

# *World Wide Web: HTTP*

## Objectives

*Upon completion you will be able to:*

- *Understand the components of a browser and a server*
- *Understand the function of the URL and cookies*
- *Understand how HTML is related to static documents*
- *Understand how CGI is related to dynamic documents*
- *Understand how Java is related to active documents*
- *Know how HTTP accesses data on the WWW*

# 22.1 ARCHITECTURE

*The WWW is a distributed client-server service, in which a client using a browser can access a service using a server. The service provided is distributed over many locations called sites.*

cuu duong than cong . com

***The topics discussed in this section include:***

***Client (Browser)***

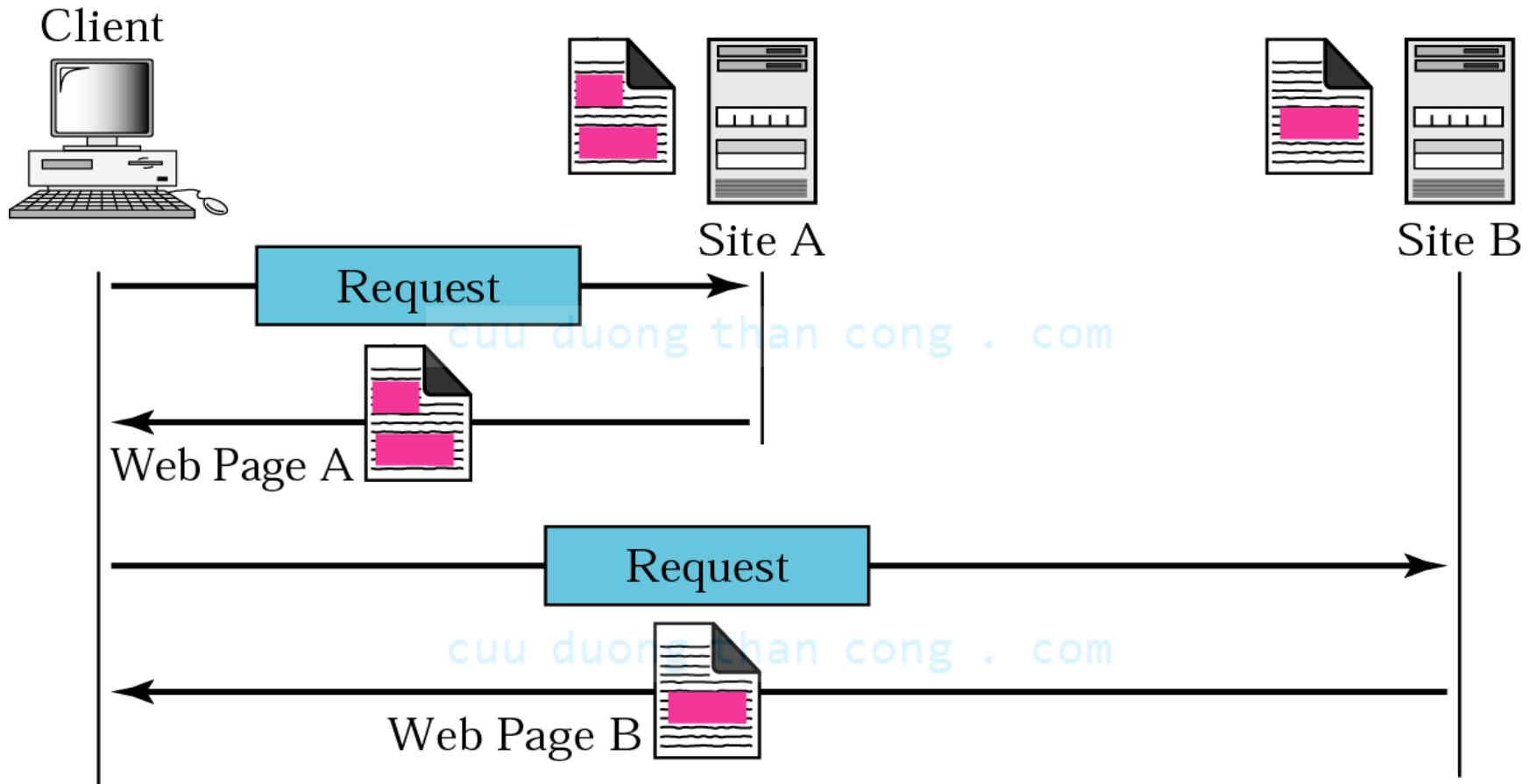
***Server***

***Uniform Resource Locator (URL)***

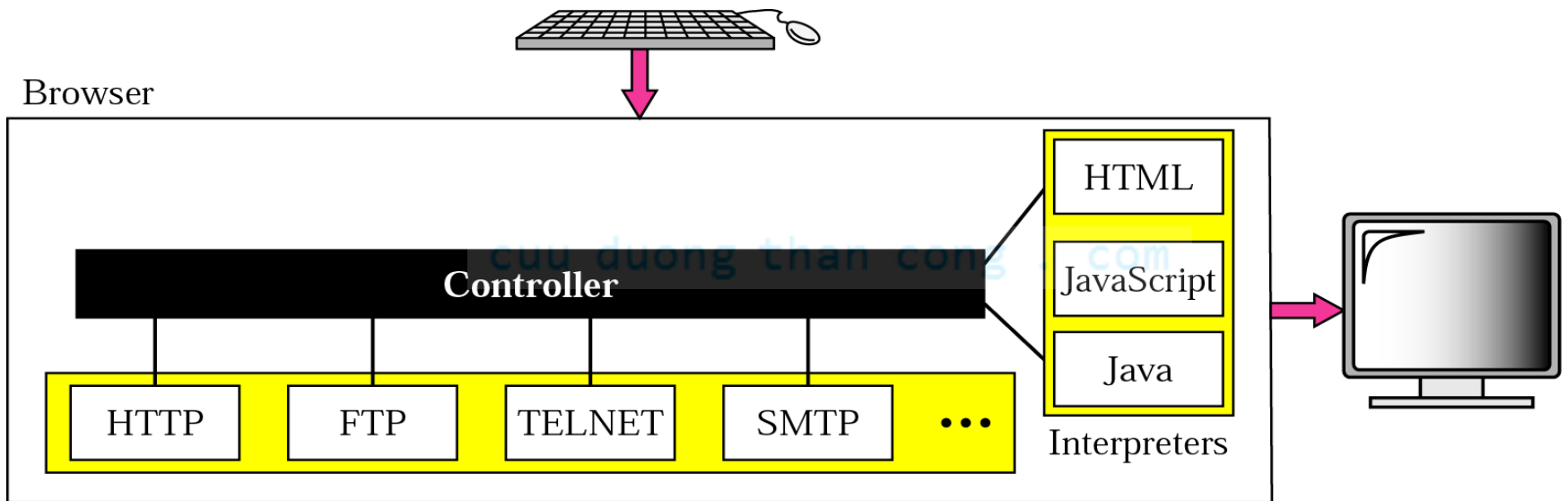
***Cookies***

cuu duong than cong . com

**Figure 22.1** *Architecture of WWW*



**Figure 22.2** *Browser*



**Figure 22.3** *URL*



## 22.2 WEB DOCUMENTS

*The documents in the WWW can be grouped into three broad categories: static, dynamic, and active. The category is based on the time the contents of the document are determined.*

