

左下角队列中的深色阴影分组必须等待。但不仅该分组要等待，左下角队列中排在该分组后面的浅色阴影分组也要等待，即使右中侧输出端口（浅色阴影分组的目的地）中无竞争。这种现象叫做输入排队交换机中的**线路前部**（Head-Of-the-Line, HOL）阻塞，即在一个输入队列中排队的分组必须等待通过交换结构发送（即使输出端口是空闲的），因为它被位于线路前部的另一个分组所阻塞。[Karol 1987]指出，由于HOL阻塞，只要输入链路上的分组到达速率达到其容量的58%，在某些假设前提下，输入队列长度将无限制地增大（不严格地讲，这等同于说将出现大量的丢包）。[McKeown 1997b]讨论了许多种解决HOL阻塞的方法。

#### 4.3.5 路由选择控制平面

在我们至此所进行的讨论和图4-6中，我们已经隐含地假设路由选择控制平面全部驻留并运行在路由器中的路由选择处理器上。网络范围的路由选择控制平面因此是分布式的，即不同部分（例如路由选择算法）执行在不同的路由器上并且通过彼此发送控制报文进行交互。实际上，今天因特网路由器和我们将在4.6节中学习的路由选择算法正是以这种方式运行的。此外，路由器和交换机厂商将它们的硬件数据平面和软件控制平面绑在一起放入封闭（但可互操作）的平台中，成为一种垂直综合的产品。

近来，一些研究人员[Caesar 2005a; Casado 2009; McKeon 2008]已经着手探索新型路由器控制平面体系结构，其中数据平面连同部分控制平面（例如，链路状态的本地测量/报告，转发表的安装和维护）在路由器中实现，部分控制平面能够在路由器外部实现（例如，在能够执行路由计算的集中式服务器中实现）。一个定义良好的API规定了这两个部分彼此交互和通信的方式。这些研究人员争辩说将软件控制平面与硬件数据平面（具有最少的路由器驻留控制平面）分离，能够以集中式路由选择计算代替分布式路由选择计算，从而简化路由选择，通过允许不同的用户定制的控制平面在快速的硬件数据平面上运行，能够进行网络创新。

#### 4.4 网际协议：因特网中的转发和编址

到目前为止，我们有关网络层编址和转发的讨论并未提及任何特定的计算机网络。在这节中，我们将注意力转向在因特网中是如何完成编址和转发的。我们将看到因特网编址和转发是网际协议（IP）的重要组件。目前有两个版本的IP在使用。我们首先研究广泛部署的IP协议版本4，这通常简称为IPv4[RFC 791]。我们将在本节结尾研究IP版本6[RFC 2460; RFC 4291]，已经提议用它替代IPv4。

在我们尝试研究IP之前，先来考虑构成因特网的网络层的一些组件。如图4-12中所示，因特网的网络层有三个主要组件。第一个组件是IP协议，它是本节的主题。第二个主要组件是路由选择部分，它决定了数据报从源到目的地所流经的路径。我们前面讲过路由选择协议计算出用于在网络中转发分组的转发表。我们将在4.6节中研究因特网的路由选择协议。网络层的最后一个组件是报告数据报中的差错和对某些网络层信息请求进行响应的设施。我们将在4.4.3节中涉及因特网的网络层差错和信息报告的协议，即互联网控制报文协议（ICMP）。