



Vakdidactiek informatica : praktijkgerichte seminars

Lesgevers :

Prof. A. Hoogewijs / Annick Van Daele

Vakgroep Pure wiskunde en computeralgebra

Galglaan 2

9000 Gent

Tel. : 09/2644900 – Fax. : 09/2644993 – E-mail :

Albert.Hoogewijs@rug.ac.be

Internet:

geschiedenis

organisatie

werking

zoekmachines

Mieke Depaemelaere

Mieke.Depaemelaere@rug.ac.be

Inhoudsopgave

Internet

1. Geschiedenis
2. Succes van het Internet
3. Structuur van het Internet
4. Een bericht versturen over het Internet
 - 4.1 Pakketten
 - 4.2 Voordelen van dit systeem
 - 4.3 Protocollen
 - 4.3.1 Het TCP/IP protocol
5. Identificatie van een computer op het Internet
 - 5.1 Domeinen
 - 5.2 DNS
6. Enkele definities

Internet

1. Geschiedenis

De wortels van het Internet gaan terug tot de late jaren zestig. In die tijd werd in de Verenigde Staten door het ministerie van Defensie een project gefinancierd dat het 'ARPANET' werd genoemd. Het doel van het Arpanet was het met elkaar verbinden van een aantal computerfaciliteiten in het hele land.

De late jaren zestig was de periode van de Koude Oorlog. Het Amerikaanse ministerie van Defensie was vooral bezorgd over de mogelijkheid van een nucleaire aanval. Daarom werd het Arpanet zo ontworpen dat het zelfs zou blijven werken als een deel van het netwerk door een bom werd getroffen.

Het project waarin deze technologie werd ontwikkeld, slaagde en tijdens de jaren zeventig en tachtig breidde het netwerk zich steeds verder uit. Zo ontstond het Internet. In eerste instantie verbond het Internet alleen militaire en universitaire computers met elkaar, maar langzamerhand kwamen er steeds meer bedrijven en individuen bij. Tegenwoordig hebben miljoenen mensen over de hele wereld toegang tot het Internet.

2. Succes van het Internet

Het Internet is dus binnen één enkele generatie uitgegroeid van een experimenteel project tot een wereldwijd netwerk dat onze cultuur blijvend heeft veranderd. Maar waarom is dat zo en hoe kon het zo snel gebeuren?

Allereerst heeft de oorspronkelijke technologie, die door het Amerikaanse ministerie van Defensie werd ontwikkeld, bewezen heel betrouwbaar te zijn. Hoewel niemand ooit een bom op het Internet heeft laten vallen, is het heel gewoon dat computers en communicatielijnen tijdelijk stoppen met werken. Op het moment dat zoiets gebeurt, wordt de informatiestroom door andere computers op het Internet automatisch omgeleid, totdat de verbroken verbindingen weer beginnen te werken. Een tweede reden waarom het Internet zo snel is gegroeid, heeft te maken met de menselijke natuur. We zoeken middelen om te communiceren. Toen in de jaren negentig de personal computers voor de gewone consument betaalbaar werden, duurde het niet lang meer voor tientallen miljoenen mensen er een kochten en gingen meedoen aan het Internet.

3. Structuur van het Internet

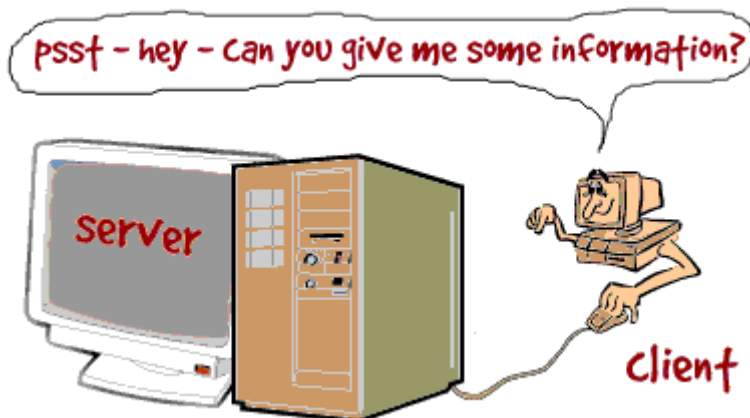
Het Internet kan eigenlijk gezien worden als het netwerk der netwerken. Het is de koppeling van een groot aantal afzonderlijke netwerken (universiteiten, leger, multinationals, overheidsinstellingen, thuisgebruikers,...), die op hun beurt uit ontelbare verschillende computersystemen bestaan. Dit betekent dat er geen eigenaar is van het Internet.

