Data gebruiken zonder ze te krijgen of te zien

Hoe we zonder privacyschending informatie verkrijgen in de zoektocht naar onvindbare veroordeelden

Voorwoord
Delen zonder delen?

In deze publicatie willen wij “delen” wat we gevonden hebben over informatiedeling, namelijk nieuwe mogelijkheden en uitdagingen. Het gaat over het uitwisselen van gevoelige gegevens in het veiligheidsdomein. Er bestaan van oudsher wetten over het doorgeven van kopieën van documenten. Digitalisering maakt het delen van data heel veel gemakkelijker; dat brengt echter andere uitdagingen met zich mee.

We leven in een land waar zelfs de overheid niet zomaar alles van iedereen weet

Omwille van de bescherming van de persoonlijke levenssfeer leven we gelukkig in een land waar zelfs de overheid niet zomaar alles van iedereen weet. Er zijn verschillende toepassingen, waarbij het delen van gegevens tussen organisaties tot betere inzichten en daardoor beter handelen van de overheid kan leiden. Het ministerie van Justitie en Veiligheid werkt aan diverse maatschappelijke vraagstukken waar het delen van informatie over betrokkenen van belang is. Het actuele voorbeeld is de achterstand in het laten uitzitten van opgelegde gevangenisstraffen: het kost veel moeite om alle mogelijke informatie over veroordeelden te verkrijgen en sommige mensen zijn daardoor langer dan nodig “onvindbare veroordeelden”.

Met verschillende partijen werken we aan faciliteiten die technisch én juridisch voldoen. In deze publicatie leest u wat het ministerie van Justitie en Veiligheid in samenwerking met TNO heeft verkend inzake privacy-verantwoord delen van data voor de onvindbare veroordeelden. In hoofdstuk 2 staat de beknopte uitleg over verschillende privacy technieken, gericht op het gebruiken van gegevens zonder dat de identiteit van betrokkenen aan de verwerkers bekend wordt gemaakt. In hoofdstuk 3 is beschreven welke twee mogelijkheden het Programma Onvindbare Veroordeelden verder uitwerkt om toe te gaan passen.

Dit is een stand van zaken maar zeker ook een uitnodiging voor iedereen die aanvullende stukjes van de oplossing kan aandragen! Vragen en opmerkingen zijn van harte welkom. Laten we de beschikbare kennis over privacy-vriendelijke gegevensuitwisseling zo veel mogelijk delen!

1. Inleiding
De uitdaging

Belang van persoonsgegevens

Het delen van data is essentieel voor het veiligheidsdomein. Het gaat er om dat mensen in hun functie voldoende geïnformeerd worden om dat werk goed te kunnen doen. De strafrechtketen draait om mensen, niet om containers of centen, boodschappen of adviesrapporten. Over die mensen valt veel te weten en te delen, zoals: persoonsidentiteit, strafblad, gegevens van mensen in bewaring en relevante medische gegevens.

Professionals in het veiligheidsdomein werken steeds vaker in discipline-overstijgende samenwerkingsverbanden. Denk aan jeugdbeschermingstafels, veiligheidshuizen, overlastbestrijding in de wijk, ZSM en Veilig Thuis. Bij dossiers zoals jeugdcriminaliteit, verwarde personen en financiële criminaliteit is het delen van informatie vanzelfsprekend nuttig. Alleen zo komt men tot integrale risico-inschattingen en effectiever optreden.

Het delen van informatie met anderen is niet vanzelfsprekend

Partijen zoals politie, departementen, gemeenten en uitvoerings-instanties zien dat zij *samen* beschikken over heel veel relevante data om dit mogelijk te maken. Een *gezamenlijk* beeld leidt naar verwachting tot de beste onderzoeken, maatregelen en interventies. Ten behoeve van de betrokkenen en ten behoeve van de maatschappij in het algemeen.

Het delen van informatie met anderen is echter niet vanzelfsprekend. De ICT-systemen van de verschillende kolommen zijn meestal niet of nauwelijks ontworpen voor informatiedeling met andere organisaties. Veel organisaties en professionals willen, mogen of kunnen hun informatie bovendien niet zomaar delen. Professionals kampen met onzekerheden over wanneer en hoe men informatie mag of moet delen. Het gaat ook om ethische afwegingen en om mogelijkheden van informatiedeling.

De uitdaging van het Programma Onvindbare Veroordeelden

Bij het ministerie van Justitie en Veiligheid onderkennen we de uitdaging onvindbare veroordeelden. Het probleem waar we voor staan is dat niet alle veroordeelden hun straf uitzitten. Na al het werk, van aangeven, opsporen, verklaren, vervolgen en berechten, blijkt een deel van de veroordeelden onvindbaar te zijn en kan de straf zodoende niet geëxecuteerd worden. Om de achterstand van veroordeelden met een primaire vrijheidsstraf weg te werken, worden allerlei acties ondernomen. We werken samen met gemeenten om gezochten te signaleren, we onderzoeken openbare bronnen. En we willen gebruik maken van méér beschikbare informatie.

