John Rauser
Velocity
June, 2010

TCP and the Lower Bound of Web Performance

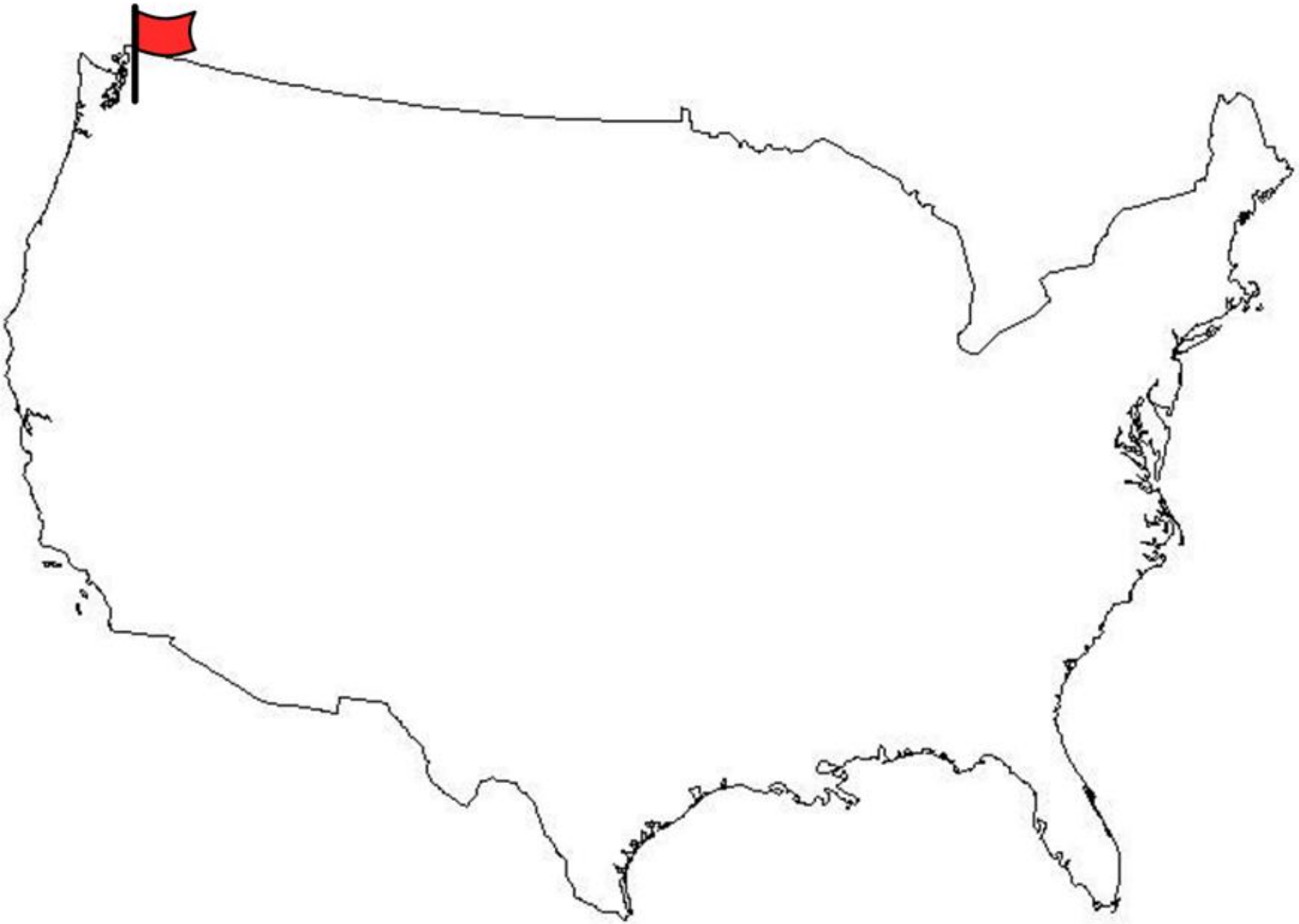
1996

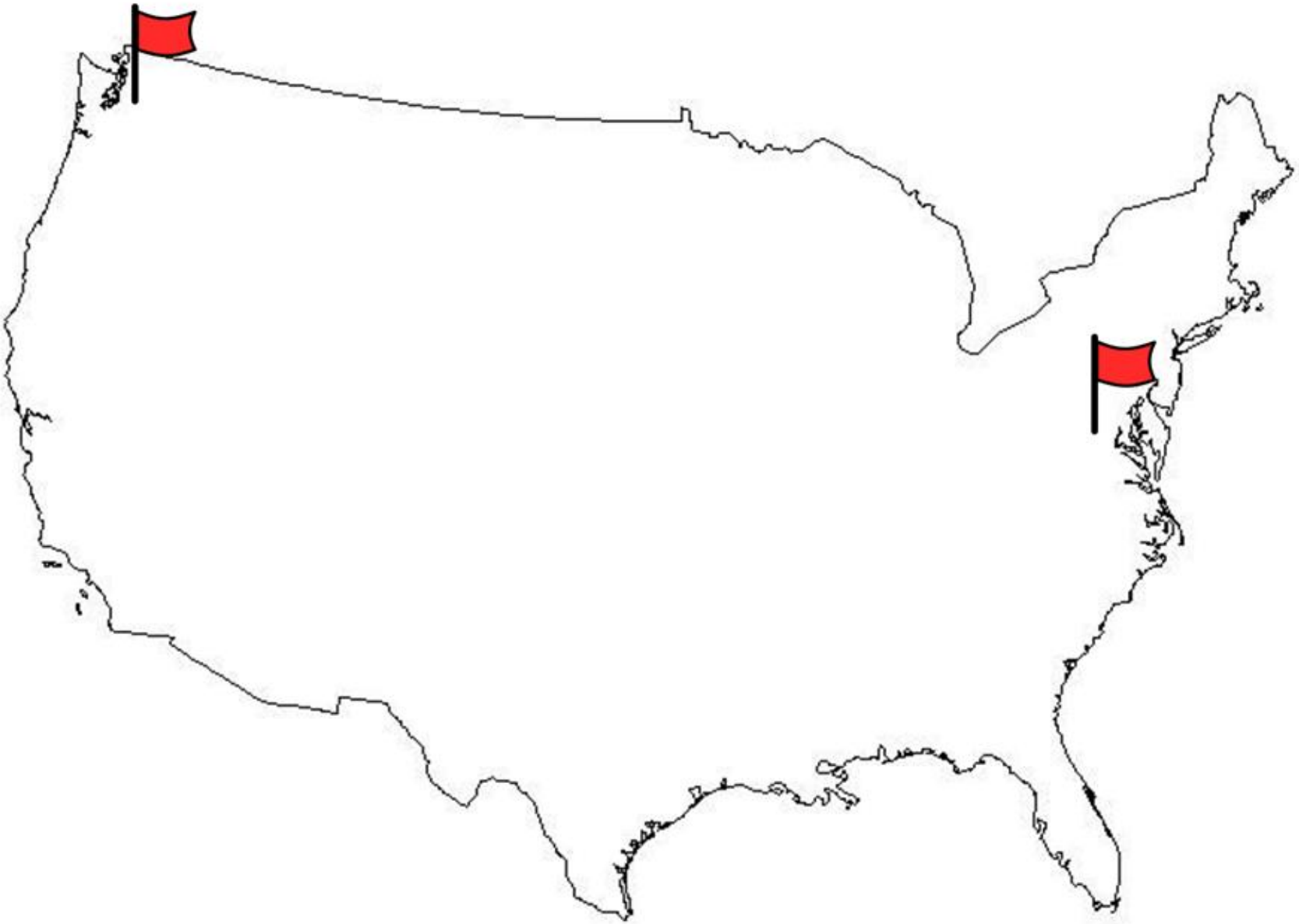
It’s the Latency, Stupid

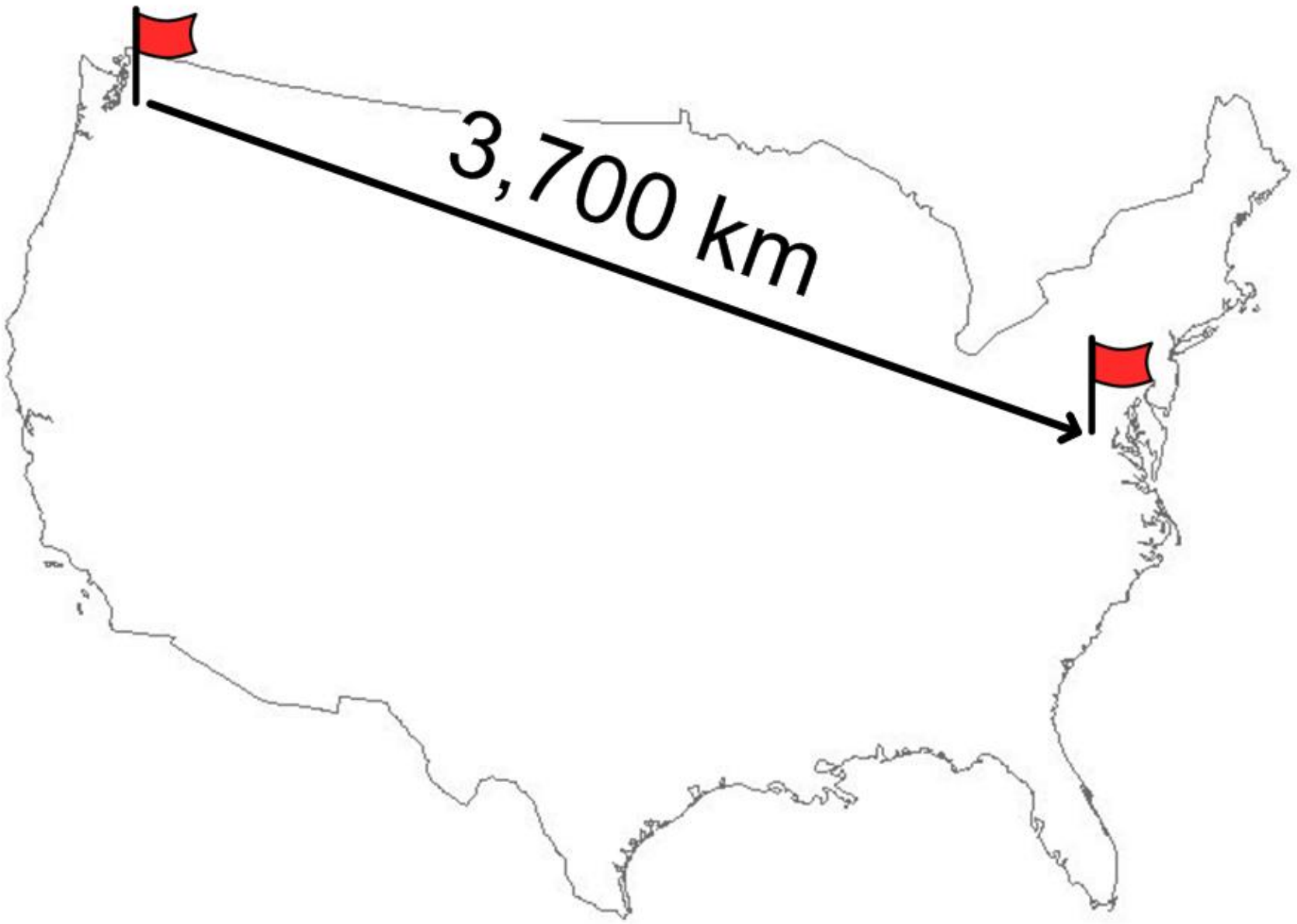
http://rescomp.stanford.edu/~cheshire/rants/Latency.html

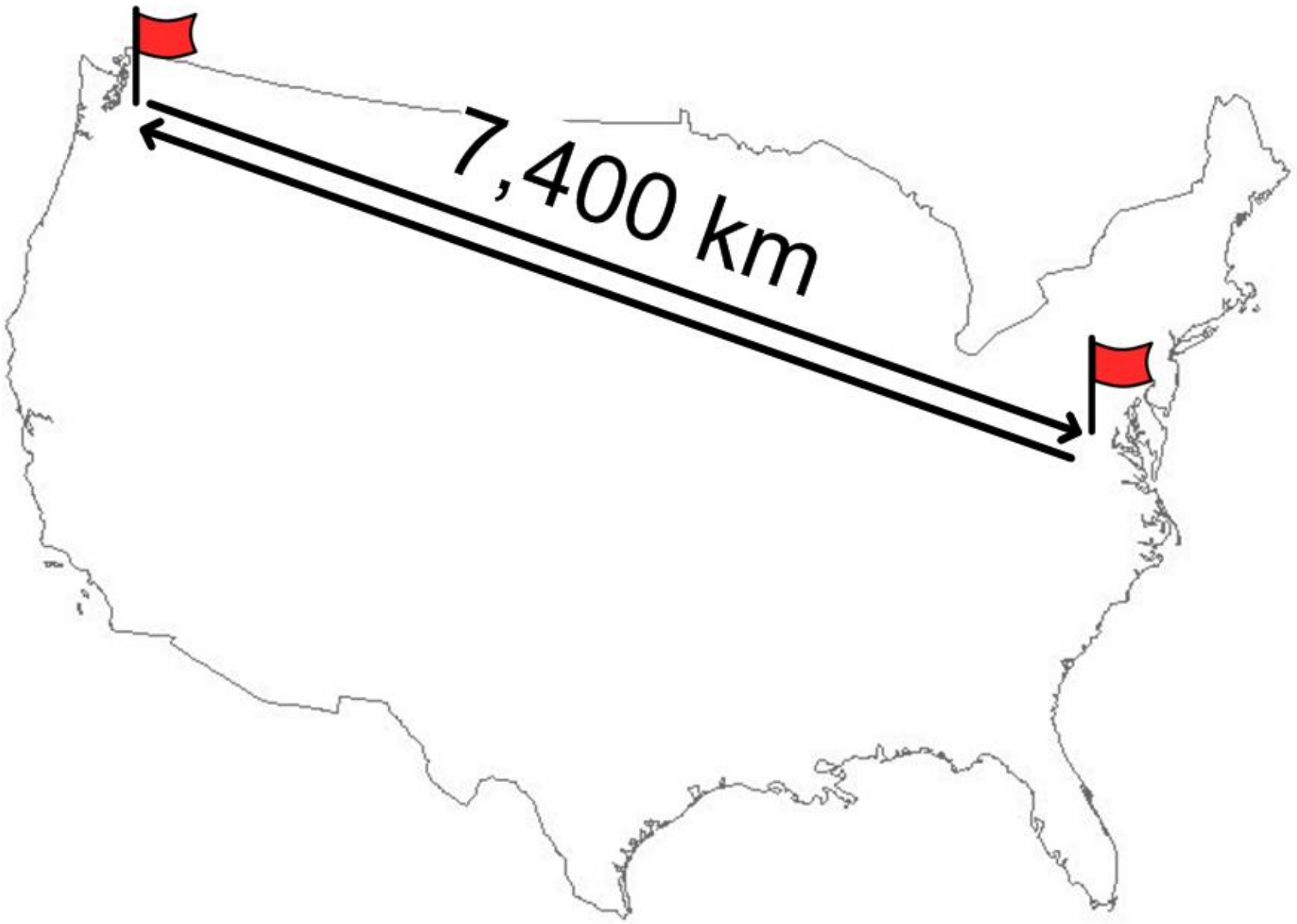
1) “Making more bandwidth is easy.”

2) “Once you have bad latency you're stuck with it.”






3,700 km


7,400 km

speed of light in km/s

Search

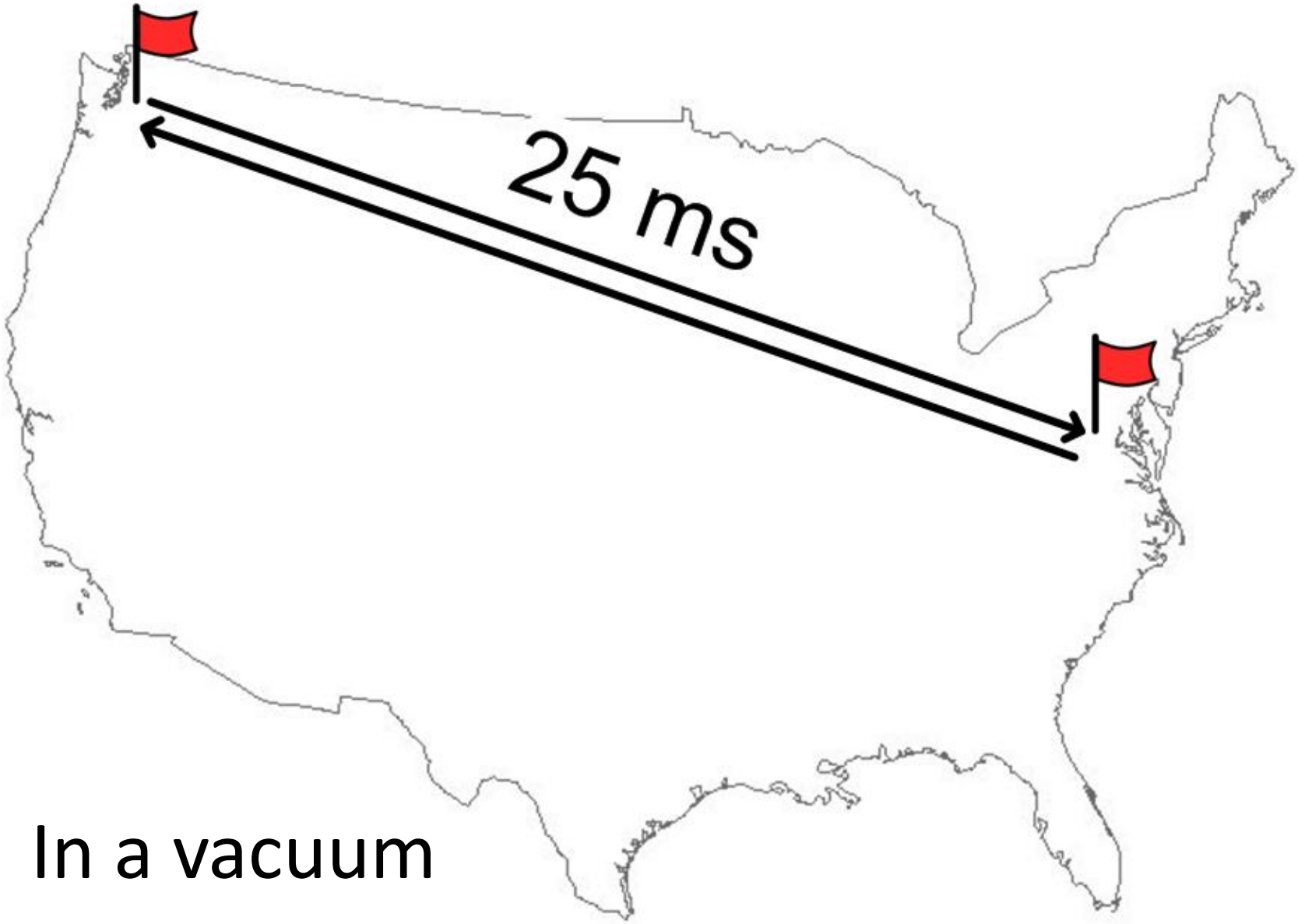
Advanced Search
Preferences

Web Show options...

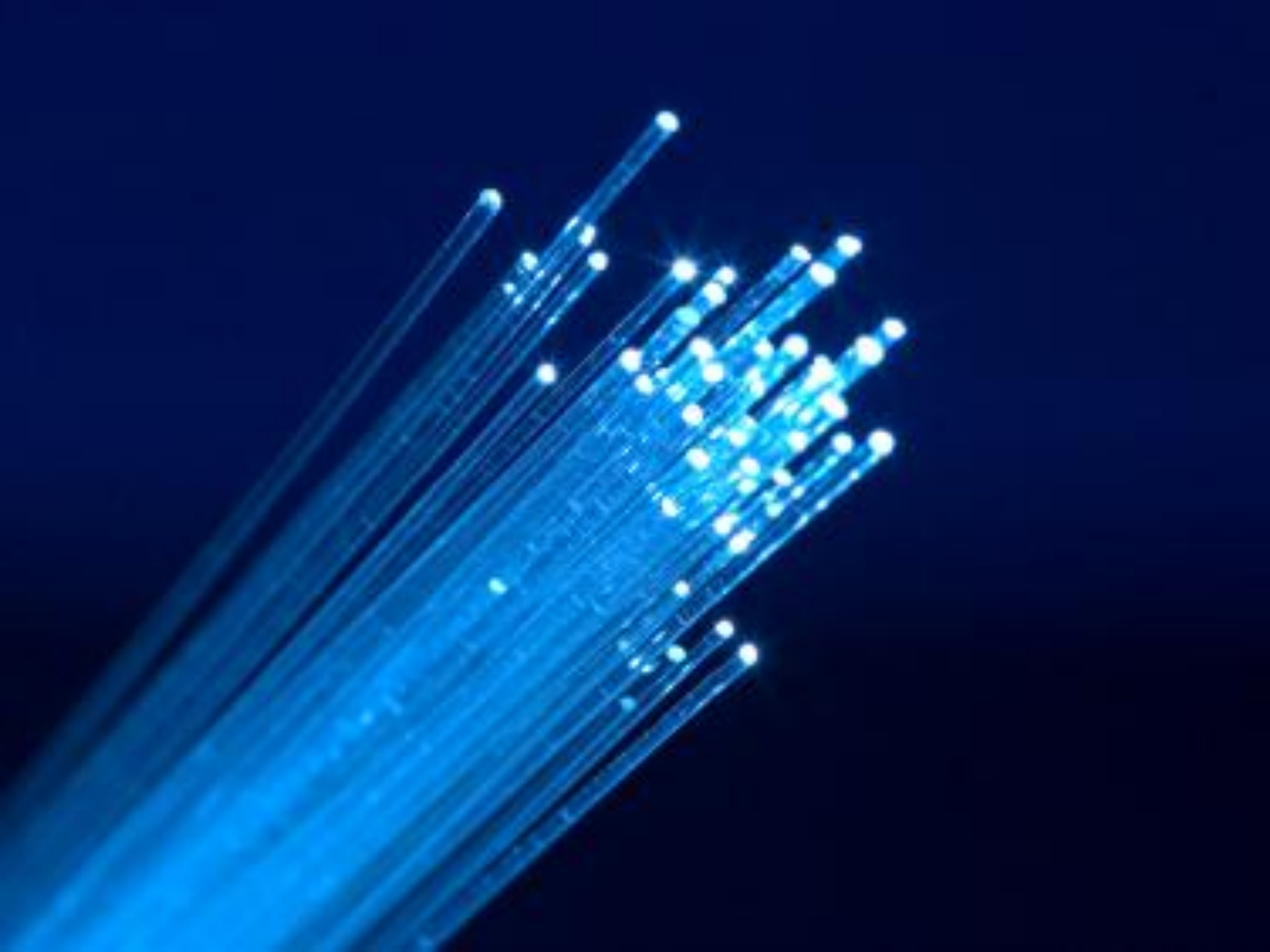
the speed of light = 299 792.458 kilometers / second

More about calculator.

