



Servicii Internet I



Notiuni de baza in Internet

- Adrese IP
- Arhitectura client server
- Porturi
- Nume de domeniu si DNS

Adrese IP (1)

- Host – calculator conectat la o retea prin intermediul careia are acces la Internet
- Adresa IP – 32 biti (4 octeti) folositi pt identificare unica in Internet
 - 193.254.231.35 – adresa IP a serverului vega.unitbv.ro
- Adresa IP se compune din
 - Adresa retea
 - Adresa hostului in cadrul retelei
- Clasele de adrese IP se obtin pe baza primului octet al adresei

Clasa adrese	Primul octet din adresa	Nr de retele	Nr de hosturi
A	1 - 126	127	16000000
B	128 - 191	16384	65534
C	192 - 223	2000000	254
D	224 - 239	adrese multicast	
E	240 - 255	Utilizare pt experimente	

Adrese IP (2)

- Clasa de adrese IP determina lungimea cimpurilor

Clasa adrese	Lungimea adresei de retea (octeti)	Lungimea adresei hostului (octeti)
A	1	3
B	2	2
C	3	1

- Pentru aflarea adresei retelei si a hostului se foloseste masa de retea – numar pe 32 de biti care contine biti de 1 corespunzator fiecarui bit din adresa care este parte din adresa retelei
- Adresa IP 193.254.231.35 cu masa 255.255.255.0 -> adresa retelei 193.254.231.0 -> adresa host 35

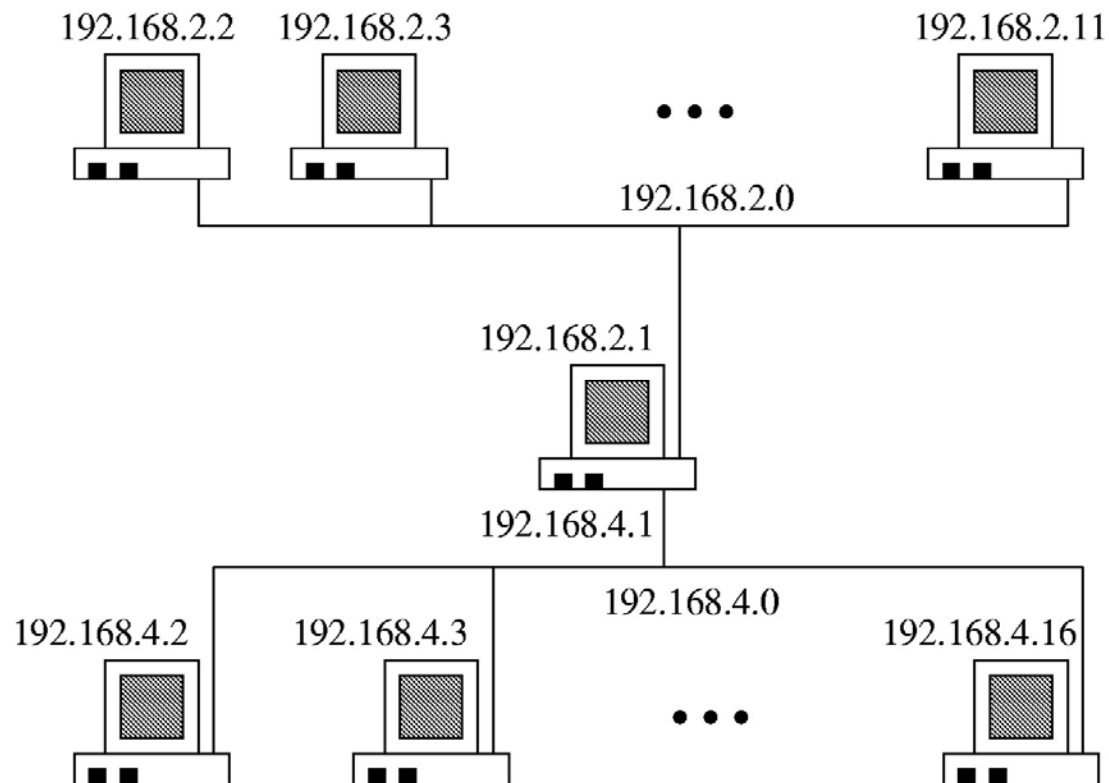
Adrese speciale

- Clasa A de adrese (127.0.0.0) – folosita pentru interfata SW a calculatorului – interfata de loop-back
- Adresa broadcast (de difuzare) – pt fiecare retea o adresa pt difuzarea datelor
 - Toti bitii corespunzatori host-ului sint 1
- Adrese IP rezervate pentru retele private
 - Neroutabile – hosturile care folosesc adresele sint parte dintr-un intranet

Clasa adrese	Adrese rezervate pentru retele private
A	10.0.0.0 - 10.255.255.255
B	172.16.0.0 - 172.31.255.255
C	192.168.0.0 - 192.168.255.255

Intranet

- Intranet – rețea privată bazată pe protocolul TCP/IP ce aparține unei organizații, care nu este neapărat conectată la Internet



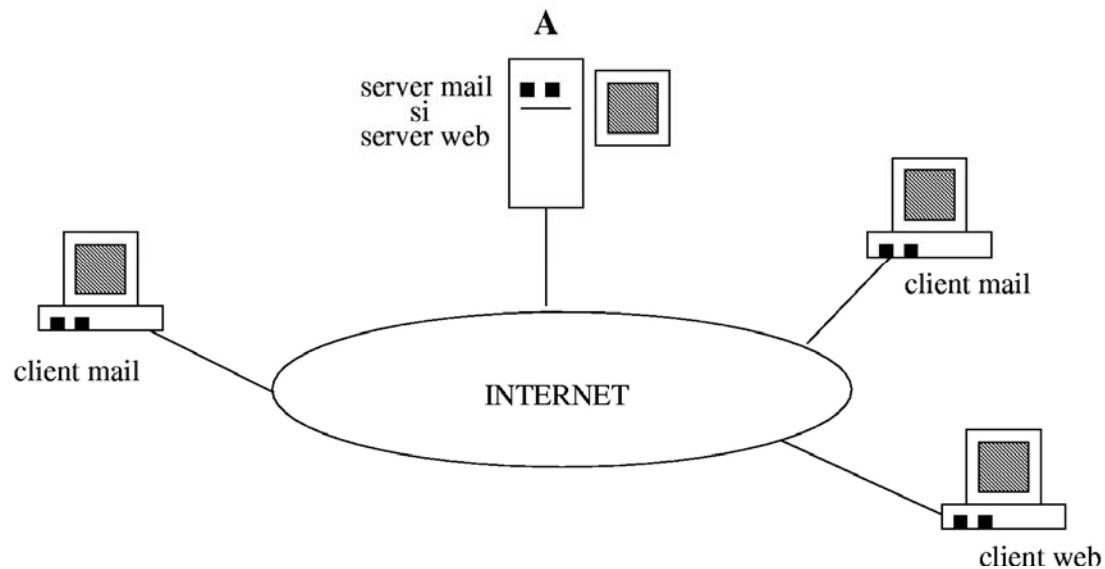


Arhitectura client-server (1)

- Arhitectura client-server este modelul care sta la baza tuturor serviciilor web
 - Permite o implementare eficienta
- Server
 - Dpdiv HW – calculator specializat/dedicat care ofera anumite servicii altor calculatoare din retea
 - Dpdiv SW – un program care ruleaza pe un calculator si ofera servicii altor programe care il apeleaza
- Client
 - Programe care ruleaza pe hosturi
 - Pornite de catre utilizatori
 - Solicita informatii/servicii de la serverele din Internet
 - Clienti web (Netscape, Firefox, etc), client mail (Thunderbird, Outlook, etc)
 - Necesita calculatoare cu performante medii

Arhitectura client-server (2)