Programma Onvindbare veroordeelden

Doelen
- Substantieel verkleinen voorraad onvindbare veroordeelden
- Voorkomen van onvindbaarheid

Wie
Ministerie JenV, OM, politie, CJIB, Justid, gemeenten en meer

Cijfers
- 20.000 vrijheidsstraffen per jaar
- 90% binnen 2 jaar ten uitvoer gelegd
- 10% ontloopt hun straf
- Hiervan 90% in het buitenland
- Voorraad ca. 10.000 op te sporen personen

Feiten van het CBS

Het Centraal Bureau voor de Statistiek (CBS) heeft voor het Programma Onvindbare Veroordeelden de beschikbare gegevens “van Nederland” geanalyseerd om een beter beeld te geven van de specifieke populatie onvindbaren. Het vinden van de onvindbare veroordeelden kost namelijk veel moeite terwijl de ervaring leert dat de meesten niet in Nederland zijn.

Het CBS ondersteunt veel werk van het ministerie van Justitie en Veiligheid met relevante statistische gegevens. Zoals ze zelf schrijven: “Het CBS maakt het mogelijk dat maatschappelijke debatten gevoerd kunnen worden op basis van betrouwbare statistische informatie.” Het gaat steeds om statistische gegevens, die volgens wetenschappelijke conventies niet te herleiden zijn tot individuen.

Het CBS onderzocht of onvindbare veroordeelden nog in Nederland verblijven

Het Programma Onvindbare Veroordeelden heeft het CBS onder andere gevraagd te onderzoeken of de bekende onvindbare veroordeelden, volgens de beschikbare gegevens, nog in Nederland verblijven. Met dergelijk inzicht kunnen betere keuzes worden gemaakt voor de opsporingsinspanning.

Het CBS heeft onderzoek gedaan naar een groep van 32.400 personen met een BSN, die op een bepaalde datum in het opsporingsregister stonden vanwege een openstaande opgelegde vrijheidsstraf of andere strafrechtelijke sanctie. Gebleken is dat na enige tijd nog slechts over 15% van deze groep, gegevens in CBS-bronnen te vinden zijn die informatie verschaffen over een woonlocatie in Nederland. Van de rest is het aannemelijk dat ze niet meer in Nederland zijn.

Zie voor een nadere toelichting en de gedetailleerde bevindingen de (vier) deelonderzoeken “Spoorloos Veroordeelden” van het CBS, d.d. 1 juli 2019.

Inzet privacy technieken

Wetende dat slechts een minderheid van de onvindbare veroordeelden in Nederland kan worden opgespoord, is het interessant per persoon in te kunnen schatten of het de moeite van het zoeken waard is. De ervaringen én specifieke data analyse hebben ons geleerd dat iemand die in Nederland verblijft, gevonden wordt. Maar als iemand niet in Nederland verblijft, is het een verspilling van tijd om gegevens te vorderen en agenten naar hem te laten uitkijken.

Daarom heeft het Programma Onvindbare Veroordeelden nader onderzoek verricht naar mogelijk in te zetten technieken op het gebied van ‘privacy-by-design’. Aan TNO is gevraagd op welke wijze de gegevens van Nederlandse instanties gebruikt kunnen worden zonder privacyregels te overtreden. TNO heeft hiervoor een verkenning gedaan van de inzet van privacy-technieken. Deze technieken zijn gericht op het kunnen analyseren van gegevens, terwijl de identiteit van de betrokkenen niet bekend wordt aan verwerkers van die gegevens. Vervolgens is een proof-of-concept uitgewerkt met “Multi Party Computation” technieken. Ook hebben we de bestaande toepassing van de techniek “Bloomfilter”, een decentrale matchingstechnologie genaamd Ma³tch, nader bekeken op mogelijke inzet voor deze uitdaging.

2. Privacy technologies De theorie

Wat verstaan we onder privacy technologies?

Privacy preserving technologies (PPT) zijn technieken om privacygevoelige gegevens (zoals persoonsgegevens) te kunnen verwerken op een privacy-vriendelijke manier. Er wordt ook wel gesproken van privacy enhancing technologies (PET). Deze omvatten een verzameling technieken die de bescherming van de persoonlijke levenssfeer van individuen binnen een informatie-systeem versterken door het voorkomen van onnodige dan wel ongewenste verwerking van persoonsgegevens. In feite gaat het om dezelfde onderliggende technologieën die helpen bij de verplichting in de privacywetgeving om "privacy-by-design" toe te passen. Met techniek kunnen we op maat persoonsgegevens beschermen.

Wij zijn geïnteresseerd in technologieën die het voor meerdere partijen mogelijk maken om relevante informatie uit hun privacygevoelige data te halen c.q. met andere partijen te delen, zonder de privacy van individuen te schenden. Recente technische ontwikkelingen maken het mogelijk om informatie te verwerken zonder dat deze informatie in voor mensen leesbare vorm bij de verwerker aanwezig is. De bedoeling is om met de inzet van zo'n techniek, zonder directe inzage van de persoonsgegevens van een specifiek individu, toch iets te leren over de data (functioneel). Voorbeelden zijn: wat is de gemiddelde leeftijd, zijn er overeenkomstige individuen in twee gegevensverzamelingen, en hoeveel contracten zijn er voor een bepaalde subset. We richten ons daarom op technieken die analyse mogelijk maken over meerdere gegevensverzamelingen heen.

Wensen daarover in het veiligheidsdomein bestaan bijvoorbeeld bij de aanpak van ondermijning of regionale misdaadthema's. Er zijn veel gegevens voorhanden die zouden kunnen bijdragen aan inzicht om criminelen aan te kunnen pakken maar hoe verzamelen we die zonder de privacy van niet-betrokken burgers te schenden?
In zorg- en veiligheidshuizen leeft de vraag hoe hun samenwerking kan worden ondersteund. Bij voorkeur gaat dat in de toekomst op efficiënte wijze met de geautomatiseerde beantwoording van de vraag of persoon a bij organisatie x bekend is.

