

高性能服务器设计与优化

9 Mar 2017

沉风

内容

网络优化

- 网络服务链路最后一公里加速

架构设计

- 利于网络服务落地与实现
- 便于网络服务扩展与升级

系统优化

- 为网络服务引擎加动力

网络优化

- 基本流程
- 网卡优化
- 协议栈与应用优化

基本流程

- 接收流程
- 发送流程

接收流程

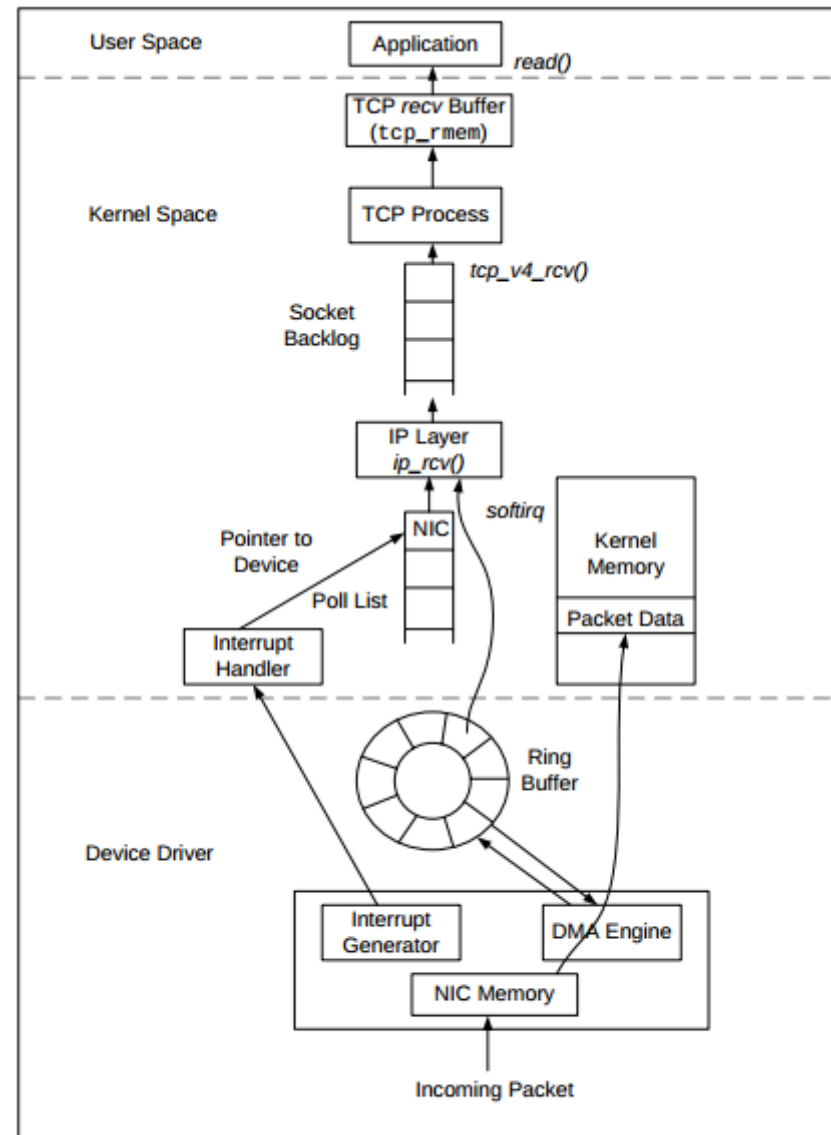


Fig. 1: Packet Reception

发送流程

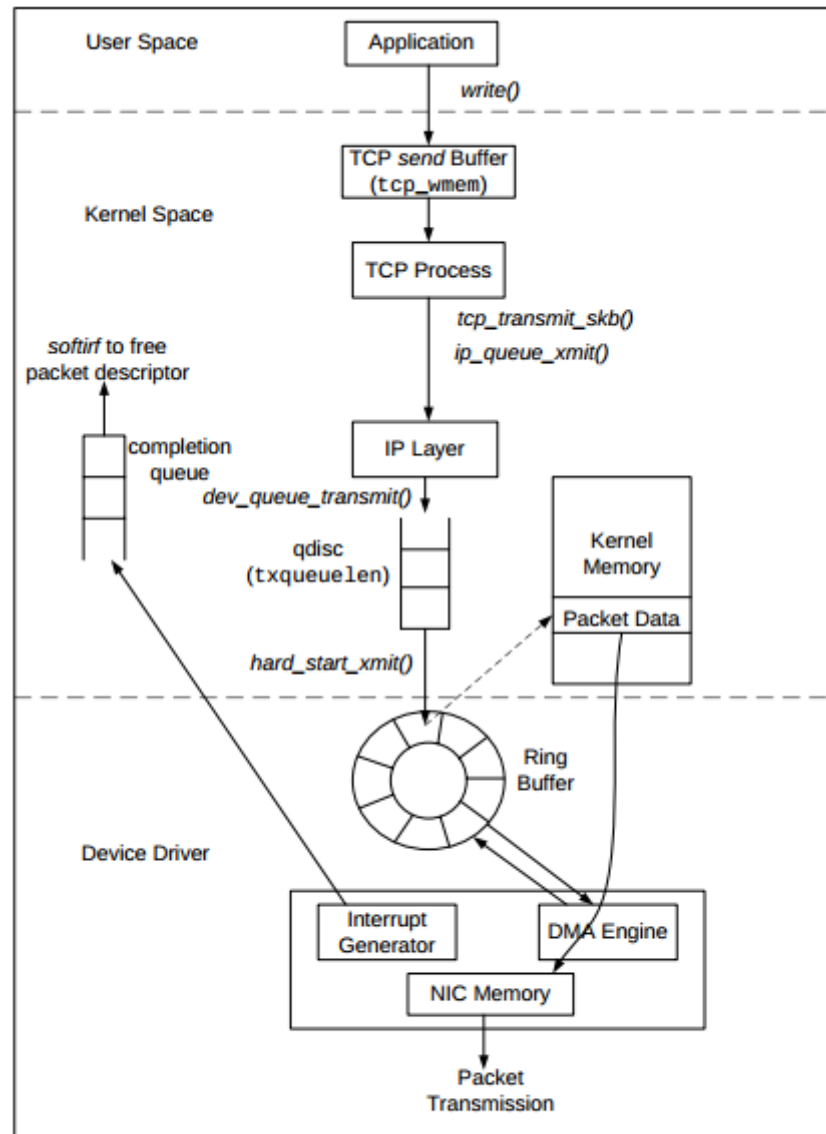


Fig. 2: Packet Transmission

网卡优化

以intel 82599网卡为例

- RSS(Receive Side Scaling)
- RPS(Receive Packet Steering)
- GRO(Generic Receive Offload)
- TSO(TCP Segmentation Offload)
- VMDq
- MSI-X
- IOMMU
- DCA
- Interrupt绑定cpu
- Interrupt负载均衡
- Interrupt合并

协议栈

- 报文上送
- IP层优化
- TCP优化

报文上送

- 中断方式
- NAPI方式

IP层优化

- 根据情况关闭iptables
- 针对性配置Qos
- 避免路由cache条目过多

TCP优化

- 更快地建立和应用TCP连接
- 支持更多的并发连接建立
- 支持更多的在线连接
- 提高TCP连接的吞吐
- 降低TCP报文的延迟
- 优化拥塞控制
- 高效读写缓冲区

更快地建立和应用TCP连接

- Tcp Fast Open
- HTTP KeepAlive
- HTTP2.0
- 适当增大cwnd
- 使用连接池

支持更多的并发连接建立

开启SO_REUSEPORT选项

适当调大backlog

- net.core.netdev_max_backlog
- net.core.somaxconn
- net.ipv4.tcp_max_syn_backlog

支持更多的在线连接

- 开启tcp_tw_reuse
- 开启tcp_tw_recycle
- 开启SO_LINGER选项
- 减小tcp_fin_timeout
- 开启tcp_sack
- 开启tcp_fack
- 增加协议栈内存

提高TCP连接的吞吐

Throughput = Buffer Size / Latency

- 对于高延迟网络传输环境下，要加大接收与发送缓冲区大小
- 增大了cwnd