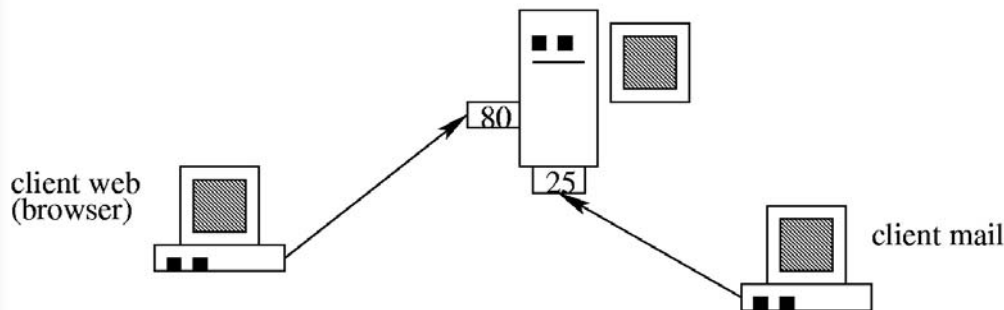
- Un host poate rula mai multe servere deservind in concomitent mai multi clienti



- Un server se activeaza doar in momentul in care un client solicita servicii apoi intra intr-o stare latentă (intre solicitari)
- Un server concurent – poate rezolva simultan mai multe cereri

Porturi

- Porturi – modalitatea de a accesa un anumit serviciu oferit de un server
 - Un server “asculta” doar un anumit port dedicat



Protocol (serviciu)	Port
ftp	21
ssh	22
telnet	23
smtp	25
nameserver	53
finger	79
http	80
pop3	110

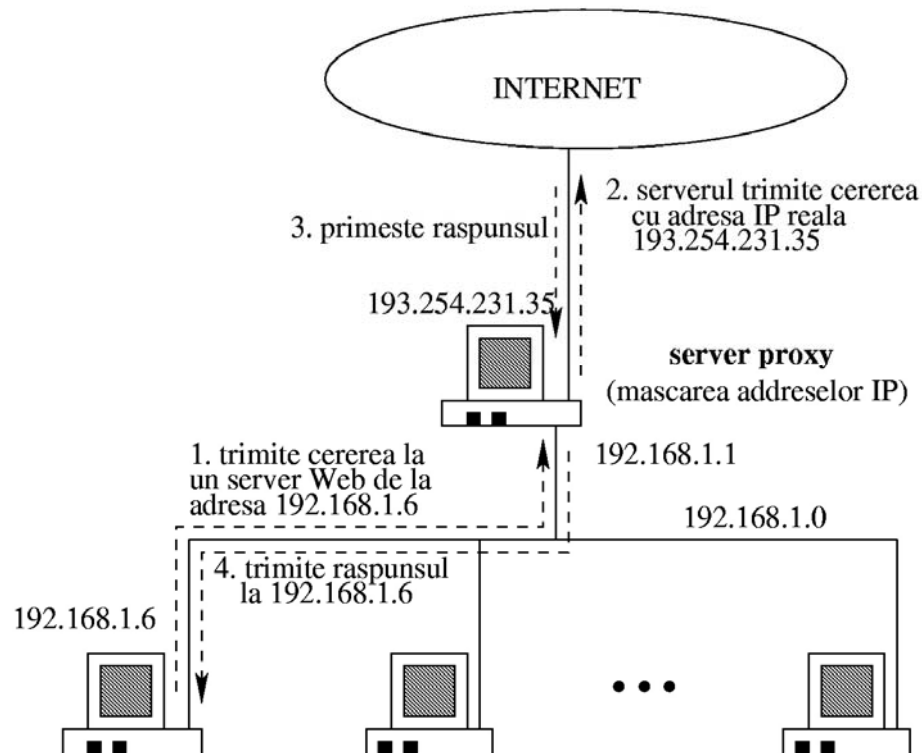


Servere proxy (1)

- Server proxy – este localizat la iesirea unei retele catre Internet
 - Partajarea unei conexiuni Internet (mascare adrese IP)
 - Filtrarea cererilor
 - Imbunatatire performantelor

Partajarea unei conexiuni Internet

- Lipsa adreselor IP -> partajarea unei singure adrese
 - Hosturile din rețeaua internă (intranet) folosesc adrese neroutabile
 - Proxy serverul este singurul care are adresa IP validă
 - Receptionează cereri din rețeaua internă și le trimite mai departe folosind adresa proprie (mascarea adresei IP)





Servere proxy (2)

■ Filtrarea cererilor

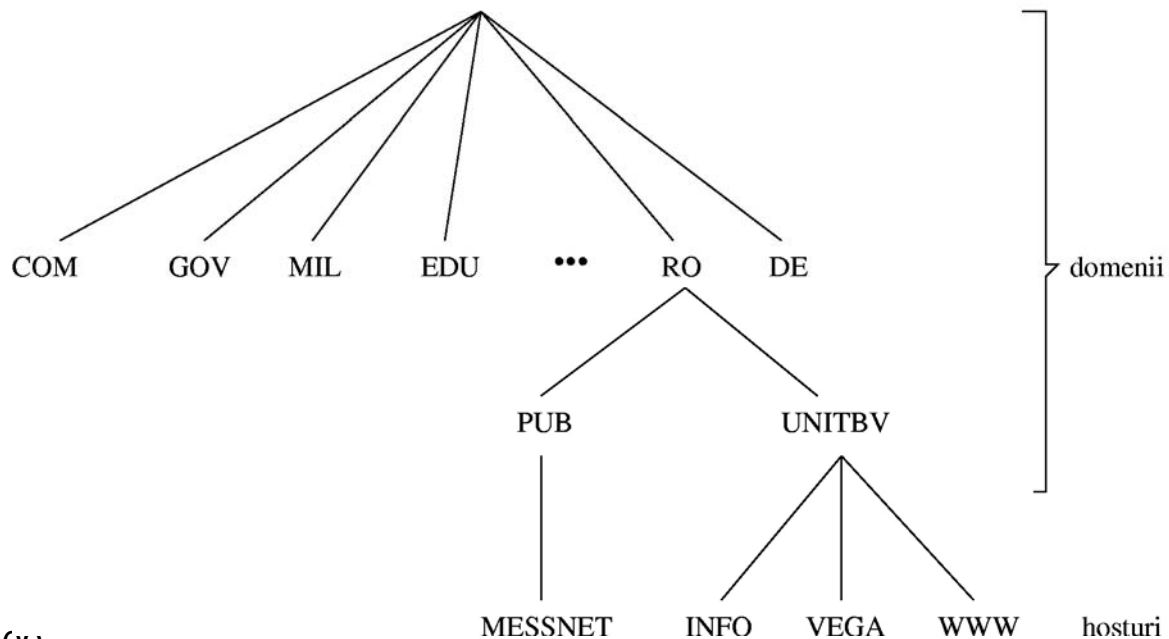
- Inspectarea traficului de iesire si intrare
- Permiterea sau interzicerea transmiterii datelor
- Reguli de filtrare bazate pe continut sau adresele sursa/destinatie/protocol

■ Imbunatatire performantelor

- Proxy cache – mentinerea paginilor cerute de utilizatori pentru accesaru ulterioare
- Daca pagina exista in cache se minimizeaza traficul

Nume de domeniu (1)

- Fiecare host este identificat de o adresa unica
 - Numele sint mult mai usor de memorat decit cifrele
 - Corespondenta intre adresa IP si numele de domeniu (ambele unice pentru un host)
- Structura ierarhica a numelor de domeniu





Nume de domeniu (2)

- Formatul numelui de domeniu :
 hostname.second_level.first_level
- Primul nivel (TLD – top level domain)
 - Cel mai puțin specific
 - Identifica țara de origine sau tipul organizației
 - Sute de astfel de domenii unice (com, ro, net, org, etc)
- Subdomenii (nivelul 2)
 - Identifica organizația (yahoo.com, google.com)
 - Pot exista duplicate la acest nivel (linux.com, linux.org)
 - Pot exista până la 127 de nivele de subdomenii
- Hostname
 - Numele calculatorului din subdomeniu respectiv (vega.unitbv.ro)