De computers die aangesloten zijn op het Internet worden hosts genoemd en deze bieden een aantal informatievoorzieningen aan.

Voorbeelden hiervan:

- e-mail (elektronische post)
- het werken op een andere computer op het net (Telnet)
- het kopiëren van bestanden van een bepaalde computer naar een andere (FTP)
- toetreden tot discussieforums
- en natuurlijk het interactief exploreren van het Internet via het World Wide Web (surfen)

Er wordt een onderscheid gemaakt tussen de client-hosts en de server-hosts. Een client-host is een computer die rechtstreeks door eindgebruikers wordt gebruikt, terwijl een server-host niet rechtstreeks met eindgebruikers wordt geassocieerd. De client is een programma dat om een dienst vraagt (aanwezig op de client-host) en de server is een programma dat een dienst verleent.



Een voorbeeldje om dit te verduidelijken:

Stel dat je naar iemand een e-mail wilt versturen. Hierbij maak je gebruik van een mailclientprogramma zoals bijvoorbeeld Outlook Express. Zodra je het bericht hebt samengesteld, gebruikt de mailclient het Internet om contact te maken met een mailserverprogramma op een computer op afstand. De client zendt het bericht naar de server, die er vervolgens voor zorgt dat het wordt afgeleverd. (details doen hier nog niet echt terzake) Je kan ook de mailclient gebruiken om te controleren of je berichten hebt ontvangen. De client neemt daartoe contact op met de server en

vraagt alle berichten op die naar je werden verstuurd. De server stuurt deze berichten naar de client en deze opent ze op het beeldscherm van de computer.

4. Een bericht versturen over het Internet

4.1 Pakketten

Het Internet functioneert doordat computers met elkaar verbinding maken om gegevens (informatie) over en weer door te geven. Om deze communicatie tussen de computers efficiënt en betrouwbaar te maken, hanteert het Internet een systeem waarin gegevens in brokken worden verdeeld. Dit werkt als volgt.

Een computer die gegevens naar een andere computer wil verzenden, verdeelt die gegevens in elektronische pakketten. De pakketten zijn genummerd en gemarkeerd met het adres van de bestemmingscomputer. Vervolgens worden ze het Internet op gestuurd om bezorgd te worden. Aan de andere kant verzamelt de bestemmingscomputer de pakketten en voegt ze bijeen om de oorspronkelijke gegevens te herstellen.

Een voorbeeldje:

Stel dat je via e-mail een foto naar een vriend(in) wil versturen. De eigenlijke foto is in een bestand op de computer opgeslagen. Het bestand is opgesplitst in pakketten en naar de computer van je vriend(in) verzonden. Bij je vriend(in) worden de pakketten automatisch weer samengevoegd tot een kopie van het originele bestand. Het mooie en gemakkelijke van het Internet is dat je vriend(in) zich niet om de details moet bekommeren. Want voor hem/haar is het alsof het bestand intact van de ene computer naar de andere is verzonden.

4.2 Voordelen van dit systeem

1. Het is efficiënt. Het doet er niet toe hoeveel informatie van de ene naar de andere computer moet worden verzonden. De gegevens worden altijd in pakketten van gelijke grootte opgesplitst en op een efficiënte manier over het Internet verzonden. Het doet er bovendien niet toe hoe de pakketten naar de bestemming worden verzonden of in welke volgorde ze aankomen. De beslissing via welke weg de pakketten gaan noemt men *routing*. (zie routers) Er wordt uiteraard gekozen voor de snelste wegen. Over de snelste lijn worden meerdere pakketten verzonden dan over een tragere lijn. Wanneer een bepaalde route uitvalt, wordt gewoon een anderen route genomen.
2. Een tweede voordeel van dit systeem is de betrouwbaarheid. Bij aankomst van elk pakket hanteert de bestemmingscomputer een methode om te testen of er onderweg eventuele fouten zijn opgetreden. Als een pakket fouten bevat, stuurt de bestemmingscomputer een bericht naar de broncomputer

met het verzoek om het pakket nog eens te verzenden. Op die manier hoeven niet alle gegevens nog eens opnieuw worden verzonden, maar alleen het pakket dat met fouten is aangekomen. Dit systeem maakt het Internet betrouwbaar zonder dat er onacceptabele vertragingen optreden.

