Individual Case Study 1

52494 MS Information Systems 2613: Database Mgmt Sys - Des, Dev
Professor: Anil Sahai

A Report on
Designing databases that enhance people's privacy without hindering organizations

Kartik Vijay Lande

Table of Contents

1. Summary

Privacy invasive technologies like **Identity related technologies** (biometric technologies, radio frequency identification, bio-implants, DNA sniffers), **location based services** (cellular systems, wireless local area networks, bluetooth, ultrawide band) and **ambient intelligence** (intelligent interfaces supported by computing and network technology which will be embedded in everything) have a lot of benefits but also carry a high risk to an individual's privacy and security with them at the same time.

Despite the existence of a variety of privacy-enhancing technologies like **identity management solutions, pseudonymous credentials, proxy services** that enable users to surf the web anonymously, **P3P protocol** which enables users to inform themselves about a website's privacy policy, **Privacy-protection at the enterprise level** in which the system consults with the privacy policy before using any data for processing, **privacy enabled database systems** in which the database stores privacy metadata for every record and only executes queries if the metadata privacy policy allows it, there is still a need for new approaches to data processing in systems that process personal data of identified individuals.

The author proposes the approach of autonomic databases which further develops existing approaches to privacy-enabled data processing. This entails a data processing system that guarantees by technological measures that data is processed in line with privacy policies. This approach adds to the privacy enhancing technology of privacy protection at the enterprise level by proposing a differentiation between personal data and transactional data. In this approach, individuals are given full access to personal data, but not to transactional data. Individuals can monitor their data through a portal in the following two ways:

1. An individual can view all personal data that is stored about him or her.
2. For every personal data, an organization must state how this data was acquired.

This portal can be a central place where an individual can monitor all the organizations that process his/her data instead of having to manage accounts with each organization.

The author proposes six new principles which if followed can ensure data privacy and security of individuals. They are as follows -

1. **Consent:** People know when their personal data are stored.
2. **Purpose:** People know for what reason their data is being used.
3. **Separation:** Personal data and business data have to be stored separately.
4. **Audit:** Transactions involving personal data must be recorded in logs. People can then follow executed transactions and retrace usage of their personal data.
5. **Participation:** Persons affected have access to their personal data, its usage and purpose specification.
6. **Ease of use:** Persons affected have the choice to bundle access to personal and audit data through portals and can define automatically applied patterns.

The author further talks about a few of the challenges that may arise resulting from the adoption of the above mentioned principles.

2. Technical Challenges

2.1 Complex environments and operations -

Organizations could have their data stored on premise, on cloud they may have a hybrid model in place. Creating a web portal where users have access to their data which is being used by various organizations could be a problem since all organizations do not follow the same method for storing data. The process of retrieving the data out of the organization and displaying it on the web portals for users to monitor it could be different for data stored on premise, on cloud or hybrid models.

2.2 Security -

The fundamental principle of ease of use proposes a portal where users can manage their personal information and the permission settings for all organizations that utilize this data. If this portal were to get compromised by a hacker who has malicious intent then it could lead to a serious data breach. The hacker could also revoke access of data to all the organizations which would disrupt the functioning of many organizations.

2.3 Data Redundancy -

Different organizations may be using different naming conventions to store the same type of data. For example, organization ABC could be storing the amount of time a person spends traveling in a day as daily_travel_time while organization XYZ could be storing the same as daily_commute_time. There would be no way to map these names with standard names to be displayed on the portal to which users have access. This will lead to data redundancy since the same kind of data will be stored under multiple columns but only with different column names.

2.4 Data Automation -

There would be no way to automate the portal in any way because changes will have to be made for every single change that every single organization makes. This would increase the administrative costs required to maintain the portal. For example, if organization Twit were to suddenly shift from MySQL to AWS RDS, the portal would no longer be able to extract the data to display to the users. The data extraction process would have to be significantly modified in order to accomplish this.

3. Operational Consequences

3.1 Memory Challenges -

Logging transactions involving personal data would take a significant effort. This would not only unnecessarily take up space but will also increase the time required to retrieve data from the database significantly.

3.2 Participation -

The fundamental principle of participations requires customers to be informed about further utilization of personal data, and to be able to give their consent for any usage purpose. To fulfill these requirements, customers need access to personal data which is stored on the organizational side. This participation can be realized in different ways, such as per telephone, forms or internet. An organization can't possibly wait for each of its customers to reply back and give their consent to using their personal data before conducting any research or analysis. If the organization were to go with the current data for which they have consent, they would possibly miss out on important data that could've been used if they waited for other customers to give their consent. Results generated using insufficient data would compromise the reliability of the results whereas time wasted while waiting for sufficient data could give a competitor an advantage.

3.3 Audit -

The fundamental principle of audit requires organizations to maintain data related to executed transactions which accessed personal data on their side. They also have to provide access of this data to users whose data has been used. If users were to gain access to someone else's transaction data, it could lead to security breaches. For example, if an employee were to gain access to another employee's salary transaction that could be a security problem because salary contracts are confidential data.

3.4 Change of Permission -

The fundamental principle of ease of use proposes a portal where users can manage their personal information and the permission settings for all organizations that utilize this data. If users were to suddenly revoke access to a particular data which organizations previously had access to, then this could lead to problems in the functioning of the organization.

4. Business Consequences

4.1 Purpose Specification -

The basic motive of every organization is to generate revenue. Giving purpose specifications to individuals on how their data is being used can have serious effects on the ability of an organization to generate revenue. An organization may lose competitive advantage by making their strategy or the way they use data public. For example, let's assume that a social media platform like facebook uses data like the kind of groups an individual is a part of and uses this data to train a machine learning classification model to predict if an individual's political views are right aligned, left aligned or center aligned. The social media platform could be selling this information to political parties so they can target their audience better for votes. If this information were to become public, competitors could adopt the same approach and hurt the social media platform's revenue generated by selling information to political parties.

4.2 Separation -

Separating personal data from business data is a good way to ensure individual privacy but significant efforts and costs will be incurred in order to carry this out. Although separating data sounds like an easy task, it could very easily create complications in the IT infrastructure. Skilled personnel will need to be hired to design the new entity relationship diagram and implement the database while avoiding redundancy and maintaining integrity at the same time.

5. Areas of Problems/Shortcomings

5.1 Consent -

With the emergence of ambient intelligence, the constant surveillance to which internet companies have acclimated their users is slowly making its way offline, where it becomes increasingly difficult to opt out of these systems [1]. Opting out of data collection systems often means not being able to use online services. Refusing to participate in these systems often means staying off the internet entirely, which comes with massive social and economic costs. The question therefore isn't whether you should opt out; it's a question of whether you can afford to [1].

6. Recommendations for improvement

6.1 Legislative Measure -

There would be no way to check if the organizations are using the data like it claims to be doing on the platform**.** An international data protection law must be passed, which requests the explicit consent of a person before personal data can be stored. Furthermore, the law stipulates that this person must have access to their data, to specify purposes and to control audit information.

6.2 Security -

Two Factor Authentication could be implemented to secure access to the portal. Encryption decryption techniques can be implemented while transferring the data from the organizations to the portal. Actively managing passwords and user access, testing database security, using real-time database monitoring, using database firewalls could be a few ways the security of databases can be improved [2].

6.3 Permissions -

Users should be allowed to set their permissions only when they first share their data with the organization. They can continue to monitor their data through the portal but not change the permissions later on. This will avoid problems discussed in 3.4

7. References

[1]https://slate.com/technology/2019/08/consent-facial-recognition-data-privacy-technology.html
[2]https://www.imperva.com/learn/data-security/database-security/