

径。这种设计可以显著地改善主机到主机的容量。为了理解该问题，重新考虑 40 条流的例子。图 5-31 中的拓扑能够处理这种流模式，因为在第 1 台二层交换机和第 2 台二层交换机间存在 4 条不相交的路径，可以一起为前两台二层交换机之间提供总和为 40Gbps 的聚合容量。这种设计不仅减轻了主机到主机的容量限制，同时创建了一种更加灵活的计算和服务环境。在这种环境中，任何未连接到同一台交换机的两个机架之间的通信在逻辑上是等价的，而不论其在数据中心的位置如何。

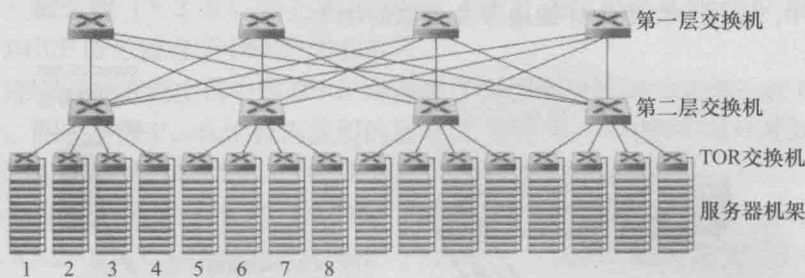


图 5-31 高度互联的数据网络拓扑

另外一个主要的趋势就是采用基于船运集装箱的模块化数据中心（Modular Data Center, MDC）[You Tube 2009; Waldrop 2007]。在 MDC 中，在一个标准的 12 米船运集装箱内，工厂构建一个“迷你数据中心”并将该集装箱运送到数据中心的位置。每一个集装箱都有多达数千台主机，堆放在数十台机架上，并且紧密地排列在一起。在数据中心位置，多个集装箱彼此互联，同时也和因特网连接。一旦预制的集装箱部署在数据中心，通常难以检修。因此，每一个集装箱都得设计为性能下降：当组件（服务器和交换机）随着时间的推移出现故障时，集装箱继续运行但是性能下降。当许多组件出现故障并且性能已经下降到低于某个阈值时，整个集装箱将会被移除，并用新的来替换。

创建由集装箱构成的数据中心提出了新的联网挑战。对于 MDC，有两种类型的网络：每一个集装箱中的内部网络和互联每个集装箱的核心网络 [Guo 2009; Farrington 2010]。在每个集装箱内部，在规模上升到数千台主机的时候，通过廉价的商用吉比特以太网交换机创建全连接的网络（如前面所描述）是可行的。然而，核心网络的设计仍然是一个带有挑战性的问题，这需要能互联成百上千的集装箱，同时能够为典型工作负载提供跨多个集装箱的主机到主机间的高带宽。[Farrington 2010] 中提出了一种互联集装箱的混合电/光交换机体系结构。

当采用高度互联拓扑的时候，一个主要的问题是设计交换机之间的路由选择算法。一种可能是采用随机路由选择方式 [Greenberg 2009b]。另一种可能是在每台主机中部署多块网络接口卡 [Guo 2009]，将每台主机连接到多台低成本的商用交换机上，并且允许主机自己在交换机间智能地为流量选路。这些方案的变种和扩展正被部署在当前的数据中心中。在数据中心设计中很可能会出现更多的创新；有兴趣的读者可以去阅读许多有关数据中心网络设计的论文。

5.7 回顾：Web 页面请求的历程

既然我们已经在本章中学过了链路层，并且在前面几章中学过了网络层、运输层和应用层，那么我们沿协议栈向下的旅程就完成了！在本书的一开始（1.1 节），我们说过“本书的大部分内容与计算机网络协议有关”，在本章中，我们无疑已经看到了情况的确如