

The background of the slide is a light beige, aged paper texture. On the left side, there are several black ink splatters of varying sizes, some with fine droplets trailing off to the right, creating a sense of movement and artistic flair.

Engineering Privacy by Design

Claudia Diaz
K.U.Leuven ESAT/COSIC

Outline

- Context
- Privacy by Design in Policy
- Data minimization
- Case study I: anonymous e-petitions
- Case study II: Electronic Toll Pricing
- Lessons learned and other considerations
- Conclusions

Context

- Implementing privacy in systems is difficult
 - privacy requirements must be integrated in systems engineering activities
- Few existing systems designed with robust privacy protection in mind
- The term “Privacy by Design” is widely used by policy makers
 - IPC Ontario
 - EU Commissions Communication: “A comprehensive strategy on data protection in the European Union”
 - FTC report: “Protecting consumer privacy in an era of rapid change”
- What do PbD principles say to engineers developing systems?

PbD principles (Cavoukian)

1. Proactive not reactive, Preventative, not Remedial
2. Privacy as the default
3. Privacy Embedded into Design
4. Full functionality - Positive Sum not Zero Sum
5. End-to-end security - Lifecycle Protection
6. Visibility and Transparency
7. Respect for User Privacy - Keep it User-Centric

Principles too vague

- Example: 3. Privacy Embedded into Design
 - “Privacy by Design is embedded into the design and architecture of IT systems and business practices. It is not bolted on as an add-on, after the fact. The result is that privacy becomes an essential component of the core functionality being delivered. Privacy is integral to the system, without diminishing functionality.”
- From the definition it is clear that something needs to be done about privacy, but... what? and how can it be translated into **systems design**?
 - *Aggressive data minimization*

Data minimization in DPD

- DP directive states that data collection should be limited
 - “personal data must be collected for *specified, explicit and legitimate purposes*” and must be “*adequate, relevant and not excessive* in relation to the purposes for which they are collected and/or further processed”
- However, no explicit mention of “data minimization”
 - Proportionality clause can be easily subverted by articulating the purpose specification to include any desired data
 - Legitimizing the collection of copious amounts of data by providing “control” and “transparency”
- Informed consent and subject access rights
 - What does “informed consent” mean for de facto “mandatory” services (energy metering, road tolling, telecom networks)?
 - Limited scope of what counts as *personal data*

Engineering perspective

- Disconnect between policy makers and engineers on what it means to technically address privacy threats
- Privacy by design in policy documents can be interpreted as the “collection and processing of any data – but with a privacy label”
- “Control” and “transparency” do not mitigate the privacy risks that arise from mass collection of data in databases
 - Single point of failure
 - Attractive target
 - Hard to secure (SP itself / malicious insiders / accidental disclosure / outsiders)
 - Risks of public disclosure, and/or “stealthy” abuses (e.g., secondary use)

Properties of digital systems

- Tendency to reproduce increasingly complicated bureaucratic systems exactly in information technology
- Properties of digital systems ignored by policy makers
 - Easy to replicate and distribute digital information
 - Statistical inferences, linkability across contexts
 - Computational capabilities: more can be done with less data
- Lack of metaphors / intuition to explain “magical” capabilities (eg, ZK protocols)
- Data minimization can be taken **much** further than what would usually be considered to be “adequate, relevant and not excessive in relation to the purpose”
 - Not just about not asking for “marital status” when subscribing to a gym

Dimensions of “data minimization”

- *Data minimization* does not refer to simply abstaining from processing personal data that is clearly irrelevant for any plausible purposes of the processing.
 - Minimize the *disclosure* of data to other entities
 - Minimize the *reliance* on (need to *trust*) other entities for guaranteeing privacy protection
 - Avoid *centralized* architectures in which one entity is in the position to massively collect user data
 - Favor *distributed* architectures in which personal data resides to a large extent in devices under the user control.
 - Minimize the *risk* of privacy *breaches*, diminishing the likelihood and impact of a database leak.
 - Minimize the *replication* of data, as the likelihood of a privacy breach increases with the number of copies.
 - Minimize the *time* the data is available in the system, by not storing it

- The interpretation of the statement “data minimization” ought to go hand in hand with the state-of-the-art technology
- Not “just” with our understanding of what data minimization may mean in the analogue world

Techniques for data minimization

- Anonymity
 - Service provider can observe access to the service
 - Cannot observe the identity of the user
 - Robust anonymization is difficult
 - Understanding anonymity sets not trivial
- Oblivious Transfer (OT) / Private Information Retrieval (PIR)
 - Service provider can identify user
 - Cannot observe details of the access to the service
 - How to convey the technical intuition to non-experts?