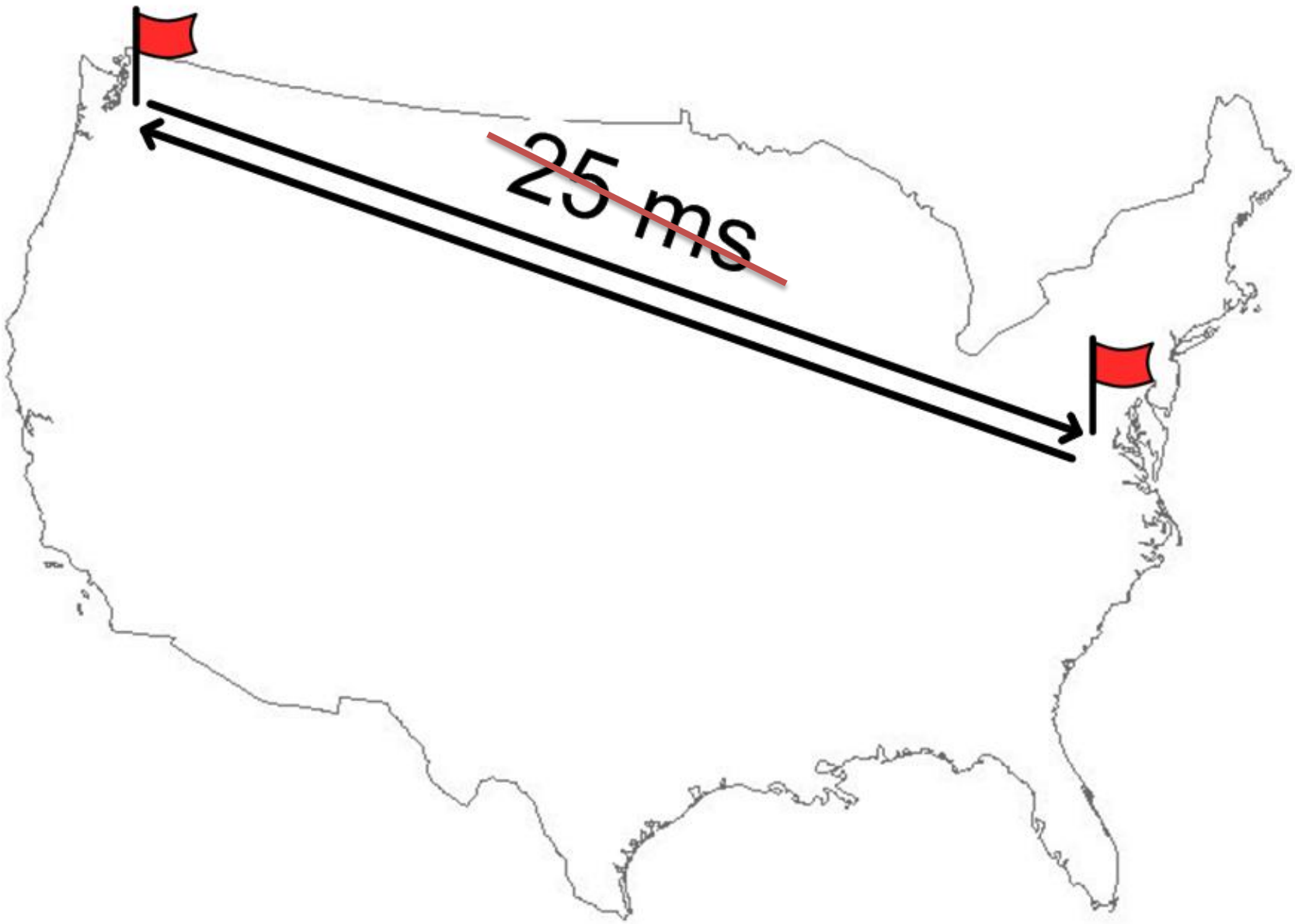
$$\frac{7{,}400 \text{ km}}{300{,}000 \text{ km/sec}} = 25 \text{ ms}$$

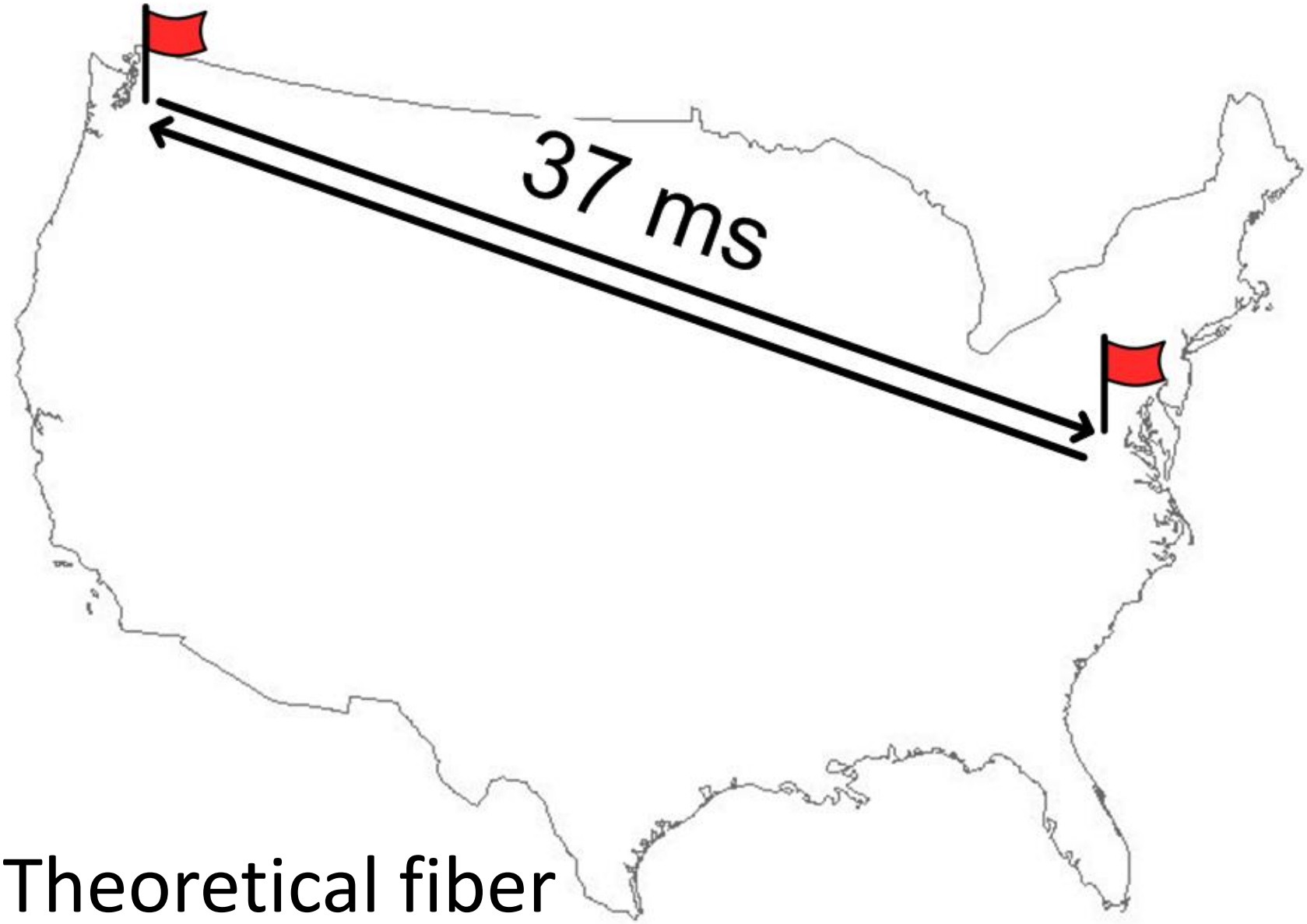


In a vacuum



$$\frac{1}{1.5} = 0.66$$

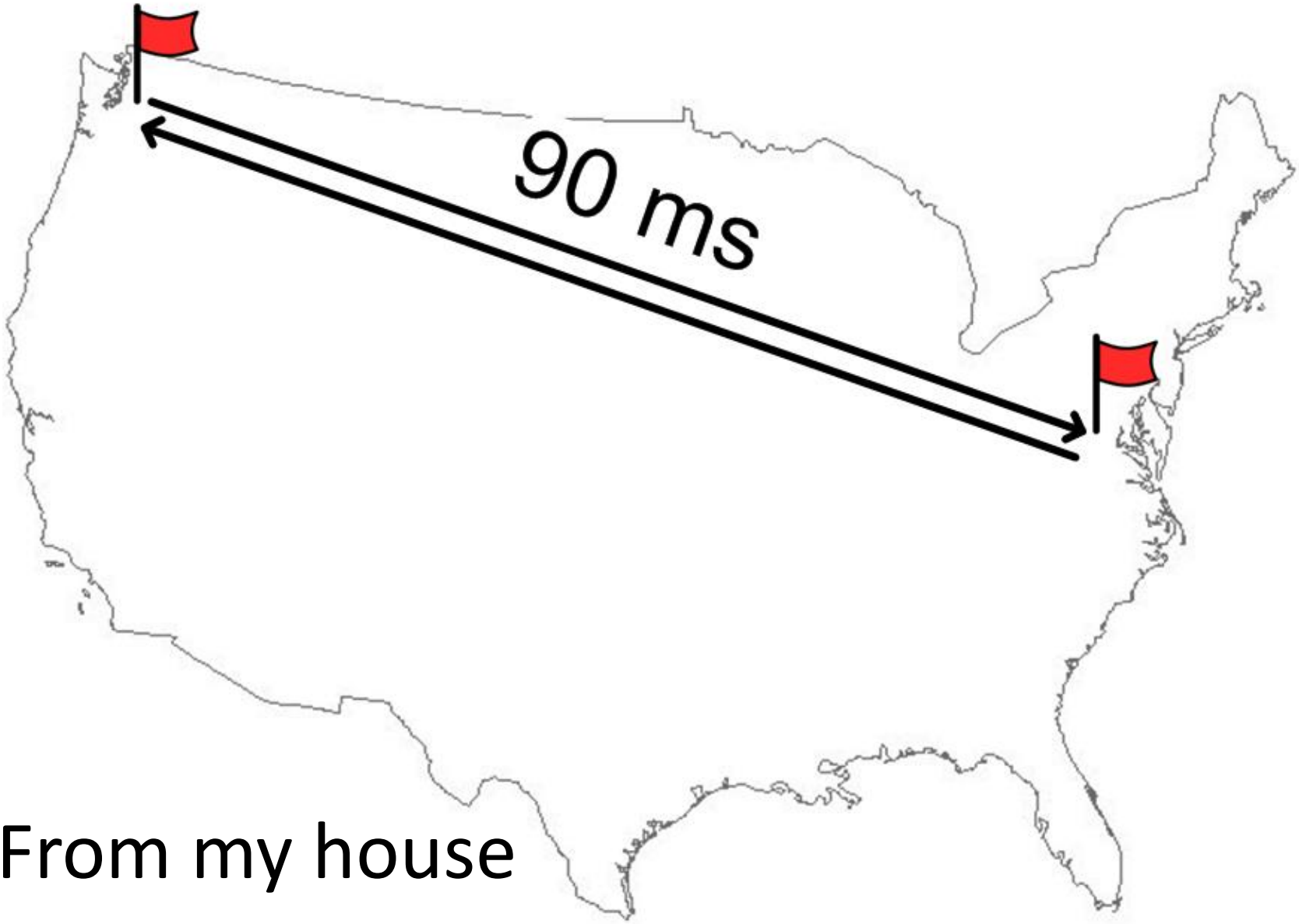

25 ms



Theoretical fiber

From my house

```
Ping statistics:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 91ms, Maximum = 98ms, Average = 93ms
```



From my house

$$\frac{\text{From my house: } 90 \text{ ms}}{\text{Theoretical fiber: } 37 \text{ ms}} \sim 2$$

It’s been this way for over a decade.

“Once you have bad latency you're stuck with it.”

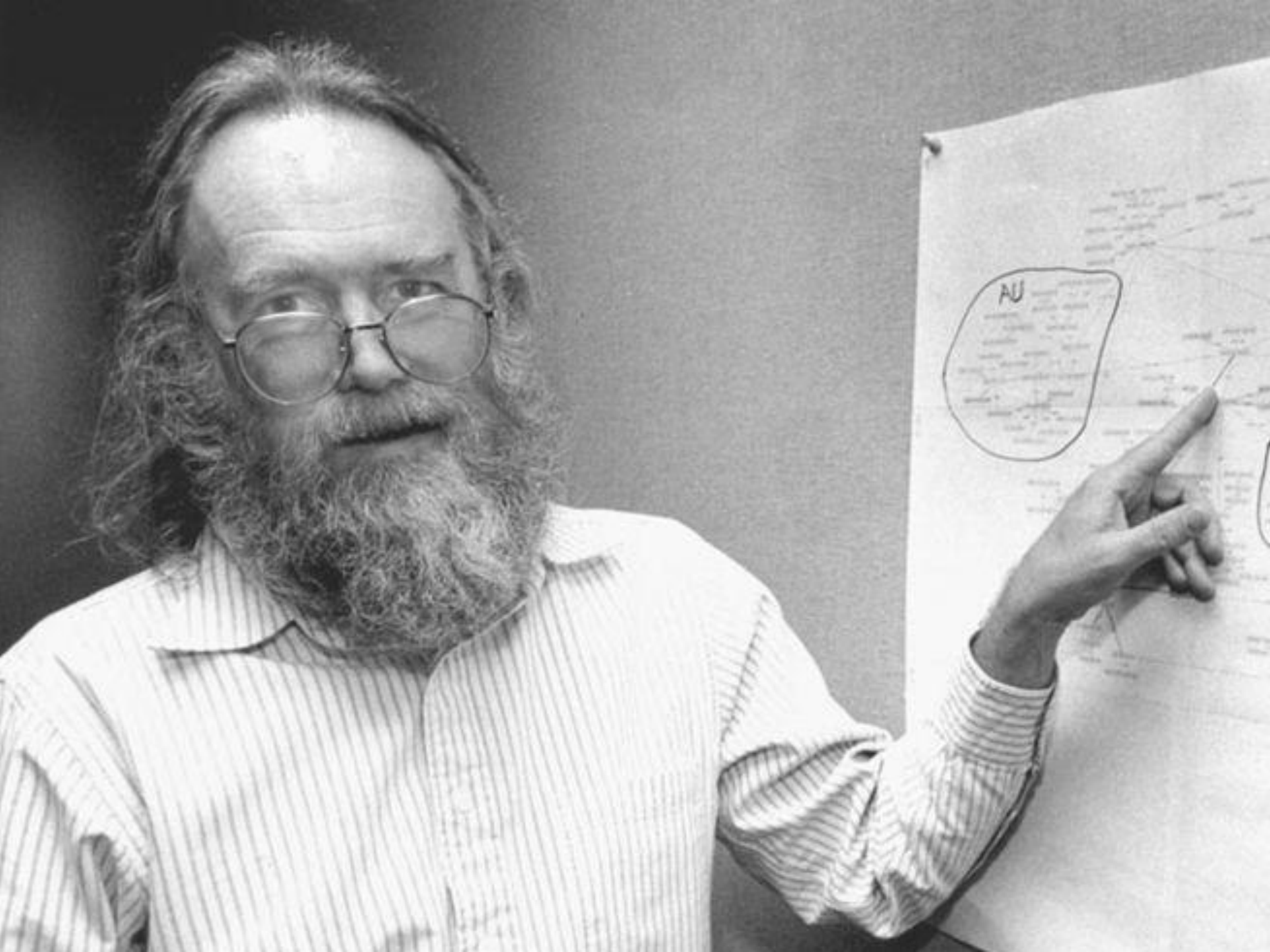

Fascinating!


WinBolo
File
Edit
WinBolo
Players
Brains
Help
Alice
Jack
0
0
ce" Alice captured a Neutral Base Alice captured a Neutral Pillbox

Network latency matters for web applications

History of the Internet

September 1981


AU

RFC 793

Transmission Control Protocol

Transmission Control Protocol

Basic Data Transfer
Reliability
Flow Control
Multiplexing
Connections
Precedence and Security