Nume de domeniu (3)

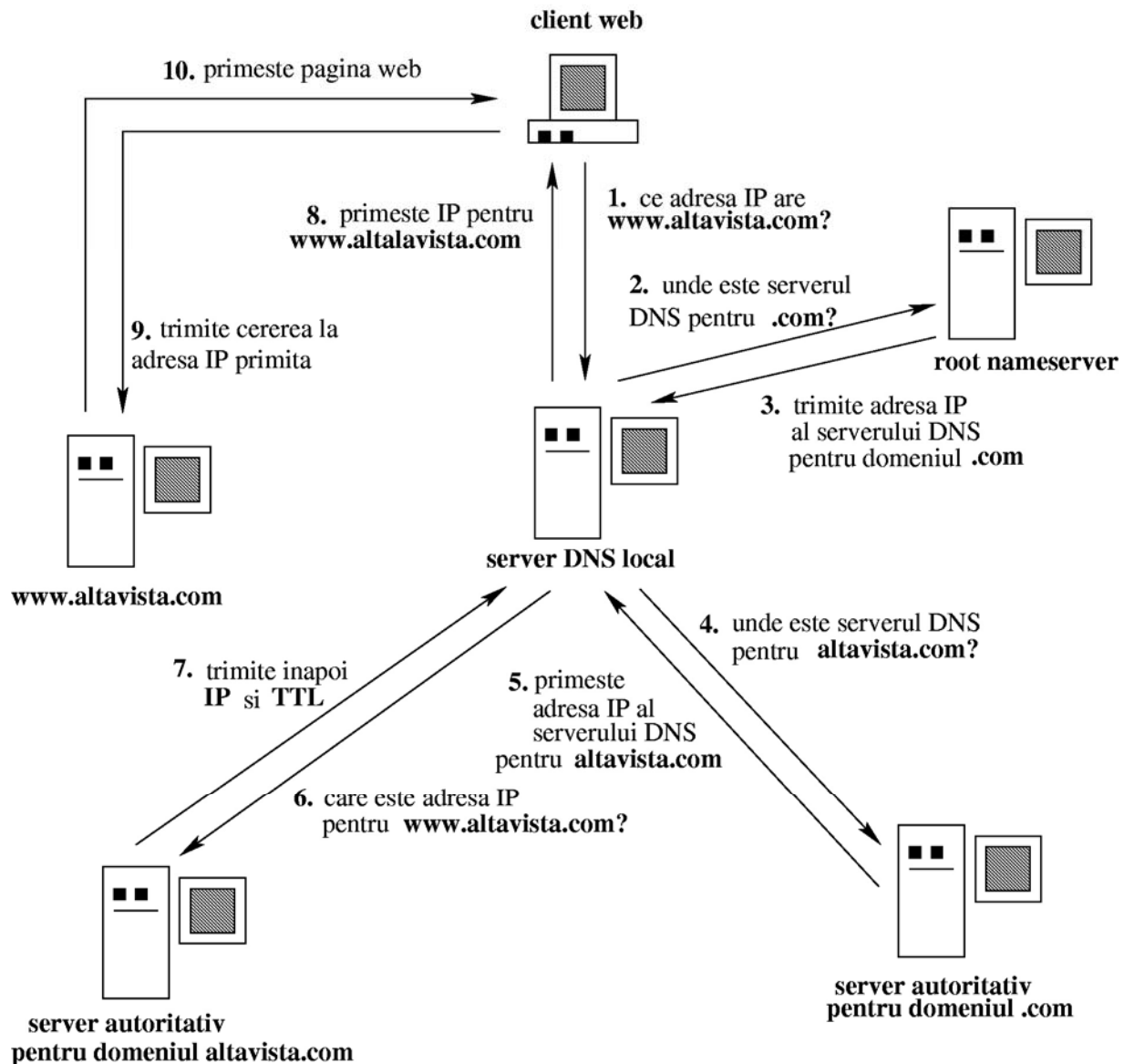
- Entitate centrala care aloca numele de domeniu pentru a pastra unicitatea
 - Structura ierarhica a organizatiilor care aloca numele de domeniu
 - In Romania RNC se ocupa de alocarea numelor de domeniu



DNS – Domain Name System (1)

- Sistemul de traducere a numelor de domeniu in adrese IP
 - Cea mai mare baza de date distribuita din lume
 - Necesitatea traducerii numelui de domeniu intrucat comunicarea intre 2 calculatoare se face pe baza adresei IP
- Initial nu exista un sistem automat
 - Fiecare host mentia o lista (host.txt) cu corespondenta nume de domeniu adresa IP
 - Fisierul era mentinut pe un server central si actualizat mereu
 - Pentru functionare el trebuia reactualizat periodic
 - Solutie functionala pt retele de dimensiuni mici
- 1984 – proiectarea sistemului ierarhic de rezolvare a numelui de domeniu
 - Fiecare nivel al numelui de domeniu este rezolvat de servere separate
 - Root name servere, servere autoritative si servere locale

DNS – Domain Name System (2)





E-mail - descriere

- Primul si cel mai des folosit serviciu – jumătate din conexiunile Internet sint folosite pentru mesaje e-mail
- Caracteristicile serviciului
 - Transmitere/receptionare simpla si eficienta
 - Viteza si usurinta de comunicare specifice unui apel telefonic
 - Siguranta livrării si permanenta specifice unei scrisori
 - Costa scazut fata de serviciul postal traditional
 - Timp de livrare extrem de scurt (minute)
- Avantajele fata de sistemul postal traditional (snail-mail)
 - Acelasi mesaj trimis mai multor persoane
 - Raspuns/retransmitere cu usurinta a unui mesaj receptionat
 - Atasamente (voce, imagini, video)
 - Permite sistem automat de raspuns (tip robot telefonic)



Mesajul e-mail

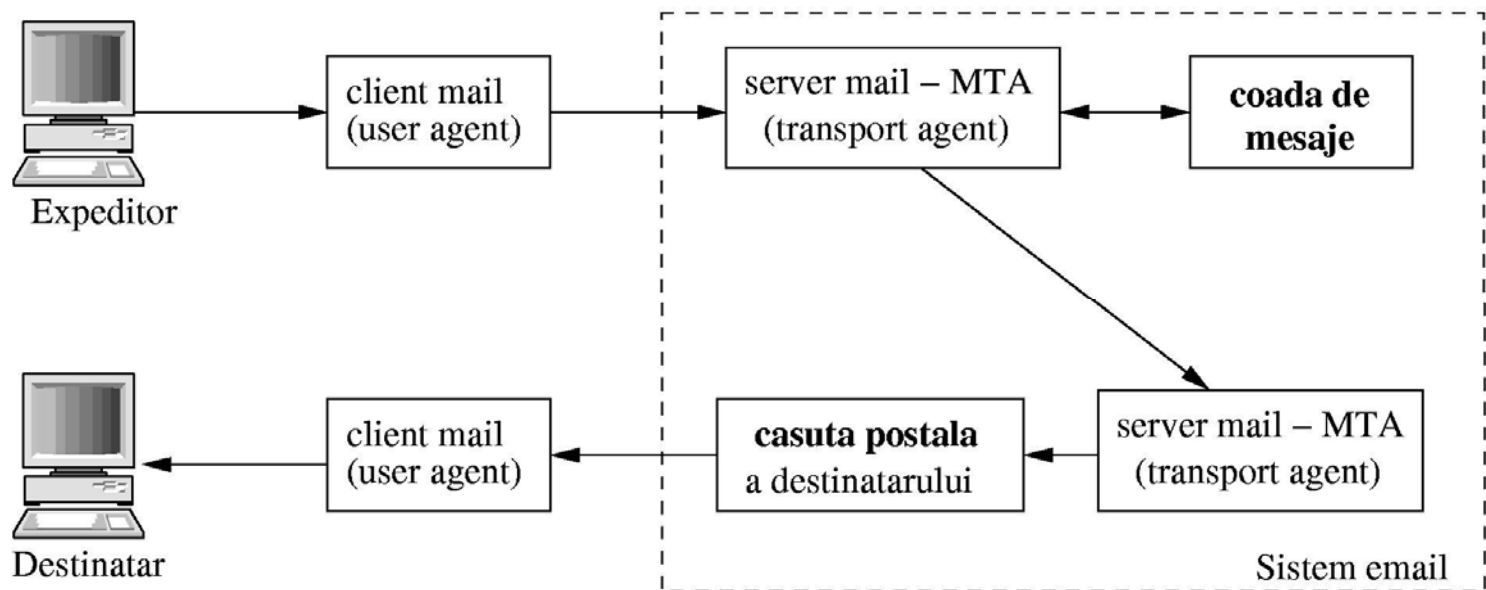
- Mesaj text de tip ASCII
- Contine 2 parti:
 - Header (antet)
 - Contine informatie de control cum ar fi server sursa, destinatie, timpul clientului
 - Contine cimpurile To, Cc, Subject care sint completate de utilizator
 - Este generat de serverul mail in momentul transmiterii mesajului
 - Corpul mesajului (message body)
 - Contine mesajul propriu zis
 - Eventualele fisiere atasate

