

Introduktion til Internettet

Peter Havskov Christensen

30. januar 2012

1 Baggrund

Internettets virkemåde er påvirket af både historiske og tekniske årsager. Lidt baggrundsviden om nettets historie og teknik kan gøre det nemmere at forstå nettets mærkværdigheder og giver indblik i dets stærke og svage sider.

1.1 Historien: et produkt af den kolde krig

Internettet er et produkt af det teknologikapløb, der var en del af den kolde krig. Da USSR i 1957 opsendte Sputnik svarede det amerikanske forsvarsministerium igen ved at oprette Advanced Research Projects Agency (ARPA). ARPA skulle give USA en førerposition indenfor videnskab og teknologi.

Under ARPA udviklede Poul Baran ideerne til et distribueret netværk. Den vigtigste bevæggrund for den decentrale struktur var ønsket om, at der ikke måtte være et uundværligt led i nettet. Under den kolde krig var frygten for, at en fjenden skulle bombe centrale kommunikationscomputere udbredt. Det var derfor naturligt at etablere et netværk uden central styring. De computere, der skal deltage i et sådant decentralt netværk, kaldes værtsmaskiner (hosts) og stiller i et eller andet omfang ressourcer til rådighed for brugere af nettet. Der er altså tale om enkeltstående computere, der kan kommunikere med hinanden.

I 1969 oprettes internettets forløber ARPANET. De 4 første værtsmaskiner var hos University of California Los Angeles (UCLA), Stanford Research Institute¹, University of California Santa Barbara (UCSB) og University of Utah. To år senere (i 1971) er ARPANET oppe på 23 værtsmaskiner og flere kendte steder er blandt de nytilkomne blandt andet RAND, Massachusetts Institute of Technology (MIT), Harvard, Stanford og NASA.

Arpanet 1969

I 1973 etableres de første forbindelser udenfor USA. Disse gik til University College of London (England) og Royal Radar Establishment (Norge).

I 1982 benyttes ordet Internet første gang i forbindelse med nogle tekniske ændringer². ARPANET når i 1982 op på 235 værtsmaskiner.

I 1983 opdeles ARPANET i ARPANET og MILNET. De militære dele fortsætter i det mere lukkede MILNET og det amerikanske civile forskningsråd National Science Foundation (NSF) får ansvaret for den civile del. I 1984 passerer antallet af værtsmaskiner 1.000. I 1986 ændres nettets navn fra ARPANET til Internet. I 1987 passerer antallet af værtsmaskiner 10.000 og i 1989 passerer det 100.000.

1986: Arpanet ændrer navn til Internet

¹Ikke at forveksle med Stanford University.

²Transmission Control Protocol (TCP) og Internet Protocol (IP) indføres på ARPANET. Disse protokoller blev taget i brug den 1. januar 1983. De bruges fortsat på Internet.

En af internettets mest populære services World Wide Web (WWW) introduceres af det fælleseuropæiske kernefysiske laboratorium CERN i Schweiz i 1991. Hvad der egentlig var tiltænkt at være et internt informationssystem viste sig at være anvendeligt på hele internettet. 1991 var også året hvor Internettet blev frigivet til kommerciel brug.

1991: WWW

I 1992 når antallet af værtsmaskiner 1 million og ved nettets 25 års jubilæum i 1994 er antallet af værtsmaskiner tæt på 4 millioner.

Primo 1997 er antallet af værtsmaskiner nået over 16 millioner.

1.2 Teknik: Ingen central styring

Internet har ingen ejere og ingen ledelse. Internet er et internationalt netværk af computere. Omkostningerne ved driften afholdes af de primære brugere: højere læreanstalter, regeringer samt de kommercielle internetudbydere.

Uanset forandringerne fra det militærprægede ARPANET til dagens mere kommercielle Internet er den grundlæggende ide om at ingen del af nettet må være uundværlig for helheden stort set bibeholdt. Det betyder blandt andet, at der ikke kan være en overordnet kontrolfunktion eller et samlet register over brugere eller information.

Herved opstår et af Internettets helt store praktiske problemer. Når en masse information er tilgængelig på en masse forskellige maskiner, ville det være praktisk med en samlet fortegnelse. Det er på grund af den decentrale struktur ikke muligt at lave en komplet fortegnelse.

