Erbgut im Ange

Je mehr wir über unsere Genomdaten wissen, desto besser können uns Ärzte künftig behandeln. Doch wie lassen sich diese sensiblen Daten nutzen, ohne dass sie missbraucht werden? Informatiker um Stefan Katzenbeisser und Kay Hamacher von der Technischen Universität Darmstadt möchten Genomdaten so geschickt verschlüsseln, dass man dennoch mathematische Analysen durchführen kann.

Von Boris Hänßler

Dirk von Gehlen heißt ein Journalist der Süddeutschen Zeitung. Kürzlich bekam er Besuch von Kollegen des Norddeutschen Rundfunks. Sie brachten ihm einen USB-Stick vorbei, auf dem die lückenlose Browser-Historie eines Monats abgespeichert war – welche Webseiten er besucht, welche Suchbegriffe er bei Google eingegeben und welche Bahnreisen er gebucht hatte. Die NDR-Reporter hatten die Daten über das Internet gekauft. Gehlen war baff: Er hatte keine Ahnung, dass eine Firma mittels einer harmlosen Browser-Erweiterung die Internet-Aktivitäten von Millionen Nutzern heimlich aufzeichnete und die Daten international zum Verkauf anbot.

Auf ähnliche Weise, so befürchten Forscher, werden künftig auch Genomdaten gehandelt, die einen immer tieferen Einblick in unsere biologische Identität gewähren. Vor einigen Jahren etwa konnten Kunden des amerikanischen Unternehmens 23andme bereits gegen Gebühr eine Speichelprobe einsenden. Die Firma wertete die geerbten genetischen Varianten aus – die sogenannten SNPs (Single Nucleotide Polymorphisms). Aus ihnen lässt sich herauslesen, ob jemand ein erhöhtes Risiko aufweist, zum Beispiel an Krebs, Huntington Disease oder Parkinson zu erkranken. Die amerikanischen Behörden verbaten zwar das Geschäft, weil sie befürchteten, dass die Kunden ohne ärztliche Beratung die Ergebnisse missverstehen könnten. Aber das Unternehmen darf nach wie vor Genomdaten sammeln, um nun die genetische Abstammung der Kunden zu ermitteln. Die Daten werden digital gespeichert und könnten theoretisch weiter verkauft werden. Für Kranken- oder Lebensversicherungen wären sie Gold wert.

„Wer einmal einwilligt, Genomdaten zu hinterlegen, kann sie nicht so einfach wieder aus dem Verkehr ziehen."

Trotz allem ist es keine gute Idee, die Speicherung und Nutzung dieser Daten zu verbieten. Denn sie könnten die Medizin revolutionieren. „Die Genomdaten sind Grundlage für die personalisierte Medizin", sagt Kay Hamacher, Bioinformatiker an der Technischen Universität Darmstadt. „Dahinter steht die Vision, dass Ärzte künftig basierend auf den Erbgutinformationen eine individuell auf Patienten zugeschnittene Therapieform anbieten können." Die Genomdaten könnten zum Beispiel Hinweise darauf liefern, ob jemand ein Medikament nicht verträgt oder eine bestimmte Therapie besonders gut funktionieren würde.

Hamacher und Stefan Katzenbeisser vom Profilbereich Cybersecurity (CYSEC) der TU möchten aus diesen Gründen Genomdaten nutzbar machen, ohne dass sie missbraucht werden können. Die Gefahr ist immer dann gegeben, wenn Ärzte und Kliniken die Daten für die Forschung frei geben. Die Genomforschung ist auf leistungsstarke Rechner angewiesen, daher müssen oft IT-Dienstleister involviert werden, die mit Hilfe von Super-Computern die Daten durchforsten. „Wir benötigen somit ein Verfahren, bei dem die Daten zwar verschlüsselt werden, bei dem aber dennoch nachträgliche Berechnungen möglich sind", sagt Stefan Katzenbeisser. „Der Dienstleister, der die Berechnung durchführt, darf also zu keinem Zeitpunkt die Möglichkeit haben, die unverschlüsselten Daten einzusehen."

Das Verfahren nennt sich homomorphe Verschlüsselung. Das folgende, vereinfachte Beispiel zeigt, wie es funktioniert: Die Zahlen „1" und „2" werden als verschlüsselte Werte A und B an einen Dienstleister geschickt. Der Dienstleister kann A und B addieren und das Ergebnis C an den Auftraggeber zurückschicken. Aber der Dienstleister kennt weder A noch B noch C. Der Auftraggeber hingegen kann C wieder entschlüsseln und sieht somit das Ergebnis, in diesem Fall die „3". Auf ähnliche Weise lassen sich hochkomplexe Berechnungen durchführen.

