Introduction to Privacy Technologies

Claudia Diaz

KU Leuven - COSIC

SecAppDev 2018

Outline

- What is privacy?

- Overview of various privacy technologies focusing on:
  - the concept of “privacy” they embed
  - their goals
  - their (trust) assumptions
  - their challenges and limitations

(Some) Definitions of Privacy

What is privacy?

- Abstract and subjective concept

- Dependent on:
  - Study discipline
  - Stakeholder
  - Social norms and expectations
  - Context

"The Right to Privacy" Warren & Brandeis (1890)

- Response to technological developments (photography and newspaper reporting on society gossip)
  - "Information which was previously hidden and private could now be shouted from the rooftops"

- Slander and libel laws (defamation) insufficient to protect the privacy of the individual because they "deal only with damage to reputation"
  - *Damnum absque injuria* ("loss without injury" and thus without legal remedy)

- Privacy as "the right to be let alone"
  - Right to prevent **publication** of information or stories about oneself
  - "protect the privacy of the individual from invasion either by the too enterprising press, the photographer, or the possessor of any other modern device for recording or reproducing scenes or sounds."

Westin (1970)

- Privacy as “the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others”

- Focus on the exercise of “control” over information about oneself

- “Informational self-determination” (German constitutional ruling, 1983)
  - “[...] in the context of modern data processing, **the protection of the individual against unlimited collection, storage, use and disclosure of his/her personal data** is encompassed by the general personal rights of the German Constitution. This basic right warrants in this respect the **capacity of the individual to determine in principle the disclosure and use of his/her personal data**. Limitations to this informational self-determination are allowed only in case of overriding public interest”

Agre and Rotenberg (1998)

- Privacy as “the freedom from unreasonable constraints on the construction of one's own identity”

- Not necessarily related to “data” or “technology” – yet useful to think about privacy in technologically-mediated environments:
  - Social media, targeted advertising
    - The construction of one's identity is mediated by “gaze of the other”
  - Profiling, surveillance
    - Chilling effects
  - Search results, clickbait, designs that create addiction exploiting behavioral biases
    - Propaganda, manipulation

- What is ‘unreasonable’? Who gets to define those boundaries? How?
  - Consumers? The market? Corporate interests and lobbyists? Regulators? Media? Experts? Civil society? Judges? Politicians? Society at large?

Solove’s taxonomy of privacy (2006)

- Harm-based approach based on Prosser’s taxonomy:
  - Intrusion upon the plaintiff’s seclusion or solitude, or into his private affairs.
  - Public disclosure of embarrassing private facts about the plaintiff.
  - Publicity which places the plaintiff in a false light in the public eye.
  - Appropriation, for the defendant’s advantage, of the plaintiff’s name or likeness.

- **Collection**
  - Surveillance
  - Interrogation

- **Invasion**
  - Intrusion
  - Decisional Interference

- **Processing**
  - Aggregation (big data inferences)
  - Identification (re-id of anon data)
  - Insecurity (lack of due care)
  - Secondary Use (purpose)
  - Exclusion (right of access)

- **Dissemination**
  - Breach of Confidentiality
  - Disclosure
  - Exposure (nudity)
  - Increased Accessibility (search)
  - Blackmail
  - Appropriation (id theft)
  - Distortion (defamation) (

Nissembaum (2004)

- Concept of privacy as “**contextual integrity**”
  - Importance of privacy not only for individuals, but for *society as a whole*

- The protection for privacy is tied to norms of specific contexts
  - Privacy is provided by *appropriate flows* of information.
  - Appropriate information flows are those that conform with *contextual information norms*
  - Contextual informational norms refer to five independent *parameters*: data subject, sender, recipient, information type, and transmission principle (e.g., confidentiality)
  - Conceptions of privacy are based on *ethical concerns* that evolve over time

- Contextual integrity is maintained when these are upheld:
  - Norms of **appropriateness**: what information is appropriate to reveal in a particular context
  - Norms of **flow** or **distribution**: what can be done with that information

- Norms may be explicit and specific, or implicit, variable and incomplete

Data Protection (European law)

- Applies to “Personal data”
  - “any information relating to an identified or identifiable natural person (data subject)”

- Principles:
  - Transparency
    - Informed consent of the data subject, access rights
  - Legitimate purpose (purpose limitation):
    - Personal data can *only* be processed for specified explicit and legitimate purposes
  - Proportionality (data minimization)
    - Data must be adequate, relevant and not excessive in relation to the purposes for which they are collected and processed
  - Accountability of the data controller

General Data Protection Regulation (GDPR)

- GDPR will replace (May 2018) the EU Data Protection Directive (90s)
- Fines up to 10M€ or 2 % of global turnover
- Stronger requirements to prove having obtained data subject consent
- Obligation to report data breaches
- Requirements for appropriate technical and organizational safeguards
    - Ensuring confidentiality, integrity, availability and resilience of systems
    - Pseudonymization, encryption of personal data
- Privacy by design and by default

Rights of data subjects under GDPR

- Consent (withdrawal thereof)
- Access
- Rectification
- Erasure (right to be forgotten)
- Restriction of processing
- Data portability
- Right to lodge a complaint with a supervisory authority
- Right to be informed of the existence of automated decision-making (including profiling) as well as possible consequences

European Convention of Human Rights (ECHR)

- Emerged as a response to the excesses of totalitarian states in the 30s and 40s
  - Spirit: protect citizens from an overbearing/intrusive *state*
  - During the cold war: ‘western’ countries distinguished themselves from the ‘eastern block’ in that the population was *not* subject to pervasive surveillance, as it was behind the Iron Curtain

- 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.

