Algoritmer regner på krypterede data

I takt med at samfundet bliver digitaliseret, registreres flere og flere fortrolige personoplysninger. Det udgør en trussel imod vores privatliv. Ny avanceret kryptografi kan beskytte os, uden at vi skal give afkald på fordelene ved indsamling af data.

Forestil dig en verden, hvor oplysninger om DNA bliver indsamlet rutinemæssigt i forbindelse med fødsler og hospitalsindlæggelse. Selv om dit DNA behandles fortroligt, slipper det ud som følge af en IT-skandale og ender på internettet. Nogle år senere lykkes det forskere at udvikle en metode, der ud fra DNA kan forudsige, hvor stor sandsynlighed en person har for at udvikle sukkersyge eller svær depression.

Hvordan vil du have det, hvis det viser sig, at du har høj risiko for at udvikle en medfødt svær depression? Forestil dig, hvilken betydning det kan have, hvis du ønsker at købe en sundhedsforsikring, søge job eller finde en kæreste.

Det er en hypotetisk situation, men den nuværende udvikling peger i retning af, at det kan blive virkelighed. Forskerne bliver hele tiden bedre til at forstå vores gener, og indsamling af DNA sker allerede i dag. Samtidig viser adskillige eksempler os, at følsomme data kan havne i de forkerte hænder på trods af beskyttende lovgivning og IT-sikkerhedsforanstaltninger.

Problemet med indsamling af DNA illustrerer et generelt dilemma. I takt med udviklingen indsamles stadig større mængder af data, eksempelvis fra forskningsforsøg, færden på nettet og smartphones. Vi tænker ofte ikke over det, og i mange tilfælde er der tale om følsomme data. På den ene side er vi interesserede i de fordele, indsamlingen giver os. På den anden side udgør indsamlingen af data en trussel mod vores privatliv.

At regne på krypterede data

Fordelen ved indsamling af datamængder skyldes ofte, at man kan *regne* på de indsamlede data. Tænk på Google, hvis forretning består i at lave beregninger ud fra enorme mængder indsamlede data, der tillader dem at målrette reklamer. Et andet eksempel er forskningsdatabaser, hvor man samler store mængder af information om borgeres sundhed. Her ligger værdien i, at forskerne kan regne på de indsamlede data, fx udregne statistikker.

Man kan til en vis grad sikre sig imod misbrug og beskytte borgernes privatliv via lovgivning. Aktuelt er EU på vej med en forordning, der strammer reglerne for, hvordan man behandler og opbevarer persondata. Hvis den bliver gennemført, vil virksomheder, der har rod i personfølsomme data, kunne straffes med op til fem procent af deres omsætning, dog maksimalt 100 millioner euro.

Lovgivning kan dog aldrig stå alene. Erfaringen viser, at der altid vil kunne ske læk som følge af IT-skandaler og kriminalitet.

Forfattere

Thomas P. Jakobsen,
ph.d. studerende
Datalogisk Institut,
Aarhus Universitet
tpj@cs.au.dk

Sebastian Cappelen,
journalist på Alexandra Instituttet
sebastian.cappelen@alexandra.dk

Kryptering kan give et ekstra niveau af sikkerhed. Hvis data krypteres, før de indsamles, så kun krypterede data opbevares, vil privatlivet være sikret, også selvom virksomheden bliver hacket. Men når først data er krypteret, kan man ikke længere regne på data, og selve idéen med indsamlingen går derfor tabt.

Men der findes en løsning, der tillader, at man kan regne på krypterede data. Teknologien hedder *secure multiparty computation* (MPC). Pointen er, at man kan regne på data fra forskellige parter på en måde, så data altid forbliver hemmelige, men kun resultatet bliver kendt.

Fra tung teori til praktisk anvendelse

De første MPC-teknikker stammer fra midten af 1980'erne. I 1987 kom et gennembrud, hvor det lykkedes forskere at vise, at alle beregninger i princippet kan udføres som MPC. Det viste teknologiens potentiale, men de første resultater var meget upraktiske. Teknologien er siden optimeret, og den MPC, vi kender i dag, er meget mere praktisk anvendelig.

En anden milepæl for teknologien blev nået i 2009. Her blev MPC for første gang anvendt i stor skala i en kommerciel sammenhæng. Det var den såkaldte sukkerroebørs, et projekt, der udsprang af et samarbejde mellem Aarhus Universitet, Alexandra Instituttet, daværende Danisco og de danske sukkerroe-producenter. Man var på udkig efter at introducere en elektronisk børs, hvor landmænd kunne købe eller sælge sukkerroe-kvoter. Men landmændene og Danisco kunne ikke enes om, hvem der skulle administrere børsen, da landmændenes bud indeholder fortrolige oplysninger.

