Mapping a Threat Model for the Aadhaar Ecosystem

Shweta Agrawal Subhashis Banerjee Subodh Sharma

[The authors are professors of computer science; Shweta is at IIT Madras, Subhashis and Subodh are at IIT Delhi]

The Aadhaar privacy debate has not made much progress, with the proponents claiming that Aadhaar is safe and the opponents claiming that Aadhaar necessarily violates privacy. As argued here and here, the discussion is uninformative without first modelling an attack surface by precisely enumerating the possible ways in which privacy may be breached. Moreover, this enumeration must be based on a comprehensive analysis of policy, legal and data security considerations, and all three are necessary. We endeavour to do such an analysis in this article (see here for a detailed version).

Privacy breaches in a setup like Aadhaar can happen through *authentication and identification without consent and beyond legal sanctions*; through *illegal profiling by correlation of identities across different data silos*; and through *illegal use of data in the central and other repositories*. All other types of privacy compromises are derived from or are special cases of the above. The onus ought to be on the designers to convince that there are reasonable protections against these.

Authentication and identification without consent

Authentication, for example to make payments, requires two independent pieces of information - identity and an authentication credential. Common examples of identity are Login or email IDs, cryptographic public keys and ATM or smart cards; some common authentication credentials are passwords (including OTPs), PINs and cryptographic private keys. Identity may be considered (limited) public information but an authentication credential must necessarily be private - a secret that is known only to the user. Moreover, authentication must be a conscious process that requires active participation by a user, but not necessarily so for identity verification. Biometrics, which are external to one's body and can easily be harvested without consent, are poor authentication credentials because of possibilities of false presentations. Use of biometrics as authentication credentials, for applications like financial transactions, is ill-conceived and requires immediate review.

Using biometrics may be excellent for identity verification under the adversarial oversight of the person or entity requiring the verification, but what is required are the legal and policy frameworks that define who has the right to verify the identity of an individual and under what circumstances.

PoS and enrolment devices are the most likely sources of leakage of biometric and other sensitive data which can be used to illegally authenticate or identify. These devices need to be registered with the UIDAI and authenticated during run-time to ensure that they have not been tampered with. We note that the UIDAI is already taking welcome steps in this direction.

Profiling by correlation of identities across data silos

Identification and profiling without consent by correlating different data silos is undoubtedly a big threat to privacy and civil liberty, as has been pointed out often. However, with the current state of affairs with digitisation, linking different data silos using Aadhaar does not seem to add significantly to the attack surface. It is true that an individual or an entity with access to multiple databases linked with Aadhaar can uniquely identify any individual in them, leading to possible illegal profiling. But such unique correlation attacks can also be carried out using other identifiers such as mobile or PAN numbers. Even if all such unique identifiers are removed from the data, linking databases for unique identification is fairly straightforward using the demographic and personal data that we provide in course of routine business. In fact, such correlation and profiling, even without unique identifiers, are common for online targeted advertising, and will be a trivial task for an entity like say NATGRID.

So, the risk of illegal profiling does not originate as much from Aadhaar as it does from the modern needs of digital record keeping in different silos, and the Aadhaar privacy debate has drawn timely attention to the issue. What is required is a thorough analysis of what kind of profiling are legitimate requirements for governance and codifying them in a law. All other kinds of profiling should be prevented.

This is not to say that using the same UID for all applications does not make it worse. It adds to the vulnerability by making unique identification easy even for a layperson. The LSE identity project report suggests cryptographic embedding of a unique global ID into separate local IDs for each application domain, thereby making cross identification using the local IDs impossible except for the ID granting authority. UIDAI should definitely consider this possibility.

A popular solution to prevent correlation is to systematically corrupt the databases using differential privacy techniques to make one indistinguishable from $k-1$ other individuals for a large enough k. However, such data corruption may impede legitimate governance requirements.

The only reasonable solution, in our contention, will be to prevent sensitive databases from coming together, except for legitimate purposes and only through automatic means. There also has to be national standards for data collection and protection, not only at UIDAI but also at other sensitive data domains.

Insider attacks

On the face of it, the data protection measures adopted by UIDAI appear to be standard and adequate against external threats, but it is not obvious that they are adequate against insider threats. Insider attacks, perhaps at the behest of powerful entities in the state machinery itself, are the biggest threat to privacy and civil liberty. Maintaining data encrypted and distributed within an organisation is not adequate protection against insider attacks if the decryption keys also reside within the same organisation.

For effective protection against insider threats it is imperative to ensure that no manual inspection of sensitive data and transaction logs is ever possible, and that data can only be accessed through pre-audited, tamper proof, digitally signed computer programs which are true to the legal and policy frameworks. Moreover, such programs must be trustworthy and do precisely and only what they are supposed to do, even when the underlying computing,

network and storage infrastructure are untrustworthy (equivalent to already been hacked).

This will require an independent third party that can play the adversarial role of an online auditor and also that of a key-keeper. The auditor has to ascertain that the programs are true to specifications, sign and seal them, and authenticate them during run-time to ensure that they have not been tampered with. The necessary policy and legal frameworks need to be put in place.

Indeed, we do believe that there are tools and techniques from computer science that may, at least for practical purposes if not provably, offer such protection in the UIDAI and other such sensitive setups. Something as important as Aadhaar definitely requires such due-diligence.