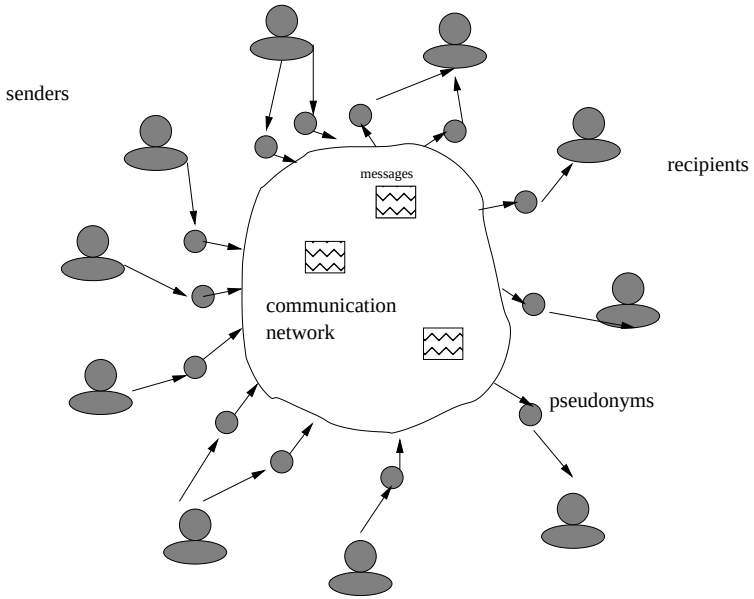
- Amandine Aurore Lucile Dupin (Paris, 1804 - Nohant-Vic, 1876): George Sand
- Caterina Albert (L'Escala, 1869 - L'Escala, 1966): Víctor Català.

Definition. A pseudonym is an identifier of a subject other than one of the subject's real names.

- The holder of the pseudonym corresponds to the subject to which the pseudonym refers to.
- Several pseudonyms may correspond to the same subject
- Several subjects can use the same pseudonym (group pseudonym). $\rightarrow$ the subjects may define an anonymity set

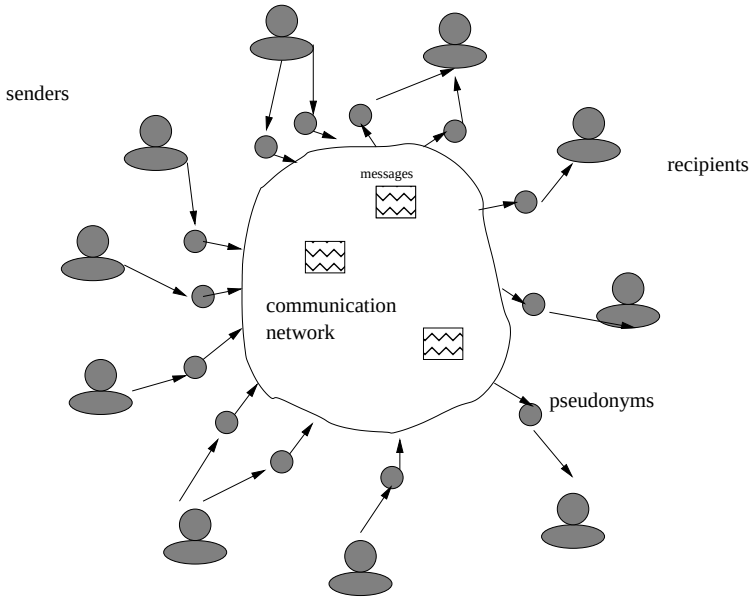
Anonymity and Pseudonyms

Pseudonyms. Several pseudonyms, one subject



Anonymity and Pseudonyms

Pseudonyms. Several pseudonyms, one subject



Pseudonyms permit to cover the range between anonymity to accountability (maximum linkability).

An individual can have a set of pseudonyms to avoid linkability.

Anonymity and Pseudonyms

Pseudonyms. Several pseudonyms, one subject

- **Problems.** Pseudonymous can be linked to individual's identity.

Anonymity and Pseudonyms

Pseudonyms. Several pseudonyms, one subject

- **Problems.** Pseudonymous can be linked to individual's identity.
- **Examples**
  - User names in email accounts, social networks
  - Identificable number in devices and software: RFID codes, Chrome browser's unique ID number
  - Communication parameters: IP address, Bluetooth number, cookie ID, User-agent string
  - Logs of internet transactions: search logs (search terms), access logs
  - Written posts and forms: posts in social networks and other types of on-line discussions

Anonymity and Pseudonyms

Pseudonyms and Pseudonymising

- Pseudonymising: the replacing of the name or other identifiers by a number in order to make the identification of the data subject impossible or substantially more difficult.
  (Federal Data Protection Act, Germany, 2001)

Anonymity and Pseudonyms

Pseudonyms and Pseudonymising

- Pseudonymising: the replacing of the name or other identifiers by a number in order to make the identification of the data subject impossible or substantially more difficult.
  (Federal Data Protection Act, Germany, 2001)

  Anonymous data are not (no longer) personal data, while pseudonymous data are personal data for the person who made them pseudonymous (i.e., the holder of the 'key') but not for the person (e.g., a researcher) who received the encoded data but who has no access to the 'key.' However, if a recipient of pseudonymized (or indeed anonymised) data for some reason has other data (or certain special means) available to him that (in spite of the anonymising or pseudonymising of the data) allow him to re-establish a link with the individuals concerned, the data again become personal for that recipient. (p.4 in (Korff, 2010))

Anonymity and Pseudonyms

Definition. Identity

Anonymity and Pseudonyms

Definition. Identity

- **Definition.** An identity is any subset of attribute values of an individual person which sufficiently identifies this individual person within any set of persons. So usually there is no such thing as "the identity", but several of them.