Adresa e-mail

- Identificarea destinatarului se face pe baza adresei e-mail
- Sintaxa generala : [utilizator@nume.de.domeniu](#)
 - Partea de utilizator (user ID) este numele contului creat pe serverul destinatie
 - Partea de host identifica numele unui calculator sau al unui intreg domeniu
- Ex [gigi@vega.unitbv.ro](#) – gigi la hostul vega
[gigi@unitbv.ro](#) – gigi la domeniul unitbv.ro
- Serverul de mail pentru un intreg domeniu este referit prin numele domeniului
 - Numele lui este obtinut de catre serverul DNS sursa
 - Pentru unitbv.ro serverul de mail este gwm.unitbv.to

Sistemul e-mail

- Arhitectura client-server si protocolul SMTP (Simple Mail Transfer Protocol)
- Transferul mesajelor este asigurat de agentii de mail
 - MTA – Mail Transfer Agent – servere de mail (sendmail)
 - MUA – Mail User Agent – clientii de mail (Outlook, Thunderbird, pine)

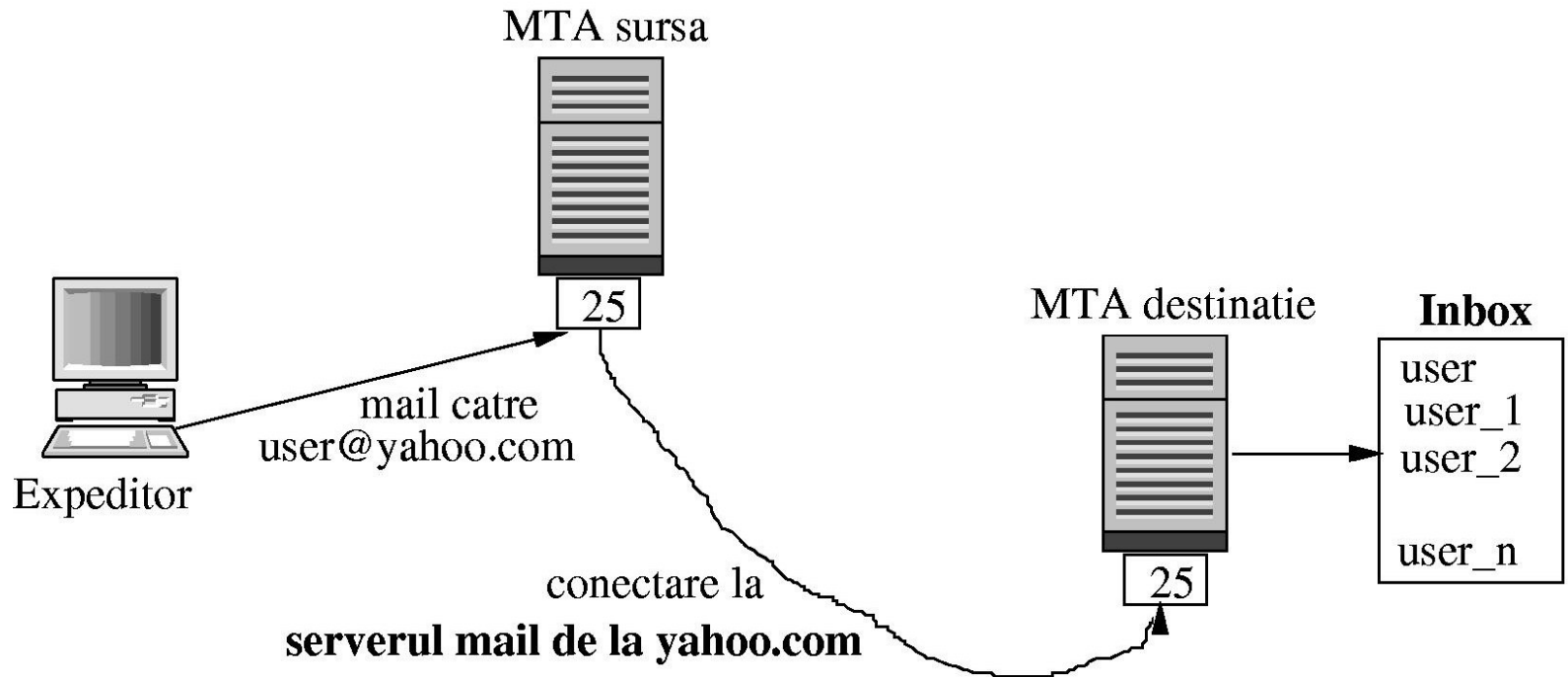


Transmiterea unui e-mail (1)

- Clientul de mail transmite mesajul serverului de mail local (MTA sursa)
- MTA sursa contacteaza DNS local pentru aflarea adresei IP a destinatiei (numele de domeniu al destinatiei este extras din adresa de mail)
- MTA sursa contacteaza MTA destinatie si transmite mesajul
 - Pentru o transmisie sigura se pastreaza local o copie a mesajului email care este stearsa dupa ce mesajul a fost livrat
- MTA destinatie identifica utilizatorul pe baza adresei de mail si stocheaza mesajul in casuta postala (Inbox)
- MTA destinatie comunica MTA sursa livrarea mesajului si acesta sterge copia locala a mesajului

Transmiterea unui e-mail (2)

- Inbox – un fisier in care sint concatenate toate mesajele e-mail
 - Fiecare utilizator are un inbox propriu





Mesaje de eroare (1)

- In caz ca transmiterea nu s-a efectuat MTA sursa incearca periodic re-transmiterea lui un anumit interval de timp (configurarea serverului de mail)
- La expirarea intervalului de timp se transmite utilizatorului un mesaj de eroare
 - MAILER_DAEMON sau postmaster apar ca expeditorii acestor mesaje
 - Mesaje contin motivul erorii
- Cele mai des intilnite erori apar din urmatoarele cauze
 - Eroare in partea de utilizator
 - Eroare in partea de host
 - Depasirea spatiului alocat inbox-ului



Mesaje de eroare (2)