4.3 Protocollen

Computers die gegevens over en weer sturen over het Internet communiceren met elkaar. Hiermee wordt bedoeld dat een programma dat op de ene computer draait, gegevens verzendt naar een programma dat op de andere computer draait.

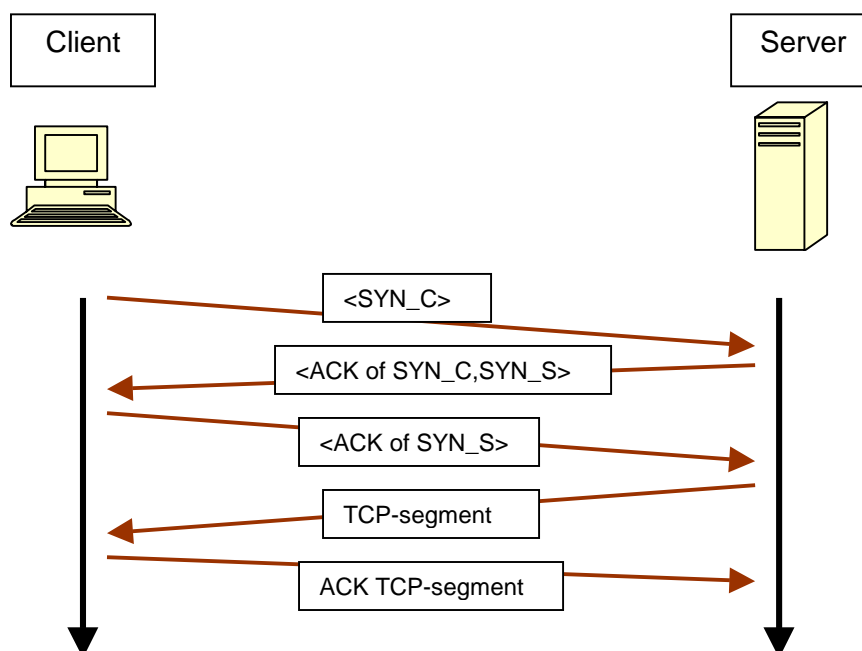
Dit systeem werkt alleen maar als de uiteenlopende computerprogramma's die op Internet actief zijn, in staat zijn gegevens te verzenden en te ontvangen volgens standaardspecificaties die protocollen worden genoemd. Een **protocol** definieert het formaat en de volgorde van de boodschappen bij communicatie tussen verschillende computers (of iets anders dat wil communiceren) en ook de acties die moeten gebeuren voor de transmissie en de ontvangst van een boodschap.

4.3.1 Het TCP/IP protocol

IP staat voor Internet Protocol en TCP staat voor Transmission Control Protocol. Computers die over het Internet met elkaar communiceren doen dit aan de hand van het TCP/IP protocol. TCP beheert de stroom pakketten en garandeert dat de gegevens intact, zonder fouten, op de plaats van bestemming aankomen. IP wordt gebruikt voor het transport van de gegevenspakketten van de ene plaats naar de andere te versturen.

