Genetic materia

The more we know about our genome data, the better our doctors will be able to treat us in the future. But how can we make use of this sensitive data, without allowing it to be misused? The IT specialists around Stefan Katzenbeisser and Kay Hamacher from the Technische Universität Darmstadt want to encrypt genome data so skilfully that it is still possible to carry out mathematical analyses.

By Boris Hänßler

Dirk von Gehlen is a journalist who works for the Süddeutsche Zeitung newspaper, which is published in Southern Germany. He was recently visited by colleagues from the North German Broadcasting Corporation, Norddeutscher Rundfunk (NDR). They brought him a USB stick on which was stored a complete month of his browser history – the websites he had visited, the search keys he had entered in Google and the train journeys he had booked. The NDR reporter had bought the data over the Internet. Gehlen was flabbergasted! He had no idea that a company was using a harmless browser add-on to secretly record the Internet activities of millions of users and offer the data for sale internationally.

"Once you agree to deposit your genome data, it is not so easy to withdraw it from circulation."

Researchers fear that something similar could happen in future with genome data that provides ever-deeper insights into our biological identity. A few years ago, for example, customers of the American company 23andme were already paying a fee and sending in a saliva sample. The company analysed the inherited genetic variations – the so-called SNPs (Single Nucleotide Polymorphisms). From this, they could discern whether someone has an increased risk of developing cancer, Huntington's disease or Parkinson's disease. Admittedly the American authorities banned the transaction, because they feared that customers could misunderstand the findings without medical advice. But the company is still able to keep on collecting genome data, so that now it can ascertain the customers' genetic parentage. The data is digitally stored and could theoretically be sold on. It would be worth its weight in gold to medical insurance and life insurance companies.

But despite everything, it is not a good idea to ban the storage and use of this data. Because it could revolutionise medicine. "Genome data is the basis for personalised medicine", says Kay Hamacher, bio-informatics specialist at the TU Darmstadt. "Behind it is the vision that in future, doctors will be able to offer their patients individually tailored forms of treatment on the basis of genetic information." The genome data could, for example, provide information about whether or not someone will tolerate certain medication, or whether a particular treatment would work really well.

Kay Hamacher and Stefan Katzenbeisser from the Cybersecurity (CYSEC) profile area are looking to harness genome data for these purposes, whilst making its misuse as impossible as current crypt-technology allows. The risk is always present, when doctors and clinics release the data for research. Genome research is assigned to powerful computers, so IT service providers often have to be involved, using super-computers to trawl through the data. "So we need a method that, although it encrypts the data, still allows its use in subsequent calculations", says Stefan Katzenbeisser. "It must never be possible for the service provider performing the calculation to see the unencrypted data."

The method is called homomorphic encryption. The simplified example below shows how it works. The numbers "1" and "2" are sent to a service provider as the encrypted values A and B. The service provider can add up A and B, and send the result C back to the client. But the service provider does not know the value of A, B or C. The client, on the other hand, can decrypt C to reveal the result, in this case "3". Highly complex calculations can be carried out in a similar way.

But that is not all there is to it. The data sets of genome data are huge. Researchers or drug manufacturers therefore concentrate only on the SNPs or mutations of the DNA that are relevant to their current question. As a result, the IT service provider can theoretically deduce from their access to the sequence, what the researchers are currently working on – even if the result of the query remains encrypted.

"The DNA string I am examining reveals a lot about the diseases and substances I am working on", says Katzenbeisser. "To prevent this, we carry out a kind of deception, the so-called oblivious RAM. The physical memory is constantly scrambled while the database is being consulted. No-one can then comprehend whether the questioner is accessing the same data several times, or has accessed a lot of different data. So the intention of the query is disguised."

Information

Cybersecurity profile area
Prof. Dr. Stefan Katzenbeisser
Email: katzenbeisser@seceng.informatik.tu-darmstadt.de
Prof. Dr. Kay Hamacher
Email: hamacher@bio.tu-darmstadt.de
www.cysec.tu-darmstadt.de

l on offer

The research is run by the Collaborative Research Centre funded by the German Research Foundation (DFG) "CROSSING – cryptography-based security solutions: enabling trust in new and next generation computing environments" and the Centre for Research in Security and Privacy (CRISP), based in Darmstadt. Hamacher and Katzenbeisser's teams first want to devise the basic techniques for the cryptographic process. As these are highly complex, the researchers are also developing tools to implement the process without error. IT staff can thus implement the necessary protocols, even if they have no cryptography expertise.

The processes must also become even more powerful. Thus far, most of the techniques have focused on smaller sets of data. Working with genome data would be beyond them. The genome of a human being has an information content of 100 megabytes to 200 gigabytes – depending on whether you store the complete DNA, or just the mutations. Added to this is the fact that for meaningful investigations, you need to take into account the genome data of as many people as possible. It all adds up.

"It is also our wish to consult small clinics any clinics and institutions where genome data is present", says Katzenbeisser. "They are more sceptical, as they only collect data from a few patients. You could easily draw conclusions about their identity. The encryption processes should make the clinics feel safe about contributing this type of data to this important research."

Greater confidence in the infrastructure of genome research is urgently required. Otherwise Germany will be at a disadvantage. American companies are already investing a lot of money in these areas. Although the details of what they actually intend to do with the data are vague. Kay Hamacher says: "Once you agree to deposit your genome data, it is not so easy to withdraw it from circulation. With genetic material, you are also making a decision for your mothers, fathers and grandchildren at the same time. Last but not least, we must remember that we are only at the start of genome research – we do not know what will be possible to glean from it in future. So it is very important that we get to grips with data security early on."

The author is a technology journalist.

Photo: Katrin Binner

Professor Stefan Katzenbeisser develops special methods to encrypt genomic data.

Names and facts

Prof. Stefan Katzenbeisser supervises the S5 Privacy-Preserving Computation project of the CROSSING Collaborative Research Centre and is head of the Security Engineering Department at TU Darmstadt. Prof. Kay Hamacher supervises the Scalable Privacy-Preserving Protocols project in the Centre for Research in Security and Privacy (CRISP), which is funded by the federal government and the state of Hesse, with substantial support from TU Darmstadt. He also leads the Computational Biology & Simulation research group at TU Darmstadt.