Elements of a privacy architecture

Summary and case studies

Subhashis Banerjee, Subodh Sharma

Privacy risks

Orwellian dangers

- Direct surveillance by unauthorised access
- Illegal profiling by linking information across silos. With or without using a digital identity.
- De-anonymisation attacks
- Insider attacks
- Mere existence of a surveillance infrastructure can disturb the balance of power between citizens and the state.

Privacy risks

Kafkaesque dangers

- Unsure about what information is being used by state and other bureaucracies, and for what purposes.
- Being mis-profiled, wrongly assessed ir even influenced by out-of-context data, without being able to control or even being aware.
- Losing control of informational self-determination.
- Being denied access to services - critical or otherwise - because of poor use cases.
    - De-registration from voter lists, PDS; denial of ration or hospital treatment because of fingerprint false negatives…

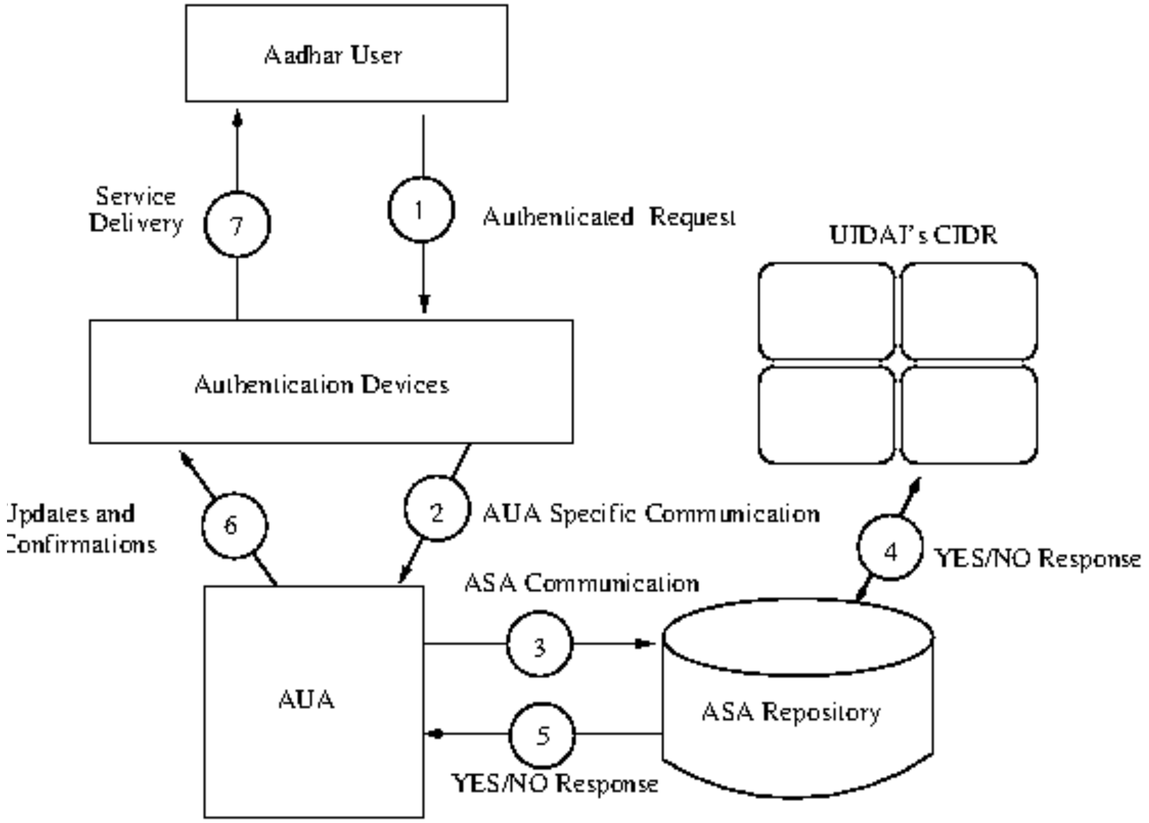
Privacy risks

Big-data analytics and machine learning

- Profiling.
- Targeting.
- Reliability.
- Bias and discrimination risks. Often non-causal.
- Opacity, scale and damage are common traits (O’Neil, Eubanks, Zuboff).
- Can exacerbate inequality (Eubanks).
- Can be both Orwellian and Kafkaesque.