cuu duong than cong . com

*The topics discussed in this section include:*

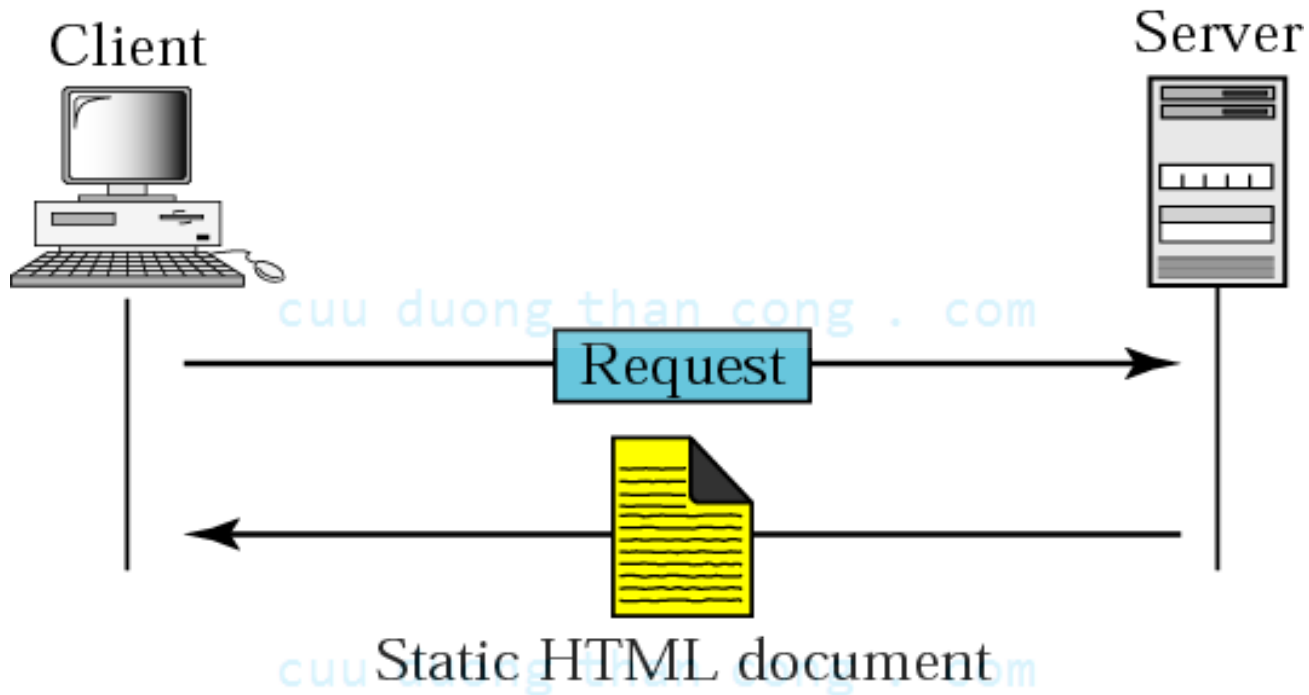
*Static Documents*

*Dynamic Documents*

*Active Documents*

cuu duong than cong . com

**Figure 22.4** *Static document*



**Figure 22.5** *Boldface tags*

Bold tag

cuu duong than cong . com

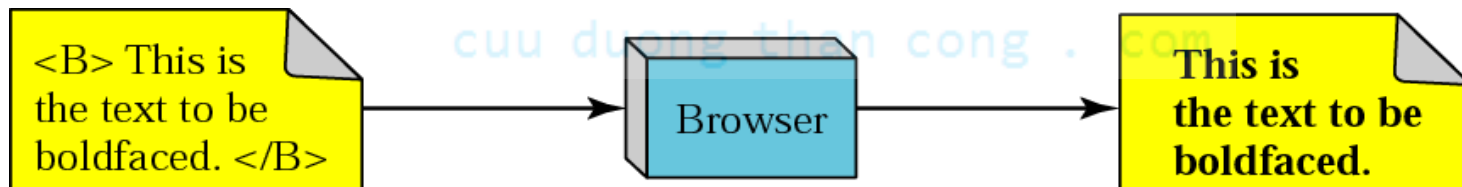
End bold

<B> This is the text to be boldfaced.</B>

cuu duong than cong . com



**Figure 22.6** *Effect of boldface tags*



**Figure 22.7** *Beginning and ending tags*

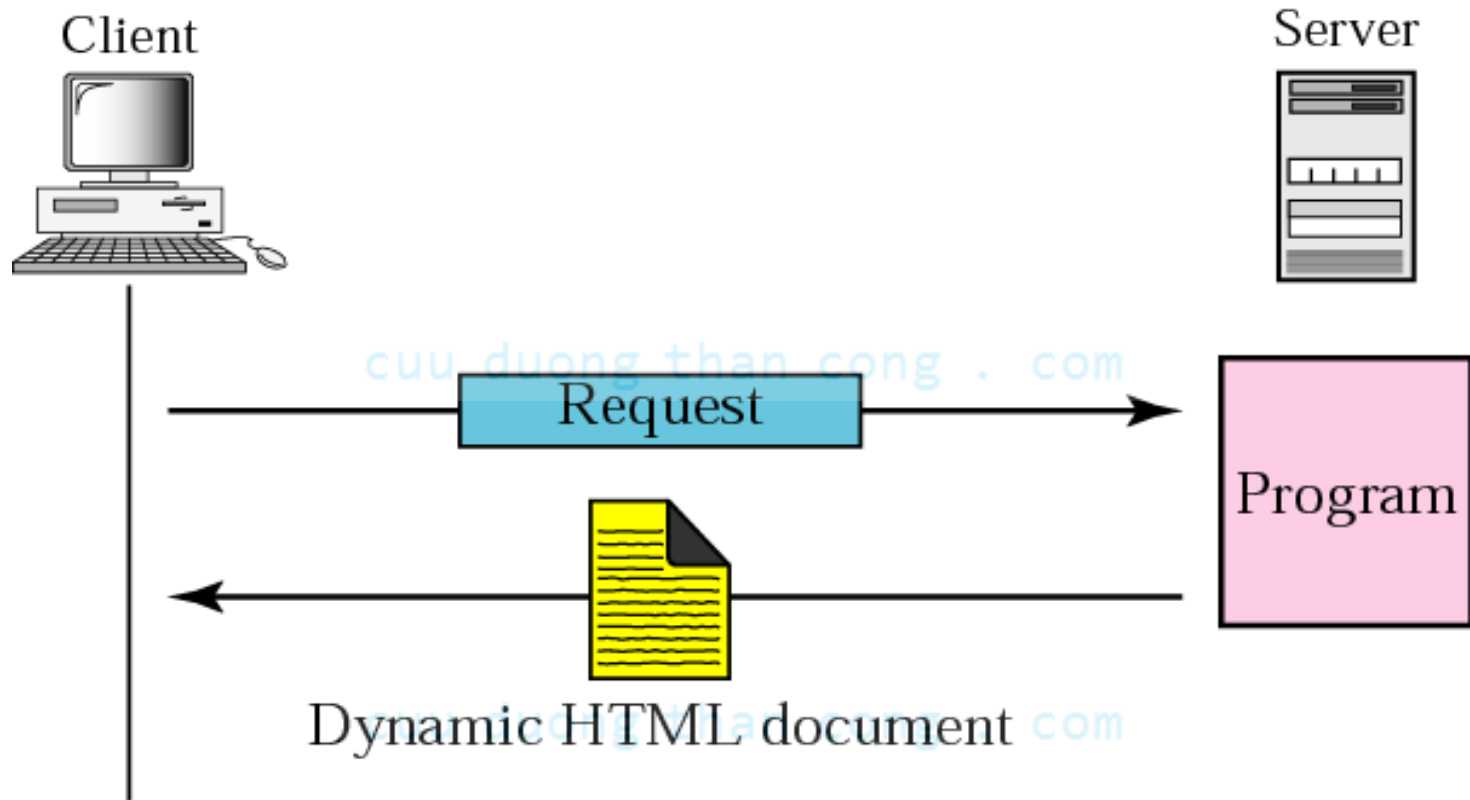
< TagName Attribute = Value Attribute = Value ... >

