Privacy Challenges for Database Systems

Prof. Johann-Christoph Freytag, Ph.D.
DBIS @ Humboldt-Universität zu Berlin
freytag@dbis.informatik.hu-berlin.de

Abstract

Over the last years the means to collect personal data implicitly or explicitly over the Web and by various kinds of sensors and to combine data for profiling individuals has dramatically increased. These improved abilities have a dramatic impact on the privacy of every individual. This talk focuses on presenting and discussing techniques, concepts, and algorithms how to prevent privacy breaches when personal data is either stored or individuals are at risk to reveal which data they access.

In the first part of the talk we introduce several examples of privacy breaches that help us to better understand what is at risk and to categorize the kind of privacy threads and privacy attacks. These examples will give us a better understanding of the different attacks on privacy as observed on the real world. They also show us that privacy is at risk in different fields such as in the field of communication and in the field of database systems.

The latter will be the focus for the rest of the presentation. To begin with we first discuss some general principles called the principles of Hippocratic database systems that should guide any privacy solution for database systems. These principles rest in part on various requirements coming from the privacy (non computer science) community and from the legal world. They will help us in better understand the different solutions and concepts to protect the privacy of individuals when storing personal data in a database system.

Based on the notion of access privacy we describe solutions to protect the users privacy when access a database such that (s)he does not reveal which data is accessed in the database. We present different solutions with different levels of privacy always assuming the same model of privacy leakage.

We then turn to the notion of data privacy. The challenge of data privacy is to prevent the linkage of stored data to individuals. For example, if we store patient and disease data, it should be impossible to find out which individual has which disease when the data is released. As a first step we introduce the notion of k-anonymity as a way to release data without revealing the identity of the individual they belong to. We show that the basic notion of k-anonymity does not suffice; it must be improved in various ways due to the different kind of privacy leakages (or possible attacks) that might occur. Throughout the talk we also discuss the ques-

tion of how to measure privacy and the leakage of privacy. Whenever appropriate solutions are known we include those in our talk.

During the last part of the talk we extend the notion of privacy to distributed systems, in our case to mobile cooperative distributed systems as found in Intelligent Transportation Systems (ITS), by presenting some initial results of the EU project PRECIOSA which we participate in.

Some of this research work was performed as part of the EU project PRECIOSA. I would like to thank Martin Kost and Lukas Dölle, Ph.D. students of the DBIS research group at Humboldt-Universitt zu Berlin who have been essential in putting together this presentation.