

1 DNS - Domain Name System

The way DNS is used is as follows:

An application program calls a library procedure called the resolver; 客户端有一个library进程叫做resolver会来帮助解析域名。

The resolver sends a UDP packet to a local DNS server, which then looks up the name and returns the ip address to the resolver. TCP / IP 协议中（就是设置的时候要配置），会有设定一个域名服务器就是这个DNS server，这个就是你本地的域名服务器，这个域名服务器可能还会问其他服务器查询目标域名。数据比较简单，所以用UDP包，如果出现丢包，就用停止等待协议。

The resolver then returns it to the caller.

1.1 The DNS Name Space

- The Internet is divided into over 200 top-level domains; Each domain is partitioned into subdomains. All domains can be represented by a tree, -[see fig]
- The top level domains come in two flavors: generic and countries.
 - Generic: com,edu,gov,int,mil,net,org ...
 - newly approved: biz (business), info, name (people's name), pro (professions), aero (aerospace), coop (co-operatives), museum.
 - countries: cn,ca,uk,jp,fr,...
- Each domain is named by the path upward from it to the (unnamed) root.. The components are separated by periods (pronounced “dot”) — cs.zju.edu.cn
- Domain names are case insensitive, so edu and EDU mean the same thing.
- Component names can be up to 63 characters long, and full path names must not exceed 255 characters.

- 域名是大小写无关的

1.2 Resource Records

- The primary function of DNS is to map domain names onto resource records.
- The format of Resource Records:
 - Domain_name, Time_to_live, Class, Type, Value
 - Type: tells what kind of record this is-[see fig]
 - The most important is A (address) record
- A portion of a possible DNS database for cs.vu.nl-[see fig]

Type	Meaning	Value
SOA	Start of Authority	Parameters for this zone
A	IP address of a host	32-Bit integer
MX	Mail exchange	Priority, domain willing to accept e-mail
NS	Name Server	Name of a server for this domain
CNAME	Canonical name	Domain name
PTR	Pointer	Alias for an IP address
HINFO	Host description	CPU and OS in ASCII
TXT	Text	Uninterpreted ASCII text

Fig . The principal DNS resource records types.

1.3 Name Server

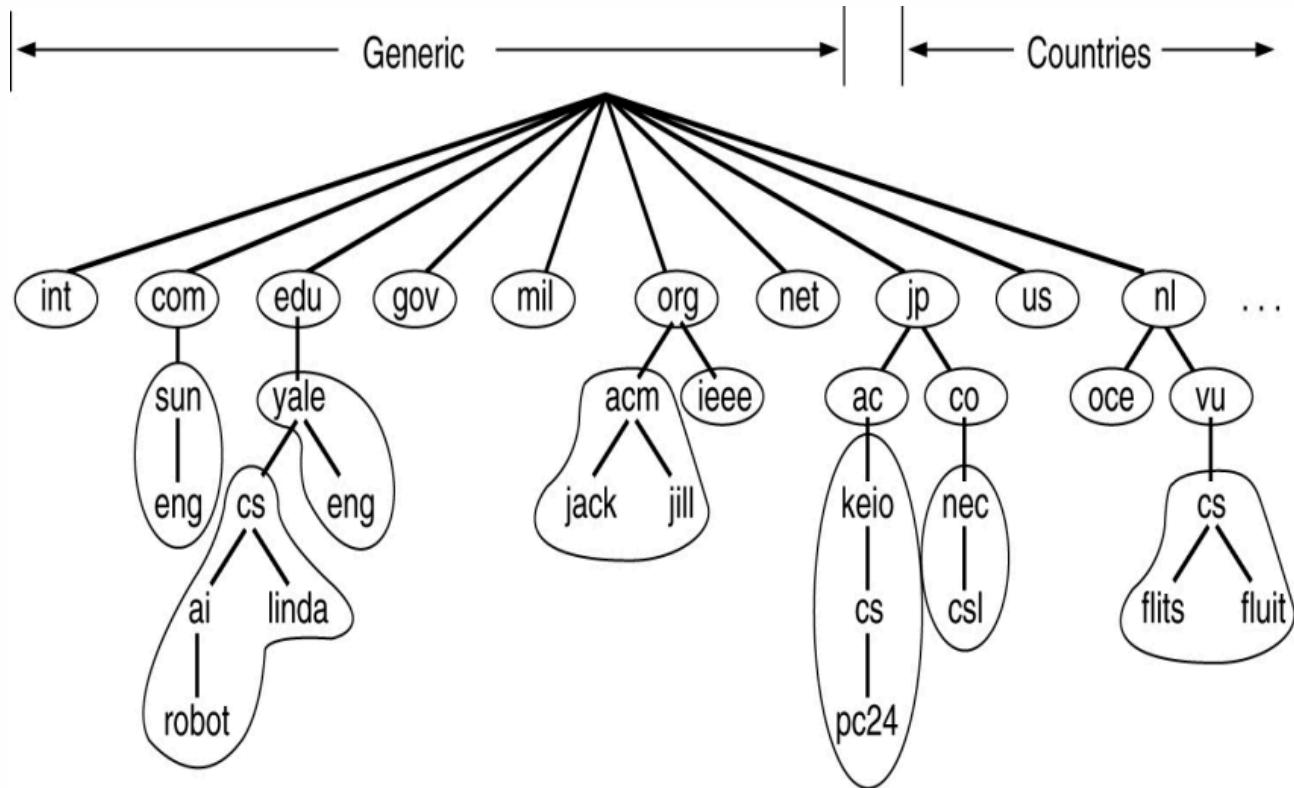


Fig . Part of the DNS name space showing the division into zones.

