计 算 机 网 络

TCP拥塞窗口作图讨论

早期TCP：RFC2001 [1997]

[RFC Home] [TEXT|PDF|HTML] [Tracker] [IPR] [Info page]

Obsoleted by: 2581 PROPOSED STANDARD

Network Working Group W. Stevens
Request for Comments: 2001 NOAO
Category: Standards Track January 1997

TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Congestion avoidance and slow start are independent algorithms with different objectives. But when congestion occurs TCP must slow down its transmission rate of packets into the network, and then invoke slow start to get things going again. In practice they are implemented together.

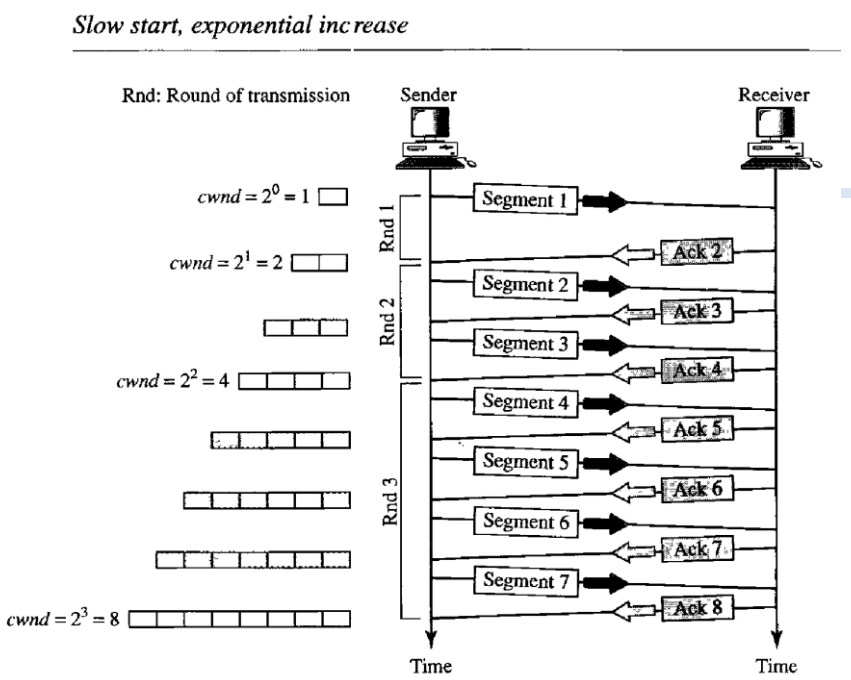
Congestion avoidance and slow start require that two variables be maintained for each connection: a congestion window, cwnd, and a slow start threshold size, ssthresh. The combined algorithm operates as follows:

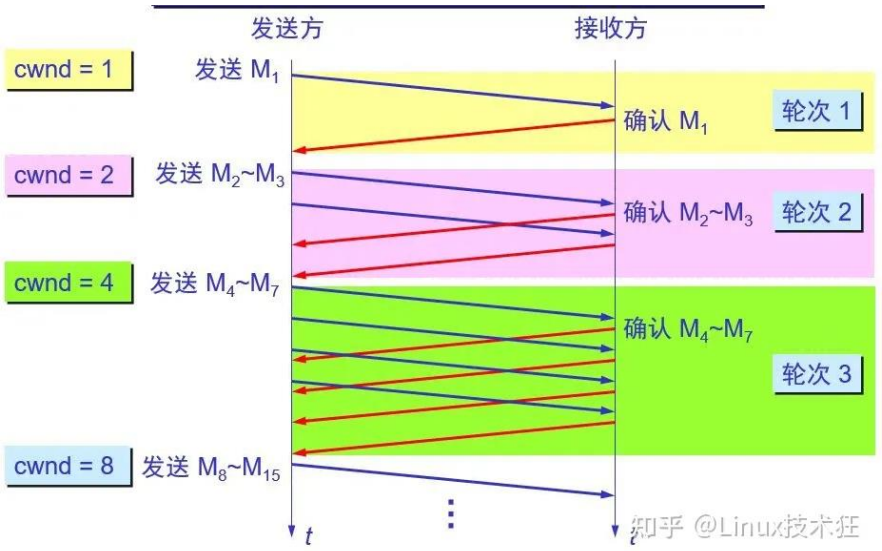
1. Initialization for a given connection sets cwnd to one segment and ssthresh to 65535 bytes.
2. The TCP output routine never sends more than the minimum of cwnd and the receiver's advertised window.
3. When congestion occurs (indicated by a timeout or the reception of duplicate ACKs), one-half of the current window size (the minimum of cwnd and the receiver's advertised window, but at least two segments) is saved in ssthresh. Additionally, if the congestion is indicated by a timeout, cwnd is set to one segment (i.e., slow start).