Basic Data Transfer

Reliability

Flow Control

Multiplexing

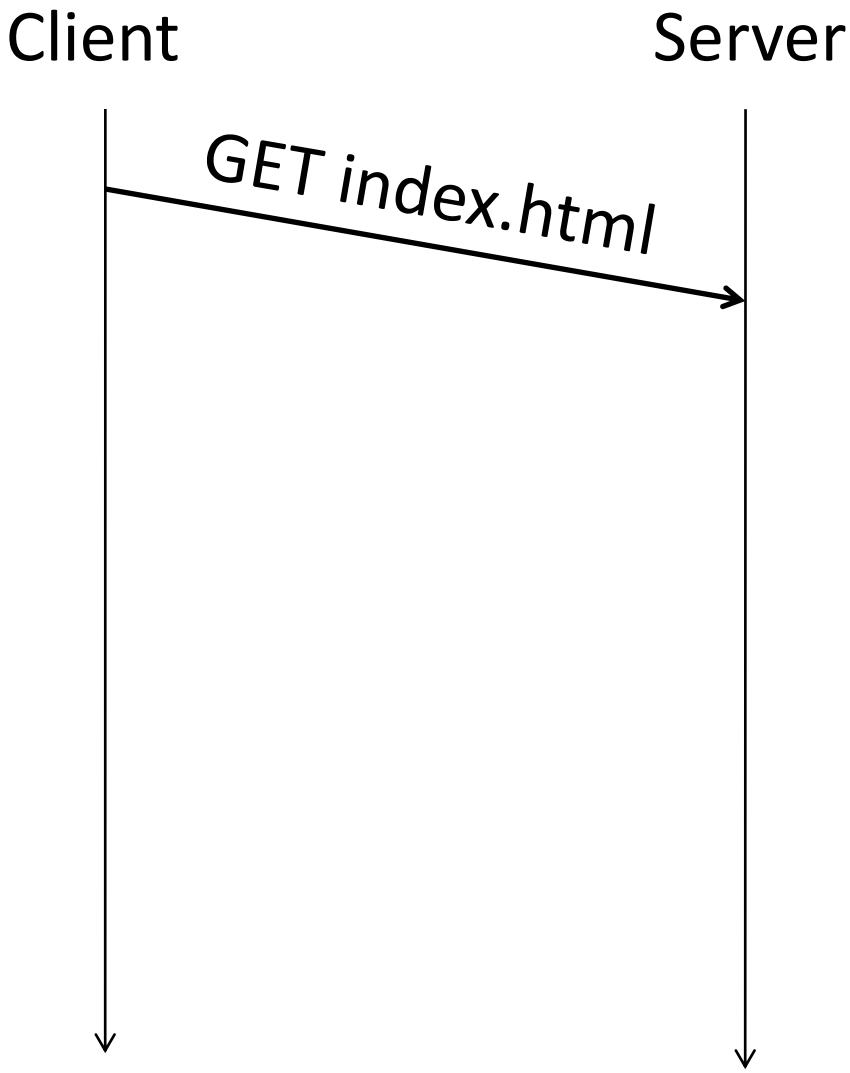
Connections

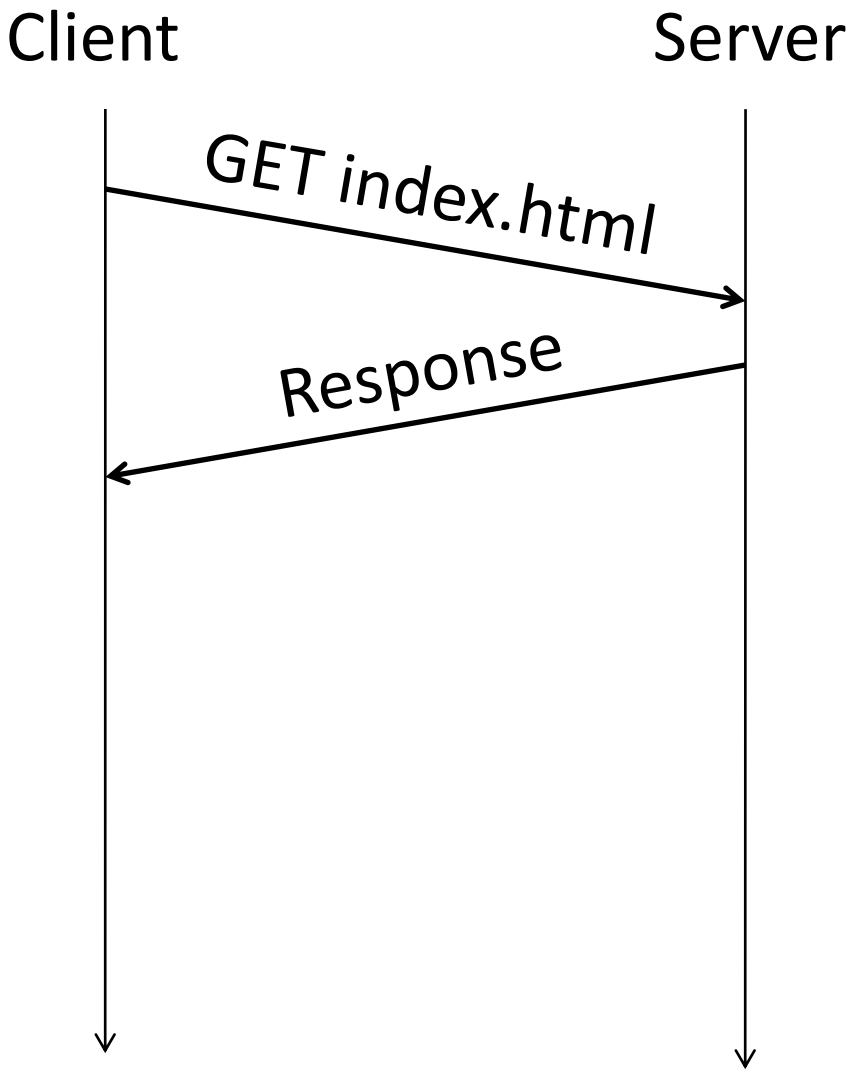
Precedence and Security

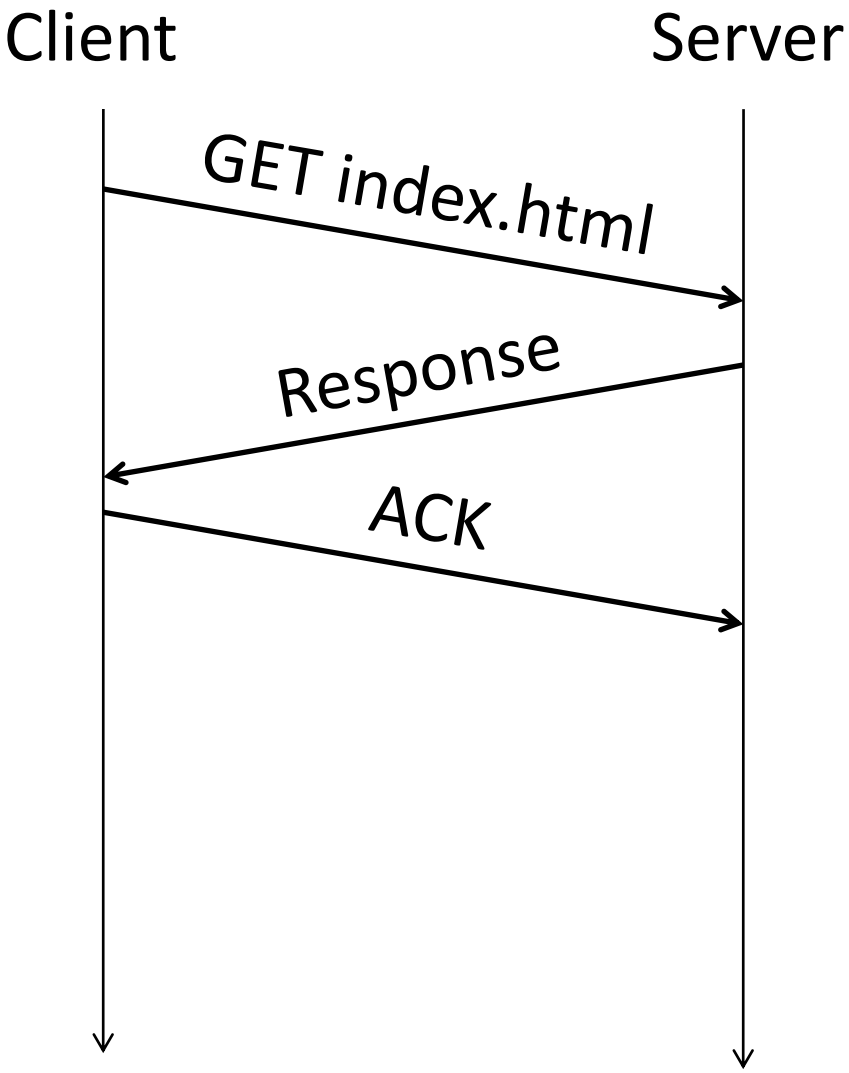
Reliability

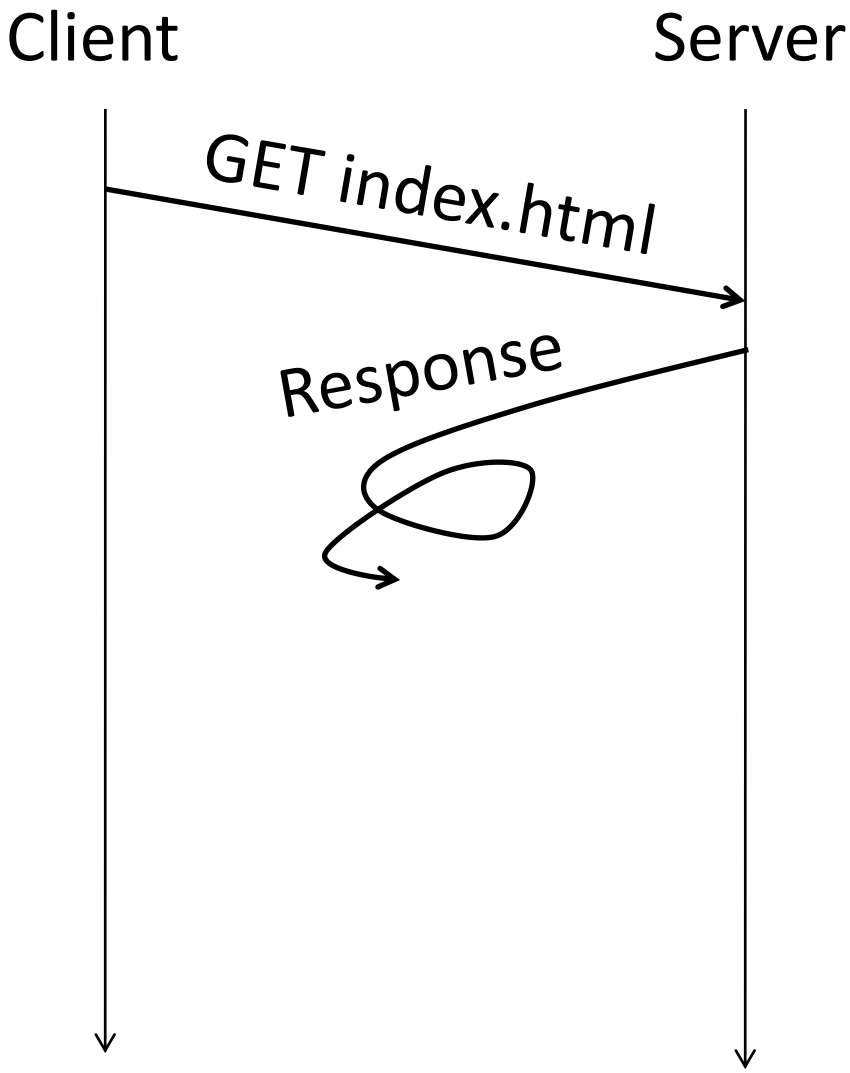
“This is achieved by… requiring a positive acknowledgment (ACK) from the receiving TCP. If the ACK is not received within a timeout interval, the data is retransmitted.”

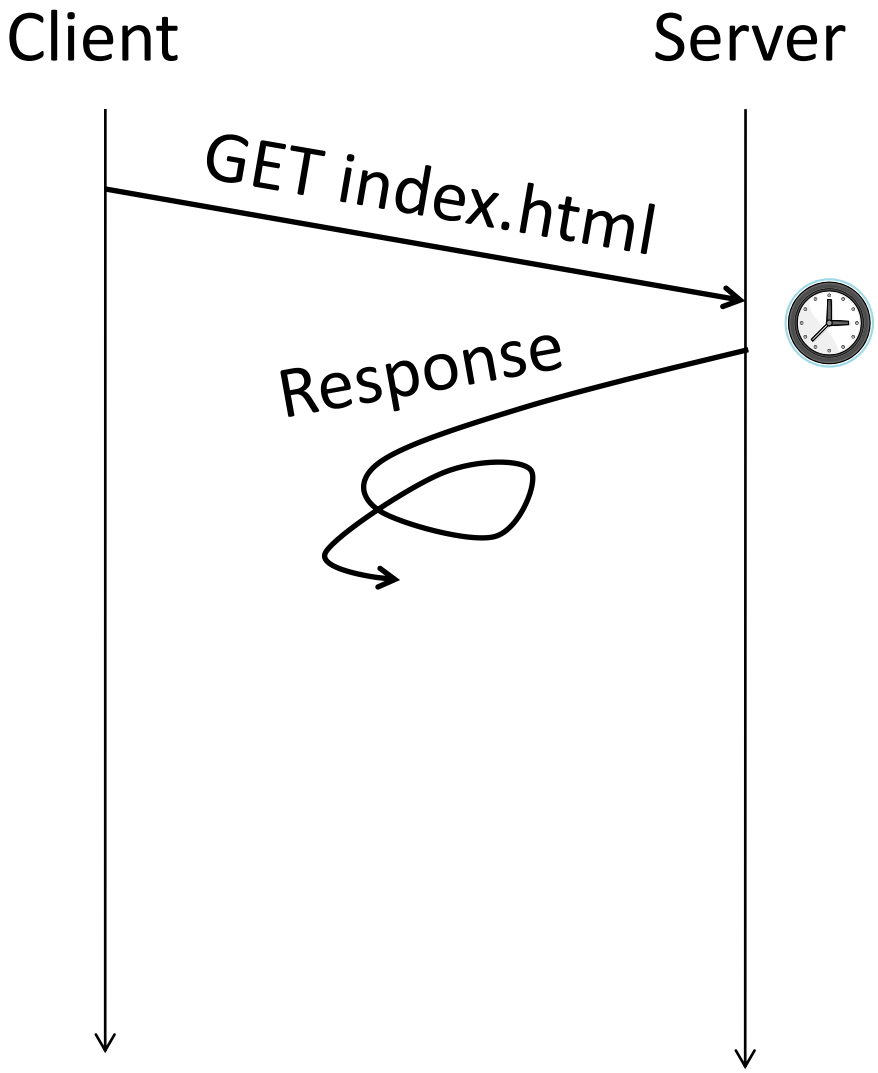
-RFC 793

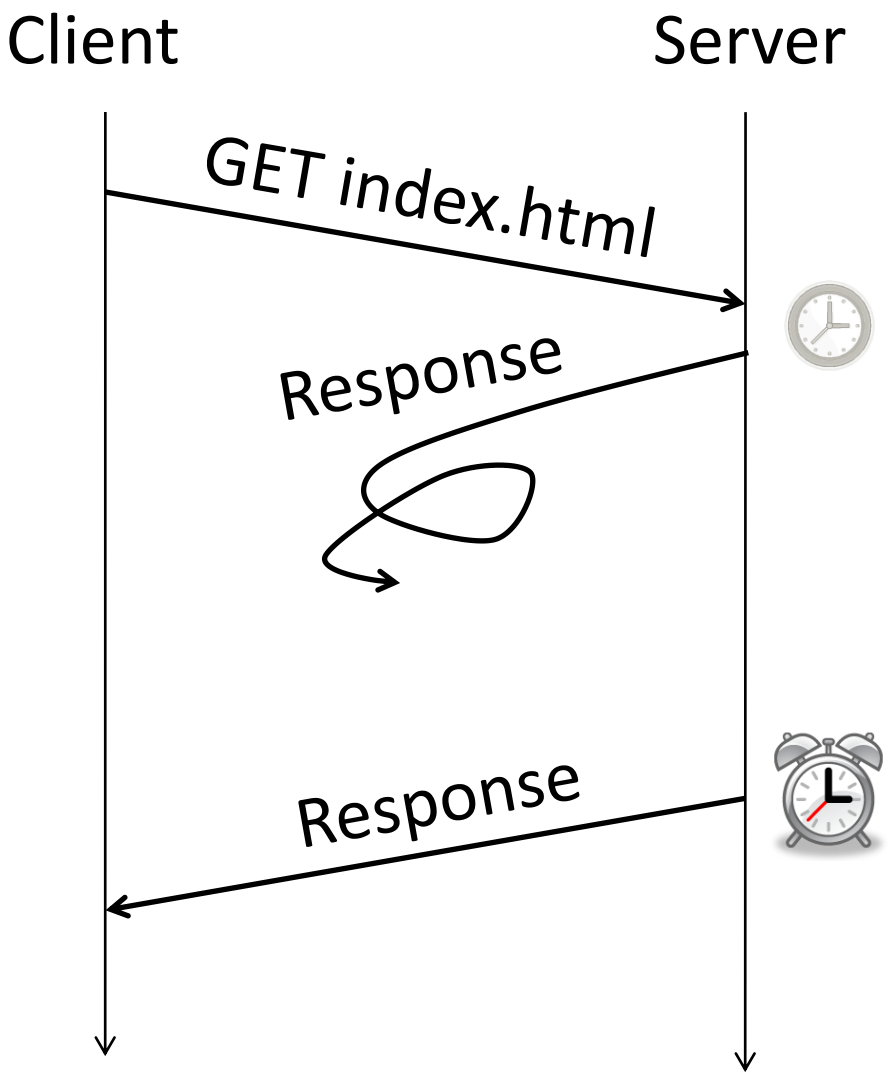

Client
Server
GET index.html

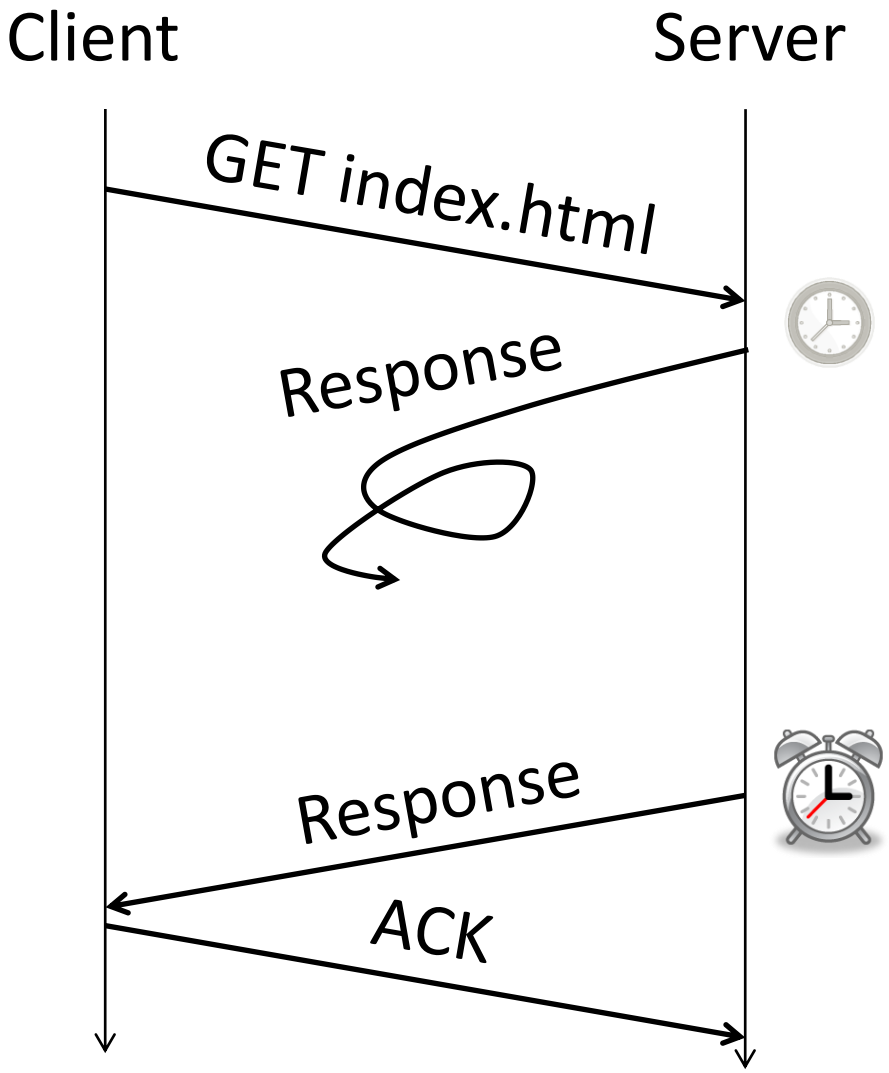

Client
Server
GET index.html
Response


Client
Server
GET index.html
Response
ACK


Client
Server
GET index.html
Response


Client
Server
GET index.html
Response


Client
Server
GET index.html
Response
Response


Client
Server
GET index.html
Response
Response
ACK

Flow Control

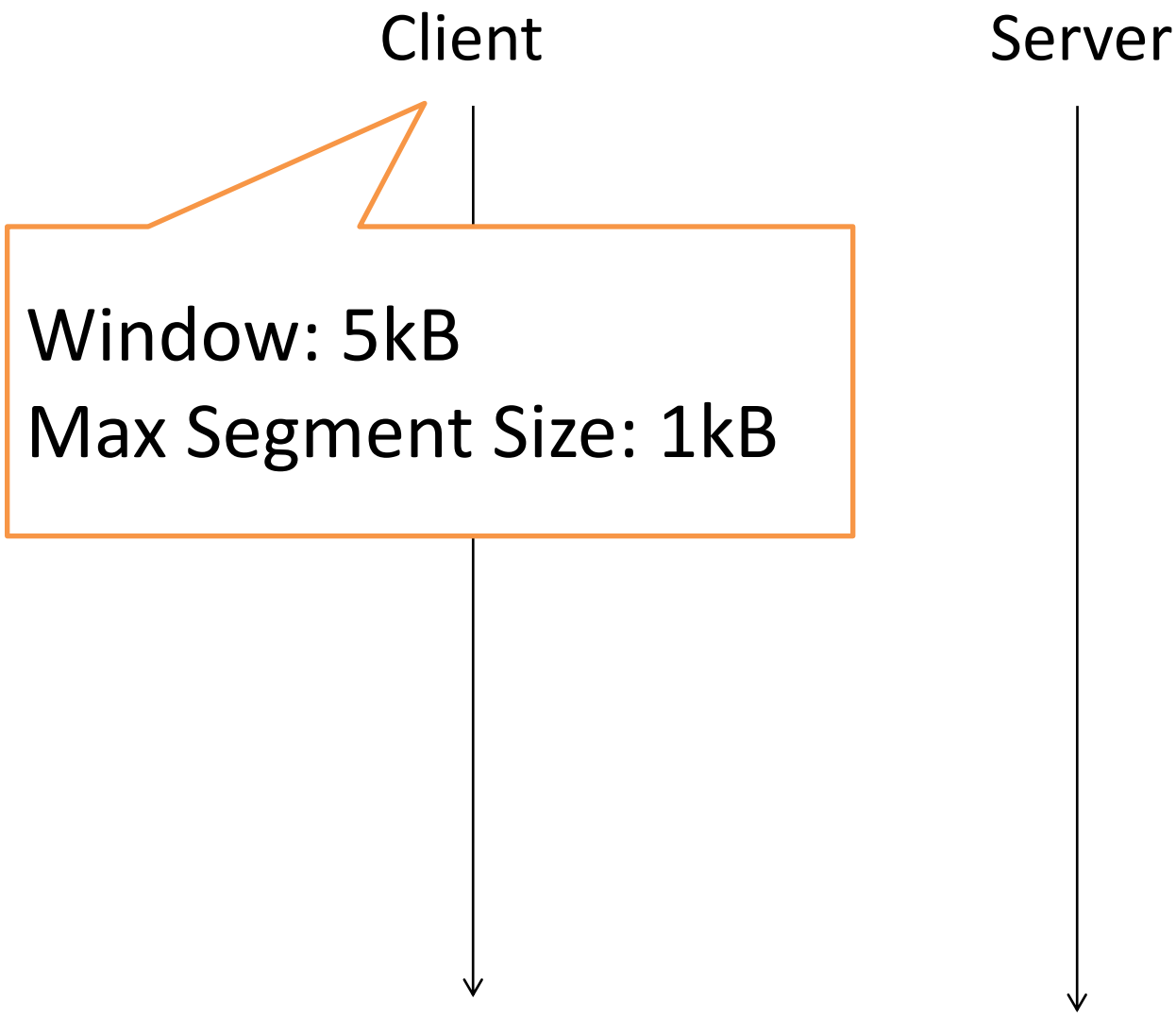
“This is achieved by returning a ‘window’ with every ACK indicating a range of acceptable sequence numbers beyond the last segment successfully received. The window indicates an allowed number of octets that the sender may transmit before receiving further permission.”

-RFC 793

"This is achieved by returning a 'window' with every ACK indicating a range of acceptable sequence numbers beyond the last segment successfully received. **The window indicates an allowed number of octets that the sender may transmit before receiving further permission."**

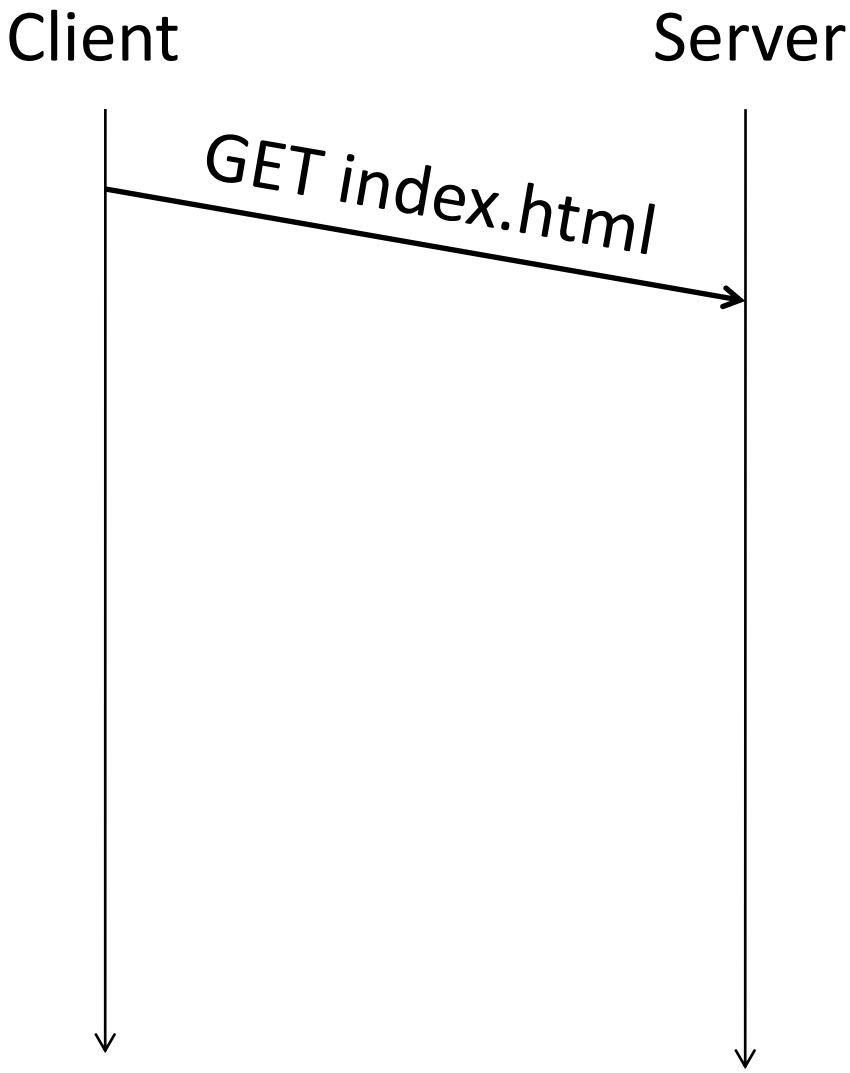
-RFC 793

TCP Window: The maximum amount of un-ACKed data in flight.