Related concepts

- Intertwined with other rights and values
  - Freedom of expression
  - Freedom of association
  - Personal dignity: airport scanners that produce naked images
  - Autonomy: censorship, filter bubble, behavioral nudges
  - (Non-)discrimination: profiling, personalization
  - Personal safety: identity theft, stalking, hate crimes
  - Democracy: targeted political messaging exploiting psychological triggers, fake news, hate speech, totalitarian state

Privacy and Technology

Offline world → Online world

- Information is hard/costly to collect, store, search, and access
  - Conversation face-to-face
  - Letters in the post
  - Papers in an physical archive
  - Paying with cash
  - Following your movements
  - Knowing who your friends are
  - Looking for info in encyclopedia

- Information is easy/cheap to collect, store search, and process
  - Skype, instant messaging
  - Emails
  - Files in digital archive
  - Paying with credit card
  - Location tracking
  - “Online” friends
  - Searching in google, wikipedia

- A new reality in terms of:
  - Available data: low cost of collection, replication, transmission, dissemination
  - Computing power and analytics: low cost of aggregation, profiling, inferences
  - Increasingly… automated decision-making (issues of “algorithmic accountability”)

Nothing to hide?

- Solove: “The problem with the ‘nothing to hide’ argument is its underlying assumption that privacy is about hiding bad things.”

- “Part of what makes a society a good place in which to live is the **extent to which it allows people freedom from the intrusiveness of others.** A society without privacy protection would be suffocation.”

- Difference between “secret” and “private”
  - Your daily routine, your movements, who your friends are, what you said in a conversation, which books you read…
  - These may not be secret, but you may not be comfortable with making it public or with third parties knowing about it, analyzing it, extracting conclusions from it, making decisions based on it…

Privacy Technologies

- Aim to address / mitigate certain privacy concerns
  - While allowing us to enjoy the benefits of modern ICTs

- Broad range of technology-based solutions that differ widely
  - Privacy concerns that they address
  - Assumptions that they make
  - Goals they intend to achieve
  - Challenges and limitations

Data security technologies

- Required for protecting personal data under GDPR

- Examples:
  - Encryption of data at rest and in transmission
  - Secure authentication and authorization of employees handling personal data
  - Secure logging of data accesses and processing for audit purposes
  - Secure deletion of data no longer necessary
  - Purpose-based access control (compliance with purpose-limitation principles)

Data security technologies

- What makes these “privacy technologies” is their application to the protection of *personal data*, rather than national secrets or corporate information assets

- Trust model
  - Architecturally, the data controller is *the* trusted party in charge of securing the personal data of subjects
  - Protection against external parties
  - Protection against errors or malicious employees working for the controller
  - No *technological* protection against malicious controllers who want to (ab)use the information for unethical or illegitimate purposes

Consent, control, access

- Privacy friendly defaults: opt-in vs opt-out
- Usable privacy settings, dashboards for privacy management
- Usable interfaces enabling the exercise of subject access rights
- Clear, concise, understandable privacy policies: user studies on readability, standardized privacy labels
- Tools to make privacy policies easier to check: P3P
- Assistance for users with privacy-relevant decision making: “privacy nudges”

Consent, control, access

- Solutions focused on the user interface and user experience (Human-Computer Interaction)

- Trust model
    - The data controller is trusted to facilitate the provision of informed consent, the exercise of access rights, and the control of data subjects over their personal data
    - Protection against user unawareness, mismatch between user preferences and system configuration, unexpected undesirable outcomes for the data subject
    - Limits on transparency posed by IP (proprietary software, algorithms, databases)
    - No *technological* protection against malicious controllers who want to mislead data subjects, disrespect their stated preferences, or (ab)use their information

Privacy in technology-mediated social contexts

- Unwanted disclosure of information to peers in social media
  - Studies on “regrets” and how to prevent them
  - Contextual feedback (e.g., audience of content, “how others see my profile”)

- Tensions between privacy and sharing
  - Attitudes are dependent on emotional states (fear and happiness)
  - Trigger reflection and self-consciousness (e.g., “timer nudge”, “sentiment nudge”)

- Emergence of privacy practices that provide adequate behavioral templates, signals and expectations in online social spaces
  - Support privacy-respectful social dynamics through technology design

Privacy in technology-mediated social contexts

- Again, solutions focused on improving user experience (HCI methods)

- Trust model
    - The social media platform is trusted with the private data of its users
    - Protection against accidental or unwanted disclosure of personal information to social peers
    - Protection against oneself making bad, regrettable decisions