cwnd **的初始值 设置为**1

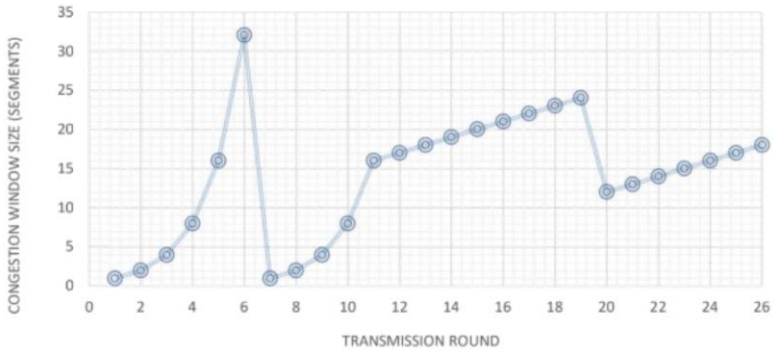
ssthresh **的最小值为**2
cwnd**在拥塞后重置为**1

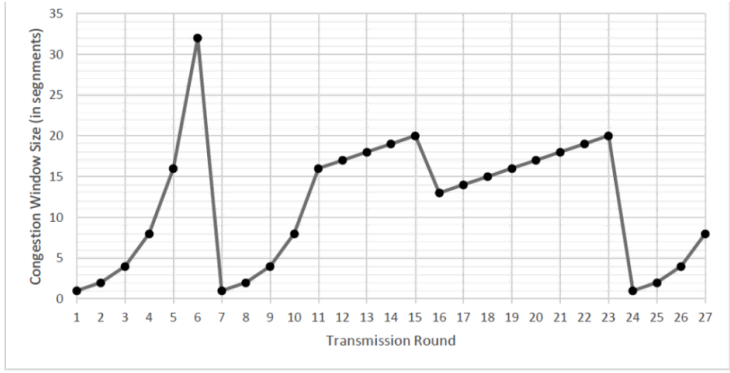
作图案例

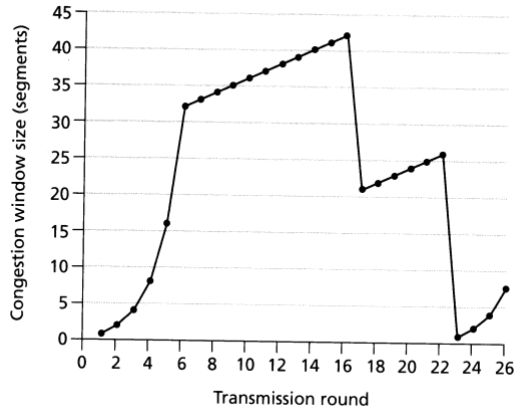

Slow start, exponential increase
Rnd: Round of transmission
Sender
Receiver
cwnd = 2^0 = 1
cwnd = 2^1 = 2
cwnd = 2^2 = 4
cwnd = 2^3 = 8
Rnd 1
Rnd 2
Rnd 3
Segment 1
Ack 2
Segment 2
Ack 3
Segment 3
Ack 4
Segment 4
Ack 5
Segment 5
Ack 6
Segment 6
Ack 7
Segment 7
Ack 8
Time
Time