Case study: Aadhaar

- Aadhaar workflow



Aadhaar workflow

Case study: Aadhaar

- For disbursal of welfare.
- Passed as a money bill (consilidated funds of India)
- Non-social, non-federated biometric-based definition of identity
- De-duplication? Combinatorial impossibility? No public report of audit
- Clear theory of exclusion due to false-negatives of biometric matching.
- DBT and AEPS.
- Unclear use cases.

Case study: Aadhaar

Privacy breaches (identity thefts and insider attacks)

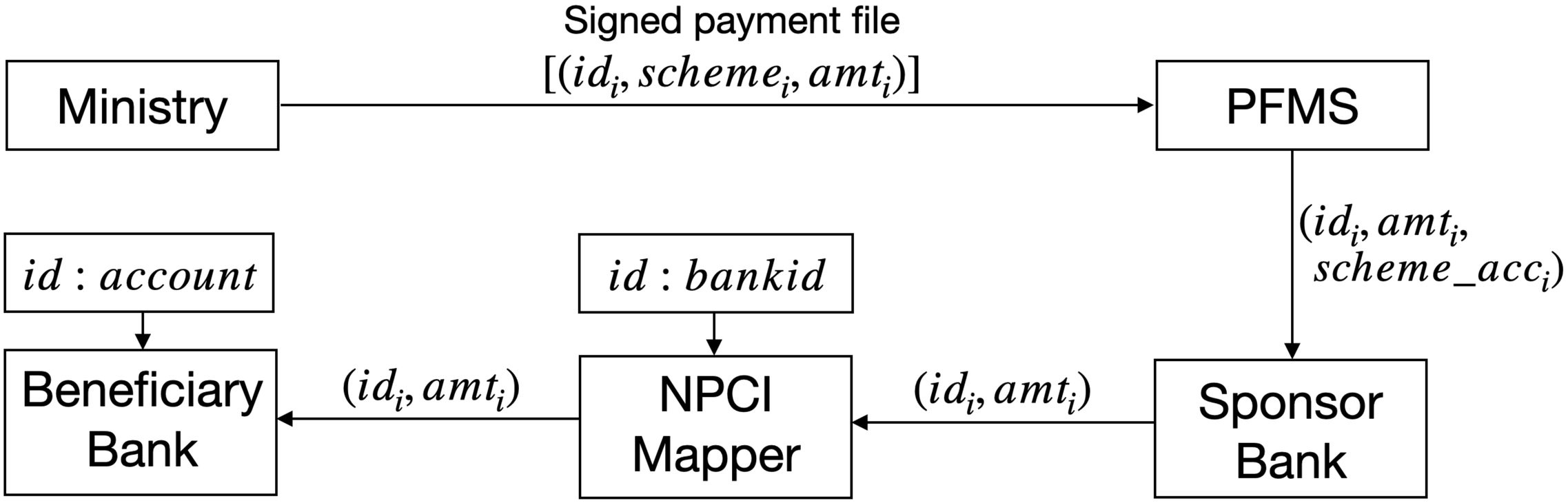
- ¨Aadhaar data hack: IIT Kharagpur grad created eKYC verification app of his own", Financial Express, 2017
- ¨Aadhaar mess: How Airtel payment banks pulled off its 190 Cr magic" - Anand Venkatanarayana and Srikanth Laxmanan, the Wire, 2017
- ¨Rs. 500, 10 minutes, and you have access to a billion Aadhaar details" - Rachna Khaira, The Tribune, 2018
- ¨UP pilferage scam reveals hollowness of government's claims about Aadhaar" - Rohan Venkataramakrishnan, Scroll.in, 2018
- ¨Riteish Deshmukh, Hanuman, Pak spy get PM Kisan funds as farmers" (also Ram Sevak Sharma!) - Sushovan Sircar and Vakasha Sachdev, The Quint, 2020.

