

接着,该应用程序对浏览器发来的数据进行处理,并输出 HTTP 格式的文档,万维网服务器把应用程序的输出作为对浏览器的响应。由于对浏览器每次请求的响应都是临时生成的,因此用户通过动态文档所看到的内容是不断变化的。动态文档的主要优点是具有报告当前最新信息的能力。例如,动态文档可用来报告股市行情、天气预报或民航售票情况等内容。但动态文档的创建难度比静态文档的高,因为动态文档的开发不是直接编写文档本身,而是编写用于生成文档的应用程序,这就要求动态文档的开发人员必须会编程,而所编写的程序还要通过大范围的测试,以保证输入的有效性。

动态文档的一个例子是我们使用携程网([www.ctrip.com](http://www.ctrip.com))购买机票。当我们打开携程网时,便看到了携程网服务器的网页,这就是一种活动文档。例如,我们选择国内机票,当我们键入出发地点、到达地点、日期、人数、舱位等信息后,携程网的服务器就可以根据你键入的数据,生成出符合你需要的各航空公司的航班表。这种航班表不是静态的,而是根据用户的需求动态生成的。当某个航班的机票已经售完时,动态航班表也会显示出某个航班的机票已经无法购买了。

✓【6-13】 浏览器同时打开多个 TCP 连接进行浏览的优缺点如何?请说明理由。

解答:浏览器同时打开多个 TCP 连接进行浏览的优点是可以同时下载好几个对象(文件或图片),加快了下载的速度。然而,由于计算机连接到网络的线路的数据率是受限的,几个下载的数据率的总和不能超过连接到网络的线路的数据率。因此,浏览器同时打开多个 TCP 连接进行浏览,有时并不能带来太多的好处。

【6-14】 请判断以下论述的正误,并简述理由。

(1) 用户点击某网页,该网页有 1 个文本文件和 3 张图片。此用户可以发送一个请求就可以收到 4 个响应报文。

(2) 有以下两个不同的网页: [www.abc.com/m1.html](http://www.abc.com/m1.html) 和 [www.abc.com/m2.html](http://www.abc.com/m2.html)。用户可以使用同一个 HTTP/1.1 持续连接传送对这两个网页的请求和响应。

(3) 在客户与服务器之间进行非持续连接,只需要用一个 TCP 报文段就能够装入两个不同的 HTTP 请求报文。

(4) 在 HTTP 响应报文中的主体实体部分永远不会是空的。

解答:(1) 错误。对于非持续 HTTP,需要使用 4 个 TCP 连接分别来发送这 4 个(请求和响应)。对于持续 HTTP,可以在一个 TCP 连接连续传送 4 个(请求和响应)。

(2) 正确。显然,这两个网页处在同一个服务器上(两个网页的域名是一样的)。如果用户使用 HTTP/1.1 持续连接,那么可以在这个连接上传送这两个网页。

(3) 错误。对于客户与服务器之间的非持续连接,每一个新的 HTTP 请求报文必须使用一个新的 TCP 连接。

(4) 错误。在某些情况下,服务器无法找到客户所请求的文件。这时,服务器返回的响应的主体实体部分就是空的。这时,在 HTTP 响应报文的行中会返回一个状态码,例如 404。

✓【6-15】 假定你在浏览器上点击一个 URL,但这个 URL 的 IP 地址以前并没有缓存在本地主机上。因此需要用 DNS 自动查找和解析。假定要解析到所要找的 URL 的





IP 地址共经过  $n$  个 DNS 服务器, 所经过的时间分别为  $RTT_1, RTT_2, \dots, RTT_n$ 。假定从要找的网页上只需要读取一张很小的图片(即忽略这张小图片的传输时间)。从本地主机到这个网页的往返时间是  $RTT_w$ 。试问从点击这个 URL 开始, 一直到本地主机的屏幕上出现所读取的小图片, 一共要经过多少时间?

**解答:** 解析 IP 地址需要的时间是:  $RTT_1 + RTT_2 + \dots + RTT_n$ 。

建立 TCP 连接和请求万维网文档需要  $2RTT_w$ 。

需要的总时间是:  $2RTT_w + RTT_1 + RTT_2 + \dots + RTT_n$ 。

**【6-16】** 在上题中, 假定同一台服务器的 HTML 文件中又链接了三个非常小的对象。若忽略这些对象的发送时间, 试计算客户点击读取这些对象所需的时间。

- (1) 没有并行 TCP 连接的非持续 HTTP;
- (2) 使用并行 TCP 连接的非持续 HTTP;
- (3) 流水线方式的持续 HTTP。

**解答:** 分别计算如下:

- (1) 所需时间 =  $RTT_1 + RTT_2 + \dots + RTT_n$  (解析 IP 地址)  
 +  $2RTT_w$  (建立 TCP 连接和读取 HTML 文件)  
 +  $3(2RTT_w)$  (读取三个对象)  
 =  $RTT_1 + RTT_2 + \dots + RTT_n + 8RTT_w$