Anonymity and Pseudonyms

Definitions related to identity and identity management

Anonymity and Pseudonyms

Definitions related to identity and identity management

- **Role.** Roles are defined as the set of actions allowed.

Anonymity and Pseudonyms

Definitions related to identity and identity management

- **Role.** Roles are defined as the set of actions allowed.
  Example. In university, professors, members of the administration, people that study.

Anonymity and Pseudonyms

Definitions related to identity and identity management

- **Role.** Roles are defined as the set of actions allowed.
  Example. In university, professors, members of the administration, people that study.
  - People in the role of a professor, an official, or a student.
  - Then, information system with these three roles.
  - Roles are not exclusive, information systems permit a user to login using different roles.

Anonymity and Pseudonyms

Definitions related to identity and identity management

- **Partial identity.** It represents the person in a specific context or role
  - The terms virtual (partial) identity and digital (partial) identity are used to refer to the data stored by a computer-based application. Virtual identities can correspond to subjects' login information into computer-based application. Systems can then have profiles for each virtual (partial) identity.

Anonymity and Pseudonyms

Definitions related to identity and identity management

- **Identity management.** Creating, storing and processing the identities of the users in an information system.

Anonymity and Pseudonyms

Definitions related to identity and identity management

- **Identity management.** Creating, storing and processing the identities of the users in an information system.
  - Includes, authentication of the user, assigning roles to users so that they can only perform authorized actions, managing the stored information about the users and their logged activities.

Anonymity and Pseudonyms

Summary of terms (I) from (Pfitzmann, Hansen, 2010)

- item of interest (IOI) <is>
- entity
- subject
- actor
- actee
- natural person (= human being)
- legal person
- computer
- sender of a message
- recipient of a message
- insider
- outsider
- object
- message

Anonymity and Pseudonyms

Summary of terms (II) from (Pfitzmann, Hansen, 2010)

- item of interest (IOI) <is>
- action
- sending of message
- receiving of message
- identifier
- name
- pseudonym
- digital pseudonym

Anonymous Communication Mechanisms

Anonymous Communication Mechanisms

Systems for anonymity and unobservability.

Anonymous Communication Mechanisms

Systems for anonymity and unobservability.

- Anonymity systems: two types

Anonymous Communication Mechanisms

Systems for anonymity and unobservability.

- Anonymity systems: two types

  High-latency anonymity systems. Interaction is not needed

Anonymous Communication Mechanisms

Systems for anonymity and unobservability.

- Anonymity systems: two types

  High-latency anonymity systems. Interaction is not needed

  Applications: email.

  Examples. mix networks.

Anonymous Communication Mechanisms

Systems for anonymity and unobservability.

- Anonymity systems: two types

  High-latency anonymity systems. Interaction is not needed

  Applications: email.

  Examples. mix networks.

  Low-latency anonymity systems. Interaction is needed

Anonymous Communication Mechanisms

Systems for anonymity and unobservability.

- Anonymity systems: two types

  High-latency anonymity systems. Interaction is not needed

  Applications: email.

  Examples. mix networks.

  Low-latency anonymity systems. Interaction is needed

  Applications: real-time as e.g. web browsing

  Examples. onion routing and crowds.

Anonymous Communication Mechanisms

Systems for anonymity and unobservability.

- Anonymity systems: two types

  High-latency anonymity systems. Interaction is not needed

  Applications: email.

  Examples. mix networks.

  Low-latency anonymity systems. Interaction is needed

  Applications: real-time as e.g. web browsing

  Examples. onion routing and crowds.

- Unobservability

Anonymous Communication Mechanisms

Systems for anonymity and unobservability.

- Anonymity systems: two types

  High-latency anonymity systems. Interaction is not needed

  Applications: email.

  Examples. mix networks.

  Low-latency anonymity systems. Interaction is needed

  Applications: real-time as e.g. web browsing

  Examples. onion routing and crowds.

- Unobservability
  - Dining cryptographer networks

Anonymous Communication Mechanisms

High-latency anonymity systems. Interaction is not needed: email

- Mix network

Anonymous Communication Mechanisms

High-latency anonymity systems. Interaction is not needed: email

- Mix network
- Centralized remailers:

Anonymous Communication Mechanisms

High-latency anonymity systems. Interaction is not needed: email

- Mix network
- Centralized remailers:
  - anon.penet.fi (in Finland)

Anonymous Communication Mechanisms

High-latency anonymity systems. Interaction is not needed: email

- Mix network
- Centralized remailers:
  - anon.penet.fi (in Finland)
- Decentralized remailers:
  - Cypherpunk (type I) and Mixminion of (type III).

