

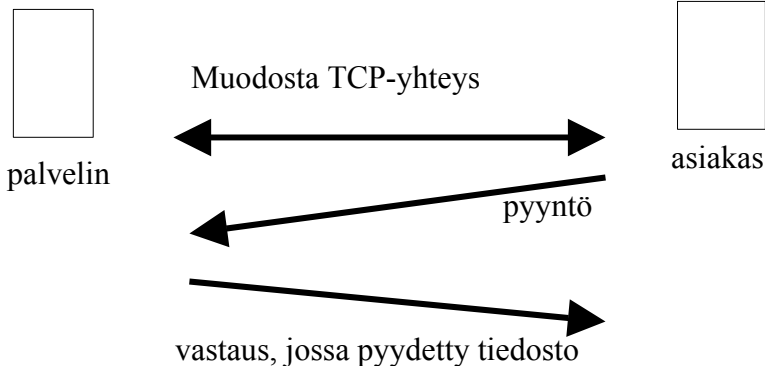
2.1. WWW

- **WWW on joukko yhteentoimivia palvelimia ja asiakkaita, jotka puhuvat samaa HTTP-kieltä (-protokollaa)**
- **graafinen asiakasohjelma selain**
 - Netscape, Explorer
 - kykenee ‘avaamaan ‘ URL-linkin**
 - = muodostamaan TCP-yhteyden ko. verkkopalvelimeen ja hakemaan sieltä ko. tiedoston

<http://www.cs.princeton.edu/index.html>

Muodostetaan TCP- yhteys koneeseen:
www.cs.princeton.edu

Pyydetään sieltä tiedosto käyttäen HTTP-protokollaa:
[index.html](http://www.cs.princeton.edu/index.html)



HTTP (Hypertext Transfer Protocol)

- **Protokolla asiakkaan ja palvelimen väliseen kommunikointiin**

- **www-sivujen hakemiseen**

- **HTML-tiedosto, jossa**
 - HTML-tiedostoja
 - JPEG-kuva, GIF-kuva
 - Java-sovelma

- **sivuilla URL-osoite**

- **tilaton protokolla**

Tekstiä:

plaa plaa

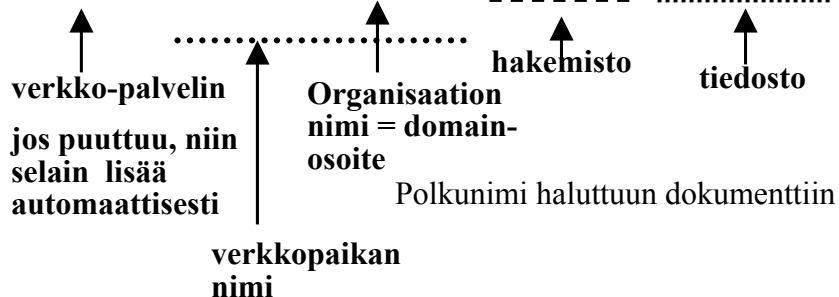
plaa



11/13/2002

URL (Uniform Resource Locator)

- **<http://www.nmib.com/glossary.index.html>**



11/13/2002

16

Muita URLeja:

- **ftp://usc.edu/pubs/myfile.doc**
 - ftp-palvelulla haettava tiedosto
- **news:uk.finance**
 - uutisryhmä
- **file:///C:/webs/html/mottle.gif**
 - tiedosto haetaan käyttöjärjestelmän avulla
(ei käyttäen http:tä)

HTTP-asiakas

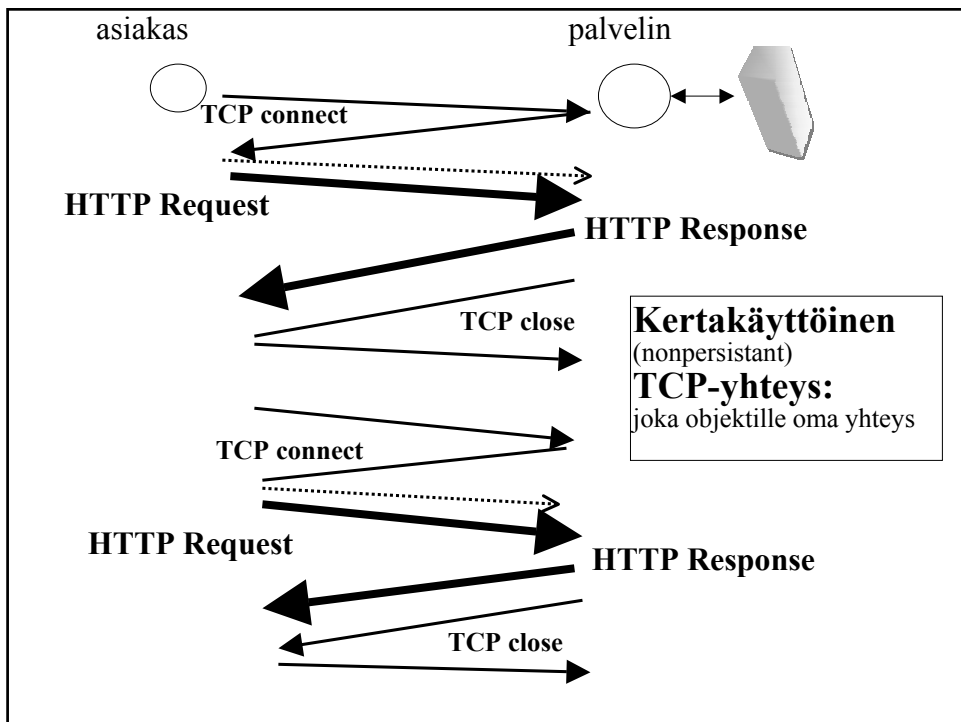
- **selain (= asiakas) pyytää verkkosivua**
 - muodostaa TCP-yhteyden palvelinkoneeseen
 - DNS auttaa domain-nimen muuttamisessa IP-osoitteeksi
 - WWW-palvelimen kuuntelema portti on 80
 - lähettää HTTP-pyyntösanoman TCP-yhteyteen liitettyyn pistokkeeseen
 - vastaanottaa palvelimen lähettämän **vastaussanoman**, jossa on sisällä pyydetty sivu
 - tämän jälkeen TCP-yhteys suljetaan
 - selain tutkii sivua ja näkee uudet viitteet kuviin, jotka se sitten hakee samalla tavalla
- **selain näyttää pyydetyn sivun käyttäjälle**

