有关排队时延的总体的、直觉的讨论；求知欲强的读者可能要浏览某些书籍（或者最终写有关这方面的博士论文）。与其他3项时延（即 d_{proc}、d_{trans} 和 d_{prop}）不同的是，排队时延对不同的分组可能是不同的。例如，如果10个分组同时到达空队列，传输的第一个分组没有排队时延，而传输的最后一个分组将经受相对大的排队时延（这时它要等待其他9个分组被传输）。因此，当表征排队时延时，人们通常使用统计量测度，如平均排队时延、排队时延的方差和排队时延超过某些特定值的概率。

什么时候排队时延大，什么时候又不大呢？该问题的答案很大程度取决于流量到达该队列的速率、链路的传输速率和到达流量的性质，即流量是周期性到达还是以突发形式到达。为了更深入地领会某些要点，令 a 表示分组到达队列的平均速率（a 的单位是分组/秒，即 pkt/s）。前面讲过 R 是传输速率，即从队列中推出比特的速率（以 bps 即 b/s 为单位）。为了简单起见，也假定所有分组都是由 L 比特组成的。则比特到达队列的平均速率是 La bps。最后，假定该队列非常大，因此它基本能容纳无限数量的比特。比率 La/R 被称为**流量强度**（traffic intensity），它在估计排队时延的范围方面经常起着重要的作用。如果 $La/R>1$，则比特到达队列的平均速率超过从该队列传输出去的速率。在这种不幸的情况下，该队列趋向于无界增加，并且排队时延将趋向无穷大！因此，流量工程中的一条金科玉律是：*设计系统时流量强度不能大于1*。

现在考虑 $La/R\leqslant 1$ 时的情况。这时，到达流量的性质影响排队时延。例如，如果分组周期性到达，即每 L/R 秒到达一个分组，则每个分组将到达一个空队列中，不会有排队时延。在另一方面，如果分组以突发形式到达而不是周期性到达，则有很大的平均排队时延。例如，假定每 $(L/R)N$ 秒同时到达 N 个分组。则传输的第一个分组没有排队时延；传输的第二个分组就有 L/R 秒的排队时延；更为一般地，第 n 个传输的分组具有 $(n-1)L/R$ 秒的排队时延。我们将该例子中的计算平均排队时延的问题留给读者作为练习。

以上描述周期性到达的两个例子有些学术味。到达队列的过程通常是*随机的*，即到达并不遵循任何模式，分组之间的时间间隔是随机的。在这种更为真实的情况下，量 La/R 通常不足以全面地表征时延的统计量。不过，直观地理解排队时延的范围很有用。特别是，如果流量强度接近于0，则几乎没有分组到达并且到达间隔很大，那么到达的分组将不可能在队列中发现别的分组。因此，平均排队时延将接近0。在另一方面，当流量强度接近1时，将存在到达率超过传输能力的时间间隔（由于分组到达率的波动），在这些时段中将形成队列。无论如何，随着流量强度接近1，平均排队长度变得越来越长。平均排队时延与流量强度的定性关系如图1-18所示。

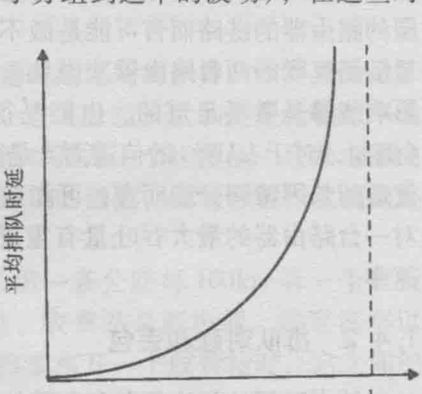

图1-18　平均排队时延与流量强度的关系

图1-18的一个重要方面是这样一个事实：随着流量强度接近于1，平均排队时延迅速增加。该强度少量的增加将导致时延大得多的增加。也许你在公路上经历过这种事。如果经常在通常拥塞的公路上驾驶，这条路经常拥塞的事实意味着它的流量强度接近于1。如果某些事件引起一个甚至稍微大于平常量的流量，经受的时延就可能很大。

为了真实地感受一下排队时延的情况，我们再次鼓励你访问本书的Web网站，该网站提供了一个有