- 比如yale里分成两个zone: zone偏向一个地理上的，一个域名只能归属于一个zone
- 有很多级的域名服务器

1.3.1 域名解析

1.3.1.1 WAY1 - recursive 递归查找

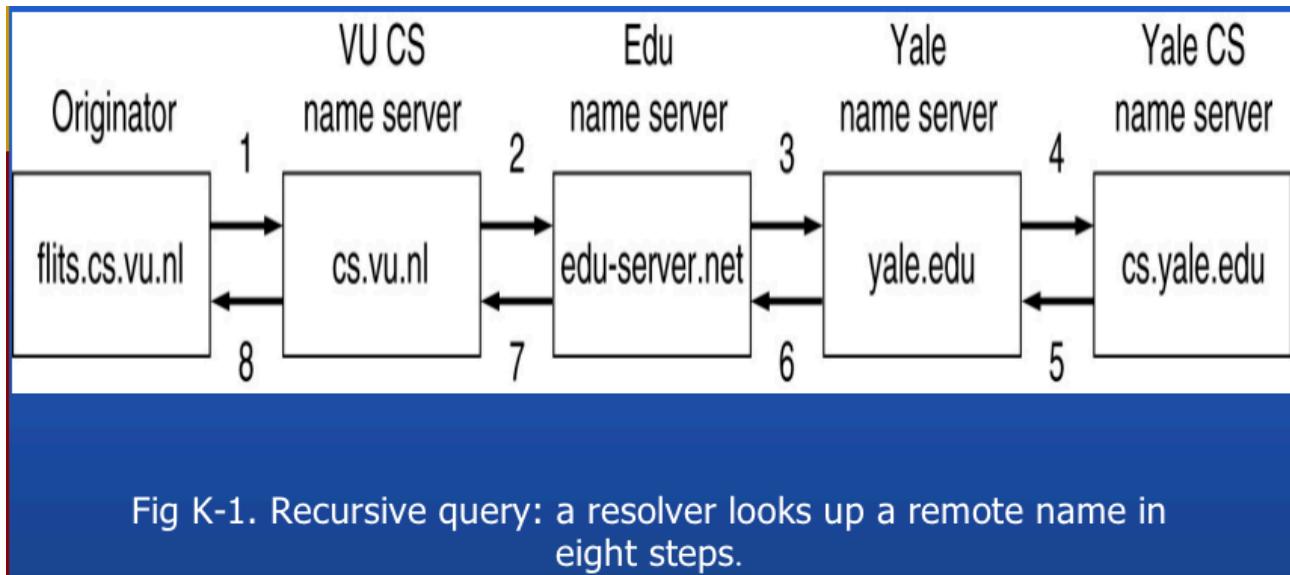


Fig K-1. Recursive query: a resolver looks up a remote name in eight steps.

- Recursive query: resolver in host flits.cs.vu.nl wants to know the IP address of host linda.cs.yale.edu. 递归查找，每个服务器会负责帮你找。
- 发起者先问本地域名服务器；本地域名服务器没有，去问顶级域名服务器；顶级域名服务器去问yale.edu二级域名服务器；二级服务器再去问三级服务器

1.3.1.2 WAY2

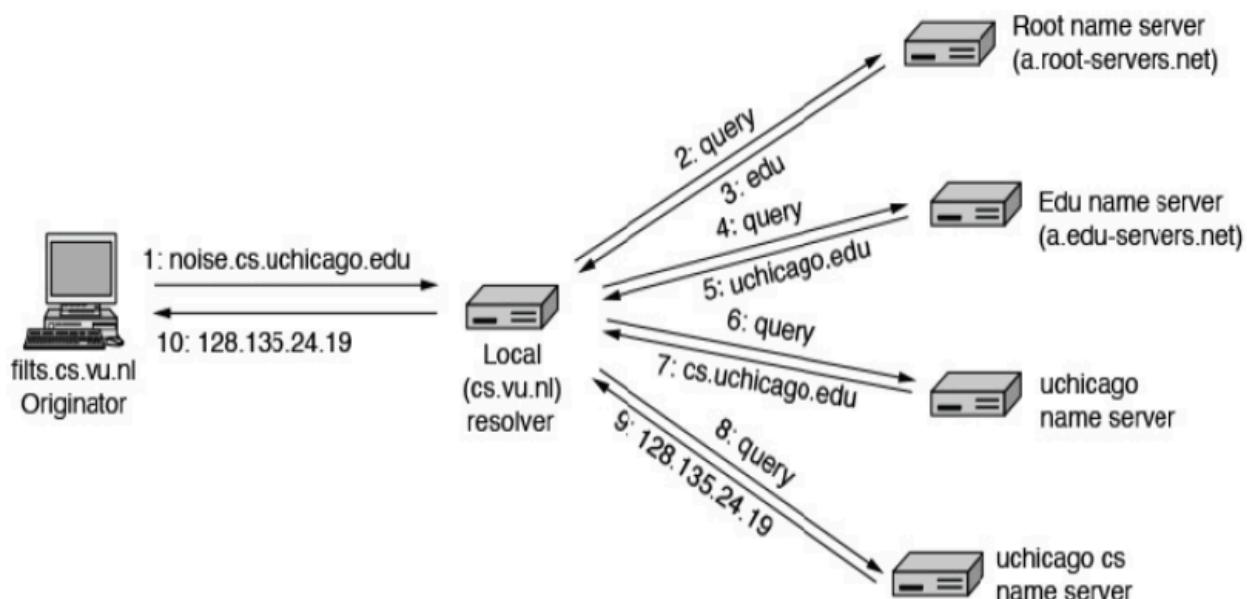


Figure 7-7

Example of a resolver looking up a remote name in 10 steps.