Doch selbst damit ist es nicht getan. Genomdaten bestehen aus immens großen Datensätzen. Forscher oder Medikamenten-Hersteller konzentrieren sich daher nur auf die SNPs oder Mutationen der DNA, die für ihre aktuelle Fragestellung relevant sind. Das führt dazu, dass der IT-Dienstleister theoretisch aus dem Zugriff auf die Sequenz schließen kann, woran die Forscher gerade arbeiten – auch wenn das Ergebnis der Abfrage verschlüsselt bleibt.

„Der DNA-String, den ich untersuche, gibt viel preis darüber, mit welchen Krankheiten und Wirkstoffen ich mich beschäftige", sagt Katzenbeisser. „Um dies zu verhindern, führen wir zusätzlich eine Art Täuschungsmanöver ein, das sogenannte Oblivious RAM. Dabei wird der physische Speicher bei der Datenbankabfrageständig durcheinander gemischt. Niemand kann dann mehr nachvollziehen, ob der Fragesteller

Informationen

Profilbereich Cybersecurity
Prof. Dr. Stefan Katzenbeisser
E-Mail: katzenbeisser@seceng.informatik.tu-darmstadt.de
Prof. Dr. Kay Hamacher
E-Mail: hamacher@bio.tu-darmstadt.de
www.cysec.tu-darmstadt.de

bot

mehrmals auf die gleichen Daten oder auf viele unterschiedliche Daten zugegriffen hat. Die Intention der Abfrage ist also verschleiert."

Die Forschungen sind Teil des von der Deutschen Forschungsgemeinschaft finanzierten Sonderforschungsbereichs „CROSSING – Kryptographiebasierte Sicherheitslösungen als Grundlage für Vertrauen in heutigen und zukünftigen IT-Systemen" und des Center for Research in Security and Privacy (CRISP) mit Sitz in Darmstadt. Die Teams von Hamacher und Katzenbeisser möchten zunächst die Basistechniken für die kryptografischen Verfahren entwerfen. Da sie sehr komplex sind, entwickeln die Forscher zudem Tools, mit denen sich die Verfahren fehlerfrei umsetzen lassen. So können IT-Mitarbeiter auch ohne Expertise in Kryptographie die notwendigen Protokolle implementieren.

Die Verfahren müssen zudem noch leistungsstärker werden. Die meisten Techniken sind bislang vor allem für kleinere Datensätze gedacht. Aber bei Genomdaten wären sie überfordert. Das Genom eines Menschen hat einen Informationsgehalt von 100 Megabyte bis 200 Gigabyte – je nachdem, ob man die vollständige DNA oder nur die Mutationen speichert. Hinzu kommt, dass für aussagekräftige Untersuchungen die Genomdaten von möglichst vielen Menschen berücksichtigt werden sollten. So kommt einiges zusammen.

„Unser Wunsch ist auch, kleine Kliniken und Einrichtungen anzusprechen, in denen Genomdaten vorliegen", sagt Katzenbeisser. „Die sind eher skeptisch, da sie nur Daten von wenigen Patienten erheben. Man könnte leicht Rückschlüsse auf ihre Identität ziehen. Die Verschlüsselungsverfahren sollen den Kliniken die Sicherheit geben, solche Daten zu dieser wichtigen Forschung beizusteuern."

Ein stärkeres Vertrauen in die Infrastruktur der Genomforschung ist dringend notwendig. Deutschland gerät sonst ins Hintertreffen. Amerikanische Firmen investieren bereits viel Geld in diesen Bereichen. Was sie im einzelnen mit den Daten vorhaben, ist allerdings ungewiss. Kay Hamacher sagt: „Wer einmal einwilligt, Genomdaten zu hinterlegen, kann sie nicht so einfach wieder aus dem Verkehr ziehen. Außerdem entscheidet man beim Erbgut für Mütter, Väter und Enkel gleich mit. Nicht zuletzt müssen wir bedenken, dass wir bei der Genomforschung erst am Anfang stehen – wir wissen nicht, was sich in Zukunft noch alles herauslesen lässt. Umso wichtiger ist es, dass wir die Sicherheit der Daten frühzeitig in den Griff bekommen."

Der Autor ist Technikjournalist.

Abbildung: Katrin Binner

Professor Stefan Katzenbeisser entwickelt Verschlüsselungsmethoden für Genomdaten.

Namen und Fakten

Prof. Stefan Katzenbeisser betreut das Projekt S5 Privacy-Preserving Computation des Sonderforschungsbereichs CROSSING und ist Leiter des Fachgebiets Security Engineering an der TU Darmstadt.
Prof. Kay Hamacher betreut in dem von Bund und Land geförderten und von der TU Darmstadt maßgeblich mitgetragenen Center for Research in Security and Privacy (CRISP) das Projekt Skalierbare Privatsphäre-schützende Protokolle. Ferner leitet er die Arbeitsgruppe Computational Biology & Simulation an der TU Darmstadt.