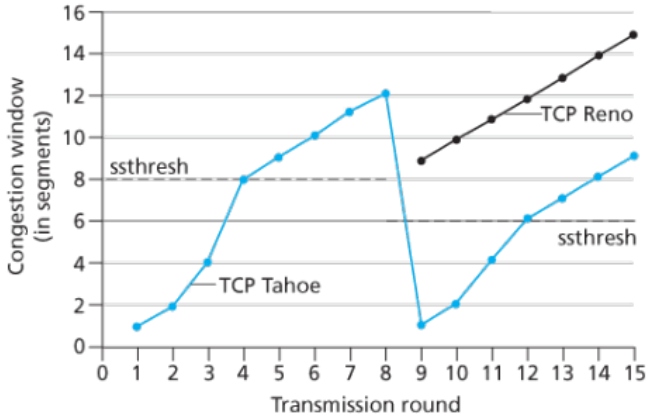

发送方
接收方
cwnd = 1
发送 M1
确认 M1
轮次 1
cwnd = 2
发送 M2~M3
确认 M2~M3
轮次 2
cwnd = 4
发送 M4~M7
确认 M4~M7
轮次 3
cwnd = 8
发送 M8~M15
t
t
知乎 @Linux技术狂

作图案例





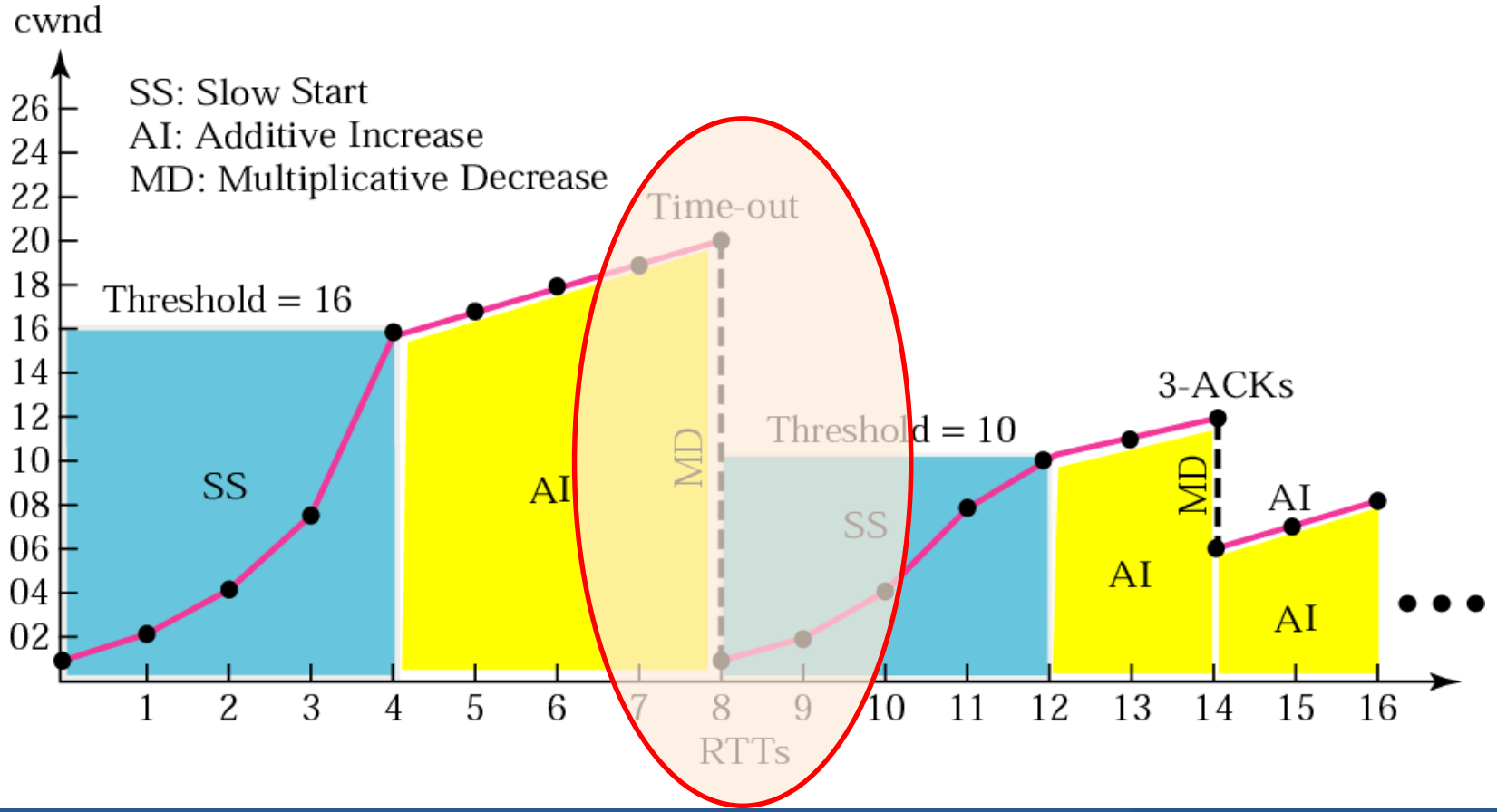




- 横轴以传输轮次 transmission round 为单位，从1 开始
- 纵轴以拥塞窗口cwnd的segment为单位，从1开始
- 作图从（1，1）开始

有问题的案例-1

（1）横轴不宜取RTT，等距RTT难以刻画TCP的变化
如果出现超时，超时时间间隔RTO应该大于等于EstimatedRTT，例如EstimatedRTT+4*DevRTT，或者n*EstimatedRTT。
所以此图RTT=8开始以后的横轴作图都有问题

