REKENEN MET VERCIJFERDE DATA

Dr.ir. Thijs Veugen is sinds 1999 werkzaam als senior scientist bij TNO in de afdeling Information Security. Tevens werkt hij vanaf 2008 aan de TU Delft als senior researcher in de Multimedia Signal Processing groep. Zijn belangrijkste onderzoeksgebied is toepassingen van cryptografie.

Een nieuwe techniek om bestaande en nieuwe toepassingen 'privacy vriendelijk' te maken is gebaseerd op het concept 'rekenen met vercijferde data'. In dit artikel wordt dit mysterieuze idee uit de doeken gedaan en wordt de kracht ervan duidelijk gemaakt.

Het privacy-probleem ontstaat doordat sommige toepassingen gevoelige persoonlijke informatie nodig hebben. Om bijvoorbeeld op internet advies te vragen over de behandeling van een bepaalde ziekte zal je eerst moeten vertellen welke ziekte je precies hebt. En om bij bol.com advies te kunnen krijgen over boeken die je mogelijk interesseren, zal bol.com eerst moeten weten wat voor soort boeken jij zoal prettig vindt. Dat kan bol.com doen door een overzicht bij te houden van de boeken die je in het verleden hebt gekocht, of door bepaalde persoonlijke voorkeuren op te vragen die bol.com vervolgens vergelijkt met voorkeuren van andere gebruikers. Als iemand die qua voorkeuren op jou lijkt een bepaald boek leuk vindt, is de kans groot dat jij dat boek ook leuk vindt.
Zou het niet prettig zijn wanneer al die toepassingen je *wel* van nuttige informatie kunnen voorzien maar *zonder* dat ze al jouw persoonlijke informatie in handen krijgen? Voor de klanten is het een hele geruststelling om te weten dat hun privégegevens nooit in vreemde handen zullen vallen, en de provider hoeft zich ook niet meer druk te maken of hij wel zorgvuldig genoeg is omgegaan met de persoonlijke gegevens van zijn klanten.

Het is bekend dat je data kunt vercijferen om te voorkomen dat anderen bij die informatie kunnen. Tot nu toe betekende dat ook dat je die vercijferde data niet kunt gebruiken om nuttige dingen uit te rekenen zoals bijvoorbeeld een aanbeveling voor leuke boeken. Maar daar is nu een einde aan gekomen! Door nieuwe technieken als homomorfe encryptie, die verderop worden uitgelegd, blijkt het mogelijk te zijn om *en* data te vercijferen *en* er toch leuke dingen mee uit te rekenen als een medisch advies speciaal voor jouw ziekte of een boek aanbevolen krijgen waar je zelf nooit op was gekomen.

De voordelen van zekerheid
Er is natuurlijk wetgeving die voorschrijft hoe er met persoonlijke gegevens dient te worden omgegaan en tot op zekere hoogte zijn je gegevens daarmee ook veilig. In de praktijk blijken

Vercijferen van data.

er echter steeds meer meldingen te komen van gevoelige gegevens die om een of andere reden toch bij onbevoegden terecht zijn gekomen. Wat gebeurt er bijvoorbeeld met je gegevens wanneer een provider als bol.com failliet gaat of wordt overgenomen door een andere partij? Of wanneer iemand weet in te breken en de gegevens steelt, of een medewerker iets te onzorgvuldig met zijn laptop is omgegaan?

Om je met zekerheid te beschermen tegen alle mogelijke manieren waarop gevoelige data kan uitlekken is encryptie een goede oplossing. Wat er ook met de data gebeurt, je weet zeker dat jouw gevoelige data nooit op straat zal komen. Door die zekerheid ontstaan er ook nieuwe mogelijkheden. Waar het gaat om zeer gevoelige gegevens van bijvoorbeeld medische of concurrentiegevoelige aard, zijn gebruikers en organisaties zeer huiverig om deze voor bepaalde toepassingen beschikbaar te stellen. Wanneer ze echter de garantie krijgen dat er niets onrechtmatigs met die data kan gebeuren, zullen ze eerder geneigd zijn om deze te overhandigen zodat er tal van nieuwe toepassingen kunnen ontstaan. Denk bijvoorbeeld aan het kunnen uitbesteden van financiële bewerkingen van bedrijfsdata, het genereren van aanbevelingen gebruikmakende van profielen van verschillende organisaties, of het gebruiken van data van contacten in een sociaal netwerk om betere aanbevelingen te kunnen krijgen [CASoN2011].

Zodra je kunt rekenen met vercijferde data opent zich een wereld van toepassingen. Een bekende is om te kunnen zoeken in vercijferde data. Zo kun je toch selectief data opvragen die vercijferd bij een derde (of in de cloud) is opgeslagen, zonder alle data te hoeven uploaden, maar ook zonder dat de server weet om welke data het gaat en waarnaar je op zoek bent. Een ander, actueel voorbeeld is het gebruik van cookies in de internetwereld. Virtuele diensten slaan allerlei persoonlijke informatie op in cookies op de computers van gebruikers om hun dienst beter te kunnen aanbieden. Dit privacy-risico zou ondervangen kunnen worden door de cookies te vercijferen zonder dat het ten koste gaat van de dienstverlening.