Client
Server
Window: 5kB
Max Segment Size: 1kB

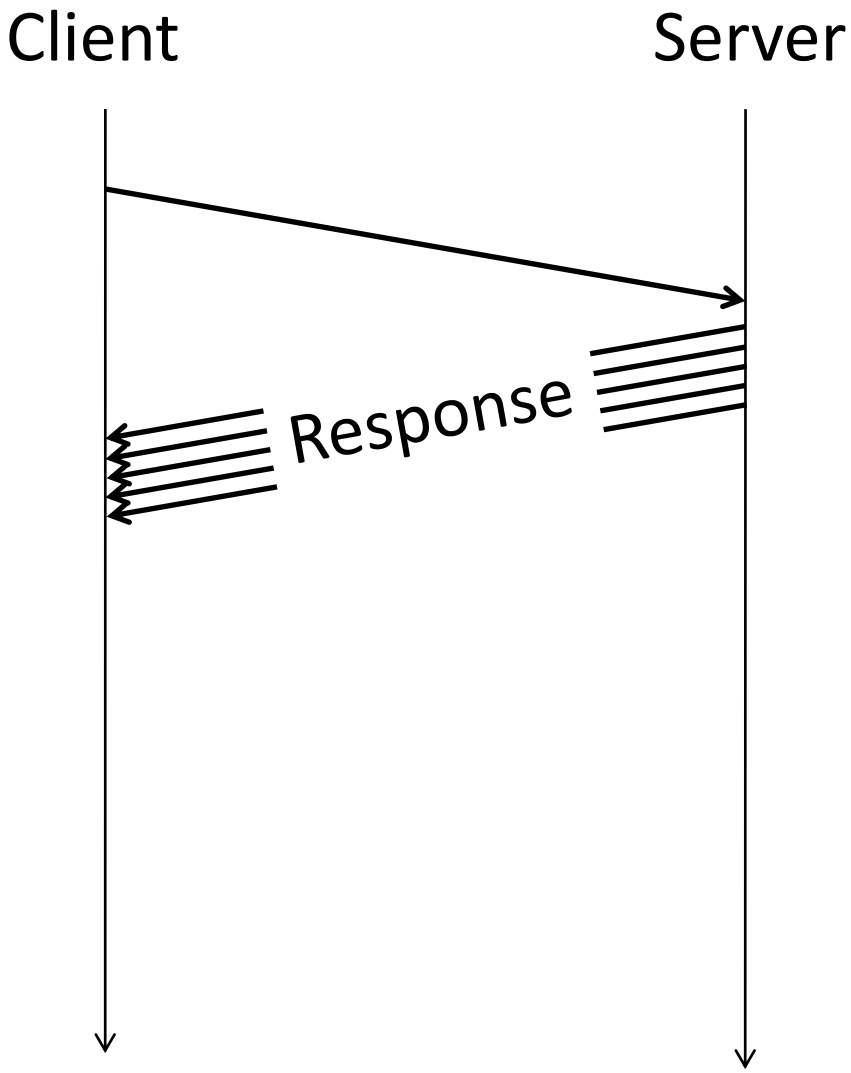
Window: 5kB

Max Segment: 1kB



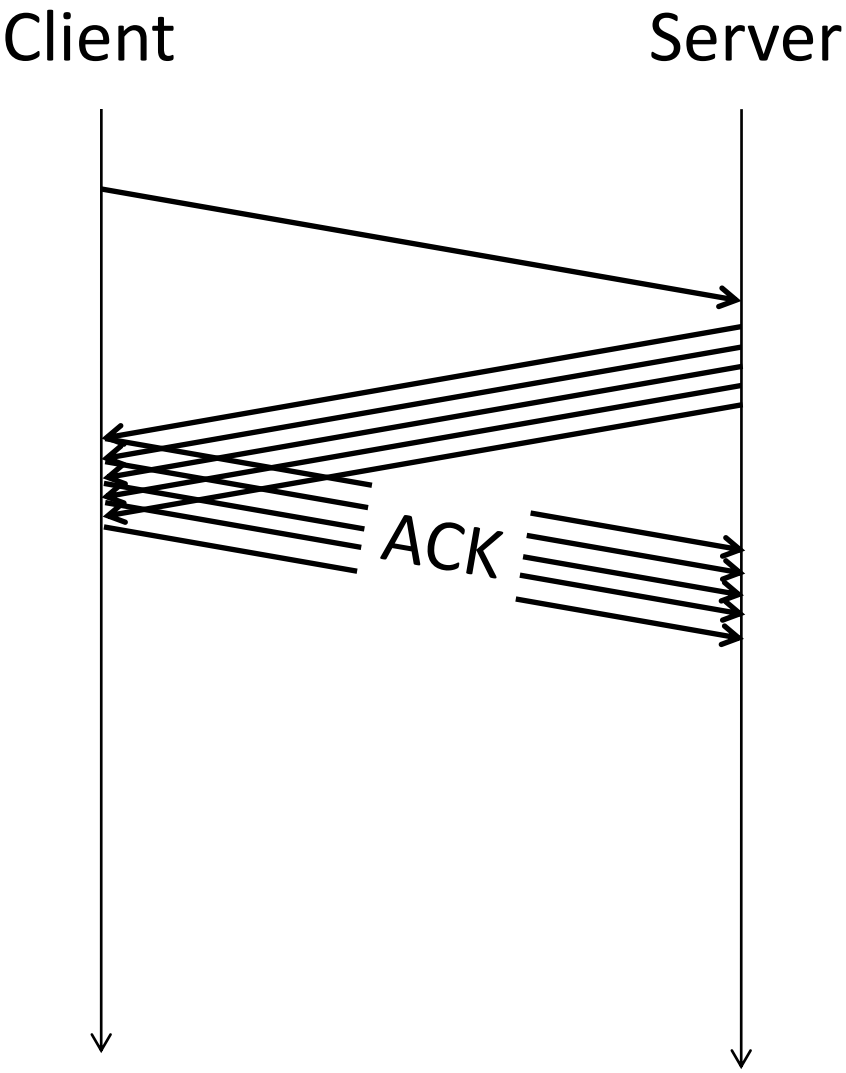
Window: 5kB

Max Segment: 1kB



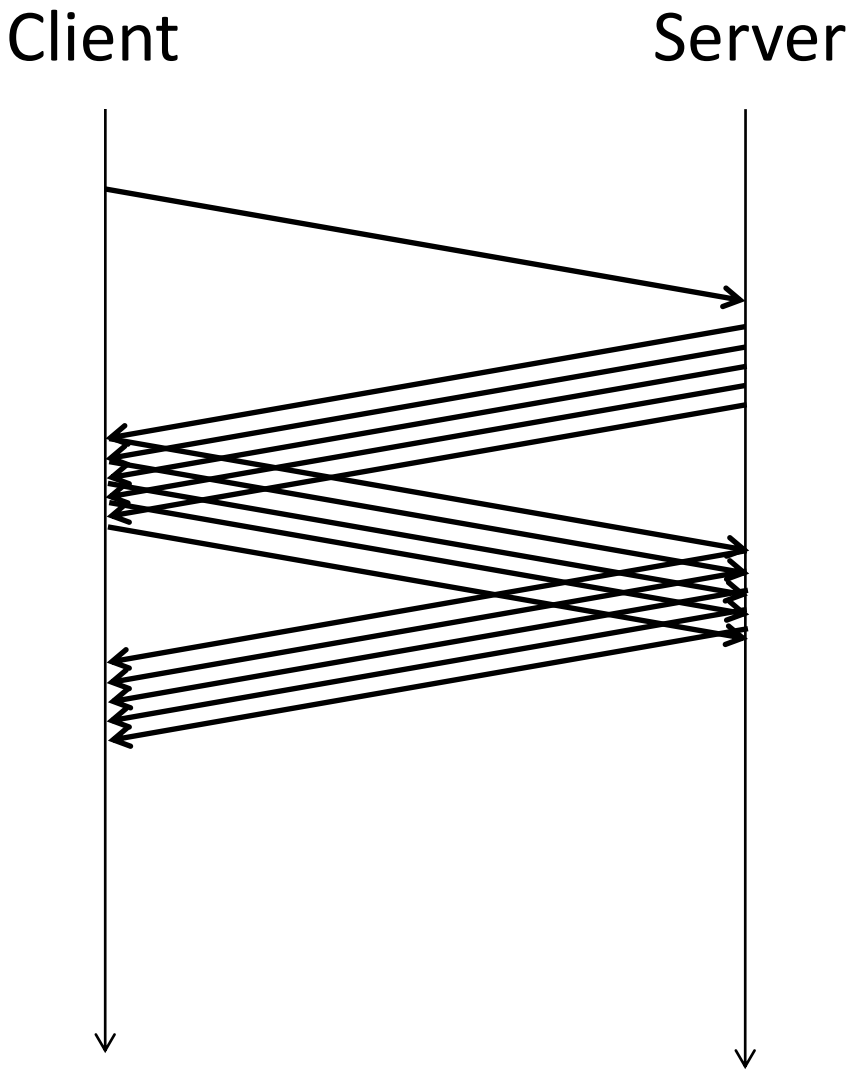
Window: 5kb

Max Segment: 1kb



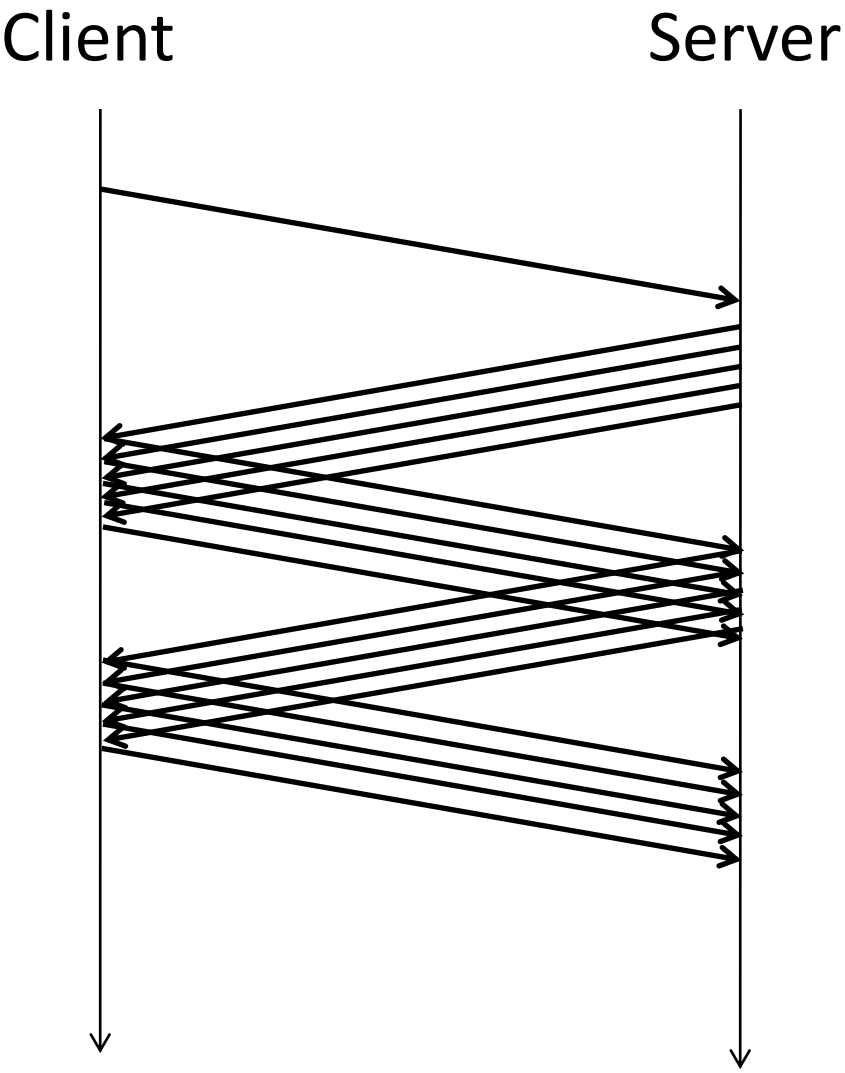
Window: 5kB

Max Segment: 1kB



Window: 5kB

Max Segment: 1kB



September 1981

RFC 791, 793

TCP/IP

September 1981

213 hosts

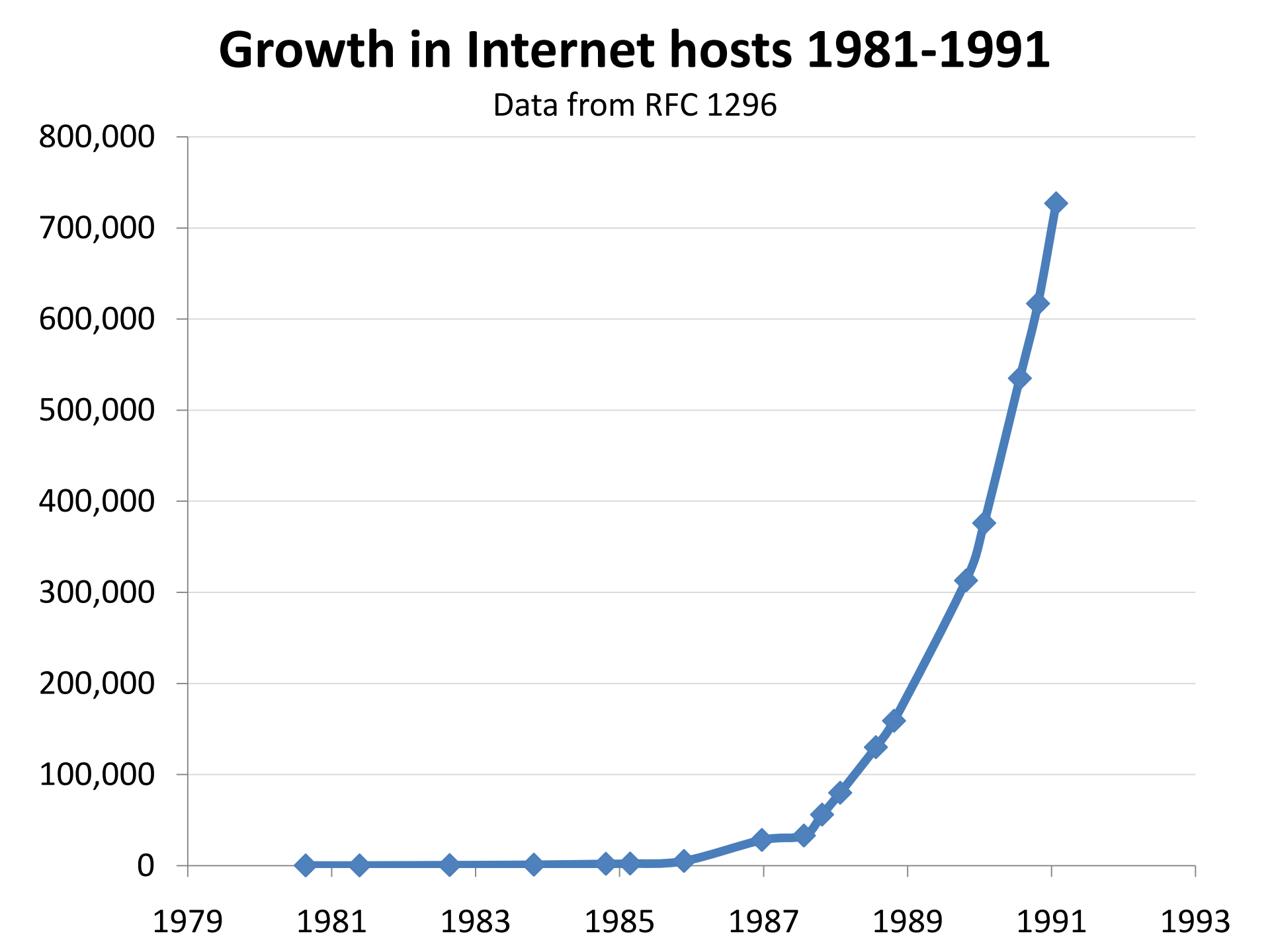
May 1982

235 hosts


MACSS
XC68000L
R9M7951

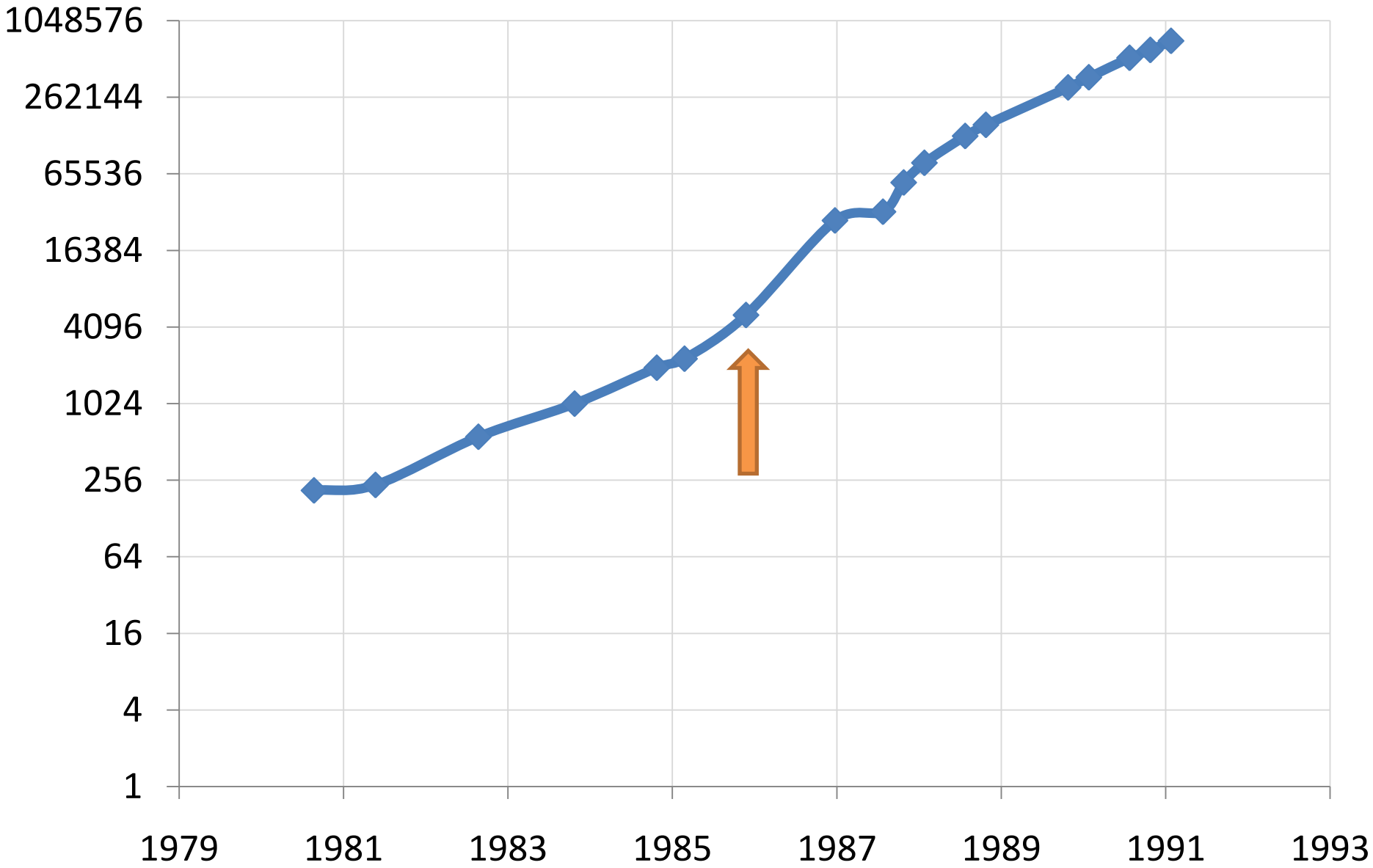
August 1983

BSD 4.2


Growth in Internet hosts 1981-1991
Data from RFC 1296
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
0
1979
1981
1983
1985
1987
1989
1991
1993