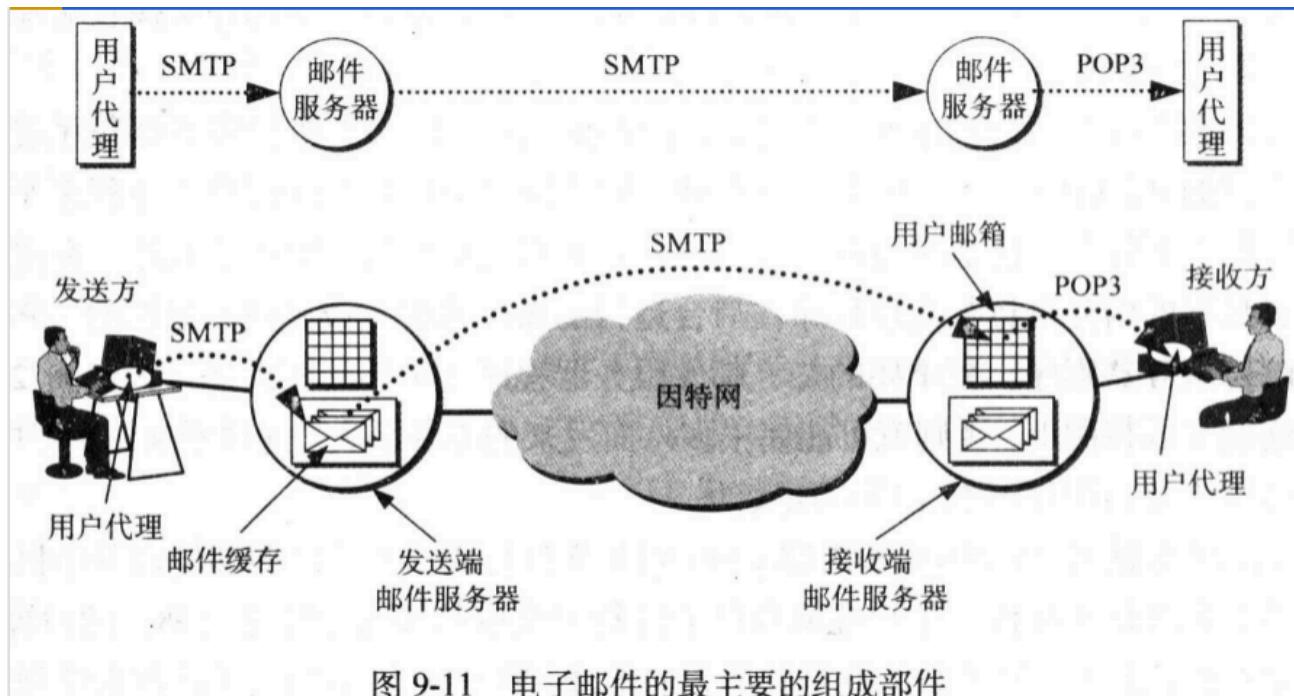
Fig K-2. Example of a resolver looking up a remote name in 10 steps.

- Another query method: 先问根服务器，根服务器找不到，根服务器会告诉你去顶级服务器那里去找；同理，某级服务器找不到，他不会帮你去找，只会告诉你去哪里找。

2 E-Mail

2.1 Architecture and Services

一些服务了解一下



- A首先把邮件发送给SMTP服务器（邮箱会设定的），SMTP服务器找到对方的邮箱，B用POP3收走
- SMTP用两次，POP3用三次

2.2 Message Formats

PPT上讲到了一些邮件内容的形式，其中首部还有一些字段的格式。有些字段的值必须要有，有些不需要。

2.2.1 MIME

原始的邮件都是ASCII编码，而汉字、图片等等是非ASCII编码（二进制编码，MIME用base-64编码）。

Basic idea: Still use the RFC 822 format, but to add structure to the message body and define encoding rules for non-ASCII messages. 会在主题部分定义编码类型。

而且MIME还会增加不同的头部来指示。

Type	Subtype	Description
Text	Plain	Unformatted text
	Enriched	Text including simple formatting commands
Image	Gif	Still picture in GIF format
	Jpeg	Still picture in JPEG format
Audio	Basic	Audible sound
Video	Mpeg	Movie in MPEG format
Application	Octet-stream	An uninterpreted byte sequence
	Postscript	A printable document in PostScript
Message	Rfc822	A MIME RFC 822 message
	Partial	Message has been split for transmission
	External-body	Message itself must be fetched over the net
Multipart	Mixed	Independent parts in the specified order
	Alternative	Same message in different formats
	Parallel	Parts must be viewed simultaneously
	Digest	Each part is a complete RFC 822 message

Fig K-7. The MIME types and subtypes defined in RFC 2045.

- Multipart功能比较神奇，它是二者选其一，比如下面的那封邮件，收件人如果有音响那就跳转放歌，否则就是显示文本，二者选一实现。

```

From: elinor@abcd.com
To: carolyn@xyz.com
MIME-Version: 1.0
Message-ID: <0704760941.AA00747@abcd.com>
Content-Type: multipart/alternative; boundary=qwertyuopasdfghjklzxcvbnm
Subject: Earth orbits sun integral number of times

This is the preamble. The user agent ignores it. Have a nice day.

--qwertyuopasdfghjklzxcvbnm
Content-Type: text/enriched

Happy birthday to you
Happy birthday to you
Happy birthday dear <b>Carolyn</b> Carolyn </b>
Happy birthday to you

--qwertyuopasdfghjklzxcvbnm
Content-Type: message/external-body;
access-type="anon-ftp";
site="bicycle.abcd.com";
directory="pub";
name="birthday.snd"

content-type: audio/basic
content-transfer-encoding: base64
--qwertyuopasdfghjklzxcvbnm--

```

Fig K-8. A multipart message containing enriched and audio alternatives.

2.2.1.1 base64 (补充)

(补充)

- base64编码方法如下：
 - 先将24bit的代码划分为4个6位组。
 - 6bit组的二进制代码共有64种不同的值，从0到63。
 - 用B表示1，等等。26个大写字母排列完毕后，接下去再排26个小写字母，再后面是10个数字，最后用“+”表示62，而用“/”表示63。再用两个连在一起的等号“==”和一个等号“=”分别表示最后一组的代码只有8或16比特。
 - A-00,B-01,C-02,D-03,E-04,F-05,G-06,H-07,I-08,J-09,K-10,L-11,M-12,N-13,O-14,P-15,Q-16,R-17,S-18,T-19,U-20,V-21,W-22,X-23,Y-24,Z-25, a-26,b-27,c-28,d-29,e-30,f-31,g-32,h-33,i-34,j-35,k-36,l-37,m-38,n-39,o-40,p-41,q-42,r-43,s-44,t-45,u-46,v-47,w-48,x-49,y-50,z-51, 0-52,1-53,2-54,3-55,4-56,5-57,6-58,7-59,8-60,9-61,+-62,/-63
 - 回车和换行都忽略，它们可在任何地方插入。
 - 作为base64编码的例子，假设有二进制代码，共24bit: 01001001 00110001 01111001。先划分为4个6bit组，即010010 010011 000101 111001，对应的十进制值为18,19,5,57。对应的base64编码为： STF5。