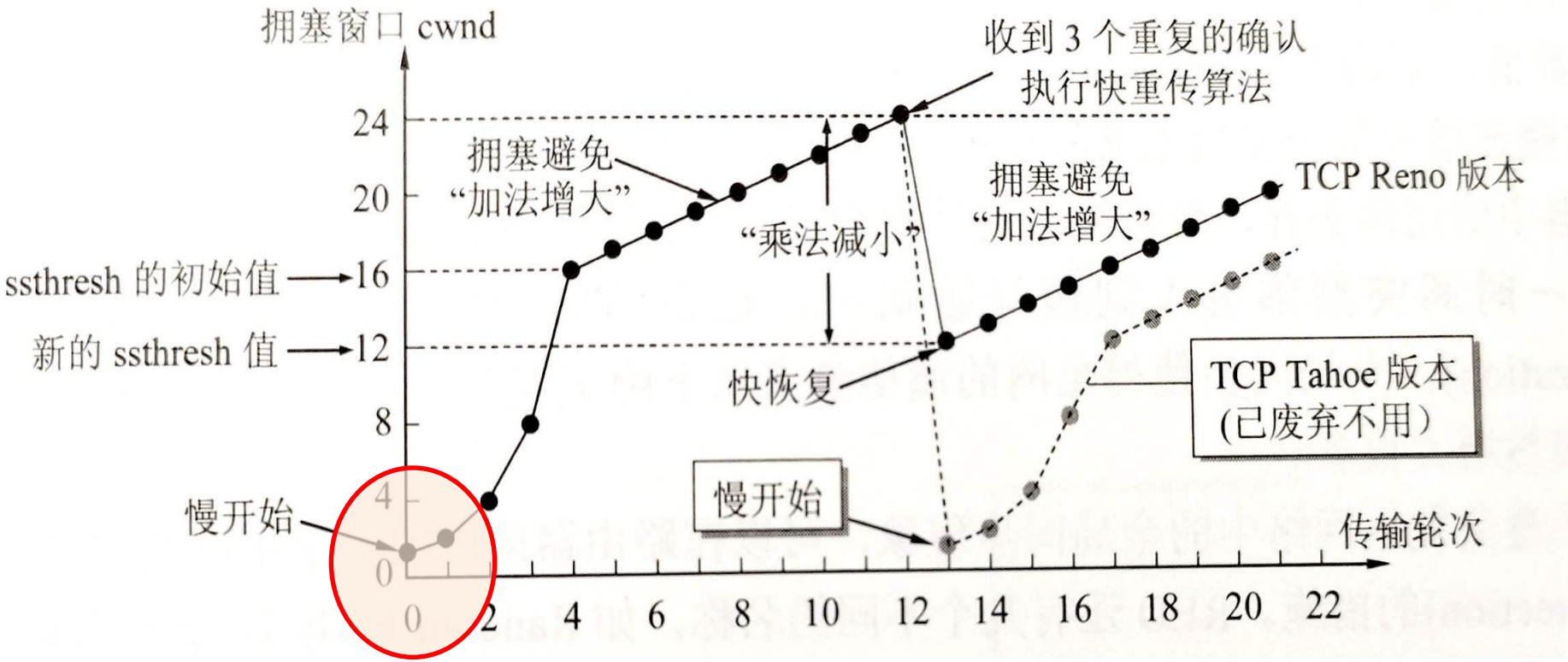


有问题的案例-2

（2）横轴取传输轮次

下图从（0，1）开始，表意为"第0次传输时cwnd=1"，表达有问题，

应改为从（1，1）开始，即"第1次传输时cwnd=1"



后续TCP改进：RFC2414 [1998]

[RFC Home] [TEXT|PDF|HTML] [Tracker] [IPR] [Info page]

Obsoleted by: 3390 EXPERIMENTAL

Network Working Group M. Allman
Request for Comments: 2414 NASA Lewis/Sterling Software
Category: Experimental S. Floyd
LBNL
C. Partridge
BBN Technologies
September 1998

Increasing TCP's Initial Window

Status of this Memo

This memo defines an Experimental Protocol for the Internet community. It does not specify an Internet standard of any kind. Discussion and suggestions for improvement are requested. Distribution of this memo is unlimited.

1. TCP Modification

This document specifies an increase in the permitted upper bound for TCP's initial window from one segment to between two and four segments. In most cases, this change results in an upper bound on the initial window of roughly 4K bytes (although given a large segment size, the permitted initial window of two segments could be significantly larger than 4K bytes). The upper bound for the initial window is given more precisely in (1):

min (4*MSS, max (2*MSS, 4380 bytes)) (1)

Equivalently, the upper bound for the initial window size is based on the maximum segment size (MSS), as follows:

```
If (MSS <= 1095 bytes)
    then win <= 4 * MSS;
If (1095 bytes < MSS < 2190 bytes)
    then win <= 4380;
If (2190 bytes <= MSS)
    then win <= 2 * MSS;
```

This increased initial window is optional: that a TCP MAY start with a larger initial window, not that it SHOULD.

规则建议变化：
cwnd 的初始值 设置为2到4

该设置是optional的，不是必须SHOULD

后续TCP改进：RFC5681 [2009]

[RFC Home] [TEXT|PDF|HTML] [Tracker] [IPR] [Errata] [Info page]

DRAFT STANDARD
Errata Exist

Network Working Group — M. Allman