Palvelinprosessi

- **Kuuntelee TCP-porttia 80 yhteydenpyyntöjä varten**
 - pyyntö yleensä selaimelta
 - kun käyttäjä on ilmoittanut haluavansa ko. dokumentin
 - esim. klikkaamalla tekstissä olevaa URL-linkkiä
 - tai antamalla URL-linkin yhteysoyennössä
 - tai selain itse huomaa linkin, josta pitää hakea

HTTP-pyyntön tullessa

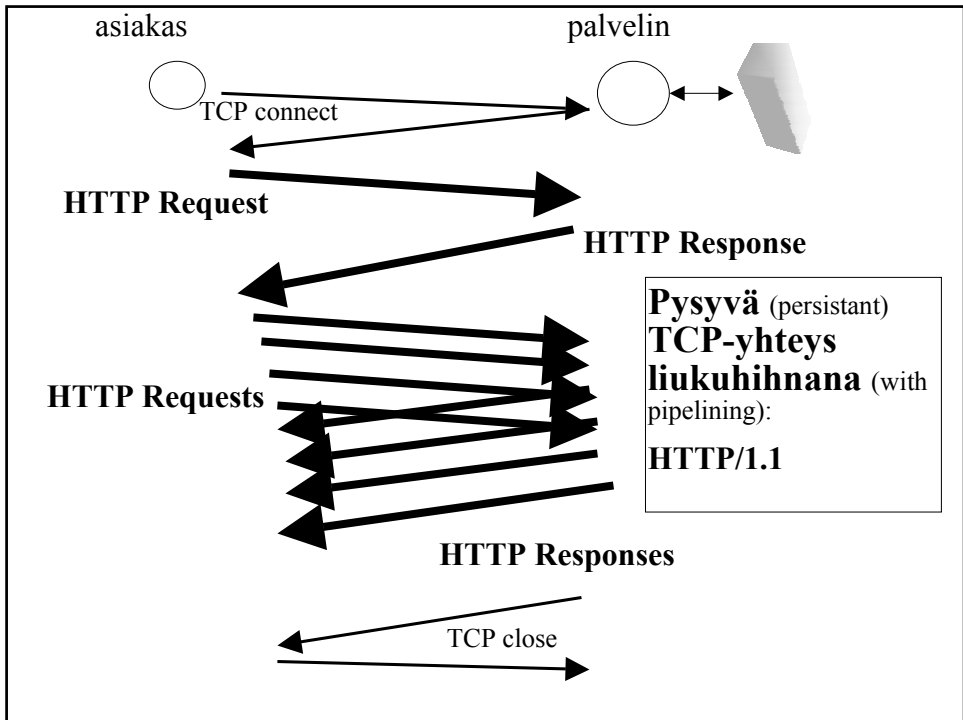
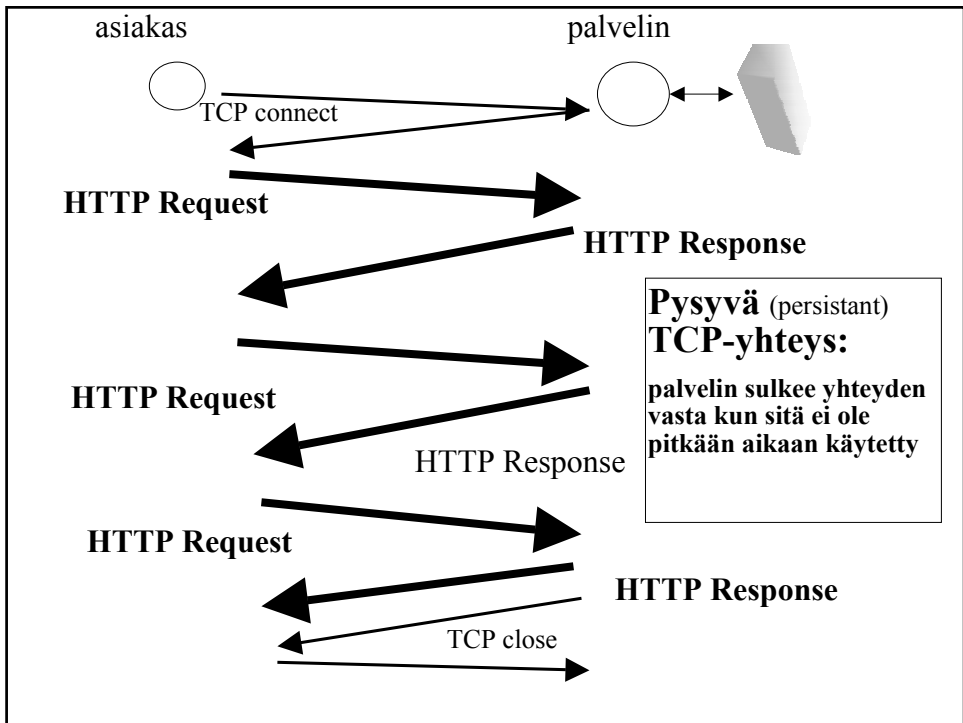
- **verkkopalvelin**
 - vastaanottaa pyyntösanoman ,
 - hakee pyydetyn sivun tai objektin omasta muististaan ,
 - liittää sen HTTP-vastauksensaamaan ja
 - lähettää sitä pyytäneelle selaimelle
 - pyytää TCP-yhteyden sulkemista



