# 百万长连接应用尝试

## 长连接应用

基于B/S的IM通信：如淘宝的阿里旺旺浏览器在线；

网页游戏：游戏中的聊天，事件的推送，实时在线人数的统计等

文字直播：如新浪门户中的NBA体育文字直播，或者两会中的总理答记者

证券系统：实时推送股市行情。

……

长连接主要用于互联网中及时性要求高的应用，该类要求应用主动更新，信息实时送达。在系统设计过程中，我们应重点关注服务器的压力承载情况，瓶颈通常会出现：

**网络IO：**网络的带宽直接决定了你的吞吐量。

**磁盘IO：**应尽量避免或者不使用磁盘，日志的输出也是会降低系统的处理能力，对于标准输入，标准输出，标准错误要关了。

**内存：**内存的分配，应提前规划，linux系统每个sock的内存都预算，同时程序中的recv缓冲区和send缓冲区都要预算。Send缓冲区应该设计为共享方式，而无须每个连接都分配一个缓冲区，减少内存。

**数据结构：**要保证性能就要选择足够快的数据结构，连接的管理使用数组，数组在所有的数据结构中最快的，采用文件描述符做为数组的下标，文件描述符的分配原则是以最小的数值来分配，它不同于进程ID的循环分配。连接超时管理采用红黑树。

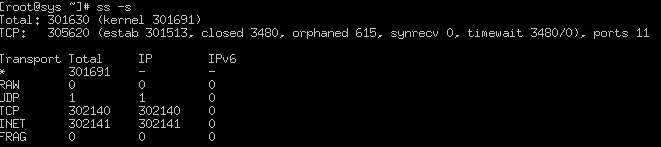
Redis lighttp nginx

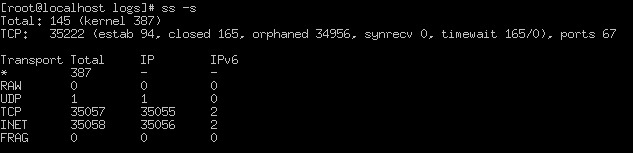
## 工具

* **查询网络状态：**

ss –s

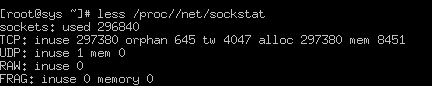
ss相对于netstat它的执行效率更高。



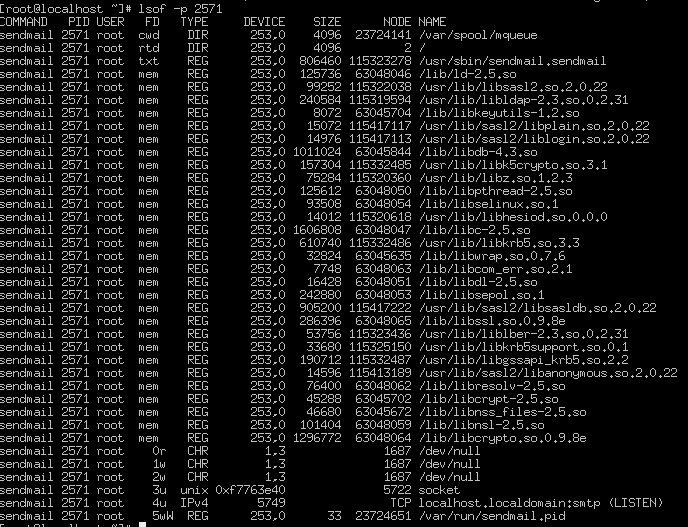


进程中止时出现大量的orphaned

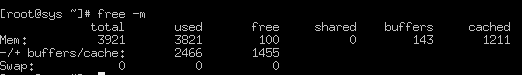
/proc/net/sockstat

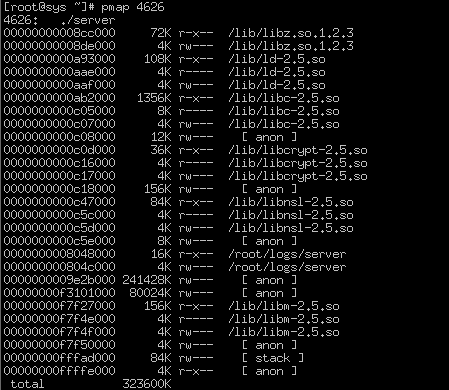


lsof观察当前进程下面的文件描述符的分配情况，观察网络的连接状况。



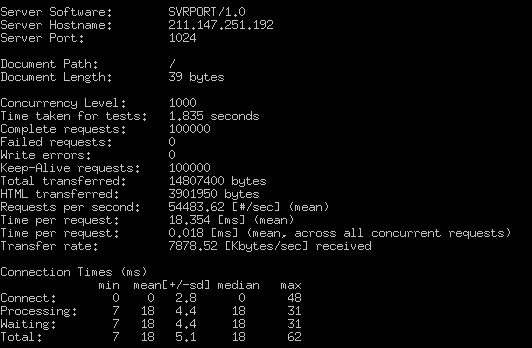
* **观察内存分配情况：**





* **测试工具：**

./ab -c 1000 -n 100000 –k http://211.147.251.192:1024/



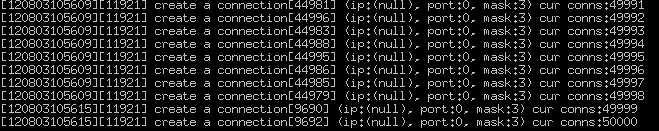
类似的还有webbench，简单看了下它的代码，简单看了下1.5版本，使用的是fork的方式，一个进程一个连接。如果并发量大，webbench不是很好的选择。

