

传输过程中到达，因此排队等待传输。在分组 1 传输后，链路调度器寻找类 2 的分组，因此传输分组 3。在分组 3 传输完成后，调度器寻找类 1 的分组，因此传输分组 2。在分组 2 传输完成后，分组 4 是唯一的排队的分组，因此在分组 2 后立刻传输分组 4。

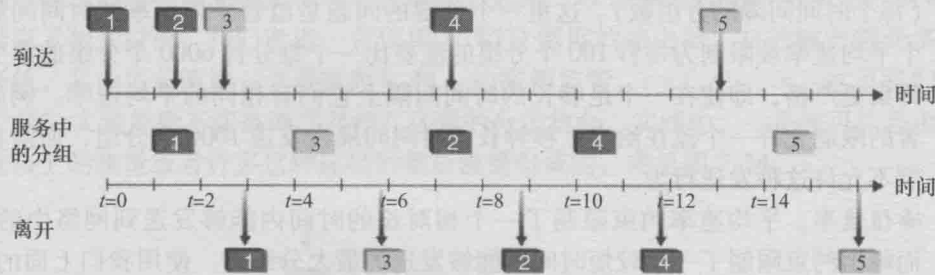


图 7-21 两类循环队列的操作

在 QoS 体系结构中已经得到大量使用的循环排队的一个通用抽象就是所谓的加权公平排队 (Weighted Fair Queuing, WFQ) 规则 [Demers 1990; Parekh 1993]。图 7-22 对 WFQ 进行了描述。到达的分组被分类并在合适的每个类的等待区域排队。与使用循环调度一样，WFQ 调度器也以循环的方式为各个类提供服务，即首先服务第 1 类，然后服务第 2 类，接着再服务第 3 类，然后（假设有 3 个类别）重复这种服务模式。WFQ 也是一个保持工作的排队规则，因此在发现一个空的类队列时，它立即移向服务序列中的下一个类。

WFQ 和循环排队不同之处在于，每个类在任何时间间隔内可能收到不同数量的服务。具体而言，每个类  $i$  被分配一个权  $w_i$ 。使用 WFQ 方式，在类  $i$  有分组要发送的任何时间间隔中，第  $i$  类将确保接收到的服务部分等于  $w_i / (\sum w_j)$ ，式中分母中的和是计算所有有分组排队等待传输的类别得到的。在最坏的情况下，即使所有的类都有分组排队，第  $i$  类仍然保证分配到带宽的  $w_i / (\sum w_j)$  部分。这样，对于一条传输速率为  $R$  的链路，第  $i$  类总能获得至少为  $R \cdot w_i / (\sum w_j)$  的吞吐量。我们对 WFQ 的描述理想化了，因为我们没有考虑这样的事实：分组是离散的数据单元，并且不能打断一个分组的传输来开始传输另一个分组；[Demers 1990] 和 [Parekh 1993] 讨论了这个分组化问题。如我们将在下面几节所见，WFQ 在 QoS 体系结构中起了重要的作用。现在的路由器产品中也用到了 WFQ [Cisco QoS 2012]。

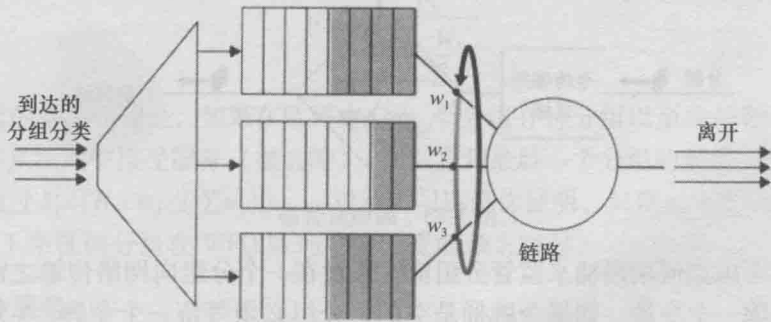


图 7-22 加权公平排队

3. 监管：漏桶

我们前面的见解之一是监管，速率的调节是一种重要的 QoS 机制，以该速率允许一类流量或流（在下面的讨论中，我们将假定监管的单位是流）向网络注入分组。但是应该对