a. Beginning tag

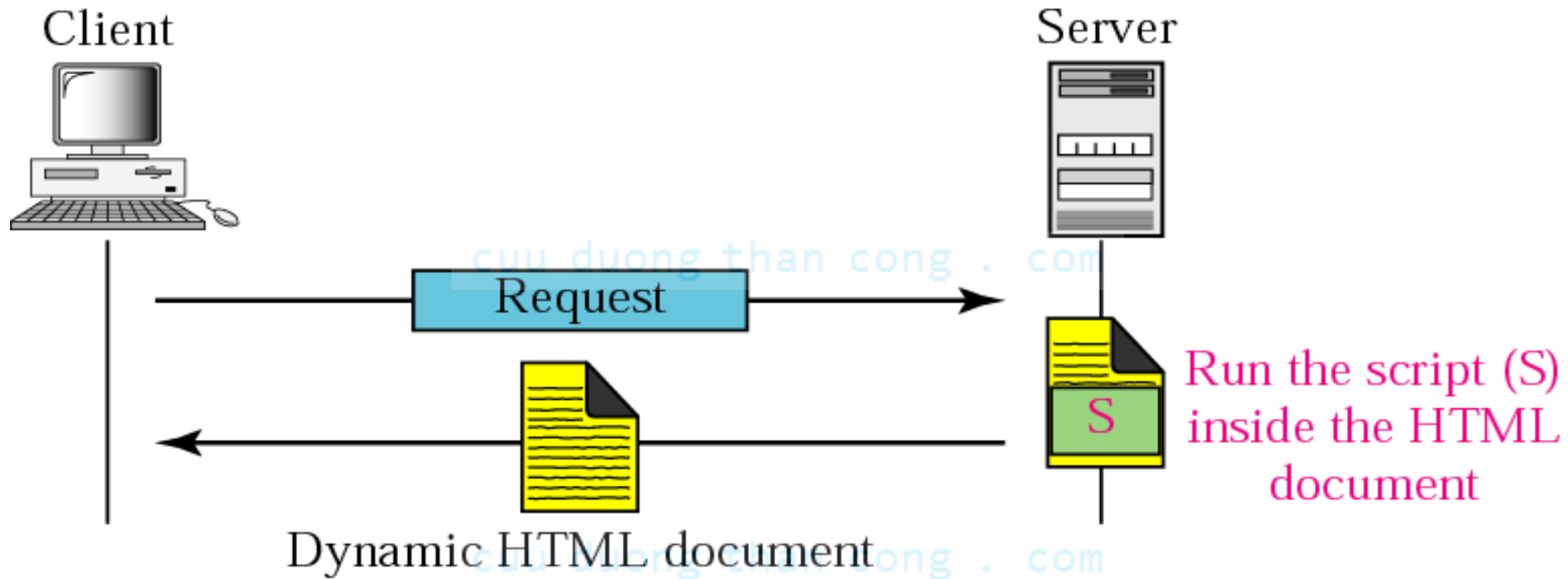
< /TagName >

b. Ending tag

**Figure 22.8** *Dynamic document using CGI*



**Figure 22.9** *Dynamic document using server-site script*



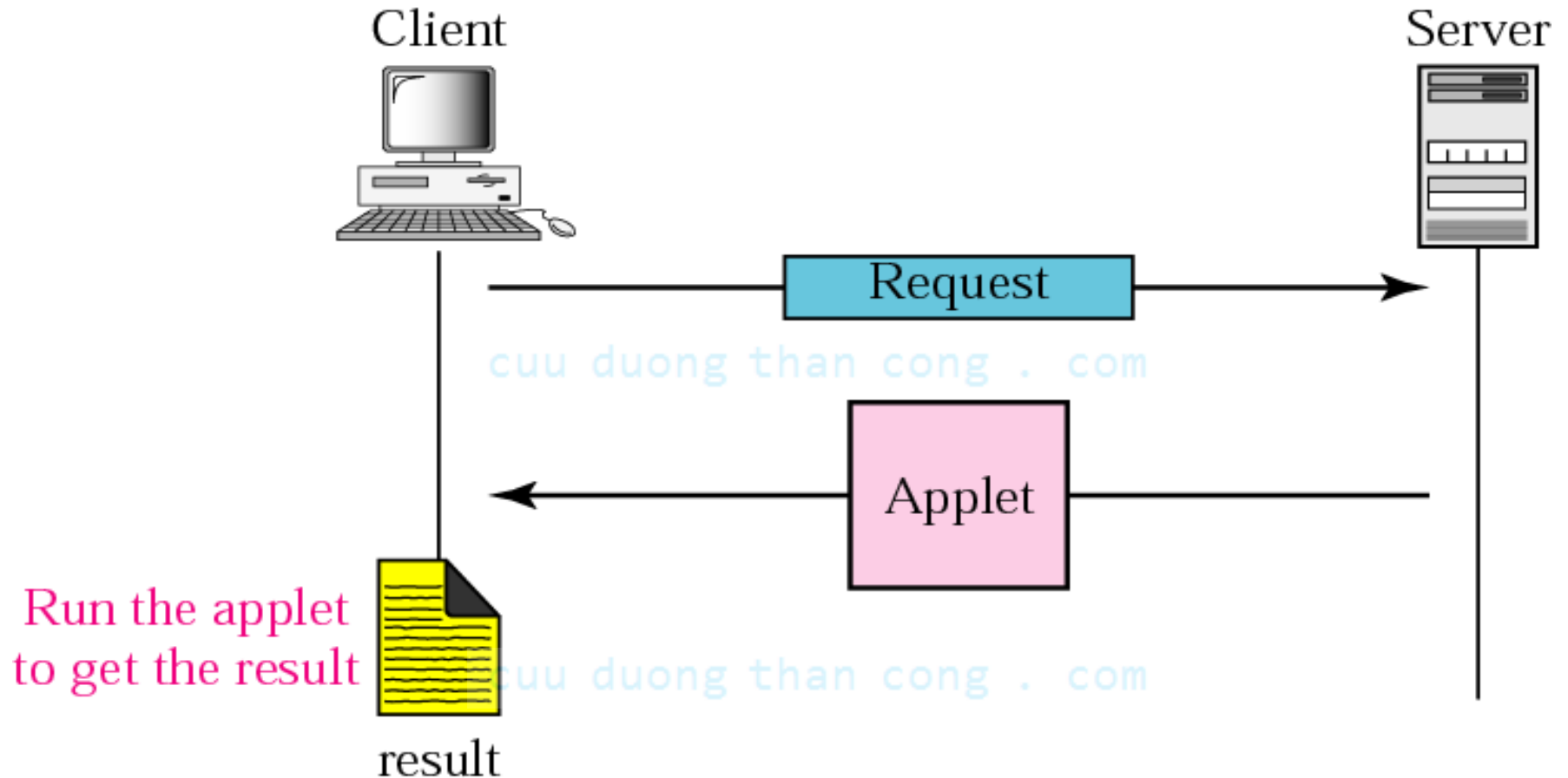


Note:

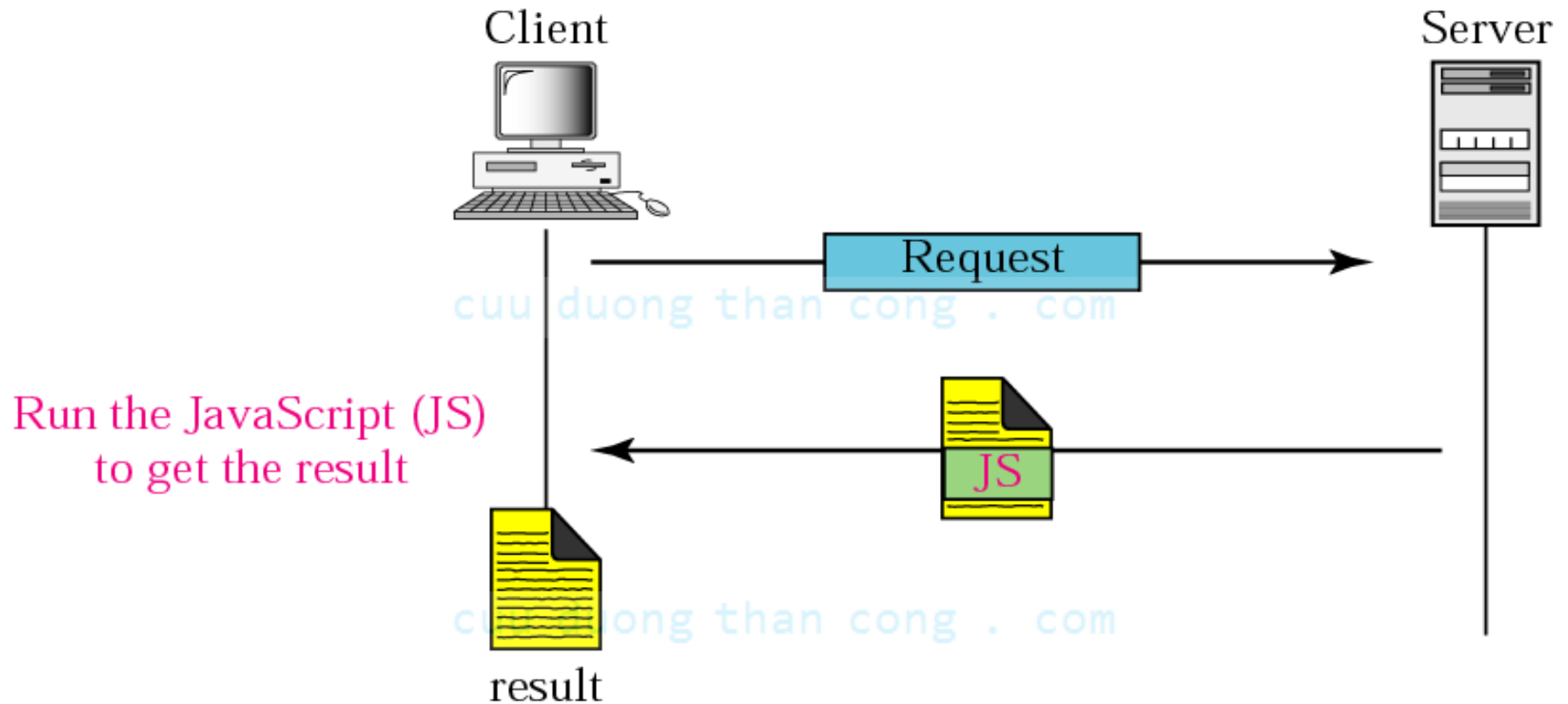
*Dynamic documents are sometimes referred to as server-site dynamic documents.*

cuu duong than cong . com

**Figure 22.10** *Active document using Java applet*



**Figure 22.11** *Active document using client-site script*





Note:

*Active documents are sometimes referred to as client-site dynamic documents.*

cuu duong than cong . com



## 22.3 HTTP

*The Hypertext Transfer Protocol (HTTP) is a protocol used mainly to access data on the World Wide Web. HTTP functions like a combination of FTP and SMTP.*

cuu duong than cong . com

*The topics discussed in this section include:*

*HTTP Transaction*

*Persistent versus Nonpersistent Connection*

*Proxy Server*

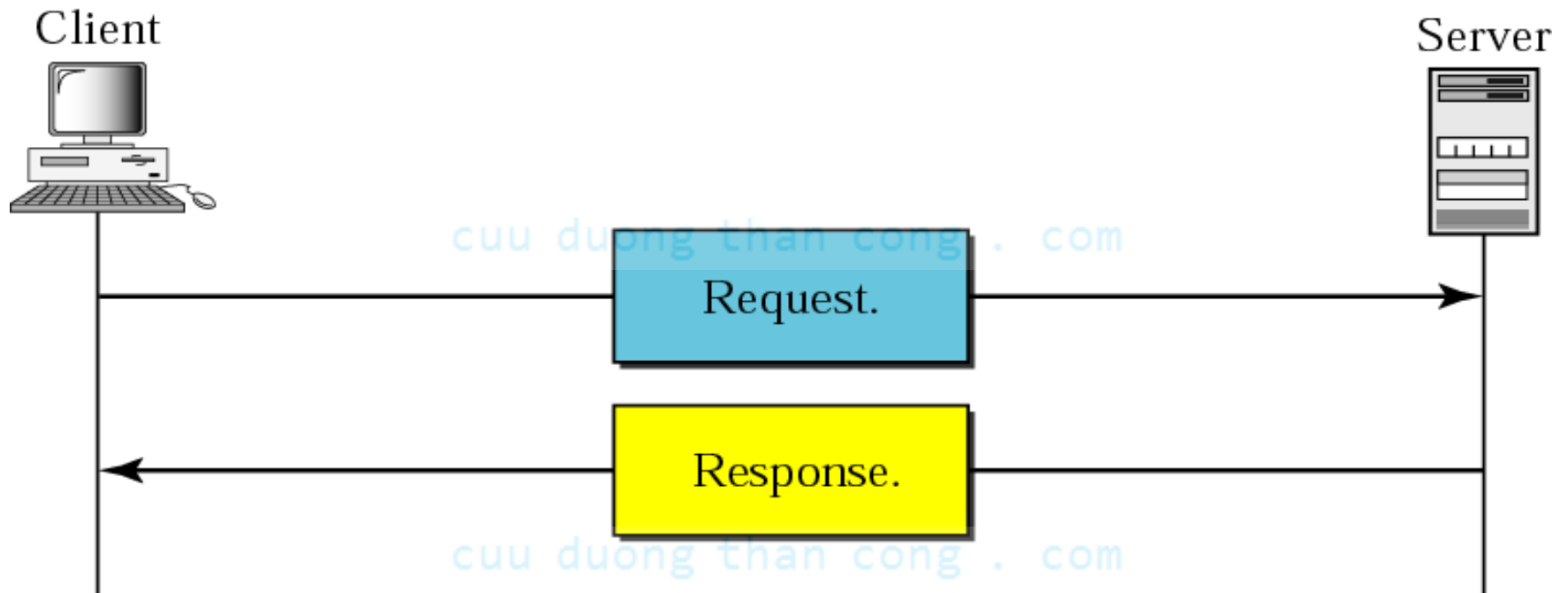


Note:

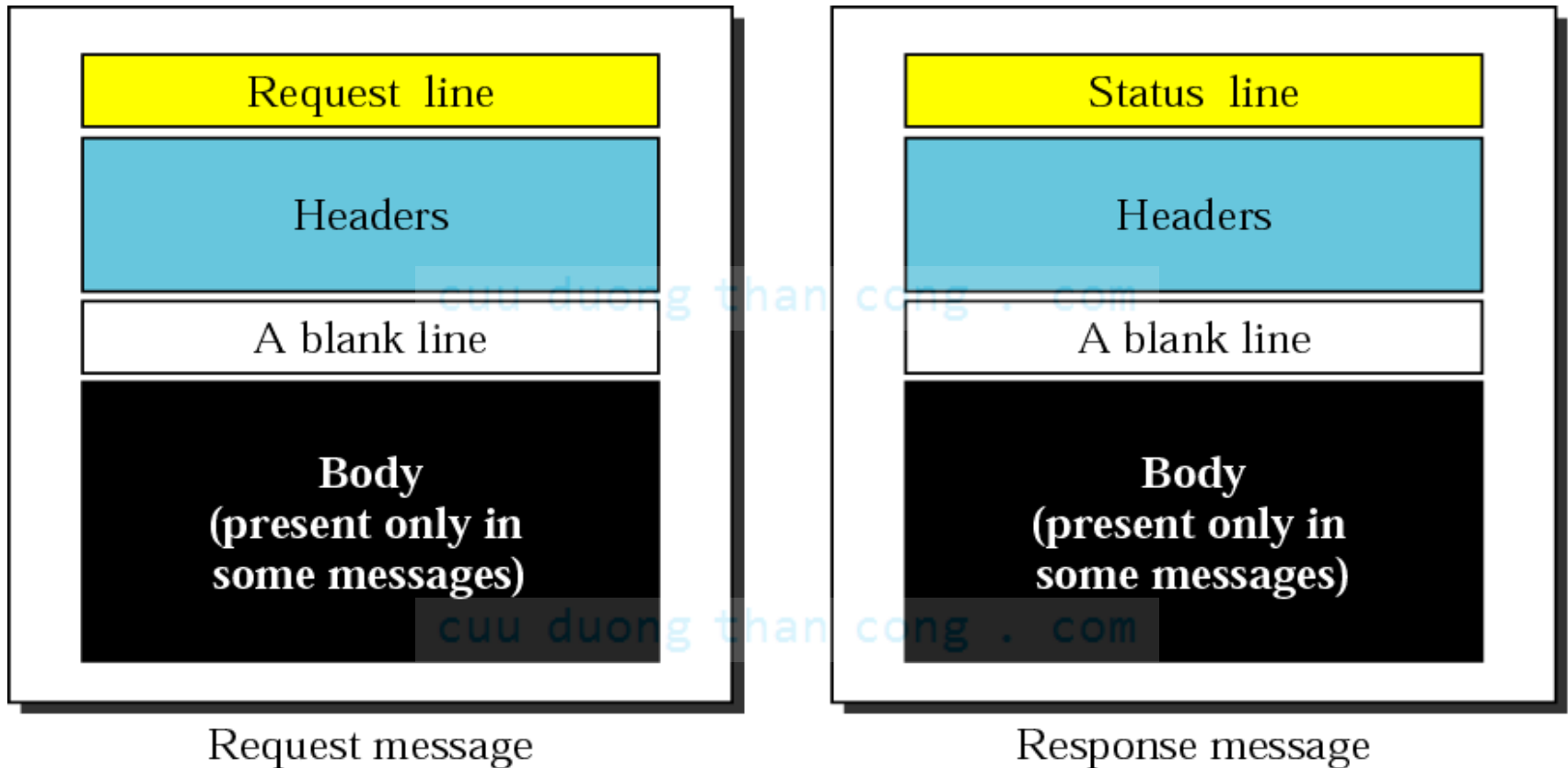
*HTTP uses the services of TCP on well-known port 80.*