25

- 将三个bytes的编码写成4个6bit的，6bit的有63种字符。写成4位的符号。
- 回车符CR，换行符LF
- Base64要求把每三个8Bit的字节转换为四个6Bit的字节（ $3 * 8 = 4 * 6 = 24$ ），然后把6Bit再添两位高位0，组成四个8Bit的字节，也就是说，转换后的字符串理论上将要比原来的长1/3。3bytes变成4bytes。

2.3 Message Transfer

SMTP工作过程

- SMTP – Simple Mail Transfer Protocol
 - A simple ASCII protocol
 - port: 25
 - TCP connection
- A simple dialog for sending the message of Fig.7-13. -[see fig k-9]
 - HELO
 - MAIL TO
 - RCPT TO
 - DATA
 - QUIT
- telnet ema.zju.edu.cn 25
- or telnet <cr>
 - help
 - set local_echo
 - open ema.zju.edu.cn 25
- ESMTP (extended SMTP): a new protocol to get around some problems of SMTP.

```

S: 220 xyz.com SMTP service ready
C: HELO abcd.com
    S: 250 xyz.com says hello to abcd.com
C: MAIL FROM: <elinor@abcd.com>
    S: 250 sender ok
C: RCPT TO: <carolyn@xyz.com>
    S: 250 recipient ok
C: DATA
    S: 354 Send mail; end with "." on a line by itself
C: From: elinor@abcd.com
C: To: carolyn@xyz.com
C: MIME-Version: 1.0
C: Message-Id: <0704760941.AA00747@abcd.com>
C: Content-Type: multipart/alternative; boundary=qwertyuiopasdfghjklzxcvbnm
C: Subject: Earth orbits sun integral number of times
C:
C: This is the preamble. The user agent ignores it. Have a nice day.
C:
C: --qwertyuiopasdfghjklzxcvbnm
C: Content-Type: text/enriched
C:
C: Happy birthday to you
C: Happy birthday to you
C: Happy birthday dear <bold> Carolyn </bold>
C: Happy birthday to you
C:
C: --qwertyuiopasdfghjklzxcvbnm
C: Content-Type: message/external-body;
C:     access-type="anon-ftp";
C:     site="bicycle.abcd.com";
C:     directory="pub";
C:     name="birthday.snd"
C:
C: content-type: audio/basic
C: content-transfer-encoding: base64
C: --qwertyuiopasdfghjklzxcvbnm
C: .
    S: 250 message accepted
C: QUIT
    S: 221 xyz.com closing connection

```

Fig k-9. Transferring a message from *elinore@abc.com* to *carolyn@xyz.com*.

27

- 先进行TCP三次握手协议，建立连接
- 先发HELO
- 然后MAIL FROM，告知服务器我自己是谁
- RCPT TO告知我要发给谁，Server会发OK，告知你可以发
- DATA告知服务器，我要发数据了
- 最后Client发送QUIT，连接释放

2.4 Final Delivery 接收

- Final Delivery – solving the message delivery problem for machines that are not on the Internet.
- POP3
 - Post Office Protocol(POP3) --- (RFC 1225), [[fig K-10](#)]
 - Server port: 110
 - A simple protocol used for fetching email from a remote mailbox
 - mail on server usually are copied to client machine and deleted from server
 - POP 3 dialog between client and server, [[fig k-11](#)]
 - telnet mail.isp.com 110
 - RETR: retrieve message.
 - DELE: delete a message

	S: +OK POP3 server ready
C: USER carolyn	S: +OK
C: PASS vegetables	S: +OK login successful
C: LIST	S: 1 2505 S: 2 14302 S: 3 8122 S: .
C: RETR 1	S: (sends message 1)
C: DELE 1	
C: RETR 2	S: (sends message 2)
C: DELE 2	
C: RETR 3	S: (sends message 3)
C: DELE 3	
C: QUIT	S: +OK POP3 server disconnecting

Fig K-11. Using POP3 to fetch three messages.

- 服务器和Client建立TCP连接
- 连接建立后， Client会给服务器发送USER命令和PASS命令，相当于在服务器上做log in操作。这

是明文的，实际上很不安全。

- 发送LIST命令，知道总共有多少邮件。
- 然后一封一封邮件收下来。
- 最后连接释放。

3 The World Wide Web

3.1 Architectural Overview

The basic model: a page on server abcd.com can contain a hyperlink to a page on the xyz.com server.

▪ The Client Side

- Web consists of a vast ,worldwide collection of documents
- Hypertext: be consist of normal text and text with hyper linker 超文本（有格式的文本，还有超媒体hypermedia）
- Browser: Mosaic、Netscape、IE 浏览器
- Hyperlink: Text string that links to other pages 超链接
- Most Web pages have dynamic content 动态网页
 - Significant amount of the content remains static 部分内容是静态网页一直存在那里的
 - 有些是动态网页，是临时生成的一个网页，比如根据客户端提交的信息提供一个网页
- Web page design
 - Use HTML or CSS (Cascading Style Sheets, 层叠样式表)
 - Use programs (Adobe Dreamweaver)

