Cryptographic methods enable analyses without privacy breaches

To advance large-scale biomedical research projects, investigators often need to share data with colleagues around the world while not violating the privacy of their study subjects. Traditionally, researchers simply stripped their data sets of personally identifiable information, such as names or Social Security numbers. But as several recent studies have shown (see go.nature.com/wtHCHS), it's now possible to deduce the identity of some research participants from clinical and genomic records once considered completely anonymous.

In response to growing privacy concerns, many research teams have begun to test various cryptographic data techniques. These secure methods enable mathematical operations on encoded data sets so that only the results of computations are revealed, but nothing else. According to Brad Malin, a computer scientist at Vanderbilt University in Nashville, Tennessee, who works on data protection issues, the hope is that such practices will facilitate data sharing and the use of online cloud computing, creating "opportunities to accomplish things that you couldn't have done otherwise."

A recent example of this approach comes from Khaled El Emam of the University of Ottawa in Canada. El Emam and his colleagues adapted a secure protocol they had previously developed for broad disease surveillance purposes to investigate the infection rates of drug-resistant bacteria among residents of long-term-care homes. "Across Canada, there had been incidents of healthcare providers trying to hide this kind of data" to protect their reputations and the privacy of their patients, El Emam says.

To get around these obstacles, El Emam's team analyzed the encrypted records from the vast majority of the nursing homes and assisted living facilities in Ontario—more than 500 in total. Their findings on the geographic distribution and prevalence of superbugs in the Ontario care homes were consistent with test results from the province's hospitals. According to El Emam, who reported the results 8 April in *PLoS One* (**9**, e93285, 2014), this congruency demonstrates the feasibility of collecting epidemiological data while still providing strong privacy and confidentiality assurances. "We have been able to do things that were once perceived to be extremely hard," he says.

Tales from the encrypt: Secure data analysis methods are proving their worth.

Other researchers are beginning to apply similar encryption strategies to the analysis of DNA data. As *Nature Medicine* went to press, a team in Switzerland was about to deploy a cryptographic scheme that enables infectious disease physicians to analyze the genetic risk profiles of HIV-infected individuals without gaining access to the patients' full genome sequences. This should allow doctors to prescribe personalized drug treatments while minimizing the potential liability risks associated with incidental genomic discoveries, says study leader Jean-Pierre Hubaux, a data protection expert at the Swiss Federal Institute of Technology in Lausanne.

Another goal, he adds, is to see "how much time [it takes] and how intrusive or how annoying it is going to be to work on data that are protected as opposed to working on the immediately accessible data."

Building trust

Hubaux plans to unveil his team's research effort next month at a one-day Workshop on Genome Privacy, which will be held in conjunction with the Privacy Enhancing Technologies Symposium in Amsterdam. The program for the event is still under development, but Hubaux, a coorganizer, expects that other presenters will also discuss new strategies for computing on encrypted data. More fundamentally, he says the meeting aims to facilitate communication between members of the genomics and information security communities, which have traditionally had difficulties understanding each other.

Ultimately, spreading the use of cryptographic programs beyond the security field could be a hard sell, as some biomedical scientists might resist platforms they don't understand and can't easily deconstruct. "To some extent, you will have to just trust that the software is correct and that no one's cheating you," says Joshua Swamidass, a bioinformaticist at the Washington University School of Medicine in St. Louis.

Swamidass hopes that, rather than simply enabling privacy-protecting versions of existing computations, encryption will create new research opportunities. For example, most geneticists currently work with relatively small numbers of patient samples that they have full access to, "so the idea of doing an analysis through a cryptologic strategy isn't yet exciting to them," he says. That may change as researchers begin to pool data from millions of patient samples housed by various repositories around the world and thus start to need more sophisticated security protections.

"In that context, there's no chance that we're going to be able to put all that data in a central repository," Swamidass says, "but maybe we could analyze it in an encrypted way so that everyone can protect the private information and still get the associations that we wouldn't have been able to get otherwise."

That idea is something that Dixie Baker thinks about all the time now. Baker is a health information technology expert who co-chairs the Security Working Group of the Global Alliance for Genomics and Health, a massive worldwide effort to develop common policies for how to manage biomedical data. In April, her team released a first set of priorities for developing security and privacy protection standards for the alliance's partner organizations, which include close to 200 public and private research institutions, healthcare providers and research funders.

"It's doubtful that we will adopt methods that people are not familiar with or comfortable with," Baker says. "When you recommend that people implement something at an international level, you really want it to be something that's fairly well proven."

It's up to security researchers to now prove their case to the biomedical community.

Nicholette Zeliadt