Ingen samlet indholdsfortegnelse

Der findes i stedet søgemaskiner på internettet. Disse søgemaskiner indsamler så meget information som muligt og lader brugerne foretage søgninger i det. Selv om de på grund af nettets struktur aldrig kan blive komplette, er de en meget stor hjælp for brugerne. Disse søgemaskiner omtales nærmere i afsnit 4.1.

Den decentrale struktur er også en enorm styrke. Internettet anses af mange for et af de mest effektive våben mod censur. Nettets struktur bevirker, at selv om myndighederne visse steder i verden spærrer for materiale de finder uacceptabelt, så betragter nettet disse spærringer som fejl. Internettet søger at korrigere for denne fejl ved at sende materialet en anden vej. På den måde kan data der skal overføres fra Danmark til mellemøsten godt gå over USA eller Australien. Grundtanken er, at selv om der er fejl (censur betragtes af Internettet som fejl), skal signalet igennem på en eller anden måde, hvilket er et enormt problem for udøvere af censur. De amerikanske myndigheder, som på en del områder er noget mindre frisindede end vi er i Skandinavien, har således også problemer med at censurere det Internet, som de selv har skabt.

Internets decentrale struktur gør censur vanskeligt

Uanset typen af information man vil finde eller stille til rådighed på Internettet, så er det nødvendigt at vide lidt om Internettets navnestruktur, da det ellers er stort set umuligt at finde rundt.

2 Internettets navnestruktur

En internetadresse kaldes også en Uniform Ressource Locator (URL). De fleste af de URL-er man ser i dagligdagen³ henviser enten til sider på World Wide Web (WWW) eller til en elektronisk postkasse.

URL

³Både aviser og TV omhandler efterhånden næsten dagligt internet.

2.1 Navne på WWW

På URL-en <http://www.kongehuset.dk/kongeraekken/kongeraekke.html> findes en oversigt over den danske kongerække. URL-en kan opdeles i flere dele og bliver derved nemmere at forstå. I eksemplet med kongerækken skal URL-en forstås således:

1: Protokollen	2: Computerens navn	3: Kataloget på computeren	4: Filens navn
http://	www.kongehuset.dk/	kongeraekken/	kongeraekke.html

De enkelte elementer i URL-en har følgende betydning:

- **Protokollen** specificerer hvorledes din computer og informationsudbyderens computer skal tale sammen. HTTP står for HyperText Transport Protocol, og er den protokol, der anvendes på World Wide Web. Andre muligheder er: ftp, gopher, news, telnet og mailto. Når man indtaster adresser i en webbrowser kan man normalt undlade at skrive <http://>, idet det er standardprotokollen i de fleste browsere.
- **Computerens navn** angives for at internettet kan lede vores forespørgsel videre til den rigtige computer. Navnestrukturen er et af de få steder, hvor internettet er hierarkisk. Computerens navn (her www.kongehuset.dk⁴) skal læses bagfra. Endelsen dk betyder at vi befinder os i hovedgruppen (top-level domainet) dk, hvilket betyder Danmark. Vi kan også se at vi befinder os i gruppen (domainet) kongehuset. Den computer som vi henvender os til hedder www. Det er meget anvendt at kalde den computer, der indeholder oplysninger til brug på World Wide Web, for www. Det gør vi også her på HHS, idet vores maskine hedder www.hhs.dk. Det kan læses som computeren www i gruppen hhs i hovedgruppen dk. Eftersom Internet er opfundet i USA, er det kun udenfor USA, at man bruger landekoder som hovedgrupper. I USA benyttes hovedgrupperne gov (regeringen på nær militæret), mil (det amerikanske militær), edu (uddannelsesinstitutioner), com (kommercielle virksomheder) og org (ikke kommercielle organisationer). Internationalt orienterede virksomheder udenfor USA vælger dog ofte hovedgruppen com i stedet for deres landekode (et eksempel herpå er www.lego.com).
- **Kataloget**⁵ angiver hvor på computeren den ønskede fil findes. I eksemplet med kongerækken hedder kataloget kongeraekken. Bemærk at der ikke må indgå danske bogstaver i URL-er.
- **Filnavnet** angiver hvilken fil i det pågældende katalog der ønskes.

I eksemplet med kongehuset er det en ret lang URL der bliver resultatet. Man kan dog også bare angive <http://www.kongehuset.dk>. Derved henvises til den pågældende computer, men ikke til en bestemt fil (et bestemt dokument). Det er så op til informationsudbyderens computer at sende standarddokumentet. Et standarddokument indeholder ofte klikbare henvisninger (links) til de øvrige dokumenter. Kongehusets hjemmeside ser således ud (<http://www.kongehuset.dk>):

Man kan her klikke på kongerækken for at komme til den tidligere omtalte side. Det er nemmere end at skrive den lange URL.