Growth in Internet hosts 1981-1991

Data from RFC 1296



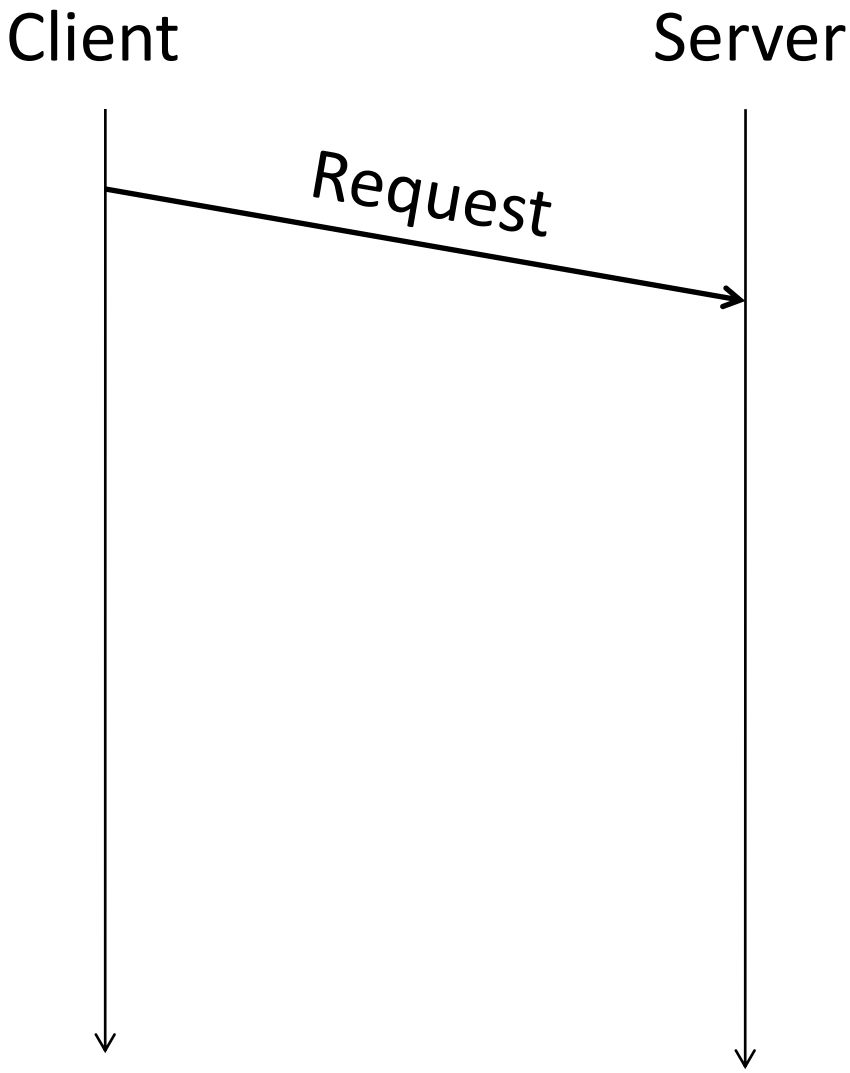
October 1986

Congestion collapse

Window size is allocated 16 bits in a TCP header, so maximum window size is 64kB

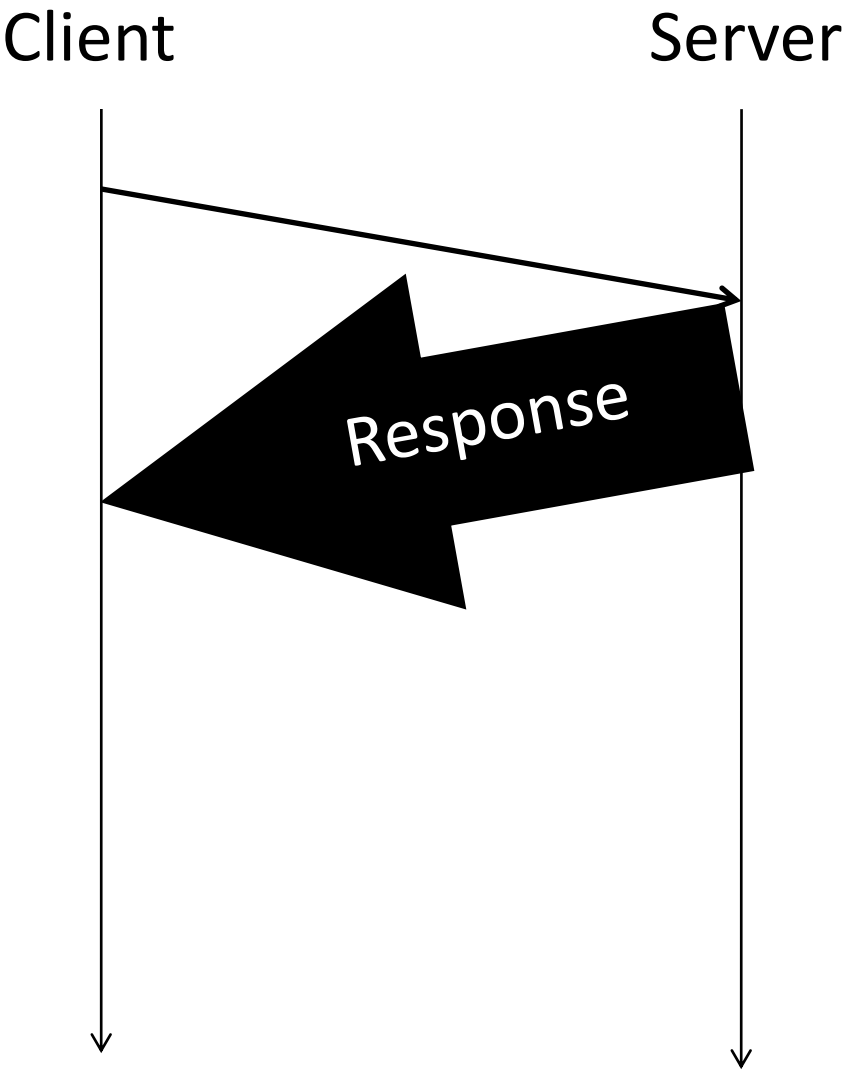
Window: 64kb

Max Segment: 1kb

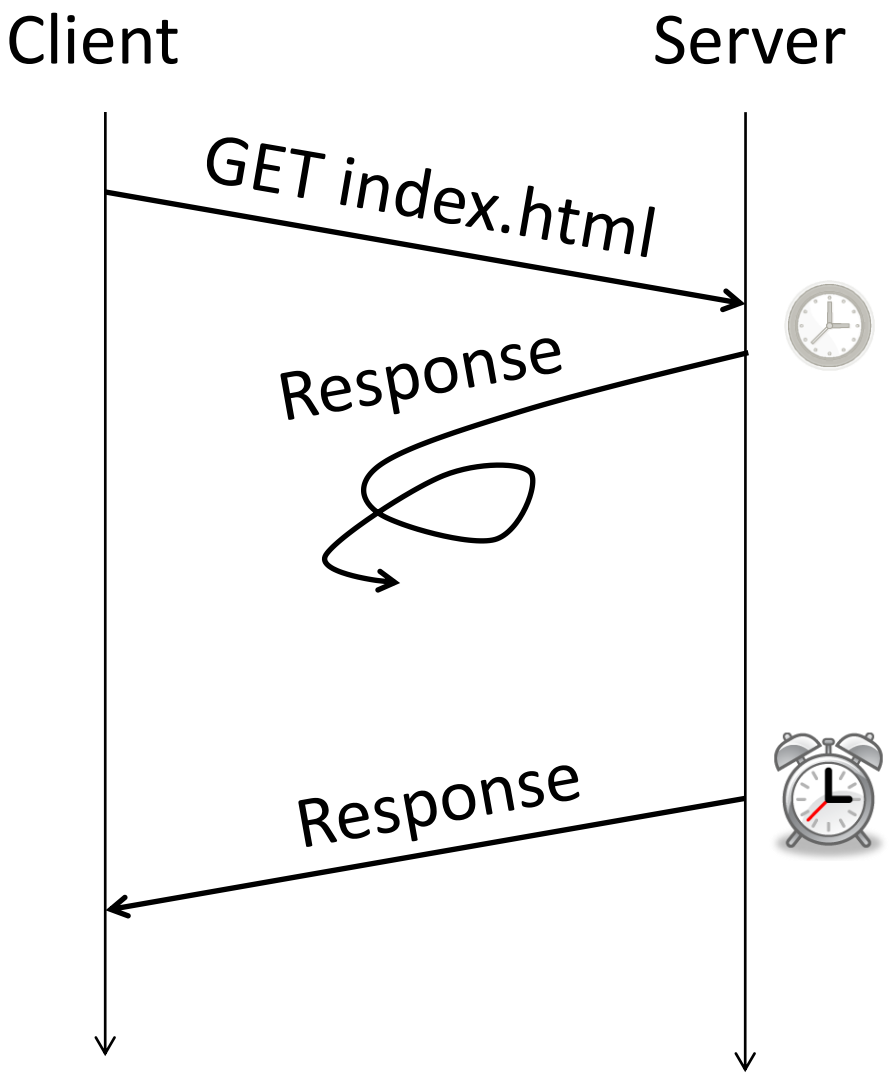


Window: 64kb

Max Segment: 1kb

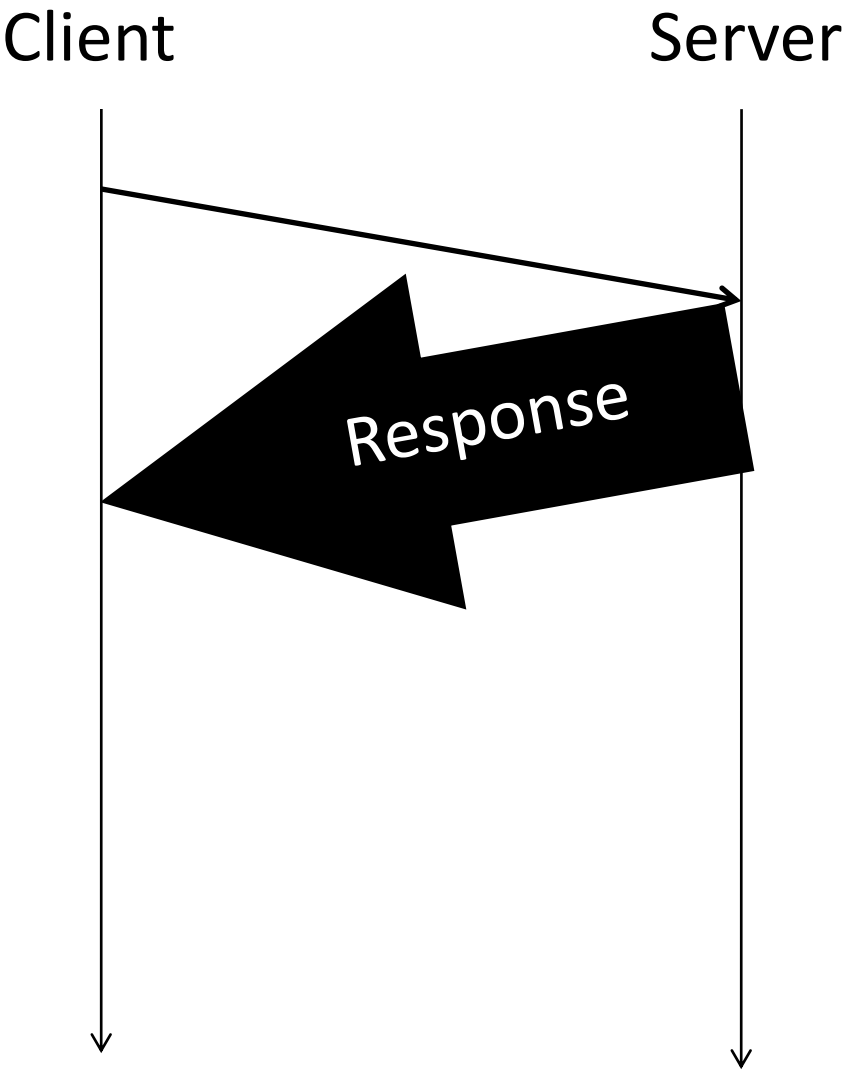



ung Hay
at Choy
PANDA EXPRESS
GOURMET CHINESE FOOD
Happy Chinese New Year
fpics.org


Client
Server
GET index.html
Response
Response

Window: 64kb

Max Segment: 1kb



January 1984

John Nagle - RFC 896

"…a sudden load on the net can cause the round-trip time to rise faster than the sending host's measurements of round-trip time can be updated."

-RFC 896

“Should the round-trip time exceed the maximum retransmission interval for any host, that host will begin to introduce more and more copies of the same datagrams into the net. The network is now in serious trouble.”

-RFC 896

"Should the round-trip time exceed the maximum retransmission interval for any host, that host will begin to introduce more and more copies of the same datagrams into the net. **The network is now in serious trouble.**"

-RFC 896

"Eventually all available buffers in the switching nodes will be full and packets must be dropped. Hosts are sending each packet several times, and eventually some copy of each packet arrives at its destination. This is congestion collapse."

-RFC 896

"Eventually all available buffers in the switching nodes will be full and packets must be dropped. Hosts are sending each packet several times, and **eventually some copy of each packet arrives at its destination.** This is congestion collapse."

-RFC 896

“This condition is stable. Once the saturation point has been reached, if the algorithm for selecting packets to be dropped is fair, the network will continue to operate in a degraded condition.”

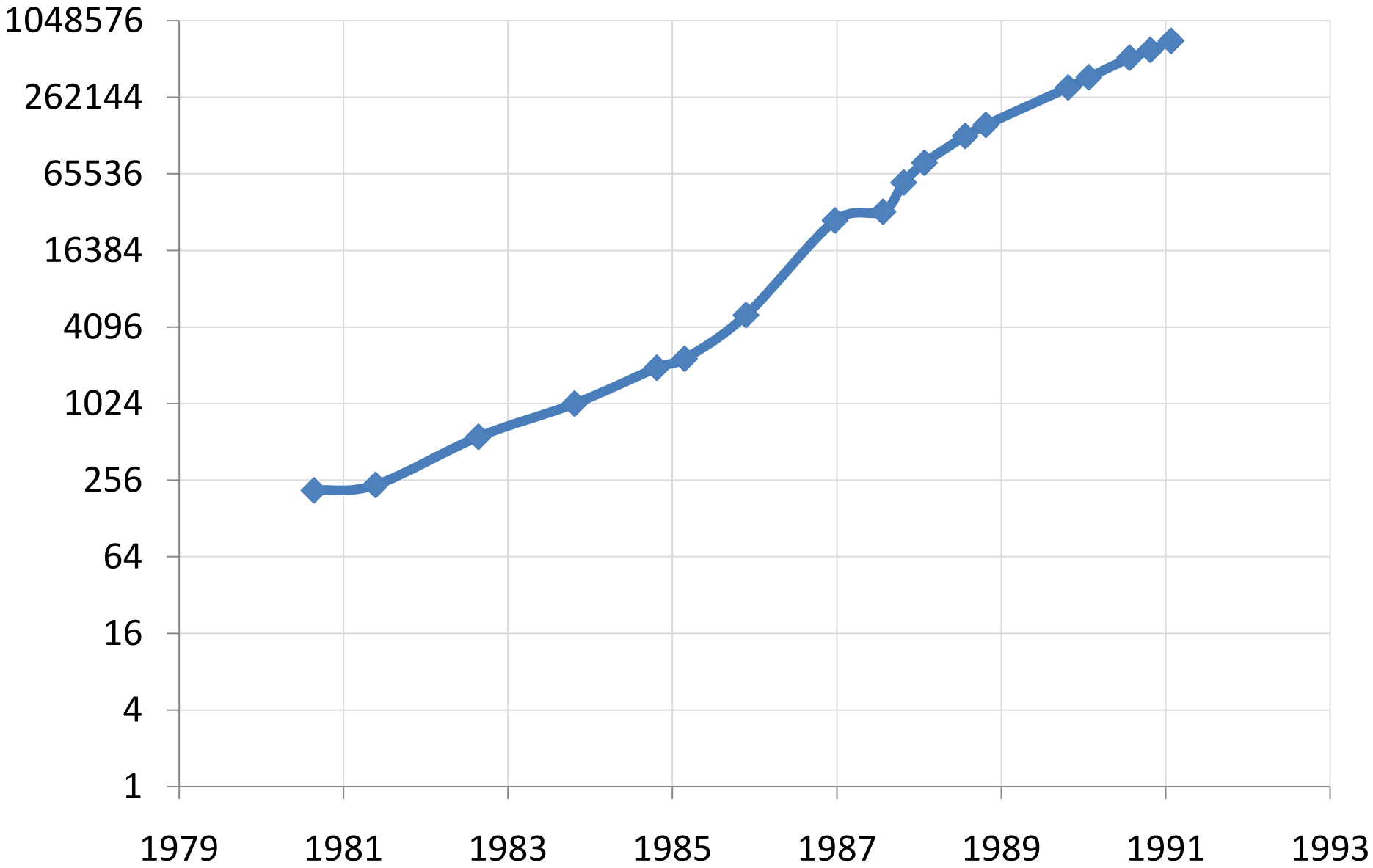
-RFC 896

“Congestion collapse and pathological congestion are not normally seen in the ARPANET / MILNET system because these networks have substantial excess capacity.”

-RFC 896

Growth in Internet hosts 1981-1991

Data from RFC 1296



“The critical congestion problems the ARPANET is experiencing causes TELNET and FTP connections to time out and mail messages from MILNET hosts to take up to 2-3 days to be delivered to BBNNET hosts.”

