UNIX Network Programming

@于龙

目录

[1 简介 2](#_Toc419032150)

[1.1 概述 2](#_Toc419032151)

[1.2 时间获取程序 2](#_Toc419032152)

[2 传输层：TCP、UDP、SCTP 2](#_Toc419032153)

[2.1 TCP 2](#_Toc419032154)

[2.2 TCP连接的建立和终止 3](#_Toc419032155)

[2.3 TIME\_WAIT状态 3](#_Toc419032156)

# 简介

## 概述

POSIX 表示可移植操作系统接口（Portable Operating System Interface ，缩写为 POSIX ），POSIX标准定义了操作系统应该为应用程序提供的接口标准.

POSIX标准意在期望获得源代码级别的软件可移植性.换句话说，为一个POSIX兼容的操作系统编写的程序，应该可以在任何其它的POSIX操作系统（即使是来自另一个厂商）上编译执行.

Linux基本上逐步实现了POSIX兼容，但并没有参加正式的POSIX认证.

微软的Windows NT声称部分实现了POSIX标准.

## 时间获取程序

read返回0表明对端关闭连接，返回负值表明发生错误

通常情况下，服务器进程在accept调用中被投入睡眠，等待某个客户连接的到达并被内核接收.TCP使用三次握手建立连接，连接建立后accept返回.

使用snprintf、strncat、strncpy来替代，防止缓冲区溢出.

close关闭与客户的连接，该调用引发每个方向上发送一二FIN，每个FIN又由对端确认.

# 传输层：TCP、UDP、SCTP

SCTP最初设计用于跨因特网传输电话信令.

SCTP与TCP类似之处在于也是一个可靠的传输协议，但它还提供消息边界、传输级别多宿支持以及将头端阻塞减少到最小的一种方法.

## TCP

TCP还提供了可靠性，当TCP向另一端发送数据时，它要求对端返回一个确认.如果没有收到确认，TCP就自动重传数据并等待更长时间.在数次重传失败后，TCP才放弃，如此在尝试发送数据上所花的总时间一般为4-10分钟.

TCP含有动态估算客户和服务器之间的往返时间的算法，以便指导等待一个确认需要多少时间.round-trip time rtt

TCP提供流量控制，总是告知对端在任何时刻一次能够从对端接收多少字节的数据，这称为通告窗口.该窗口指出接收缓冲区中当前可用的空间，从而确保发送端发送的数据不会使接收缓冲区溢出.

TCP是全双工连接，需要时可以转换成一个单工连接.

## TCP连接的建立和终止

三路握手：

SYN J

SYN K + ACK J + 1

ACK K + 1

TCP选项：

MSS最大分节大小

窗口规模选项 65535

时间戳选项 忽略

连接禁止：

FIN M

ACK M + 1

FIN N

ACK N + 1

当一个进程自愿或非自愿关闭时，所有打开的描述符都被关闭，这仍然导致FIN发送.

TCP状态转换图

服务器对客户请求的确认是伴随其应答发送的，这称为捎带piggybacking，它通常在服务器处理请求并产生应答的时间少于200ms时发生.如果服务器耗用更长时间，那么将先确认后应答.



## TIME\_WAIT状态

执行主动关闭的那端经历了这个状态.被动关闭是 CLOSE\_WAIT.

该端点停留在这个状态的持续时间是最长分节生命期MSL的两倍.2MSL

RFC 1122建议MSL为2分钟.源自Berkeley的实现传统上是30秒.MSL是任何IP数据包能在因特网中存活的最长时间.

分组在网络中迷途通常是由路由异常的结果.

TIME\_WAIT状态有两个存在的理由：

可靠地实现TCP全双工连接的终止；

允许老的重复分节在网络中消逝.

## 端口号

Unix系统有保留端口，小于1024的任何端口，使用这些端口的服务器必须以超级用户特权启动.

## 缓冲区大小及限制

IPv4 65535

IPv6 65575

MTU 以太网1500

当一个IP数据包超过相应链路MTU时，IPv4和IPv6都将执行分片，这些分片到达最终目的地之前通常不会被重组.v4路由器会对转发的数据进行分片，v6路由器不会.

IPv4 和 IPv6都定义了最小重组缓冲区大小，它是任何实现都必须保证支持的最小数据包大小.IPv4 576 IPv6 1500

TCP有一个MSS，最大分节大小，用于向对端TCP通告对端在每个分节中能发送的最大TCP数据量.试图避免分片.