Case study I: Anonymous e-petitions

12

e-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
- Benefits of going electronic:
 - Many resources are needed in order to physically collect the signatures
 - Manual signature verification is a costly and tedious process
- European Citizens' Initiative (ECI):
 - Introduced by the Lisbon Treaty
 - Allows citizens to request new EU legislation once a million signatures from a significant number of member states have been collected

The straightforward e-petition implementation

- Have users sign the petitions with their e-ID
 1. Go to e-petition website and 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 (public disclosure or stealthy abuse)
 - Leak sensitive information on political beliefs, religious inclinations, or other inferences
 - Potential of abuse of this information to profile, categorize, discriminate, or stigmatize people based on their ideas
 - Through unique identifiers, petition signatures can be linked to other data
 - These risks are greater than in paper-based petitions due to the properties of digital systems

e-petition functionality

- Basic requirements
 - Authentication: citizen is who claims to be (i.e., no impersonation)
 - Required attributes: citizen is entitled to sign (e.g., age ≥ 18 and nationality \in EU)
 - 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)
 - The point is to know *how many* citizens support an initiative, not *who* they are

PKI vs Anonymous Credentials

PKI

Signed by a trusted issuer
Certification of attributes
Authentication (secret key)

No data minimization
Users are identifiable
Users can be tracked
(Signature linkable to other
contexts where PK is used)