- Eroare in partea de utilizator
 - Utilizatorul specificat nu are cont pe serverul de mail destinatie
 - Eroarea este detectata de catre MTA destinatie
 - Mesajul de eroare este transmis de catre MTA sursa dupa ce a fost informat de catre MTA destinatie
- Eroare in partea de host
 - Hostul/numele de domeniu nu exista in baza de date DN
 - Eroare este detectata de serverul DNS local
 - Mesajul de eroare este trimis de catre MTA sursa dupa ce a contactat DNS local
- Depasirea spatiului de stocare (over quota)
 - Nu se poate scrie mesajul in Inbox-ul destinatie intrucit s-a depasit spatiul alocat acelu utilizator
 - Eroare este detectata de catre MTA destinatie
 - Mesajul de eroare este trimis de catre MTA sursa



Citirea mesajelor

- Client de mail – aplicatia folosita pentru a citi mesaje stocate pe serverul mail
 - Aplicatie de sine statatoare (Outlook, Netscape, pine, etc)
 - Intefata WEB (Yahoo, G-mail, etc)
- Indeplineste urmatoare functiuni
 - Afiseaza mesajele primite impreuna cu antetul lor
 - Permite citirea/compunerea/retransmiterea mesajelor
 - o Transmiterea se face prin protocolul SMTP
 - o Citirea se face prin protocolul POP3 sau IMAP
 - Atasarea la mesaj a diferitelor tipuri de fisiere (audio, video)
- Conectarea la server-ul de mail
 - Conectare on-line
 - Conectare POP3
 - Conectare IMAP

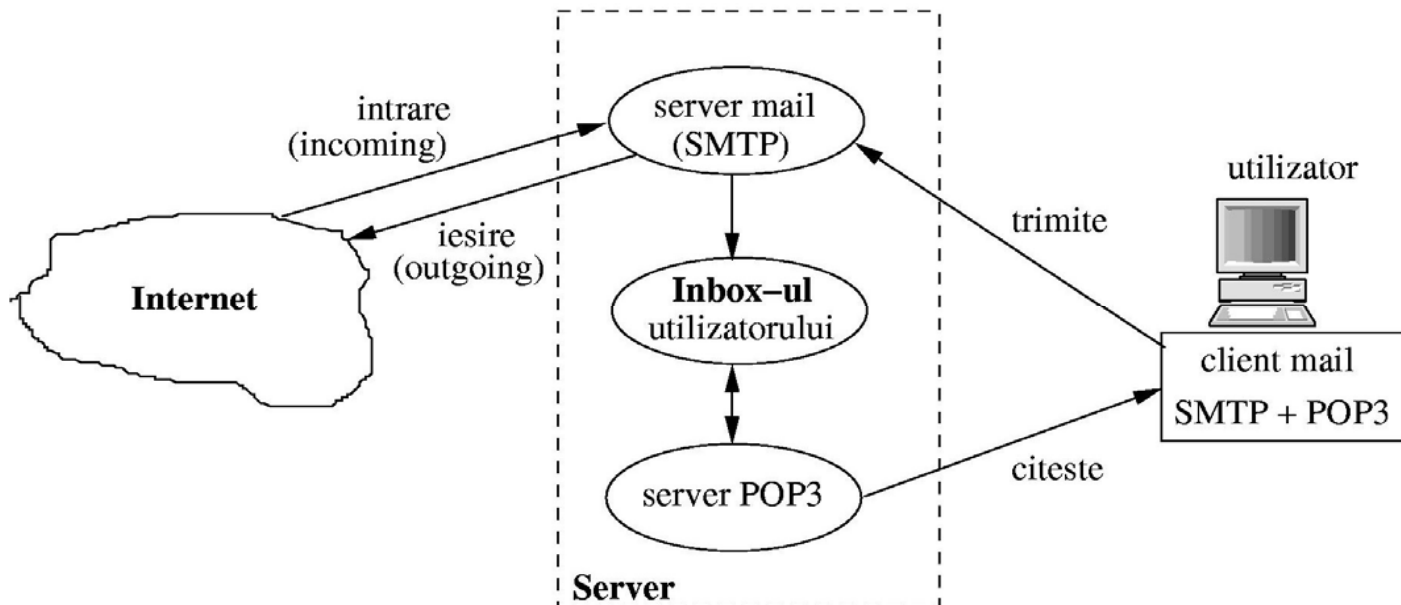


Conectarea on-line (1)

- Conectare directa pe serverul de mail (pt servere Unix-like) folosind clienti care ruleaza pe server (e.g pine) sau clienti web
- Avantaje
 - Se poate accesa de la orice calculator din retea si nu necesita setari speciale
 - Interfata este aceasi indiferent de sistemul de operare local
 - Performantele calculatorului local sint irelevante, totul se proceseaza pe server
- Dezavantaje
 - Nu se preteaza prea bine la conexiuni dial-up
 - Atasamentele trebuie tranferate in prealabil pe server

Acces POP3 (1)

- POP3 (Post Office Protocol) – protocol folosit pentru citirea mail-urilor de pe un server aflat la distanta
- Clientul ruleaza pe calculatorul local si foloseste protocolul PO3
 - Se conecteaza pentru a transfera mesajele
 - Citirea/compunere mesajelor se poate face off-line





Acces POP3 (2)

■ Avantaje

- Minimizeaza timpul conexiunii (pt conexiuni dial-up)
- Toate mesajele e-mail se pot stoca local (spatiu disponibil mai mare decit cel oferit de server)
- Se poate realiza o conectare periodica pentru verificarea mesajelor noi

■ Dezavantaje

- Nu permite lucrul de pe calculatoarele publice datorita setarilor si a faptului ca mesajele trebuie pastrate
- Nu se permite decit accesul cutiei postale de pe server, mesajele mail ne-putind fi organizate pe server ci doar local



Acces IMAP

■ IMAP

- Protocol pentru citirea mailurilor de pe un server de la distanta mai complex decat POP3
- Neceista o conexiune permanenta cu serverul
- Permite operatii mai complexe nu doar transferul mesajelor din Inbox
 - o transferul mesajelor salvate sau deja trimise
 - o transferul doar al antetelor
 - o Selectarea mesajelor ce se doresc transferate local



SPAM (1)

- Reprezinta toate mesajele e-mail receptionate fara a fi solicitate
 - Sint denumite si UCE (Unsolicited Commercial Email)
- Poate mari costul accesului la Internet
 - Creasterea traficului (nedorit) limiteaza latimea efectiva de banda ceea ce conduce la necesitatea cresterea latimii de banda concomitent cu marirea costurilor
 - Scade viteza de procesare a mail-urilor datorita filtrelor anti-spam care previn livrarea acestor mesaje
- In anul 2002, providerul AOL anunta ca 30% din traficul e-mail este spam



SPAM (2)

- Metode de evitare care nu reprezinta neaparat solutii definitive
 - Sa nu raspunda la mesaje necunoscute care au ca rol doar confirmarea validitatii casutei postale
 - o Mesajul trebuie citit si sters
 - Filtrarea automata a mesajelor
 - o Mesajele tot vor ocupa latime de banda
 - o Trebuie receptionate pentru a instrui filtrul anti-spam



Netiquette

- Norme de eticheta folosite la screirea mesajelor text (mesaje –mail, instant messaging)
 - Mesaje scurte si la obiect
 - Subiectul sa aiba legatura cu mesajului in sine
 - Lungimea unei linii sa nu depaseasca 70-80 de caractere pentru a nu se afisa mesajul in mod ciudat
 - NU FOLOSISTI majuscule, insemna ca strigati
 - Mesajul trimis ca raspuns este bine sa includa numai paragrafele relevante din mesajul original pentru a nu mari artificial corpul mesajului
 - Nu folositi in exces abreviereile si emoticoanele



Abrevieri

- TLA – Three Letter Acronym
 - BTW - by the way
 - FYI - for your information
 - IMHO - in my humble opinion
 - ROTFL - rolling on the floor laughing
 - 2 ¢/\$0.02 - my 2 cents = my small contribution
- Emoticoane – folosite pentru a exprima sentimente altfel exprimate prin limbajul trupului
 - :-) - zambet
 - ;-) - a face cu ochiul
 - :-| - indiferenta
 - :-> - ranjet (sarcasm)
 - :-/ - ramas perplex\\
 - :-(- incruntare (furie sau suparare)