- Nancy Cassidy in mod.risks, September 22 1986

Figure 7: Multiple conversation test setup

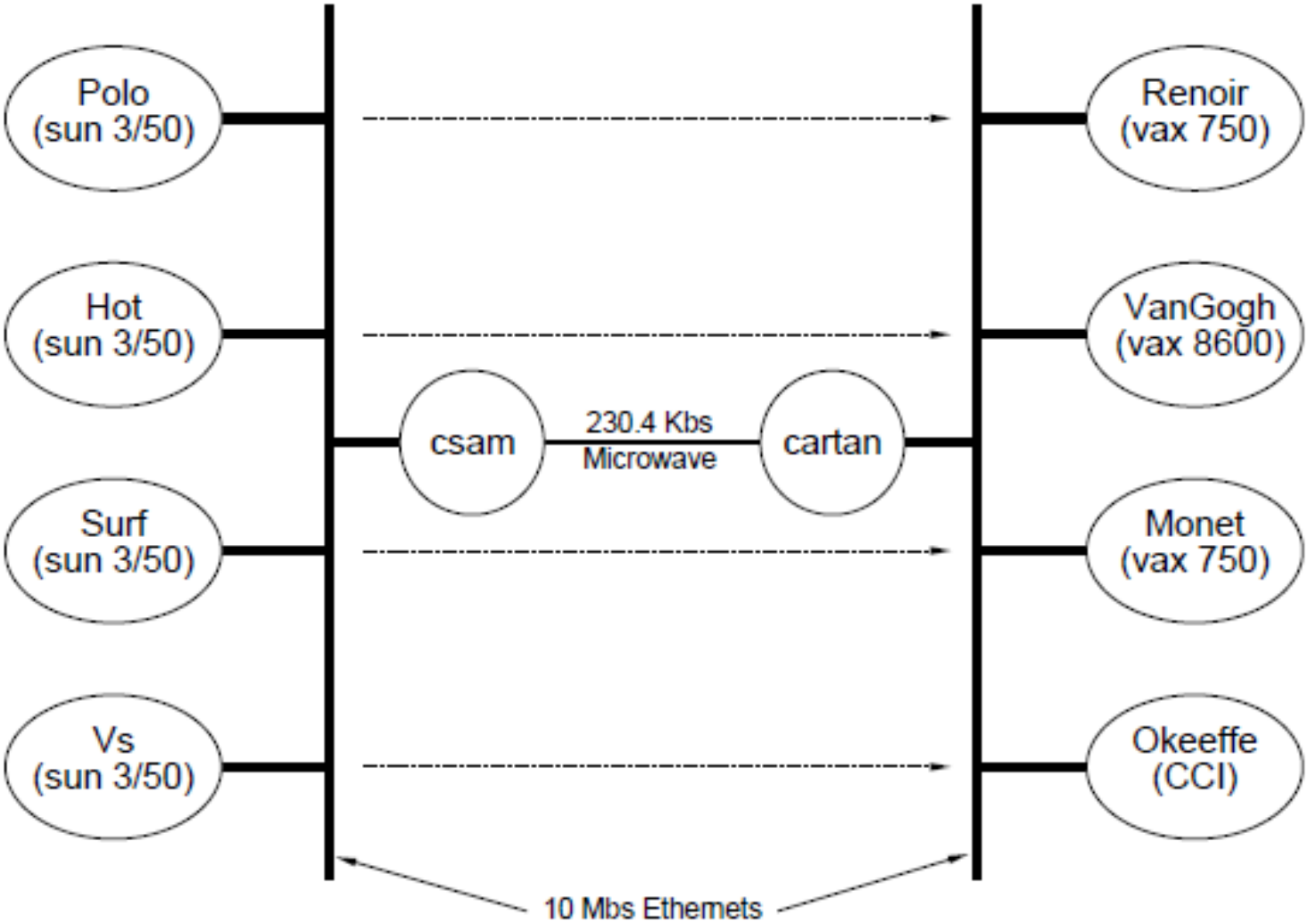


Figure 7: Multiple conversation test setup

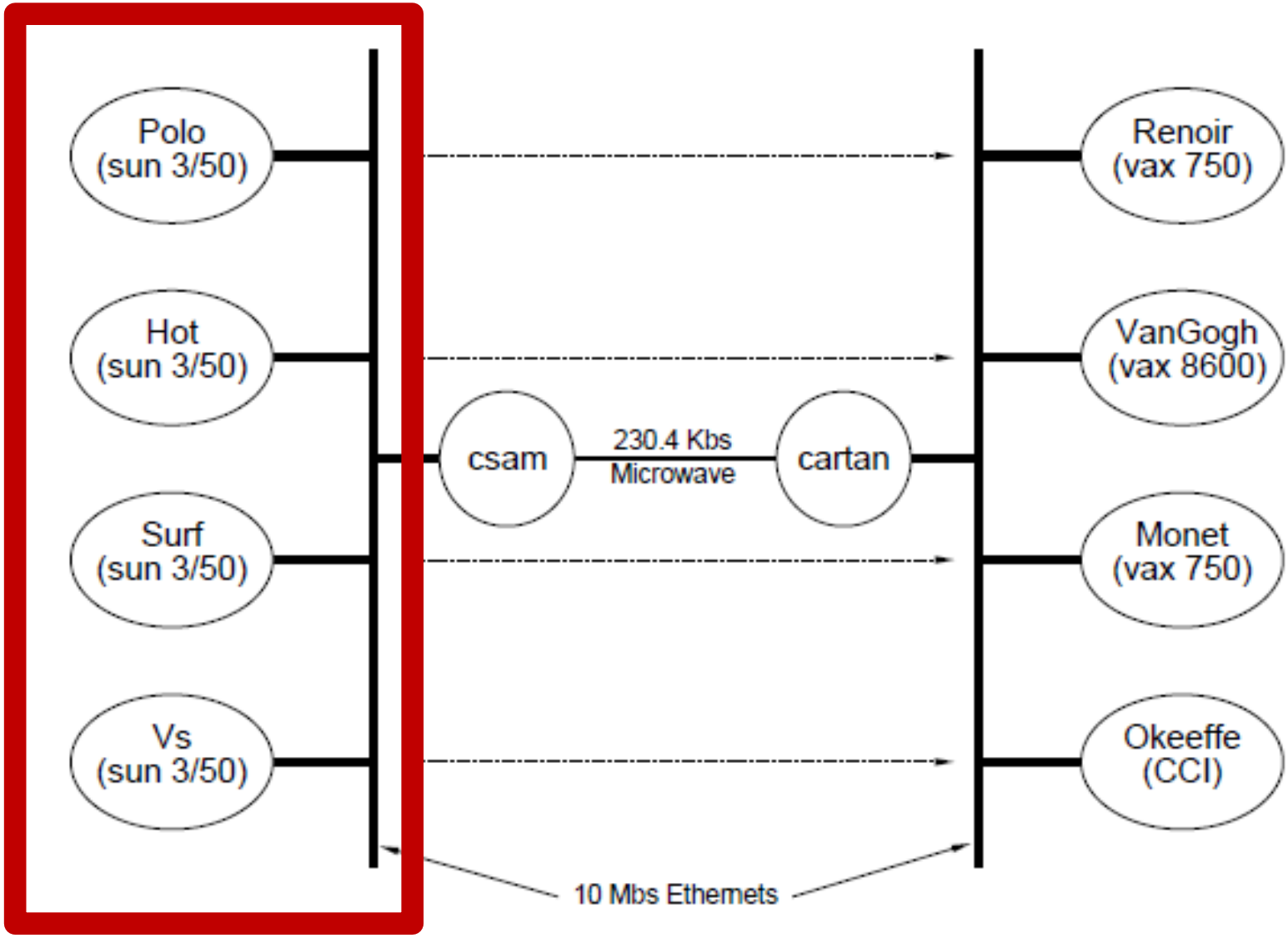


Figure 7: Multiple conversation test setup

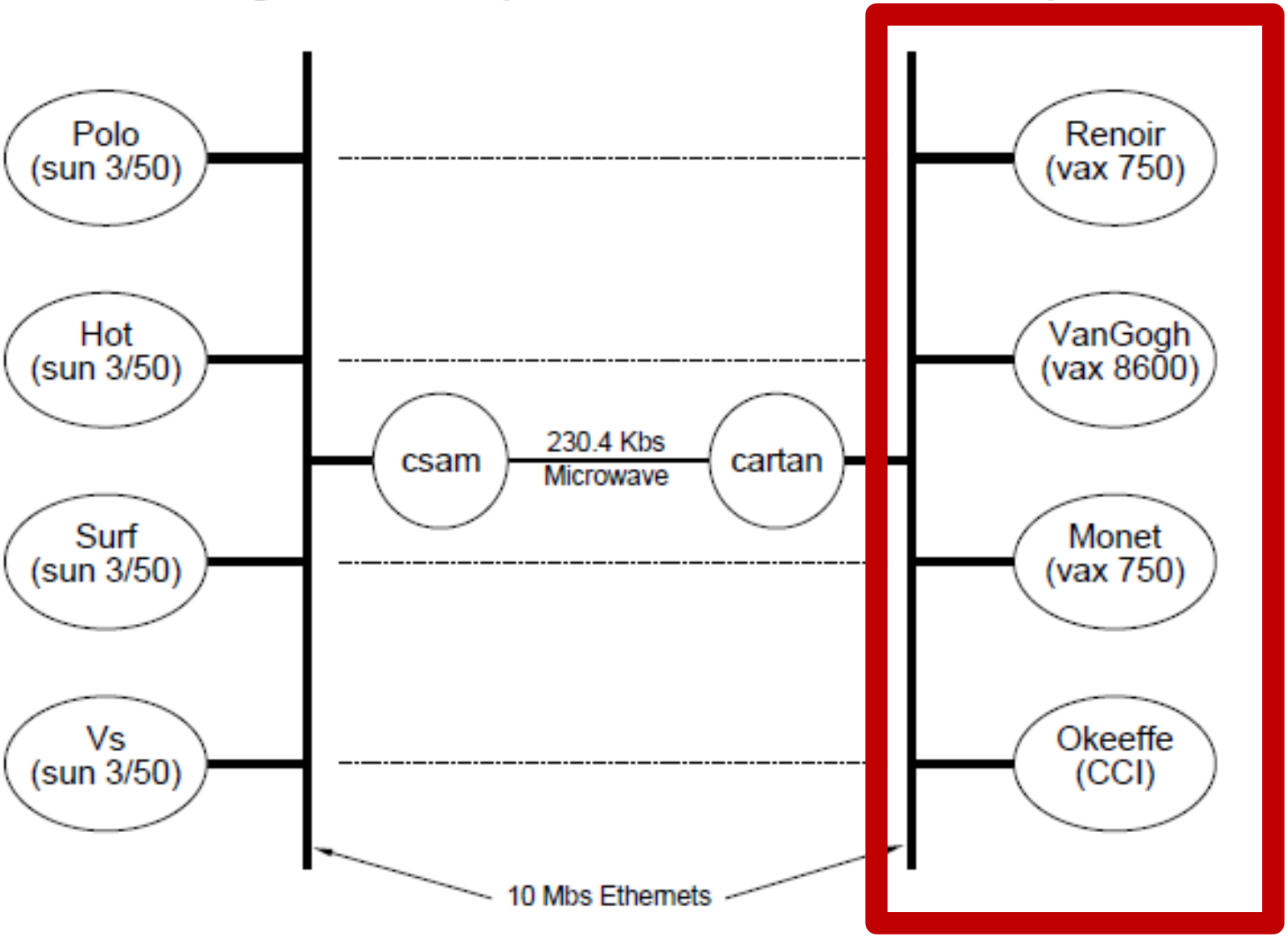


Figure 7: Multiple conversation test setup

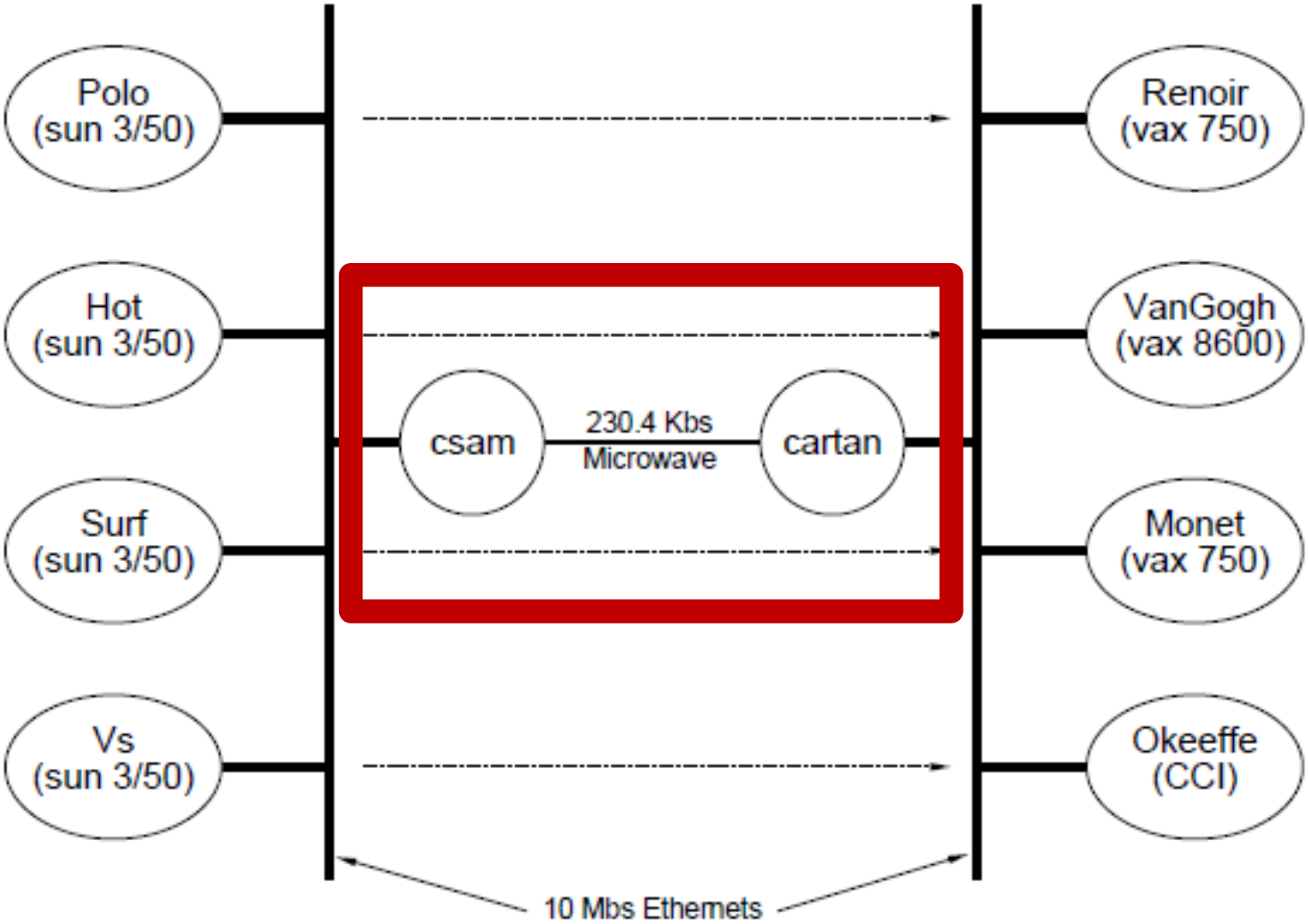


Figure 3: Startup behavior of TCP without Slow-start

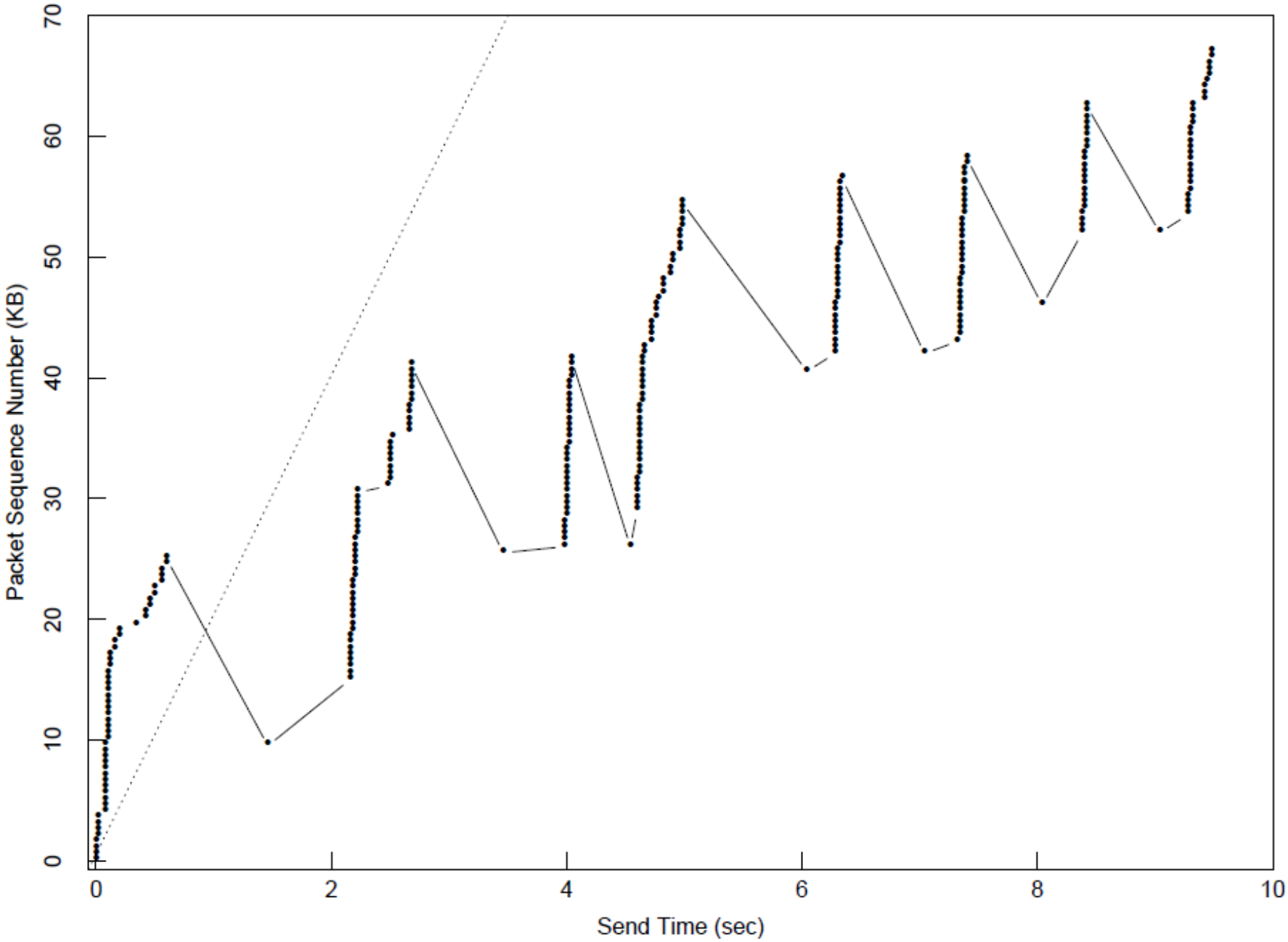


Figure 3: Startup behavior of TCP without Slow-start

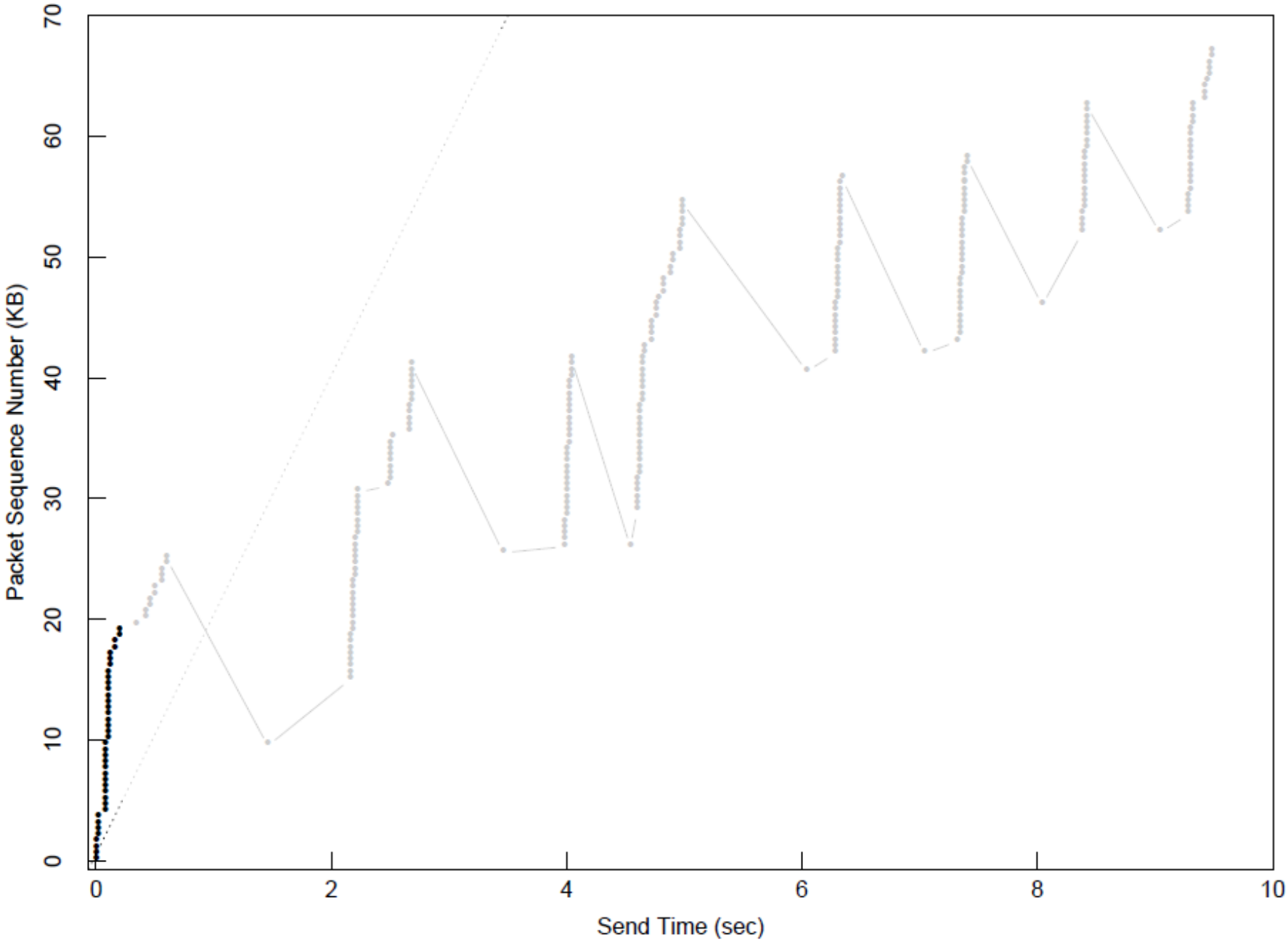


Figure 3: Startup behavior of TCP without Slow-start

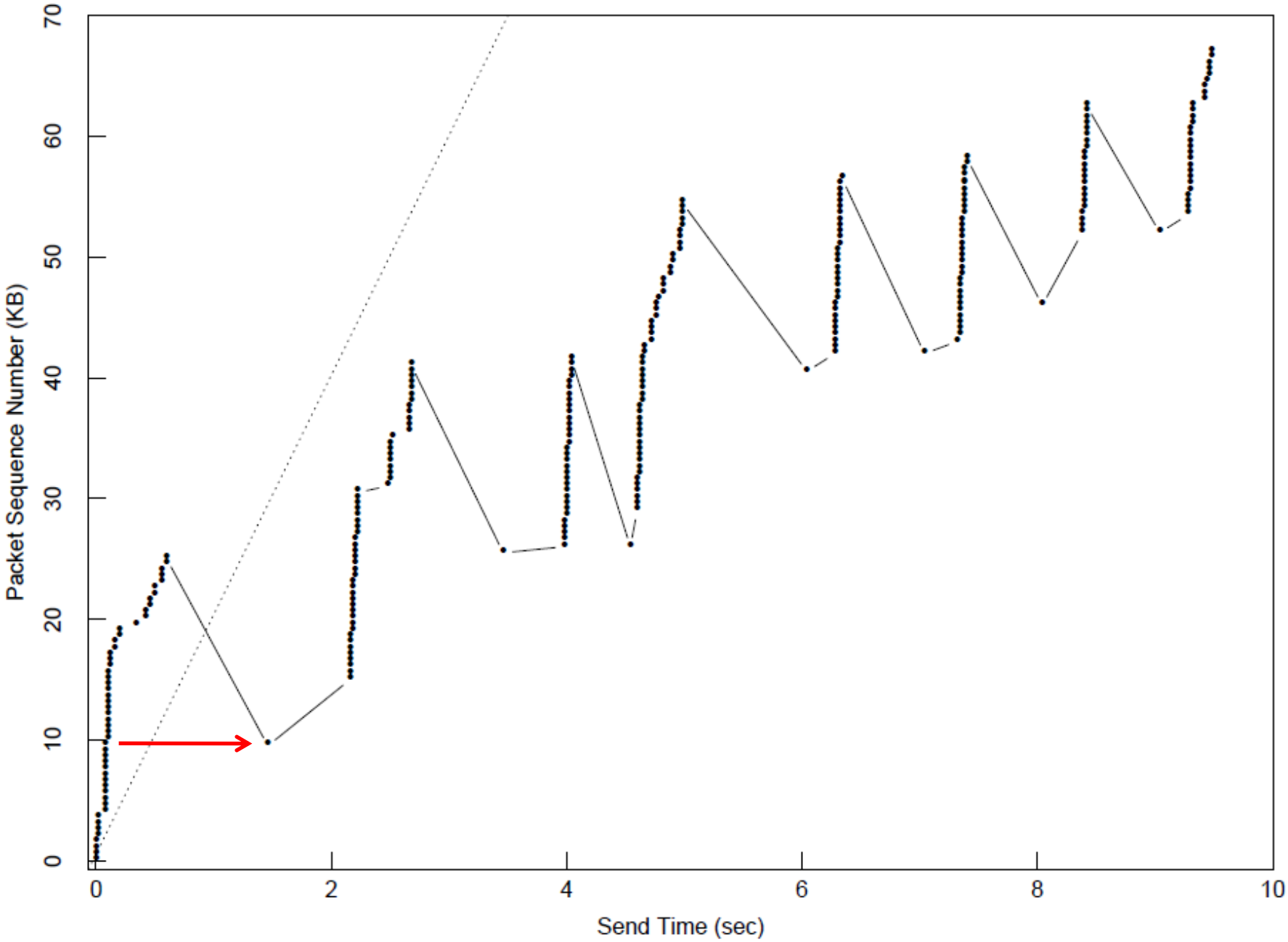


Figure 3: Startup behavior of TCP without Slow-start

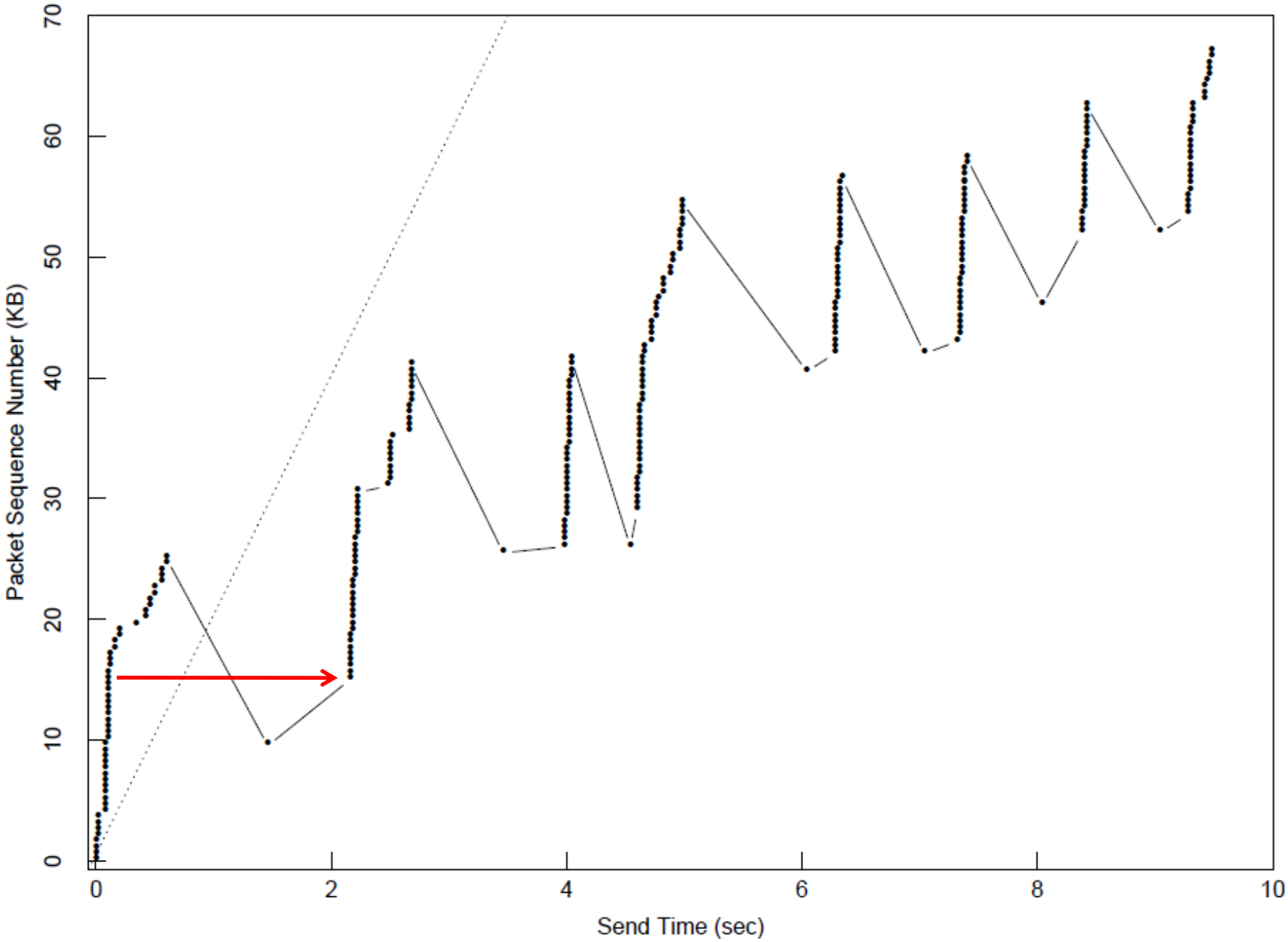


