

意到,为了预测两个网络端点之间的应用级性能,必须处理下列问题,并因此提供充足的容量来满足应用的性能要求。

- 网络端点之间的流量要求模型。这些模型可能需要定义在呼叫层次(例如,用户“到达”网络并启动端到端应用)和分组层次(例如,由进行中的应用所产生的分组)。注意负载可能随着时间而变化。
- 定义良好的性能要求。例如,为支持诸如会话式多媒体应用等时延敏感的流量,其性能要求可能是:分组的端到端时延大于最大可容忍时延的概率要小于某个很小的值 [Frleigh 2003]。
- 对给定的负载模型预测端到端性能的模型,以及求出最小成本带宽分配(该带宽分配将导致满足所有用户的需求)的技术。这里,研究人员正忙于研发能够量化给定负载下的性能的性能模型,以及能求出满足性能要求的最小成本带宽分配的优化技术。

假定今天尽力而为的因特网能够(从技术角度讲)以适当的性能水平支持多媒体流量(如果它被定制成这样的话),自然的问题是为什么今天的因特网满足不了这样的要求。答案基本上是经济上和组织上的原因。从经济角度看,用户将愿意向其 ISP 支付足够多的费用,使 ISP 安装充足的带宽经尽力而为的因特网来支持多媒体应用吗?组织问题也许更为令人气馁。注意到在两个多媒体端点之间的端到端路径将通过多个 ISP 的网络。从组织角度看,这些 ISP 将愿意合作(也许以收入共享方式)以确保端到端路径被适当地定制来支持多媒体应用吗?对这些经济和组织问题的展望,参见 [Davies 2005]。对于供给第一层主干网络以支持时延敏感流量的展望,参见 [Frleigh 2003]。

7.5.2 提供多种类型的服务

也许对今天因特网中的以不变应万变的尽力而为服务而言,一种最简单的强化是将流量划分为多种类型,并为这些不同类型的流量提供不同等级的服务。例如,某 ISP 可能要为时延敏感的 VoIP 或电信会议流量比为电子邮件或 HTTP 等弹性流量提供更高的服务类型(并对该服务收取更高的费用!)。另一种做法是,ISP 可能直接向愿意对这种改进服务支付更多费用的顾客提供更高质量的服务。一些住宅有线接入 ISP 和蜂窝无线接入 ISP 已经采用了这样的梯次等级服务,即铂金卡服务用户比金卡服务用户或银卡服务用户享有更好的服务性能。

我们都从日常生活中熟悉了不同类型的服务,如航班上头等舱乘客比公务舱乘客得到更好的服务,公务舱乘客又比经济舱乘客得到更好的服务;VIP 在活动中能够立即进入,而所有其他人都必须排队等待;在某些国家中老年人被尊重,提供了荣誉座位和最精细的食物。重要的是注意到在聚合流量中(即在多种流量类型中而不是单个连接中)提供了这种有差别的服务。例如,所有头等舱乘客被一视同仁(没有哪个头等舱乘客得到了比其他头等舱乘客更好的服务),就像所有的 VoIP 分组在网络中得到了相同的对待,与它们所属的特定的端到端连接无关。如我们将见到的那样,通过处理少量的流量聚合,而不是大量的单个连接,提供好于尽力而为服务所需要的新型网络机制能够保持相对简单。

早期因特网设计者的心中清晰地具有这种多种类型服务的概念。回想在图 4-13 中 IPv4 首部中的服务类型 (ToS) 字段。IEN123 [ISI 1979] 描述也呈现在 IPv4 数据报的原型中的 ToS 字段时说:“服务类型 [字段] 提供了所希望的服务质量的抽象参数的指示。