Anonymous Communication Mechanisms

Mix network. (Chaum, 1981) to unlink sender and receiver

- Short description: A proxy server receives an encrypted message from the sender and forwards it to the receiver. Sender and receiver are unlinked when the server receives and sends messages from different senders to different receivers after suffling them.

Anonymous Communication Mechanisms

Public-key cryptography. It requires two separate keys, one of which is private and one of which is public.

- Also known as asymmetric cryptography.

Anonymous Communication Mechanisms

Public-key cryptography. It requires two separate keys, one of which is private and one of which is public.

- Also known as asymmetric cryptography.

A message for A is encrypted using the public key of A

A will decrypt the message with his private key.

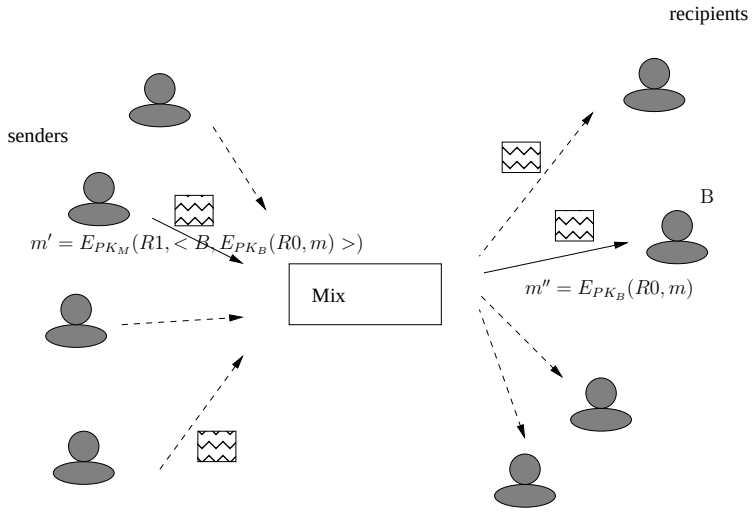
Anonymous Communication Mechanisms

Mix network. The mix has a public key PK and a private key SK.

- $E_{PK}(m)$ denotes encryption using the public key PK of message m,
- D_{PK} denotes decryption mechanism.

Anonymous Communication Mechanisms

Mix network. Representation



Anonymous Communication Mechanisms

Mix network. Steps of the process (I)

Message preparation. A sender wants to send a message m to receiver B. Using the public key PK_M of the mix and the public key of B computes:

$$m' = E_{PK_M}(R1, < B, E_{PK_B}(R0, m) >)$$

where $R0$ and $R1$ are random strings, B denotes the address of the receiver, and $< a, b >$ denotes a message with the pair a and b.

Anonymous Communication Mechanisms

Mix network. Steps of the process (II)

Message sent. The message m' is sent to the mix. The mix decrypts m' using his private key.

$$\begin{aligned} m' = D_{SK_M}(m') &= D_{SK_M}(E_{PK_M}(R1, < B, E_{PK_B}(R0, m) >)) \\ &= (R1, < B, E_{PK_B}(R0, m) >) \end{aligned}$$

Then, the mix discards the random string $R1$, and sends the message $E_{PK_B}(R0, m)$ to B. Let m'' denote this message.

$$m'' = E_{PK_B}(R0, m)$$

Anonymous Communication Mechanisms

Mix network. Steps of the process (III)

Reception. The receiver B uses his private key to decrypt the message. That is, it computes:

$$D_{SK_B}(m'') = D_{SK_B}(E_{PK_B}(R0, m)) = (R0, m).$$

Then, it discards the random string $R0$ to find m.

Anonymous Communication Mechanisms

Flushing algorithms. Algorithms used by the mix to order the messages

Anonymous Communication Mechanisms

Flushing algorithms. Algorithms used by the mix to order the messages and attacks taking into account the flushing algorithm used.

Example.

Anonymous Communication Mechanisms

Flushing algorithms. Algorithms used by the mix to order the messages and attacks taking into account the flushing algorithm used.

Example.

- **Threshold mixes** (Chaum, 1981): collect n encrypted messages, then decrypt all of them, an forward them to destination in random order. (n a parameter)

Anonymous Communication Mechanisms

Flushing algorithms. Algorithms used by the mix to order the messages and attacks taking into account the flushing algorithm used.

Example.

- **Threshold mixes** (Chaum, 1981): collect n encrypted messages, then decrypt all of them, an forward them to destination in random order. (n a parameter)
- **Attack:** $(n-1)$attack (or flooding attack) (Serjantov et al. 2002)

Anonymous Communication Mechanisms

Flushing algorithms. Algorithms used by the mix to order the messages and attacks taking into account the flushing algorithm used.

Example.

- **Threshold mixes** (Chaum, 1981): collect n encrypted messages, then decrypt all of them, an forward them to destination in random order. (n a parameter)
- **Attack:** $(n-1)$attack (or flooding attack) (Serjantov et al. 2002)
  - An adversary delays a legitimate message m,

Anonymous Communication Mechanisms

Flushing algorithms. Algorithms used by the mix to order the messages and attacks taking into account the flushing algorithm used.

Example.

- **Threshold mixes** (Chaum, 1981): collect n encrypted messages, then decrypt all of them, an forward them to destination in random order. (n a parameter)
- **Attack:** $(n-1)$attack (or flooding attack) (Serjantov et al. 2002)
  - An adversary delays a legitimate message m,
  - generates and sends dummy messages until the mix flushes,

Anonymous Communication Mechanisms

Flushing algorithms. Algorithms used by the mix to order the messages and attacks taking into account the flushing algorithm used.