cuu duong than cong . com

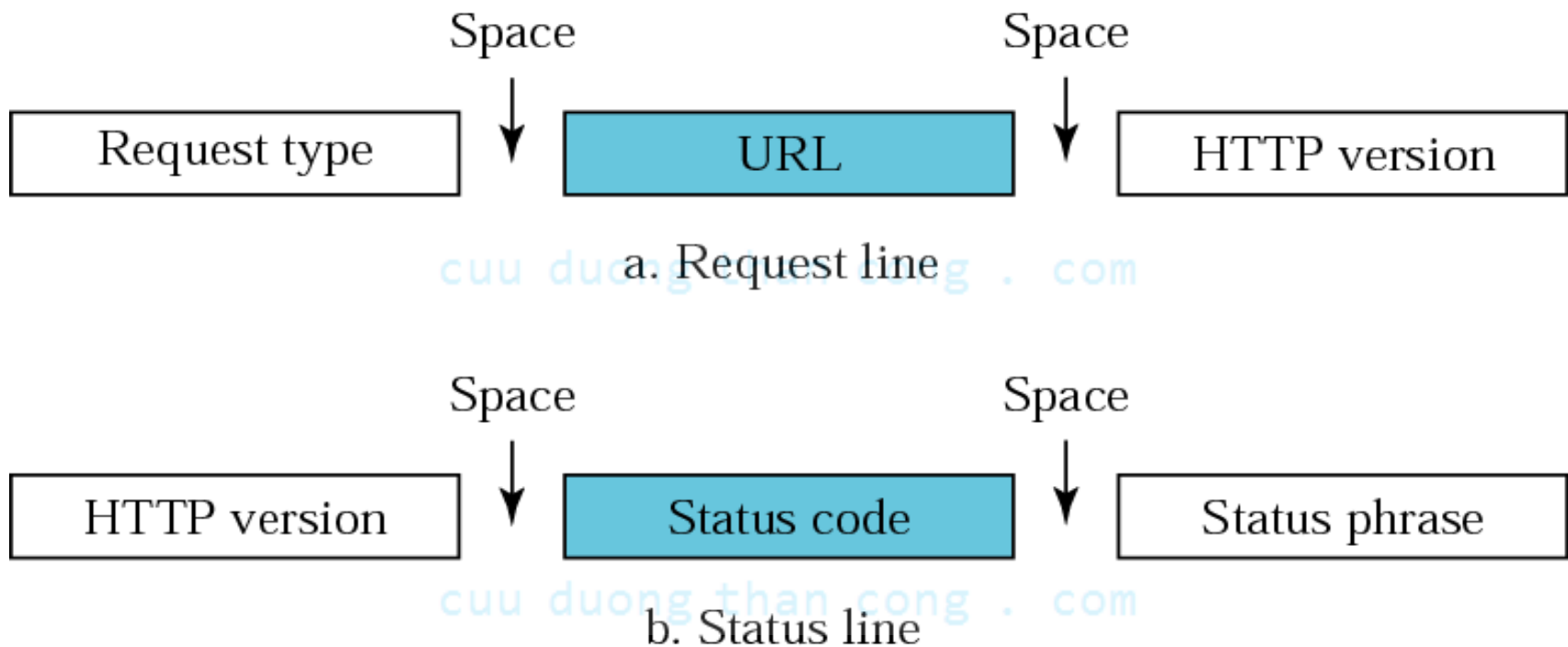
**Figure 22.12** *HTTP transaction*



**Figure 22.13** *Request and response messages*



**Figure 22.14** *Request and status lines*



**Table 22.1 Methods**

<i>Method</i>	<i>Action</i>
GET	Requests a document from the server
HEAD	Requests information about a document but not the document itself
POST	Sends some information from the client to the server
PUT	Sends a document from the server to the client
TRACE	Echoes the incoming request
CONNECT	Reserved
OPTION	Enquires about available options

**Table 22.2 Status codes**

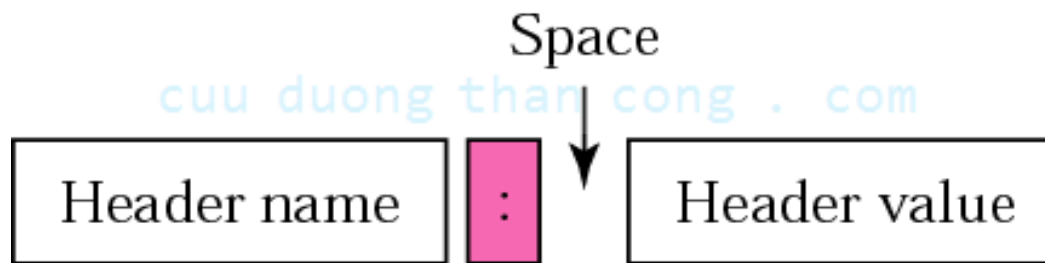
Code	Phrase	Description
<b>Informational</b>		
<b>100</b>	Continue	The initial part of the request has been received and the client may continue with its request.
<b>101</b>	Switching	The server is complying with a client request to switch protocols defined in the upgrade header.
<b>Success</b>		
<b>200</b>	OK	The request is successful.
<b>201</b>	Created	A new URL is created.
<b>202</b>	Accepted	The request is accepted, but it is not immediately acted upon.
<b>204</b>	No content	There is no content in the body.