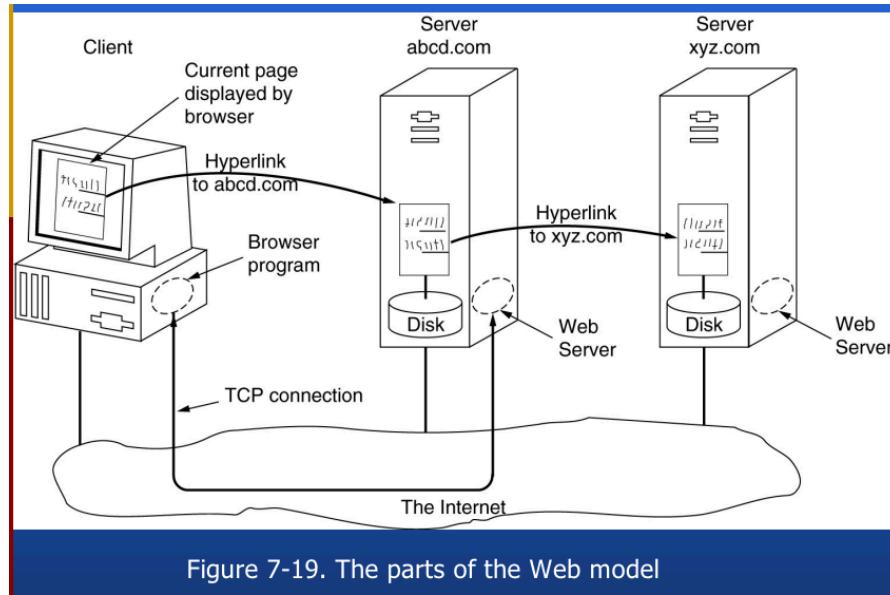


Figure 7-19. The parts of the Web model

- 上图是超链接的工作情况

3.1.1 浏览器工作流程

- When a user click on a hyperlink, <http://www.itu.org/home/index.html>, the browser carries out a series of steps in order to fetch the page pointed to.
 - 1. The browser determines the URL
 - 2. The browser asks DNS for the IP address of www.itu.org
 - 3. DNS replies with 156.106.192.32
 - 4. THE BROWSER MAKES A TCP CONNECTION TO PORT 80 ON 156.106.192.32
 - 5. It then sends over a request asking for file /home/index.html
 - 6. The www.itu.org server sends the file /home/index.html.
 - 7. The TCP connection is released.
 - 8. The browser displays all the text in /home/index.html.
 - 9. The browser displays all images in this file.

- 1 解析URL
- 2 浏览器会向DNS服务器找域名主机的IP地址
- 3 DNS服务器会返回IP地址
- 4 浏览器选择一个端口号和目标域名主机建立TCP连接
- 5 浏览器发出 GET / POST请求，如果访问到一个html里面有图片文件，浏览器会自动发送GET 请求
- 6 WEB服务器返回
- 7 TCP 连接释放

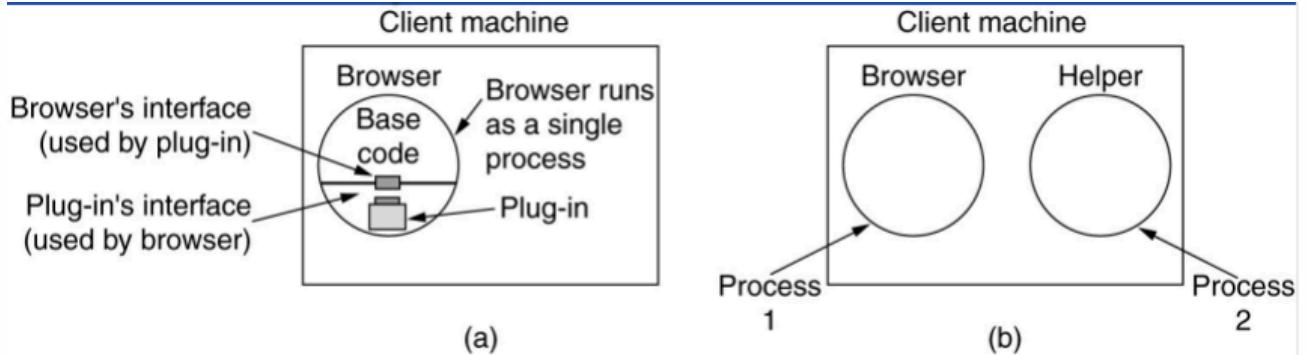


Figure 7-20. (a) A browser plug-in. (b) A helper application.

如果一个网页包含 PDF, GIF, Jpeg这样的文件，浏览器可以用插件展示（plug-in） / 或者调用其他软件（helper application）

另外，如果浏览器把获取的所有东西都执行，那么有安全漏洞，可以勾选选项（如禁用 ActiveX），使得浏览器不自动执行文件

3.1.2 提高WEB性能

- 应用Cache作为硬盘的补充
- 多线程作并发服务器
- 集群服务器

3.1.3 URL

- For example, <http://www.cs.vu.nl/video/index-en.html>, consists of:
 - Protocol, for example, http
 - The DNS name of the machine on which the page is located
 - A file path name uniquely indicating the specific page
- An empty file name defaults to the site's main home page.
- When the file named is a directory, this implies a file named *index.html*, default.htm, etc.
- Common protocols in URL:

Name	Used for	Example
http	Hypertext (HTML)	http://www.cs.vu.nl/~ast/
ftp	FTP	ftp://ftp.cs.vu.nl/pub/minix/README
file	Local file	file:///usr/suzanne/prog.c
news	Newsgroup	news:comp.os.minix
news	News article	news:AA0134223112@cs.utah.edu
gopher	Gopher	gopher://gopher.tc.umn.edu/11/Libraries
mailto	Sending e-mail	mailto:JohnUser@acm.org
telnet	Remote login	telnet://www.w3.org:80

- 协议 + DNS域名 + 文件路径

```

1 protocol://hostname:port/path
2 http://127.0.0.1/path1
3 ftp://127.0.0.1/ftp_path1

```

3.1.4 Cookie

浏览器会记录一些用户输入的用户名、密码之类的东西，一些小文件

3.2 Static Web Objects

Tag	Description
<html> ... </html>	Declares the Web page to be written in HTML
<head> ... </head>	Delimits the page's head
<title> ... </title>	Defines the title (not displayed on the page)
<body> ... </body>	Delimits the page's body
<h n> ... </h n>	Delimits a level <i>n</i> heading
 ... 	Set ... in boldface
<i> ... </i>	Set ... in italics
<center> ... </center>	Center ... on the page horizontally
 ... 	Brackets an unordered (bulleted) list
 ... 	Brackets a numbered list
	Starts a list item (there is no)
 	Forces a line break here
<p>	Starts a paragraph
<hr>	Inserts a Horizontal rule
	Displays an image here
 ... 	Defines a hyperlink