Løsningen var at introducere en børs baseret på MPC, hvor landmændene kunne sende deres bud ind, uden at andre kunne dekryptere dem. Børsen var med til at sikre et optimalt marked, hvor udbud bedst muligt møder efterspørgsel, uden at fortrolige informationer bliver afsløret. I første kørsel indgik ca. 1200 fortrolige bud, og børsen har været aktiv lige siden.

Teknologi med potentiale

Secure multiparty computation er et redskab, der er relevant i alle situationer, hvor man ønsker at udføre en beregning på fortrolige data, og hvor der ikke er nogen oplagt udenforstående part, som alle deltagere stoler på, der kan indsamle de fortrolige data og udføre beregningen.

Sukkerroe-børsen var med til at kickstarte virksomheders interesse for MPC. Flere virksomheder, bl.a. Microsoft, IBM, SAP og danske Partisia arbejder i dag målrettet med MPC.

Partisia samarbejder bl.a. med Aarhus Universitet og Alexandra Instituttet i projektet PRACTICE om at lave en prototype, kaldet Secure Survey. Det handler om at gøre elektroniske spørgeskemaer fortrolige og kan fx anvendes, hvis en virksomhed vil vide, hvor tilfredse medarbejderne er med chefen. Den enkelte medarbejders oplysninger forbliver krypteret, samtidig med at man kan offentliggøre, hvor mange der er utilfredse.

I COBE-projektet (Confidential Benchmarking) handler det om, at alle banker er interesseret i at lave en kreditvurdering af deres kunder. Bankerne har hver deres metode til at beregne kreditværdigheden af en bankkunde. Vurderingen ville dog blive meget bedre, hvis også andre bankers kundedata kunne indgå, men hver banks kundedata er fortrolige. Med MPC vil bankerne kunne forbedre deres kreditvurderinger uden at afsløre fortrolige kundedata.

Anonymisering er ikke nok

Det er ikke kun ved hackerangreb, at data utilsigtet ender i offentligheden. I nogle tilfælde har man forsøgt at sikre fortrolighed ved at anonymisere indsamlede data, fx ved at fjerne åbenlyse markører som navne og CPR-numre. Det er eksempelvis ofte tilfældet ved medicinske databaser, der frigives til forskning. Ofte har deltagerne i et forskningsforsøg kun ønsket at deltage under forudsætning af, at data anonymiseres.

Eksempler fra USA viser, at anonymisering alene kan være en tvivlsom løsning. I 2008 viste dr. David W. Craig, der er genetiker ved forskningsinstitutionen TGEN i Phoenix, hvordan man kan identificere en person ved at matche en prøve af personens DNA op imod indholdet af en anonymiseret forskningsdatabase. Opgaven burde være umulig og svarer til at lede efter en nål i en høstak.

Efterfølgende har Yaniv Erlich fra Whitehead Institute for Biomedical Research i Massachusetts vist, hvordan det er muligt at identificere personer, hvis DNA indgår i forskning, ved at krydsreferere den anonymiserede DNA med data, der er offentlig tilgængelige, alene via en internetopkobling.

Tilfældene har været øjenåbnere i USA. Forsøgspersoner, der deler deres DNA, risikerer tab af ikke bare deres eget privatliv, men også af deres børns og børnebørns, som vil arve mange af de samme gener, siger Mark B. Gerstein, professor ved Yale, der studerer genetiske databaser.

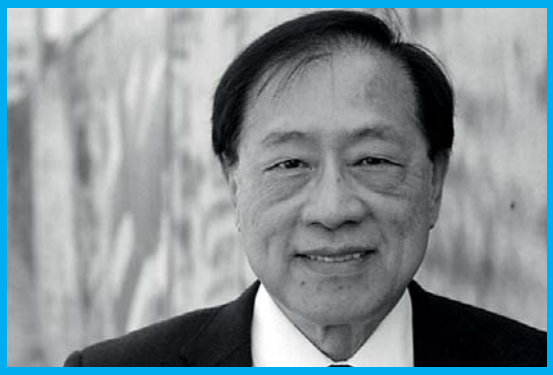
Fra teori til virkelighed

Det var et teoretisk gennembrud i 1980'erne, der gjorde det muligt at udføre vilkårligt avancerede beregninger på fortrolige data. De første resultater var upraktiske, men gav dog andre forskere blod på tanden. I dag arbejder adskillige forskningsgrupper bl.a. i USA, Israel, England og Danmark på at gøre teknikken mere effektiv.

