

# Privacy Technologies

Claudia Diaz  
K.U.Leuven COSIC

# Outline

- What is privacy?
- Hard and soft privacy
- Privacy properties
- Privacy metrics
- Research challenges
- Conclusions

# What is privacy?

- Abstract and subjective concept, hard to define
- Dependent on cultural issues, study discipline, stakeholder, context
- Popular definitions:
  - “The right to be let alone”
    - Focus on freedom from intrusion
  - “Informational self-determination”
    - Focus on control

# Data protection

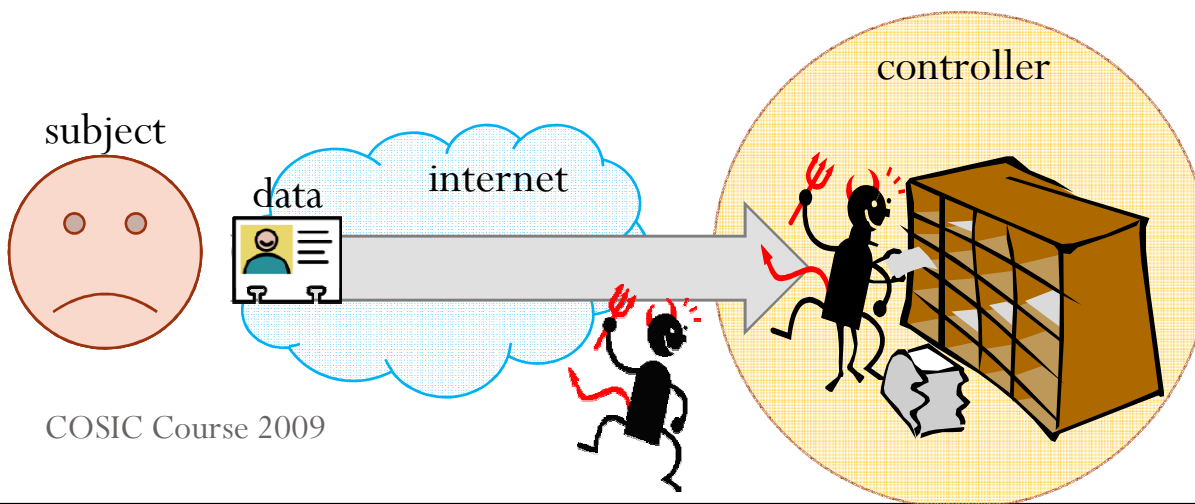
- Data collected for specific and legitimate **purpose**
- **Proportional**: adequate, relevant and not excessive (data minimization)
- With the subject's awareness and **consent**
  - Unless data is necessary for...
- Data subject's right to access, correct, delete her data
- Data security
  - Integrity, confidentiality of the data
  - Unfortunately, millions of records with personal data are breached every year
- Weak enforcement, low penalties
- USA: fair information practices
  - Many individual laws (HIPAA, California disclosure laws)

# Solove's taxonomy of privacy

- Information Collection
  - Surveillance
  - Interrogation
- Information Processing
  - Aggregation
  - Identification
  - Insecurity
  - Secondary Use
  - Exclusion
- Information Dissemination
  - Breach of Confidentiality
  - Disclosure
  - Exposure
  - Increased Accessibility
  - Blackmail
  - Appropriation
  - Distortion
- Invasion
  - Intrusion
  - Decisional Interference

