

多应用程序包括许多文件共享系统，在下载期间其用户界面显示了其瞬时吞吐量，也许你已经观察过它！) 如果该文件由 F 比特组成，主机 B 接收到所有 F 比特用去 T 秒，则文件传送的平均吞吐量 (average throughput) 是 F/T bps。对于某些应用程序如因特网电话，希望具有低时延和在某个阈值之上的一致的瞬时吞吐量。例如，对某些因特网电话是超过 24kbps，对某些实时视频应用程序是超过 256kbps。对于其他应用程序，包括涉及文件传送的那些应用程序，时延不是至关重要的，但是希望具有尽可能高的吞吐量。

为了进一步深入理解吞吐量这个重要概念，我们考虑几个例子。图 1-19a 显示了服务器和客户这两个端系统，它们由两条通信链路和一台路由器相连。考虑从服务器传送一个文件到客户的吞吐量。令 R_s 表示服务器与路由器之间的链路速率； R_c 表示路由器与客户之间的链路速率。假定在整个网络中只有从这台服务器到那台客户的比特在传送。在这种理想的情况下，我们现在问该服务器到客户的吞吐量是多少？为了回答这个问题，我们可以想象比特是流体，通信链路是管道。显然，这台服务器不能以快于 R_s bps 的速率通过其链路注入比特；这台路由器也不能以快于 R_c bps 的速率转发比特。如果 $R_s < R_c$ ，则由该服务器注入的比特将顺畅地通过路由器“流动”，并以速率 R_s bps 到达客户，给定了 R_s bps 的吞吐量。在另一方面，如果 $R_c < R_s$ ，则该路由器将不能够以接收它们那样快的速率来转发比特。在这种情况下，比特将以速率 R_c 离开该路由器，从而得到端到端吞吐量 R_c 。(还要注意的，如果比特继续以速率 R_s 到达该路由器，继续以 R_c 离开路由器的话，在该路由器中等待传输给客户的积压比特将不断增加，这是一种非常不希望的情况！) 因此，对于这种简单的两链路的网络，其吞吐量是 $\min\{R_c, R_s\}$ ，这就是说，它是瓶颈链路 (bottleneck link) 的传输速率。在决定了吞吐量之后，我们现在近似地得到从服务器到客户传输一个 F 比特的大文件所需要的时间是 $F/\min\{R_c, R_s\}$ 。对于一个特定的例子，假定你正在下载一个 $F = 32 \times 10^6$ 比特的 MP3 文件，服务器具有 $R_s = 2\text{Mbps}$ 的传输速率，并且你有一条 $R_c = 1\text{Mbps}$ 的接入链路。则传输该文件所需的时间是 32 秒。当然，这些吞吐量和传送时间的表达式仅是近似的，因为它们并没有考虑分组层次和协议的问题。

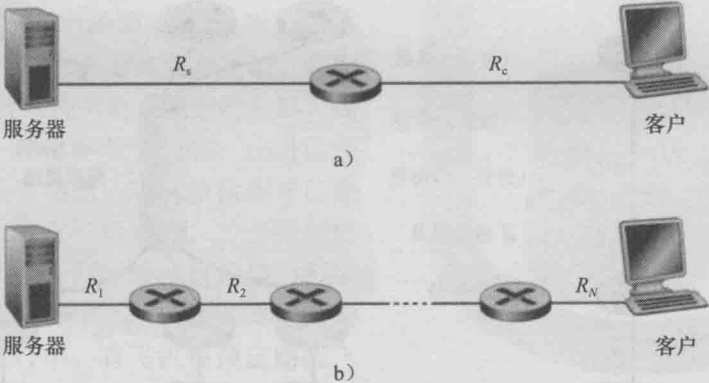


图 1-19 一个文件从服务器传送到客户的吞吐量

图 1-19b 此时显示了在服务器和客户之间具有 N 条链路的一个网络，这 N 条链路的传输速率分别是 R_1, R_2, \dots, R_N 。应用与对两条链路网络的分析相同的方法，我们发现从服务器到客户的文件传输的吞吐量是 $\min\{R_1, R_2, \dots, R_N\}$ ，这同样仍是沿着服务器和客户之间路径的瓶颈链路的速率。