TCP

Beschouw het volgende schema:



Uitleg:

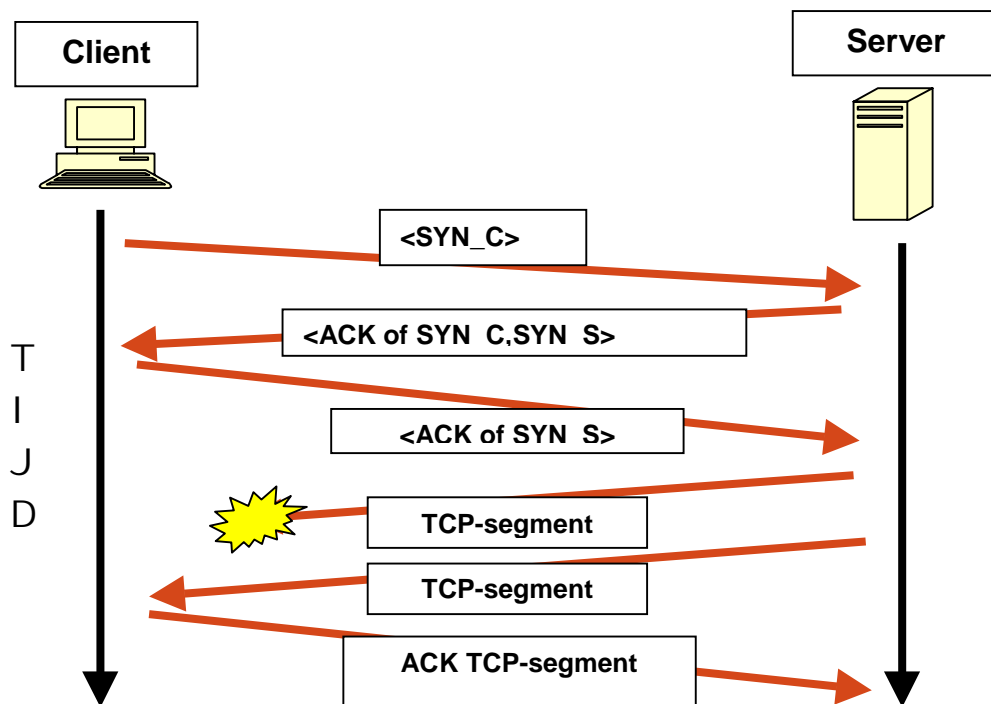
Vooraleer data van de server naar de client kan worden gestuurd, moet er eerst een connectie worden opgezet tussen beide.

1. De client (die de aanvraag doet) stuurt een SYN-segment (synchronisatie segment) naar de server. Hierin staat over welk soort applicatie het gaat. (FTP, e-mail)
2. De server antwoordt dan met een acknowledgement van het SYN-segment van de client (=goed ontvangen en OK) en voegt er een eigen SYN-segment aan toe.
3. De client antwoordt op de server met een acknowledgement van het SYN-segment van de server.

Dan kan de eigenlijke data-overdracht gebeuren. Dit gebeurt onder de vorm van TCP-segmenten. Een TCP-segment bevat een deel (vaste grootte) van het bestand,... dat moet worden doorgestuurd en een header. In deze header staat informatie over de applicatie.

Retransmissie (=het opnieuw doorsturen van een TCP-segment)

Als er na een bepaalde tijd geen ACK wordt ontvangen van een TCP-segment, dan wordt dit segment opnieuw doorgestuurd. Dit scenario treedt op als het TCP-segment niet of niet correct wordt ontvangen door de client of als de ACK van een bepaald segment niet wordt ontvangen in de server.



IP

Dit protocol zorgt ervoor dat IP-datagrammen van de ene naar de andere computer worden verstuurd. Een IP-datagram is opgebouwd uit een TCP-

segment en een header. De header bevat informatie over het segment (hiermee wordt bedoeld welk soort segment het is, want TCP is niet het enige mogelijke protocol (bv. UDP)) en adres van de bestemming en van de afzender.

5. Identificatie van een computer op het Internet

Elk van de computers op Internet aangesloten, heeft wereldwijd een uniek adres: het IP-adres. Een IP-adres bestaat uit vier getallen van 0 tot 255 die gescheiden zijn door punten. (bvb. 159.193.100.46) Dit IP-adres is noodzakelijk om een communicatie tot stand te brengen. Daar een nummer moeilijk is om te onthouden, koppelt men namen aan deze nummers. Deze namen (hostnamen) hebben een structuur als "www.microsoft.com".

