The Working Set Model for Program Behavior

这篇文章介绍了内存与外存的交换模型，工作集模型。

再页式管理模型中，只有存在在内存中的数据才可以被处理器访问，其他数据都存在在外存中。当处理器要访问的数据不在内存中时，就需要将内存中的数据换出到外存，再将需要的数据换入内存。这样的时间称为交换时间（traverse time）。一个基本的问题就是，决定将内存中哪一个数据换出。基本的策略就是，最小化内存与外存的交换流量（page traffic）。

在本文之前，已经有一些方案被提出。1随机选择。当需要一个新的页时，随机选择一个页换出去，将需要的换进来。这个方法实现简单，但是会频繁的换出有用的页面，这些页面不得不再次换进来，增大交换流量。2FIFO。当需要的页面不在内存中时，选择最先进入的页交换出去。显然这也不是很好。3LRU，最近最少使用原则。每个页都有一个使用位，使用的时候标为ON。一个周期间隔内，所有页面会被查找并更新使用位。当需要新的页时，将交换出最久没有使用的页。这种方法，在只有一个进程并且所有的页面不能同时装入内存时，才是合理的；但是，当有很多进程竞争内存时，这种方法易受影响。4ATLAS Loop Detection Method。在页访问模型中，这种方法检测循环行为，然后移出将会被最久访问的页面。由于程序需要禁止随机访问行为，因此性能并不是很好。并且实现代价较大。

工作集模型。对于程序员来说，信息工作集是为了保证高效执行程序而必须在内存中的最少的信息集合。我们不能从编译器或者程序员中得知哪些信息必须在内存中，必须依赖于操作系统来决定。因此，从操作系统的角度，一个进程的工作集就是最近访问的页面。1工作集的大小是单调增的。2直觉上，越接近当前时刻的页面对预测的作用越大。3再次换入率。考虑过去的时间τ越短，则在有用的页面不在工作集中的概率越大。4τ敏感性。函数σ衡量再次换入率和时间的关系。5内存分配。只有足够的内存给程序的工作集，这个程序才会被运行。

调度。1由于进程和他的工作集的关系密切，因此内存管理和进程调度必须紧密配合活动。2效率是首要考虑的。3设计的机制要能够提供当前工作集大小和每个进程的处理器时间的消耗。

共享。共享是必要的。当不同进程有相同的页面时，则工作集将会重叠。当某个页面被多个进程共享是，则设定使用位来标记，当需要交换时，这些页面不会被交换出去。

资源分配是权衡各方面的问题。作者给出了内存需求和处理器需求的定义。平衡状态，1所有进程的内存需求等于内存资源提供量，2处理器资源需求等于处理器资源提供量。然后给出了权衡策略。策略的目的是使的系统处于平衡之中。即最小化需求减去提供量。