# Collecting digital user data without invading privacy

March 6 2014

Saarbrücken computer scientists enable computing aggregate statistics about digital user data without undermining the privacy of users. Credit: Oliver Dietze

The statistical evaluation of digital user data is of vital importance for analyzing trends. But it can also undermine the privacy. Computer scientists from Saarbrücken have now developed a novel cryptographic method that makes it possible to collect data and protect the privacy of the user at the same time. They present their approach for the first time

at the computer expo CeBIT in Hannover at the Saarland University research booth (hall 9, booth E13).

"Many website providers are able to collect data, but only a few manage to do so without invading users' privacy", explains Aniket Kate, who leads the research group "Cryptographic Systems" at the Cluster of Excellence "Multimodal Computing and Interaction" (MMCI) in Saarbrücken. Two aspects threaten privacy during data aggregation: On the one hand, where and how is the data aggregated? For example, website owners are interested in the age and gender of their visitors. Therefore, they store data files (cookies) on their computers that observe which other websites they visit. "But this wealth of sensitive information allows them also to reconstruct detailed profiles of each individual", says Kate. On the other hand, it is important to publish aggregated data in a privacy-preserving way. "Researchers have already demonstrated that precise information about the habits of citizens can be reconstructed from the electricity consumption information collected by so-called smart meters", explains Kate.

In cooperation with his colleagues Fabienne Eigner and Matteo Maffei from the Center for IT-Security, Privacy and Accountability (CISPA) and Francesca Pampaloni from the Italian IMT Institute for Advanced Studies Lucca, Kate developed a software system called "Privada". It is not only able to resolve the dilemma between the desire for information and the protection of data, but it can also be easily applied in different domains. "For example, with Privada website owners are still able to observe that their websites are mainly visited by middle-aged women, but nothing more", Kate explains.

To achieve this, users split up the requested information and send parts of it to previously defined servers performing multi-party computation: Each server evaluates its data without being aware of the data of other parties. So together they compute a secret, but are not able to decode it

on their own. Moreover, each party adds on a value corresponding to a probability distribution to make the data a little bit imprecise. The perturbated partial results are assembled into the actual analysis. The perturbation ensures that the identity of the individual person is protected, while trends are still significant in the aggregated statistic about user data.

The privacy is even guaranteed if all but one of the servers collaborate. Hence, according to the researchers, it is even conceivable that companies could provide such servers. If only servers, and not users, perturb the data with a certain amount of noise, that has two advantages: Firstly, not much computational power is necessary on the user's side. Hence, even a mobile phone could send the partial result to a particular server. Also, in total, there is only a minimal amount of noise attached to the aggregated data. Hence, the resulting statistic about user data is as accurate as possible.

The computer scientists from Saarbrücken have already implemented their concept. "The computation is fast; the servers just need a few seconds", says Fabienne Eigner, part of the research group "Secure and Privacy-preserving Systems" at Saarland University. She also worked on the software system. The architecture is constructed in such a way that it would not make any difference if someone were to analyze the data of a thousand or a million people", explains Eigner.

Provided by Saarland University