Met dit soort vragen in gedachte hebben we een aantal technieken bekeken.

1. Anonimiseren
2. Pseudonimiseren
3. Bloom filter inzetten
4. Differential privacy door kansberekening toepassen
5. Secure set intersection gebruiken
6. Homomorfe encryptie in berekeningen toepassen
7. Secret sharing voor berekeningen toepassen

Tabel 1 Functionaliteit per soort

	Soort techniek	Functionaliteit
1	Anonimiseren	Onherleidbare gegevens in statistiek gebruiken
2	Pseudonimiseren	Koppelen
3	Bloom □lter inze□en	Matchen
4	Di□erential privacy door kansberekening toepassen	Beschermen bij herhaald gebruik
5	Secure set intersection gebruiken	Overeenkomende gegevens signaleren
6	Homomorfe encryptie in berekeningen toepassen	Berekening na versleuteling ontsleutelen
7	Secret sharing voor berekeningen toepassen	In ongevoelige stukjes delen

De beknopte uitleg per soort

1. Anonimiseren:
Anonimiseren is niets anders dan een verzameling technieken om data te ontdoen van persoonsgegevens. Dan is de data niet meer herleidbaar naar personen. Dit is vooral nuttig als de geaggregeerde, statistische informatie van belang is.

2. Pseudonimiseren:
Bij pseudonimiseren worden identificeerbare eigenschappen vervangen, zoals de achternaam van een persoon in een dataset, door andere waarden: pseudoniemen. Zonder aanvullende informatie (welk pseudoniem verwijst naar wie) is de data niet meer direct te herleiden naar personen. Dit wordt ingezet als databronnen gekoppeld worden.

3. Bloom filter inzetten:
Een bloom filter is een verkorte weergave (samenvatting) van een aantal objecten om te controleren of een object in een lijst aanwezig is. In het bloom filter zijn de afzonderlijke oorspronkelijke objecten niet herkenbaar. Bloom filters worden ingezet om matching te doen: gecheckt wordt of een persoon ook in een andere database voorkomt.

4. Differential privacy door kansberekening toepassen:
Bij differential privacy wordt een database bevraagd maar wordt een kleine verstoring, op basis van kansen, in het antwoord toegevoegd op zo'n manier dat gegevens nooit kunnen worden herleid naar personen. Differential privacy is nuttig bij statistische analyses en is veilig bij herhaald bevragen van de database. In het geheel van gegevens gaat geen informatie verloren maar de input is niet meer herleidbaar.

5. Secure set intersection gebruiken:
Met secure set intersection kunnen meerdere partijen nagaan welke objecten in hun datasets overeenkomen (een "hit" geven), zonder dat ze informatie over die objecten met elkaar delen.

6. Homomorfe encryptie in berekeningen toepassen:
Homomorfe encryptie is encryptie (versleuteling) waarmee specifieke berekeningen kunnen worden uitgevoerd op versleutelde data zonder deze eerst te hoeven ontcijferen. Dit kan gebruikt worden als men niet geïnteresseerd is in de data zelf maar de data wel in een berekening wil meenemen.

7. Secret sharing voor berekeningen toepassen:
Secret sharing is het in stukjes delen van gegevens waardoor geen gevoelige informatie wordt prijsgegeven maar wel in berekeningen kan worden gebruikt. Net als bij homomorfe encryptie is de functie dat men geïnteresseerd is in de uitkomst maar niet in de onderliggende data zelf.

Keuze voor techniek

Voor de keuze welke techniek het best kan worden ingezet, is in de eerste plaats de gewenste functionaliteit van belang: zie tabel 1. Met functionaliteit wordt bedoeld: wat kun je er mee doen, wat levert deze techniek op?

Vervolgens is interessant welke voor- en nadelen aan de verschillende technieken kleven; die zijn kort in tabel 2 weergegeven. Het gaat om de aspecten beveiliging, snelheid en volwassenheid van de toepassing. Daaronder verstaan we:

- Beveiliging: hoe goed zijn de persoonsgegevens beschermd, hoe makkelijk is de identiteit "te kraken", te achterhalen?
- Snelheid: hoe snel werkt de techniek, wat is de benodigde rekentijd en rekencapaciteit, hoe goed toepasbaar is dit voor grote gegevensverzamelingen?
- Volwassenheid: in hoeverre is de techniek in de praktijk toepasbaar?

Tabel 2 Voor- en nadelen per soort

	Soort techniek	Beveiliging	Snelheid	Volwassenheid
1	Anonimiseren	0	++	+
2	Pseudonimiseren	-	+	+
3	Bloom filter inzetten	+	++	+
4	Differential privacy door kansberekening toepassen	+	+	0
5	Secure set intersection gebruiken	++	0	0
6	Homomorfe encryptie in berekeningen toepassen	++	-	-
7	Secret sharing voor berekeningen toepassen	++	+	0

(+ betekent een positieve score op dit aspect, 0 betekent neutraal, - betekent negatief beoordeeld)

Voor een uitgebreidere beschrijving: zie bijlage 1. In het volgende hoofdstuk wordt het vraagstuk van de onvindbare veroordeelden beschreven en de toepassing van gecombineerde technieken daarvoor.

3. Het Programma Onvindbare Veroordeelden De toepassing

Wat bieden privacy technologies voor het Programma Onvindbare Veroordeelden?