Eksemplerne nedenfor giver et indtryk af, hvad man kan med moderne MPC. De varierende udførselstider skyldes bl.a., at MPC-teknikkerne har forskellig grad af sikkerhed. Som det ses, er det endnu ikke alle slags beregninger, der effektivt kan udføres med MPC, men med den hastige udvikling, forventer man at være meget længere om blot få år.

AES er en moderne krypteringsalgoritme. Med en nøgle k og en besked m beregnes en krypteret besked som c = AES(k, m). Når AES beregnes via MPC betyder det, at man kan få krypteret beskeder, hvor kun den ene part kender beskeden og kun den anden part kender nøglen. Det har mange anvendelser, og AES bruges derfor ofte som benchmark i forbindelse med MPC. En besked, fx et tal mellem 0 og 2^{128}, kan via MPC krypteres på ca. 0,5 sekund.

Fællesmængde (*private set intersection*) er en anden nyttig MPC-funktion. Kan fx bruges af to organisationer, der hver har et fortroligt register over personer. De vil kunne identificere personer, der optræder i begge registre, uden at nogen af dem behøver afsløre for den anden, hvem der ellers findes i registeret. I en lempelig sikkerhedsmodel er det muligt at beregne fællesmængden af to mængder med hver ca. 300.000 CPR-numre på 14 sekunder.



Andrew Yao betragtes som grundlæggeren af den diciplin indenfor kryptografi, der kaldes secure multiparty computation.

Levenshtein-distance (edit distance) L(a,b) for to tekststrenge a og b er det minimale antal simple operationer (insert, delete, modify), der skal til for at komme fra a til b. Kan bl.a. bruges til at sammenligne to DNA-sekvenser uden at kende sekvenserne. For et alfabet med 4 bogstaver og a og b på hver 200 tegn kan L(a,b) beregnes via MPC på 16 sekunder.

Videre læsning

Du kan finde ekstramateriale med et simpelt eksempel på, hvordan Secure Multiparty Computation fungerer.: aktuelnaturvidenskab.dk/nyeste-numre/2-2015/

En aktuel konkurrence med henblik på at finde de mest effektive teknikker til genetisk analyse uden at lække information: www.humangenomeprivacy.org

Beskrivelse af den danske auktion for sukkerroekvoter: http://eprint.iacr.org/2008/068

Confidential Benchmarking (COBE) projektet: http://ifro.ku.dk/english/research/centres/msap/research/research_projects/cobe/

Practice projektet: http://practice-project.eu

Mange anvendelsesmuligheder

Teknologien kan også bruges indenfor mange andre områder:

- Man kunne beregne det gennemsnitlige antal patientdødsfald som følge af lægefejl pr. hospital i Danmark, uden at et enkelt hospital er tvunget til at løfte sløret for antallet af sine fejlbehandlinger.
- Man vil kunne lave en elektronisk afstemning, hvor de fortrolige input er stemmer og resultatet af beregningen viser, hvem der har fået flest stemmer, uden risiko for at de enkelte stemmer bliver kendt.
- To konkurrerende forskerteams vil kunne afgøre, om de er nået frem til det samme forskningsresultat, fx en ny kemisk formel, uden at afsløre deres formler til hinanden eller andre.
- Man vil kunne matche forskellige personers DNA eller trække specifikke egenskaber ud af DNA, uden at nogen behøver at afsløre sit DNA.
- Man vil kunne undgå kollisioner mellem fjendtlige satellitter, droner, eller militærfly, da en potentiel kollision vil kunne detekteres, uden at nogen behøver at afsløre sin position eller kurs.
- Man vil kunne lave en fortrolig dating-side, hvor man ikke afslører sine præferencer, men alligevel vil kunne blive matchet med en partner med samme præferencer.

Behov for at kontrollere data

I takt med at vores liv bliver mere digitalt, bliver secure multiparty computation stadig mere aktuel. Med fænomener som big data og cloud computing opstår der et behov for en mere finkornet kontrol med, hvilke data der er fortrolige, og hvilke der er offentlige. Grundlæggende handler det om at holde fast i vores privatliv og fortrolighed uden at give afkald på fordelene ved big data, bl.a. de muligheder, der ligger inden for DNA-forskning og andre sundhedsområder.

Der er kort sagt et kæmpe potentiale i teknologien, og man kan forvente, at MPC inden for de næste 20 år vil være med til at skabe en revolution. Situationen svarer på mange måder til 1970'erne, hvor man opdagede public-key kryptering. I starten havde det mest teoretisk interesse, men i dag bliver public-key kryptering brugt mange steder i vores dagligdag, bl.a. til at sikre hjemmesider og til NemID. ■