Lecture 2.

Internet:
who talks with whom?

An application layer view,
with particular attention to the
World Wide Web

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

Basic scenario

Client
(local PC)

Internet

Server
(remote host)
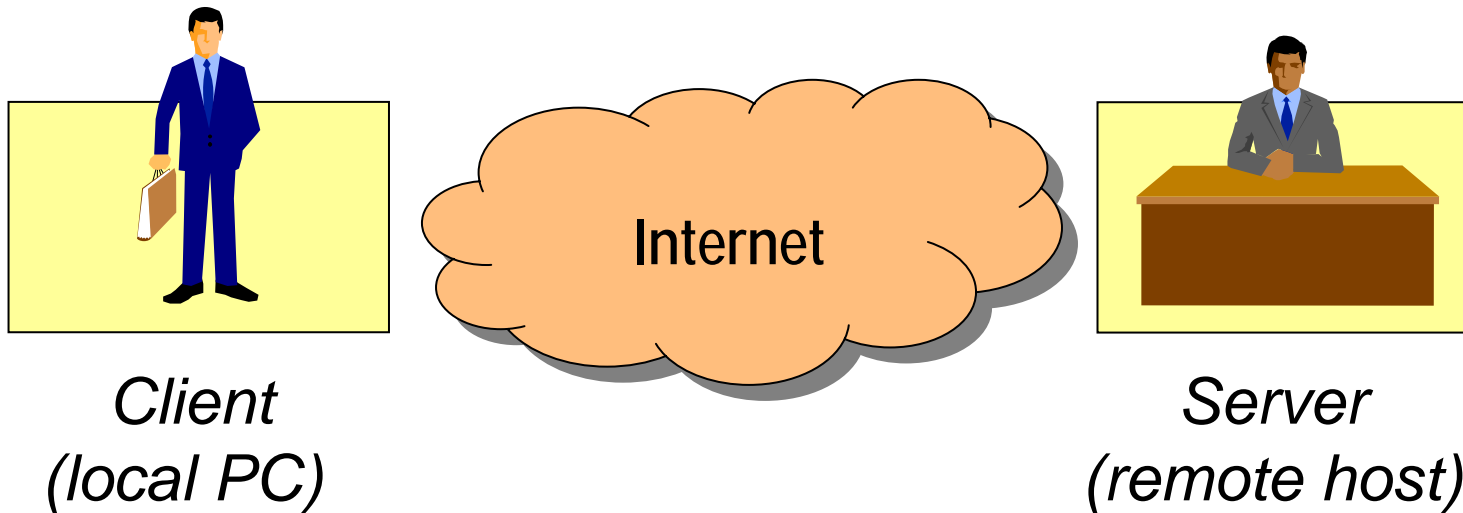
Client wants to retrieve a web page. What happens?

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

What is a "page" on the web?

a resource (i.e. a file), specified by a
URL: Uniform Resource Locator.

e.g. my home page:

HTTP://www.tti.unipa.it/~vincenzo/index.html

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

The three components of an URL

1. ***Protocol (also called "scheme")***
   - ⇨ *how can a page be accessed?* (application protocol used)
     - → **http**://cerbero.elet.polimi.it/people/bianchi/index.html

2. ***Host name***
   - ⇨ *Where is the page located?* (symbolic or numeric location)
     - → http://**cerbero.elet.polimi.it**/people/bianchi/index.html

3. ***File (resource) name***
   - ⇨ *What is the page called?* (with full path)
     - → http://cerbero.elet.polimi.it**/people/bianchi/index.html**

Università degli Studi di Palermo

1. Protocol

➔ **HTTP: the protocol of the WWW**

➔ version 1.0, RFC 1945, may 1996

➔ version 1.1, RFC 2068 (jan97), obsoleted by RFC 2616 (jun99), updated by RFC 2817 (may00)

➔ **many other protocols may be specified on the browser command line:**

⇨ **FTP**: file transfer protocol

⇨ **TELNET**: opens a telnet window

⇨ **FILE**: access local file

⇨ **NEWS**: Usenet Newsgroups

⇨ ...etc...

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

2. Location – host name

Specifies where is the page located:

⇨ on which host

→ Humans understand names;

→ Machines prefer numbers!