Het landelijk Programma Onvindbare Veroordeelden dient om veroordeelde personen die zich onttrekken aan hun vrijheidsstraf te vinden en aan te houden voor het ondergaan van hun straf. De minister schrijft begin 2019 aan de Kamer : “In 2018 zijn 1.100 van de totale voorraad van ongeveer 11.000 dossiers van onvindbare veroordeelden op bovengenoemde manier onderzocht. Dit heeft in 10 procent van de gevallen geleid tot een aanhouding.” Dat aantal is gering en heeft ermee te maken dat veel van de veroordeelden zich in het buitenland bevinden. Soms is dan de openstaande strafmaat te laag voor uitlevering. Of de persoon bevindt zich in een land waar Nederland geen uitleveringsverdrag mee heeft.

Een deel van de veroordeelden doet waarschijnlijk gewoon bankzaken in Nederland of heeft een telefoonabonnement

Om dit percentage te verhogen zou het helpen als men al in een vroegtijdig stadium bij het dossieronderzoek een indicatie heeft of de gezochte veroordeelde zich in Nederland bevindt. Dat maakt gerichter zoeken in Nederland mogelijk en vergroot de kans op een succesvolle aanhouding.

Een deel van de onvindbare veroordeelden doet waarschijnlijk “gewoon” bankzaken in Nederland, heeft een telefoonabonnement of is zelfs bekend bij andere overheidsinstanties. Daarvoor kan het OM gegevens vorderen. Het is niet goed werkbaar om voor alle onvindbaren, per persoon met een vordering, gegevens op te vragen bij al dit soort derde partijen. Daarom zou het handig zijn om, zonder de persoonsgegevens bekend te maken, te kunnen checken óf derde partijen gegevens hebben van onvindbare veroordeelden. We zitten dus met de vraag “waar vinden we informatie of iemand in Nederland is”. Verder willen we weten “van wie is het meest te vinden”. Voor antwoorden hebben we een gekwalificeerde bevraging nodig van instanties, zonder bekendmaking van de persoonsgegevens.

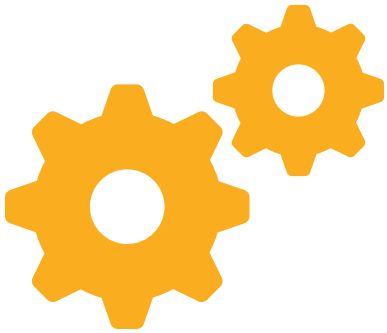
Met TNO hebben we verkend hoe we daar privacy enhancing technologies zoals Multi Party Computation (MPC)-technieken voor in kunnen zetten.

Met FCInet (Financial Criminal Investigation Network) hebben we verkend hoe dat met het PET bloom filter zou kunnen. FCInet is als organisatie ontstaan met de ontwikkeling van software om de internationale samenwerking tussen ‘financial and criminal investigation services’ (zoals FIOD) te ondersteunen. Die software, Ma³tch (spreek uit: match3), is al jaren in gebruik bij de 28 Financial Intelligence Units (FIU’s) in de Europese Unie om op een veilige en privacy-vriendelijke manier gegevens te delen.

Hierna zijn de mogelijkheden van MPC en van Ma³tch beschreven.

Multi Party Computation om instanties gekwalificeerd te bevragen

De Multi Party Computation (MPC)-technieken Secure Set Intersection en Homomorfe encryptie kunnen goed ingezet worden voor de uitdaging van de onvindbare veroordeelden. Hieronder wordt toegelicht wat deze technieken inhouden en te bieden hebben bij het verzamelen van gegevens van personen uit verschillende bronnen.



MPC maakt het mogelijk dat meerdere partijen op verzoek gegevens vergelijken en er samen mee kunnen rekenen, zonder dat ze elkaars data kunnen inzien en dus zonder dat ze vertrouwelijke persoonsgegevens bekend hoeven te maken. Zowel de te leveren gegevens, als de berekeningen, als de totaaluitkomst zijn versleuteld met encryptie, zodat ze zonder sleutel niet ingezien kunnen worden.

Het effect van MPC is vergelijkbaar met een vertrouwde tussenpartij die alle relevante inputdata verzamelt, op basis hiervan de benodigde berekeningen doet en dan de uitkomst bekend maakt. Echter, het gebruik van een echte vertrouwde tussenpartij is vaak duur, tijdrovend en door de wetgeving niet zomaar toegestaan. De inzet van MPC-techniek maakt de tussenpartij overbodig. MPC is voor allerlei privacygevoelige analyses te gebruiken. In theorie is het mogelijk om iedere berekening of analyse toe te passen op data van meerdere partijen, zonder deze data daadwerkelijk te hoeven delen.

TNO heeft een proof-of-concept (POC) uitgewerkt met inzet van MPC voor de onvindbare veroordeelden. Die ziet er samengevat als volgt uit.

Met behulp van een rekenformule wordt bepaald welke personen het meest waarschijnlijk opgespoord kunnen worden

De lijst van onvindbare veroordeelden wordt met behulp van MPC vergeleken met de relevante databases van overheids- en niet-overheidsinstellingen, op overeenkomende personen. Ook wordt er met behulp van een rekenformule bepaald van welke personen de meest relevante gegevens beschikbaar zijn, waardoor die het meest waarschijnlijk opgespoord kunnen worden.

