Architecture for privacy

Subhashis Banerjee

April 11, 2018

The debate engendered by the identity project has propelled us from being a predominantly pre-privacy society to one in which privacy protection in digital databases has emerged as a major national concern. The welcome and scholarly Supreme Court judgment on the right to privacy has made it abundantly clear that privacy protection is imperative, and any fatalistic post-privacy world-view is untenable. Informational self-determination and the autonomy of an individual in controlling usage of personal data have emerged as central themes across the privacy judgment.

This provides us with a unique opportunity to take a fresh look at design of digital services in India. On the one hand we should have stricter provisions than the sector specific standards in the US, where not only are identity theft rates unacceptably high, but also from where some of the world's largest corporate panopticons like Google and Facebook have grown more or less unchecked. On the other hand India should ideally have a more innovation friendly setup than what the European GDPR can offer, which perhaps is unduly restrictive but is unlikely to be commensurately effective. Moreover, our designs need to be specially sensitive to our large under-privileged population which may not have the necessary cultural capital to deal with overly complex digital setups.

Recording transactions with a digital identity projects an individual into a data space, and any subsequent loss of privacy can happen only through the data pathway. Hence data protection is central to privacy protection insofar as databases are concerned. The critical challenge in design of a data protection framework is that the main uses of digitisation - long term record keeping and data analysis - are seemingly contradictory to the privacy protection requirements.

The most common fear of digitisation is that of mass surveillance. Databases linked by unique identities can possibly create an infrastructure for totalitarian observation of citizens' activities across different domains. The mere existence of such infrastructure, if unrestricted, can potentially disturb the balance of power between the citizens and the state, stifle dissent and be a threat to civil liberty and democracy. Several commentators have used clichéd metaphors like the *Orwellian big brother* or *panopticon* to describe the situation.

A more common and subtle manner of erosion of privacy is by the way of losing control of informational self-determination both to the state and to other seemingly mysterious, uncaring and opaque bureaucracies. Often there is no obvious invasion of privacy, but one may sometimes become unsure about what information about her is being used by the state and other bureaucracies and for what purposes. *Kafkaesque* is an appropriate metaphor that has sometimes been used to describe the situation. Not only can personal information leach out and be used by unpredictable entities in unpredictable ways, but one can also be misprofiled, wrongly assessed or even influenced using out-of-context data, without being able

to control such actions or sometimes even being aware of them. Indiscriminate or unethical use of machine learning can also lead to profiling and privacy violations whose consequences are not immediately obvious.

Exclusions and denials because of poorly thought out use cases, for example because fingerprints do not match, are more direct violations. Such callous omissions can even be threats to liberty and life. Traditional approaches to protection of rights have been less than effective, anywhere, mainly because the enforcement methods have been weak.

When participation is voluntary, privacy self-management through notice and informed consent; collection, purpose and storage limitation; transparency; and individual participation through opt-in and opt-out have been advocated as foundational principles for privacy protection. However, notice and consent are usually ineffective because of information overload, choice limitation and consent fatigue - as often demonstrated by the customary negligent clicking of 'I Agree'.

Mandatory use of digital identities requires clearly establishing a legitimate state interest and enacting a proportional and just law. The role of consent in such situations is minimal, but collection and purpose limitation are important operative principles and citizens' basic rights still need to be protected. However, the state's understanding of and respect for this principle has often been questionable.

Also, in either case, recognition and acknowledgement of purpose extensions have almost always been problematic.

The European GDPR proposes *right to explanation* as a countermeasure to indiscriminate and biased machine learning applications. However, predictive analytics rarely support causal reasoning, and, without expert audit of algorithmic and data biases, the explanations will most likely turn out to be inane.

Effective data protection in India will require a strong regulatory framework with a hierarchy of data regulators that can protect our basic rights irrespective of our understanding of complex digital setups and levels of consent. Also, any solution that is solely based on detection of privacy violations and subsequent punitive actions is unlikely to be effective, mainly because the causal effects of privacy violations, especially of the *Kafkaesque* types, cannot be easily and immediately determined. What is required is an online architectural solution that prevents privacy invasions in the first place.

Not only do the data regulators require independent authority, but they also need to actively participate in the data protection architecture. Apart from determining fairness of algorithms and use cases, they need to play two other main roles. The first should be to determine, and explicitly list out, authorisations for data accesses for various data processing functions based on a rights-based principle in addition to consent. Purpose limitation needs to be built into such authorisations, and all purpose extension requirements must be explicitly considered. The second role should be to ensure that data can be accessed only through audited, pre-approved and digitally signed computer programs after online authentication and verification of the authorisations presented. Both the data regulator and the data controller should maintain non-repudiable logs of all data accesses, and neither should be able to access the data independent of the other.

The technology to support such regulatory functions exists, what are necessary now are an effective and rights-based data protection law, and the will to build the required regulatory capacity.

Note: This article is a summary of the author's submission (https://arxiv.org/abs/1801.05313) in response to B N Srikrishna committee's white paper on a data protection framework.