Figure M-2 A selection of common HTML tags.
Some can have additional parameters.

```

1  <HTML>
2    <HEAD>      <TITLE> AMALGAMATED WIDGET, INC. </TITLE>
3    </HEAD>
4  <BODY>
5    <H1> Welcome to AWI's Home Page </H1>
6    <IMG SRC=http://WWW.Widget.com/images/logo.gif ALT="AWI Logo"> <BR>
7    We are so happy that you have Chosen to Visit <B> Amalgamated
8    Widget's </B> home page. We hope <I> you </I> Will find all the
9    information you need here.
10   <P>Below we have links to information about our many fine products.
11   You can order electronically (by WWW), by telephone, or by fax.<HR>
12   <H2> Product information </H2>
13   <UL><LI><A HREF="http://widget.com/products/big"> Big widgets </A>
14     <LI><A HREF="http://www.widget.com/products/little"> Little
15     Widgets </A>
16   </UL>
17   <H2> Telephone numbers </H2>
18   <UL> <LI> By telephone: 1-800-WIDGETS
19     <LI> By fax: 1-415-765-4321
20   </UL>
21 </BODY>
22 </HTML>

```

- 图片的表示形式 SRC
- 超链接表示形式 HREF
- 下面是效果

Welcome to AWI's Home Page



We are so happy that you have chosen to visit **Amalgamated Widget's** home page. We hope *you* will find all the information you need here.

Below we have links to information about our many fine products. You can order electronically (by WWW), by telephone, or by FAX.

Product Information

- Big widgets
- Little widgets

Telephone numbers

Figure M-1(b)

- 1-800-WIDGETS
- 1-415-765-4321

3.3 Dynamic Web Documents

3.3.1 动态网页的表单和流程

FORM(表单)

- An order form -[see fig M-4]
- Starting in HTML 2.0
- Contain boxes or buttons that allow users to fill in information or make choices and then send the information back to the page's owner
- When using POST method, the format of data that FORM in client sends to server is: "&"---separating multiple variables, "+" --- representing space character, example data of fig.7-29:
 - *customer=John+Doe & address=100+Main+st. & city=White+Plains & state=NY & country=USA & cardno=1234567890 & expires=6/98 & cc=mastercard & product=cheap & express=on*
- 定义一个接口，即通用网关接口CGI，用来使万维网服务器将浏览器发来的数据传送给这个应用程序，然后万维网服务器能够解释这个应用程序的输出，并向浏览器返回HTML文档。
 - GET--- CGI程序从Request.QueryString获取上传数据
 - POST--- CGI程序从Request.Form获取上传数据

- 就是列出一些空，让用户填进去，然后传送给处理程序，生成一个动态网页

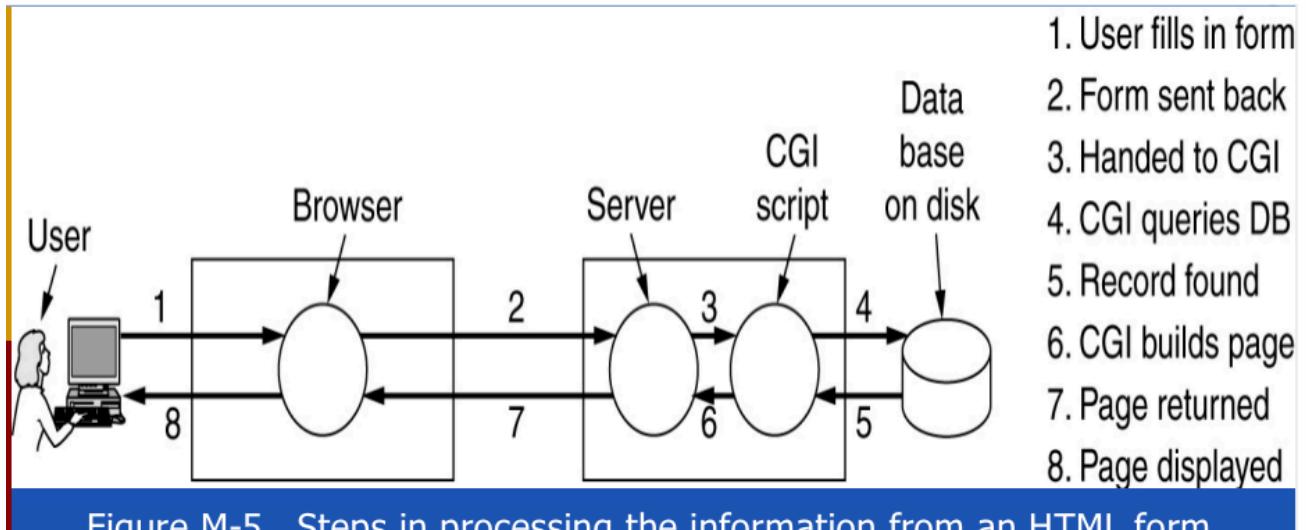


Figure M-5. Steps in processing the information from an HTML form.

- 比如淘宝提交购物车，提交后查询各个商品的定价，然后新生成一个付款界面

3.3.2 动态网页后端

3.3.2.1 PHP

Figure M-7 .

