Privacy Enhancing Technologies and Statistical Disclosure Control Methods

Jouni Markkula
University of Oulu, Finland

INTRODUCTION

Privacy is gaining increased attention in our current world of advancing Internet services and extending applications of pervasive technology. The rapid development of information and communication technology has provided expansive computational power and introduced new means of collecting more data about individuals, together with sophisticated methods to process it. The recent trend of Big Data demonstrates the understanding of various private and public sector organizations of the value of extensive data collection for developing innovative services and for business purposes. In many cases, personal data represents an essential, and often the most valuable, element of the data for the organizations. This is evident, for example, from the growth of Google and Facebook and their position in almost everybody's life. Collected personal data can be used for our benefit, for example by providing us with novel innovative services. However, and conversely, the exhaustive personal data collection and processing is a serious challenge to people's privacy sphere. For these reasons, personal data privacy has become an increasingly important concern for both service users and data intensive system developers.

The data privacy issue is based on two aspects: the systematic collection of data about individuals and the development of technology. The first one has its old roots in statistical data collection by governments, e.g. census data, for public purposes. The second is related to the information technology that enables the more efficient collection and processing of data. The modern information technology, to which our present day privacy concerns relate, dates back to the 1960s. When its potential power for personal data collection was realized, there emerged a need to define practices of data management, which led to the formulation of Fair Information Principles (FIPs) and later to data privacy legislation in many countries. The conceptualization of privacy in the form of FIPs, and emphasizing its fundamental significance in the form of regulation, has led to the development of methods to protect personal data in two disciplines that correspond to the two aforementioned aspects underlying data privacy: computer science and statistics. Although these disciplines and their solutions have similarities and overlap in many respects, they take a different view of, and approach to, privacy issues. In computer science, the approach is usually more technological and information system-related. The computer science solutions are typically referred to as Privacy Enhancing Technologies (PETs). In statistics, the view is essentially data centric, with data seen as a register of personal information, and the technological aspects are not in focus. In statistics, the solutions are known as Statistical Disclosure Control (SDC) methods.

The objective of this article is to present the two distinct viewpoints of the aforementioned disciplines for solving present day highly challenging data privacy issues. The article discusses the relationship between the approaches, suggesting that future development of better and more comprehensive data privacy solutions would gain from a wide exchange of the specialized knowledge of these disciplines. Integration of the views and approaches of the different fields would have a potential impact on the development of improved and more generally usable data privacy methods to fit the changing technologies and organizational environments of the future.

BACKGROUND

Discussion of data privacy has a history that dates back to the 1960s. A seminal work in this area is Alan Westin's book Privacy and Freedom (Westin, 1967). In 1962, the Special Committee on Science and Law of the Association of the Bar of the City of New York

DOI: 10.4018/978-1-4666-5888-2.ch430

proposed a more formal study of the impact of modern technology upon privacy. Professor Westin was selected to organize the committee's studies and to direct its research, which finally led to the publication of the book. Westin (1967) noted that problems of privacy were posed by familiar and increasingly pervasive items: the miniature battery-powered microphone, the extension telephone, the portable (and concealable) tape recorder, and the small high-resolution camera. This statement also sounds familiar and topical today if the older examples of technology are replaced with smartphones and sensor networks.

Since then, privacy principles have been expressed in Fair Information Principles (FIPs) and in privacy-related regulation. The FIPs were first specified by the U.S. Department of Health, Education, and Welfare (1973), which defined privacy in the following way:

An individual's personal privacy is directly affected by the kind of disclosure and use made of identifiable information about him in a record. A record containing information about an individual in identifiable form must, therefore, be governed by procedures that afford the individual a right to participate in deciding what the content of the record will be, and what disclosure and use will be made of the identifiable information in it. Any recording, disclosure, and use of identifiable personal information not governed by such procedures must be proscribed as an unfair information practice unless such recording, disclosure or use is specifically authorized by law.

The original FIPs were defined in the report for record-keeping organizations, which stated that the organizations should follow certain fundamental principles of fair information practice:

- There must be no personal-data record-keeping systems whose very existence is secret
- There must be a way for an individual, to find out what information about him is in a record and how it is used
- There must be a way for an individual to prevent information about him obtained for one purpose from being used or made available for other purposes without his consent
- There must be a way for an individual to correct or amend a record of identifiable information about him
- Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take reasonable precautions to prevent misuse of the data

Since then, the FIPs have spread widely internationally and transformed in form. The Organization for Economic Cooperation and Development (OECD) presented its own FIPs in 1980, which were updated as recently as 2013 (OECD, n.d.). In the U.S., FIPs are currently presented by the Federal Trade Commission (n.d.), and they are also underlying in European Union data protection legislation (EC Directive 95/46/EC; EC Directive 2002/58/EC).

Underpinning data privacy regulation and FIPs is an understanding that personal data is valuable for society at large and ultimately also for individuals, and its collection should not be prevented. Therefore, personal data collection, processing, and disclosure should only be controlled by a means to ensure it does not compromise personal privacy. This is explicated, for example, in the original FIPs (U.S. Department of Health, Education, and Welfare, 1973):

The safeguards we recommend require the establishment of no new mechanisms and seek to impose no constraints on the application of electronic data-processing technology beyond those necessary to assure the maintenance of reasonable standards of personal privacy in record keeping. They aim to create no obstacles to further development, adaptation, and application of a technology that, we all agree, has brought a variety of benefits to a wide range of people and institutions in modern society

The value of personal data in many cases is clear, for example for statistical information production to aid prediction and planning, as well as for business purposes. However, appropriate measures and technologies should be used to protect personal data privacy, and further developed in response to information technology development, e.g. Big Data processing and data analytics.