- Other limitations
    - Top-down decision making regarding technological design of the medium
    - Focus on “privacy expectations”: slippery slope if expectations erode
    - Representativeness of user studies (mostly conducted in Europe and North America, mostly students)
        - Global networks with diverse subpopulations, diverse uses and practices, diverse privacy concerns and requirements

Data anonymity / Statistical Disclosure Control

- Tension between benefits of large-scale studies and private information of individuals being disclosed
  - Utility vs Privacy

- Data de-identification and sanitization
  - Releases the modified dataset: suppression, generalization of attributes
  - k-anonymity: ensure that a record may correspond to at least k individuals
  - Quasi-identifiers vs sensitive attributes

- Differential privacy mechanisms
  - Releases computations over the dataset (dataset kept with a trusted curator)
  - Add specially crafted noise to computations over data that bound what can be learnt about an individual from the result

Data anonymity / Statistical Disclosure Control

- Trust model
  - The curator or dataset holder is trusted with the raw data
  - Protection towards further data analysts or towards the public (open data)
  - No *technological* protection against malicious controllers who want to make unrestricted invasive private inferences about individuals in the dataset

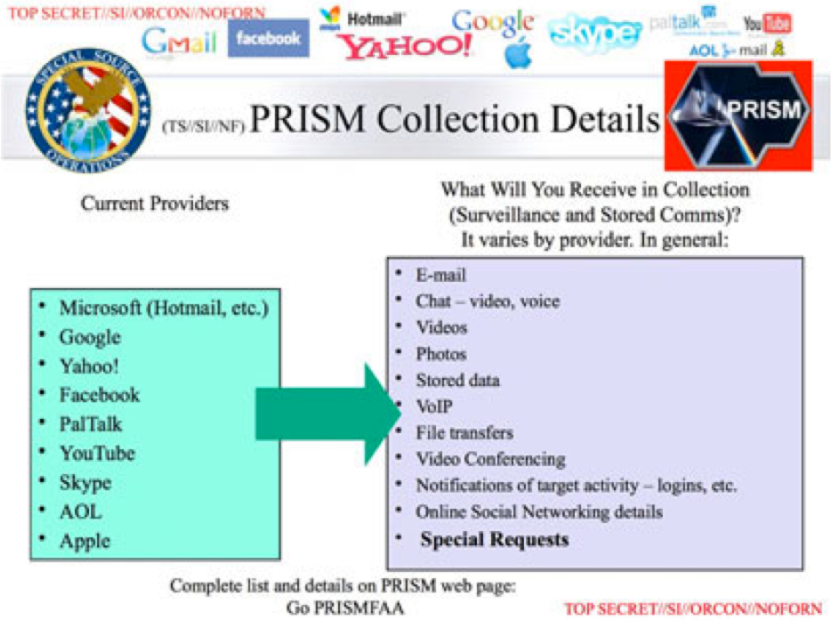
- Issues with k-anonymity
  - Sensitive attributes may also act as quasi-identifiers
  - Adversarial auxiliary knowledge not known
  - Weak guarantees in practice (data may be re-identifiable)

- Issues with differential privacy
  - No assumptions about adversarial auxiliary knowledge
  - Relative guarantees: privacy breaches can still occur if they derive from the utility of (what we can learn from) the dataset analysis results

Moving away from trusted single points of failure

- Previous technologies rely on the service provider for privacy protection
- Service providers may be
  - Malicious
  - Compromised
  - Coerced
  - Driven by incentives that are not in the best interest of the data subject
- Many privacy technologies aim to provide protection towards providers