Request for Comments: 5681 — V. Paxson
Obsoletes: 2581 — ICSI
Category: Standards Track — E. Blanton
Purdue University
September 2009

TCP Congestion Control

Abstract

This document defines TCP's four intertwined congestion control algorithms: slow start, congestion avoidance, fast retransmit, and fast recovery. In addition, the document specifies how TCP should begin transmission after a relatively long idle period, as well as discussing various acknowledgment generation methods. This document obsoletes RFC 2581.

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Beginning transmission into a network with unknown conditions requires TCP to slowly probe the network to determine the available capacity, in order to avoid congesting the network with an inappropriately large burst of data. The slow start algorithm is used for this purpose at the beginning of a transfer, or after repairing loss detected by the retransmission timer. Slow start additionally serves to start the "ACK clock" used by the TCP sender to release data into the network in the slow start, congestion avoidance, and loss recovery algorithms.

IW, the initial value of cwnd, MUST be set using the following guidelines as an upper bound.

If SMSS > 2190 bytes:
  IW = 2 * SMSS bytes and MUST NOT be more than 2 segments
If (SMSS > 1095 bytes) and (SMSS <= 2190 bytes):
  IW = 3 * SMSS bytes and MUST NOT be more than 3 segments
if SMSS <= 1095 bytes:
  IW = 4 * SMSS bytes and MUST NOT be more than 4 segments

As specified in [RFC3390], the SYN/ACK and the acknowledgment of the SYN/ACK MUST NOT increase the size of the congestion window. Further, if the SYN or SYN/ACK is lost, the initial window used by a sender after a correctly transmitted SYN MUST be one segment consisting of at most SMSS bytes.

