Abstract: "Need to know" security for data analysis

The massive amounts of information that are being collected about each of us will only increase as sensors become ever cheaper and more powerful. Analysis of this wealth of data supports advances in medicine and public health, improved software and services through user pattern analysis, and more efficient economic mechanisms. At the same time, the potential for misuse of such data is significant. A long-term research question is how best to support beneficial uses while inhibiting misuse.

One approach is to enable individuals to maintain tighter control of their own data while still supporting the computation of group statistics. Currently, analysts are usually given access to all data in order to compute statistics, and often use a third party service provider to store, or even process, such data. Either the third party has access to all data or the data are encrypted, in which case the third party does no processing. An interesting research question is how to provide mechanisms to support "need to know" security in which an individual has full access to her own data, the third party learns nothing about the data but can nevertheless contribute to the processing, and the analyst learns only the desired statistics. We have explored "need to know" security in connection with MyUnity, a prototype awareness system.

MyUnity collects data from a variety of sources and displays summary presence states, such as ``in office'' or ``with visitor,'' computed from the received data. MyUnity was deployed in a small research lab and has been in use by over 30 people for more than a year. To avoid concerns about misuse, the system did not store any data initially. The researchers developing the system were interested, however, in analyzing usage patterns, and users expressed interest in seeing personal trends, activity patterns of coworkers, and long-term data pooled across groups of users, all requiring data to be stored. At the same time, users continued to be concerned about misuse of stored data. We looked at ``need to know'' security for cases in which, at each time step, each member of a group of users has a value (i.e., a presence state) to contribute, and the group would like to provide only an aggregate view of those values to people outside their group.

We designed and implemented an efficient protocol that enables each user to encrypt under her own key in such a way that a third party can compute an encryption of a sum across values encrypted under different keys without the need for further interactions with the individuals. The protocol provides means for an analyst to decrypt the encrypted sum. We designed key structures and extensions to provide a family of efficient non-interactive ``need to know'' protocols for time series data in which the analyst learns only the statistics, not the individual data values, and the third party learns nothing about the values.