- (a) A Web page containing a form.
- (b) A PHP script for handling the output of the form.
- (c) Output from the PHP script when the inputs are "Barbara" and 24 respectively

```
<html>
<body>
<form action="action.php" method="post">
<p> Please enter your name: <input type="text" name="name"> </p>
<p> Please enter your age: <input type="text" name="age"> </p>
<input type="submit">
</form>
</body>
</html>
```

(a)


```
<html>
<body>
<h1> Reply: </h1>
Hello <?php echo $name; ?>.
Prediction: next year you will be <?php echo $age + 1; ?>
</body>
</html>
```

(b)


```
<html>
<body>
<h1> Reply: </h1>
Hello Barbara.
Prediction: next year you will be 25
</body>
</html>
```

(c)

- 一个很简单的动态网页

3.3.2.2 JSP - Java Server Pages

3.3.2.3 ASP - Active Server Pages

3.3.3 Client-Side Dynamic Web Page Generation

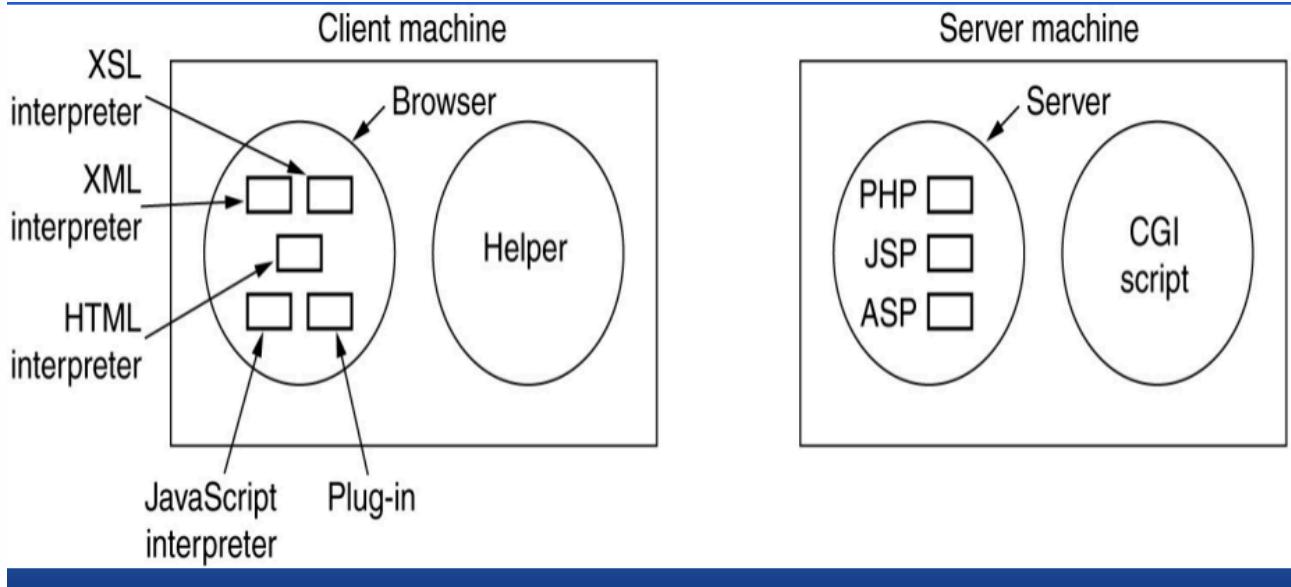
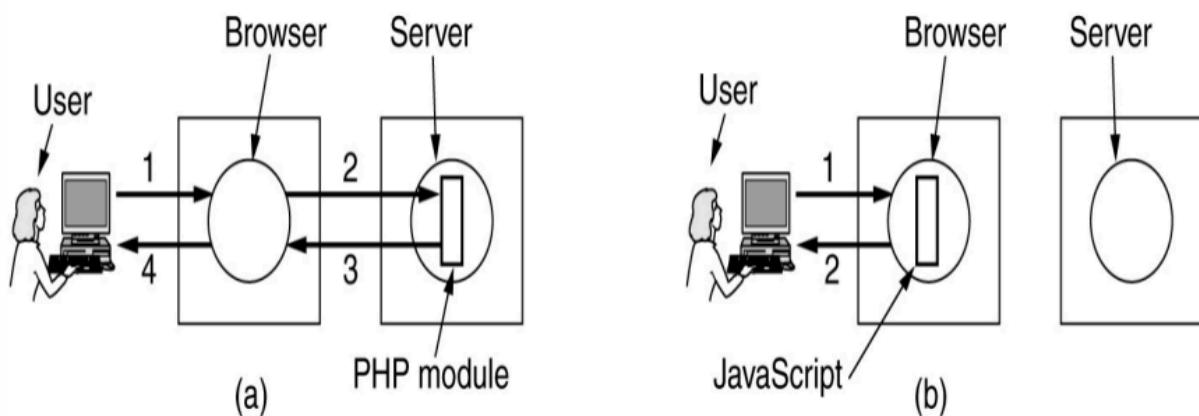


Figure M-11. The various ways to generate and display content

Figure M-9.(a) Server-side scripting with PHP.
(b) Client-side scripting with JavaScript

- PHP是后端处理好了，传送给浏览器
- JavaScript是把程序给浏览器，让浏览器处理

3.4 HTTP and HTTPS

HTTP可以用MIME编码、RFC、甚至二进制编码，传输层用TCP，可靠传输。

```
Trying 4.17.168.6...
Connected to www.ietf.org.
Escape character is '^].
HTTP/1.1 200 OK
Date: Wed, 08 May 2002 22:54:22 GMT
Server: Apache/1.3.20 (Unix) mod_ssl/2.8.4 OpenSSL/0.9.5a
Last-Modified: Mon, 11 Sep 2000 13:56:29 GMT
ETag: "2a79d-c8b-39bce48d"
Accept-Ranges: bytes
Content-Length: 3211
Content-Type: text/html
X-Pad: avoid browser bug

<html>
<head>
<title>IETF RFC Page</title>

<script language="javascript">
function url() {
    var x = document.form1.number.value
    if (x.length == 1) {x = "000" + x}
    if (x.length == 2) {x = "00" + x}
    if (x.length == 3) {x = "0" + x}
    document.form1.action = "/rfc/rfc" + x + ".txt"
    document.form1.submit
}
</script>
</head>
```

MAC帧+IP头+TCP头+HTTP首部+HTTP数据（html文件）。

HTTPS是HTTP的安全版，其在传输层运用的是TLS协议

3.4.1 request

Figure 7-25

The built-in HTTP request methods.