合并收发

- 开启Nagle
- 开启TCP_CORK

降低TCP报文的延迟

- 开启TCP NODELAY
- 开启TSQ(TCP small queue)
- 关闭TSO
- 关闭GSO

优化拥塞控制

- Vegas vs Reno vs CUBIC vs BBR
- TSQ
- 开启tcp_sack
- 开启tcp_window_scaling

高效读写缓冲区

- readv, writev
- sendfile
- 应用协议友好的代码
- 非阻塞

架构设计

- 技术
- 模型

技术-1

并发

Epoll

RPC

异步

协程

- libco

无锁化

- CAS, memory barrier
- Lock-free(Queue, Ring Buffer, Stack)

技术-2

缓存

- redis
- memcache
- 进程内缓存

数据库

- 数据描述
- 数据分布
- 数据读写
- 数据迁移
- 数据扩容

模型

- Reactor
- Proactor
- 流水线

系统优化

- CPU
- 内存
- 算法
- 工具

CPU

减少context切换

- cpu affinity
- 业务分配

进程调度优化

利用硬件，降低CPU负载

内存-1

ptmalloc vs tcmalloc vs jemalloc

- 内存使用场景决定

内存池

代码对cache友好

- 时间有效性
- 空间有效性

降低cache冲突

- 数据结构优化
- 算法优化
- 代码级优化

内存-2

站在巨人肩膀上参考nginx式用户态内存管理

减少内存拷贝

关闭SWAP

算法

- 空间复杂度
- 时间复杂度

工具

工欲善其事，必先利其器

- googleperf
- perf
- OProfile
- 更多

参考 [Brendan D. Gregg](http://www.brendangregg.com/) (<http://www.brendangregg.com/>)

欢迎交流与指正

Thank you

9 Mar 2017

Tags: [performance](#) [web-server](#) (#ZgotmplZ)

沉风

myself659@163.com (mailto:myself659@163.com)

<http://myself659.github.io/> (http://myself659.github.io/)

<https://github.com/myself659> (https://github.com/myself659)