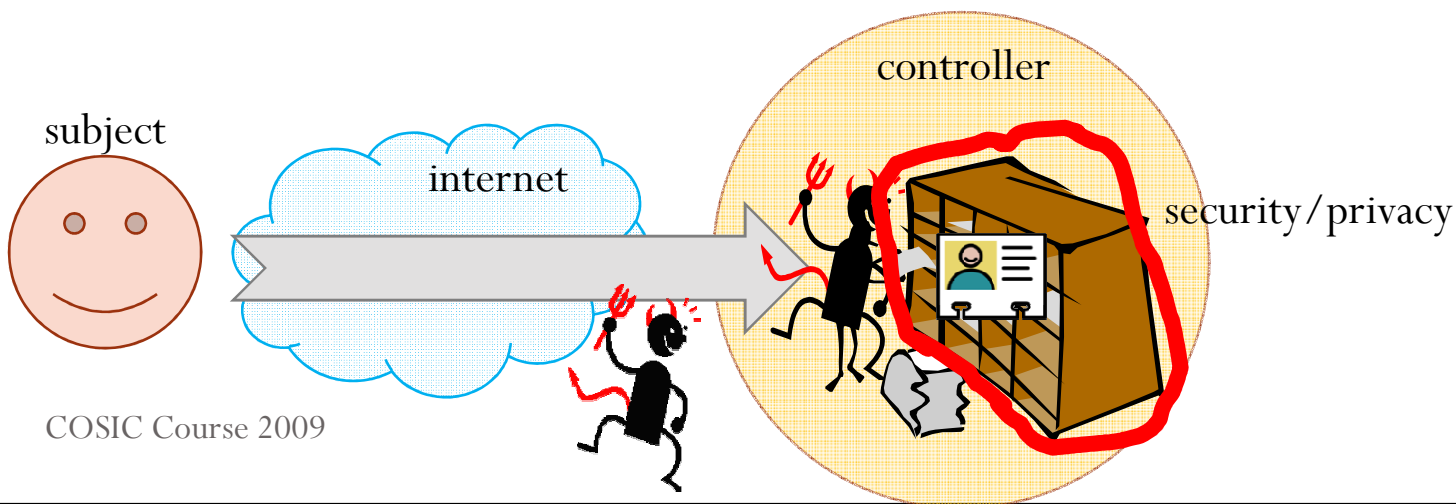
# Soft privacy

- System model
  - Data subject provides her data
  - Data controller responsible for its protection
- Threat model
  - External parties, errors, malicious insider



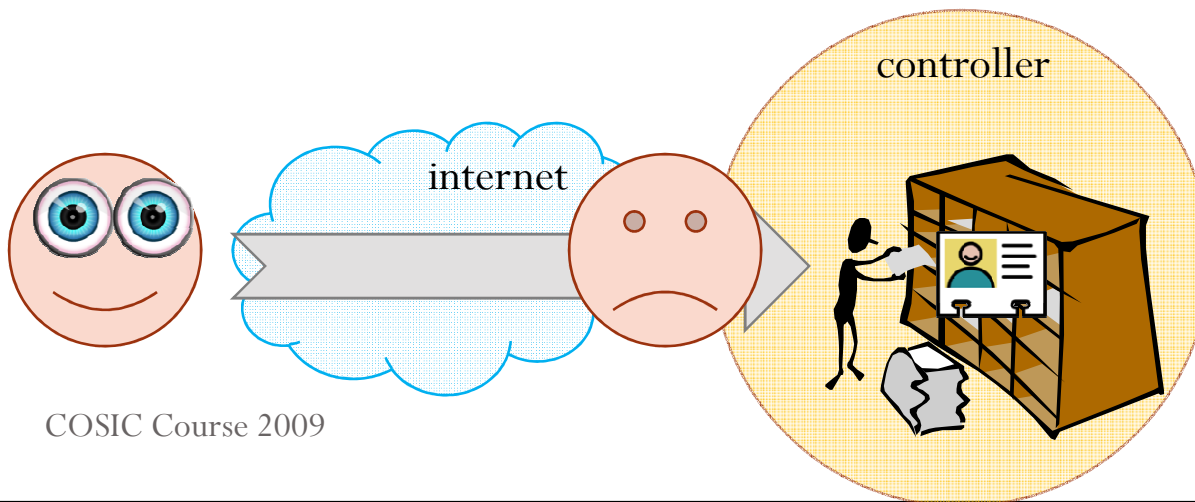
# Soft privacy

- Controller: main security “user”
- Policies, access control, audits (liability)
- Goal (data protection): purpose, consent, data security