-
-
-

Suorituskyky?

- **Jos haetaan 10 objektia**
 - 10 TCP-yhteyden muodostusta ja purkua
 - kukin 2 sanomaa = RTT (round-trip time) => 20 RTT
 - hidas aloitus (slow start) hidastaa lähetysnopeutta
 - voidaan avata useita rinnakkaisia yhteyksiä
 - puskuritilat yhteyksille
- **käytetään pysyvää TCP-yhteyttä**
 - palvelin jättää yhteyden sulkematta
 - muut pyynnöt ja vastaukset käyttävät samaa yhteyttä

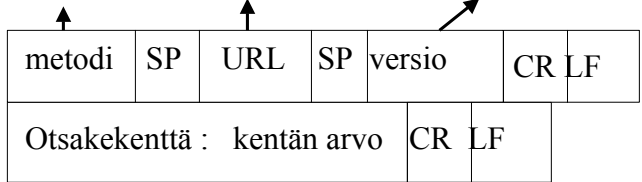


HTTP-sanomat (esimerkki)

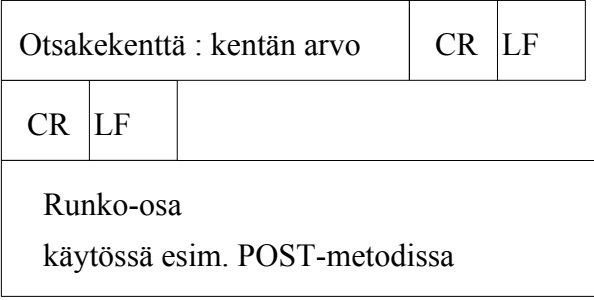
- HTTP Request Message:**
GET /jokuhakemisto/sivu.html HTTP/1.1
Host: WWW.jokupaikka.fi
Connection: close
User-agent: Mozilla/4.0
Accept-language:fi
(ylimääräinen CR ja LF)

Pyyntösanomien yleinen rakenne

GET /jokuhakemisto/sivu.html HTTP/1.1



... Lisää otsakerivejä



Otsakekenttä : kentän arvo	CR	LF
-----------------------------------	----	----

Host: WWW.jokupaikka.fi kone, jossa dokumentti on

Connection: close sulje yhteys lähetyksen jälkeen

User-agent: Mozilla/4.0 selaimen tyyppi

Accept-language:fi dokumentin kieli

○
○
○

HTTP-sanomat (esimerkki)

- **HTTP Response Message:**

HTTP/1.1 200 OK

Connection: close

Date: Thu, 19 Oct 2000 12:00:15 GMT

Server: Apache/1.3.0 (Unix)

Last-Modified: Mon, 22 Jun 2000 09:23:24 GMT

Content-Length: 6821

Content-Type: text/html

(data data data data data ...)