Zoals hiervoor beschreven is, kan met de inzet van secure set intersection 'geheim' een vergelijking tussen bestanden worden gedaan, op overeenkomende elementen. Hier worden bovendien meerdere instanties tegelijk bevraagd. De instanties blijven onwetend voor welke personen de vergelijking wordt gedaan.

In de POC van TNO wordt, in geval van een "hit" (overeenkomst), door middel van een functie een zogenaamde vindbaarheidsscore berekend voor die persoon. Het idee is dat de vindbaarheidsscores aangeven van welke persoon het meest bekend is. Deze functie maakt gebruik van de techniek homomorfe encryptie. Ingrediënten voor deze rekenformule kunnen bijvoorbeeld zijn:

1) Score wordt verhoogd voor iedere overeenkomst (dus als een persoon voorkomt) in overheidsbestanden, bijvoorbeeld per database hit 1 punt erbij.
2) Score wordt verhoogd als een persoon een bankrekening heeft bij een bank (1 punt, en 4 extra punten als de bankrekening recent (afgelopen 2 weken) nog is gebruikt).
3) Score wordt verhoogd als een persoon een telefoonabonnement heeft bij een telecomprovider (1 punt, en 4 extra punten als de telefoon recent (afgelopen 2 weken) nog is gebruikt).

Dit levert per onvindbare veroordeelde dus een score op van 0 tot misschien wel 30 punten.

De gebruikte lijst wordt daarna gesorteerd op puntenaantal en getoond aan de betrokken ambtenaren. Hierbij zijn verschillende varianten van 'ontsleuteling van de versleutelde gegevens' mogelijk. Hieronder beschrijven we er twee:

- De complete lijst wordt getoond inclusief het totaal aantal punten, hit/no-hit op bestanden, de namen van bank/telecomprovider en het wel/niet recent gebruik van bankrekeningen/telefoons. Bij deze variant kan men zien bij welke instanties punten zijn gescoord maar nog niet wat de onderliggende gegevens zijn, zoals het bankrekeningnummer of wanneer er gebeld is.
- Een tweede variant zou met nog meer afscherming van gegevens kunnen werken, door op de lijst slechts de totaalscore per onvindbare veroordeelde weer te geven. Daarmee is er informatie of een reguliere vordering van gegevens kansrijk is; dat draagt bij doordat er voorrang gegeven kan worden aan die vorderingen voor personen waar nu bekend van is dát er gegevens beschikbaar zijn.

Afhankelijk van de mate van implementatie en gewenste rol van het Openbaar Ministerie kunnen varianten, voor het weergeven van met MPC gevonden scores, worden ingezet.

Bij deze aanpak is het sleutelbeheer van groot belang. Uitgegaan wordt van een vertrouwde partij in het justitiedomein, die uitsluitend in opdracht van het OM verzamelde gegevens ontsleutelt.

Evaluatie MPC POC

Functioneel gezien voldoet de POC in ieder geval voor de eerste stap (is iemand bekend bij bepaalde instanties?). De geheime check geeft daar inzicht in. Als is gebleken dat een derde partij over gegevens van de onvindbare veroordeelde beschikt, kan justitie deze gericht opvragen. Tot die tijd weet de derde partij niet op wie de check betrekking had. Als een derde partij geen gegevens heeft van onvindbaren, wordt er niets opgevraagd en blijft de check anoniem.

MPC ondersteunt het vinden van onvindbare veroordeelden met antwoorden óf er een indicatie is dat iemand in Nederland is en wáár relevante gegevens te vorderen zijn. Dit voldoet aan privacy-by-design: de informatiepositie wordt verbeterd zonder dat dit ten koste gaat van het privacybelang.

Doordat bij gebruik van technieken als MPC de data, berekeningen en uitkomsten onleesbaar zijn, zou men kunnen stellen dat deze onleesbare gegevens geen persoonsgegevens zijn en dus buiten de privacywetgeving vallen. Zie ook G. Spindler en P. Schmechel, „Personal Data and Encryption in the European General Data Protection Regulation," *JIPITEC*, vol. 163, nr. 7, 2016. Hier is echter nog geen definitieve (juridische) uitspraak over, en het zal afhangen van de specifieke gebruikte techniek, in verband met de benodigde bescherming van middelen om te ontcijferen.

Na afronding van de verborgen analyse moet men de uitkomst en daarmee het gewenste inzicht bekendmaken, bijvoorbeeld door de uitkomst te ontcijferen. De deelnemende partijen bepalen wie de uitkomst van de berekening mag inzien.

Door, na herkenning, een score weer te geven in plaats van de detailgegevens, kan in het proces een stap of meer worden ingebouwd om zo privacy-vriendelijk mogelijk te handelen. De gegevens over welke bank of bankrekeningnummer het gaat, en de achterliggende detailgegevens van gebruik, kunnen zo voor iedereen worden afgeschermd, behalve voor de betrokken Officier van Justitie en het opsporingsteam.

De MPC-technieken bieden geen garanties voor de (on-)gevoeligheid van de ontcijferde uitkomsten. Het is dus mogelijk dat een uitkomst nog steeds herleidbaar is naar een individu. Daarom combineert men MPC soms ook met andere PET, zoals differential privacy. Voor deze POC voor de onvindbare veroordeelden achten we de kans op herkenbaarheid aan de hand van veel voorkomende zaken als bankrekeningnummers zodanig klein dat hier geen rekening mee is gehouden.

