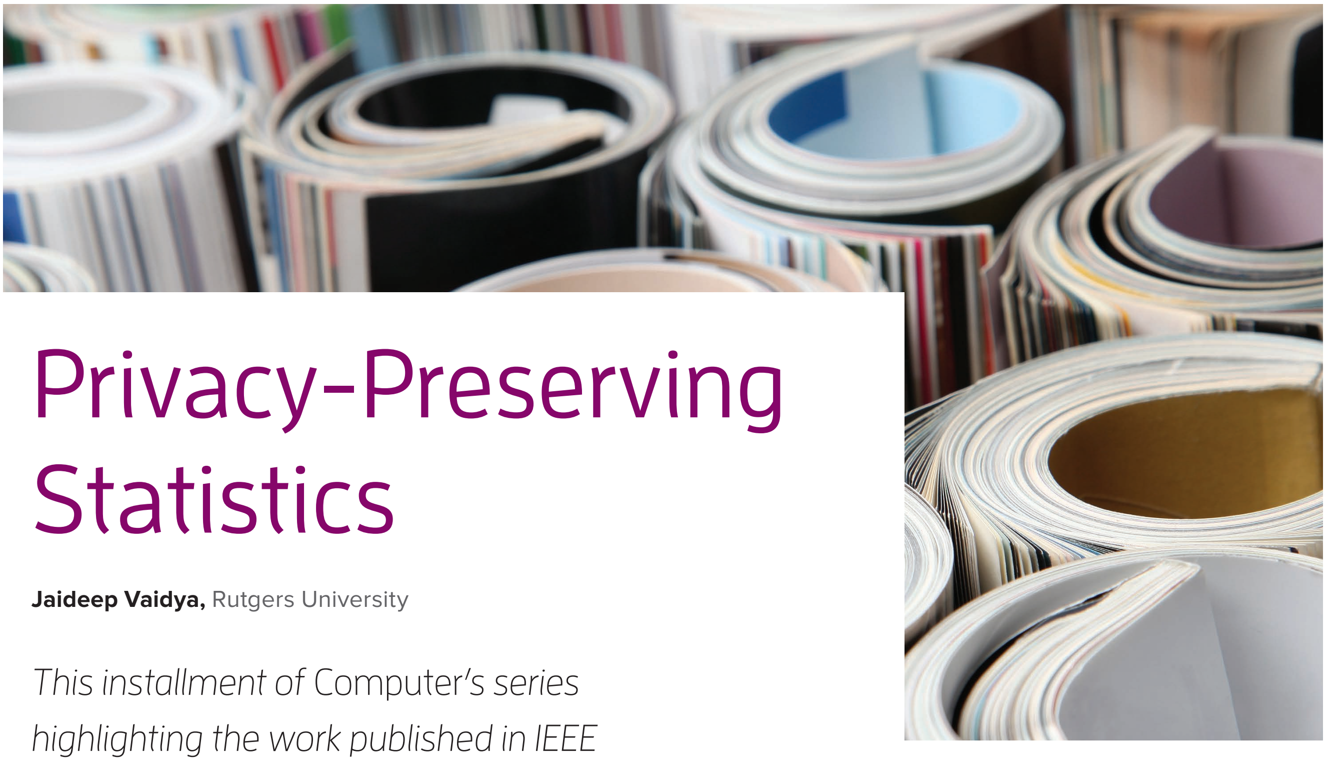

Privacy-Preserving Statistics

Jaideep Vaidya, Rutgers University

This installment of Computer's series highlighting the work published in IEEE Computer Society journals comes from IEEE Transactions on Dependable and Secure Computing.

In 1982, Andrew Yao formulated what has come to be known as Yao's Millionaires' Problem, wherein two parties (millionaires), each holding a number (their respective net worth), want to determine who has the larger number (that is, larger net worth) without revealing the actual values. This was one of the seminal works in computing and was noted in Yao's Turing award citation. In later years, this was generalized to the problem of multiple parties that wanted to collaboratively compute any general function over their private inputs, and became known as secure multiparty computation, or SMC. There has been significant theoretical work in this area, with general solutions developed that are resilient to different adversarial models. However, although SMC has been well established now for nearly four decades, it remained primarily of academic interest for many years.

Recently, there has been great interest in using these academic solutions for practical applications. Surprising applications have emerged, such as the use of an SMC-enabled auction to find the market clearing price of sugar beets in Denmark. DARPA also asked cryptographers to develop SMC protocols for satellite orbital data sharing to avoid collisions such as the one in 2009 involving the US's Iridium 33 satellite and the Russian's Cosmos-2251.

In "Rmind: A Tool for Cryptographically Secure Statistical Analysis," Dan Bogdanov, Liina Kamm, Sven Laur, and Ville Sokk describe the first privacy-preserving statistical analysis environment supporting a complete SMC-based data analysis process in which data are collected from various sources, linked, and statistically analyzed (*IEEE Trans. Dependable and Secure Computing*, vol. 15, no. 3, 2018, pp. 481–495). Rmind aims to provide a familiar user interface (similar to R) while enabling private computation.

Rmind is based on the Sharemind platform (developed by the same team), which enables secure computation using additive secret sharing and hardware isolation. Optimizations made in the Sharemind SMC system that Rmind uses for private arithmetic have made Rmind significantly faster, reducing the running time of certain statistical workflows by nearly 75 times.

0018-9162/18/$33.00 © 2018 IEEE

Furthermore, the Rmind team reports that, at the request of Rmind users, additional privacy-preserving features—including grouping with various aggregation functions, date support, outer database joins, multicolumn joins, logistic regression, principal component analysis, and various ease-of-use and access control features—have since been implemented for real-world studies.

The first landmark use of Rmind on real-world data took place in 2015, when the Estonian Centre for Applied Research (CentAR) used Rmind to collect, link, aggregate, and analyze 10 million personal real-world tax records with more than half a million education records to analyze the relationship between working during studies and failing to graduate. This study also set a data protection precedent. After reviewing the procedures and technology behind the use of Rmind, the Data Protection Authority ruled that no personal data was processed as defined in the Data Protection Act. This precedent created a separate branch of legal research that suggested that a similar claim could be made under the EU's General Data Protection Regulation (GDPR).

The future of privacy-preserving statistics is bright. For example, in the US, the "Student Right to Know Before You Go Act," introduced by Senators Ron Wyden, Marco Rubio, and Mark Warner, suggests using SMC to analyze data about graduation rates, debt levels, and postgraduation earning potential. Thus, this area of research remains important, and tools such as Rmind have significant potential for widespread adoption. C

JAIDEEP VAIDYA is the RBS Dean's Research Professor in the Management Science and Information Systems Department at Rutgers University. Contact him at jsvaidya@business.rutgers.edu.