

Behov for algoritmer

FORSKNING. Komplekse data i hastigt voksende mængder er et stigende problem. Og datamasserne kan behandles bedre, hvilket gør grundforskning i algoritmer væsentligt, mener leder af nyt center på Århus Universitet.

Af Jens Schnabel

jls@ncom.dk

»Vi er blevet bedre til at indsamle data, end vi er blevet til at behandle det. Vi samler data hurtigere, end vores computere bliver hurtigere, og så når vi til et punkt, hvor softwaren ikke kan følge med,« fortæller Lars Arge, professor i datalogi ved Århus Universitet og leder af det nye center for algoritmik, MADALGO.

Traditionelt har der ikke været meget fokus på at løse problemer med virkeligt store datasæt. Hvis problemet blev større, købte man bare en endnu større maskine. Men det er vi gået væk fra,« vurderer Lars Arge og beskriver tendensen til, at vi bruger flere mobiltelefoner og bærbare computere, men stadig forventer at de stadig kan løse alle opgaver.

Drukner i data

»COWI laserskanner for tiden hele landet til nye højdekort. Hvis du måler Danmark med et eller flere punkter på hver meter, så får du så store datasæt, at der faktisk ikke er noget software, der kan håndtere det - det går simpelthen dødt,« siger Lars Arge og forklarer dermed, hvordan bedre algoritmer kan være



Lars Arge er professor på Datalogisk institut ved Århus Bibliotek og leder af det nye »Center for Massive Data ALGORithmics«, kaldet MADALGO. Foto: Morgen Fauerby /www.apella.dk

afgørende for, om så store projekter kan lykkes. »COWI havde på et tidspunkt skannet hele jernbanenettet. De skulle bruge det til at lave støjberegninger, så de kunne finde ud af, hvor meget toget støjede, når det kørte igennem landskabet. Men det kunne de ikke, fordi deres software døde på datasættet – det var ganske enkelt for stort til, at de kunne beregne det. De måtte smide halvdelen af deres data væk, før de kunne lave beregningerne.«

»Et eksempel fra it-området er »streaming,« hvor data kommer konstant. Det kunne være »denial-of-service«-angreb, hvor en masse maskiner giver sig til at spørge den samme router eller maskine om noget, så netværket går ned. Der er mange penge i at opdage, hvornår sådan noget sker. Derfor er der masser af routere på nettet, der beretter om, hvilken tilstand de er i lige i øjeblikket og hvilke datapakker, der passerer. I sådan et tilfælde er

Lars Arge

- Professor ved Datalogisk Institut og leder af MADALGO ved Århus Universitet. 39 år.
- Lars Arge arbejdede i otte år på Duke University i USA, der gerne ville hente ham tilbage og lokkede med professorat og dobbelt løn. Men han valgte at blive i Århus, da det nye center blev en realitet.
- Han har desuden en Ph.D. grad fra Århus Universitet, 1996.

der så meget data, at det ikke kan gemmes, og man kan kun se et øjebliksbillede af, hvordan netværket ser ud. Så må du lave algoritmer til at finde ud af, om der er et »denial of service«-angreb i gang. De data, man gerne vil finde, er der. Problemet er bare at finde dem, for der er så mange, at man ikke ved hvordan,« forklarer Lars Arge.

I USA bruges sensorer til at måle jordrystelser, så det næste store jordskælv i Kalifornien kan forudsiges. Det giver mængder af data, som bruges til at overvåge udviklingen. Men Lars Arge mener, at de mange data er ved at kvæle os. Løsningen er, at vi får nogle hurtigere algoritmer til at løse problemerne.