Example.

- **Threshold mixes** (Chaum, 1981): collect n encrypted messages, then decrypt all of them, an forward them to destination in random order. (n a parameter)
- **Attack:** $(n-1)$attack (or flooding attack) (Serjantov et al. 2002)
  - An adversary delays a legitimate message m,
  - generates and sends dummy messages until the mix flushes,
  - allows m to reach the mix,

Anonymous Communication Mechanisms

Flushing algorithms. Algorithms used by the mix to order the messages and attacks taking into account the flushing algorithm used.

Example.

- **Threshold mixes** (Chaum, 1981): collect n encrypted messages, then decrypt all of them, an forward them to destination in random order. (n a parameter)
- **Attack:** $(n-1)$attack (or flooding attack) (Serjantov et al. 2002)
  - An adversary delays a legitimate message m,
  - generates and sends dummy messages until the mix flushes,
  - allows m to reach the mix,
  - sends more dummy messages until the mix flushes again,

Anonymous Communication Mechanisms

Flushing algorithms. Algorithms used by the mix to order the messages and attacks taking into account the flushing algorithm used.

Example.

- **Threshold mixes** (Chaum, 1981): collect n encrypted messages, then decrypt all of them, an forward them to destination in random order. (n a parameter)
- **Attack:** $(n-1)$attack (or flooding attack) (Serjantov et al. 2002)
  - An adversary delays a legitimate message m,
  - generates and sends dummy messages until the mix flushes,
  - allows m to reach the mix,
  - sends more dummy messages until the mix flushes again,
  - uses knowledge on dummy messages to identify ms destination

Anonymous Communication Mechanisms

Mix network. Mix Cascade

- Before we have considered a single mix, but a series of them (a cascade) can also be used. The advantage of a cascade is that all mixes have to collaborate to break the anonymity. In other words, a single mix is able to provide secrecy.

Anonymous Communication Mechanisms

Low-latency anonymity systems. Interaction is needed: web browsing

- Crowds
- Onion routing

Anonymous Communication Mechanisms

Crowds. (Reiter, Rubin, 1998)

Anonymous Communication Mechanisms

Crowds. (Reiter, Rubin, 1998)

- A crowd is defined as a collection of users.

Anonymous Communication Mechanisms

Crowds. (Reiter, Rubin, 1998)

- A crowd is defined as a collection of users.
- When a user needs to transmit a transaction, it is either submitted directly or passed to a another member of the crowd.

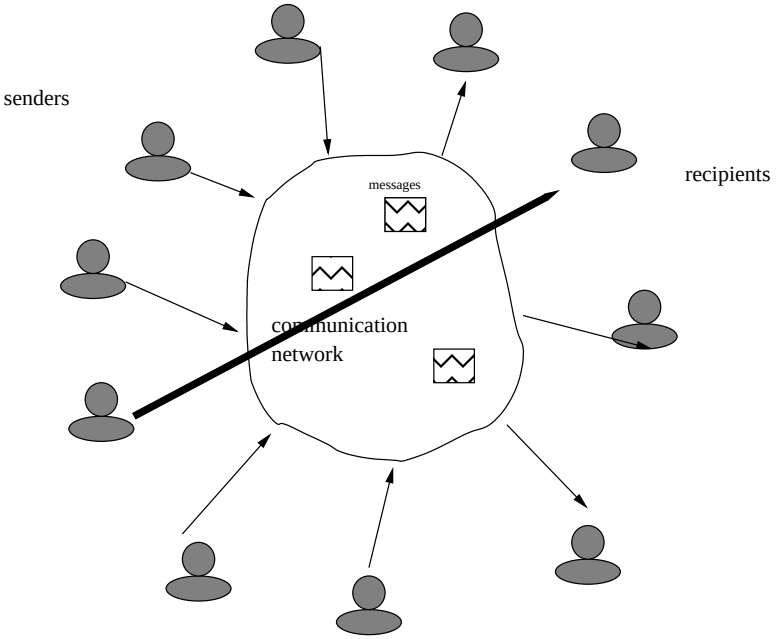
Anonymous Communication Mechanisms

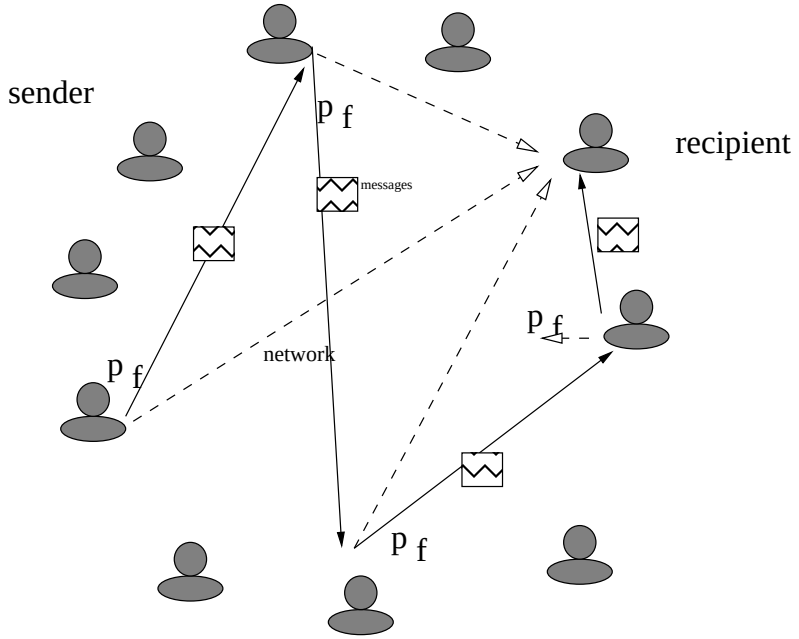
Crowds. (Reiter, Rubin, 1998)