Hostnamen zijn een onderdeel van een systeem dat het Domain Name System (DNS) wordt genoemd (later wordt hier nog op terug gekomen). DNS is het systeem dat het mogelijk maakt aan elke computer op het Internet een unieke naam toe te kennen.

Beschouw de volgende URL

<http://www.harley.com/25-things/index.html>

1. **http://**

Het eerste deel slaat op het protocol dat gebruikt wordt. Hier wordt het 'http' protocol gebruikt. (Hier wordt op terug gekomen). Dan volgt een dubbele punt en een dubbele 'slash' om het eerste stuk van de rest af te scheiden.

2. **www.harley.com/**

Dit is Internetadres van de computer (server) waarop de informatie te vinden is.

Het meest rechtse deel wordt het top-leveldomein genoemd en het geeft algemene informatie over de host.

3. **25-things/**

Dit is de directory of het pad waar het document zich bevindt op de host.

4. **index.html**

Dit is de naam van het document zelf. De extensie van zulke documenten is vaak html (Hypertext Markup Language) waarmee gewoon bedoeld wordt dat het om hypertext documenten gaat opgesteld met een speciale taal

geschikt hiervoor. Soms staat er geen bestandsnaam, dan wordt er standaard index.html genomen.

5.1 Domeinen

Beschouwen we nu alleen het Internetadres van de computer (server).

www.harley.com

Het meest rechtse deel is het toplevel-domein. Hier is dit com.

Er wordt een onderscheid gemaakt tussen twee soorten top-level domeinen: organisatorische domeinen en geografische domeinen. Organisatorische domeinen beschrijven een categorie, terwijl geografische domeinen een specifiek land aanduiden. Het organisatorisch domein edu staat bijvoorbeeld voor onderwijsorganisaties, terwijl het geografische domein au het land Australië aanduidt.

Tabel 1 geeft een overzicht van enkele organisatorische domeinen weer. Tabel 2 somt enkele geografische domeinen op.

<u>Domein</u>	<u>Beschrijving</u>
com	commercieel
edu	Amerikaanse Universiteiten (onderwijs)
gov	Amerikaanse federale overheid
int	internationaal
mil	Amerikaanse leger
net	netwerkproviders
org	diverse organisaties

Tabel 1: Organisatorische top-leveldomeinen

In de praktijk komt het regelmatig voor dat hostnamen met com, net of org geenszins aan de gestelde criteria voldoen.

<u>Domein</u>	<u>Beschrijving</u>
at	Oostenrijk
au	Australië
be	België
ca	Canada
ch	Zwitserland
de	Denemarken

fr	Frankrijk
it	Italië
jp	Japan
nl	Nederland
uk	Verenigd Koninkrijk
us	Verenigde Staten

Tabel 2: Voorbeelden van geografische top-leveldomeinen

Noot: Waarom die twee soorten top-leveldomeinen?

In zijn beginjaren was het Internet een Amerikaanse aangelegenheid en waren er maar een paar top-leveldomeinen: edu, com, gov, mil, org, net en int. Toen het Internet zich naar andere landen ging uitbreiden, werden de geografische domeinnamen ingesteld, zodat elk land ook een eigen top-leveldomein kreeg. Het was de bedoeling dat de organisatorische top-leveldomeinen gingen wegvallen, maar ze waren toen al zo ingeburgerd dat ze zijn gebleven.

Beschouw volgende hostnamen:

Treasury.gov.au

Health.gov.au

Deze twee hostnamen hebben de twee meest rechtse delen van de namen gemeen. Ze behoren tot hetzelfde second-leveldomein. Hier behoren ze allebei tot gov.au.

In zulke gevallen duiden we het specifiekere domein aan als een subdomein van het meer algemene domein. gov.au is het subdomein van au.

Op dezelfde manier behoort de hostnaam www.royal.gov.uk tot het domein royal.gov.uk dat een subdomein is van gov.uk dat zelf een subdomein is van uk.

Op deze manier zijn alle hostnamen op het Internet ingedeeld in één groot hiërarchisch systeem dat op domeinnamen gebaseerd is. Dit systeem is het Domain Name System of DNS.