Coercion and hacking by nation-state adversaries



TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) PRISM Collection Details

Current Providers

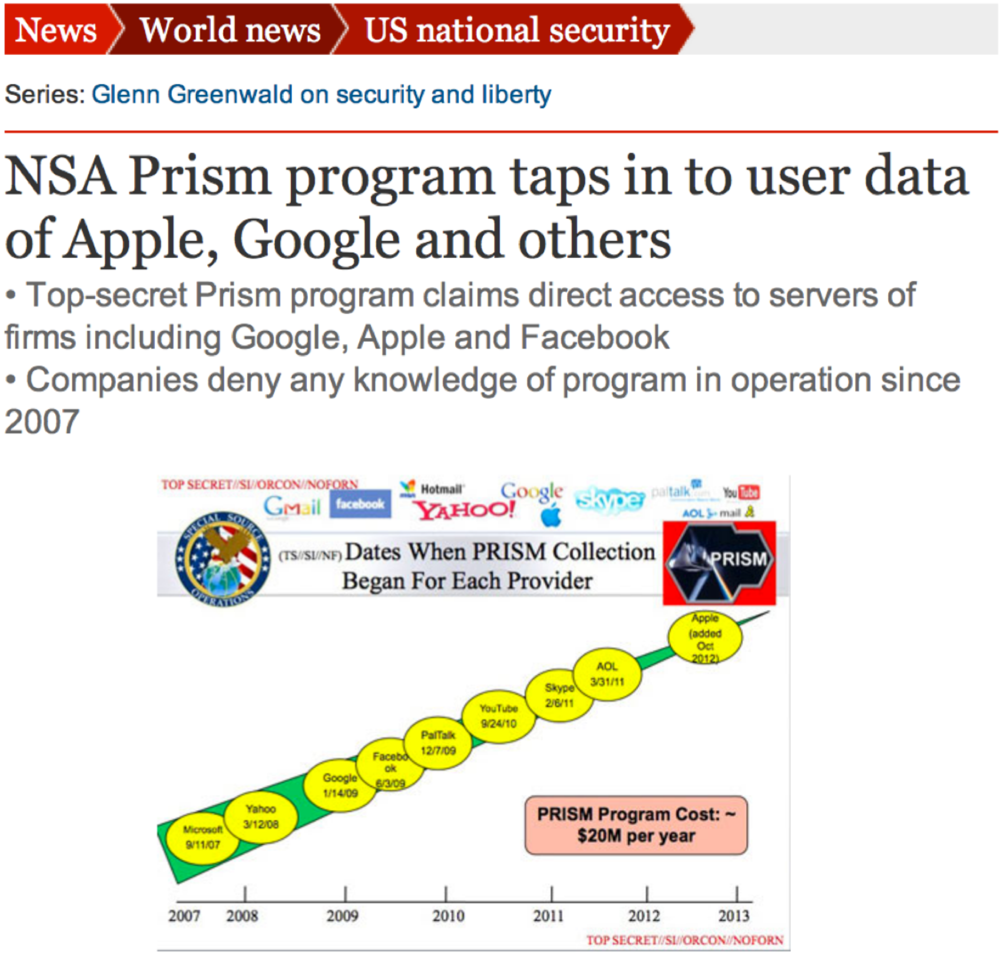
- Microsoft (Hotmail, etc.)
- Google
- Yahoo!
- Facebook
- PalTalk
- YouTube
- Skype
- AOL
- Apple

What Will You Receive in Collection (Surveillance and Stored Comms)? It varies by provider. In general:

- E-mail
- Chat – video, voice
- Videos
- Photos
- Stored data
- VoIP
- File transfers
- Video Conferencing
- Notifications of target activity – logins, etc.
- Online Social Networking details
- **Special Requests**

Complete list and details on PRISM web page:
Go PRISMFAA

TOP SECRET//SI//ORCON//NOFORN



News › World news › US national security

Series: Glenn Greenwald on security and liberty

NSA Prism program taps in to user data of Apple, Google and others

- Top-secret Prism program claims direct access to servers of firms including Google, Apple and Facebook
- Companies deny any knowledge of program in operation since 2007

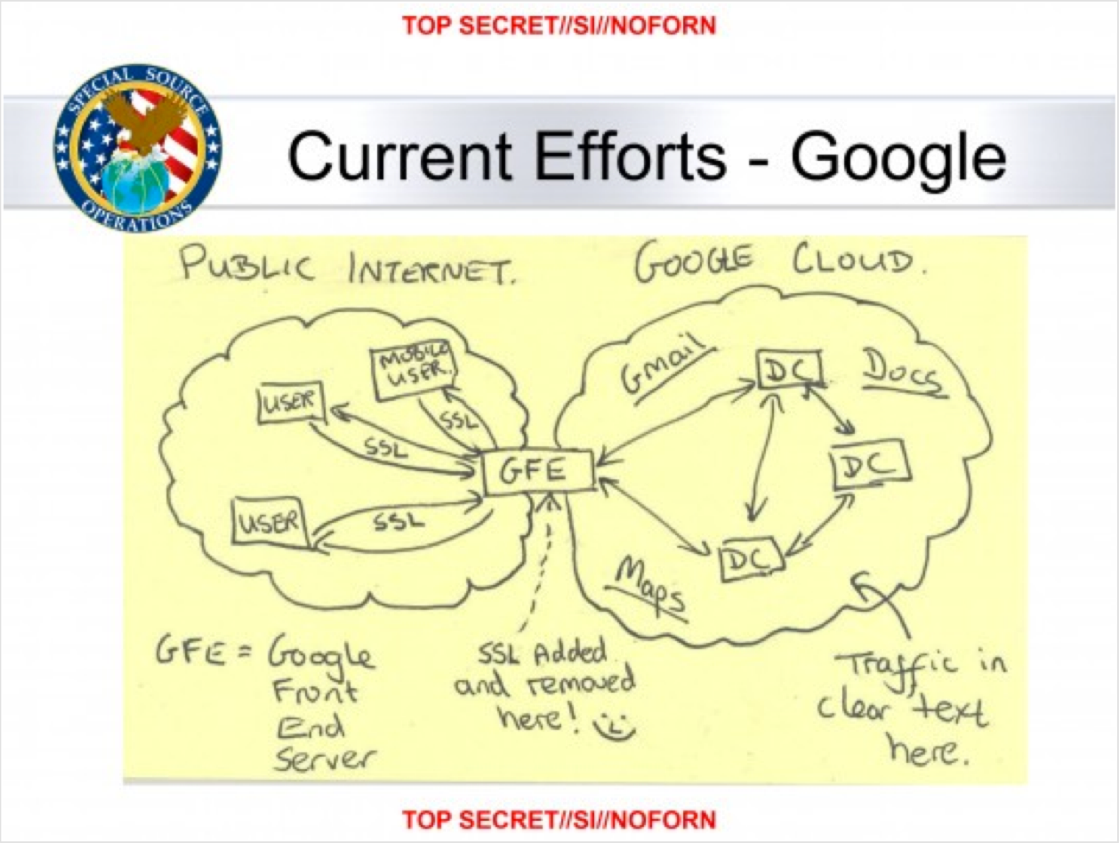
TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) Dates When PRISM Collection Began For Each Provider