Figure 3: Startup behavior of TCP without Slow-start

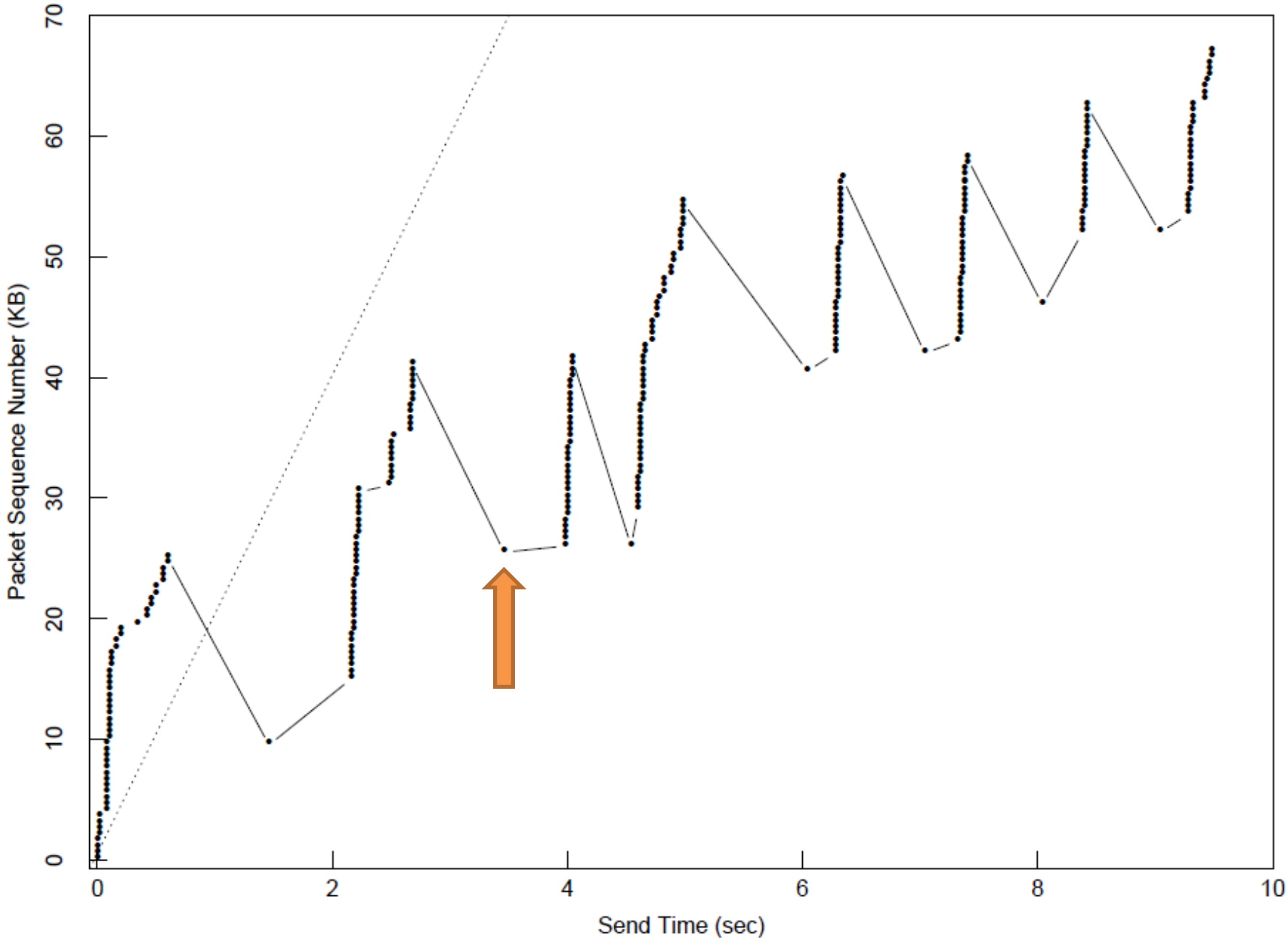


Figure 3: Startup behavior of TCP without Slow-start

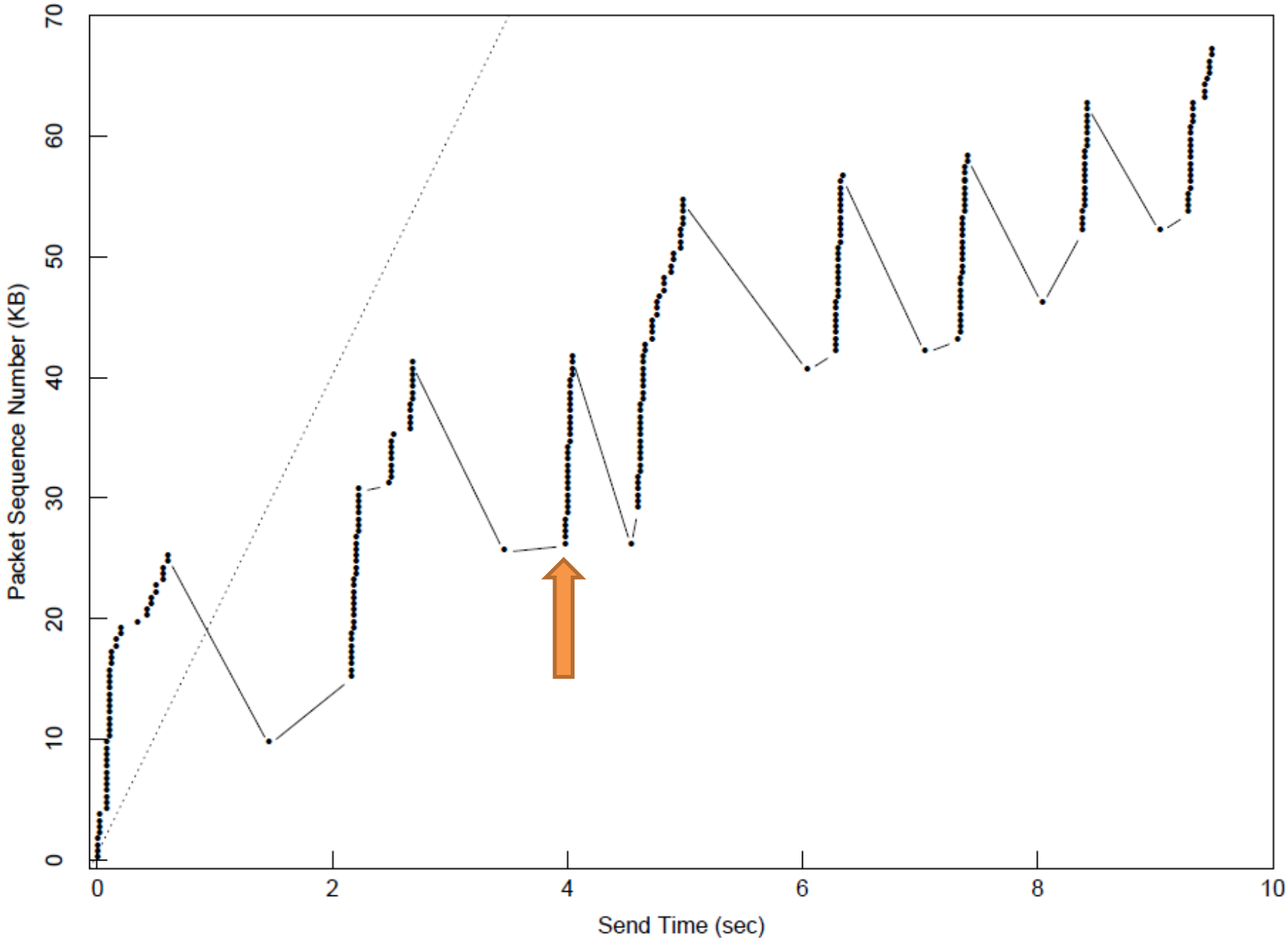


Figure 3: Startup behavior of TCP without Slow-start

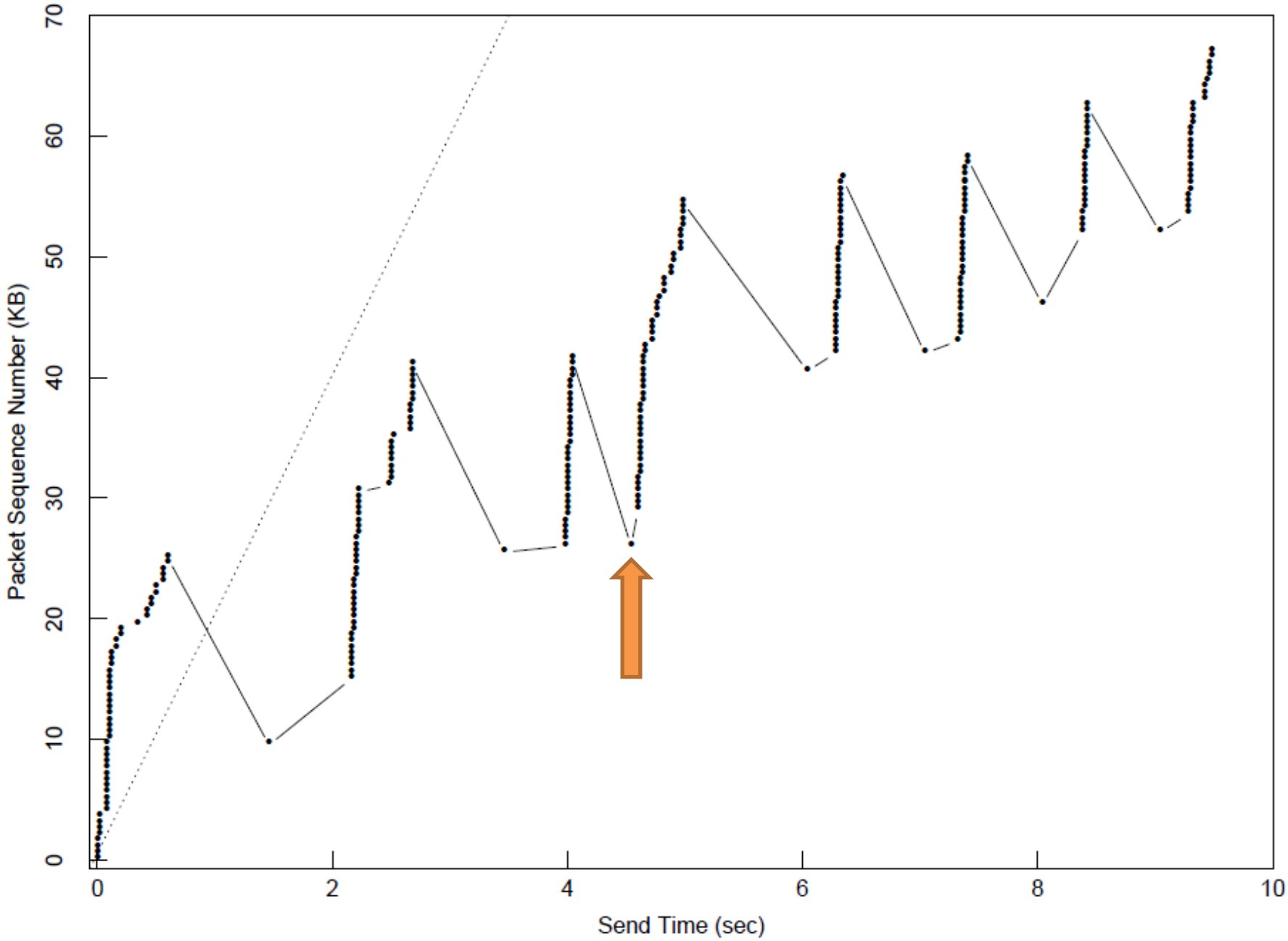


Figure 3: Startup behavior of TCP without Slow-start

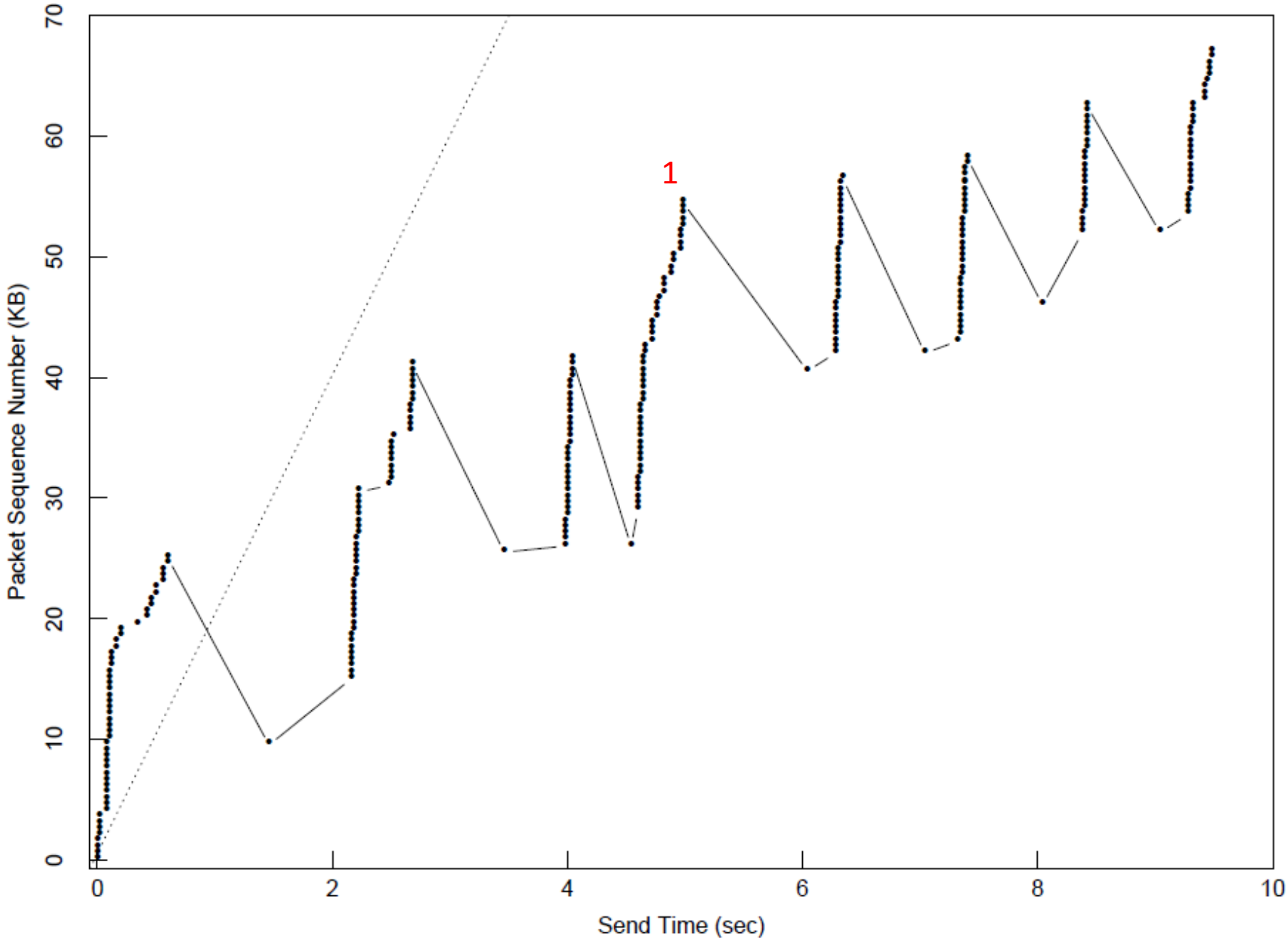


Figure 3: Startup behavior of TCP without Slow-start

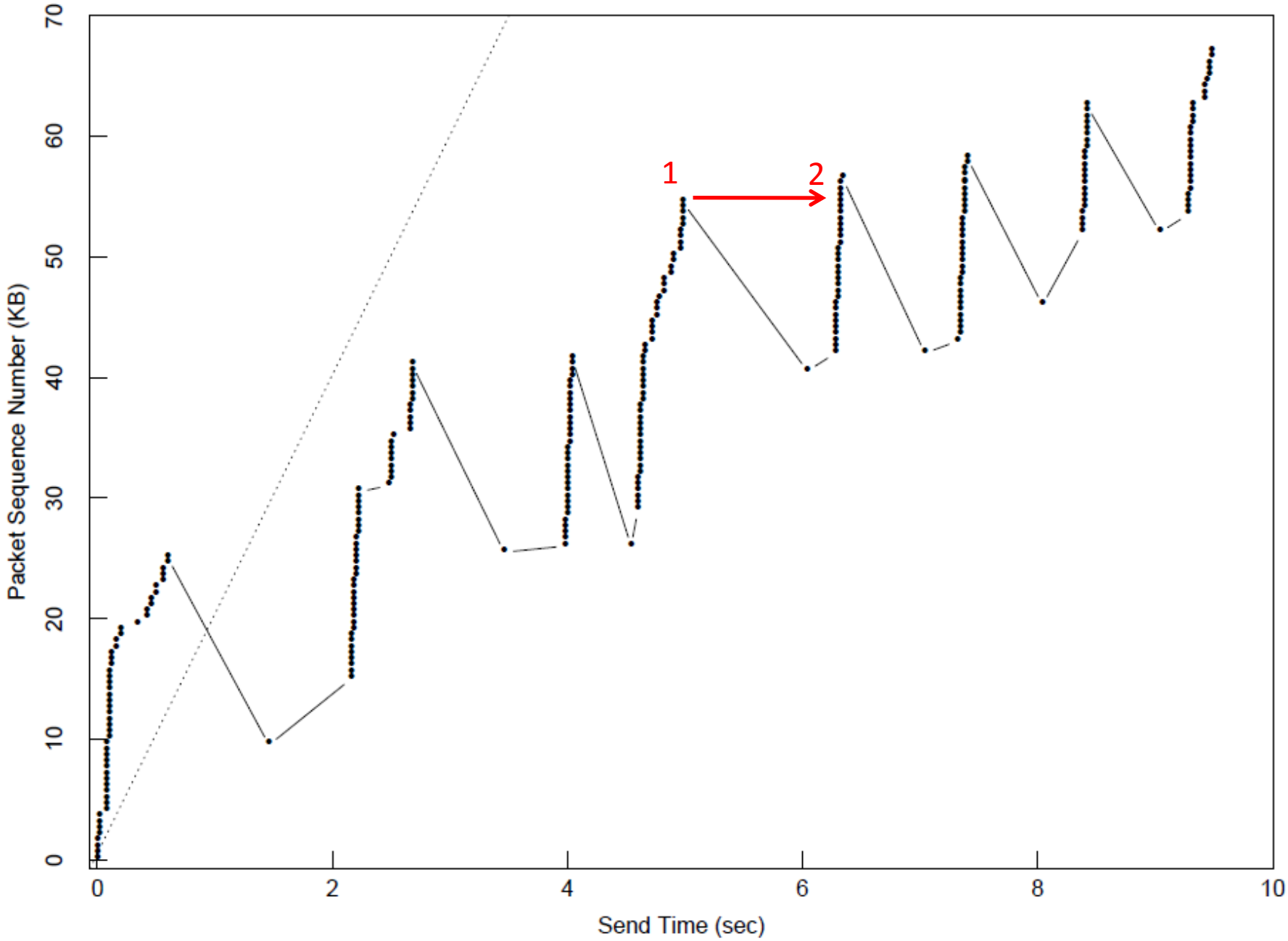


Figure 3: Startup behavior of TCP without Slow-start

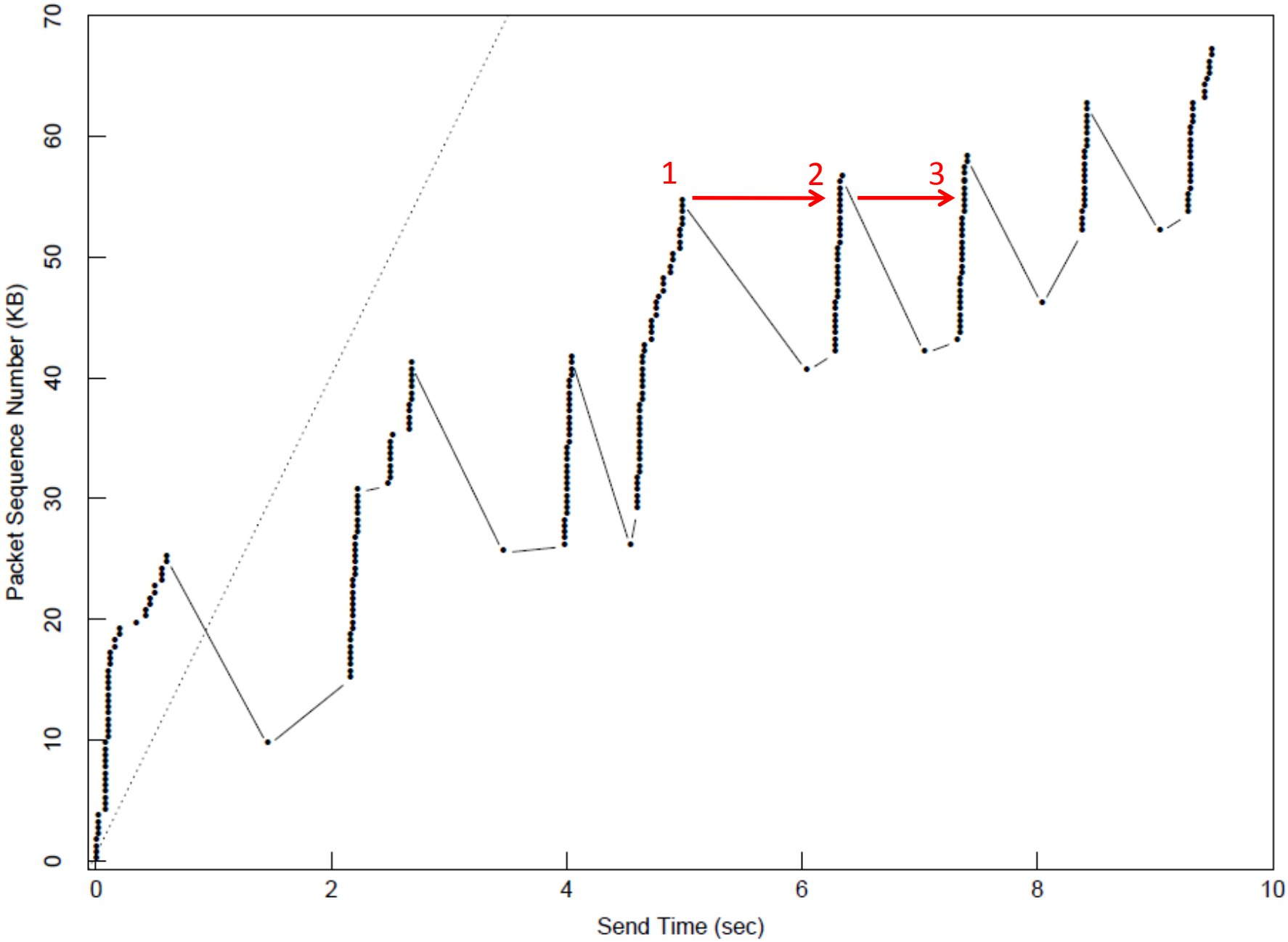


Figure 3: Startup behavior of TCP without Slow-start

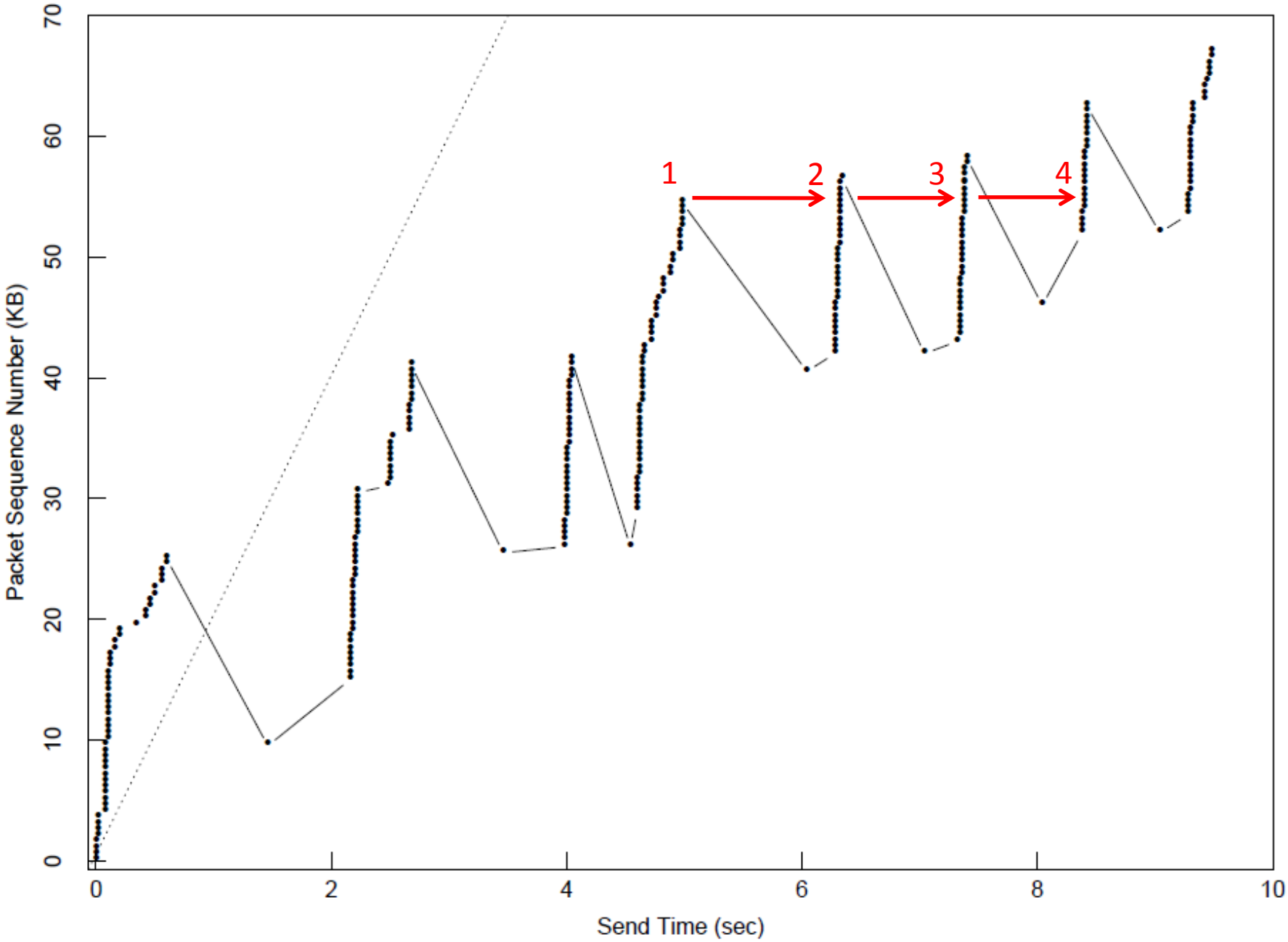
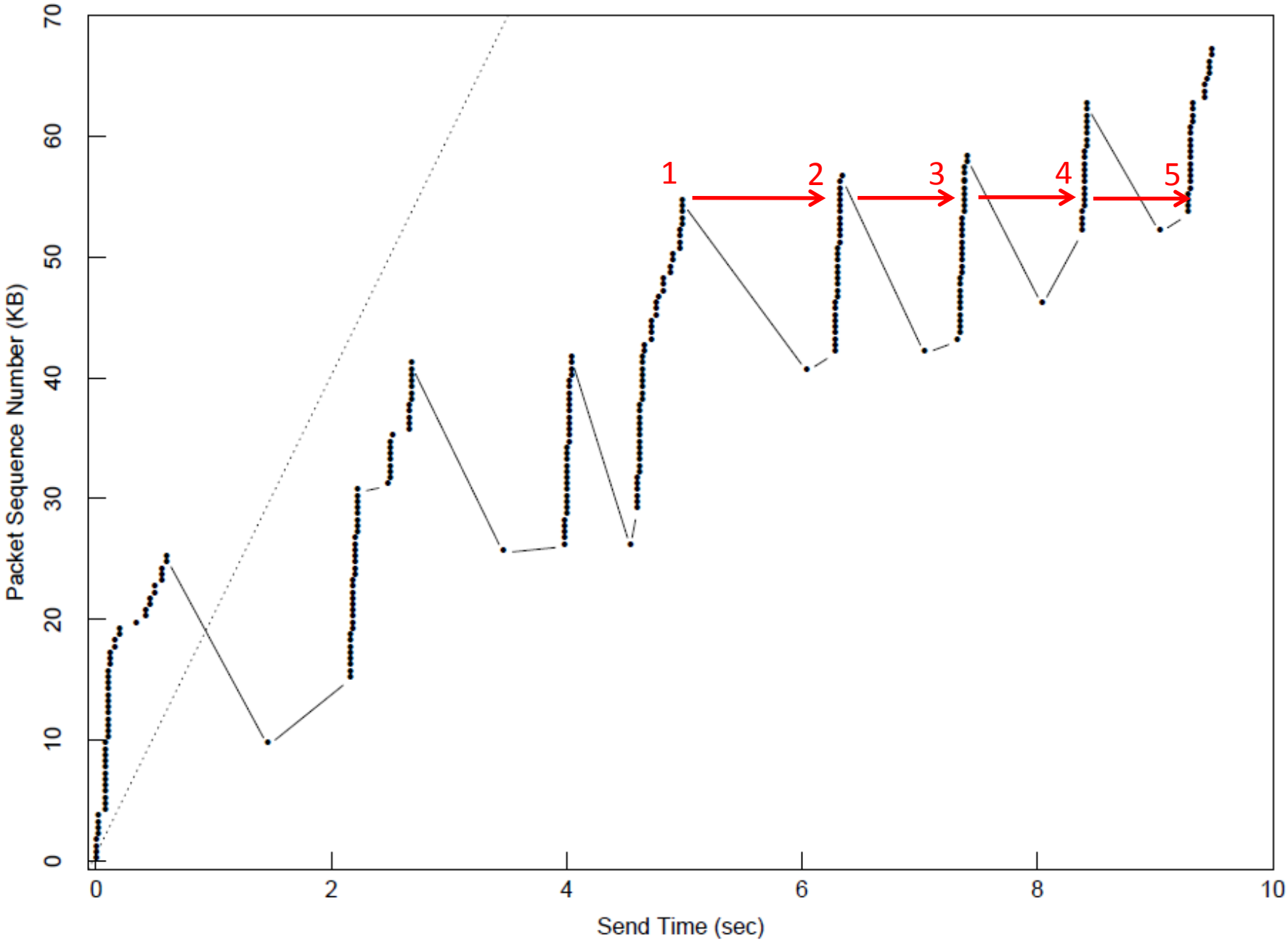
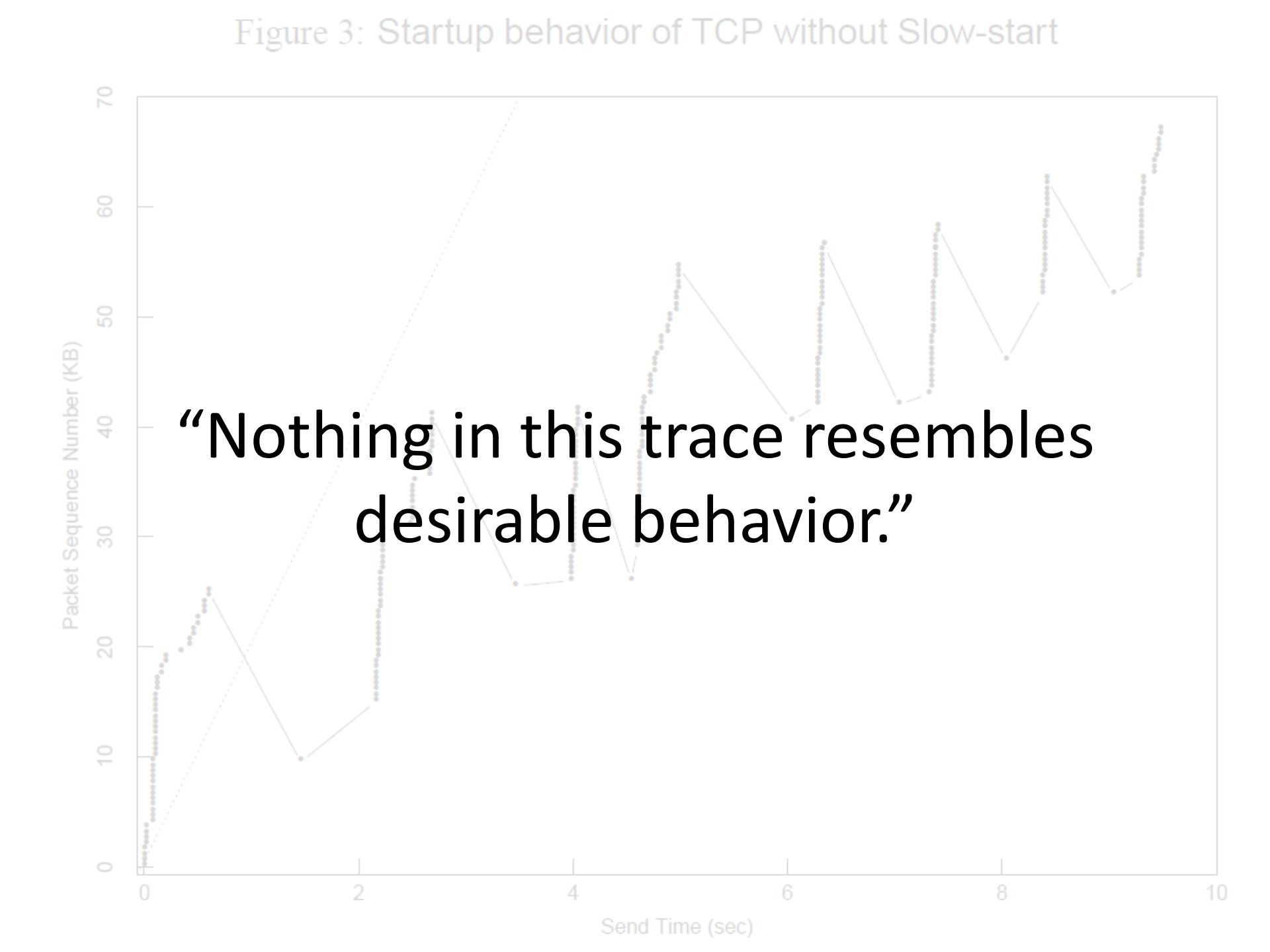