Een nieuwe veelbelovende techniek

Het veilig kunnen rekenen met een aantal partijen samen is in de academische wereld al langer bekend als secure multiparty computation. Er zijn allerlei cryptografische protocollen bekend voor specifieke problemen, elk met hun eigen voor- en nadelen. Over het algemeen vragen dergelijke oplossingen echter een grote hoeveelheid rekenkracht en communicatie waardoor ze nog nauwelijks in de praktijk worden toegepast. Daar lijkt nu een einde aan te komen door de opkomst van *homomorfe encryptie*. Dat zijn encryptiesystemen die het mogelijk maken om bepaalde berekeningen te kunnen doen met vercijferde data zonder de tussenkomst van protocollen. Ook de Amerikaanse Defensie ziet het belang van deze techniek en heeft vijf miljoen dollar beschikbaar gesteld voor onderzoek [TheHN2011].
Onder andere via het COMMIT [COMMIT2011] programma, dat dit jaar is gestart, steekt de Nederlandse overheid veel geld in ICT-onderzoek. In een van de COMMIT-projecten, namelijk Trusted Healthcare Services (P15), worden expliciet technieken ontwikkeld om homomorfe encryptie praktisch toepasbaar te maken. In dit geval voor het gezondheidszorgdomein.
Dat betekent dat de komende vier jaar in Nederland hard kan worden gewerkt aan de doorontwikkeling van deze privacy-beschermende technieken.

Toepassing

Zoals gezegd is homomorfe encryptie een nieuwe techniek die het mogelijk maakt om te rekenen met vercijferde data. Het gaat te ver om hier uit te leggen hoe die asymmetrische encryptie precies werkt, maar stel bijvoorbeeld dat [x] de vercijfering voorstelt van bericht x, en [y] de vercijfering van bericht y. De belangrijkste eigenschap van homomorfe encryptie is dan dat je bijvoorbeeld de berichten x en y bij elkaar kunt optellen zonder ze te hoeven ontcijferen:

$$[x] * [y] = [x + y]$$

Door de cijferteksten [x] en [y] te vermenigvuldigen, krijg je de vercijfering van x + y. Encryptiesystemen met deze eigenschap worden *additief* homomorfe systemen genoemd. Op dezelfde manier zijn er ook *multiplicatief* homomorfe systemen waarmee je berichten met elkaar kunt vermenigvuldigen zonder ze te hoeven ontcijferen:

$$[x] * [y] = [x * y]$$

Het nadeel is dat een homomorf encryptiesysteem nooit beide eigenschappen tegelijk heeft. Het is ofwel additief homomorf, ofwel multiplicatief homomorf. Het kunnen rekenen met vercijferde data is dan beperkt tot optellen dan wel vermenigvuldigen. Om andersoortige rekenkundige bewerkingen te kunnen doen heb je weer cryptografische protocollen nodig en dat betekent interactie met de partij die in staat is om te ontcijferen.
Het ideale homomorfe encryptiesysteem kan berichten zowel optellen als vermenigvuldigen, en je kunt laten zien dat je daarmee alle mogelijke bewerkingen kunt uitvoeren die je maar wilt. Er wordt in de academische wereld momenteel gewerkt aan dergelijke systemen, die *volledig* homomorf worden genoemd, maar de eerste oplossingen vragen nog teveel rekenkracht.
Daarmee wordt dus ook duidelijk waar momenteel nog de schoen wringt. Homomorfe encryptie is namelijk een veelbelovende techniek om te kunnen rekenen met vercijferde data, maar in de praktijk ontkom je er nog niet aan om op een bepaald moment toch interactie te hebben met de partij die de decryptiesleutel heeft. Afhankelijk van de toepassing is dat de eigenaar van de data. Maar om gebruikers zoveel mogelijk te ontlasten kan het ook een tweede serviceprovider zijn die samen

met de eerste serviceprovider alle berekeningen uitvoert.
Een dergelijk model is te zien in fig. 1. Ter initiatie van de service sturen users hun vercijferde persoonlijke data naar een van de serviceproviders. Data richting serviceprovider 1 wordt vercijferd met de publieke sleutel van provider 2 en omgekeerd. De service providers, twee onafhankelijke organisaties, gaan via een protocol samen met die data rekenen zonder dat een van beide providers de data daadwerkelijk leert. Alle rekenkracht en communicatie concentreert zich tussen de twee providers. De users worden ontlast. Uiteindelijk wordt de output ter beschikking gesteld aan een van de partijen. In fig. 1 zijn de twee providers vergelijkbare rollen toegedicht. Het is echter ook mogelijk dat users alleen data naar serviceprovider 1 uploaden, en dan krijgt serviceprovider 2 meer een rol van onafhankelijke privacy serviceprovider die serviceprovider 1 helpt om op een privacy-vriendelijke manier diensten aan te bieden.