Cryptografie zorgt ervoor dat het doen van berekeningen door middel van MPC een stuk trager is dan in het niet-versleutelde domein. MPC is daarom niet voor iedere toepassing even geschikt. In de afgelopen jaren heeft MPC wel grote stappen gezet naar betere technieken en snellere protocollen. In een proefimplementatie zal nader onderzocht moeten worden wat de mogelijkheden zijn, afhankelijk van de omvang van de lijsten en het aantal betrokken instanties. Bij het ontwerpen van een MPC-oplossing moet rekening worden gehouden met dergelijke randvoorwaarden met betrekking tot schaalbaarheid en veiligheid.

Ma³tch toepassing van een bloom filter

Een bloom filter is een verkorte weergave (samenvatting) van een aantal objecten om te controleren of een object in een lijst aanwezig is. In het bloom filter zijn de afzonderlijke oorspronkelijke objecten niet herkenbaar. De elementen worden als het ware versleuteld en samengevat waardoor ze niet leesbaar zijn maar wel bij bevraging herkend worden.

In het bloom filter zijn de afzonderlijke oorspronkelijke objecten niet herkenbaar

Zo'n lijst van objecten kan bijvoorbeeld een database van gezochte personen zijn. Bloom filters zijn erg efficiënt. Een lijst die hiermee is samengevat neemt namelijk weinig geheugen in beslag en het controleren of een element aanwezig is vraagt niet veel rekenkracht. Vanwege deze efficiëntie is er echter een nadeel: *false positive* antwoorden zijn niet uitgesloten. Dit betekent dat een filter in sommige gevallen zal aangeven dat een object aanwezig is in de lijst (een 'hit'), terwijl dit in werkelijkheid niet zo is. Het is mogelijk om de kans op *false positives* te verkleinen; het benodigde filter wordt dan groter en dat kan in heel omvangrijke toepassingen minder goed werkbaar zijn.

Een specifieke praktijk toepassing van Bloom filters is Ma³tch. De Ma³tch-technologie is ontwikkeld binnen het ministerie van Justitie en Veiligheid voor de FIU's in de Europese Unie en wordt momenteel doorontwikkeld in FCInet: een generiek gedecentraliseerd technologie-platform waarmee nationale en internationale overheidsorganisaties veilig data kunnen koppelen, standaardiseren, analyseren en uitwisselen, waarbij data decentraal en onder volledige controle van de eigenaar blijft.

Organisaties checken hiermee of een bepaald individu in hun eigen database voorkomt en in die van een andere organisatie. Het kan financiële opsporingsdiensten helpen om te bepalen of ook andere opsporingsdiensten specifieke verdachte individuen onderzoeken; bij een match kan men een officieel verzoek doen voor meer informatie bij de andere dienst.

Globaal werkt Ma³tch als volgt. Organisaties besluiten samen te werken door elkaar strikt vertrouwelijk te informeren indien er gegevens van bepaalde personen beschikbaar zijn. Het proces kent dan de stappen:

1. Een deelnemende organisatie selecteert een aantal *velden* in haar database, bijvoorbeeld voornaam, achternaam en geboortenaam van alle verdachte individuen.
2. De Ma³tch-technologie hasht en aggregeert deze gegevens in een *filter*, een binaire vector, zodat er in principe geen persoonsgegevens herkenbaar zijn.
3. Men stuurt het filter op naar de andere deelnemende organisaties.
4. Een onderzoekende organisatie kan zelf bij (een selectie van) haar eigen *records* de bijbehorende filters bepalen via de Ma³tch-technologie.
5. Er volgt een vergelijking van deze filters met het eigen filter. Met de toegepaste techniek kunnen in een fractie van een seconde duizenden records worden gecheckt.
6. Bij een match kan de onderzoekende organisatie besluiten om meer informatie te vragen aan de organisatie met wiens filter een match is gevonden.

Het filter is zo in te richten dat bijvoorbeeld spelfouten of andere schrijfwijzen van dezelfde naam op hetzelfde filter te zien zijn. Een match wordt alleen bekend bij de onderzoeker, niet bij de andere deelnemende organisaties; er is ook niet bekend wie in de filters van andere organisaties zijn weergegeven.

Zie voor nadere beschrijving en een beoordeling van de bescherming van persoonsgegevens:

- U. Kroon, „Ma³tch: Privacy AND Knowledge - Dynamic Networked Collective Intelligence," in *Big Data, 2013 IEEE International Conference on Big Data, 2013*.
- P. Balboni, Macenaite M., „Privacy by design and anonymisation techniques in action: case study of Ma³tch technology," *Computer law & security review*, vol. 29, nr. 4, pp. 330-340, 2013.
- W. Geelhoed, R. A. Hoving, K. Lindenberg en A. Renshof, „FCInet: Legal Context and Data Protection," University of Groningen, 2018.

Evaluatie Ma³tch

Ma³tch heeft een aantal eigenschappen die informatie beschermen. Een deelnemende organisatie leert niets over individuen van andere organisaties. Men stuurt de lijst (filter) en daar blijft het bij, behalve wanneer een partij besluit contact op te nemen over specifieke individuen; dan is duidelijk dat daar iets mee aan de hand is. Ook leert de onderzoekende partij niets over individuen in de database van de zender die *niet* in zijn eigen filter zijn verwerkt. Een organisatie kan ervoor kiezen om slechts een selectie van de eigen database te verwerken in een filter en deze te delen met anderen.

