Overview of Privacy Enhancing Technologies

Claudia Diaz
K.U.Leuven ESAT/COSIC
Outline

- Privacy by design: building blocks
- Privacy by design: example applications
  - Buying digital goods
  - Electronic petitions
  - Electronic Toll Pricing
Privacy by design

- What are the data necessary for the provision of the service?
  - *Data minimization principle*

- Is it necessary to collect any other data?
Anonymity vs Oblivious Transfer

- **Anonymity**
  - Service provider can observe access to the service
  - Cannot observe the identity of the user

- **Oblivious Transfer (OT) / Private Information Retrieval (PIR)**
  - Service provider can identify user
  - Cannot observe details of the access to the service
    - Which data records were accessed
    - Which search keywords were used
    - Which data was downloaded
  - ...

- All parties have assurance that the other participants in the protocol are cannot cheat
Authentication

- Entity authentication often first step of a transaction
- Makes sense in an organizational environment (government, military, even commercial)
- ...but what if there is no closed group?
- The **Identity Management** concept
The privacy-invasive way

- **Usual way:**
  - **Identity provider** certifies attributes of a **subject**.
  - **Identity consumer** checks those attributes
  - Match credential with **live person** (biometric)

- **Examples:**
  - E-passport: signed attributes, with lightweight access control.
    - Attributes: nationality, names, number, pictures, ...
  - Identity Cards: signatures over attributes
    - Attributes: names, date of birth, picture, address, ...
Idea behind credentials

- Many transactions involve attribute certificates
  - ID docs: state certifies name, birth dates, address
  - Letter reference: employer certifies salary
  - Club membership: club certifies some status

- Do you want to show all attributes for each transaction?

- Credential: token certifying attributes
  - Prover proves to the Verifier that she holds a credential with certain properties certified by the Issuer
Anonymous credentials

The players:
- Issuer (I) = Identity provider
- Prover (P) = subject
- Verifier (V) = identity consumer

Properties:
- The prover convinces the verifier that he holds a credential with attributes (certified by the Issuer) that satisfy some boolean formula:
  - Simple example “salary>30,000 AND contract= permanent”
- Prover cannot lie
- Verifier cannot infer anything else aside the formula
- Anonymity maintained despite collusion of V & I

Slide credit: George Danezis
1. Issuing protocol:
   Prover gets a certified credential.

2. Showing Protocol:
   Prover makes assertions about some attributes
   Name=Peggy, age=25, address=Cambridge, Status=single

   Cannot learn anything beyond age>18

Slide credit: George Danezis
# PKI vs anonymous credentials

<table>
<thead>
<tr>
<th>PKI</th>
<th>Anonymous credentials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signed by a trusted issuer</td>
<td>Signed by a trusted issuer</td>
</tr>
<tr>
<td>Certification of attributes</td>
<td>Certification of attributes</td>
</tr>
<tr>
<td>Authentication (secret key)</td>
<td>Authentication (secret key)</td>
</tr>
<tr>
<td>No data minimization</td>
<td>Data minimization</td>
</tr>
<tr>
<td>Users are identifiable</td>
<td>Users are anonymous</td>
</tr>
<tr>
<td>Users can be tracked (Signature linkable to other contexts where cert is used)</td>
<td>Users are unlinkable in different contexts</td>
</tr>
</tbody>
</table>
Anonymous e-cash [Chaum]

- Secure and private payments
  - Cannot forge money or payments
  - With the anonymity of cash
  - Not just cash: public transport

- 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!
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
What is Traffic Analysis

- Making use of (merely) the traffic data of a communication to extract information.
- As opposed to ‘interception’ or ‘cryptanalysis’.

- What are traffic data?
  - Identities or call signs of communicating parties.
  - Time, duration or length of transmissions.
  - Location of emitter or receiver.
  - No content – it may be encrypted.
Traffic analysis

- Diffie & Landau statement – ‘Privacy on the line’ on the politics of encryption.
  - “Traffic analysis, not cryptanalysis, is the backbone of communications intelligence.”

- Could this become true on the Internet? Is it already?
  - Different transport technologies: cable, wireless, satellite, ATM, ethernet, . . .
  - Common routing protocols – they expose traffic data.
  - New technologies: wireless, overlay networks, convergence with telephone – more opportunities for collecting traffic data.
Can the Secure SHe11 (SSH) protect your privacy?

- SSH is used for secure remote login and file transfer. All data is encrypted and authenticated. What information can we extract about a password typed in a protected session?
  - Key observation: each key pressed is transmitted separately.
  - Depending on the position of the key on the keyboard, different inter key timings.
  - Attack (Song et al.): observe the inter key timings (many times if you wish) – infer what keys have been pressed.
  - Result: reduce the entropy of password – fewer guesses required.