→ Domain Name System (DNS) protocol:

» translates names in numbers

www.elet.polimi.it

DNS

131.175.120.33

Università degli Studi di Palermo

3. File names
(several shortcuts handled server side)

page name non mandatory

> http://cerbero.elet.polimi.it/people/bianchi/
> http://www.yahoo.com
> http://www.sun.com/products

Unix style user id shortcuts

> http://www.cs.columbia.edu/~hgs

Various page extensions (with server side meaning:
no client side interpretation!)

> *.htm *.html *.asp *.jsp *.php ……..

Università degli Studi di Palermo

Case sensitivity

➔ **Protocol: non case-sensitive**

➔ **Location: non case-sensitive**

➔ **File name: case-sensitive**

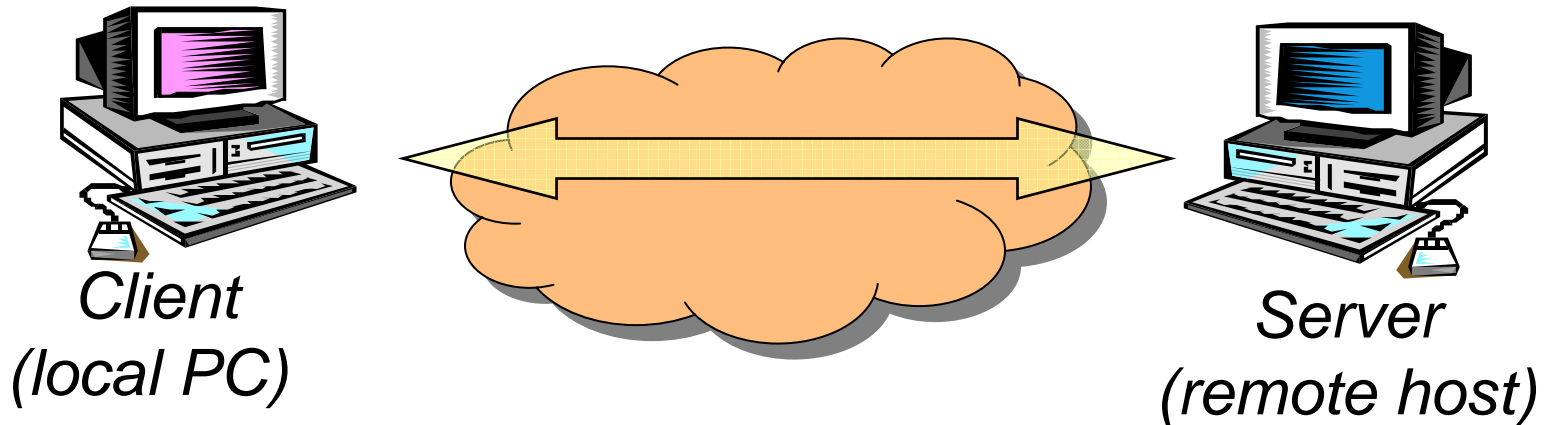http://cerbero.elet.polimi.it/people/bianchi/index.html

equal to

Http://Cerbero.Elet.PoliMi.it/people/bianchi/index.html

different from

http://cerbero.elet.polimi.it/people/bianchi/Index.html

Università degli Studi di Palermo

Refined scenario

Client
(local PC)

Server
(remote host)

Client wants to retrieve a remote resource:

/people/bianchi/index.html

On the server

cerbero.elet.polimi.it

By using the application layer protocol

http

What happens?