Vastaussanomien yleinen rakenne

versio	SP	statuskoodi	SP	fraasi	CR	LF
Otsakekenttä : kentän arvo				CR	LF	

... Lisää otsakerivejä

Otsakekenttä : kentän arvo		CR	LF
CR	LF		
Runko-osa			


Pyyntömetodeja

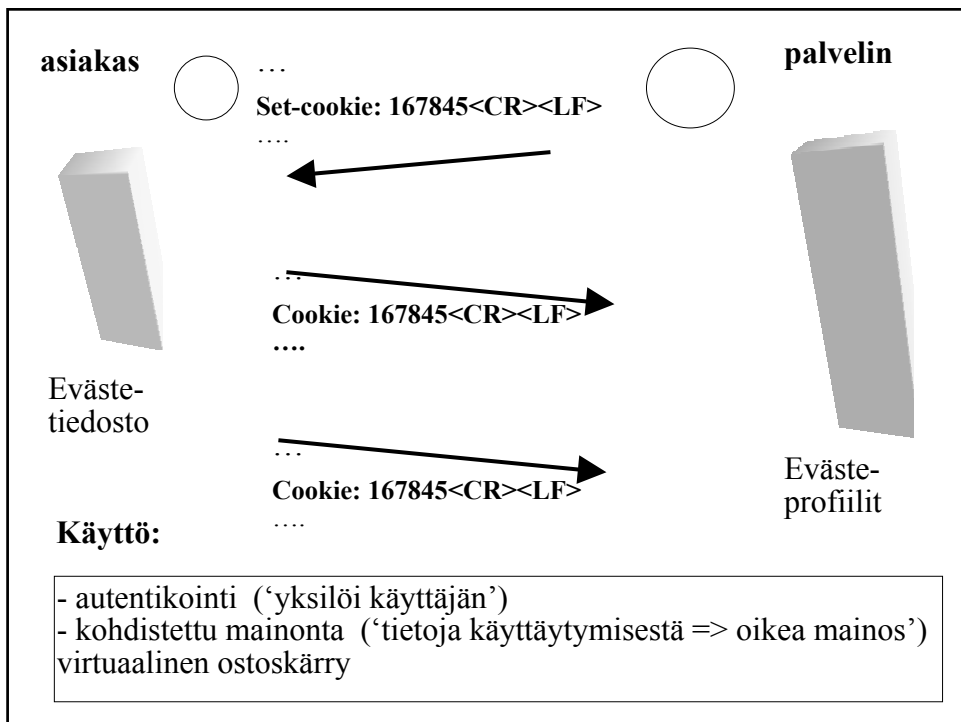
- **GET** sivun lukeminen
 - aina tai vain tietyn päiväyksen jälkeen muutettu sivu
- **HEAD** pelkän otsikon lukeminen
- **PUT** sivun tallettaminen
- **POST** lisäys verkkosivulle
- **DELETE** sivun poistaminen
- **LINK** uusi linkki
- **UNLINK** linkin poistaminen

Statuskoodeja ja fraaseja

- **200 OK**
 - pyyntö onnistui , pyydetty sivu vastauksessa
- **301 Moved Permanently**
 - uusi URL on otsakekentässä Location
- **400 Bad Request**
- **404 Not Found**
- **505 HTTP Version Not Supported**

Käyttäjän tunnistaminen

- **Autentikointi**
 - omat koodit ja otsakkeet tätä varten
 - 401 Authorization Required -statuskoodi
 - WWW-Authenticate -otsake kertoo kuinka**
 - asiakas lähettää autentikointitiedot joka pyynnössä (Authorization- otsakerivi)
- **Eväste (pipari, cookie)** 
 - asiakkaalle talteen käyttäjäkohtainen tunnuskoodi, joka esitetään joka pyynnön yhteydessä



Evästeitä: (Windows: cookies.txt)

www.download.com lähettäjän osoite
FALSE / FALSE turvayhteyksien käyttö
946598400 vanhenemispäivä
csr 'piparin' nimi
/=/PC/FrontDoor/0^1^0- 0^01.html data

www.harmony-central.com **FALSE / FALSE**
1034625738 s 13013000876945738607

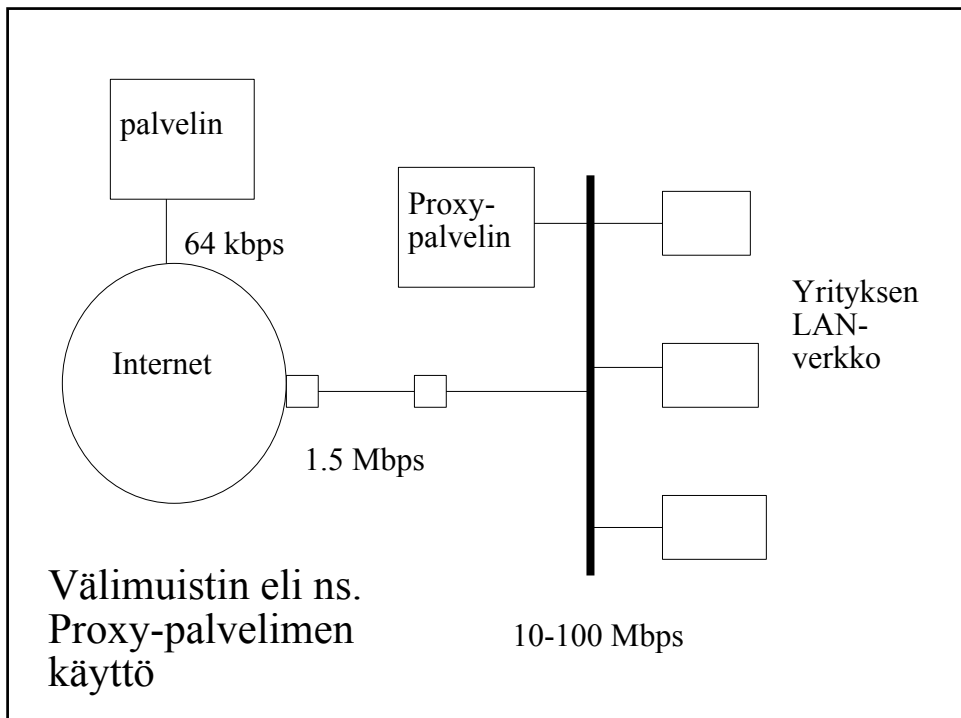
www.illuminatus.com **TRUE / FALSE**
945734399 Count 3