PRISM Program Cost: ~ $20M per year

TOP SECRET//SI//ORCON//NOFORN

NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say



TOP SECRET//SI//NOFORN

Current Efforts - Google

PUBLIC INTERNET.
GOOGLE CLOUD.
USER
MOBILE USER.
USER
SSL
SSL
SSL
GFE
Gmail
DC
Docs
DC
DC
Maps
GFE = Google Front End Server
SSL Added and removed here! ☺
Traffic in clear text here.

TOP SECRET//SI//NOFORN

In this slide from a National Security Agency presentation on "Google Cloud Exploitation," a sketch shows where the "Public Internet" meets the internal "Google Cloud" where user data resides. Two engineers with close ties to Google exploded in profanity when they saw the drawing.

PrivEx: privacy preserving statistical data collection

- Based on distributed differential privacy and secure multiparty computations

- Add noise to individual inputs at collection time
  - The noise protects individual measurements
  - The noice cancels out with large-scale aggregation for accurate statistics

- Not possible to determine an individual value even if all other values are known
  - As long as at least one data collector and one tally server remain honest
  - Distribution of trust

Private messaging

- End to end encryption
  - Now available in popular messaging apps such as Signal or Whatsapp
  - Solutions for email (PGP) not as widespread or easy to use
- Service provider not able to read the information exchanged between users
  - Protection of the service provider itself towards coercion to disclose user communications
  - Law enforcement complaints of “going dark”
- Challenges with authentication and key management

Traffic analysis

- Even if content is encrypted…
- Valuable information revealed by meta-data:
  - Identities or call signs of communicating parties
  - Time, duration or length of transmissions
  - Location of emitter or receiver
- Can be exploited to infer content:
  - Web page fingerprinting (defeat TLS encryption)
  - SSH analysis to infer passwords
  - Reconstruction of encrypted VoIP conversations

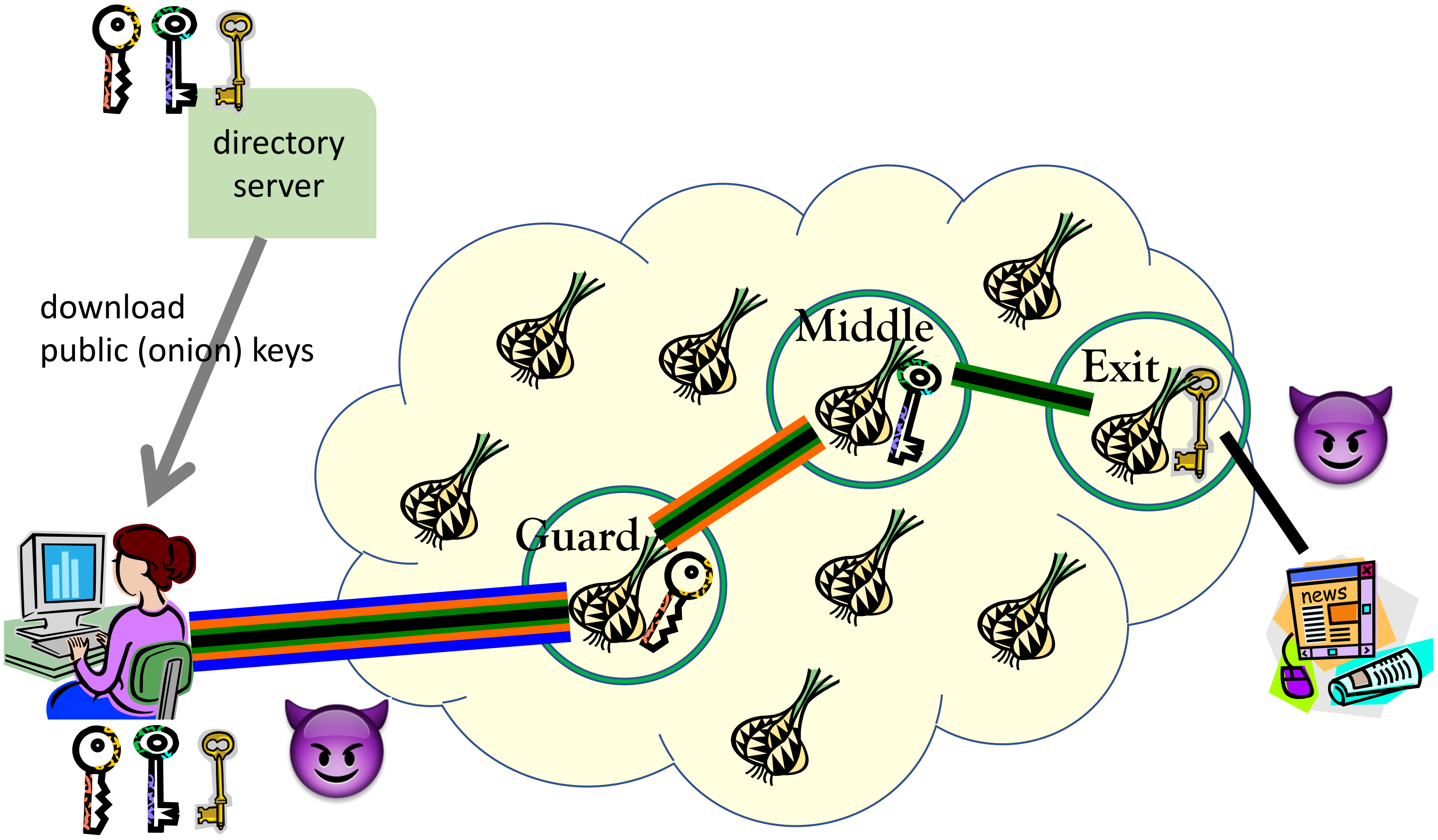
“Just Metadata”