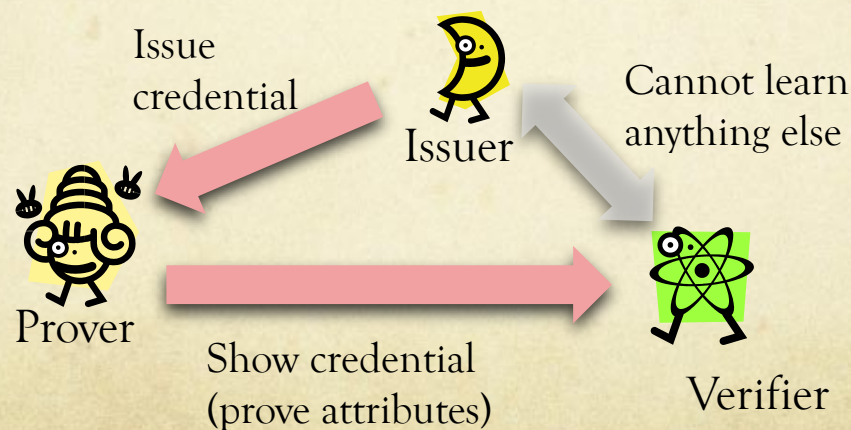
Anonymous credentials

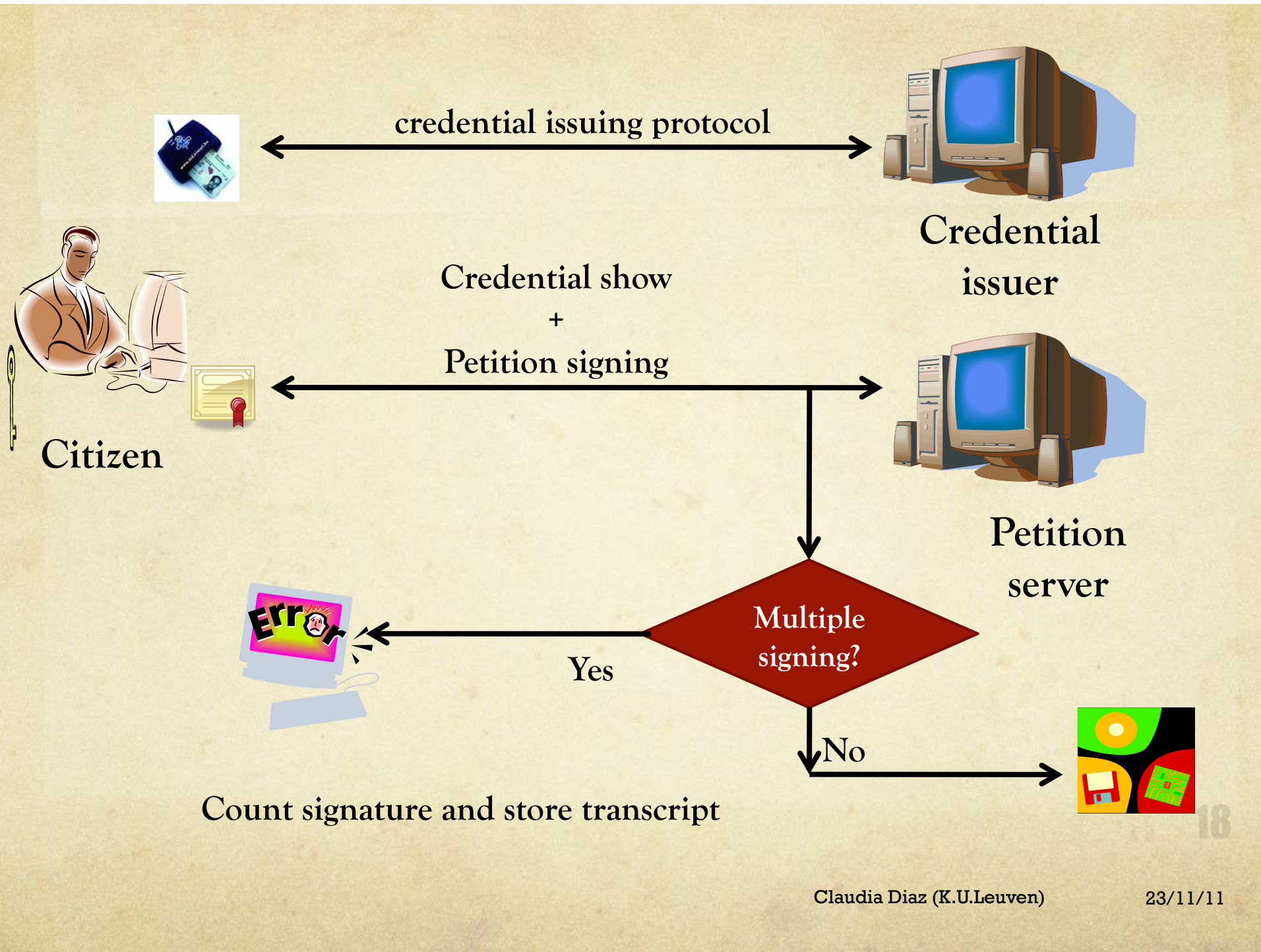
Signed by a trusted issuer
Certification of attributes
Authentication (secret key)

Data minimization
Users can be anonymous
Users can be unlinkable in
different contexts

Anonymous credentials

- Properties:
 - The prover convinces the verifier that he holds a credential with (certified) attributes that satisfy some conditions:
 - 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