A detailed rationale and discussion of the IW setting is provided in [RFC3390].

cwnd **的初始值 设置为**2**到**4

后续TCP改进：RFC5681 [2009]

ssthresh **的最小值为**2
cwnd**在拥塞后重置为**1

When a TCP sender detects segment loss using the retransmission timer and the given segment has not yet been resent by way of the retransmission timer, the value of ssthresh MUST be set to no more than the value given in equation (4):

ssthresh = max (FlightSize / 2, 2*SMSS) (4)

where, as discussed above, FlightSize is the amount of outstanding data in the network.

On the other hand, when a TCP sender detects segment loss using the retransmission timer and the given segment has already been retransmitted by way of the retransmission timer at least once, the value of ssthresh is held constant.

Implementation Note: An easy mistake to make is to simply use cwnd, rather than FlightSize, which in some implementations may incidentally increase well beyond rwnd.

Furthermore, upon a timeout (as specified in [RFC2988]) cwnd MUST be set to no more than the loss window, LW, which equals 1 full-sized segment (regardless of the value of IW). Therefore, after retransmitting the dropped segment the TCP sender uses the slow start algorithm to increase the window from 1 full-sized segment to the new value of ssthresh, at which point congestion avoidance again takes over.

后续TCP改进：RFC6928 [2013]

[RFC Home] [TEXT|PDF|HTML] [Tracker] [IPR] [Info page]

EXPERIMENTAL

Internet Engineering Task Force (IETF)
Request for Comments: 6928
Category: Experimental
ISSN: 2070-1721

J. Chu
N. Dukkipati
Y. Cheng
M. Mathis
Google, Inc.

April 2013

Increasing TCP's Initial Window

Abstract

This document proposes an experiment to increase the permitted TCP initial window (IW) from between 2 and 4 segments, as specified in RFC 3390, to 10 segments with a fallback to the existing recommendation when performance issues are detected. It discusses the motivation behind the increase, the advantages and disadvantages of the higher initial window, and presents results from several large-scale experiments showing that the higher initial window improves the overall performance of many web services without resulting in a congestion collapse. The document closes with a discussion of usage and deployment for further experimental purposes recommended by the IETF TCP Maintenance and Minor Extensions (TCPM) working group.

A complementary set of slides for this proposal can be found at [CD10].

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. TCP Modification

This document proposes an increase in the permitted upper bound for TCP's initial window (IW) to 10 segments, depending on the maximum segment size (MSS). This increase is optional: a TCP MAY start with an initial window that is smaller than 10 segments.

More precisely, the upper bound for the initial window will be

```
min (10*MSS, max (2*MSS, 14600))          (1)
```

This upper bound for the initial window size represents a change from RFC 3390 [RFC3390], which specified that the congestion window be initialized between 2 and 4 segments, depending on the MSS.