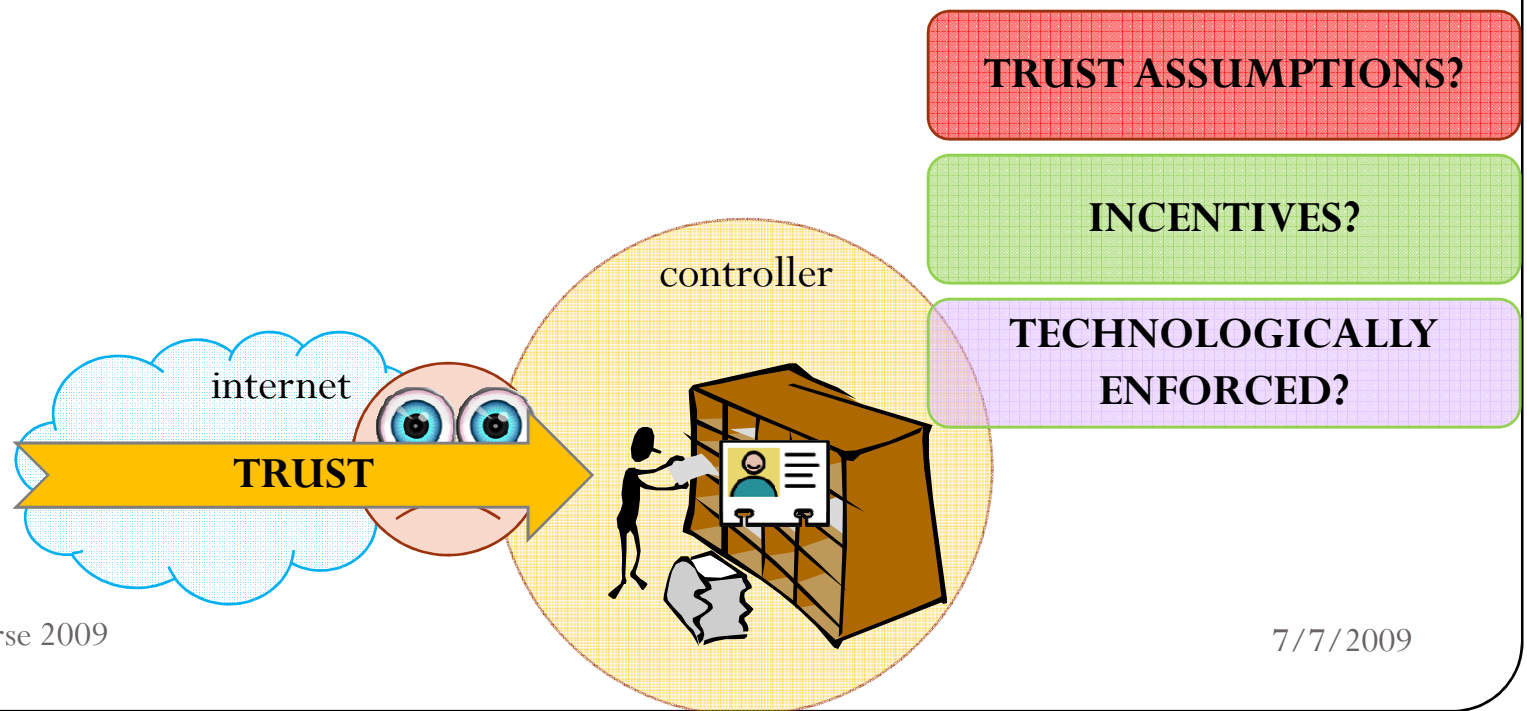
# Soft privacy

- Data subject has already lost control of her data
  - In practice, very difficult for data subject to verify how her data is collected and processed



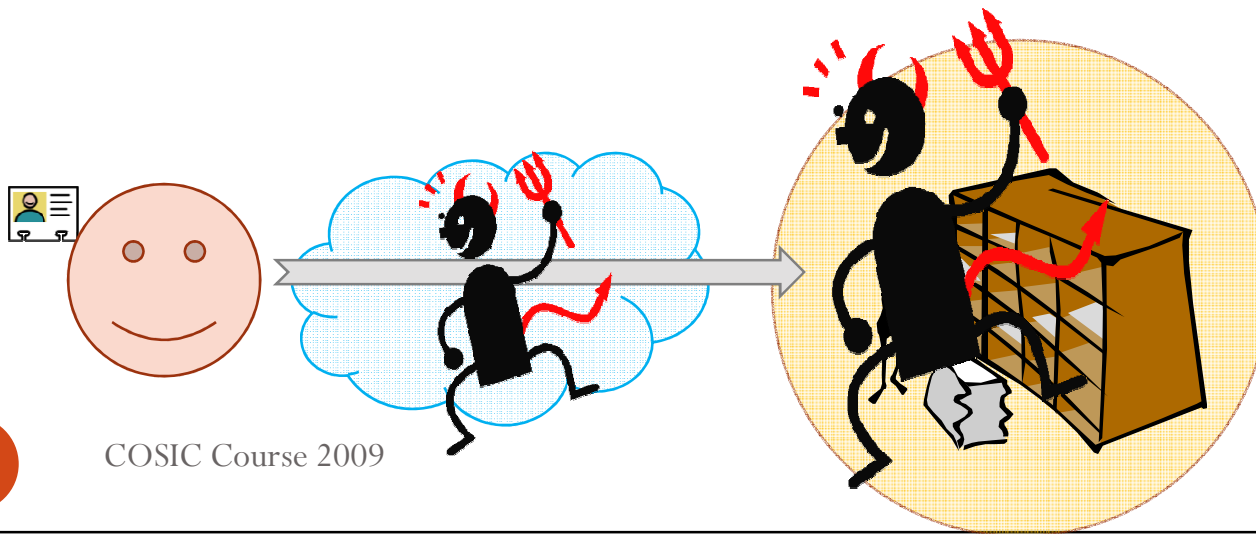
# Soft privacy

- Data subject has already lost control of her data
  - In practice, very difficult for data subject to verify how her data is collected and processed
  - Need to trust data controllers (honesty, competence) and hope for the best