Verkkovälimuisti

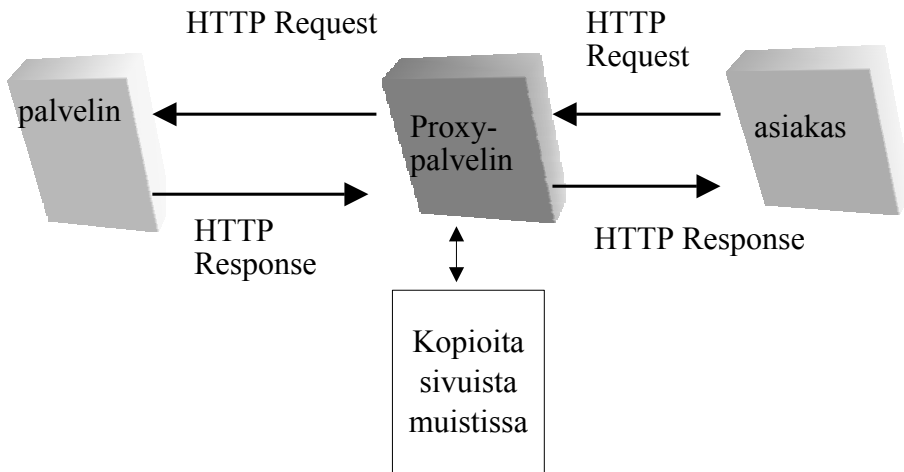
- **Säilyttää kopioita haetuista sivuista**
 - viimeksi haetut sivut
- **pyyntö ohjautuu ensin välimuistiin**
 - haetaan verkon yli vasta, jos ei löydy välimuistista
- **etuja:**
 - lyhentää kyselyaikaa
 - vähentää Internetin liikennettä
 - tuo nopean jakelukanavan hitaille palvelimille

11/13/2002

35



Haku palvelimelta vain, jos ei ole jo muistissa



Conditional GET

- **Välimuistiin talletettu objekti haetaan verkosta, jos objektia on muutettu**
 - tallessa vanha versio
- **Get**
 - If-modified-since: Mon, 22 Jun 2000 09:23:24**
- **HTTP/1.0 304 Not Modified**

HTML (HyperText Markup Language)

- **SGML (Standard Generalized Markup Language)**
 - merkkäuskieli
 - kertoo, kuinka dokumentit muotoillaan
 - TeX, troff,
 - ladontamerkinnot
 - WYSIWYG
- **XML (Extensible Markup Language)**
- **Näistä enemmän:**

582302 Digitaalisen median tekniikat (2ov)

582304 XML-metakieli (2ov)

11/13/2002

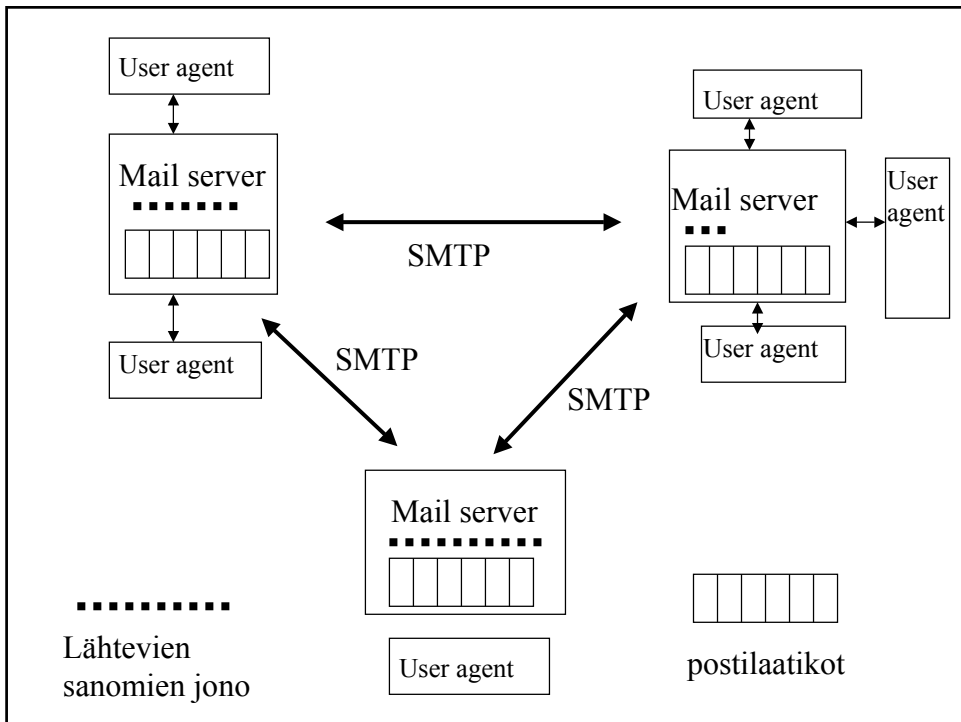
39

2.2. Sähköposti

- **Käyttäjäliitäntä (user agent)**
 - sanomien kirjoittaminen, lukeminen ja lähettäminen
- **Postipalvelin (mail server)**
 - postilaatikat
- **Postiprotokolla SMTP**
 - sanomien toimittaminen lähettäjän postijärjestelmästä vastaanottajan postijärjestelmään
 - sanoman muoto
 - tekstisanoma, MIME-sanoma

11/13/2002

40

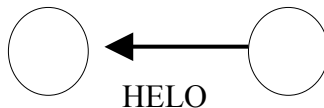


SMTP (Simple Mail Transfer Protocol)

- **Postipalvelimet käyttävät SMTP-protokollaa sanomien siirtoon**

- sanomat lähetävä postipalvelu muodostaa TCP-yhteyden vastaanottajan postipalvelimeen
- 'kättelyjen' jälkeen siirretään sanomat