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

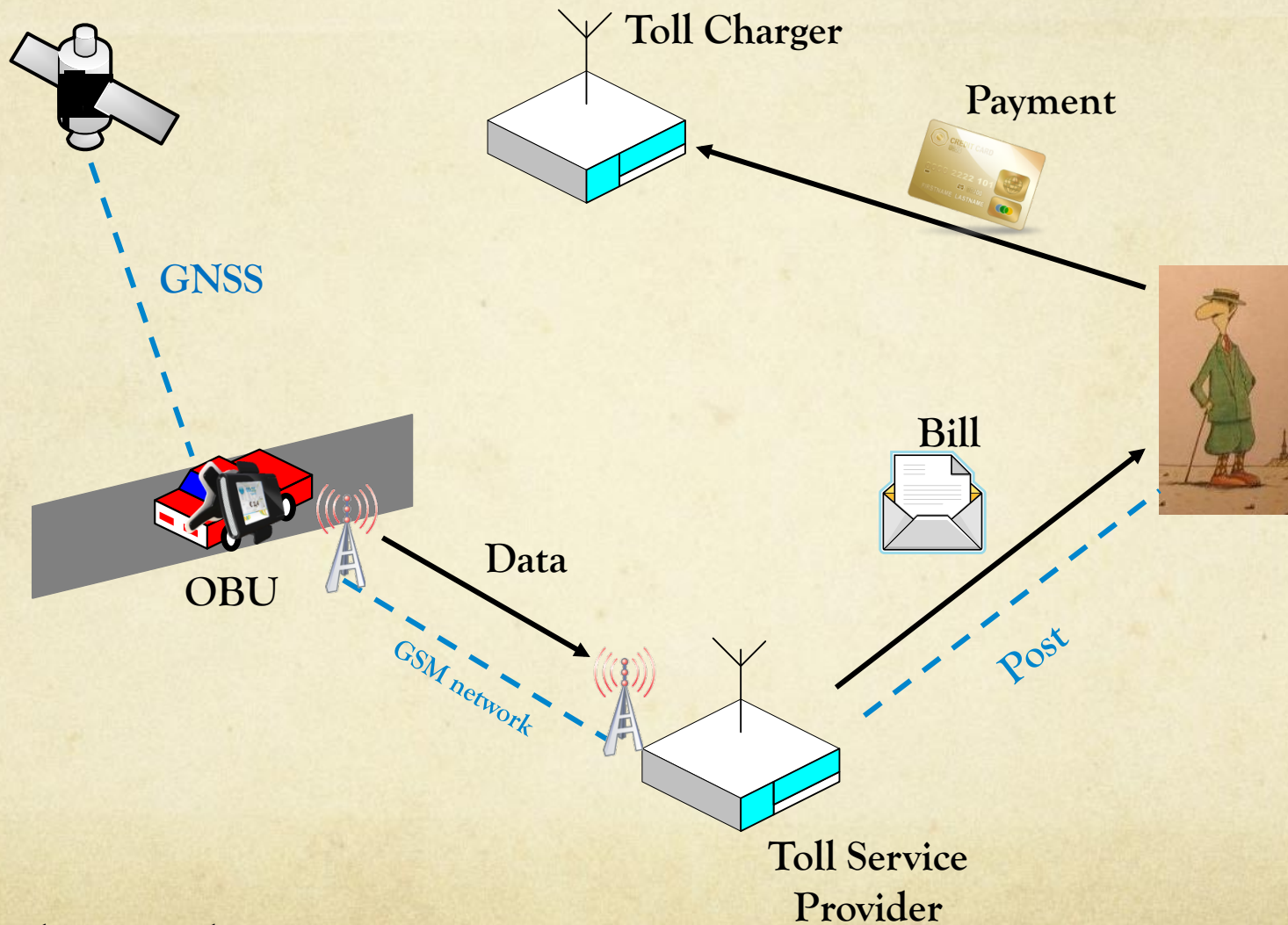
Case study II: Electronic Toll Pricing

20

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

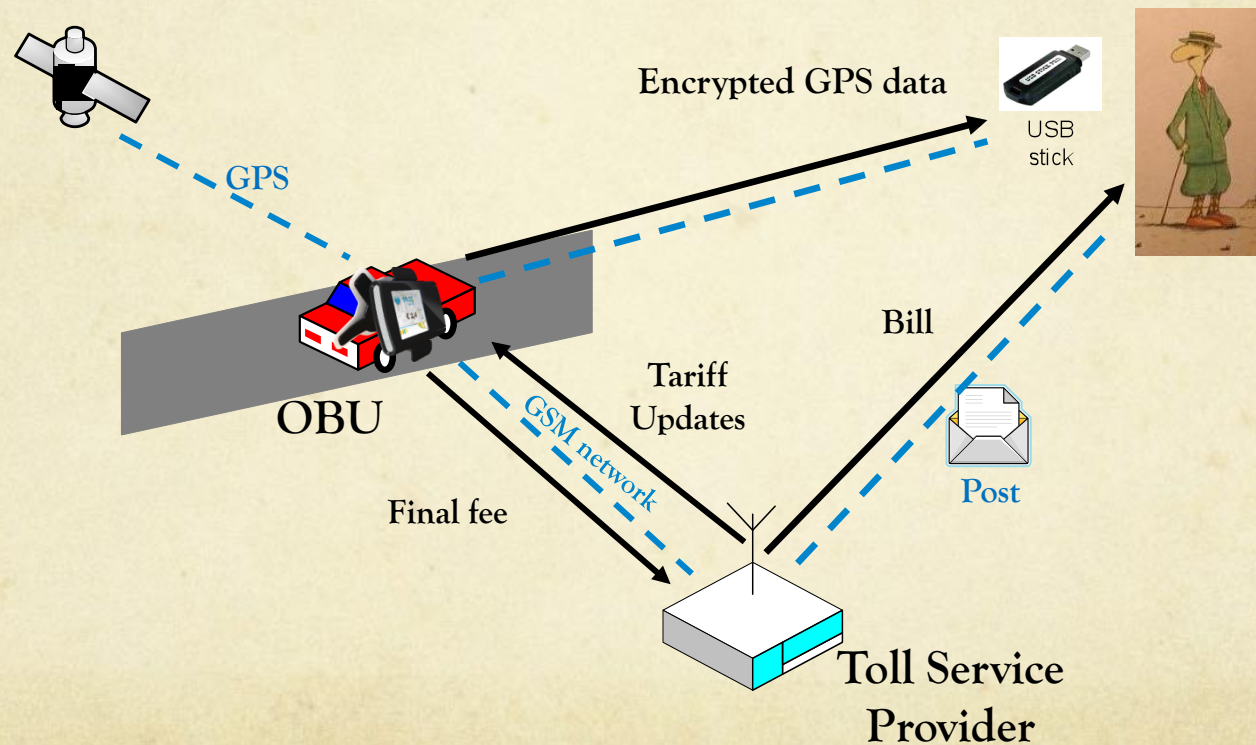


Privacy for Electronic Toll Pricing

- Privacy issues?
 - *Pay as you drive*
 - Fine grained GPS data allows for all kinds of inferences
- What data is necessary?
 - Final fee that the user must pay to the provider/government
 - This is the actual purpose of the whole system – and not collecting everyone's detailed location data
 - Enormous **reduction of risk and cost** by eliminating the need to store all the raw data
- Legal / service integrity issues
 - Actors must not be able to cheat