# Hard privacy

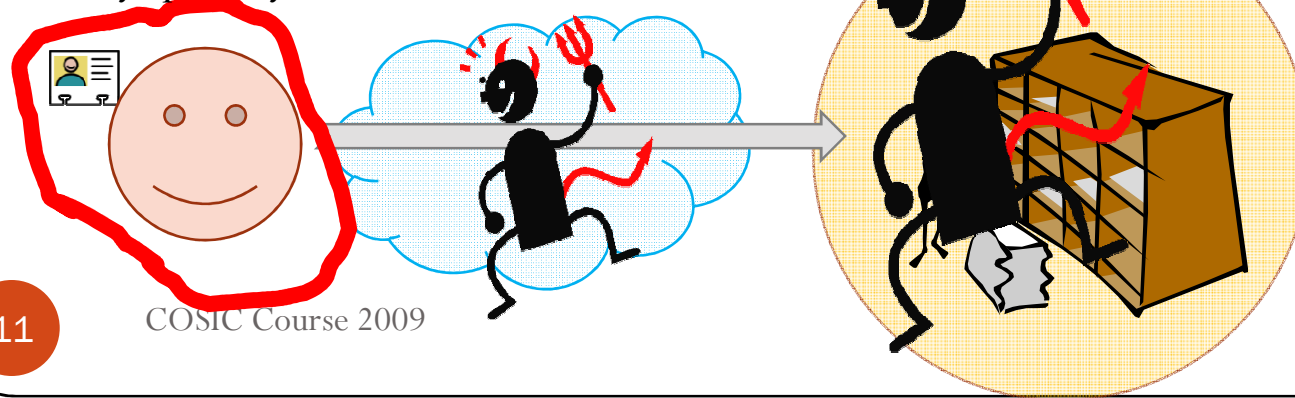
- System model
  - Subject provides as little data as possible
- Reduce as much as possible the need to “trust” other entities
- Threat model
  - Adversarial environment: communication provider, data holder
  - Strategic adversary with certain resources motivated to breach privacy (similar to security systems)



# Hard privacy

- Subject is an active security “user”
- Goal (data protection): data minimization
- Goal (Solove): protect against surveillance, interrogation, aggregation, identification
- Overview of hard privacy solutions