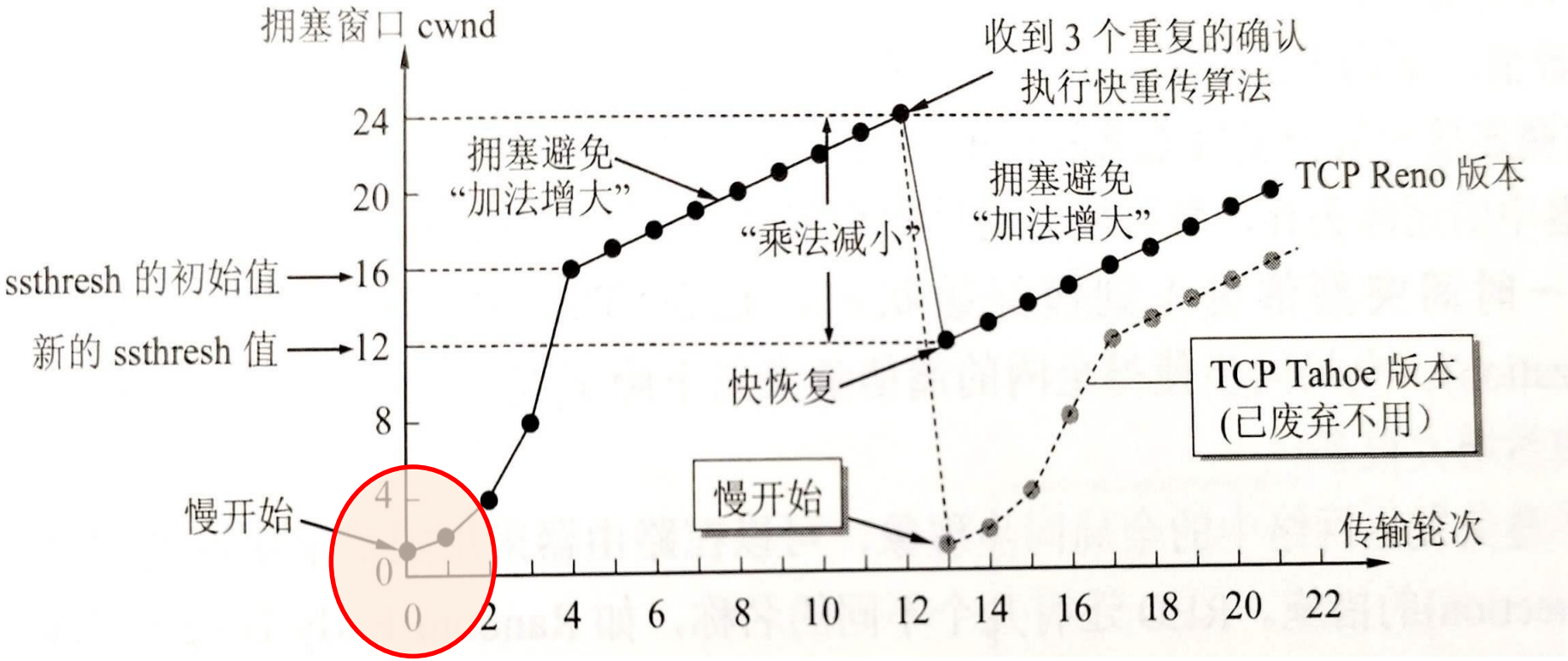
This change applies to the initial window of the connection in the first round-trip time (RTT) of data transmission during or following the TCP three-way handshake. Neither the SYN/ACK nor its ACK in the three-way handshake should increase the initial window size.

规则建议变化：
cwnd **的初始值 设置为**2**到**10

有问题的案例的修改

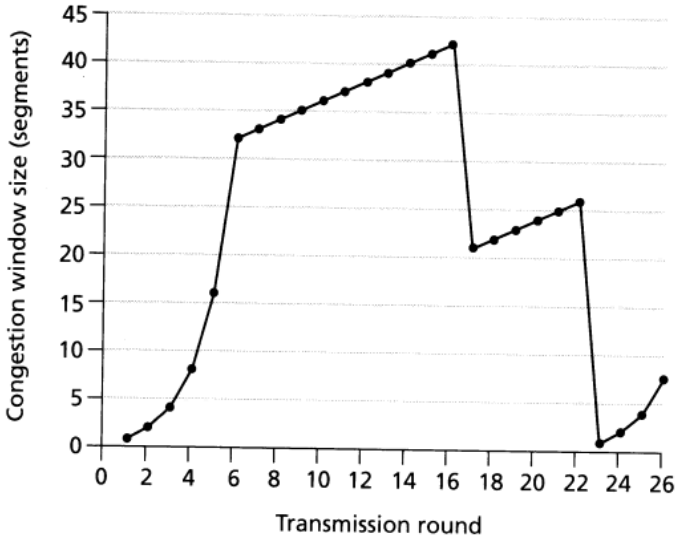
下图的修改：去掉（0，1）点，改为从（1，2）开始，并注明“cwnd初始窗口从2开始”，也是对的

——在本课程不推荐这种作图



结论

- 本课程中推荐笔试中TCP拥塞窗口作图方法



- 横轴以传输轮次 transmission round 为单位，从1 开始
- 纵轴以拥塞窗口cwnd的segment为单位，从1开始
- 作图从（1，1）开始