

TCP/IP 详解卷 1：协议 - 14

阅读目的：了解 TCP 数据流与窗口管理。
阅读时间：2 小时
阅读概况：第 15 章

第 15 章 TCP 数据流与窗口管理

1. 概述

交互式通信

在一定时间内，互联网的不同部分传输的网络流量（通常以字节或包来计算）也存在相当大的差异。例如，局域网与广域网以及不同网站之间的流量都会有所不同。TCP 流量研究端口表明，通常 90% 或者更多的 TCP 报文段都包含大批量数据（如 Web、文件共享、电子邮件、备份），其余部分则包含交互式数据（如远程登录、网络游戏）。批量数据段通常较大（1500 字节或更大），而交互式数据段则会比较小（几十字节的用户数据）。

延时确认

在许多情况下，TCP 并不对每个到来的数据包都返回 ACK，利用 TCP 的累积 ACK 字段就能实现该功能。累积确认可以允许 TCP 延迟一段时间发送 ACK，以便将 ACK 和相同方向上需要传的数据结合发送。这种捎带传输的方法经常用于批量数据传输。TCP 不能任意时长地延迟 ACK；否则对方会误认为数据丢失而出现不必要的重传。

TCP 实现 ACK 延迟的时延应小于 500ms。实践中时延最大取 200ms。

2. 流量控制与窗口管理

滑动窗口

TCP 连接的每一端都可收发数据。连接的收发数据量是通过一组窗口结构（window structure）来维护的。每个 TCP 活动连接的两端都维护一个发送窗口结构（send window structure）和接收窗口结构（receive window structure）。图中显示了一个假设的 TCP 发送窗口结构。

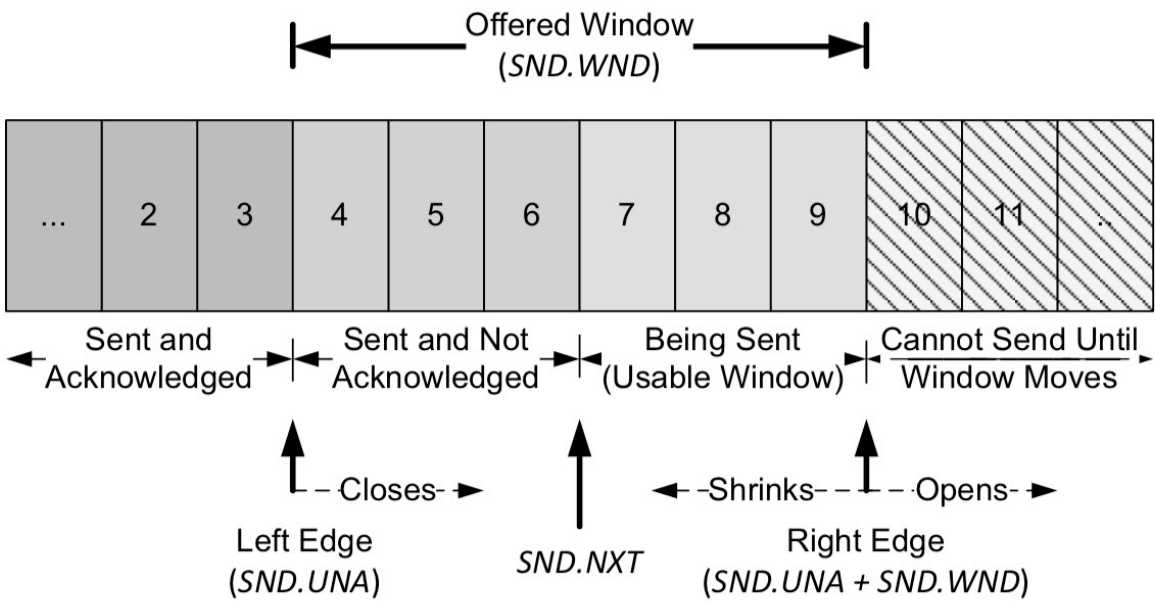


Figure 15-9 The TCP sender-side sliding window structure keeps track of which sequence numbers have already been acknowledged, which are in flight, and which are yet to be sent. The size of the offered window is controlled by the *Window Size* field sent by the receiver in each ACK.