- Diffie and Landau: ‘Privacy on the line’
  - “Traffic analysis, not cryptanalysis, is the backbone of communications intelligence”

- NSA General Counsel Stewart Baker:
  - *“Metadata absolutely tells you everything about somebody’s life. If you have enough metadata, you don’t really need content.”*

- General Michael Hayden, former director of the NSA and the CIA:
  - *“We kill people based on metadata.”*

Tor


directory server
download public (onion) keys
Middle
Exit
Guard
news

Anonymous communications

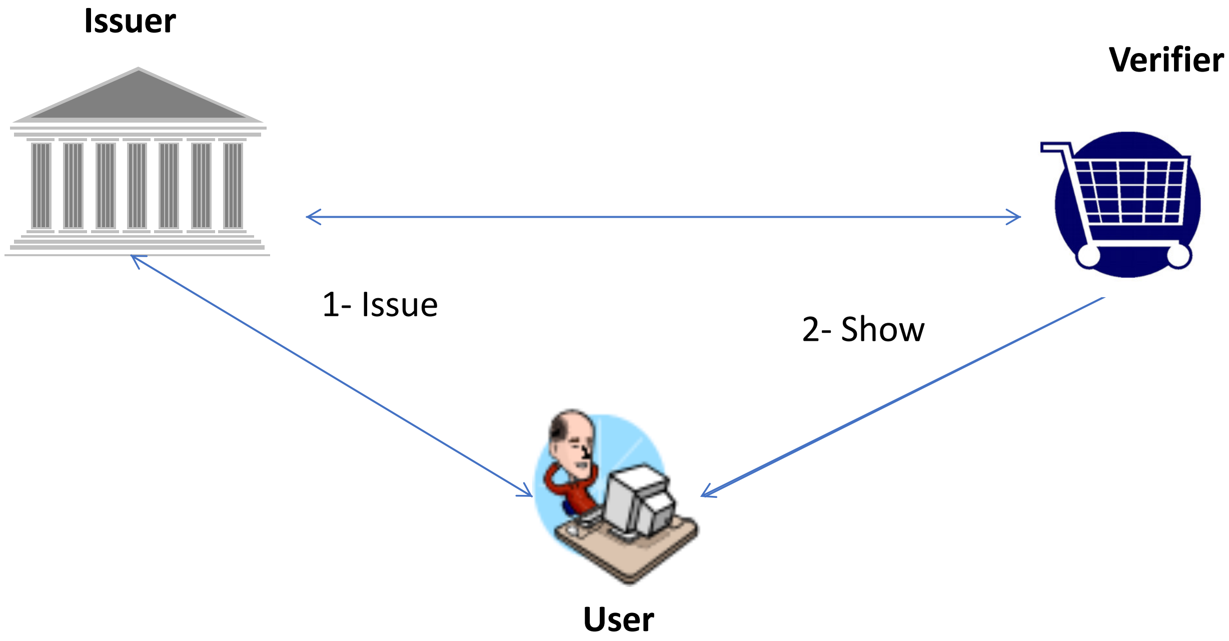
- Trust model
  - Distribute trust over multiple relays (no single point of failure)
  - At least some operators and some users are honest
  - The network infrastructure may be controlled by the adversary

- Limitations
  - Needs a critical mass of users to build an anonymity set
  - Needs diversity: of users, or operators, of uses
  - Anonymity is fragile: need to provide it at all layers
  - Statistical protection guarantees
  - Tradeoffs with performance and cost

Advanced crypto protocols

- Broad range of solutions offering strong privacy guarantees
- Minimizing information disclosure and trust assumptions
- Counterintuitive properties