security/privacy



# Entity and attribute authentication

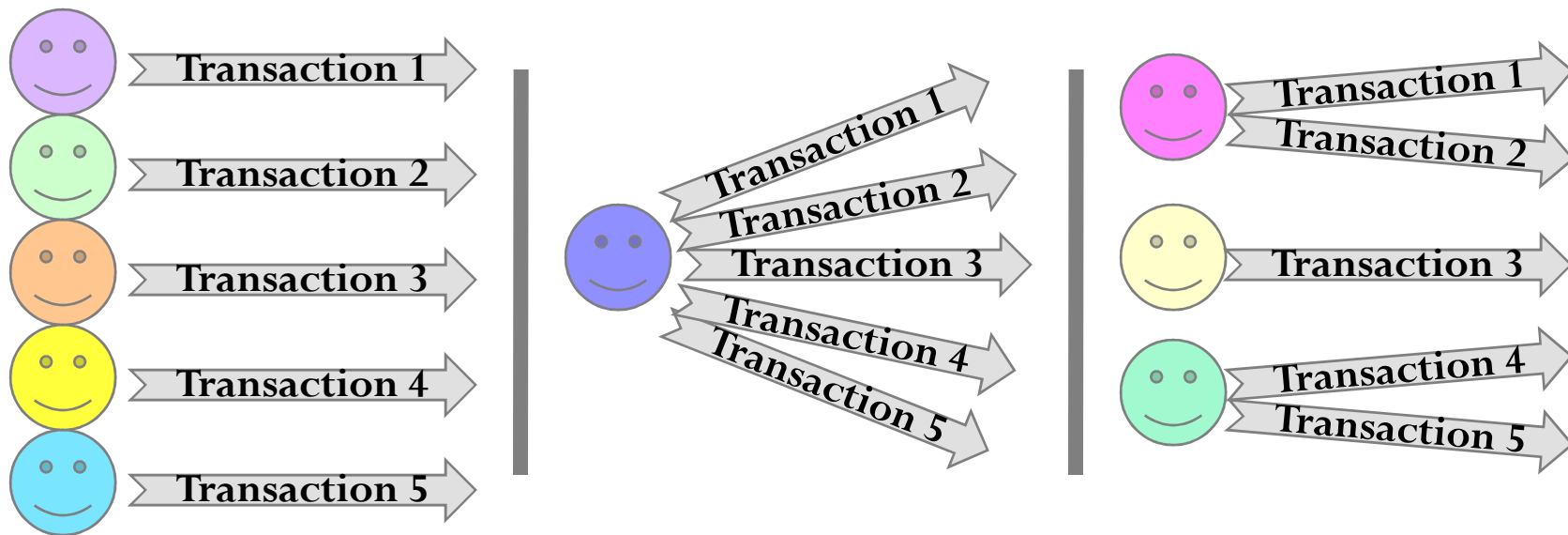
- Entity authentication often first step of a transaction
- Many transactions involve attribute verifications
  - ID documents: state certifies name, birth date, address, ...
  - Letters of reference: employer certifies your salary for your landlord
  - Club membership: the club certifies you are a gold member
- Credential: token that allows you to certify an attribute
- Entities
  - Issuer (State, Employer, Club)
  - Prover (holder of the credential)
  - Verifier (anyone)
- Property: Prover proves to the Verifier that she holds a credential with certain properties certified by the Issuer
- Properties: Unforgeability and Privacy

# Anonymous credentials

- Cryptographic protocols between  $\langle \text{Issuer, Prover, Verifier} \rangle$ 
  - Prover can prove that he holds a credential with certain attributes
  - or any expression on them (simple arithmetic, boolean) (e.g.  $\text{salary} > 30.000$  and  $\text{contract} = \text{permanent}$ )
  - Verifier gains no more information
  - Secure even if Issuer and Verifier collude (single/multiple show)
  - Security: cryptographic (Hard Privacy)

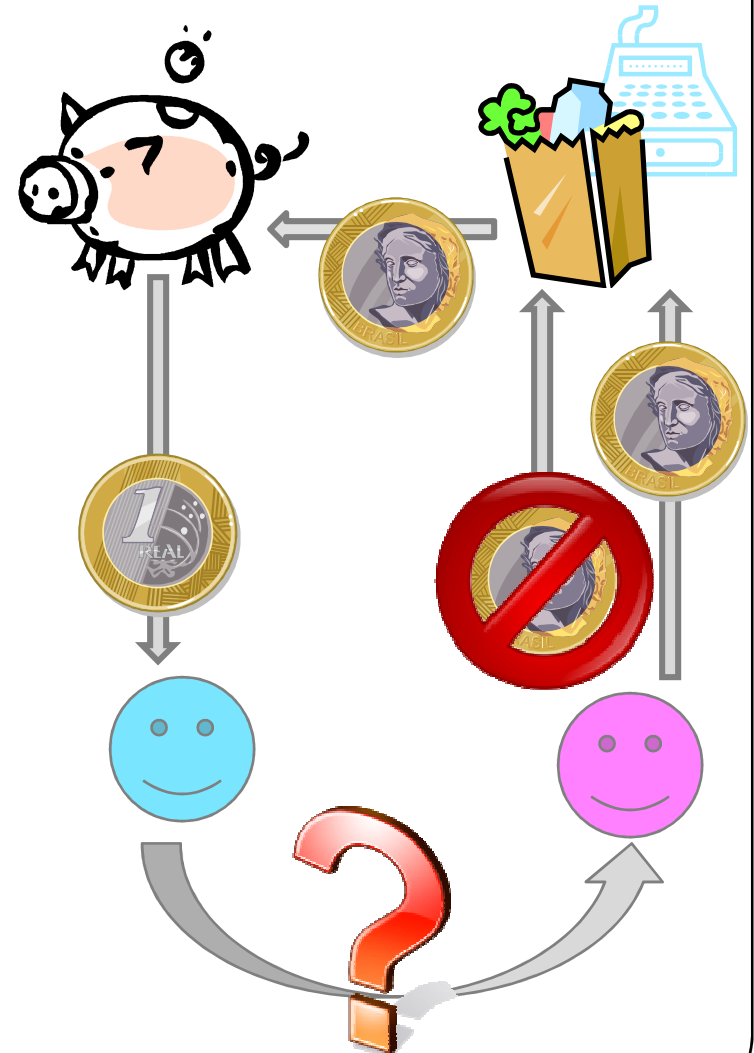
# Pseudonymous identity management

- One-time pseudonyms: anonymity
- Persistent pseudonyms: they become an identity
- Solutions in between: partial identities