5.2 DNS

Aan de hand van het systeem DNS (Domain Name System) worden hostnamen (en ook e-mailadressen) afgebeeld op IP-adressen. DNS is een hiërarchisch, op domeinen gebaseerd naamgevingssysteem plus een gedistribueerd databanksysteem (dit wil zeggen verspreid over verschillende hosts) die deze naamgevingmethode toelaat.

Als je bijvoorbeeld “www.microsoft.com” intikt in de adresbalk van je browser, dan roept dit applicatieprogramma een bepaalde procedure aan met de hostnaam als parameter. Hierdoor wordt een pakketje over het Internet naar een DNS-server gestuurd, die de naam opzoekt en het IP-adres aflevert aan de procedure die het aan de aanvrager doorgeeft. Pas dan kan het programma een verbinding met de gevraagde host tot stand brengen.

Hoe wordt de hostnaam vertaald in het IP-adres?

Het Internet is zodanig geconstrueerd dat elke organisatie zijn eigen deel van het Internet beheert. Daarom is het logisch dat elke organisatie ook zijn eigen hostnamen beheert. Het komt erop neer dat specifieke domeinnamen worden gebruikt door organisaties die dan ook verantwoordelijk zijn voor alle hostnamen en IP-adressen binnen die domeinen. Een voorbeeld: de organisatie IBM is verantwoordelijk voor ibm.com. Elke organisatie moet minstens twee computers inrichten, *naamserver*s genoemd, om voor alle hostnamen binnen hun domein in adresinformatie te kunnen voorzien. (Twee servers voor het geval er eentje tijdelijk uitvalt.)

Omdat DNS het beheer van de verschillende domeinen over het Internet verdeelt, hoeft er geen gigantische lijst van alle computers op het Internet te worden bijgehouden. Elke keer dat een programma een vertaling van een hostnaam in een IP-adres nodig heeft, is het voldoende hiervoor de naamserver van dat specifieke domein te benaderen.

Hoe vindt het programma nu de juiste naamserver onder de vele servers op het Internet? Het antwoord is dat vanaf de top-leveldomeinen naar beneden wordt gewerkt, totdat de juiste naamserver is gevonden.

Op verschillende plaatsen op het Internet staan speciale computers die rootnaamserver worden genoemd. Elke rootnaamserver houdt een lijst bij van naamserver s die top-leveldomeinen afhandelen. Elk van deze naamserver s heeft op zijn beurt een lijst met naamserver s die second-leveldomeinen beheren, enzovoort.

Als een programma het IP-adres voor een hostnaam nodig heeft, begint het zoeken vanaf een rootnaamserver en wordt van server tot server naar beneden gewerkt totdat de naamserver die het vereiste IP-adres bevat, gevonden is.

Een voorbeeldje om dit te verduidelijken:

Stel dat een programma op zoek is naar het IP-adres van de hostnaam www.ibm.com. Het programma begint met het raadplegen van de rootnaamserver om het adres van de naamserver com te verkrijgen. Vervolgens maakt het programma contact met deze com-naamserver om het IP-adres van de naamserver ibm.com te achterhalen. Als laatste wordt verbinding gemaakt met de naamserver ibm.com om het IP-adres van www.ibm.com te verkrijgen.

Om tijd te besparen, houdt elke server op elk niveau een lijst bij met recentelijk aangevraagde namen en adressen. Als dan een volgende aanvraag voor dezelfde informatie binnenkomt, kan de server onmiddellijk reageren.

6. Enkele definities:

URL's

Een URL of Uniform Resource Locator is de beschrijving van de locatie van webpagina's. Maar URL's kunnen ook naar andere bronnen verwijzen dan webpagina's, bijvoorbeeld mailadressen, ftp-bestanden,...