Vastaanottava on palvelin



Lähetävä on asiakas

Lähetävä postipalvelin

- muodostaa TCP-yhteyden
- esittelee itsensä
- kertoo lähettäjän ja vastaanottajan
- lähettää sanoman, jos vastaanottava palvelin antaa luvan

Vastaanottava postipalvelin

- kuuntelee TCP-porttia **25**
- hyväksyy yhteydenmuodostuspyynnöt
- vastaanottaa ja siirtää sanomat postilaatikoihin
- palauttaa virheilmoitukset

```
S: 220 helsinki.fi
C: HELO princeton.edu
S: 250 Hello princeton.edu
C: MAIL FROM: <Bob@princeton.edu>
S: 250 <Bob@princeton.edu> OK
C: RCPT TO: <pekka.puupaa@cs.helsinki.fi>
S: 250 <pekka.puupaa@cs.helsinki.fi> OK
C: DATA
S: 354 Enter mail, end with "." on a line by itself
C: dataa ... dataa
C: dataa ... dataa

C: .
S: 250 Message accepted for delivery
C: QUIT
S: 221 princeton.edu closing connection
```


○
○
○

- **TCP-yhteys on pysyvä**

- yhteydellä voidaan lähettää useita sanomia

- **Yhteydellä voidaan lähettää vain ASCII-sanomia**

- sanomatekstin lopettaa omalla rivillään irrallisena oleva piste (“.”) (*Tämä on vanha protokolla!*)

- eli ASCII-merkit **CRLF.CRLF**

- jos sanomassa on binääridataa, esim. kuvia, niin siellä voisi esiintyä irrallinen piste eli CRLF.CRLF => sanoma katkeaa. => binääridata koodataan ASCIIksi siten, ettei siinä esiinny CRLF.CRLF : iä

11/13/2002

45

○ ○ ○ ○ ○ ○ ○ ○ ○ ○

S: 220 helsinki.fi

C: HELO princeton.edu

S: 250 Hello princeton.edu

C: MAIL FROM: <Bob@princeton.edu>

.....

S: 250 Message accepted for delivery

C: HELO princeton.edu

S: 250 Hello princeton.edu

C: MAIL FROM: <Alice@princeton.edu>

.....

S: 250 Message accepted for delivery

.....

C: QUIT

S: 221 princeton.edu closing connection

Push- ja pull-protokollat

- **työntöprotokolla (PUSH)**
 - asiakas lähettää tiedoston
 - SMTP
- **noutoprotokolla (PULL)**
 - asiakas pyytää tiedostoa
 - HTTP
 - poikkeuksena PUT- ja POST-metodit

11/13/2002

47

Sähköpostisanoman muoto

To:
CC:
From:
Message-Id:
Received:
Date:
Reply-To:
Subject:

Yleisiä sanoman
otsakekenttiä

kentät erotettu
rivinvaihdolla

RFC 822 sallii myös omien
kenttien käytön

11/13/2002

48

MIME (Multipurpose Internet Mail Extension)

- **Alunperin vain tekstisanomia (ASCII)**
- **nykyisin myös**
 - erityyppisiä tekstisanomia
 - Word-dokumentteja
 - kiinankielistä tekstiä
 - ääntä,
 - videota
 - kuvia

MIME-laajennus

- **Lisää kenttiä otsikkoon**

MIME-versio:
Content-Transfer-Encoding:
Content-Type:

From: alice@crepes.fr
To: bob@hamburger.edu
Subject: Picture of yummy crepe
MIME-Version: 1.0
Content-Transfer-Encoding: base64
Content-Type: image/jpeg
(base64 encoded data
..... base64 encoded data)
.

◦
◦
◦

MIME-sisältötyyppiä

- text/plain; charset=us-ascii
- text/html
- image/gif, image/jpeg, video/mpeg
- application/postscript,
application/msword,
application/octetstream
- **multipart/mixed**

Content-Type:
multipart/mixed;Boundary=StartOfNextPart

-- StartOfNextPart

Hei Allu,

sinulle kaunis kuva kissastani Villestä.

-- StartOfNextPart

Content-Transfer-Encoding: base64

Content-Type: image/jpeg

base64 encoded data

.....

.....base64 encoded data

-- StartOfNextPart

Haluatko muita kuvia!

.

o
o
o



- **tyyppien koodaukset siirtoa varten**

- sähköpostisanomassa saa olla vain ASCII-merkkejä

- ei esimerkiksi ääkkösiä

- **base64-koodaus**

Base64-koodaus

- **Sanoman 24 bitin ryhmät jaetaan 6 bitin osiksi, jotka kukin koodataan ASCII-merkiksi => 64 eri vaihtoehtoa**