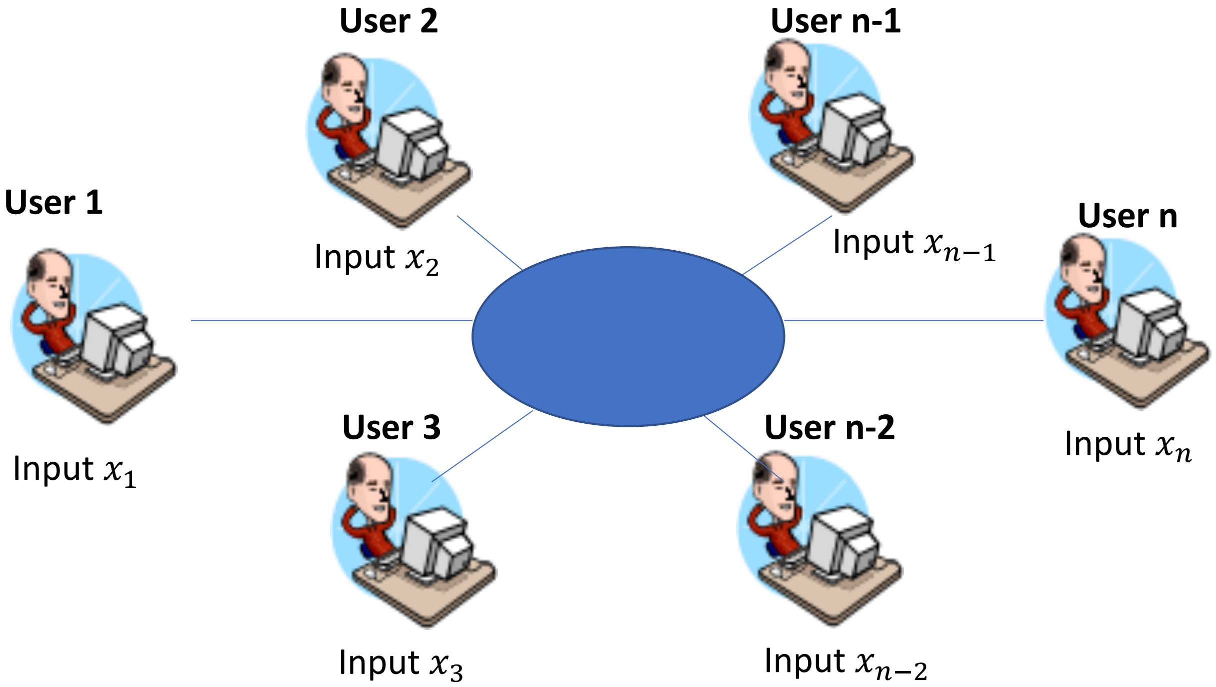
Attribute Based Credentials



Properties

- Unlinkability: Issue and Show, or several shows, cannot be linked by Issuer and Verifier.
- Unforgeability: Users cannot forge credentials
- Non-transferability: Users cannot pool their credentials
- Anonymity/Pseudonimity for users

Secure Multiparty Computation



User 2

Input x_2

User n-1

Input x_{n-1}

User 1

Input x_1

User n

Input x_n

User 3

Input x_3

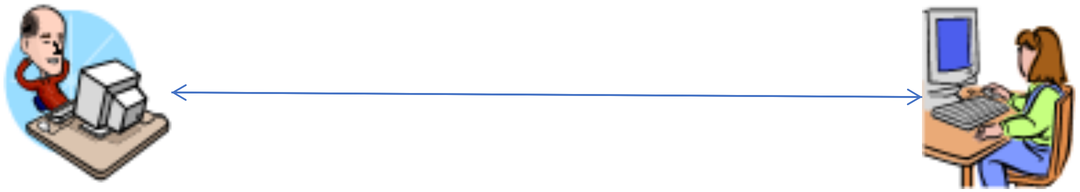
User n-2

Input x_{n-2}

Output $f(x_1, \ldots, x_n)$ for (a subset of) users

Property: **User i** does not learn $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$