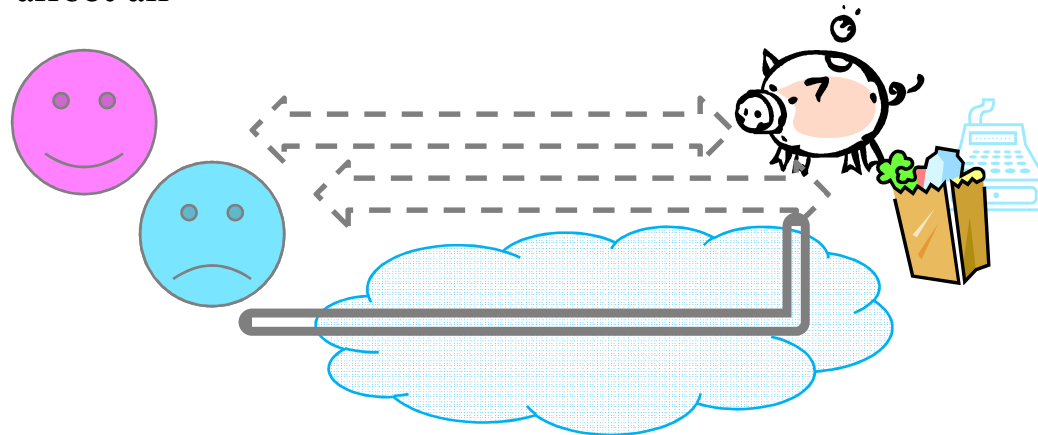
# Anonymous e-cash

- Secure and private payments
  - Cannot forge money or payments
  - with the anonymity of cash
  - Not just cash: cinema tickets
- Anonymous credentials can provide this
  - The bank certifies I have one euro
  - Payment: prover shows the credential, verifier accepts it
  - Verifier goes to the bank to deposit the coin
- Security properties:
  - Unforgeability
  - Privacy (for payer)
  - Double spending prevention!



# Communication infrastructure

- Applications assume that the **communication** channels are secured / maintain privacy properties
  - Example: previous protocols are useless if the adversary can link transactions based on traffic data (e.g., IP address)
- Secure channels
- Data confidentiality and integrity: same as traditional security
- Confidentiality of identities (**anonymity**) and relations (**unlinkability**):
  - Cryptographically: credential protocols
  - Network: protection against traffic analysis
- The infrastructure is **shared** by individuals, business, government, military, etc: privacy threats affect all



# Anonymous communications

- Anonymity / unlinkability **not** provided by default by the communication infrastructure
- **Traffic** data (origin, destination, time, volume): side channel information
  - Less volume than content: coarser, but highly valuable information
  - Formats that are easy to process for machines
  - Can be used to select targets for more intensive surveillance
  - Hard to conceal
- Adversarial:
  - **Third party** with access to the communication channels
  - **Recipient**: adversarial or trusted (subject can authenticate over the anonymous channel)

# Systems for anonymous communications

- Theoretical / Research
  - Mix networks (1981)
  - DC-networks (1985)
  - ISDN mixes (1992)
  - Onion Routing (1996)
  - Crowds (1998)
- Real world systems
  - Single proxy (90s): anon.penet.fi, Anonymizer, SafeWeb
  - Remailers: Cipherpunk Type 0, Type 1, Mixmaster(1994), Mixminion (2003)
  - Low-latency communication: Freedom Network (1999-2001), JAP (2000), Tor (2005)

# Attacks against anonymity systems

- Traffic Analysis: against vanilla or hardened systems
  - Extract information out of patterns of traffic (no content)
- Many adversary models are possible and realistic
- Passive attacks
  - Long-term intersection attacks
  - Traffic correlation / confirmation
  - Fingerprinting
  - Epistemic attacks (route selection)
  - Predecessor attacks
- Active attacks
  - N-1 attacks
  - Sybil
  - Traffic watermarking
  - Tagging
  - Replay
  - DoS