⁴Hvis man bare henviser til computeren og ikke til et bestemt katalog eller en bestemt fil er den afsluttende skråstreg valgfri.

⁵I Windows 95 og Windows NT kaldes kataloger for mapper eller folders.

Udenfor USA benyttes
landekoder

Korte adresser

Når HHS's homepage har adressen `http://www.hhs.dk`, er det også kun en henvisning til den pågældende computer, som så sender sit standarddokument, som er HHS's homepage. Metoden med at definere et standarddokument virker også nede i de enkelte kataloger. Når jeg oplyser at min homepage findes på URL-en `http://www.hhs.dk/~phc`, er det egentlig ikke helt præcist. Den komplette URL er `http://www.hhs.dk/~phc/index.htm`, men den korte udgave virker, fordi filen `index.htm` er standardfil i kataloget `~phc`.

På den måde gøres URL'er ofte så korte som muligt. Herved bliver de enkelte adresser nemmere at huske, men det bliver også sværere at lære URL-ernes logik og opbygning.

Efter ovenstående retningslinier er det logisk, at URL-en `http://www.stanford.edu` peger på Stanford University's homepage. Efter de netop beskrevne retningslinier burde `http://www.monash.au` være URL-en for Monash University i Australien's homepage. Det er desværre ikke sådan. De engelsktalende lande vil gerne efterligne den amerikanske måde, hvor man kan se hvilken type institution der gemmer sig bag URL-en. Derfor indskydes et niveau mere således at adressen `http://www.monash.edu.au` er den korrekte for Monash University. I England bruges `ac` for academic, og London School of Economics' homepage findes derfor på URL-en `http://www.lse.ac.uk`. Hvorledes der underopdeles i de enkelte lande er helt op til landet selv.

Det er vigtigt at erindre, at det er op til indehaveren af gruppenavnet at bestemme computernavnet. HHS kunne lade sin WWW-server have et andet navn end `www.hhs.dk`. HHS har også andre WWW-servere. En af disse har adressen `http://www3.hhs.dk`, men den indeholder kun homepages for ansatte og studerende i Sønderborg. En bruger af `http://www.hhs.dk` skal være meget opmærksom for at opdage, at nogle af dokumenterne ligger på `http://www3.hhs.dk`.

WWW-adresser kan, når de er skrevet komplet, genkendes på at de starter med `http://`. Brugeren behøver dog ikke skrive `http://` når man vil se den pågældende side, men når man oplyser URL'er bør det medtages.

Underopdeling er et nationalt anliggende

2.2 Mailadresser

En anden meget væsentlig type internetadresser er mail-adresser. De kan kendes på at de altid indeholder et @. En mailadresse er opbygget af et brugernavn, et @ og en identifikation af mailservoren. Strukturen er `brugernavn@computernavn.gruppe.hovedgruppe` og min mailadresse er derfor `phc@ko.hhs.dk`. Heraf kan ses at der i gruppen `hhs` er en computer, der modtager mails, der hedder `ko`. Der findes andre maskiner i gruppen `hhs` der kan modtage mails. Disse hedder `sb`, `es` og `fl`.

I forbindelse med mailadresser kan computernavnet nogle gange undlades. Den amerikanske præsident's⁶ mailadresse er `president@whitehouse.gov`. Der er gruppen `whitehouse` og hovedgruppen `gov`. Computernavnet er ikke angivet.

2.3 Andre adresser

Et eksempel på en internetadresse der ikke hører i en af ovennævnte kategorier kan være `ftp://ftp.dante.de` som blandt andet indeholder gratis software. Her benyttes ftp-protokollen. En FTP-adresse kan godt kontaktes af de moderne www-browsere (de programmer man kigger på WWW med), men man bør huske `ftp://`⁷.

⁶Lad være med at sende dumme og provokerende mails til personer du ikke kender godt. Du risikerer at blive flamet (du får så mange mails der fortæller dig hvor dum du er, at du ikke kan lave andet end at tømme din elektroniske postkasse).

⁷Mange FTP-servere svarer efterhånden også hvis de kontaktes med http-protokollen. Betydningen af at huske `ftp://` er således faldende.