$8RTT_w$  的图解如图 T-6-16(a)所示。

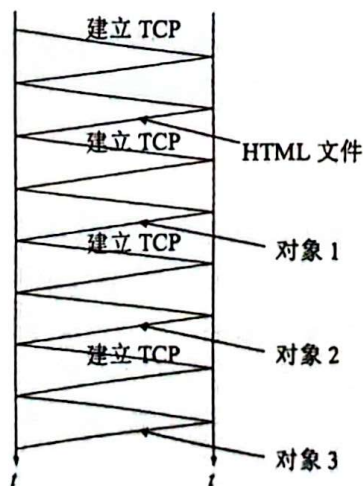


图 T-6-16(a)  $8RTT_w$  的图解

- (2) 所需时间 =  $RTT_1 + RTT_2 + \dots + RTT_n$  (解析 IP 地址)  
 +  $2RTT_w$  (建立 TCP 连接和读取 HTML 文件)  
 +  $2RTT_w$  (并行地建立 TCP 连接和并行地读取三个对象)  
 =  $RTT_1 + RTT_2 + \dots + RTT_n + 4RTT_w$

$4RTT_w$  的图解如图 T-6-16(b)所示。



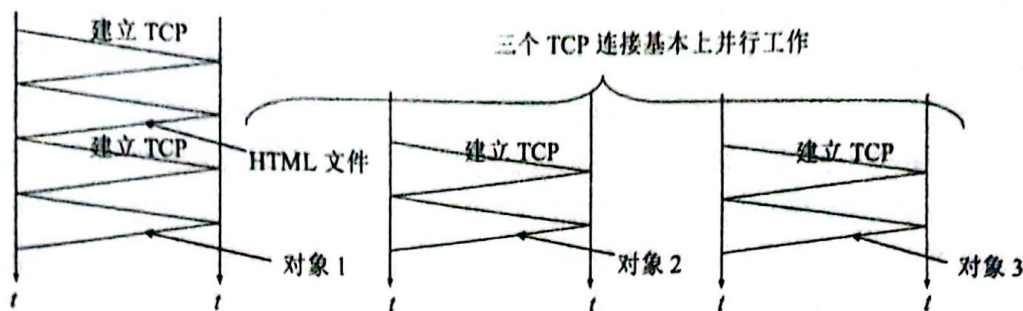


图 T-6-16(b)  $4RTT_w$  的图解

- (3) 所需时间 =  $RTT_1 + RTT_2 + \dots + RTT_n$  (解析 IP 地址)  
 +  $2RTT_w$  (建立 TCP 连接和读取 HTML 文件)  
 +  $RTT_w$  (连续读取三个对象)  
 =  $RTT_1 + RTT_2 + \dots + RTT_n + 3RTT_w$

$3RTT_w$  的图解如图 T-6-16(c)所示。

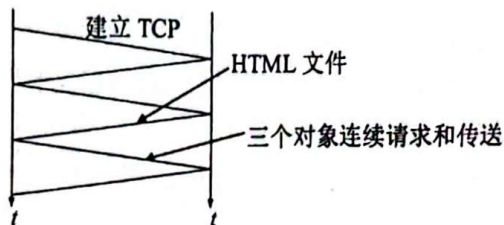


图 T-6-16(c)  $3RTT_w$  的图解

【6-17】在浏览器中应当有几个可选解释程序。试给出一些可选解释程序的名称。

解答：在浏览器中的解释程序数目并无明确规定，但一般的浏览器都有 HTML 解释程序和 Java 小应用程序解释程序。

【6-18】一个万维网网点有 1000 万个页面，平均每个页面有 10 个超链。读取一个页面平均要 100 ms。请问：要检索整个网点所需的最少时间是多少？

解答：依题意，一个页面上有 10 个超链，和本题并无关系，因为题目并未指出是否还要点击这 10 个超链（也没有给出点击一个超链需要多少时间），以及是否要在点击超链后再继续点击下去，等等。本题实际上就是问，读取这 1000 万个网页需要多少时间。

既然读取一个页面平均要 100 ms，那么读取 1000 万个页面，就需要的时间为：

$$T = 10^7 \times 100 \times 10^{-3} = 10^6 \text{ s} \quad \text{即约 11.6 天。}$$

【6-19】搜索引擎可分为哪两种类型？各有什么特点？

解答：搜索引擎的种类很多，但大体上可划分为两大类，即全文检索搜索引擎和分类目录搜索引擎。

全文检索搜索引擎是一种纯技术型的检索工具。它的工作原理是通过搜索软件（例如一种叫作“蜘蛛”或“网络机器人”的 Spider 程序）到互联网上的各网站收集信息，找到一个网站后，又可以从这个网站再链接到另一个网站，像蜘蛛爬行一样。然后按照一定的规则建立一个很大的在线数据库供用户查询。用户在查询时只要输入关键词，就从已经建立的索引数据库