

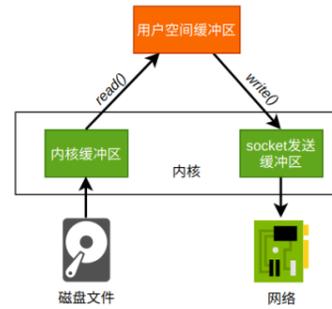
Linux零拷贝

场景

在Web服务器中，经常需要将硬盘中的文件原封不动地通过已连接的套接字发送出去，典型的应用就是图片、文档等数据的发送

其中一种处理方法就是循环读取文件内容，并将读取的内容发送至socket

```
while ((n = read(diskfilefd, buf, BUF_SIZE)) > 0) {  
    write(socketfd, buf, n);  
}
```



上述方式虽然仅有寥寥两行代码，但是在内部却涉及了多次用户态与内核态的切换，以及数据在用户空间和内核空间的拷贝

当在大量传输磁盘文件时，CPU状态的切换以及数据的拷贝将会造成较大的性能影响

因此，系统调用sendfile()被设计用来去除这些无用的动作，可将硬盘文件直接传输至套接字上

sendfile()系统调用

sendfile()系统调用允许将磁盘文件内容直接发送给socket，而不必进行CPU状态切换

`ssize_t sendfile(int out_fd, int in_fd, off_t *offset, size_t count);` 成功时将返回传输的数据量，失败则返回-1

参数释义
`out_fd` 必须指向一个套接字，不能是普通文件或者管道。在Linux 2.4版本之前，`out_fd`允许指向文件，但在后续版本中被移除。如果在未来的版本中重新添加了文件描述符的支持，文件复制又将会得到一部分的性能提升

`in_fd` 指向的文件必须能够进行mmap()操作，通常是文件，但不可以是套接字

`offset_t` 文件偏移量，需要注意的是这里传递的是个指针，也就意味着当sendfile()调用结束时，将会把新的文件偏移量结果写入至该指针中。sendfile()调用并不会改变文件的偏移量