- Note that there is still variability across different people. Adds noise – but also opportunities (Rubin et al.)!
  - Monitor a user session and record the timings of key presses.
  - Use existing profiles to infer their identities according to the leaked timing.

- Can extract both information and identity from a ‘secure’ session.

Slide credit: George Danezis
Do Secure Sockets (SSL/HTTPs) protect your privacy?

- SSL is used to ‘hide’ sensitive web information (HTTP encrypted and authenticated) – but does it hide everything?
- HTTP retrieves many resources per request (HTML page, style, images, . . . )
- SSL does not disturb timing much – doesn’t hide length well.
- Attack: profile the website using SSL. For each possible request make a list of retrieved resources and their lengths.
- Observe the sequence of retrieved resource lengths of the victim – make a (good) guess about which page their correspond to.

- We do better if we observe a sequence of requests (Danezis).
- Note that users are most likely to follow links on pages.
- Try to guess not only one request but a sequence – can use hidden Markov models to do this efficiently.
- Even if SSL is used web click streams can be revealed.

Slide credit: George Danezis
Can I guess which pages you visited before? (Without observing you!)

- Have you visited my competitor’s website before visiting mine?
- Adversary is a hostile website that tries to determine browsing behaviour.
- Cannot directly observe the victim. The victim only makes one request to the hostile site.
- Key observation: modern browsers have caches of pages visited – good for efficiency.
- A resource in the cache will load much faster than if requested from the network.
- Attack: embed in my website a sequence of pictures from my competitor’s site. Note how long it takes the browser to load these resources. Estimate if they were in the cache. Bingo!

- Anonymizing proxies do not help! (Attack by Felten et al.)

Slide credit: George Danezis
Location information

- Traffic data from cellular/GSM phones, WiFi base station registration and DF can be mined. Results from early studies:
  - Pascual et al. studied WiFi access point data at HAL and KTH. Could infer talks/lectures attended by owner of machine. Could infer relationships by common patterns of movements.
  - Intel Cambridge ran a bluetooth discovery experiment. Devices would record what other devices they see. Two members of staff’s devices were seeing each other at night.
  - MIT Reality Mining: 100 Media Lab staff and students were given mobile phones and traffic data was recorded. Could infer friends (Saturday 8pm), could infer status (entropy of location), could predict movements.

- Location data can be used to infer movements, relationships, status, . . . not just location!

Slide credit: George Danezis
Anonymous communications: abstract model

- Objective: hide the identity of the sender (or receiver, or both)
- Make the bit patterns of inputs and outputs different (bitwiseunlinkability)
- Destroy the timing characteristics (traffic analysis resistance)

Slide credit: George Danezis
Basic Anonymity Properties

- 3rd party anonymity
  - Alice and Bob trust each other but do not want other parties to learn that they are communicating

- Sender anonymity
  - Alice sends to Bob, and Bob cannot trace Alice’s identity

- Receiver Anonymity
  - Bob can contact Alice, without knowing her identity.

- Bi-directional Anonymity
  - Alice and Bob communicate without knowing each other's identities.

Slide credit: George Danezis
Deployed anonymous communication systems

- Email:
  - Anon.penet.fi (Helsingius 1993)
  - Cypherpunk remailers (Hughes, Finney 1996)
  - Mixmaster (Cottrell, evolving since 1995)
  - Mixminion (Danezis, 2003)

- Low-latency, bidirectional circuits (web browsing)
  - ZK Freedom network (1999-2001)
  - Java Anon Proxy (since 2002)
  - Tor (since 2003)
Main limitations and attacks

- Hard to conceal persistent communications
  - Disclosure attacks

- Vulnerable to adversaries that can create many “identities”
  - Sybil attacks

- Low latency systems: end-to-end correlation attacks
Unobservability

- With “just” anonymity, information still leaks
  - Volume of information received or transmitted.
  - Type of traffic.
  - Time of communications, or presence.
  - \[\Rightarrow\] Can be used for attacks, or Target Selection.

- Solution: Unobservability
  - Presence is not visible
  - Participation in, and volume of communications hidden.
Oblivious Transfer (OT) [Rabin 81]

- A inputs two bits, B inputs the index of one of A’s bits
- B learns his chosen bit, A learns nothing
  - A does not learn which bit B has chosen;
  - B does not learn the value of the bit that he did not choose
- Generalizes to bitstrings, M instead of 2, etc.