Ma³tch voldoet functioneel voor de informatievraag. Een partij leert, met grote kans, welke individuen in de eigen database matchen met filters van andere organisaties, en daarmee het al dan niet voorkomen in die databases. Er is geen 100% zekerheid wegens een kans op false positives; dit vraagt dat een hit wordt opgevolgd door detailbevragingen via overeengekomen procedures.

De inzet van deze techniek kent wel een risico. De vrager kan, als deze kwaadwillend is, ook besluiten om persoonsgegevens van individuen door het Ma³tch-algoritme te halen die niet per se in de eigen database voorkomen. Dit zou kunnen worden gebruikt om te weten te komen welke andere individuen mogelijk voorkomen in de database van de zender. Hoe dit precies kan hangt af van de ingestelde entropie en precisie. Het met brute computerkracht natrekken van 'alle' personen zou leiden tot een groot aantal valse matches en is daarom niet zo zinvol, maar specifieke relevante namen checken is wel mogelijk.

Deze eigenschappen van de Ma³tch-methode (de niet volledige zekerheid en de niet volledige bescherming) kunnen te gevoelig zijn voor sommige toepassingen. Het is dan mogelijk om gebruik te maken van (een deel van) het Ma³tch-algoritme in combinatie met Multi-Party Computation-technieken (MPC). De combinatie is dan een snelle en efficiënte manier om het filter te gebruiken voor matching. Wat ook nog kan: versleuteling van de filters van Ma³tch als bij MPC.

4. Slotwoord
Het vervolg

Met dit verhaal hopen we een interessant kijkje in de wereld van privacy-technologie te hebben gegeven. Met “kennis van deze kennis” kunnen meer mensen waarschijnlijk hun weg vinden om traditionele vragen met nieuwe middelen te beantwoorden.

Het vertrouwelijk delen van data speelt breder, bij allerlei onderwerpen van het ministerie van Justitie en Veiligheid. We proberen hiervoor kennis en ervaring te delen met geïnteresseerden. Recent is een netwerk gestart om JenV-breed onderzoeks- en experimenteerprogramma’s rond privacy technologies te ondersteunen. Langs potentieel waardevolle use cases wordt de toegevoegde waarde onderzocht en technische en juridische kennis verder opgebouwd en gedeeld. De uitdaging van de onvindbare veroordeelden is één van die use cases.

Wat weten wij samen over de persoon, waardoor we een beter beeld hebben?

Er zijn veel uitdagingen in het veiligheidsdomein waar gevoelige persoonsgegevens een grote rol spelen. Denk bijvoorbeeld aan ZSM-tafels, Zorg- en Veiligheidshuizen, personen met verward gedrag, DNA-veroordelingen, ondermijning, onvindbaar vermogen, en zo verder. De vraag luidt steeds min of meer: wat weten wij samen over die persoon waardoor we een beter beeld hebben?

En dat binnen de kaders: welke informatie mogen we over die persoon uitwisselen?

Iedere oplossing heeft juridische en technische aspecten. Alleen al daarom is samenwerking tussen deskundigen geboden. En capaciteit is schaars dus laten we de mogelijke synergie tussen elkaars oplossingen zo goed mogelijk benutten.

Verder lezen over deze ontwikkelingen kan: zie de verwijzingen en de bijlage. Zowel het ministerie van Justitie en Veiligheid als TNO gaan verder met onderzoek en kennisdeling van privacy-zaken. Daarnaast bevelen we van harte aan: de publicatie van collega Wim Borst, “De verdachte in de ketens” van Wim Borst, 2019.

Bijlage
Toelichting privacy technologies

In deze bijlage vindt u per soort privacy technology een iets uitgebreidere beschrijving. De technieken verschillen van elkaar voor wat betreft toepassingsmogelijkheden en complexiteit. Het gaat hier niet om de vergelijking tussen deze technieken; ze zijn naar voren gekomen in verband met de mogelijke inzet voor privacy-vriendelijke gegevensanalyses.

Deze tekst over privacy technologies is een weergave van de tekst van het TNO-document "Boekje Privacy TNO V0.9docx"

Zie ook de introductie over privacy enhancing technologies en privacy preserving technologies in: D. Bachlechner, M. Friedewald, J. Weitkamp en N. Martin, „e-Sides D3.1 Overview of Existing Technologies," 2018.

1. Anonimiseren:

Anonimiseren is niets anders dan een verzameling technieken om data te ontdoen van persoonsgegevens. Dan is de data niet meer herleidbaar naar personen. Er kan wel mee worden gerekend.

Enkele voorbeelden:

- *Suppressie*: verwijdering van gevoelige of identificeerbare informatie, zoals een kolom namen.
- *Generalisatie*: verruwing van gevoelige of identificeerbare informatie, bijvoorbeeld door de vervanging van geboortedata door leeftijdscategorieën of enkel de opslag van het numerieke deel van een postcode.
- *Perturbatie*: verruwing van gevoelige of identificeerbare informatie, bijvoorbeeld door het toevoegen van kleine hoeveelheden ruis.
- *Permutatie*: husselen van gevoelige informatie, waardoor de koppeling tussen een individu en gevoelige informatie ook op toeval gebaseerd kan zijn.

2. Pseudonimiseren:

Pseudonimiseren omvat het vervangen van identificeerbare attributen in een dataset door andere attributen (pseudoniemen). Zonder aanvullende informatie (welk pseudoniem verwijst naar wie) is de data niet meer te herleiden naar personen.