# Steganography and covert communications

- Encryption: hide data content
- Anonymity/unlinkability: hide identities / relations
- **Unobservability**: hide existence
- Communications:
  - Hide the fact that there is any communications
  - Embed a communication within another
  - Covert channels: hide secrets within public information
- Storage:
  - Hide the existence of files
  - Under coercion can deny there are any files to decrypt

# Censorship resistance

- How is that privacy technology?
  - Communities, tools, and techniques overlap.
  - Second definition: informational self-determination
  - Freedom (and techniques) to communicate, publish or access information
- Censorship resistance in communications:
  - Firewall busting techniques (national firewalls)
  - Peer-to-peer networking and file sharing (combine anon.comms, replicated storage, ...)
- Censorship resistance is the new availability!

# Other properties

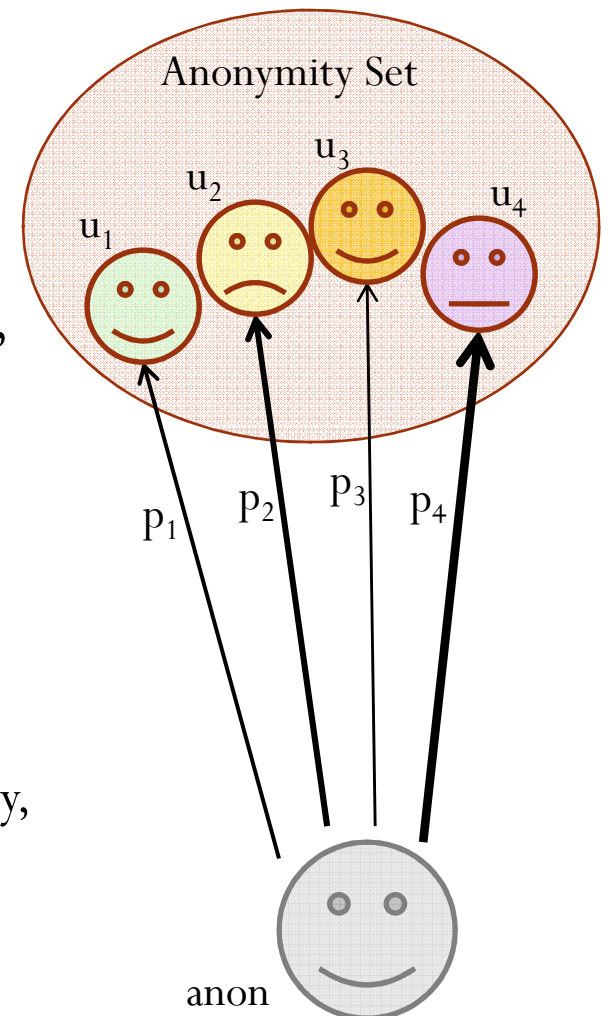
- **Forward security**
  - Ephemeral keys
  - Minimize consequences of security breach
  - Compulsion
- **Deniability** (repudiation)
  - Not possible to prove user knows / has said or done something
  - Communication: off-the-record property
  - Storage: compulsion resistance

# Data anonymization

- Anonymized data can be very useful, for example, for research purposes
  - Incidence of diseases: medical research
  - Social network structures: epidemiology, sociology
  - Optimization of services (e.g., transport or computer infrastructures)
- Measure the risk of **re-identification** of anonymized data:
  - Note: data protection does not apply to anonymized data
  - Records in an anonymized database
    - Medical data
    - Internet searches (AOL case)
  - Nodes in an anonymized social graph
- K-anonymity techniques

# Defining anonymity

- Definitions [PH00]
  - “*Anonymity is the state of being not identifiable within a set of subjects, the anonymity set.*”
  - “The *anonymity set* is the set of all possible subjects who might cause an action or be addressed.”
  - “Anonymity is the stronger, the larger the respective anonymity set is and the more evenly distributed the sending or receiving, respectively, of the subjects within that set is.”
- Probabilistic definition
  - Probabilistic definitions also possible for unlinkability, unobservability, deniability, ...