- A crowd is defined as a collection of users.
- When a user needs to transmit a transaction, it is either submitted directly or passed to a another member of the crowd.
- Anonymity comes from the fact that users send some of their transactions but also transactions from other members of the crowd.

Anonymous Communication Mechanisms

Crowds. Graphical representation.





Anonymous Communication Mechanisms

Crowds. Steps of the process (I)

Anonymous Communication Mechanisms

Crowds. Steps of the process (I)

Starting. The user starts his local process (the jondo), that will represent in the crowd.

Anonymous Communication Mechanisms

Crowds. Steps of the process (I)

Starting. The user starts his local process (the jondo), that will represent in the crowd.

Contacting the server. The jondo contacts a server called the blender to request admitance in the crowd.

Anonymous Communication Mechanisms

Crowds. Steps of the process (I)

Starting. The user starts his local process (the jondo), that will represent in the crowd.

Contacting the server. The jondo contacts a server called the blender to request admitance in the crowd.

Admission. The blender admits the jondo and sends him the information needed to participate in the crowd.

Anonymous Communication Mechanisms

Crowds. Steps of the process (I)

Starting. The user starts his local process (the jondo), that will represent in the crowd.

Contacting the server. The jondo contacts a server called the blender to request admitance in the crowd.

Admission. The blender admits the jondo and sends him the information needed to participate in the crowd.

Jondo as proxy. The user selects this jondo as his web proxy for all services.

Anonymous Communication Mechanisms

Crowds. Steps of the process (II)

Request sent. Any request from the browser, is sent to the jondo.

Anonymous Communication Mechanisms

Crowds. Steps of the process (II)

Request sent. Any request from the browser, is sent to the jondo.

Processing. When a jondo receives a request, it forwards the request to another jondo with probability p_f. The receiver jondo is selected at random. Otherwise (with probability $1 - p_f$), the jondo submits the request to the end server (the final destination of the request). Note that this step is applied either for requests sent to the jondo by the user, or sent to the jondo by other jondos.

Anonymous Communication Mechanisms

Onion Routing. (Reed, Syverson, Goldschlag, 1998)

Anonymous Communication Mechanisms

Onion Routing. (Reed, Syverson, Goldschlag, 1998)

- Data and communication anonymous

Anonymous Communication Mechanisms

Onion Routing. (Reed, Syverson, Goldschlag, 1998)

- Data and communication anonymous
- Tor is the current implementation of onion routing
  http://www.onion-router.net/Publications/tor-design.pdf

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (I)

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (I)

Retrieval of known routers. The user accesses a directory server that provides known routers and their current state.

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (I)

Retrieval of known routers. The user accesses a directory server that provides known routers and their current state.

Construction of the path. The list of routers is used to define a path. That is, an ordered list of routers that will be traversed by the message.

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (II)

Onion. A data structure, called onion, is built for the message. Each layer of the onion defines one hop in the route. Public-key cryptography is used for this purpose. Let $E_{PK}(m)$ denote the encryption using the public key PK of the message m. Let $< i, m >$ denote that we send message m to router i. Then, if the message m is sent through the path traversing nodes 3, 5, and 6 with public keys PK_3, PK_5, and PK_6, the onion denoted by o will be something like:

$$o = E_{PK_3}(< 5, EP_{PK_5}(< 6, EP_{PK_6}(m) >) >)$$

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (III)

Message passed. The message is passed to the entry funnel, an onion router with a longstanding connection. The outermost layer of the onion is intended to this router. The router peels off its layer, identifies the next hop, and sends the embedded onion to that onion router.

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (III)

Message passed. The message is passed to the entry funnel, an onion router with a longstanding connection. The outermost layer of the onion is intended to this router. The router peels off its layer, identifies the next hop, and sends the embedded onion to that onion router.

- Peeling off the layer consists of applying D_{SK}, the decryption mechanism, to the message. Here SK is the private key of the router.

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (IV)

Message passed. Example

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (IV)

Message passed. Example

- In the case above, router 3 is the entry funnel. This router applies D_{SK_3} to the onion o obtaining:

$$\begin{aligned} D_{SK_3}(o) &= D_{SK_3}(E_{PK_3}(< 5, EP_{PK_5}(< 6, EP_{PK_6}(m) >) >)) \\ &= < 5, EP_{PK_5}(< 6, EP_{PK_6}(m) >) > \end{aligned}$$

Then, the onion $o' = EP_{PK_5}(< 6, EP_{PK_6}(m) >)$ is passed to router 5.