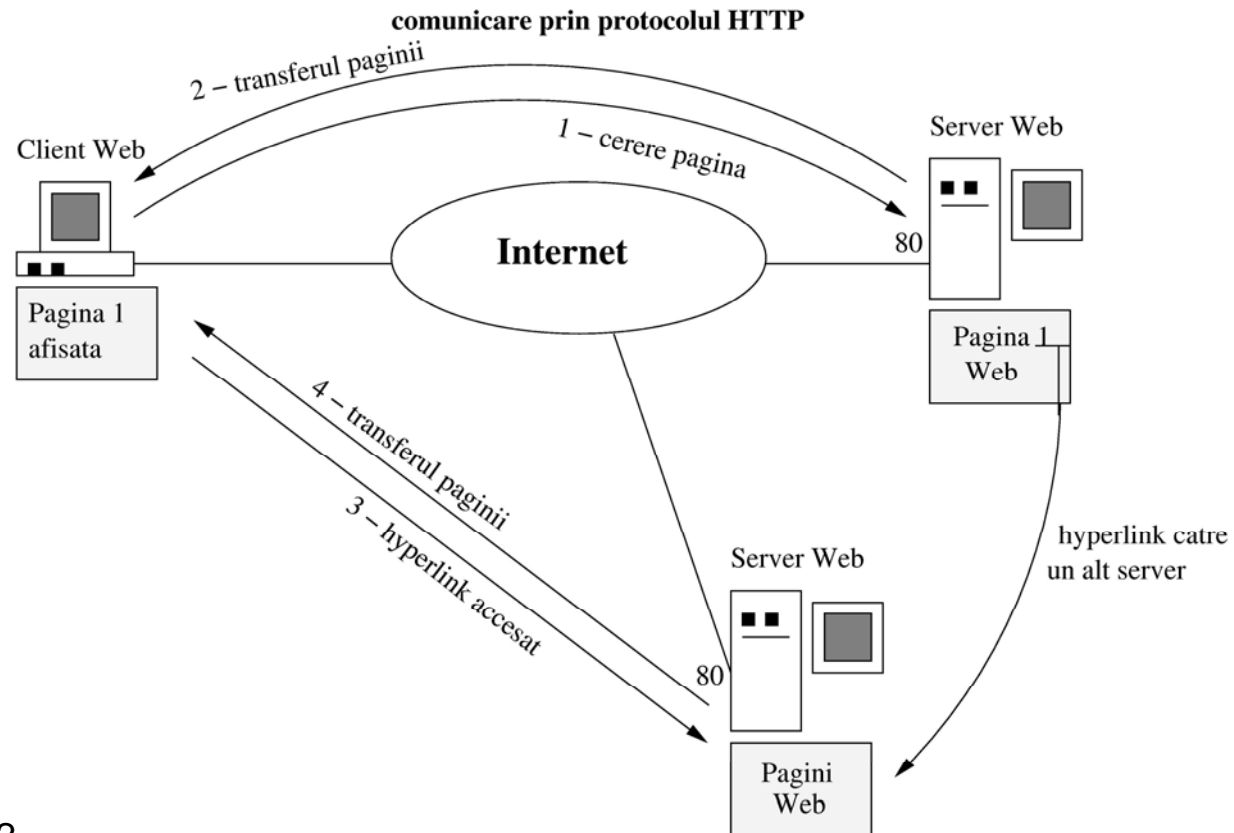


World Wide Web

- Serviciu complet care ofera acces la documente raspindite la nivel mondial:
 - Accesarea informatiei de pe hosturi aflate la distanta in mod interactiv
 - Afisarea de text si imagini
 - Transmisii audio si video in direct
 - Accesarea mai multor servicii folosind un singur mecanism

Notiuni web

- WWW sau W3 a aparut in 1989 la CERN
 - Tim Berners-Lee
- Documentele pot contine legaturi care alte documente - hyperlegaturi





URL

- Functionarea WWW are la baza 2 elemente cheie
 - Fiecare pagina are o adresa unica URL
 - Paginile sint descrise in limbajul HTML (limbaj de marcare a textului) – navigatorul formateaza textul pe baza unor cuvinte cheie
- URL – Uniform Resource Locator
 - Adresa serverului web (nume de domeniu)
 - Adresa paginii in cadrul serverului (calea pina la pagina)
 - Protocolul folosit pentru accesarea paginii
 - `<protocol>://<nume.de.domeniu.server>/<cale catre pagina>/<nume pagina>`
 - Referinta la un director -> referite la o pagina index.html, index.php, index.asp



HTML

- HyperText Markup Language
 - Descrie modalitatea de formatare a textului dintr-o pagina web
 - Versiuna 1.0 era unidirectionala (clientul doar cerea pagini)
 - Versiunea 2.0 adauga interactivitatea
 - User-ul poate transmite informatii suplimentare catre server
- CSS – Cascading Style Sheets – imbunatatesc aspectul unei pagini
 - Proprietati pt culori si fundal
 - Proprietati pt fonturi
 - Proprietati pt text (spatiere, paragraf)
 - Proprietati pt blocuri
 - Clasificari (controlul listelor, etc)

XML

■ eXtensible Markup Language

- Meta limbaj care permite definirea “tag”-urilor pe baza carora se formateaza textul
- Folosita pentru structurarea informatiei in cadrul unei pagini web
- Este o modalitate independenta de produsele SW pentru a transmite si formata datele

```
<nota>
```

```
<to>Ion</to>
```

```
<from>Radu</from>
```

```
<heading>Reminder</heading>
```

```
<body>Nu uita sa-mi trimiti actele pana la week-  
end!</body>
```

```
</nota>
```



Transferul paginilor web

- 1. Navigatorul rezolva numele de domeniu al destinatiei cu ajutorul serverului DNS local
- 2. Navigatorul primeste adresa IP si stabileste conexiune cu serverul web folosind portul 80 sau cel specificat in URL
- 3. Clientul(navigatorul) trimite cererea pt fisierul HTML
 - Daca exista cookie-uri se trimit si acestea
- 4. Serverul transmite fisierul cerut si cookie-urile daca exista
- 5. Serverul web inchide conexiunea
 - Pentru o alta pagina se deschide o alta conexiune
- 6. Navigatorul afiseaza pagina formatat pe baza tag-urilor existente in fisierul HTML

