Protecting data privacy: Authorisation and access control

Subhashis Banerjee Subodh Sharma

Computer Science and Engineering, IIT Delhi
New Delhi 110016

June 27, 2018

The Committee of Experts on a Data Protection Framework, with Justice B. N. Srikrishna as the chairman, was constituted in August 2017 to examine issues related to data and privacy protection in digital applications, recommend methods to address the issues, and draft an effective data protection law. The committee's white paper on a data protection framework [Srikrishna et al., 2017] is based on the broad principles of informed consent; collection, storage and purpose limitation [1]; enforcement; accountability and penalties; uniform application across all sectors and technology neutrality. The report outlines the roles of data controllers and data processors, and their responsibilities and accountability. It recommends independent data regulators for enforcement, in either a command-and-control or a co-regulatory structure. While the report also suggests privacy-by-design, it does not elaborate on possible methodologies for such an approach. In this note we advocate authorisation and access control as a viable framework for privacy-by-design.

Detection vs prevention

Privacy protection in digital databases has been less than effective, anywhere, mainly because the enforcement methods have been weak. In most cases, the enforcement strategies have been based on post-facto punitive and corrective measures after detection of violations. Even the recently enacted European General Data Protection Regulation (GDPR) [The European Parliament and the Council of European Union, 2016] does not clearly specify any standards for enforcement. We argue that an architectural solution that prevents privacy invasions in the first place is more likely to succeed than strategies based on detection of violations and subsequent punitive measures. Detection of privacy infringements will often be uncertain because the causal effects of invasions are usually hard to establish.

For example, it may turn out to be impossible to know for sure whether a person has lost their job because their medical data was accessed without authorisation and used to discriminate against them, or if some other reason given as the official explanation was the actual determining factor. Causal links of privacy violations due to indiscriminate and unethical use of machine learning are also hard to establish, and the *right to explanation* proposed in the GDPR is unlikely to be an effective countermeasure. Hence, *ex-ante*, rather than *ex-post* [Raghavan, 2018], ought to be the preferred approach.

Privacy protection does not demand that personal data should not be collected, stored or used, but that there should be provable guarantees that the data cannot be used for unauthorised purposes.

Rights-based approach

[1] which require that data should be used only for the purpose for which it was collected.

When user participation in a digital application is voluntary, informed consent has often been advocated as the foundational principle for privacy protection. However, information overload and choice limitation often makes *consent* ineffective. Also, considering that a large fraction of India's population may not have the necessary cultural capital to deal with complex digital setups, a rights-based approach [Matthan, 2017] that shifts a significant part of the accountability from an individual to the data controller will be more appropriate. This should not be in lieu of but in addition to individual consent. Any mandatory digitisation with personal identifiers, such as in income tax or welfare, needs to be backed by a just and proportional law, and must balance the potential loss of individual privacy with the expected public good. In either case, purpose limitation must be a fundamental operative principle for protecting privacy rights of individuals.

Identifiers and privacy

Use of unique personal identifiers, or even phone numbers, in application databases may enable unauthorised profiling of individuals by correlation of identities across different application domains and lead to identification without consent. Moreover, it is well known [Dwork and Roth, 2014] that anonymisation with provable guarantees against re-identification attacks is a difficult task. Hence, it is imperative not only to use different virtual identifiers for each application domain making linking impossible [Agrawal et al., 2017], but also to architecturally prevent unauthorised access of information from siloed databases.

Obligation of regulators

The main task of a data regulator should be to ensure that all data accesses are legitimate and that they do not violate consent, purpose limitation or any other rights-based principles. It should be obligatory for all controllers and processors to present their data access and processing requirements to the data regulator for scrutiny. As a part of the process, the associated computer programs for accessing and processing of data must also be audited and pre-approved by the data regulator. Both the data regulator and the data controller should maintain independent and non-repudiable logs - perhaps in a public blockchain (cryptographically secured, distributed ledger) - of all requests and approvals, and the data regulator should issue an authorisation token for each such access request. It should be incumbent on the data regulator to ensure that the data accesses are according to the authorisations granted. The data regulators should also approve and authorise any purpose extension requirements after verifying the legitimacy of the requests based on a rights-based or a consent renewal principle.

For example, if the data controller for the health ministry wants to make parts of an individual's electronic health record available to a requesting health professional, the regulator of the health record database must examine and pre-approve the entire protocol for consent generation and case-by-case verification, in addition to the architecture for data access. There ought to be a data regulator for every database maintained by a data controller, and a data processor may need to obtain approval from multiple regulators. If a longitudinal study to understand the causes of high rate of stunting in India needs access to electronic health records of individuals, and the PDS (public distribution system) purchase records and consumption data of their families, the computer programs for data access and analytics must be scrutinised and pre-approved by the data regulators of the health, PDS, and other consumption-related databases.

All data accesses must be monitored by the regulators, neither the regulator nor the controller should access the data independent of the other, and both must maintain independent non-repudiable logs of all data accesses. The regulatory control should not be so lax that it is ineffective, neither should it be so overbearing or paralytic with inertia that it stifles innovation.

The technology to support such regulatory functions exist [Agrawal et al., 2017], and it may be possible for the data regulators to digitally sign both the authorisations and the computer programs responsible for the data access and processing. In such situations, it should be possible for the regulators to even verify the authorisations and the authenticity of the accessing and processing programs, online and in real-time, before granting access to the data. We believe that building such regulatory capacity and a strong data protection law will be key to effective privacy protection in India.

References

Shweta Agrawal, Subhashis Banerjee, and Subodh Sharma. Privacy and Security of Aadhaar: A Computer Science Perspective. *Economic and Political Weekly*, Vol. 52(Issue No. 37), 16 2017.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3–4):211–407, August 2014. ISSN 1551-305X. doi: 10.1561/0400000042. URL http://dx.doi.org/10.1561/0400000042.

Rahul Matthan. Beyond Consent: A New Paradigm for Data Protection - Discussion Document 2017-03, July 2017. URL http://takshashila.org.in/wp-content/uploads/2017/07/TDD-Beyond-Consent-Data-Protection-RM-2017-03.pdf.

Malavika Raghavan. Before The Horse Bolts, January 2018. URL https://www.thinkpragati.com/think/brainstorm/3180/before-the-horse-bolts/.

B. N. Srikrishna, Aruna Sundararajan, Ajay Bhushan Pandey, Ajay Kumar, Rajat Moona, Gulshan Rai, Rishikesha Krishnan, Arghya Sengupta, and Rama Vedashree. White Paper of the Committee of Experts on a Data Protection Framework for India, 2017. URL http://meity.gov.in/writereaddata/files/white_paper_on_data_protection_in_india_171127_final_v2.pdf. [Online; Accessed January 9, 2018].

The European Parliament and the Council of European Union. Regulation (EU) no 2016/679, 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679.