Method	Description
GET	Read a Web page
HEAD	Read a Web page's header
POST	Append to a Web page
PUT	Store a Web page
DELETE	Remove the Web page
TRACE	Echo the incoming request
CONNECT	Connect through a proxy
OPTIONS	Query options for a page

- HEAD 是获取HTTP的头部
- POST是更改web网页，一般不要同意，DELETE同理
- 还有一些请求

3.4.2 response

Code	Meaning	Examples
1xx	Information	100 = server agrees to handle client's request
2xx	Success	200 = request succeeded; 204 = no content present
3xx	Redirection	301 = page moved; 304 = cached page still valid
4xx	Client error	403 = forbidden page; 404 = page not found
5xx	Server error	500 = internal server error; 503 = try again later

Figure 7-26. The status code response groups

- 100: 同意连接
- 200: 请求成功； 204:无内容。
- 301: 页面跳转； 304:cache中的内容仍有用，继续执行。
- 403: 禁止访问； 404:页面找不到。 (client problem)
- 500: 服务器内部错误； 503:稍后尝试。 (server problem)

4 Content Delivery

4.1 caching

访问过的网页放在本地中。

可以放在本机，存了浏览器的缓存。

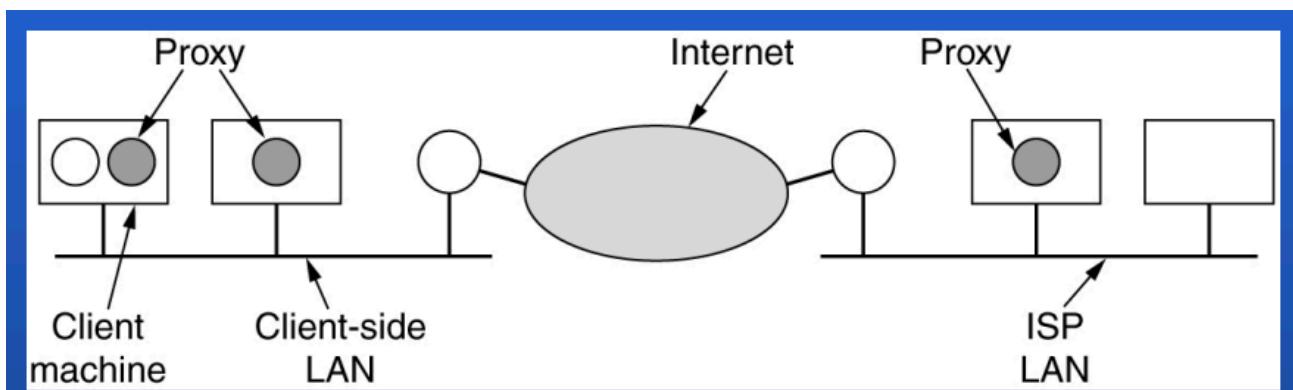


Figure M-13.Hierarchical caching with three proxies

- 1 也可以用代理服务器，本地代理服务器缓存了局域网LAN内经常访问的一些网页，可以节约国干网上的流量，本机访问代理服务器会比直接访问目标主机快很多。

ISP代理服务器是互联网上的，互联网上的运营商提供代理服务。

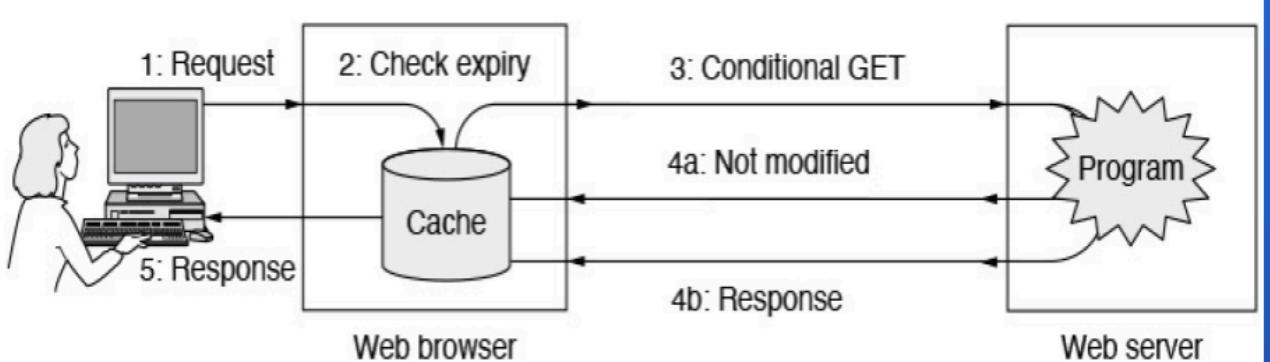


Figure 7-28
HTTP caching.

但是也有些网页的时效性要求很高。可以在HTTP头中加一个 `If-modified-since` 请求，在cache中会保存一个网页过期时间，如果访问时间还没过期，那就正常显示页面；如果已经过期了，那么cache就会发送条件请求，如果web服务器说还没过期，那就返回not modified，如果更新就回response。这样一来，保证用户看的一直都是最新的。

4.2 server replication 镜像站点

缓存是一种客户端技术。

提高性能的服务器端技术：镜像站点。

4.3 负载均衡



- 一个DNS域名可以对应多个服务器，每个服务器有不同的IP地址
- www.apusapp.com IN A 114.100.20.201;
www.apusapp.com IN A 114.100.20.202;
www.apusapp.com IN A 114.100.20.203;

- 如果一个服务器崩溃了，DNS服务器可以返回其他的IP地址，以供用户使用
- 这样就可以提高用户使用效率，降低出错率

4.4 content delivery network

- 将客户端的请求重定向到最近的CDN服务器
- 工作模式：
 - 小页面（仅HTML文本）存储在内容提供商的服务器上
 - 大型文件（图像、音频、视频）由CDN分发。
 - CDN将内容提供商的有价值内容放在许多ISP的服务器上，并为此付费。
- CDN预处理内容提供商的页面[fig.M-14]，将较大文件的hyperlink修改到CDN的假服务器，CDN的假服务器使用状态301（页面移动）和位置响应头将这些访问重定向到ISP的服务器。

就是把较大的一些文件都放在一个专属服务器上，需要访问的时候，给提示给其访问。

