

第6章 死 锁

在计算机系统中有许多独占性的资源，在任一时刻它们都只能被一个进程使用。常见的有打印机、磁带以及系统内部表中的表项。打印机同时让两个进程打印将造成混乱的打印结果；两个进程同时使用同一文件系统表中的表项会引起文件系统的瘫痪。正因为如此，操作系统都具有授权一个进程（临时）排他地访问某一种资源的能力。

在很多应用中，需要一个进程排他性地访问若干种资源而不是一种。例如，有两个进程准备分别将扫描的文档记录到CD上。进程A请求使用扫描仪，并被授权使用。但进程B首先请求CD刻录机，也被授权使用。现在，A请求使用CD刻录机，但该请求在B释放CD刻录机前会被拒绝。但是，进程B非但不放弃CD刻录机，而且去请求扫描仪。这时，两个进程都被阻塞，并且一直处于这样的状态。这种状况就是死锁（deadlock）。

死锁也可能发生在机器之间。例如，许多办公室中都用计算机连成局域网，扫描仪、CD刻录机、打印机和磁带机等设备也连接到局域网上，成为共享资源，供局域网中任何机器上的人和用户使用。如果这些设备可以远程保留给某一个用户（比如，在用户家里的机器使用这些设备），那么，也会发生上面描述的死锁现象。更复杂的情形会引起三个、四个或更多设备和用户发生死锁。

除了请求独占性的I/O设备之外，别的情况也有可能引起死锁。例如，在一个数据库系统中，为了避免竞争，可对若干记录加锁。如果进程A对记录R1加了锁，进程B对记录R2加了锁，接着，这两个进程又试图各自把对方的记录也加锁，这时也会产生死锁。所以，软硬件资源都有可能出现死锁。

在本章里，我们准备考察几类死锁，了解它们是如何出现的，学习防止或者避免死锁的办法。尽管我们所讨论的是操作系统环境下出现的死锁问题，但是在数据库系统和许多计算机应用环境中都可能产生死锁，所以我们所介绍的内容实际上可以应用到包含多个进程的系统。有很多有关死锁的著作，《Operating Systems Review》中列出了两本参考书（Newton, 1979; Zobel, 1983），有兴趣的读者可以参考这两本书。死锁方面的大多数研究工作在1980年以前就完成了，尽管所列的参考文献有些老，但是这些内容依然是很有用的。

6.1 资源

大部分死锁都和资源相关，所以我们首先来看看资源是什么。在进程对设备、文件等取得了排他性访问权时，有可能会出现死锁。为了尽可能使关于死锁的讨论通用，我们把这类需要排他性使用的对象称为资源（resource）。资源可以是硬件设备（如磁带机）或是一组信息（如数据库中一个加锁的记录）。通常在计算机中有多种（可获取的）资源。一些类型的资源会有若干个相同的实例，如三台磁带机。当某一资源有若干实例时，其中任何一个都可以用来满足对资源的请求。简单来说，资源就是随着时间的推移，必须能获得、使用以及释放的任何东西。

6.1.1 可抢占资源和不可抢占资源

资源分为两类：可抢占的和不可抢占的。可抢占资源（preemptable resource）可以从拥有它的进程中抢占而不会产生任何副作用，存储器就是一类可抢占的资源。例如，一个系统拥有256MB的用户内存和一台打印机。如果有两个256MB内存的进程都想进行打印，进程A请求并获得了打印机，然后开始计算要打印的值。在它没有完成计算任务之前，它的时间片就已经用完并被换出。

然后，进程B开始运行并请求打印机，但是没有成功。这时有潜在的死锁危险。由于进程A拥有打印机，而进程B占有了内存，两个进程都缺少另外一个进程拥有的资源，所以任何一个都不能继续执行。不过，幸运的是通过把进程B换出内存、把进程A换入内存就可以实现抢占进程B的内存。这样，进程A继续运行并执行打印任务，然后释放打印机。在这个过程中不会产生死锁。