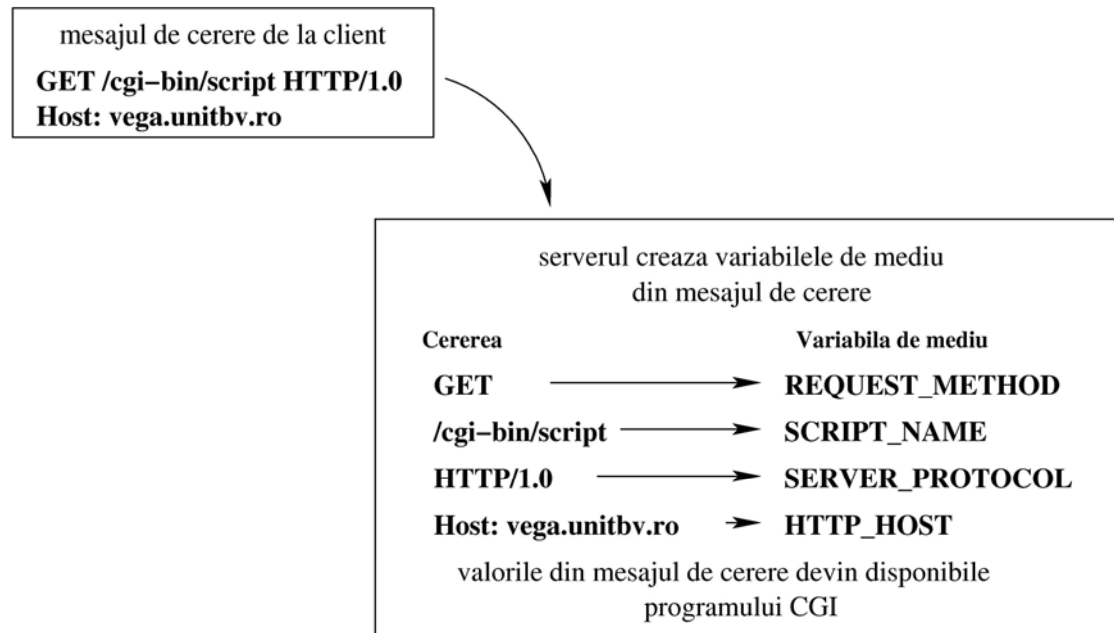


CGI – Common gateway interface

- Pagini web care prezinta interactiune cu user-ul
 - Cererea user-ului este procesata de o aplicatie locala pe server
 - CGI – metoda standard a serverului web de a transmite informatiile spre/dinspre aplicatie
- Cererea trimisa de catre client contine mai multe **campuri**
 - Campul <Request> este singurul obligatoriu
- La primirea cererii serverul efecturaza mai multe operatii
 - Localizeaza in sistemul de fisiere local pagina solicitata
 - Determina daca este necesara autentificare pt acea pagina
 - Decide daca pagine necesita o procesare speciala
 - Rularea unei aplicatii externa – scrip CGI
 - Dupa efectuarea procesarilor speciale se trimite pagina catre client

CGI – Server parsing

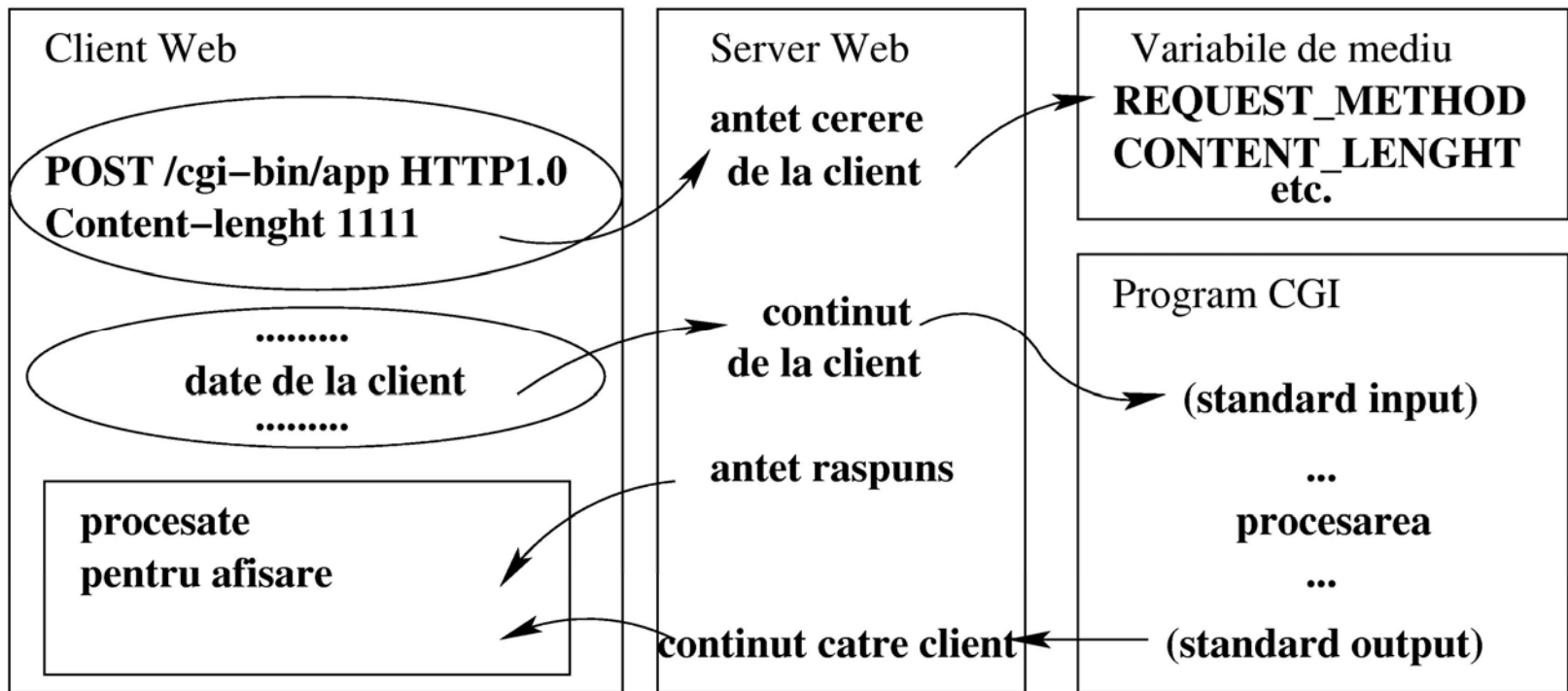
- La primirea cererii serverul initializeaza variabile de mediu cu valorile cimpurilor din cererea clientului
- Argumentele scriptului CGI sint transmise prin variabile QUERY_STRING si sint extrase din URL unde sint depuse de catre navigator:
 - <http://vega.unitbv.ro/cgi-bin/scrip?param>
 - Solutie limitata la 1024 de caractere



CGI – Post method

- Pentru date mai lungi de 1024 de caractere clientul trebuie sa incapsuleze aceste date intr-un anumit format (metoda POST):
 - Antet cu metoda POST
 - Linie goala pt delimitare
 - Datele propriu-zise
- Serverul rezolva cererile prin metoda POST in urmatoorii pasi
 - Determina ca este vorba despre metoda POST (mesajul de cerere)
 - Salveaza datele si determina lungimea datelor
 - Creaza variabilele de mediu (CONTENT_LENGTH)
 - Se trimite datele catre scriptul CGI care le proceseaza si genereaza o pagina web in format HTML
 - Pagina generata este preluata de serverul web si transmisa clientului
 - Serverul inchide conexiunea

CGI – Post method (2)





Aternative CGI

- Scripturile CGI pot fi scrise in C/C++, Perl, TCL, Python
- Ca si alternative exista PHP si ASP de la Microsoft
- ASP- Active Server Pages
 - Limbaj scriptic incapsulat in paginile HTML si procesat de un server web Microsoft (IIS) iar rezultatul este incorporat in pagina web inainte de a fi trimisa utilizatorului
 - Contine componente care pot fi folosite cu limbaje diferite VBScript, Jscript sau PerlScript
- PHP – PHP: Hypertext Procesor
 - Alternativa ASP pentru Linux (poate fi rulat de serverele web Apache)
 - Codul PHP este incorporat in pagina web si este interpretat de serverul web inainte de a genera fimra finala a paginii



JAVA Script

- Limbaj interpretat orientata pe obiect dezvoltat de catre Netscape pentru aplicatii web
 - Asemnator ca structura cu Java
 - A devenit mai popular decat Java
 - Codul este incapsulat in pagina web si este executata de serverul web inainte de a transmite pagina
- JScript un limbaj similar creat de Microsoft
 - Incompatibil cu Netscape
- ECMAScript (European Computer Manufacturer Association) – o versiune creata pentru a unifica JavaScript cu JScript
- Aplicatii unde este folosit cod JavaScript:
 - Continut dinamic (text, imagini)
 - HTML dinamic (modificarea pozitiei elementelor in pagina)
 - Validarea si preluarea datelor dintr-un formular



