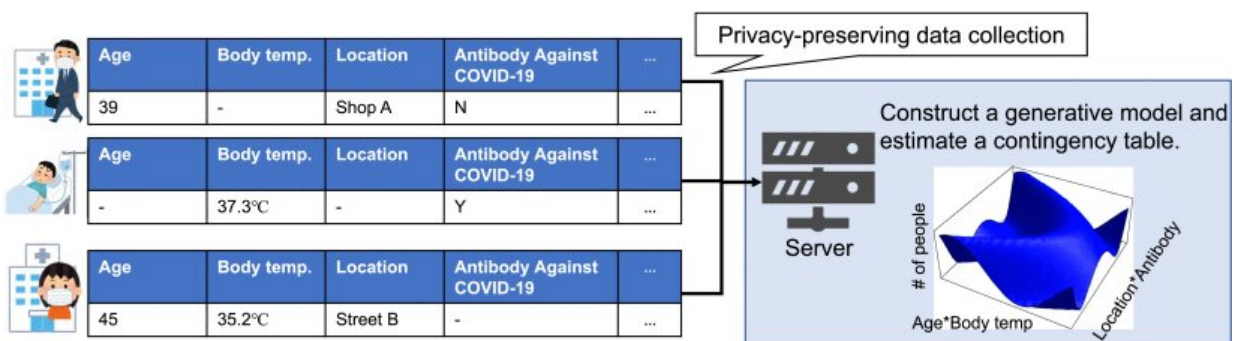


# Study addresses privacy-preserving collaborative data collection and analysis with many missing values

July 7 2023



Example application of the proposed privacy-preserving data collection method. Credit: *IEEE Transactions on Dependable and Secure Computing* (2022). DOI: 10.1109/TDSC.2022.3174887

To control pandemics like the novel coronavirus infection (COVID-19), data such as the age, gender, family composition, and medical history of infected individuals are required. While patients themselves may provide this information to medical institutions, these details are highly confidential.

If the data is properly handled for [privacy protection](#), it can be shared with researchers worldwide without identifying the infected individual, which can help clarify the state of the pandemic and more accurately

predict its progression.

There may be missing values in the information provided by patients, and existing methods do not take these missing values into account when collecting personal data while ensuring privacy. This is leading to a significant reduction in the accuracy of data analysis.

Differential Privacy, the privacy protection metric addressed in this paper, is adopted by many organizations, including Apple, Google, Microsoft, and LINE. Numerous methods have been proposed to collect and analyze [personal data](#) based on Differential Privacy. However, none of the existing methods take into account the presence of missing values.

When considering medical data, as during the COVID-19 pandemic, it is conceivable that different hospitals can obtain different information, and many patients may feel comfortable providing only some data after privacy protection processing. Under the current methodology, the accuracy of analysis is greatly reduced in such scenarios, which has prevented sufficient data analysis for pandemic mitigation.

Professor Sei has demonstrated that using the Copula model, primarily used in the finance field, can restore the true statistical model from data processed by Differential Privacy technology even in situations with many missing values, enabling highly accurate data analysis. Of course, he mathematically proves that each individual's [privacy](#) is strictly protected at the exact same level as existing methods.

In real society, data typically has various missing elements. By using the proposed method, not only [medical information](#) but also various societal and [personal information](#) with missing values can be safely analyzed with high accuracy. Therefore, this research is expected to have a significant impact on society.

The study is published in the journal *IEEE Transactions on Dependable and Secure Computing*.

**More information:** Yuichi Sei et al, Privacy-Preserving Collaborative Data Collection and Analysis with Many Missing Values, *IEEE Transactions on Dependable and Secure Computing* (2022). [DOI: 10.1109/TDSC.2022.3174887](https://doi.org/10.1109/TDSC.2022.3174887)

Provided by The University of Electro-Communications

Citation: Study addresses privacy-preserving collaborative data collection and analysis with many missing values (2023, July 7) retrieved 14 April 2024 from <https://techxplore.com/news/2023-07-privacy-preserving-collaborative-analysis-values.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.