**Table 22.2 Status codes (continued)**

Code	Phrase	Description
<b>Redirection</b>		
<b>301</b>	Multiple choices	The requested URL refers to more than one resource.
<b>302</b>	Moved permanently	The requested URL is no longer used by the server.
<b>304</b>	Moved temporarily	The requested URL has moved temporarily.
<b>Client Error</b>		
<b>400</b>	Bad request	There is a syntax error in the request.
<b>401</b>	Unauthorized	The request lacks proper authorization.
<b>403</b>	Forbidden	Service is denied.
<b>404</b>	Not found	The document is not found.
<b>405</b>	Method not allowed	The method is not supported in this URL.
<b>406</b>	Not acceptable	The format requested is not acceptable.
<b>Server Error</b>		
<b>500</b>	Internal server error	There is an error, such as a crash, at the server site.
<b>501</b>	Not implemented	The action requested cannot be performed.
<b>503</b>	Service unavailable	The service is temporarily unavailable, but may be requested in the future.



**Figure 22.15** *Header format*



**Table 22.3** *General headers*

<i>Header</i>	<i>Description</i>
Cache-control	Specifies information about caching
Connection	Shows whether the connection should be closed or not
Date	Shows the current date
MIME-version	Shows the MIME version used
Upgrade	Specifies the preferred communication protocol

**Table 22.4 Request headers**

<i>Header</i>	<i>Description</i>
Accept	Shows the media format the client can accept
Accept-charset	Shows the character set the client can handle
Accept-encoding	Shows the encoding scheme the client can handle
Accept-language	Shows the language the client can accept
Authorization	Shows what permissions the client has
From	Shows the e-mail address of the user
Host	Shows the host and port number of the client
If-modified-since	Send the document if newer than specified date
If-match	Send the document only if it matches given tag
If-non-match	Send the document only if it does not match given tag
If-range	Send only the portion of the document that is missing
If-unmodified-since	Send the document if not changed since specified date
Referrer	Specifies the URL of the linked document
User-agent	Identifies the client program

***Table 22.5 Response headers***

<i>Header</i>	<i>Description</i>
Accept-range	Shows if server accepts the range requested by client
Age	Shows the age of the document
Public	Shows the supported list of methods
Retry-after	Specifies the date after which the server is available
Server	Shows the server name and version number

***Table 22.6 Entity headers***

<i>Header</i>	<i>Description</i>
Allow	Lists valid methods that can be used with a URL
Content-encoding	Specifies the encoding scheme
Content-language	Specifies the language
Content-length	Shows the length of the document
Content-range	Specifies the range of the document
Content-type	Specifies the media type
Etag	Gives an entity tag
Expires	Gives the date and time when contents may change
Last-modified	Gives the date and time of the last change
Location	Specifies the location of the created or moved document

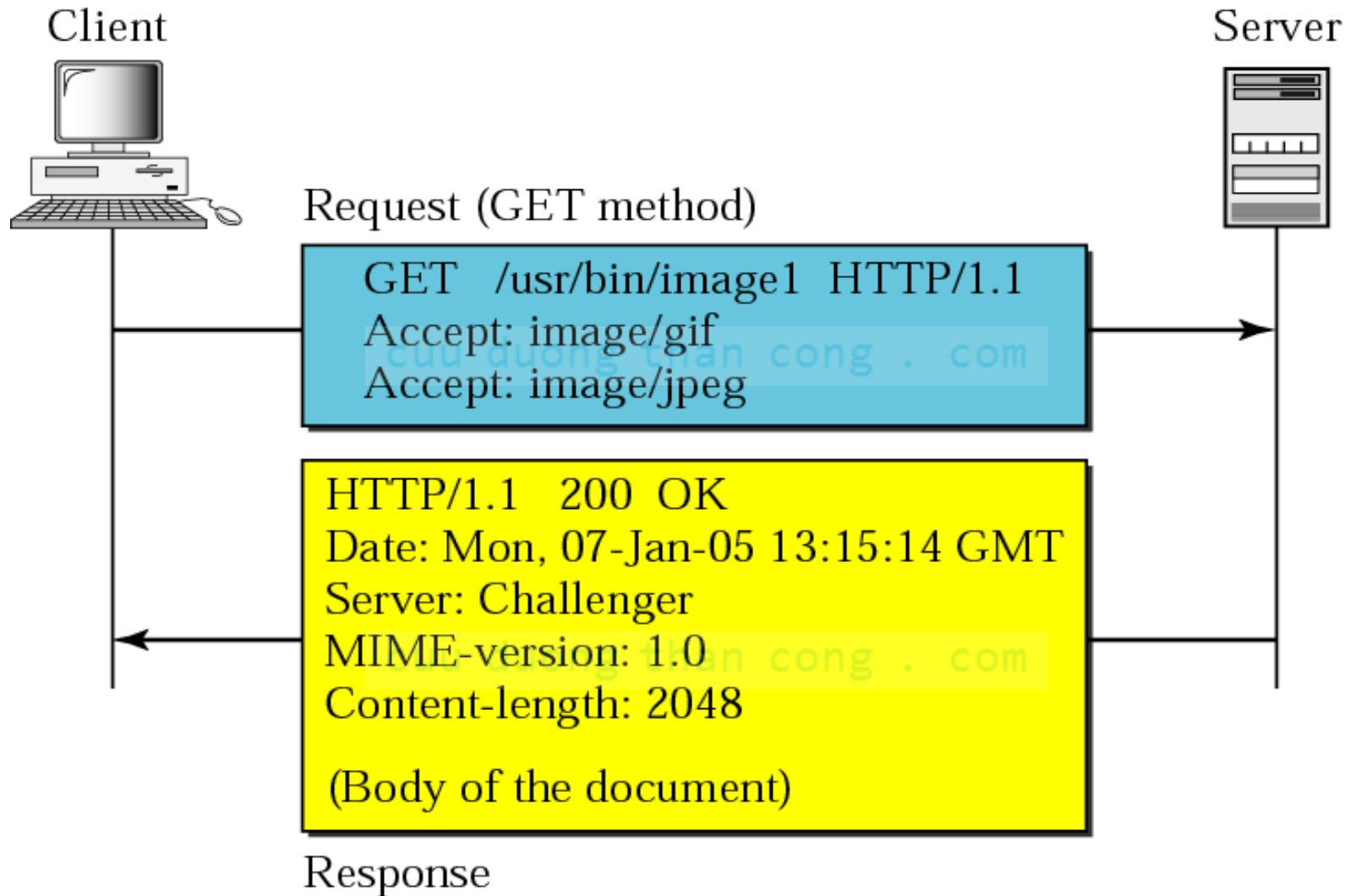


## ***EXAMPLE 1***

*This example retrieves a document. We use the GET method to retrieve an image with the path /usr/bin/image1. The request line shows the method (GET), the URL, and the HTTP version (1.1). The header has two lines that show that the client can accept images in the GIF or JPEG format. The request does not have a body. The response message contains the status line and four lines of header. The header lines define the date, server, MIME version, and length of the document. The body of the document follows the header (see Figure 22.16).*