Slide credit: Vitaly Shmatikov

Claudia Diaz - Interdisciplinary Privacy Course  June 23, 2009
Private Search (Ostrovsky, Skeith)

- Alice ("searching party"): stores documents (in clear)
- Bob ("decoding party"): wants to retrieve documents matching some keywords
- Solution based on homomorphic cryptosystem: $E(x) \cdot E(y) = E(x+y)$
- Properties:
  - Bob gets documents containing the keywords
  - Alice does not learn Bob’s keywords
  - Alice does not learn the results of the search

June 15, 2010
Claudia Diaz (K.U.Leuven)
Privacy by design: 
buying digital goods
Comprando bonos o suscripciones

Estimado usuario asedee:

Para acceder al texto completo de las noticias, necesitas comprar un bono o suscripción. Pincha en el ícono del bono o suscripción que desees comprar:

- **BONO**: 10 euros
  - 20 euros (pagas sólo 19)
- **SUSCRIPCIÓN**: 6 meses 50 euros
  - 1 año 60 euros

Todos los importes incluyen el IVA.

**Bonos**

Para acceder a un número limitado de noticias del Archivo, compra un bono. El precio de las noticias, salvo que se indique lo contrario, es 0,75 euros, IVA incluido.

**Datos de la tarjeta**

Introduce el número de tarjeta (sin espacios en blanco) con el que deseas realizar la operación. Se aceptan como medio de pago las tarjetas Master, VISA, AMEX, además de cualquier tarjeta de "la Caixa".

- **Número de la tarjeta**: [ ]
- **Fecha de caducidad**: [ ]
- **Código Seguridad**: [ ]

3. **Aznar quiere ser el Tony Blair español**

Claudia Diaz (K.U.Leuven) / June 15, 2010
Privacy problem

- Payment => Vendor authenticates the reader
- Purchase => He learns which news are sold
  - Political membership
  - Religion
  - Hobbies
- Vendor can create personal profiles
- Privacy disrespectful vendor can sell them to companies that make targeted advertising
Priced Oblivious Transfer (POT)

Newspaper Archives
Webserver

Authentication

Deposit

Oblivious purchases

1, 2, 3, ..., N

Reader

Slide credit: Alfredo Rial
Properties of POT

- Privacy of the buyer:
  - Vendor does not learn what she buys
  - Vendor learns neither the amount of money paid nor the new value of the deposit (Newdeposit = OldDeposit – price)

- The vendor is assured that:
  - Buyer does not learn anything about news for which she did not pay.
  - Buyer pays the right price for the item she buys and updates the deposit correctly.
Privacy by design: e-petitions
Petitions

- Formal requests addressed to an authority and signed by numerous individuals

- Typically citizens provide
  - Unique identifier (name, national ID number)
  - Signature

- Verification:
  - Validating that the signatures correspond to the identifiers
  - Discarding multiple/invalid signatures
Electronic petitions

- Benefits of going electronic:
  - Many resources are needed in order to physically collect the signatures
  - Manual signature verification is a costly and tedious process
The naive e-petition implementation

- Have users sign the petitions with their e-ID
  1. Select petition
  2. Sign using the e-ID (2-factor authentication)
  3. Check that the petition has not yet been signed with that e-ID
  4. Count (or discard) the signature

- Privacy risks
  - Leak sensitive information on political beliefs, religious inclinations, etc.
  - Through unique identifiers, petition signatures can be linked to other data
e-petition requirements

- Basic requirements
  - Authentication: citizen is who claims to be (i.e., no impersonation)
  - Required attributes: citizen is entitled to sign (e.g., age > 18)
  - Uniqueness: citizens sign a petition only once
  - Correctness: all valid signatures are counted

- Privacy requirements
  - Citizen unlinkable to petition (i.e., not possible to identify who are the signers)
Citizen

credential issuing protocol

Credential issuer

Credential show

Petition signing

Petition server

Multiple signing?

Yes

Count signature and store transcript

No

Claudia Diaz (K.U.Leuven)
June 15, 2010
Properties

- Only citizens entitled to sign can do so
  - Possession of e-ID + knowledge of PIN
  - Attribute verification (e.g., age, locality)
  - One credential per citizen

- Citizens can sign only once (multiple signing is detectable so that repeated signatures can be deleted)

- Collusion of credential issuer and e-Petition server does not reveal the identity of a signer

- Need for anonymous communication channel to preserve privacy properties
Privacy by design: PrETP (Electronic Toll Pricing)
Electronic Toll Pricing

- Differentiated payment for mobility: Congestion pricing
  - Users will pay depending on their use of the car and roads

- European Electronic Toll Service (EETS) Decision (Oct 2009)
  - Defines EETS architecture and interfaces
  - Within three years for vehicles above 3.5 tons, all other vehicles within five years.
EETS straightforward implementation