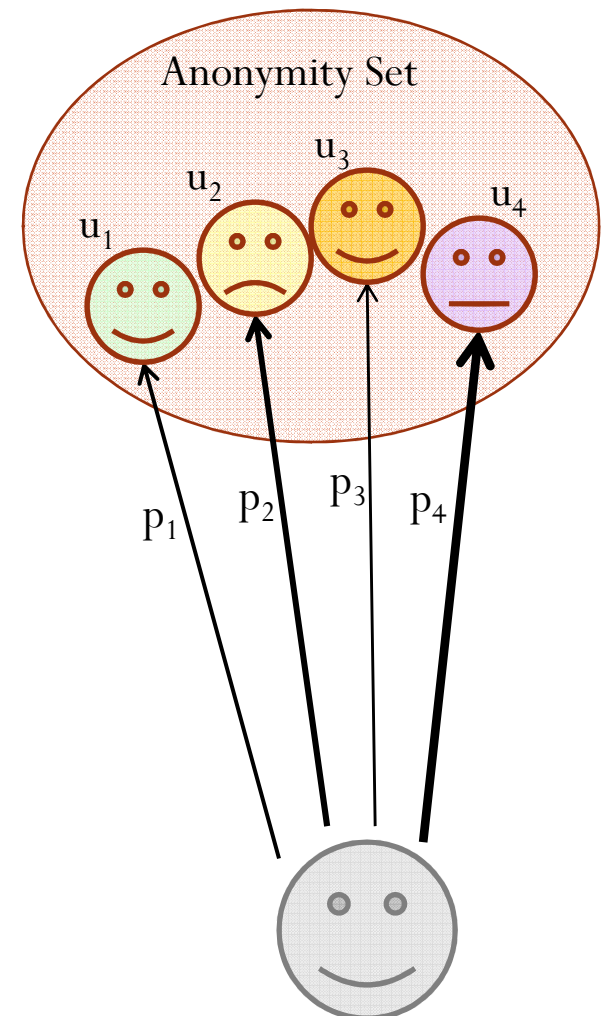


# Quantifying anonymity

- Anonymity depends on *both*:
  - The number of subjects in the anonymity set
  - The probability distribution of each subject in the anonymity set being the target
- Entropy: measure of the amount of *information* required on average to describe the random variable

$$H = -\sum_{i=1}^N p_i \cdot \log_2(p_i)$$

- Measure of the *uncertainty* of a random variable
- Increases with number N of possible values and with the uniformity of the distribution



# Privacy challenges

- Privacy requirements and privacy by design
- Finding robust and secure mechanisms
  - Proposed techniques keep on getting broken
  - Secure implementation is even harder
- Usability issues: ease of use, performance
- Economic incentives: tradeoffs privacy/cost (overhead, usability)
- Awareness and transparency

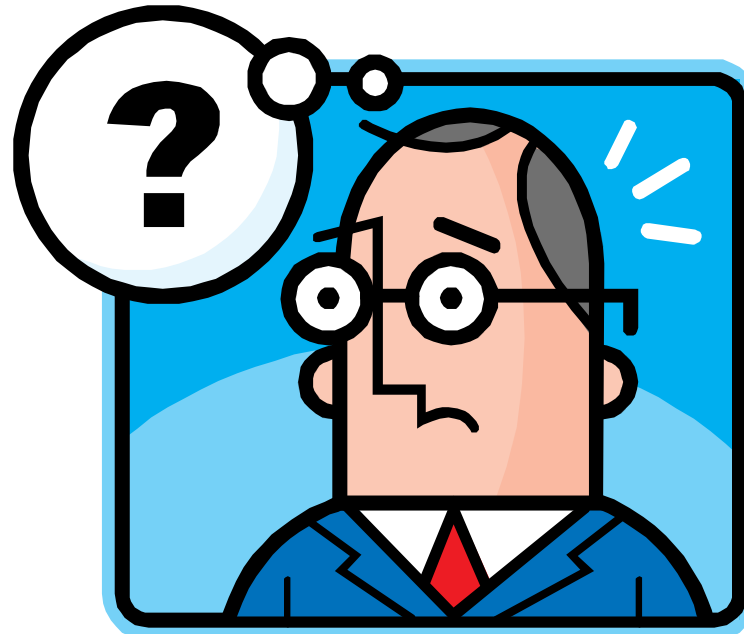
# New challenging scenarios

- Location privacy
- Ubiquitous environments
  - Principle of data maximization
  - Constrained devices
  - Securing the physical link
- Social networks: tension with data sharing
- Cloud computing: outsourcing of storage/computations

# Conclusions

- Privacy is not “opposed” to security, but rather a security property
- Compliance is a strong driver
  - Data Protection
  - US disclosure legislation
- Soft Privacy is the state of the art
  - Hidden costs of securing the data silos
- Hard Privacy solutions:
  - Active research
  - Poor deployment (cost)

Thanks !



<http://homes.esat.kuleuven.be/~cdiaz/>