Università degli Studi di Palermo

Being more precise

➔ **Who is the "client"?**
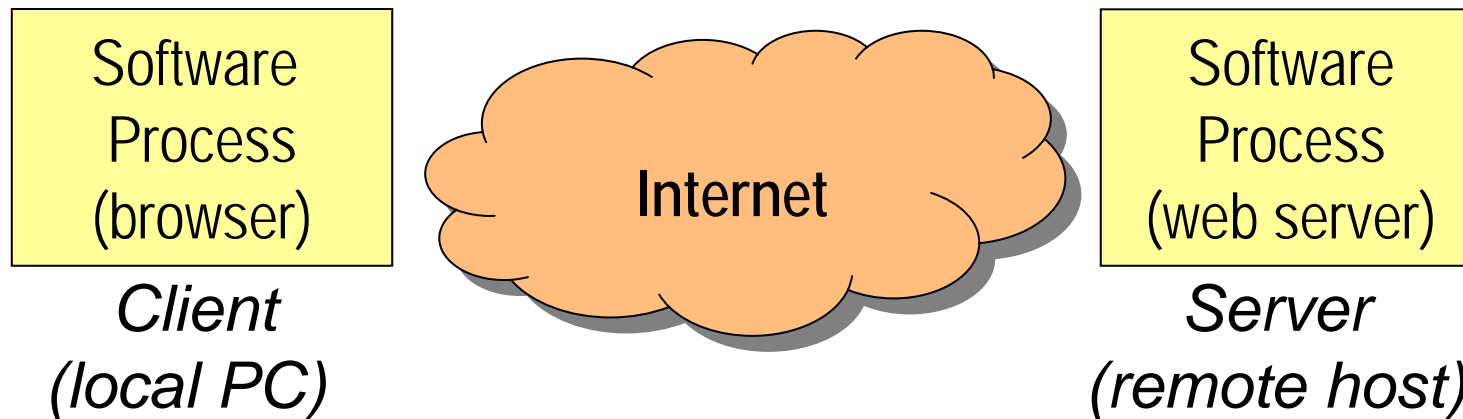 ⇨ The web browser
➔ **Who is the server?**
  ⇨ The web server
➔ **What is the Networking application?**
  ⇨ The WWW
➔ **What is the application layer protocol?**
  ⇨ HTTP

A more precise communication model:

Software Process (browser)	Internet	Software Process (web server)
Client (local PC)		*Server (remote host)*

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

Networking Application
vs
Application Layer Protocol

➔ **Networking application**

⇨ software processes

⇨ on possibly different end-systems

⇨ that communicate each other by exchanging messages

➔ **Application Layer Protocol**

⇨ define format of messages

⇨ define order of messages exchanged

⇨ define actions taken on receipt of a message

An application-layer protocol: is only one piece (although a big piece!!) of a network application

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

What is a protocol

➔ **A protocol defines the format and the order of messages exchange among some entities (two or more), as well as the actions related to the transmission and reception of such messages *(Jim Kurose)***

G.Bianchi, G.Neglia, V.Mancuso

Example: the WWW application
many components, including:

⇨ standard for document formats

→ HTML & HTML interpreters

⇨ Web browsers

→ Netscape Navigator, Internet Explorer, Lynx, Opera,…

⇨ Web servers

→ Apache, Microsoft and Netscape servers, …

⇨ Back-end DB connectivity, programming/scripting languages

→ Public domain (e.g. MYSQL) or commercial (Oracle, …)

→ ODBC, JDBC DB connectivity

→ ASP, JSP, PHP scripting embedded in html; CGI, NSAPI, ISAPI to connect to external program

⇨ An application-layer protocol

→ the HyperText Transfer Protocol (HTTP)

Università degli Studi di Palermo

Application-layer protocol vs networking application

➔ **Example2: Internet electronic mail application. Components:**

⇨ mail servers that house user mailboxes

⇨ mail readers that allow users to read and create messages

⇨ a standard for defining the structure of an email message (i.e., MIME)

⇨ application-layer protocols (the principal is SMTP, Simple Mail Transfer Protocol, RFC 821) that define:

➔ how messages are passed between servers

➔ how messages are passed between servers and mail readers

➔ how contents of certain parts of the mail message (e.g., header) are to be interpreted.