3 Services

Services på internettet opdeles normalt efter hvilke protokoller de benytter. I forbindelse med gennemgangen af navnestrukturen blev www-adresser og email-adresser gennemgået og det passer fint med, at disse 2 services idag er de vigtigste på nettet. Andre services har dog også betydning, så for fuldstændighedens skyld skal de alle nævnes her. For ikke at gøre det for teknisk bruges de populære begreber i stedet for de tekniske (og mere præcise) begreber:

3.1 FTP

FTP står for File Transfer Protocol. Protokollen har været anvendt længe og benyttes til at overføre filer fra en maskine på internettet til en anden. Eftersom http-protokollen også kan overføre filer, har den vundet indpas på steder, hvor der tidligere blev brugt ftp. Det er muligt at lave brugervenlige programmer der benytter ftp, men da de http-baserede programmer var blandt de første brugervenlige programmer har ftp fået et lidt teknisk rygte.

3.2 Mail

Email er sammen med WWW de helt store hit på Internet. Email er både hurtigt og billigt. En email vil normalt ikke bruge mere end et lille minut om at nå til Australien eller USA.

Efterhånden som email-programmerne er blevet mere brugervenlige, må email siges at være for alle. Email giver mulighed for at vedhæfte filer. Det betyder at man kan sende filer uanset type (tekstbehandling, regneark, billeder m.v.) som "bilag" til en email. Det kan i nogle email-systemer virke lidt teknisk, men har man behov for at udveksle data med andre, er det en effektiv løsning.

3.3 WWW

World Wide Web (WWW) benytter HyperText Transfer Protocol (HTTP). Teknikken blev frigivet i 1991. Selv om protokollen hentyder til hypertext kan den overføre andre typer filer end hypertext og har derfor delvis fortrængt ftp.

WWW er den service, der har gjort Internettet tilgængeligt for alle. Det smarte er, at man utrolig nemt kan følge henvisninger til bestemte dokumenter også selv om det er på en helt anden computer. Normalt kan man klikke på den tekst der er blå og understreget og i en del tilfælde også figurer og knapper.

Et godt startsted på dansk er de emneopdelte sider hos <http://www.jubii.dk> og de kan være til stor nytte for begyndere på nettet, idet de giver henvisninger til en nogle populære danske sider som er gode at øve sig på.

3.4 News

Nyhedsgrupper er steder på internettet hvor der foregår diskussioner. De enkelte indlæg gemmes på såkaldte nyhedsservere. Diskussion er emneopdelt og der findes omkring 20.000 grupper. Det er dog ikke alle nyhedsservere der har alle grupper.

Indlæg i nyhedsgrupper kan både være svar på andres indlæg og nye indlæg. De fleste af de programmer, der bruges til at læse nyheder med, kan sortere de enkelte beskeder således, at det er nemt at se hvilke indlæg der er svar på hvad.

I ovenstående billede ser vi et svar fra Yuri Franken til Crispin Chapman om benyttelse af en funktion i regnearket Excel. Newsgrupper bruges både af brugere, der hjælper hinanden og folk, der debatterer elektronisk.

På WWW findes, på adressen <http://www.dejanews.com>, et arkiv over indlæg i news-grupper. På den måde supplerer de enkelte services hinanden.

3.5 Telnet

Telnet kan populært betragtes som en ”forlængelse af tastatur- og skærmmkabel”. Du bruger internettet til at videresende dine tastetryk og modtager skærbilleder retur. Der er mange anvendelsesmuligheder. Protokollen anvendes stadig og nye brugere af Internettet vil på et tidspunkt støde på den. En del af de ting, der tidligere kun var tilgængelige via Telnet, bliver med tiden tilgængelige via WWW. Normalt er WWW-adgangen noget mere brugervenlig, så den bør foretrækkes hvor det er muligt.

3.6 Gopher

GOPHER er en protokol der blev lanceret omkring 1990. Den var tænkt som et mere brugervenligt og menustyret alternativ til FTP. Da WWW og http-protokollen også tilbød disse faciliteter har Gopher aldrig vundet stor udbredelse. Nye internetbrugere behøver ikke bruge tid på den.

3.7 IRC

Internet Relay Chat (IRC) er ikke en ny protokol, men på grund af nye og smarte programmer der benytter denne protokol, anvendes den mere og mere. Den benyttes som skrivetelefon og mange kan deltage på en gang. Brugere etablerer via Internettet forbindelse til en IRC-server, og skriver så til hinanden.