Voorbeelden van technieken om data te pseudonimiseren: Article 29 Working Party, „Opinion 05/2014 on Anonymisation Techniques".

- *Maskeren*: het verbergen van een belangrijk deel van de data met willekeurige karakters of andere data. Bijvoorbeeld: een creditcardnummer 5500 0000 0000 0004 kan worden opgeslagen als XXXX XXXX XXXX 0004.
- *Hashen*: de hashfunctie vertaalt een document in een reeks bytes, met als bijzondere eigenschap dat die omzetting maar in één richting kan plaatsvinden, namelijk van document naar hash en niet andersom. Hashen van privacygevoelige informatie als een naam maakt het mogelijk om data in verschillende databases aan elkaar te koppelen zonder rechtstreeks privacygevoelige data prijs te geven. Hoewel een hash er uitziet als willekeurige data is het belangrijk je te realiseren dat dit niet zo is: de hashfunctie is volledig deterministisch. Iedere keer dat we van een bepaalde tekst de hash berekenen volgt hetzelfde resultaat! Deze techniek wordt ook ingezet om te controleren of een gegeven of tekst niet is veranderd, als controlegetal.
- *Encryptie*: vercijfering en deling van gevoelige of identificeerbare informatie. De ontvangende partij heeft geen toegang tot deze informatie, maar het proces is wel omkeerbaar voor degene die de sleutel heeft om informatie te ontcijferen. Voor de AVG is het noodzakelijk dat de sleutel gescheiden blijft van de gepseudonimiseerde data.

3. Bloom filter inzetten:

Een bloom filter is een verkorte weergave (samenvatting) van een aantal objecten om te controleren of een object in een lijst aanwezig is. In het bloom filter zijn de afzonderlijke oorspronkelijke objecten niet herkenbaar. De elementen worden als het ware versleuteld en samengevat waardoor ze niet leesbaar zijn maar wel bij bevraging herkend worden.

Traditioneel gebruikt men bloom filters in databases om efficiënt te controleren of een rij of kolom aanwezig is. Die filters spelen steeds vaker een rol in een dataminimalisatie-strategie: men maakt geen gebruik van de lijst zelf om te controleren of deze een object bevat of niet, maar gebruikt hiervoor het bloom filter. In het bloom filter zitten de oorspronkelijke objecten (bijvoorbeeld de namen van gezochte personen) niet meer.

Zie ook de eerdere verwijzingen naar literatuur over het principe achter de Ma³tch-technologie, in de paragraaf Ma³tch toepassing van een bloom filter.

4. Differential privacy door kansberekening toepassen:

Differential privacy is nuttig bij statistische analyses waarbij de uitkomsten niet te herleiden zijn tot personen. Op grond van statistiek kunnen met de verstrekte gegevens wel analyses worden gemaakt over de gehele populatie. Ook kunnen individuele resultaten met toevallige variaties worden gewijzigd waardoor ze niets met zekerheid zeggen over een individu.

Zie ook, inzake Google-RAPPOR, over gebruikersstatistieken zonder individuele verwijzingen de publicatie: Ú. Erlingsson, V. Pihur en A. Korolova, „Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014.

5. Secure set intersection gebruiken:

Met secure set intersection kunnen meerdere partijen informatie delen uit hun datasets zonder gegevens over personen weer te geven.

Vertaald naar databases komt het erop neer dat partijen, elk met hun eigen database, gezamenlijk bepalen welke *records* ze gemeenschappelijk hebben. Dit kan leiden tot bijvoorbeeld de volgende resultaten:

- Het aantal gemeenschappelijke *records*.
- De inhoud van de gemeenschappelijke *records*: welke items, of welke attributen van die items.
- De uitkomst van een functie op de gemeenschappelijke *records*. Bijvoorbeeld: de gemiddelde leeftijd van de personen die in beide databases zitten, de regio waar de meeste gemeenschappelijke personen wonen, enz.

6. Homomorfe encryptie in berekeningen toepassen:

Homomorfe encryptie is encryptie waarmee specifieke berekeningen kunnen worden uitgevoerd op versleutelde data zonder deze te hoeven ontcijferen. Het voordeel daarvan is dat de privacy gewaarborgd blijft – de gegevens zelf blijven immers onbekend. De uitkomst van zo'n berekening is ook versleuteld, maar na ontcijfering krijgen we de correcte uitkomst.

Zie ook, over het aantonen van Fully Homomorfe Encryptie: C. Gentry, „Fully homomorphic encryption using ideal lattices," *Stoc.*, vol. 9, 2009.

7. Secret sharing voor berekeningen toepassen:

Secret sharing is het in stukjes delen van gegevens waardoor geen gevoelige informatie wordt prijsgegeven maar deze wel in berekeningen kan worden gebruikt.

Secret sharing is een techniek waarmee partijen enkel ongevoelige (tussen-)uitkomsten met elkaar delen, om daarmee toch gezamenlijk berekeningen te kunnen doen. Het berust op opdeling van geheime inputdata van een partij (een *secret*) in meerdere stukjes (*shares*), en wel zo dat ieder stukje geen informatie geeft over de oorspronkelijke data. Deze *shares* gaan naar de verschillende partijen, die hier dan mee kunnen rekenen.

Ministerie van Justitie en Veiligheid

TNO innovation for life

Colofon
Auteurs: Freek Bomhof en Paula Giezeman
Contactpersoon Programma Onvindbare Veroordeelden:
Aad den Boer

September 2019