Limbajul JAVA

- Limbaj de programare orientat pe obiect dezvoltat de catre Sun Microsystems pentru dezvoltarea aplicatiilor Internet
 - Programe independente de sistemul de operare si platforma hardware
 - Applet – aplicatii JAVA incapsulate in pagina web care este rulata de catre clientul web nu de catre server
 - o Permite interactivitate cu userii (jocuri)
 - o Formulare complexe de calcul
 - o Animatii fara necesitatea unor programe externe

Adobe Flash (Macromedia/Shockwave)

- Exista 2 tipuri de imagini folosite pe web
 - Imagini raster – compuse din pixeli de diferite culori, numarul lor care defineste rezolutia imaginii (jpg, bmp)
 - Se pierde din calitate daca se mareste imaginea
 - Imagini vectoriale – compuse din vectorii descrise cu ecuatii matematice (Flash)
 - Independente de rezolutie pot fi marite fara a pierde din calitate
 - Utile in animatii
 - Se bazeaza pe un limbaj scriptic – ActionScript



Portal

- Desemneaza un site web folosit ca si punct principal de pornire pentru utilizatorii care navigheaza, datorita continutului variat de informatie
 - Posibilitatea efectuarii cautarilor web
 - Informatii din domenii variate (sport, finainte)
 - Oferirea ultimelor stiri si a vremii probabile
 - Cautari de adrese si numere de telefon
 - Oferire servicii e-mail
- Oferă posibilitatea parametrizării formei de afişare şi a conţinutului
- Datorita numarului mare de vizitatori reprezinta adevarate centre multimedia folosite in scopuri comerciale



Motoare de cautare

- Asigura modalitatea de regasire a informatiei distribuita in intreg Internetul
- Principiile de baza sint aceleasi insa difera implementariile si implicit relevanta rezultatelor
 - Cautarea informatiilor pe web
 - Indexarea cuvintelor gasite si a contextului in care au fost gasite (relevanta)
 - Re-gasirea rapida in baza de date proprie a informatiei solicitate de client
- Un motor de cautare indexeaza zeci de miliarde de pagini si raspunde zilnic la zeci de milioane de cereri
 - 20 miliarde de pagini – Google
 - 150 milioane de cereri zilnice - Google

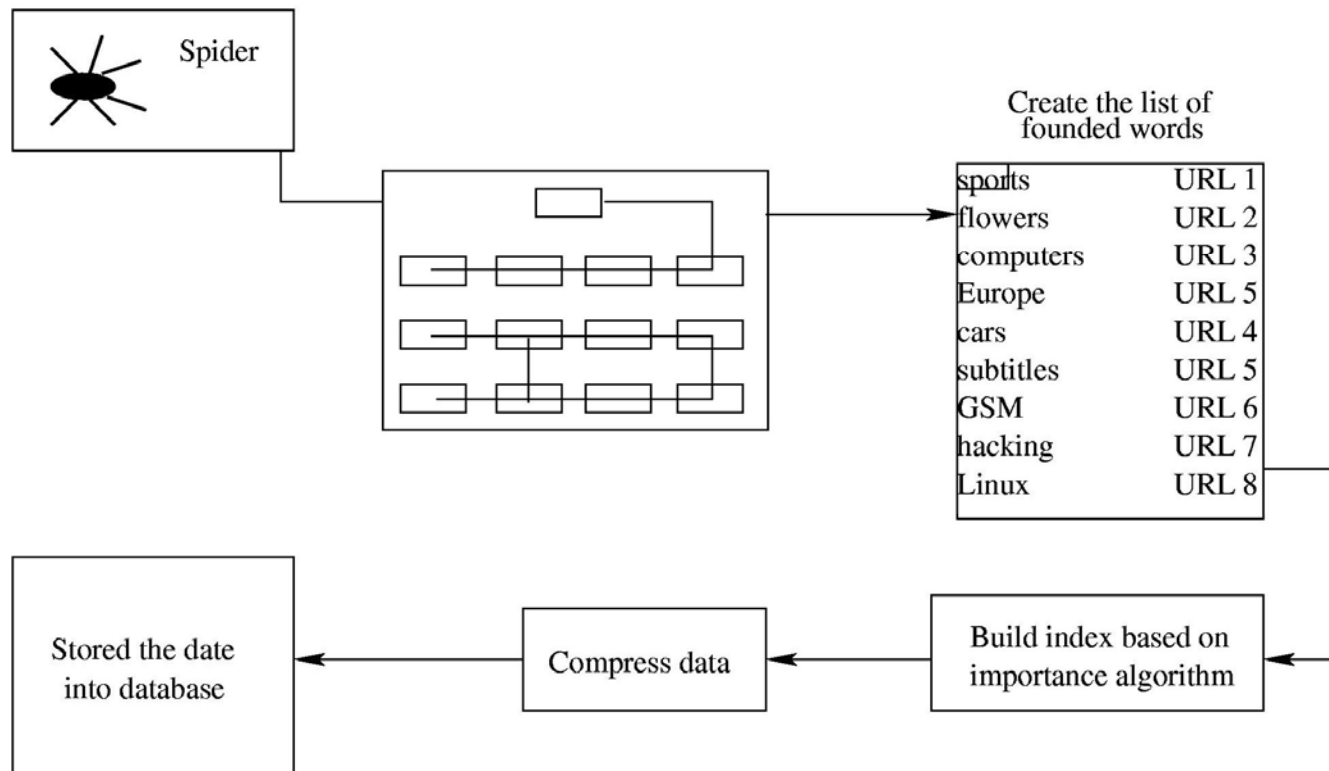


Web crawling

- Procedeeul de adunare a informatiilor relevante din cadrul paginilor web
 - Realizate de catre programe specializate (spider)
 - Pornesc de la o baza de date proprie si continua cu link-urile pe care le gasesc in paginile pe care le navigheaza
 - Se elimina cuvintele irelevante (prepozitii, conjunctii)
 - Re-vizitarea paginilor se face periodic; paginile care conti informatii cerute mai des sint vizitate mai des
 - Tag-urile META prezinta spider-ului cuvintele cheie ale paginilor
 - o Folosit original pentru indexare
 - o Exista posibilitatea excluderii robotului de cautare

Modalitatea de functionare

- Accesibilitatea datelor de catre utilizatori
 - Informatia propriu-zisa care este cautata
 - Modalitatea de indexare a informatiei





Indexarea informatiei

- Este important sa se determine relevanta rezultatului
 - Numarul de cuvintele si asezarea lor in cadrul paginii
 - Numarul de link-uril relevante, etc
 - Modalitatiile de determinare a relevantei unei informatii sint tinute secrete de catre companii
- Informatia trebuie stocata intr-o forma care permite o regasire rapida
 - Hash table sau functii de dispersie – impartirea informatiei in mode egal intr-un numar predeterminat de sectiuni
- Cautarea se realizeaza folosind operatori binari (AND, OR, +, -)
 - Cautare literala, nu dupa sensul expresiei
 - Imbunatatirea rezulotatelor prin cautare pe baza de concept



Cookies

- Un sir de caractere stocat pe hard-diskul clientului pentru a usura navigarea
 - Pereche nume-valoare
 - Se transmite serverului in momentul conectarii la acesta
 - In Windows sint stocate in directorul Cookies
- Sint folosite pentru
 - Identificarea numarului de vizite ale unui user
 - Stocarea anumite informatii specifica unui anume utilizator (produse cumparate, parametrizare paginii de start, continutul cosului de cumparaturi, etc)
 - Realizarea unui profil al utilizatorului in materie de cumparaturi pe baza paginilor vizitate
 - o Ridica problem etice intrucit culegerea de informatii se face fara acordul si stirea user-ului
 - o Serviciul DoubleClick furnizeaza profile de utilizatori diferitor site-uri pentru a oferi doar informatia relevanta pentru acel utilizator