Case study: Aadhaar

Function creep (more Kafkaesque)

- ``Aadhaar cards''
    - Bank accounts, Covid testing, hospital treatment, blood bank, rent agreement, property deals, car purchase, insurance, municipal corporations,…KYCs
- NPCI and banking
- DBT and AEPS - ``incomplete Aadhaar''
- micro finance
- SIM cards
- Covid vaccinations?

Case study: DBT



Case study: Contact tracing

Basic idea

- BLE and GPS (optional) on smartphones
- Smartphone exchange anonymous tokens when in proximity, and each phone keeps a record of all sent and received tokens
- When an individual is infected all tokens are uploaded to a central service
- Central service
    - Trusted - can decrypt and link tokens and alert the users of potential risks
    - Untrusted -
        - publish a list of `risky' tokens and risk is computed on each cellphone.
        - *Private set intersection computation between server and tracing app.*

Case study: contact tracing

Privacy risks

- Individuals learning about other individuals as high risk spreaders
- malicious claims by individuals forcing quarantine on others
- Insiders at the central service learning about individuals at high risk
- exposure of social graphs of individuals
- Other purpose limitation violation at the central service

Case study: contact tracing

Privacy risks of various approaches

System	System Req.		Privacy Protection Against				Client
	Trusted Server	#	Infection Status		Social	False-positive	Comm.
			By User	By Server	Graph	User	Cost
TraceTogether [5]	Yes	1	Yes	No	No	Yes	$O(n)$
Baseline*	**No**	1	No	No	Most	Some	$O(N)$
Private Messaging [2]		3	No	Yes	Yes	No	
Epione		2	**Yes**	**Yes**	**Yes**	**Yes**	$O(n\log(N))$

- Baseline systems include Private Kit, Covid-watch, CEN, DP-3, PACT
- AarogyaSetu uses a central server, static tokens, GPS locations and is vulnerable to all.
- All except TraceTogether and AarogyaSetu compromise utility.

CS techniques: utility and limitations

Encryption

- Addresses only secrecy aspects of privacy
- Vulnerable to insider attacks
- A **must** for almost all privacy protection

CS techniques: utility and limitations

Data minimisation

- Access control
    - often within the same organisation
- Zero-knowledge proofs
- Anonymity and anonymous credentials
    - linkable and unlinkable anonymity
    - blind signatures
- Anonymous networks (mixnets)
- Database anonymisation
    - seldom works

CS techniques: utility and limitations

Inferential and differential privacy

- **Inferential privacy**: the notion that no information about an individual should be learnable with access to a database that could not be learnt without any such access.
    - No such guarantee possible if adversary has access to unrestricted auxiliary information
- **Differential privacy**: minimises the *additional privacy risk* that each individual incurs by participating in the database.
    - A weaker notion and applicable only to statistical databases

.

CS techniques: utility and limitation

Formal specification and static analysis of purpose limitation

- Purpose specification languages
- Static analysis of programs
    - often involve tedious hand tagging
- Log analysis
    - completeness and soundness doubtful

CS techniques: utility and limitations

Homomorphic encryption

$$E(m_1) \times E(m_2) \times \ldots \times E(m_n) = E(m_1 + m_2 + \ldots + m_n)$$

- Fully homomorphic encryption is both additive and multiplicative
- Performance not good enough for practical deployment as yet

CS techniques: utility and limitations

Secure multiparty computation

- allows multiple parties to compute a function of their private inputs such that no party learns about others' private inputs, other than what the function's output reveals
- often requires significant re-engineering
- may not always fit organisational realities

CS techniques: utility and limitations

Hardware solutions like SGX

- Condifentiality of enclave
- Integrity
- Remote attestation
- Secure provisioning of cryptographic assets
    - based of public/private key pairs and Diffie-Hellman key exchange
- No end-to-end privacy preserving application solutions as yet

Elements of an Architectural solution

- Regulatory oversight
- Regulated access control
- Appropriate data minimisation when data crosses regulated boundary
    - anonymous credentials (virtual identities), blind signatures
    - (un)linkable and (un)traceable depending on use case
- Strict purpose limitation under regulatory oversight.

Elements of an Architectural solution

- Data encrypted in storage or transit
- **Only** programmatic access
- Pre-audited, tamper-proof computer programs which do only what they are supposed to do.
- Online consent and authorisation architecture.