Voorbeelden:

<http://www.ibm.com>

<ftp://ftp.microsoft.com/Products/msmq/demos.zip>

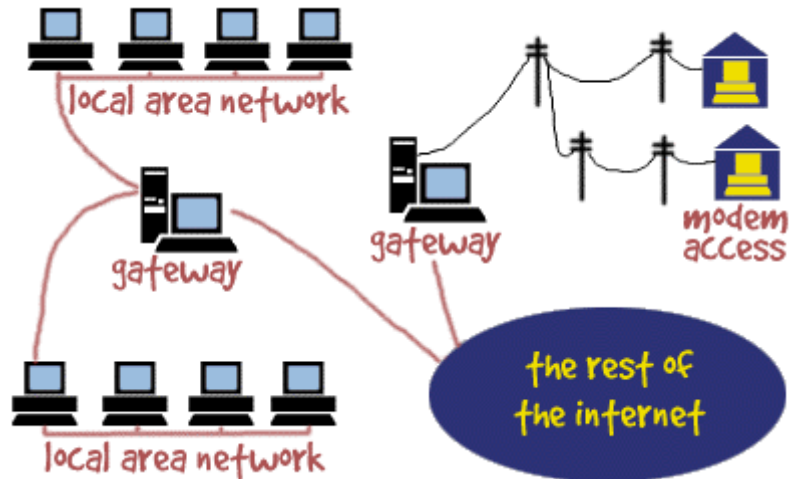
<mailto:billg@microsoft.com>

Mailto is een URL soort die geen hostnaam specificieert, omdat het zich ook niet op een specifieke computer bevindt. Als je op deze URL klikt dan wordt een mailprogramma geopend.

Opmerking: waarschijnlijk heb je al ondervonden dat het niet nodig is op <http://> in te tikken, www.ibm.com volstaat.

Gateway

Dit is een knooppunt in een netwerk (hier in het Internet) dat toegang biedt tot een andersoortig netwerk en daarbij alle formaat- en adresseringsproblemen oplost. Het is een programma dat een automatische verbinding tot stand brengt tussen twee soorten informatiesystemen.



Een specifiek voorbeeld hiervan is het versturen van een e-mail naar een systeem die een ander protocol dan TCP gebruikt. De email-gateway zal zorgen dat het bericht toch kan verstuurd worden.

Bridge

In een netwerk verbindt twee netwerken met elkaar zodat ze toegankelijk worden voor elkaar. Bridges kunnen gebruikt worden om twee verschillende types netwerken met elkaar te verbinden maar worden meestal gebruikt om één groot netwerk op te splitsen in twee kleinere netwerken voor performantie redenen. Een bridge kent alle adressen aan beide kanten en kan zo informatie doorzenden.

Router

Een computer die belast is met de verzending van pakketjes tussen verschillende netwerken en knooppunten van een packet-switching netwerk zoals het Internet. De router beschikt over (actuele) informatie over andere routers in het netwerk, en bepaalt op grond hiervan welke pakketjes naar welke knooppunten worden verzonden. Een router beschikt over een routingstabel met daarin de Internet-adressen die gebruikt worden voor het doorzenden van pakketjes naar een volgend knooppunt in Internet.

Internet

Wereldomspannend netwerk van computers (verbinding van kleinere netwerken) die communiceren volgens gemeenschappelijke protocollen en toepassingen. Kenmerkend voor het Internet is het ontbreken van een centraal computersysteem. Het gemeenschappelijke protocol is TCP/IP. Belangrijke toepassingen zijn e-mail, FTP, telnet en het World Wide Web.

Intranet

Een computernetwerk binnen een besloten eenheid (bijvoorbeeld een bedrijf) gebaseerd op de toepassingen, technieken en protocollen die ten grondslag liggen aan Internet. Het intranet kan verdeeld zijn over verschillende vestigingen en een gateway naar Internet bevatten. Een intranet wordt meestal gebruikt voor het distribueren van bedrijfsnieuws en -informatie, interne e-mail en discussieplatforms.

WWW

World Wide Web. Wereldwijde verzameling van sites gebaseerd op het Hypertext Transfer Protocol. Multimedia en hypertext zijn de belangrijkste kenmerken van het World Wide Web.

Hypertext

Systeem waarin gegevens koppelingen bevat naar andere documenten.

Hyperlink

Hypertext link: = koppeling: Verwijzing in een document naar een ander document. Dit kan tekst zijn, maar bijvoorbeeld een foto.

http

Hypertext Transfer Protocol – Protocol voor de uitwisseling van hypertext documenten op het WWW. Het is een toepassingsprotocol binnen TCP/IP.