TCP 以字节（而非包）为单位维护其窗口结构。

随着时间的推移，当接收到返回的数据 ACK，滑动窗口也随之右移。窗口两端的相对运动使得窗口增大或减小。可用三个术语来描述窗口左右边界的运动：

- 1. 关闭（close），即窗口左边界右移。当已发送数据得到 ACK 确认时，窗口会减小。
- 2. 打开（open），即窗口右边界右移，使得可发送数据量增大。当已确认数据得到处理，接收端可用缓存变大，窗口也随之变大。
- 3. 收缩（shrink），即窗口右边界左移。但并不建议这一做法。

每个 TCP 报文段都包含 ACK 号和窗口通告信息，TCP 发送端可以据此调节窗口结构。窗口左边界不能左移，因为它控制的是已确认的 ACK 号，具有累积性，不能返回。当得到的 ACK 号增大而窗口大小保持不变时（通常如此），我们就说窗口向前“滑动”。

若随着 ACK 号增大窗口却减小，则左右边界距离减小。当左右边界相等时，称之为零窗口。此时发送端不能再发送新数据。这种情况下，TCP 发送端开始 探测（probe）对方窗口，伺机增大提供窗口。

接收端也维护一个窗口结构，但比发送端窗口简单。该窗口结构记录了已接收并确认的数据，以及它能够接收的最大序列号。该窗口可以保证其接收数据的正确性。特别是接收端希望避免存储重复的已接收和确认的数据，以及避免存储不应接收的数据（超过发送方右窗口边界的数据）。

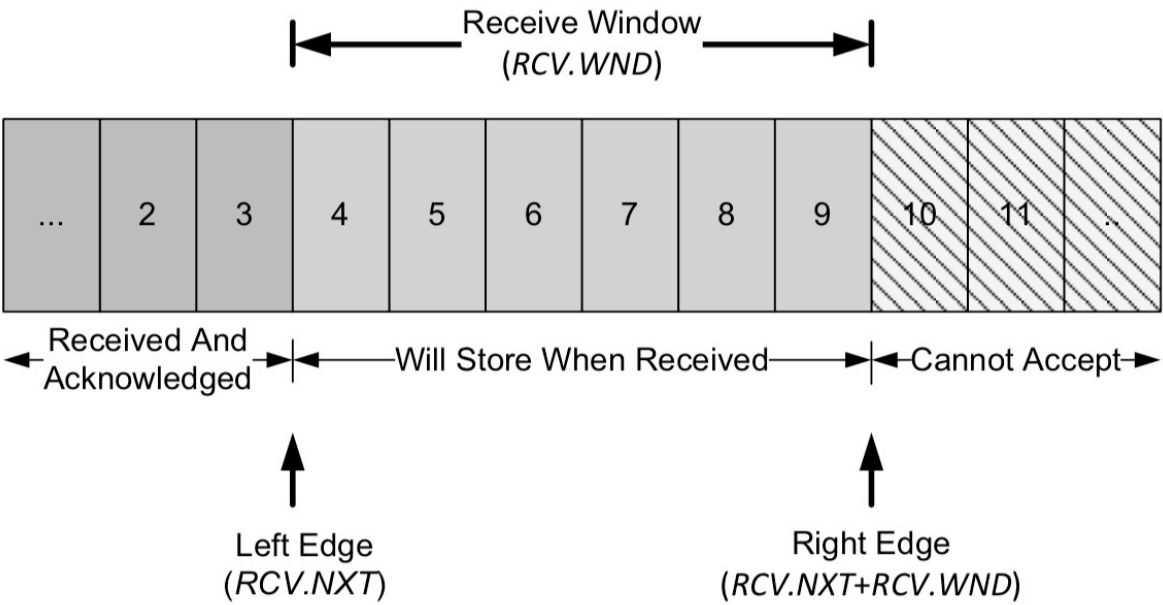


Figure 15-10 The TCP receiver-side sliding window structure helps the receiver know which sequence numbers to expect next. Sequence numbers in the receive window are stored when received. Those outside the window are discarded.

零窗口与 TCP 持续计时器

TCP 是通过接收端的通告窗口来实现流量控制的。通告窗口指示了接收端可接收的数据量。当窗口值变为 0 时，可以有效阻止发送端继续发送，直到窗口大小恢复 非零值。当接收端重新获得可用空间时，会给发送端传输一个 窗口更新（window update），告知其可继续发送数据。这样的窗口更新通常都不包含数据（为“纯 ACK”），不能保证其传输的可靠性。因此 TCP 必须有相应措施能处理这类丢包。

如果一个包含窗口更新的 ACK 丢失，通信双方就会一直处于等待状态：接收方等待接收数据（已将窗口设为非零值），发送方等待收到窗口更新告知其可继续发送。为防止这种死锁的发生，发送端会采用一个持续计时器间歇性地查询接收端，看其窗口是否已增长。持续计时器会触发 窗口探测（window probe）的传输，强制要求接收端返回 ACK（其中包含了窗口大小字段）。协议建议在一个 RTO 之后发送第一个窗口探测，随后以指数时间间隔发送。