**See Next Slide**

**Figure 22.16** *Example 1*





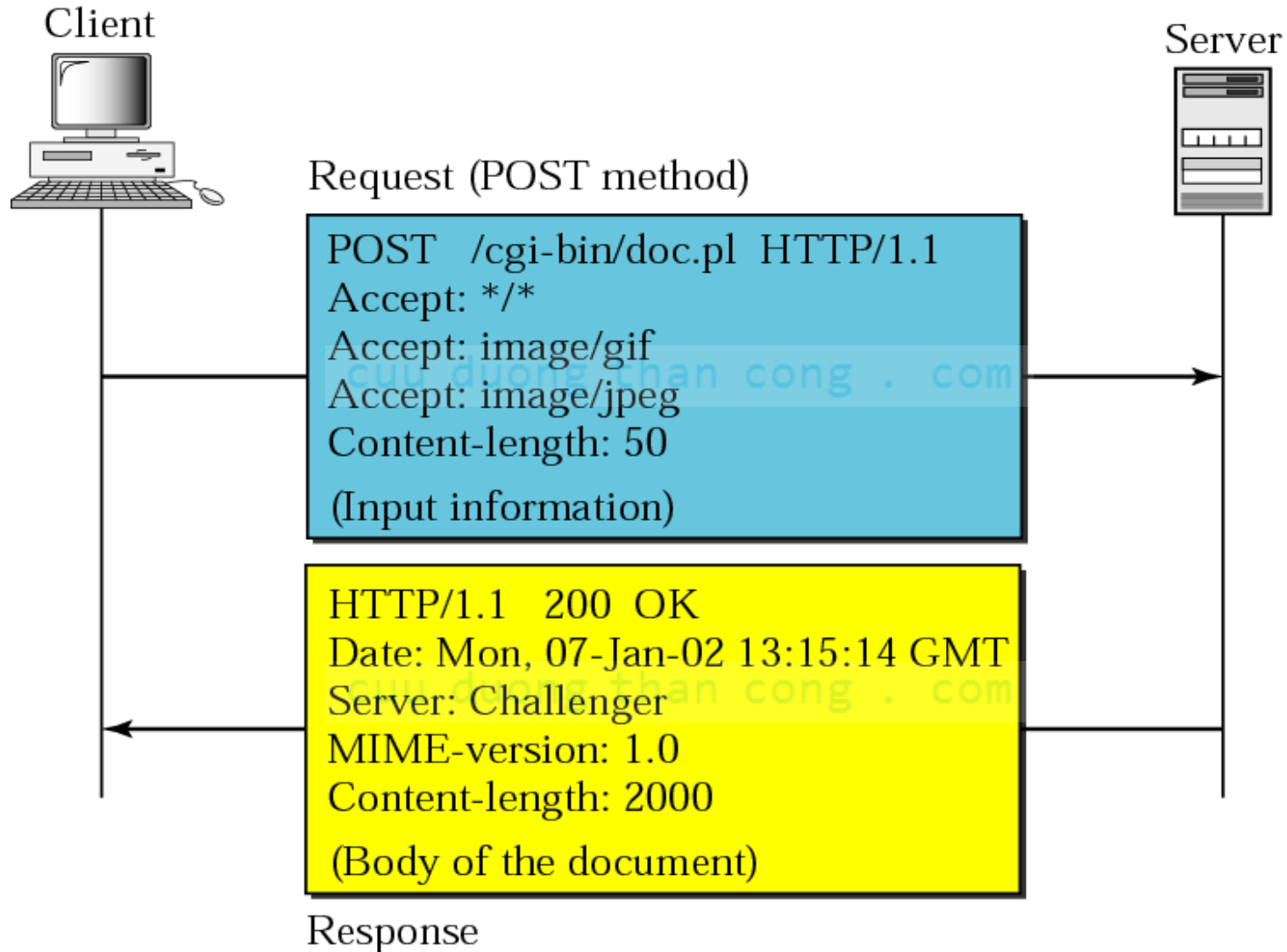
## ***EXAMPLE 2***

*In this example, the client wants to send data to the server. We use the POST method. The request line shows the method (POST), URL, and HTTP version (1.1). There are four lines of headers. The request body contains the input information. The response message contains the status line and four lines of headers. The created document, which is a CGI document, is included as the body (see Figure 22.17).*

**See Next Slide**



**Figure 22.17** *Example 2*





## ***EXAMPLE 3***

*HTTP uses ASCII characters. A client can directly connect to a server using TELNET, which logs into port 80. The next three lines shows that the connection is successful. We then type three lines. The first shows the request line (GET method), the second is the header (defining the host), the third is a blank terminating the request. The server response is seven lines starting with the status line. The blank line at the end terminates the server response. The file of 14230 lines is received after the blank line (not shown here). The last line is the output by the client.*

**See Next Slide**



## ***EXAMPLE 3***

***\$ telnet www.mhhe.com 80***

*Trying 198.45.24.104...*

*Connected to www.mhhe.com (198.45.24.104).*

*Escape character is '^]'.*

*GET /engcs/compsci/forouzan HTTP/1.1*

*From: forouzanbehrouz@fhda.edu*

***HTTP/1.1 200 OK***

***Date: Thu, 28 Oct 2004 16:27:46 GMT***

***Server: Apache/1.3.9 (Unix) ApacheJServ/1.1.2 PHP/4.1.2 PHP/3.0.18***

***MIME-version:1.0***

***Content-Type: text/html***

***Last-modified: Friday, 15-Oct-04 02:11:31 GMT***

***Content-length: 14230***

*Connection closed by foreign host.*



*HTTP version 1.1 specifies a persistent connection by default.*

cuu duong than cong . com