Private Information Retrieval



User

Input i

Output m_i

Database

Input $m_1, \ldots, m_n$

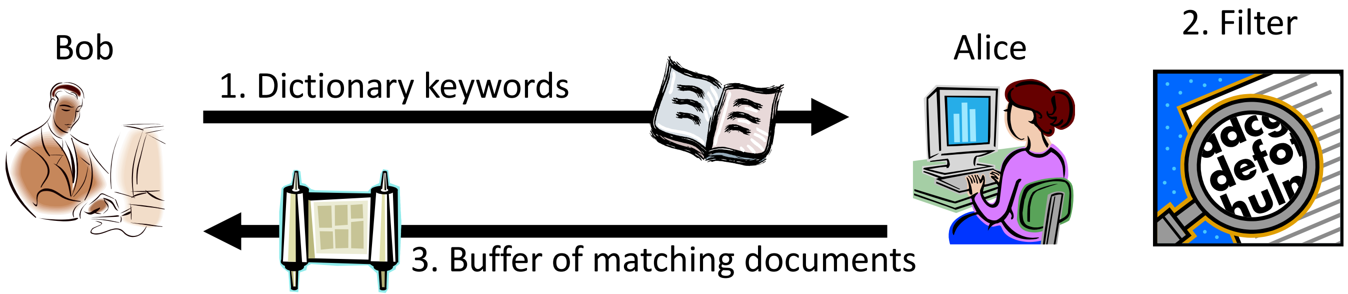
Output $\perp$

Privacy property: Database does not learn i

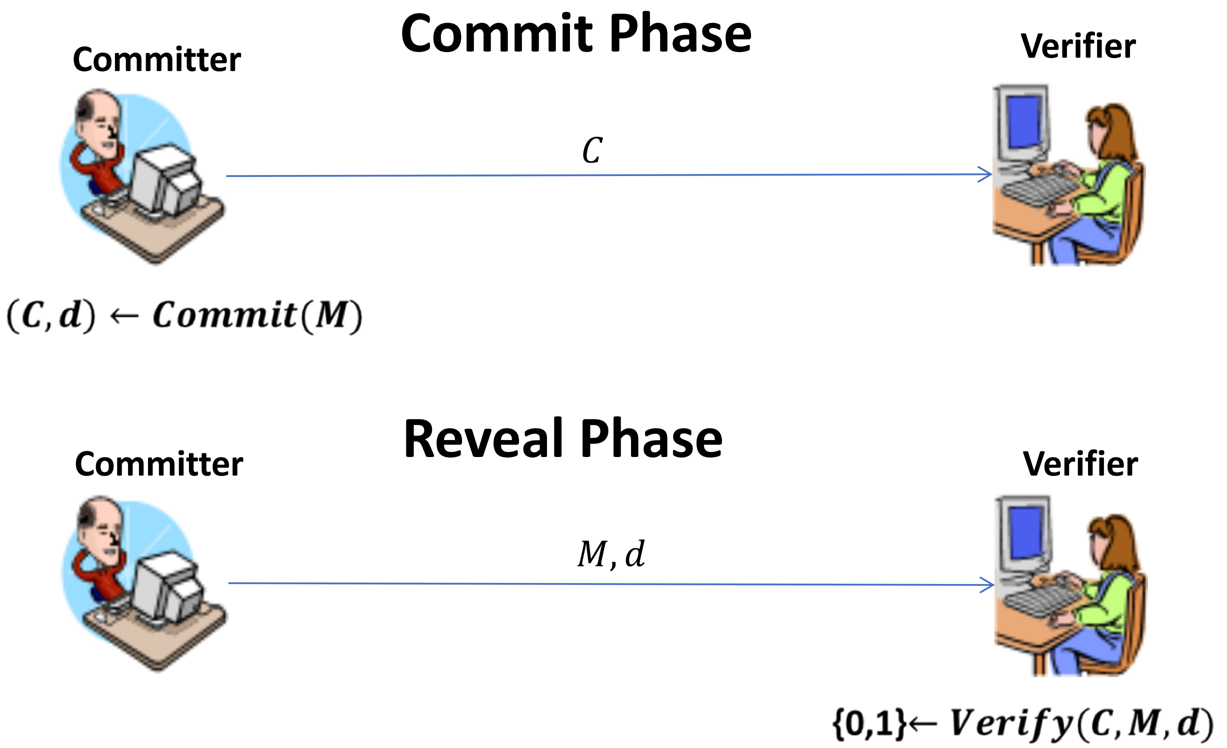
In Oblivious Transfer (aka ”symmetric PIR”): User *only* learns m_i

Private Search

- Alice stores documents
- Bob wants to retrieve documents matching some keywords
- Properties:
  - Bob gets documents containing the keywords
  - Alice does not learn Bob's keywords
  - Alice does not learn the results of the search
- Based on the Paillier cryptosystem (additive homomorphic)



Commitment Schemes



Commit Phase

Committer

Verifier

C

$(\boldsymbol{C}, \boldsymbol{d}) \leftarrow \boldsymbol{Commit}(\boldsymbol{M})$

Reveal Phase

Committer

Verifier

M, d

$\{0,1\} \leftarrow \boldsymbol{Verify}(\boldsymbol{C}, \boldsymbol{M}, \boldsymbol{d})$

Properties:

- Hiding
- Binding

M: message being committed to
d: opening value (randomness used for computing the commitment)
C: commitment

Advanced crypto protocols

- Trust model
  - Possibly adversarial service provider (honest-but-curious, malicious)
  - Limit information disclosure to what is strictly required by the functionality
  - Strong (cryptographic) guarantees
- Limitations
  - Complexity and limitations on functionality
  - Need for providers to integrate in their services
    - Different from unilateral approaches based on obfuscation

Summary

- Defining privacy is difficult: multidimensional concept
  - Do not assume that you ‘know’ what someone means when they say “privacy”
  - Technology has a huge impact on privacy
  - But privacy is not “just” about technology

- Privacy technologies: a diverse landscape in terms of
  - Privacy “concept”
  - Assumptions
  - Goals
  - Limitations

- Privacy technologies can enable online services while offering much better protection than currently deployed systems