Et meget populært program, der benytter IRC, hedder mIRC. Det ser således ud:

Diskussionen foregår i såkaldte ikke eksisterende (virtuelle) rum. Det viste skærbillede er fra rummet ”#beginner”. Det øverste vindue viser diskussionen og til højre ses listen over deltagere. Det nederste vindue er et statusvindue.

Der holdes indenfor visse faggrupper diskussionsmøder via IRC på faste ugentlige eller månedlige tidspunkter, men ellers anvendes IRC mest til underholdningsbrug.

4 Eksempler på anvendelse

Dette afsnit giver eksempler på de mangeartede former for oplysninger der findes på internettet. Det er en noget blandet omgang, men det er valgt for både at vise nogle tekniske finesser og steder med brugbar information.

4.1 Søgmaskiner

Teknisk set er det at stille information om internettets indhold til rådighed ikke anderledes end at stille anden information til rådighed. Det specielle ved disse søgemaskiner er, hvorledes de indsamler information om internettet.

Søgmaskiner kan principielt opdeles i 2 grupper:

- Søgmaskinen får en tilmelding (oftest via email eller WWW) med oplysninger om din side. Din side bliver set igennem og medtaget i de relevante kategorier. Det er måske lidt "finere" at få sin side registreret i en sådan søgemaskine fordi der er foregået en eller anden form for (overfladisk) review.
- Søgmaskinen (eller en hjælpemaskine) surfer rundt på nettet, og registrerer de sider den finder. For at blive optaget skal der være nogen, der har lavet link til din side, eller du skal selv gøre maskinen opmærksom på din sides eksistens.

En del af de mest populære søgemaskiner benytter en kombination af de 2 teknikker. Den første type søgemaskiner er gode for begyndere, eller hvis du blot er på nettet for en hyggetur. De kan også have deres berettigelse, hvis du skal se hvem, der er "store" på internet indenfor et givet område.

De punkter der hedder "Det sker", "EDB", "Erhverv" med videre hos den danske søgemaskine Jubii (<http://www.jubii.dk>) er eksempler på sider der først er reviewet inden de optages. Den viste side er et udmærket startsted for begyndere på nettet.

En anden søgemaskine, der drives af de samme som driver Jubii, men som optager alle de danske sider den støder på er Jori (<http://jori.jubii.dk>). Af efterfølgende billede kan ses, at der ikke er emner, men kun et felt til at indtaste søgeord. Jori kender til gengæld mange flere sider end Jubii. Selv om Jubii også har mulighed for indtastning af søgeord, så er det begrænset hvad nytte denne funktion giver, idet den kun søger i de sider der er med i de angivne grupper. Som bruger må man vælge mellem flot ordnede søgemaskiner, der ikke kender så mange sider (eksempelvis Jubii), eller søgemaskiner der kender mange sider, men som ikke er systematiseret (eksempelvis Jori).

Begge ovennævnte søgemaskiner optager kun danske sider. Når man vil søge udenfor Danmark, er der gode muligheder på følgende adresser:

URL	Kort beskrivelse
http://www.altavista.telia.com	Hurtig. Sponsoreret af Digital. Kender mange sider. Hurtigt svar. Fri søgning (ikke systematiseret).
http://www.yahoo.com	Jubii's udenlandske forbillede. Kender mange sider. Fri søgning samt systematiseret indeks.
http://www.lycos.com	En klassiker. Kender mange sider. Fri søgning samt systematiseret indeks.

For de fleste søgemaskiner gælder, at de har ekstra faciliteter. Det er derfor i praksis mere værd at have et rimeligt kendskab til 2-3 søgemaskiner end et overfladisk kendskab til 20.

4.2 Biblioteker

En del biblioteker stiller information til rådighed på internet. Dette gølder også HHS's bibliotek (<http://www.hhs.dk/bibliotek/>). En oversigt over biblioteker på Internet findes hos Jubii (<http://www.jubii.dk/referenc/biblio.htm>).

4.3 Nyheder

En del tidsskrifter og aviser er gået på Internettet. En oversigt er tilgængelig hos Jubii (<http://www.jubii.dk/erhverv/aviser.htm>).

Danmarks Radio (<http://www.dr.dk>) er også på Internet. Her kan man også læse tekst-tv siderne online (<http://www.dr.dk/cgi-bin/ttx/100>).

4.4 Netkameraer

Netkameraer er noget af det mindre nyttige. Det er mest en demonstration af teknikken muligheder. Hvis du vil se et billede fra London (opdateres hvert 5 minut) så kig på <http://www.fujiint.co.uk/street/index.html>.