 - Actors must be held liable when misusing the system

Privacy-Friendly Electronic Toll Pricing

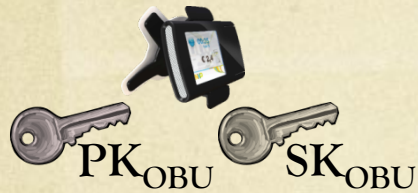
- No personal data leaves the domain of the user



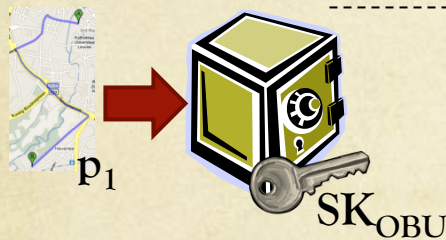
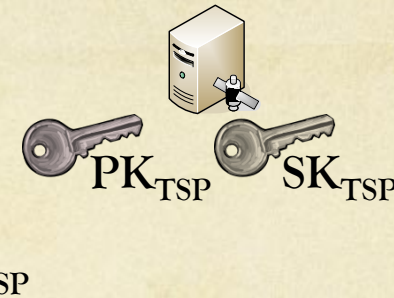
Enforcement

- OBU in hands of the user
 - Incentive to cheat (paying less)
 - Even if the box is tamper-resistant, the input is easy to spoof
- We need to:
 - Detect vehicles with inactive OBUs
 - Detect vehicles reporting false location data
 - Detect vehicles using incorrect road prices
 - Detect vehicles reporting false final fees

Non-Interactive Commitment Schemes



	00u00 – 07u00	22u00 – 00u00
Highway	p_1	p_2
Primary	p_3	p_4
.....
Residential	p_{n-1}	p_n



HIDING PROPERTY

Where you at....?