– 0-25 = A-Z,

– 26-51 = a-z,

– 52-61 = 0-9,

– 62= +, 63 =/

- **Tekstisanoma voidaan koodata myös 7-bitin ASCIIlla.**

11/13/2002

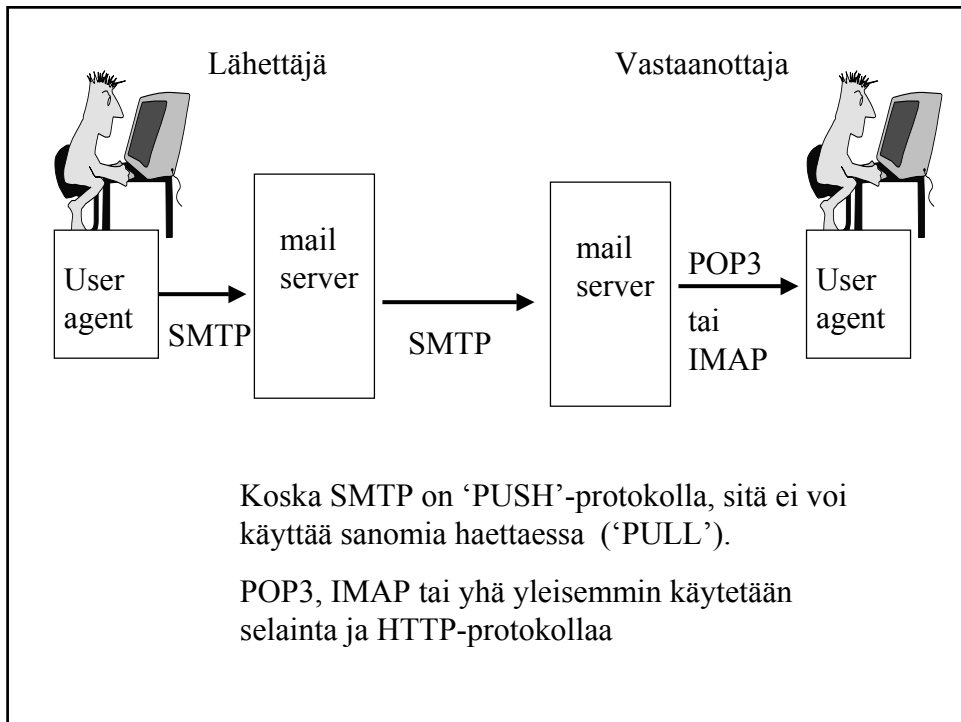
55

Postinhakuprotokollat (mail access protocols)

- **Käyttäjän postiohjelma siirtää postinhakuprotokollan avulla sanomat postipalvelimen postilaatikoista käyttäjän PC:lle**
 - POP3: yksinkertainen ohjelma, joka siirtää sanomat vastaanottajan PC:lle
 - IMAP: monipuolinen, käyttäjä voi käsitellä postipalvelimen postilaatikkoo kuin se olisi omassa koneessa
- **Entä lähetettäessä?**

11/13/2002

56



-
-
-

Postiyhdyskäytävä (Mail Gateway)

- **Yleensä postipalvelin ottaa suoraan yhteyttä vastaanottajan postipalvelimeen**
- **välissä voi olla postiyhdyskäytäviä**
 - tallettavat ja toimittavat eteenpäin sanomia
 - yrittävät lähettämistä useita päiviä
 - **Miksi tarvitaan?**
 - ohjaa oikeaan koneeseen: liisa.marttinen@helsinki.fi
 - vastaanottajan kone usein pois päältä

2.3. DNS (Domain Name System)

- **Internetin hakemistopalvelu**
- Ihmisten ja ohjelmien käyttämien merkkijonomuotoisten nimien muuttaminen binäärisiksi IP-verkko-osoitteiksi
 - **hydra.cs.helsinki.fi => 218.214.4.29**
- hierarkkinen nimentätapa + hajautettu hakemisto IP-osoitteiden selvittämiseksi
 - **esim. sähköpostiosoitteita, koneiden IP-osoitteita**

11/13/2002

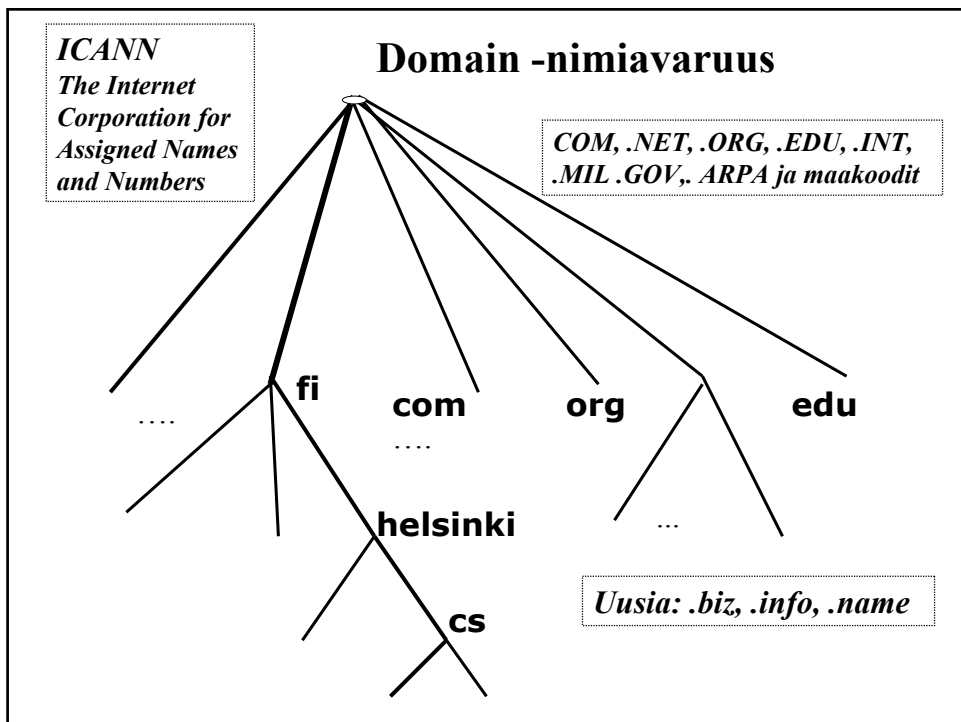
59

DNS- nimiavaruus

- **Domain (nimihallinta-alue)**
 - maiden tunnukset
 - fi, fr, de, uk, ...
 - yleiset (yleensä USA)
 - com, edu, gov, int, mil, net, org
 - esimerkkejä:
 - cs.helsinki.fi
 - www.cnn.com
 - www-dept.cs.ucl.ac.u