Università degli Studi di Palermo

Hypertext Transfer Protocol (HTTP)
a "Client-Server" protocol
a "Request-Response" protocol
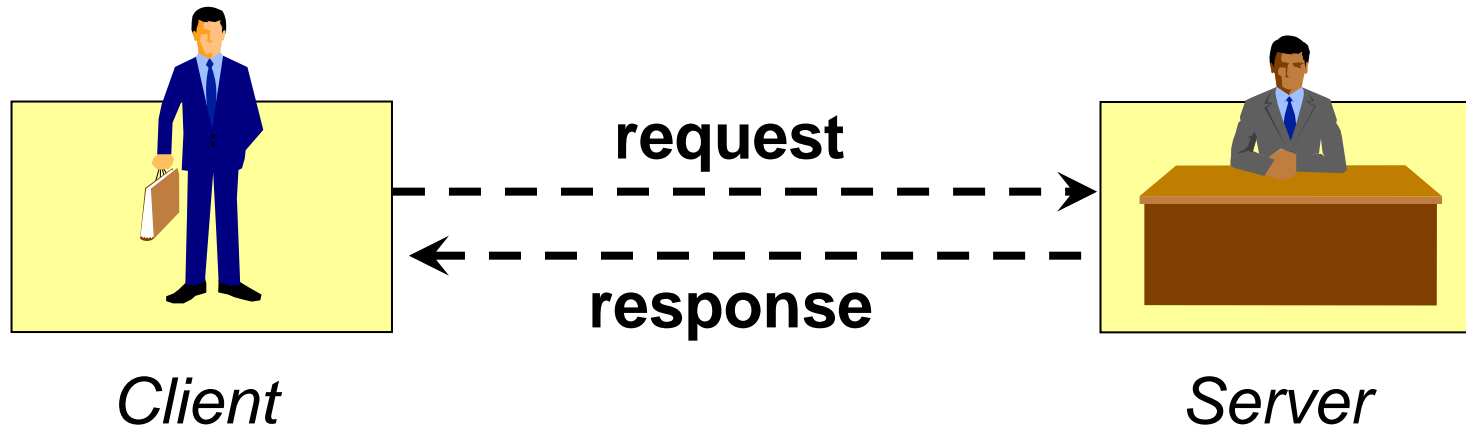
Client

Server

software process:
Web browser

software process:
Web server

G.Bianchi, G.Neglia, V.Mancuso

Client-Server approach

Network application protocol divided into two parts (sides):

→client side: asks

→server side: responds

⇨HTTP:

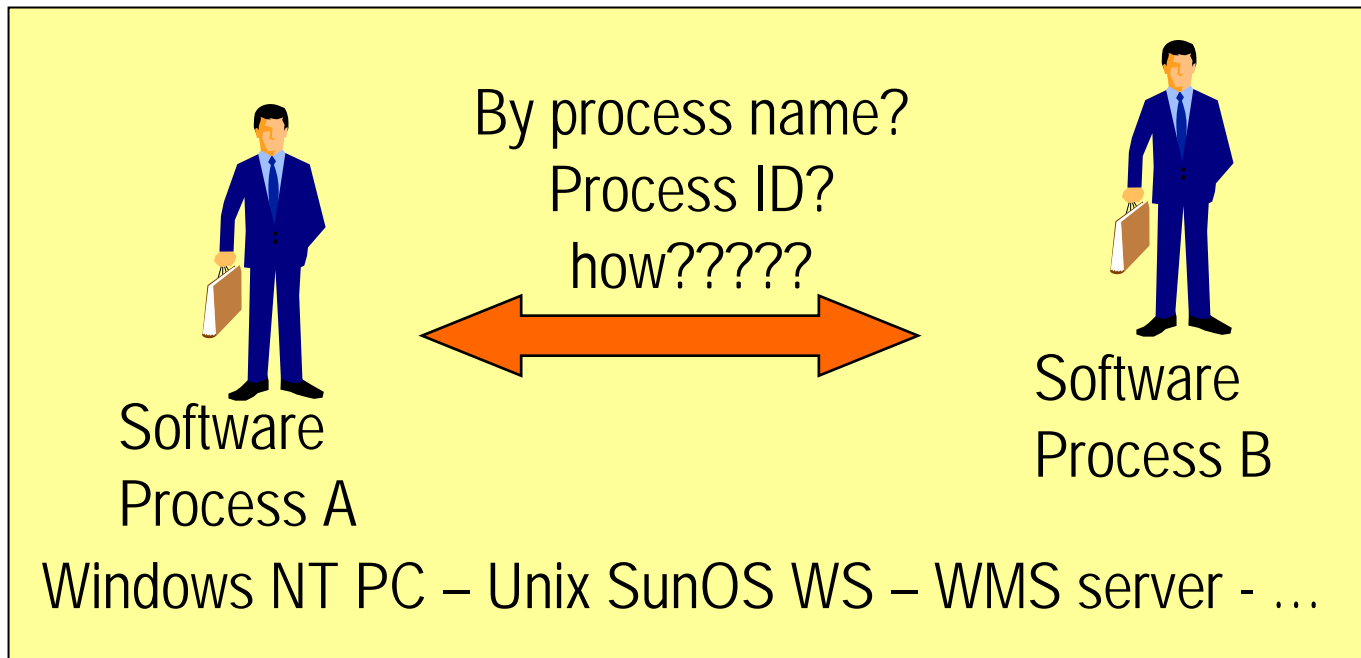» Server = web server

» Client = browser

⇨Telnet, FTP, SMTP:

» host acts as both a client and a server at the same time

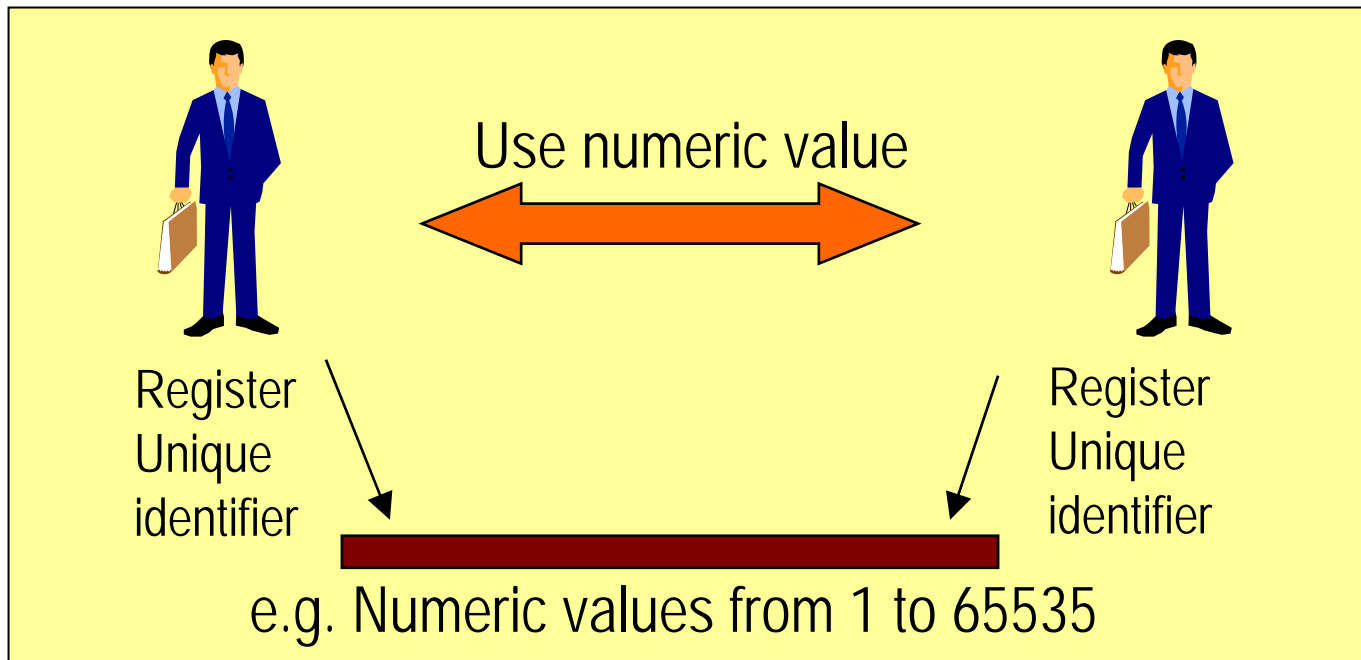» ***conventionally, host that initiates the session is labeled the client***.

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

Addressing software processes
Inside the same machine: an Operating Systems issue

By process name?
Process ID?
how?????

Software
Process A

Software
Process B

Windows NT PC – Unix SunOS WS – WMS server - …

Different OSs = different process naming!

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

Addressing SW processes
very old solution adopted in most OS

Use numeric value

Register Unique identifier

Register Unique identifier

e.g. Numeric values from 1 to 65535

To talk each other, the SW processes need to know their numeric Values.