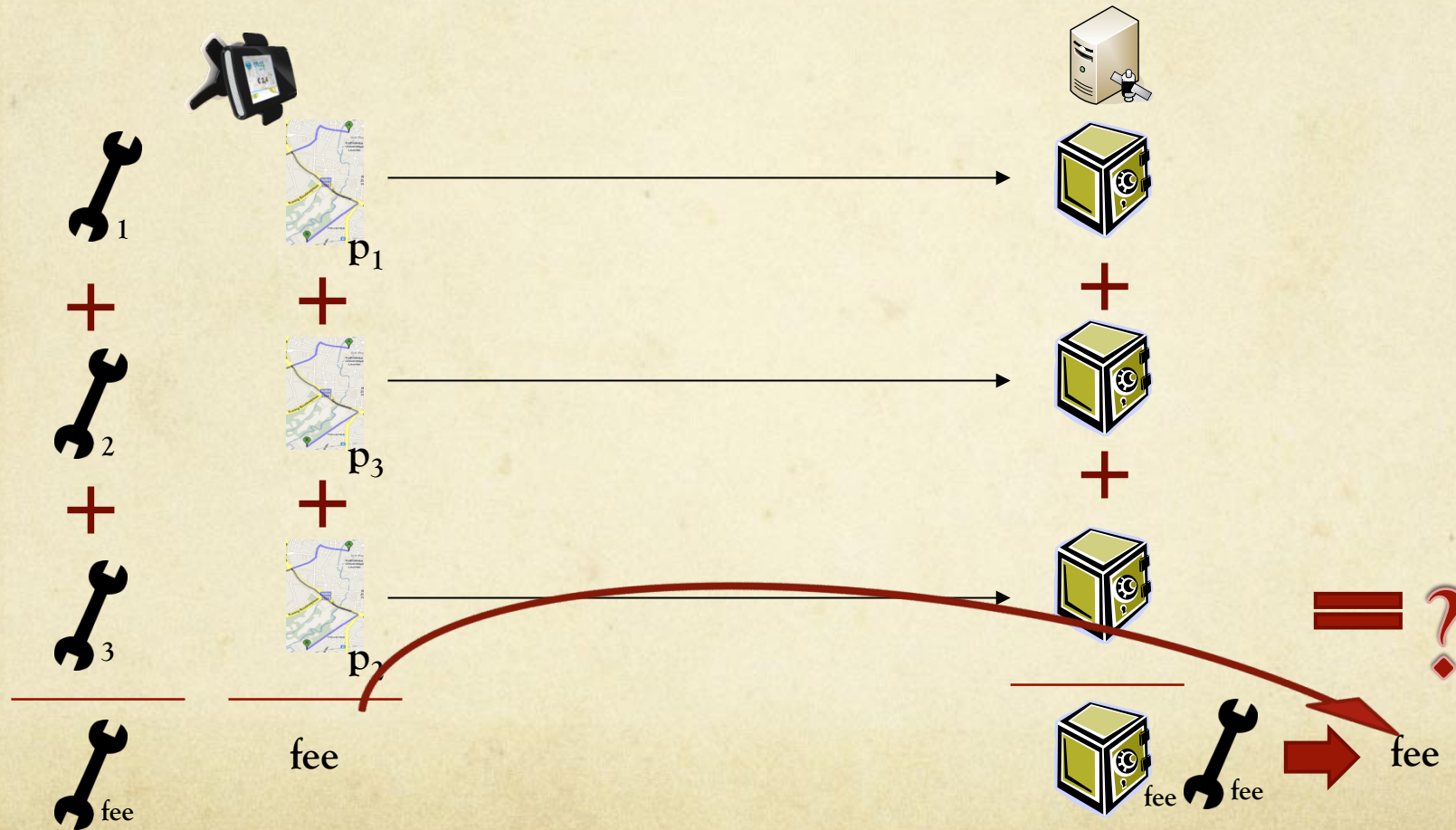
BINDING PROPERTY



= ? 26

Homomorphic commitments

- The content of the vaults can be added up without being known

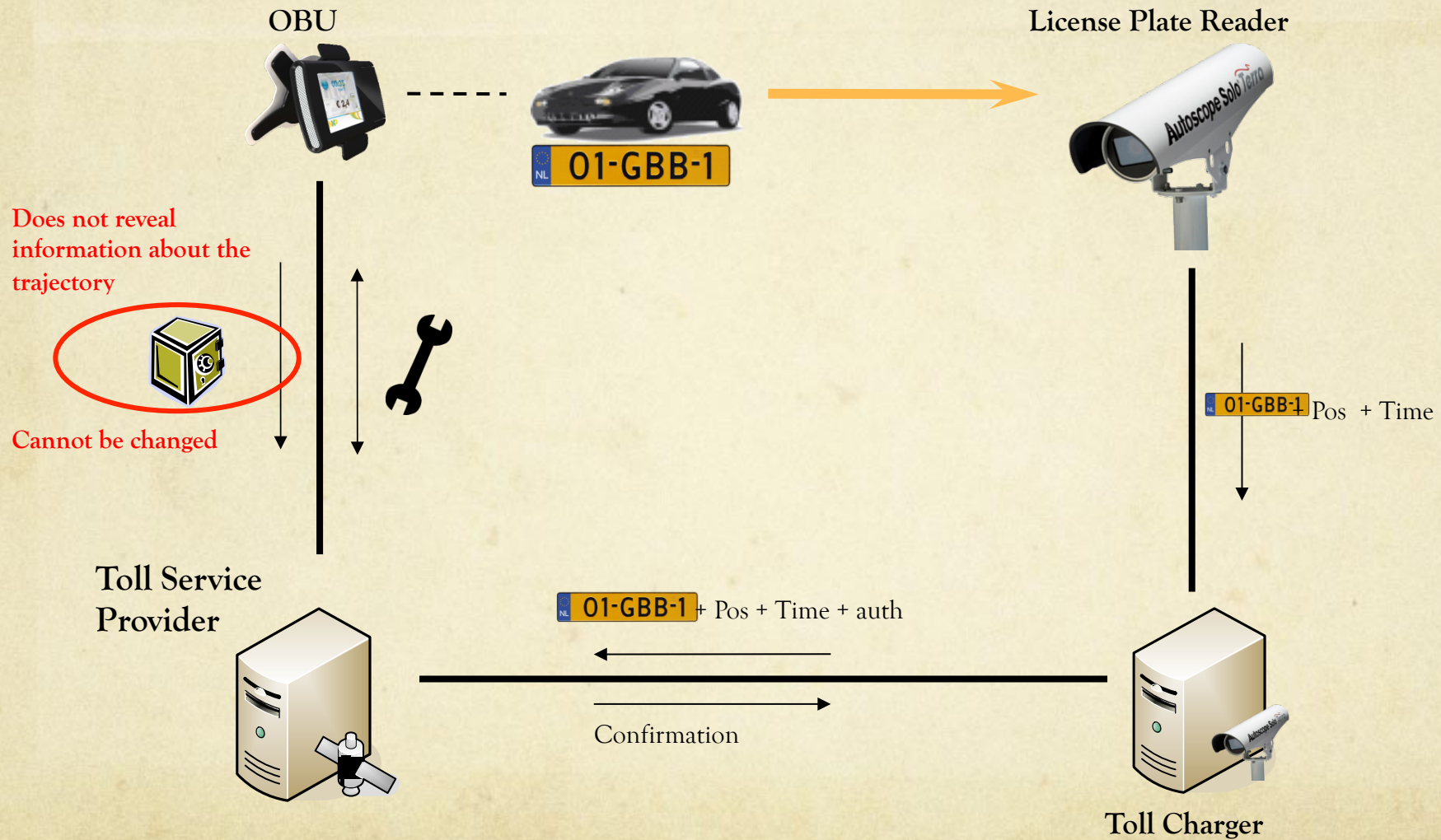


Slide credit: Carmela Troncoso

Claudia Diaz (K.U.Leuven)

23/11/11

How does it work?



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

Towards PbD methodology

- Functional requirements analysis
 - It is critical to provide a precise description of what the system should do
- Data minimization (several dimensions)
 - Find the minimum set of data that is strictly necessary to fulfill the functionality, and the integrity of the system
 - Data may reside in the system but in the user device instead of in centralized database
 - Anonymity / ZK protocols / both: requires knowledge of the state of the art in privacy technologies
- Modeling attackers, threats, and risks
 - Some threats (eg, secondary use, inferences, abuse derived from *authorized* access) may not be obvious to non-privacy-expert systems designers
- Multilateral Security Requirements Analysis: security requirements of the different stakeholders (eg, integrity)
- Implementation and testing of the design (re-iteration and re-evaluation of risks and threats)

Other considerations

- Ethical, legal and political analysis of proportionality
 - “Legitimacy” of the desired system given its burden on privacy: “the establishment that the application goals would be useful for the intended use population”
- Privacy by design and population surveillance
 - If the purpose of the system is to do intrusive surveillance of populations, then putting a privacy by design label on these systems is misleading (white-washing of intrusive systems)
- Risks and social norms
 - Non-technical risks (e.g., discrimination of populations)

Conclusions

- Data minimization must have a central role for PbD
- Data minimization not only about anonymity
- Need for better intuition / metaphors to convey to non-experts what state-of-the-art privacy technologies can do
 - Specific expertise is needed
- Need to deploy robust privacy systems that can be used as a reference
- Need to develop an engineering methodology for PbD
 - Avoid reducing PbD to checklists that can be easily ticked away for compliance

Thanks!

- For more info:
- *Engineering Privacy by Design*. Seda Gürses, Carmela Troncoso, and Claudia Diaz. In Conference on Computers, Privacy & Data Protection (CPDP), 2011
- <https://www.cosic.esat.kuleuven.be/publications/article-1542.pdf>