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (V)

Message forwarded. The same approach is applied subsequently by the other routers in the path.

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (V)

Message forwarded. The same approach is applied subsequently by the other routers in the path.

Message delivery. Once the message arrives to the last router in the path, this router (known as the exit funnel) using the information in m delivers the message to the appropriate address.

Anonymous Communication Mechanisms

Onion Routing. Steps of the process (VI)

On the reply. The process is reversed for data moving back to the original sender. In this case, each router encrypts the data using the private keys. The receiver (the individual that initiated the communication) will decrypt the message using the original path and the public keys of the routers in the path. Following with the example above, the following answer will be received by the original sender:

$$m' = E_{SK_3}(E_{SK_5}(ES_{SK_6}(answer)))$$

Applying

$$m' = D_{PK_6}(D_{PK_5}(D_{PK_3}(E_{SK_3}(E_{SK_5}(ES_{SK_6}(answer))))))$$

Anonymous Communication Mechanisms

Onion Routing. Onion Routing vs. Mix (cascades of mixes)

Anonymous Communication Mechanisms

Onion Routing. Onion Routing vs. Mix (cascades of mixes)

- Routers of onion routing "are more limited in the extent to which they delay traffic at each node".

Anonymous Communication Mechanisms

Onion Routing. Onion Routing vs. Mix (cascades of mixes)

- Routers of onion routing "are more limited in the extent to which they delay traffic at each node".
- In onion routing, all routers are entry points, and traffic entering or exiting the nodes may not be visible.

Anonymous Communication Mechanisms

Onion Routing. Onion Routing vs. Mix (cascades of mixes)

- Routers of onion routing "are more limited in the extent to which they delay traffic at each node".
- In onion routing, all routers are entry points, and traffic entering or exiting the nodes may not be visible.
- Another difference is the use of cryptographic operations. To be fast, onion routing uses also symmetric cryptography

Anonymous Communication Mechanisms

Onion Routing. Attacks to Tor (September 2013)

Wired: http://www.wired.com/threatlevel/2013/09/freedom-hosting-fbi/

FBI Admits It Controlled Tor Servers Behind Mass Malware Attack

- By Kevin Poulsen
- 09.13.13
- 4:17 PM

Follow @kpoulsen

Anonymous Communication Mechanisms

Onion Routing. Attacks to Tor (September 2013)

Wired: http://www.wired.com/threatlevel/2013/09/freedom-hosting-fbi/

Photo: Andrew Hart/Flickr

It wasn't ever seriously in doubt, but the FBI yesterday acknowledged that it secretly took control of Freedom Hosting last July, days before the servers of the largest provider of ultra-anonymous hosting were found to be serving custom malware designed to identify visitors.

Freedom Hosting's operator, Eric Eoin Marques, had rented the servers from an unnamed commercial hosting provider in France, and paid for them from a bank account in Las Vegas. It's not clear how the FBI took over the servers in late July, but the bureau was temporarily thwarted when Marques somehow regained access and changed the passwords, briefly locking out the FBI until it gained back control.

Anonymous Communication Mechanisms

Unobservability. Undetectability and anonymity against other subjects

- Dining cryptographer networks

Anonymous Communication Mechanisms

Dining Cryptographer network. DC-net (Chaum, 1985)

Anonymous Communication Mechanisms

Dining Cryptographer network. DC-net (Chaum, 1985)

- Sender anonymity, or a secure multi-party computation of the function OR.