11/13/2002

60



- o
- o
- o

IP-nimen selvittäminen

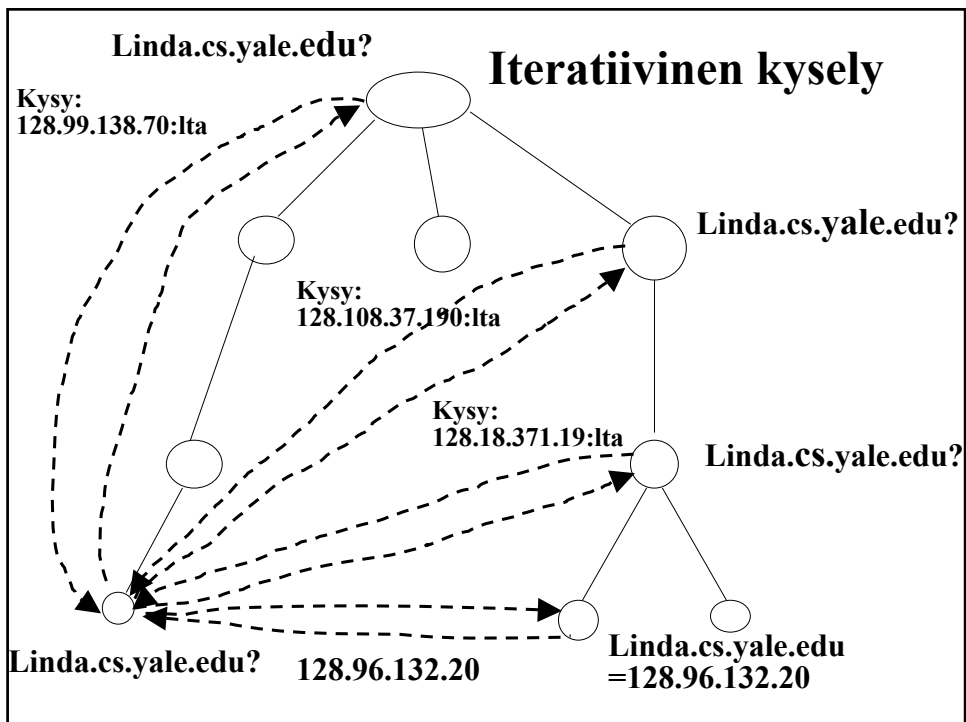
- **sovellusohjelma kutsuu kirjastorutiinia parametrina nimi merkkijonona**
 - esim Unix:ssa gethostbyname()
- **kirjastorutiini lähettää UDP-datasähkeen paikalliselle DNS-palvelimelle, joka etsii nimeä vastaavan IP-osoitteen ja palauttaa sen kirjastorutiinille**
 - etsinnässä tarvitaan usein monien palvelimien yhteistyötä

Nimipalvelimien hierarkia

- **DNS-nimiavaruus jaettu vyöhykkeisiin (zone)**
 - kukin vyöhyke kattaa osan nimipuusta
 - vyöhykkeellä on yksi siitä vastaava nimipalvelija (primary) ja yksi tai useita apunimipalvelijoita (secondary)
- **vyöhykejako on hallinnollinen**
 - tarpeen mukaan nimipalvelijoita vastaamaan omasta alueestaan

• Kysely ensin paikalliselle nimipalvelimelle

- jos nimi on sen vastuulla => **authoritative record**
- jos sillä ei ole mitään tietoa nimestä, se lähettää kyselyn ylimmän tason ns. juurinimipalvelijalle (DNS root server,
<http://www.wia.org/pub/rooterv.html>,
<ftp://ftp.rs.internic.net/domain/named.root>)
 - esim. Kysely "linda.cs.yale.edu" lähetetään nimipalvelimelle "edu-server.net", joka vastaa vyöhykkeestä "edu".



○
○
○

Hakemistotiedot

- **DNS-hakemistotietue (resource record)**

- (nimi, arvo, tyyppi, tietueen elinaika)
- Tyyppi määrää nimen ja arvon merkityksen
 - **A-tyyppi**
 - koneen domain nimi , arvo on IP-osoite
 - **NS-tyyppi**
 - domain-nimi, arvo on alueesta vastaavan nimipalvelijan koneen domain-nimi
 - **CNAME-tyyppi**
 - alias-nimeä vastaava 'kanoninen' nimi
 - **MX-tyyppi**
 - aliasnimeä vastaava postipalvelin

DNS-sanomat

Identification	Flags
Number of questions	Number of answers
Number of authority RRs	Number of additional RRs
Questions (name and types for queries)	
Answers (RRs in response to queries)	
Authority (RRs for authoritative servers)	
Additional information	

11/13/2002

69

DNS-välimuisti

- **Suorituskyvyn parantamiseksi nimipalvelijat varastoivat välimuistiinsa DNS-tietueita**
- **ei tarvitse hakea uudestaan**
- **elinaika kertoo voimassaoloajan**

11/13/2002

70