Apache项目源代码带庞大的基础类，ab的代码没有分析了，目前主流的测试工具，strace跟踪的话，可以看到使用了epoll非阻塞的单进程。

考虑到以后的项目测试，一些非使用http协议的客户端测试，编写了一个简单客户端。完成并发5万，提交一个http request。设计思路同ab，epoll+非阻塞。

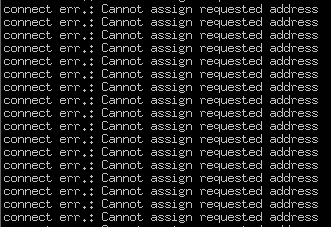
./client\_test

Epoll模式管理自己的客户端连接。



## 系统优化

一、客户端分配足够的端口



# /etc/sysctl.conf

net.ipv4.ip\_local\_port\_range = 1024 65535

二、服务器分配足够的文件描述符



# /etc/rc.local

ulimit -SHn 2097152

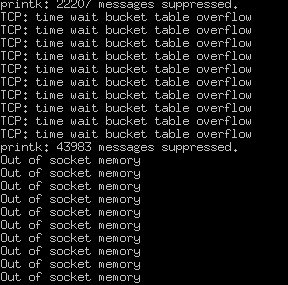
首先要查询最大的文件描述符，具体数目可以cat /proc/sys/fs/file-max察看,一般来说这个数目和系统内存关系很大。

三、合理管理time wait状态的socket连接

Time wait：该状态存在于主动发起FIN端。大量出现该问题，

在测试过程中，也出现了一大量的close\_wait状态的连接，该问题可以规避，很简单close(fd)就行，这个是代码问题。代理中回调函数出了点问题，出现了该问题。

dmesg:发现系统的内核错误。



net.ipv4.tcp\_tw\_reuse = 1

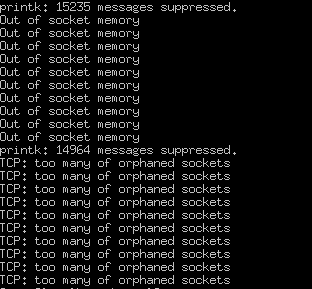
net.ipv4.tcp\_tw\_recycle = 1

net.ipv4.tcp\_max\_tw\_buckets = 96000

#单位页

net.ipv4.tcp\_mem = 786432 1048576 1572864

四、大量无主socket和依旧依旧存在的socket内存不足。



net.ipv4.tcp\_rmem = 4096 4096 4194304

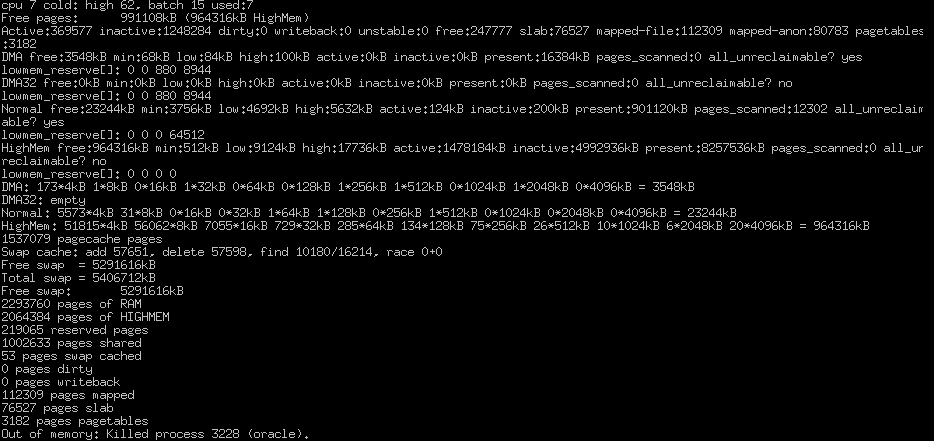
net.ipv4.tcp\_wmem = 4096 4096 4194304

net.ipv4.tcp\_max\_orphans = 162768

net.core.netdev\_max\_backlog = 50000

五、Linux OOM Killer

dmesg

OOM Killer是一层保护机制，用于避免 Linux 在内存不足的时候不至于出太严重的问题，把一些进程杀掉，OOM Killer 在 2.4 与 2.6 上表现是不一样的。2.4 的版本中是把新进来(新申请内存)的进程杀掉。而 2.6 上是杀掉占用内存最厉害的进程(这是很危险的，很容易导致系统应用瘫痪)。

Linux 内核定义了三个区域：

# DMA: 0x00000000 - 0x00999999 (0 - 16 MB)

# LowMem: 0x01000000 - 0x037999999 (16 - 896 MB) - size: 880MB (NORMAL ZONE不足OOM Killer)

# HighMem: 0x038000000 - <硬件特定>

/proc/meminfo

/proc/buddyinfo

Oracle数据库进程被强制停止

解决办法：升级到64bit

## 参考

杨建：网站加速--服务器编写篇（上）（下）

<http://blog.sina.com.cn/s/blog_466c66400100bi2n.html>

<http://blog.sina.com.cn/s/blog_466c66400100bi2s.html>

http长连接200万尝试及调优

<http://rdc.taobao.com/blog/cs/?p=1062>

sysctl 配置