Anonymous Communication Mechanisms

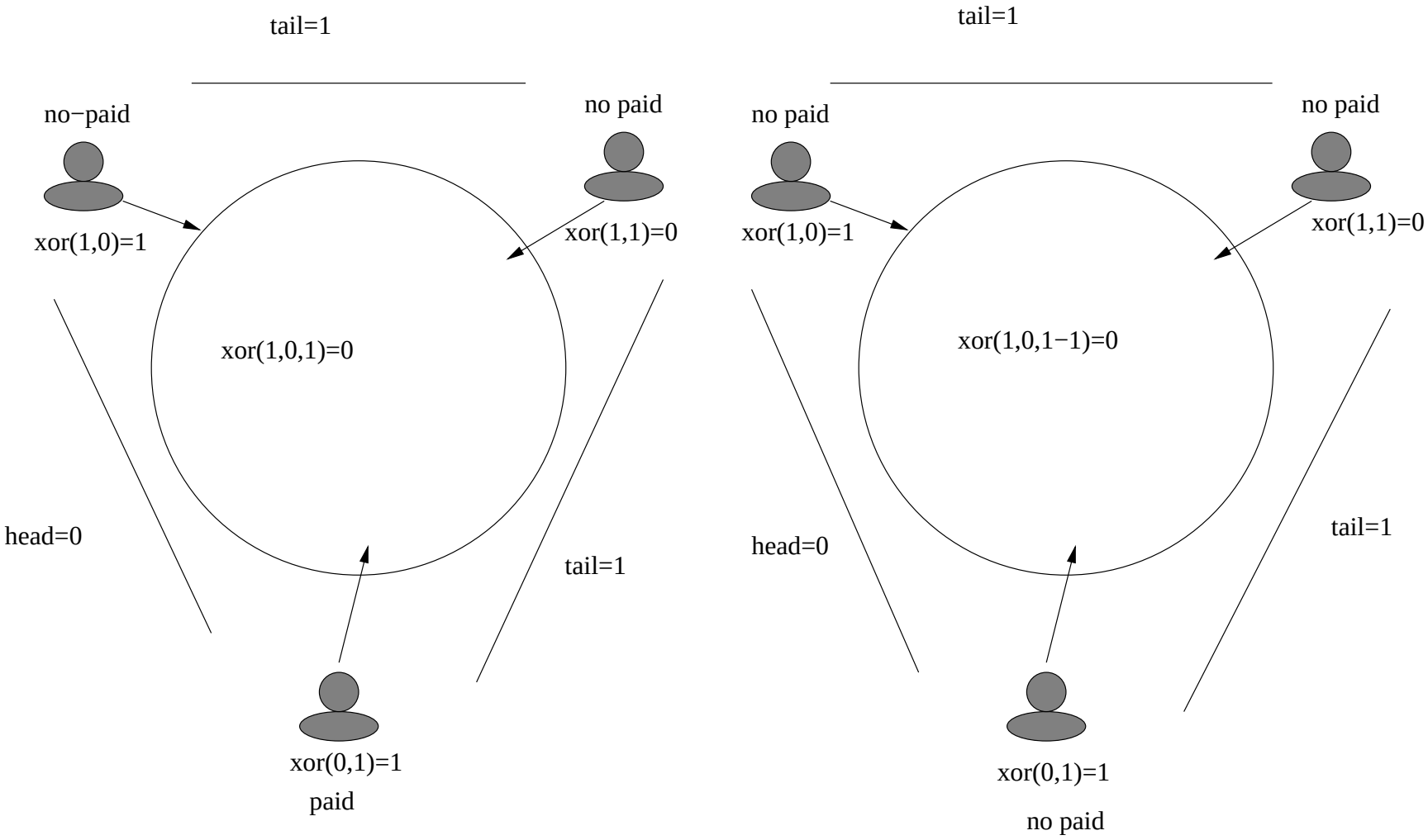
Dining Cryptographer network. DC-net (Chaum, 1985)

- Sender anonymity, or a secure multi-party computation of the function OR.
- **Problem.** Three cryptographers are sitting down to dinner at their favorite three-star restaurant. Their waiter informs them that arrangements have been made with the maître d'hôtel for the bill to be paid anonymously. One of the cryptographers might be paying the dinner, or it might have been NSA (U.S. National Security Agency). The three cryptographers repect each other's right to make an anonymous payment, but they wonder if NSA is paying.

Anonymous Communication Mechanisms

Dining Cryptographer network.

- Graphical representation of the solution
  (A cryptographer pays (left) and none of them pay (right))



Anonymous Communication Mechanisms

Dining Cryptographer network. Steps of the process (I)

Step 1. Each cryptographer flips a coin and shares its outcome with the crytographer on his right. Let us represent tails and heads by 1 and 0, respectively. Let $coin_i$ be the outcome of the coin of the ith cryptographer.

Anonymous Communication Mechanisms

Dining Cryptographer network. Steps of the process (I)

Step 1. Each cryptographer flips a coin and shares its outcome with the crytographer on his right. Let us represent tails and heads by 1 and 0, respectively. Let $coin_i$ be the outcome of the coin of the ith cryptographer.

Step 2. Each cryptographer finds whether the two coins he knows about (the one he flipped and the one his left-hand neighbor flipped) fell on the same side or not. Let us use the xor on the results of the two coins to represent the computation of the cryptographer: $c_i = xor(coin_1, coin_2)$.

Anonymous Communication Mechanisms

Dining Cryptographer network. Steps of the process (II)

Step 3. If a cryptographer is the payer, then he states the opposite of what he sees. Otherwise, says what he sees. Formally, let us represent the statement of the ith cryptographer by c_i', then

$$c_i' = \begin{cases} c_i & \text{if the } i\text{th cryptographer did not pay the meal} \\ 1 - c_i & \text{if the } i\text{th cryptographer paid the meal.} \end{cases} \quad (1)$$

Anonymous Communication Mechanisms

Dining Cryptographer network. Steps of the process (II)

Step 3. If a cryptographer is the payer, then he states the opposite of what he sees. Otherwise, says what he sees. Formally, let us represent the statement of the ith cryptographer by c'_i, then