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

Port numbers
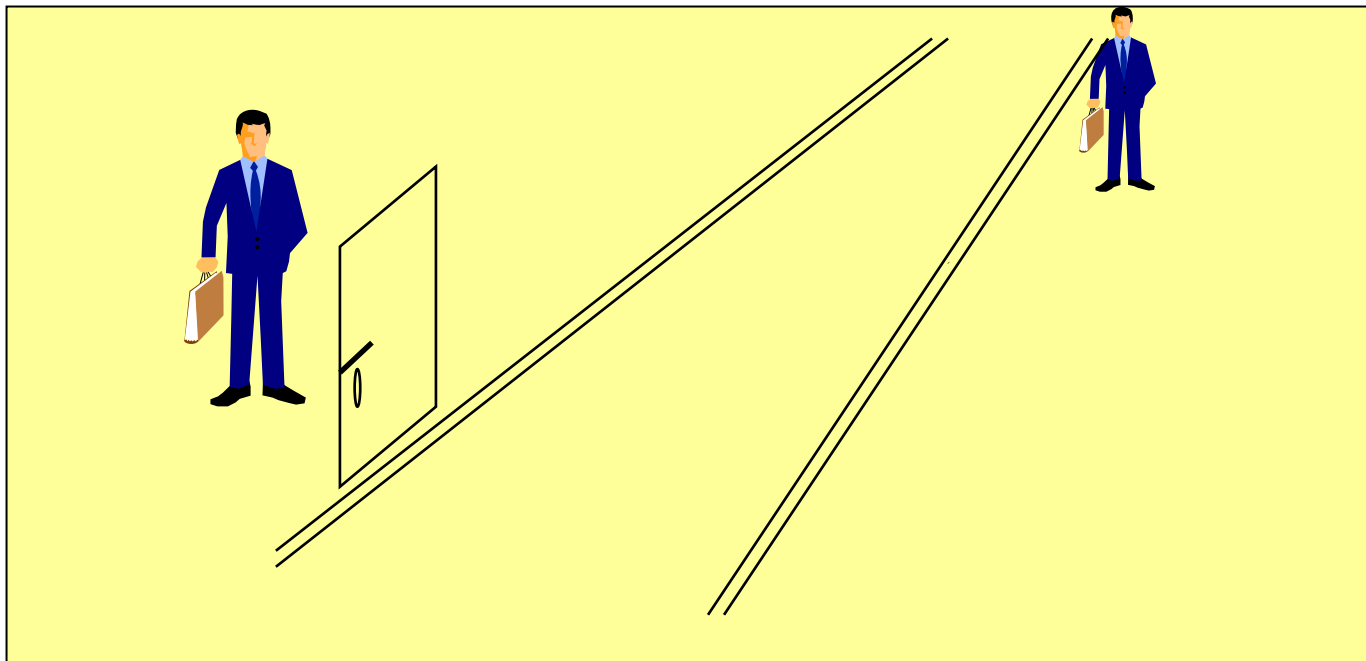
the "address" of the SW process inside the computer!

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

Same addressing scheme works for different machines!!

Port number X

Computer address Y

Want to talk to application process that resides at port X on computer Y

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo
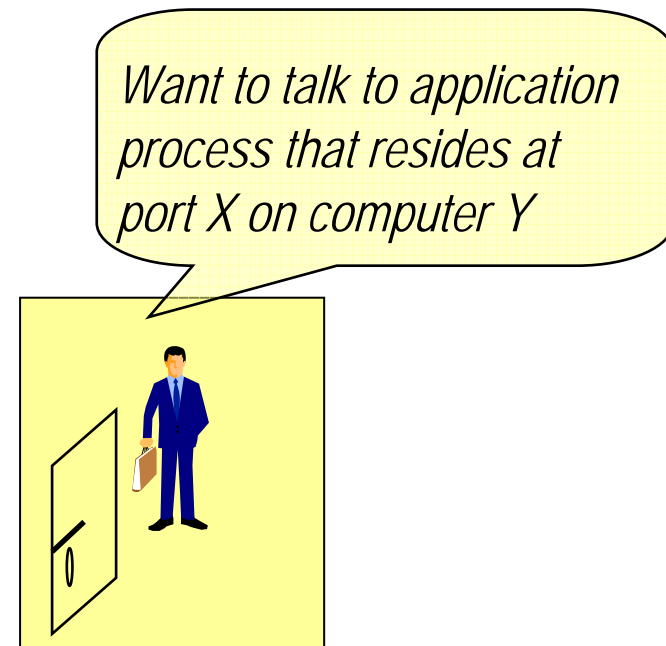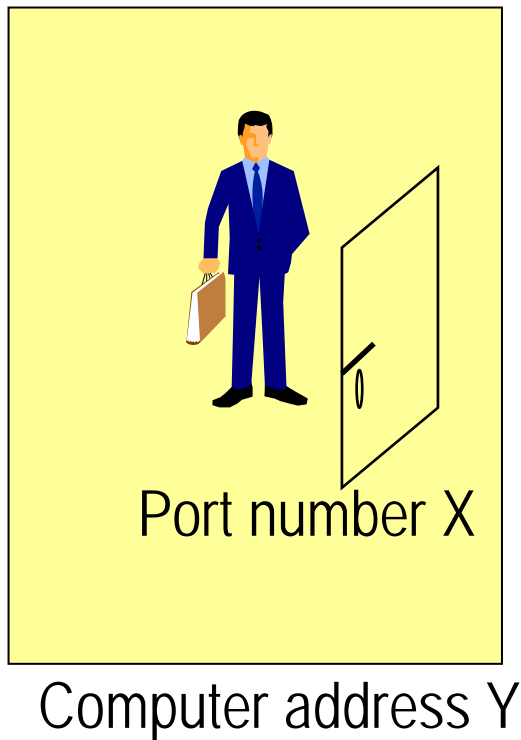
Re-understanding URLs