Slide credit: Carmela Troncoso

Claudia Diaz (K.U.Leuven)  June 15, 2010
Privacy for Electronic Toll Pricing

- **Privacy issues?**
  - Pay as you drive
  - Fine grained GPS data allows for inferences
    - Medical issues (visits to a Cancer specialized clinic)
    - Political affiliation (visits to the headquarters of a political party)
    - Industry espionage (visits to other companies)

- **What data is necessary?**
  - Final fee that the user must pay to the provider/government

- **Legal issues**
  - Actors must not be able to cheat
  - Actors must be held liable when misusing the system
Privacy-Friendly Electronic Toll Pricing

- No personal data leaves the domain of the user.
Law enforcement

- Technical means do not suffice
  - OBU in hands of the user

- Instead technology can help:
  - Detect vehicles with inactive OBUs
  - Detect vehicles reporting false location data
  - Detect vehicles using incorrect road prices
  - Detect vehicles reporting false final fees

- Combination of law + technology
Non-Interactive Commitment

Schemes

<table>
<thead>
<tr>
<th></th>
<th>00u00 – 07u00</th>
<th></th>
<th>22u00 – 00u00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highway</td>
<td>p₁</td>
<td>p₂</td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>p₃</td>
<td>p₄</td>
<td></td>
</tr>
<tr>
<td>Residential</td>
<td>p₅, ..., pₙ</td>
<td></td>
<td>pₙ, ..., pₙ</td>
</tr>
</tbody>
</table>

HIDING PROPERTY

BINDING PROPERTY

Slide credit: Carmela Troncoso

Claudia Diaz (K.U.Leuven)     June 15, 2010
Homomorphic commitments

The content of the vaults can be added up without being known.
Creating commitments

- Slice trajectory in segments (e.g., 1 Km)

- Each segment has assigned a price per Km $p_i$

- This price is specified by the policy, example:
  - $p_i = f\left(\text{type road, time day}\right)$

- A commitment per segment is created
How does it work?

- OBU
- License Plate Reader
- Toll Service Provider
- Toll Charger

Does not reveal information about the trajectory
Cannot be changed

01-GBB-1 + Pos + Time + auth

Confirmation
What can we prove?

- OBU was active
  - A commitment with the committed location and time must be available

- OBU used correct prices
  - Prices in the table signed by Toll Service Provider
  - Check correct pricing upon commitment opening

- OBU was at reported location
  - Compare photo location with committed location

- OBU made correct operations
  - Homomorphic commitments: prices in the “vaults” can be added to verify that they correspond to the reported final fee without being opened
Holistic analysis

- From a theoretic point of view
  - The cryptography in the system ensures both privacy and law enforcement

- From a legal point of view
  - No personal data involved
  - Data minimization by design

- From a practical point of view
  - Prototype (ARM7)
  - Performance analysis
    - Computation
    - Communication

Slide credit: Carmela Troncoso
Overview of Privacy Research Challenges
Problems not quite solved yet...

- Defining the privacy requirements
- Including privacy principles in the design phase
- Hard to “add privacy” later on
Problems not quite solved yet...

- Finding robust and secure mechanisms
  - Proposed techniques keep on getting broken
  - Secure implementation is even harder
  - Complex systems require privacy at every level: the chain is as strong as its weakest link

- Easy to defeat by “changing” abstraction layer
  - camera’s, RFID tags, unique device properties, singulation protocols, traffic analysis, ...
Problems not quite solved yet...

- Usability issues (e.g., design of privacy settings)
- What can we expect users to understand and manage?
- Automatization vs control: can we redefine all the possible situations that may arise?
Problems not quite solved yet...

- Economic incentives
  - Who pays for privacy?
  - Privacy techniques very costly: complexity, overhead, lower QoS, diminished functionality
  - Privacy invasive technologies are better funded than privacy enhancing technologies
  - Tragedy of the commons
Problems not quite solved yet...

- Awareness and transparency
  - Do we know what happens to our data?
  - Who collects it?
  - For which purposes is it used?
  - What profiles do they build on us?
  - What are the consequences?
Problems not quite solved yet...

- Legal compliance
  - Some hard privacy technologies may not be compliant
    - Legal systems are often national, while technologies are transnational
  - Implementation of soft privacy techniques (e.g. privacy policies and access control)
Conclusions

- Current privacy technologies are able to reconcile requirements that seem (intuitively) incompatible.

- Feasible to design applications with very strong privacy protection.
  - For specific and well defined purposes.

- Privacy by design:
  - What are the data strictly necessary for the provision of the service?
  - Check existing building blocks.
  - Privacy properties do not compose: need for taking into account multiple system layers.