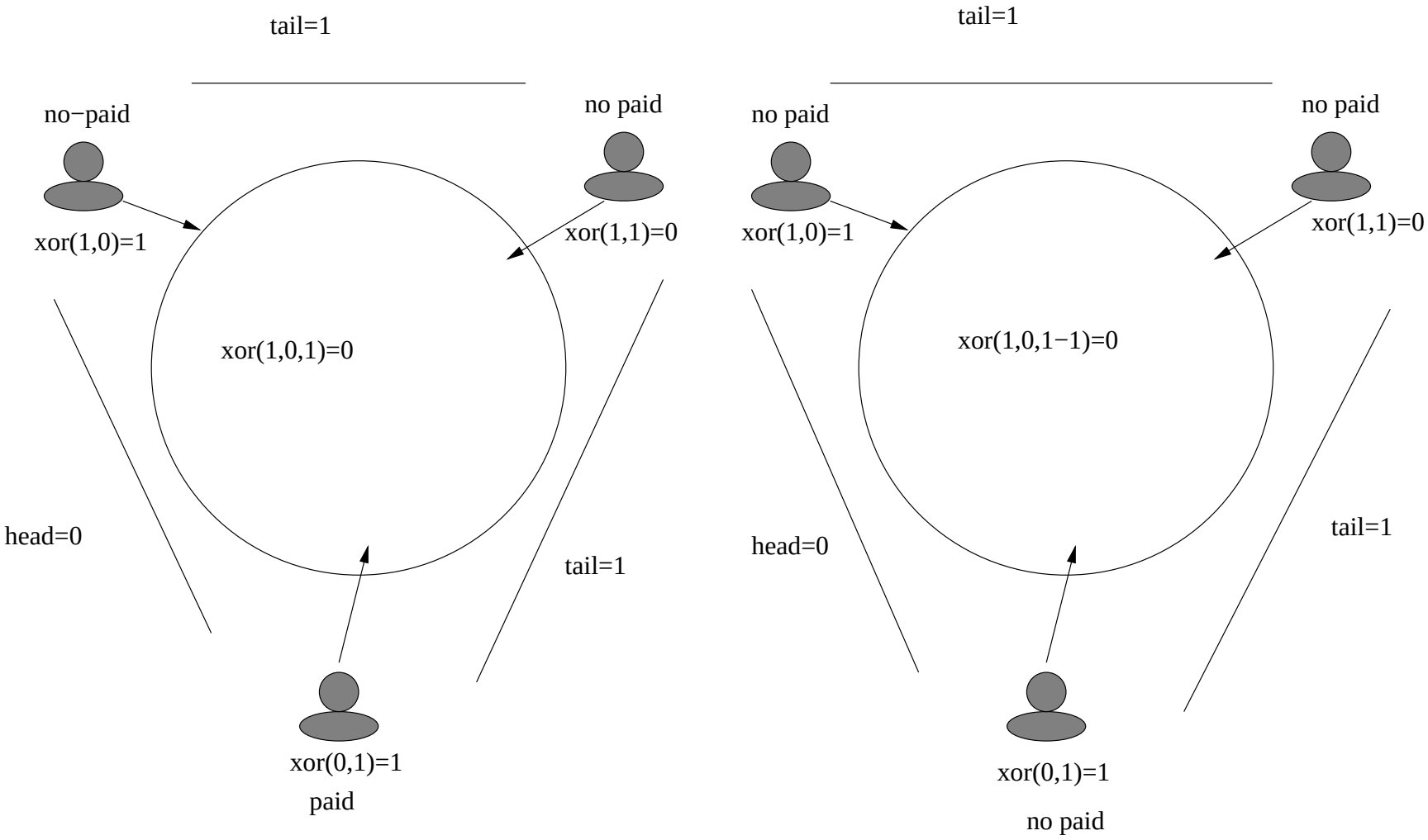
$$c'_i = \begin{cases} c_i & \text{if the } i\text{th cryptographer did not pay the meal} \\ 1 - c_i & \text{if the } i\text{th cryptographer paid the meal.} \end{cases} \quad (1)$$

Step 4. Then, let s be the sum of the values c'_i. If the sum is even, no one paid. If odd, one crytographer paid. The xor function can be used for this purpose.

Anonymous Communication Mechanisms

Dining Cryptographer network.

- Graphical representation of the solution
  (A cryptographer pays (left) and none of them pay (right))



Anonymous Communication Mechanisms

Dining Cryptographer network. Properties (I)

- This protocol can be generalized to an arbitrary number of crytographers.

Anonymous Communication Mechanisms

Dining Cryptographer network. Properties (I)

- This protocol can be generalized to an arbitrary number of crytographers.
  - **Protocol.** Each crytographer needs a secret bit with each other participant. Each cryptographer computes the sum modulo two (or the xor function of all the bits). Then, the ith cryptographer applies the function above to determine c_i' from c_i (as above). Then, as in Step 4 above, let s be the sum of the values c_i'. If the sum is even, no one paid. If odd, one crytographer paid.

Anonymous Communication Mechanisms

Dining Cryptographer network. Properties (II)

- Main problems: (i) malicious participants make the output useless; (ii) for n participants we need n^2 communications (one for each pair of participants).
- Only one participant can transmit a bit at a time. Two bits from different participants would cancel each other and would not be detected.

An Introduction to an ongoing research area in the field of privacy

Ongoing research

Areas. Privacy and security, privacy preserving data mining, statistical disclosure control

Topics. New types of data (data streams, time series, graphs), alternative definitions of privacy (k-anonymity, differential privacy)

Application-oriented. Social networks, documents (document sanitization)

Summary

Summary

Part I: Introduction to Privacy and Privacy-Enhancing Technologies

1. Introduction to Privacy
   - Definition of Privacy
   - Legal Aspects
   - Research Areas in Privacy
2. Anonymity and Pseudonyms
   - Identity and Identifiers
   - Types of Pseudonyms
   - Anonymity
3. Anonymous Communication Mechanisms
   - The Dining Cryptographers
   - MIX Networks
   - Crowds
   - Onion Routing and TOR
4. An Introduction to an ongoing research area in the field of privacy