Voorbeelden van toepassingen met twee serviceproviders zijn:

- de Nederlandse overheid en de Duitse overheid die onderlinge databases van gezochte personen met een criminele achtergrond willen combineren om te kijken of er overeenkomsten tussen zitten;
- twee concurrerende aanbieders van diensten die userdata van de ander willen gebruiken om betere aanbevelingen te kunnen genereren;
- de overheid die gevoelige data van burgers ter beschikking stelt aan commerciële bedrijven die daar op een privacy-gevoelige manier waarde aan kunnen toevoegen;
- een provider van een sociaal netwerk gericht op bepaalde patiënten, die bij een content provider kan zoeken naar media die relevant kan zijn voor de patiënten.

Zoals gezegd heeft het ontbreken van volledig homomorfe encryptie tot gevolg dat af en toe intensieve cryptografische protocollen nodig zijn tussen twee onafhankelijke partijen. Dat kunnen de serviceprovider en de gebruiker zijn, maar wanneer de gebruiker intensieve berekeningen moet verrichten, gaat dat ten koste van de toepasbaarheid. Bovendien dragen ingewikkelde cryptografische protocollen die alleen door experts zijn te doorgronden doorgaans niet bij aan de transparantie richting gebruiker. Het model met twee serviceproviders biedt een mogelijkheid om die extra belasting niet bij de gebruiker neer te hoeven leggen. Wanneer de twee serviceproviders zouden samenwerken is het mogelijk om de gevoelige gebruikersdata te achterhalen, dus toepassing van dat model is voorlopig beperkt tot situaties waarbij de twee providers van nature niet geneigd zijn om informatie uit te wisselen maar toch het voordeel zien van samenwerking. Meer onderzoek is nodig om 'rekenen met vercijferde data' breder toepasbaar te maken en de benodigde hoeveelheid berekeningen en communicatie tot een minimum terug te brengen.

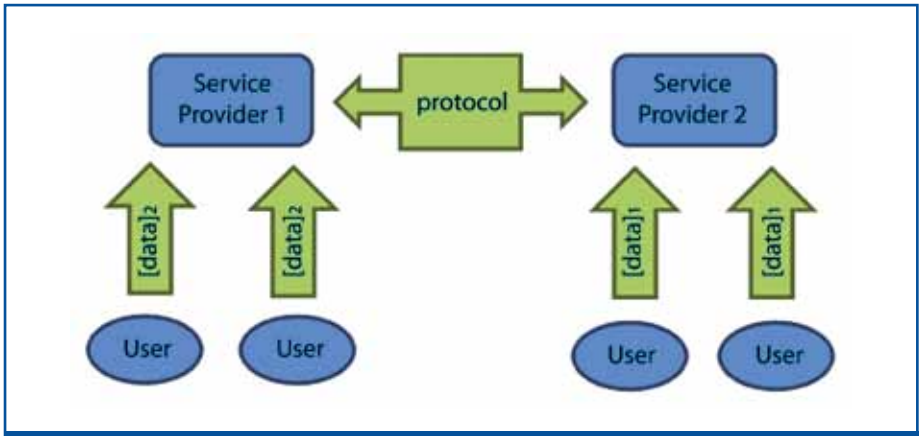

Fig. 1. Toepassing met twee service providers.

Conclusies

Door het toenemend aantal incidenten met gevoelige persoonlijke data groeit de behoefte aan privacybescherming. Er bestaat wetgeving die bedrijven dwingt om zorgvuldig met persoonlijke data om te gaan, maar bepaalde risico's lijken onvermijdelijk. Om zeker te zijn dat gevoelige data nooit uit zal lekken, zonder de kwaliteit van de dienstverlening aan te tasten, is rekenen met vercijferde data een veelbelovende techniek. Nieuwe toepassingen met zeer gevoelige data behoren dan tot de mogelijkheden.
Rekenen met vercijferde data is gebaseerd op een techniek die we homomorfe encryptie noemen. Deze veelbelovende techniek is ontstaan in de academische wereld en staat op het punt om zijn intrede te gaan maken in de hedendaagse praktijk. Bepaalde toepassingen met twee serviceproviders behoren al tot de mogelijkheden, maar omdat nog niet alle berekeningen met vercijferde data efficiënt kunnen worden gedaan, is er onderzoek nodig om de laatste stap richting toepassing te realiseren. De Nederlandse overheid erkent dit en investeert in de techniek door middel van het COMMIT-programma. Daar zal de toepassing worden toegesneden op de medische gezondheidszorg en zullen oplossingen worden bedacht om de transparantie richting gebruikers te verbeteren.

Referenties

CASoN, The International Conference on Computational Aspects of Social Networks, IEEE, *'Generating Private Recommendations in a Social Trust Network'*, Zekeriya Erkin, Thijs Veugen en Inald Lagendijk, isplab.tudelft.nl/content/generating-private-recommendations-social-trust-network, 2011.
COMMIT, www.commit-nl.nl/index.htm, 2011.
The Hosting News.com, DARPA invests 5 million towards solving homomorphic encrytion, www.thehostingnews.com/darpa-invests-5-million-towards-solving-homomorphic-encryption-17158.html, april 2011.