HTTP://cerbero.elet.polimi.it/people/bianchi/index.html

protocol   location   Filename

HTTP request

Web browser

HTTP response
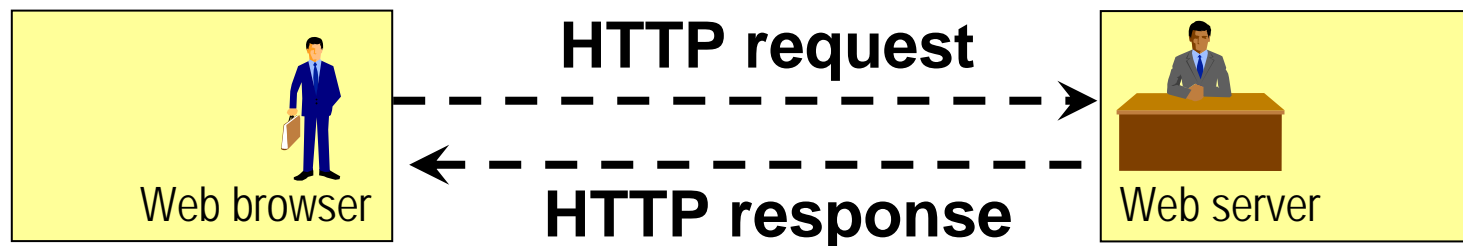
Web server

Web browser (SW process) needs to send HTTP Request to Web Server.
Location not enough (it is just the computer address!)

Università degli Studi di Palermo

Addressing web servers: (wrong) idea

HTTP://cerbero.elet.polimi.it/….

protocol   location   Filename

Location = computer address!
Protocol = SW process address (HTTP = goto Web Server)
…. No need for port numbers??   …

What if more than one Web Server installed???

Università degli Studi di Palermo

URL structure (corrected!)

HTTP://cerbero.elet.polimi.it : *portnumber*/….

protocol   location   Filename

Default value for HTTP: 80

HTTP://cerbero.elet.polimi.it/…

equal to

HTTP://cerbero.elet.polimi.it:80/…

different from
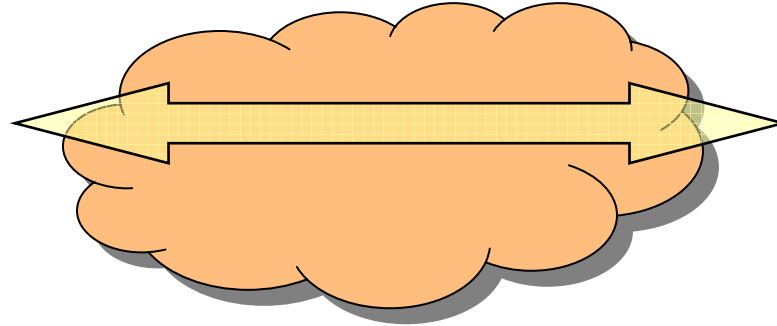
HTTP://cerbero.elet.polimi.it:8080/…   (if exists!)

G.Bianchi, G.Neglia, V.Mancuso

Università degli Studi di Palermo

URL semantics
...finally...

Client
web browser

Server
Web server

Client, via a web browser, wants to retrieve a remote resource:

/people/bianchi/index.html

From an application sw program, the web server, whose address is

131.175.15.1 : 80

Where, for convenience, client uses the DNS name instead:

cerbero.elet.polimi.it : 80

Retrieval will occur using the application layer protocol rules named

http

What happens?

Università degli Studi di Palermo