Figure 3: Startup behavior of TCP without Slow-start





"Nothing in this trace resembles desirable behavior."

TCP Slow Start

window : receiver

window : receiver

+

congestion window : sender

window =
the maximum amount of un-ACKed data in flight.

min(window, cwnd) =
the maximum amount of un-ACKed data in flight.

Congestion window:

Slow start
Congestion avoidance
Fast retransmit
Fast recovery

Tahoe
Reno
Vegas
New Reno
Westwood
BIC/CUBIC (Linux 2.6.19)
Compound TCP (Vista)

RFC 2851 -

TCP Congestion Control

RFC 3390 -

Increasing TCP's Initial Window

Congestion window:

Slow start

Congestion avoidance

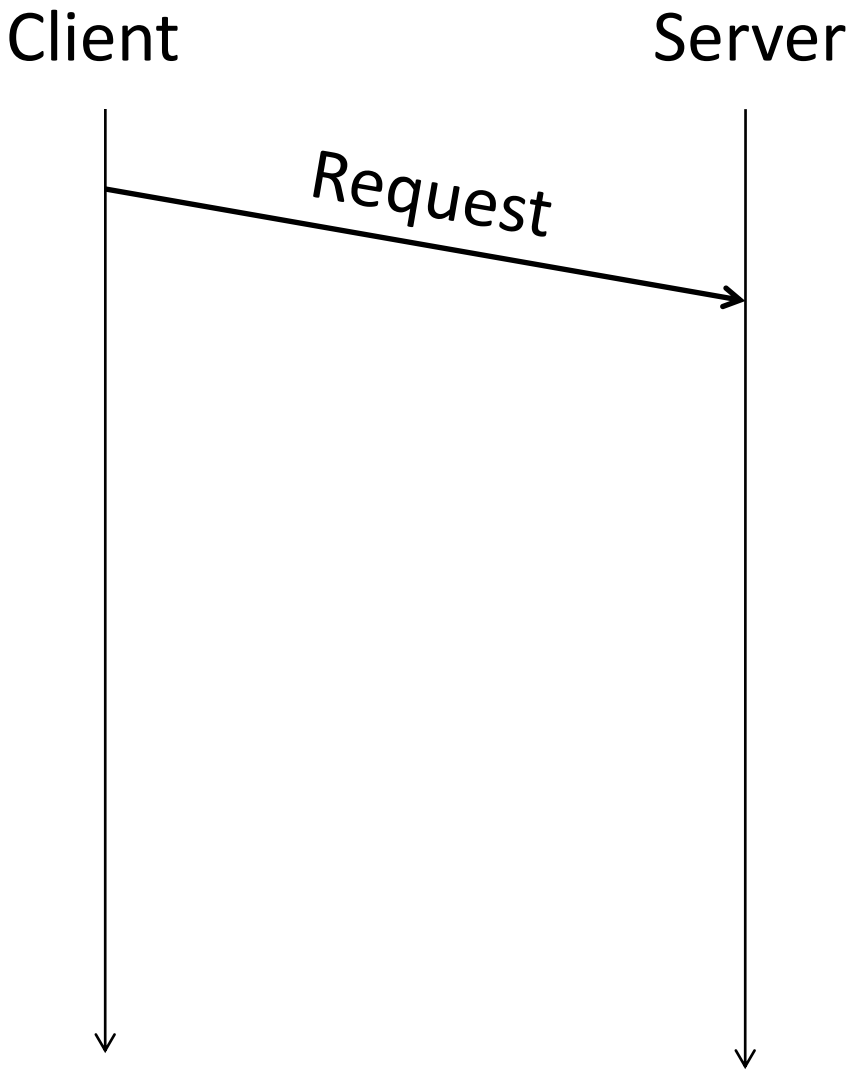
Fast retransmit

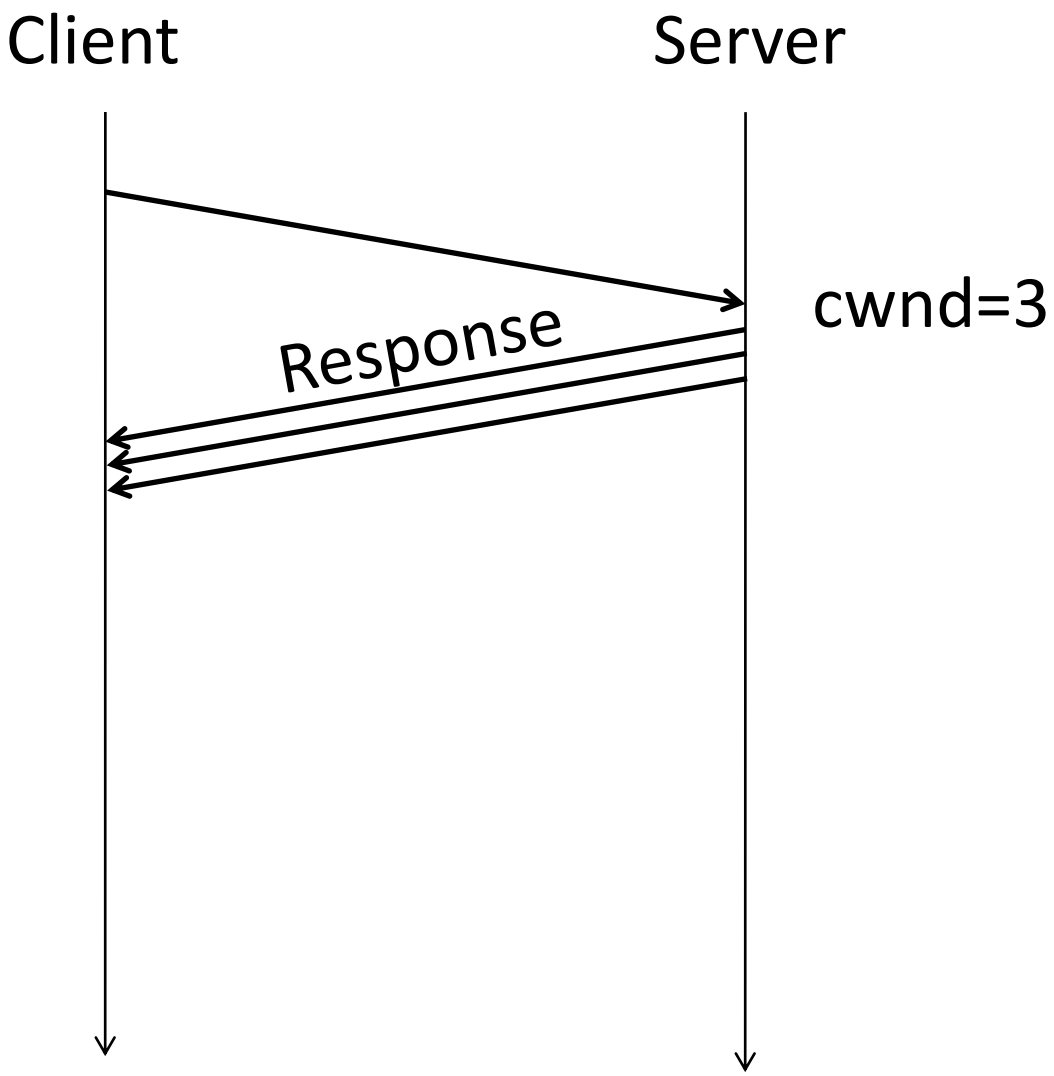
Fast recovery

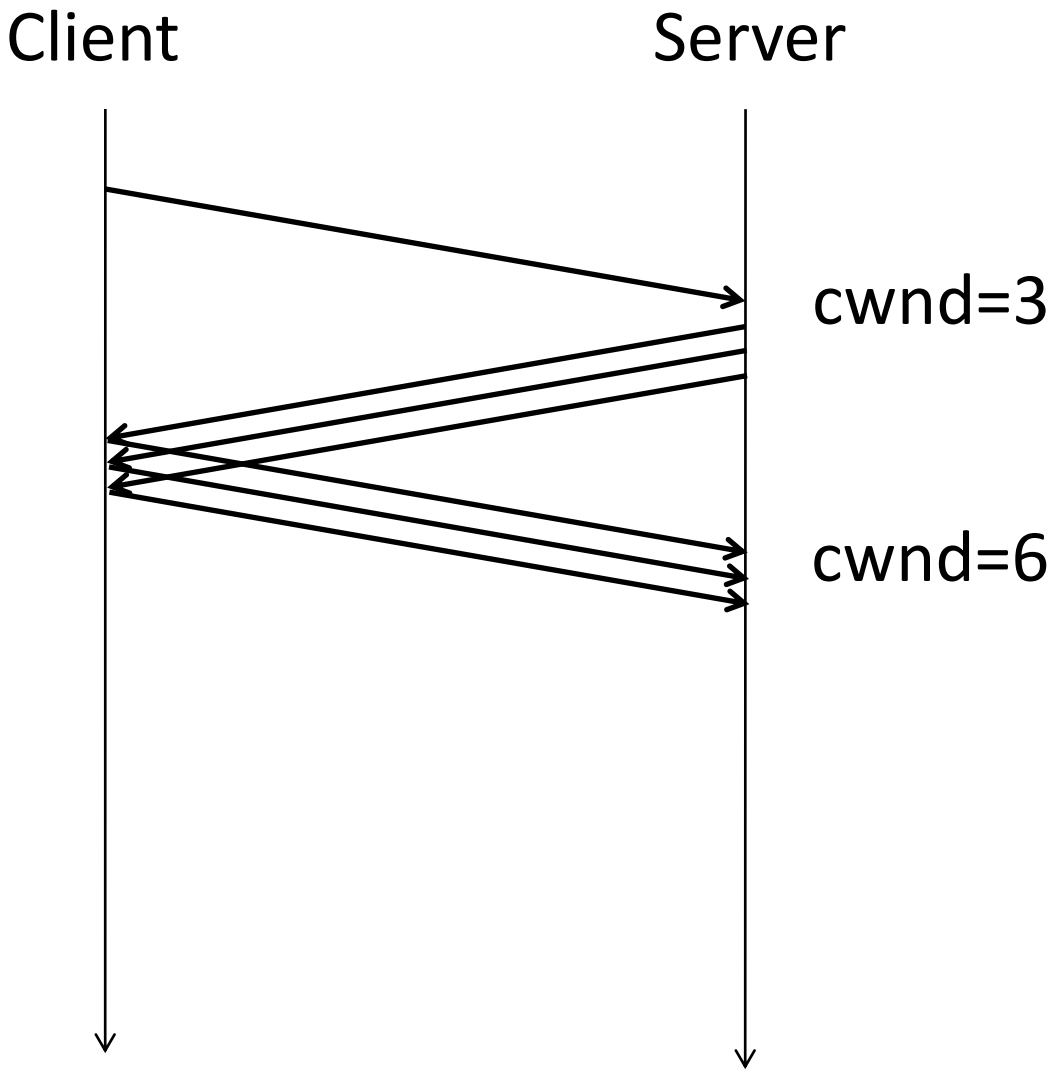
Slow start:

1) Initialize cwnd to three full segments

2) Increment cwnd by one full segment for each ACK


Client
Server
Request


Client
Server
cwnd=3
Response


Client
Server
cwnd=3
cwnd=6

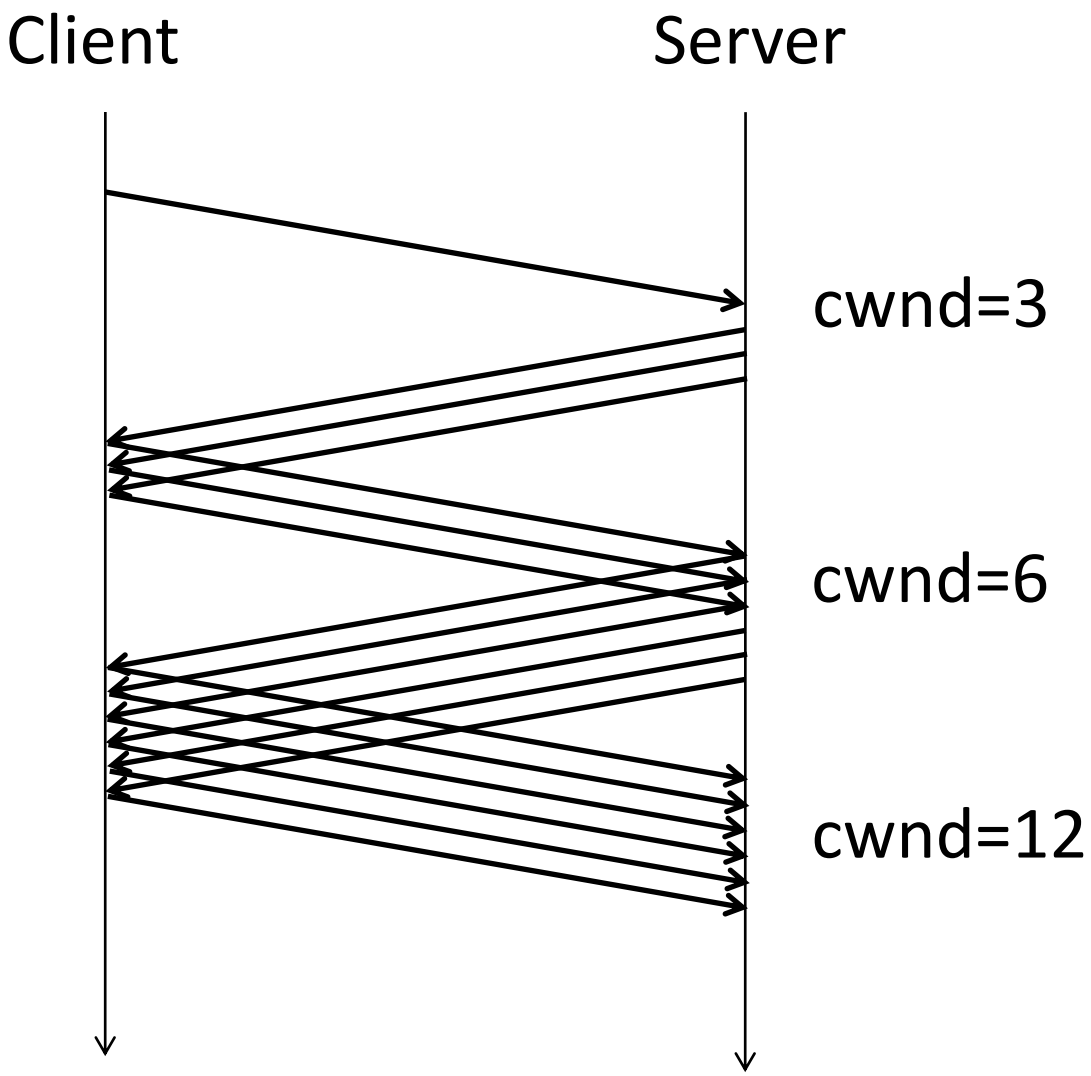

Client
Server
cwnd=3
cwnd=6
cwnd=12

Figure 3: Startup behavior of TCP without Slow-start

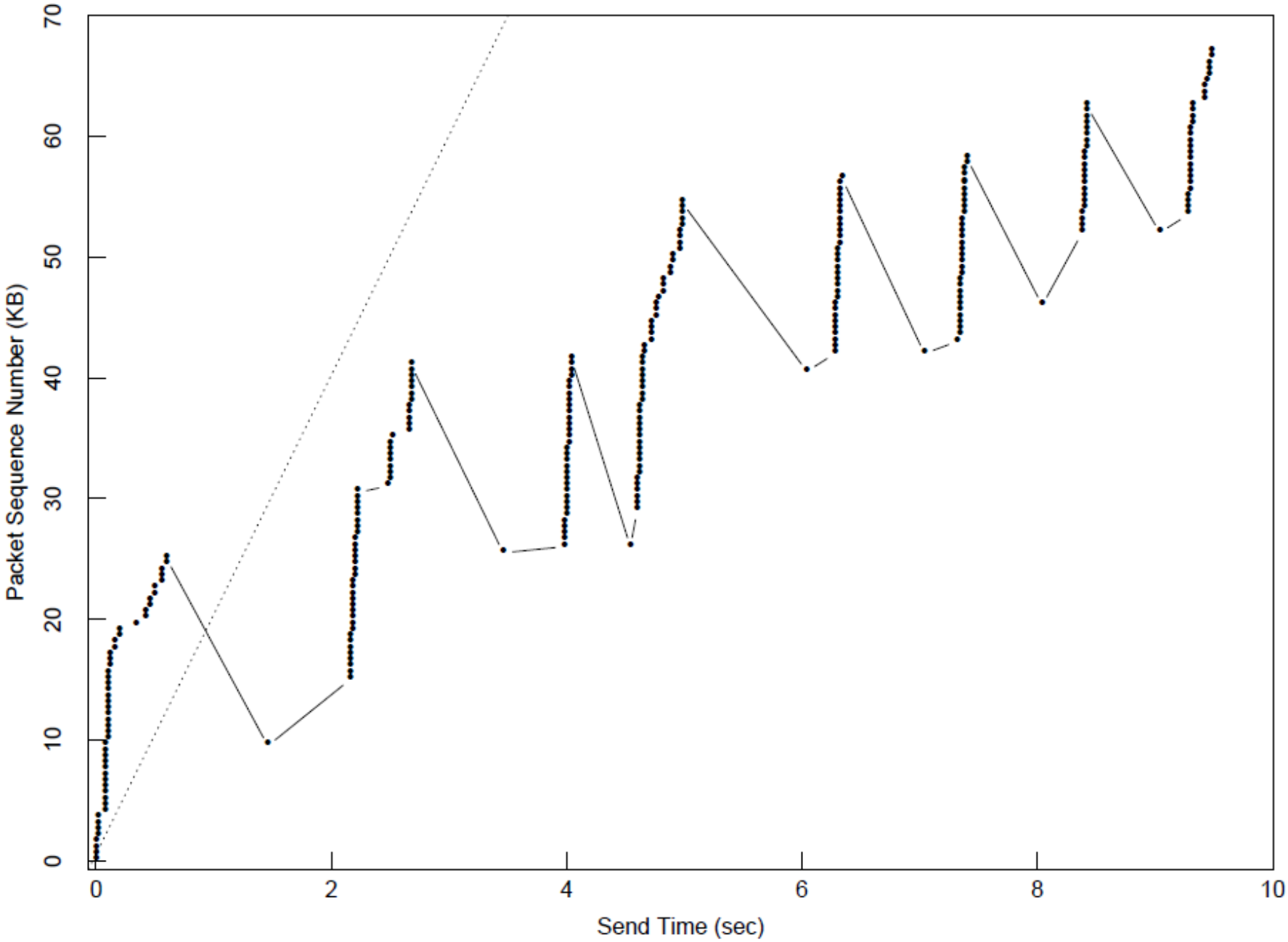
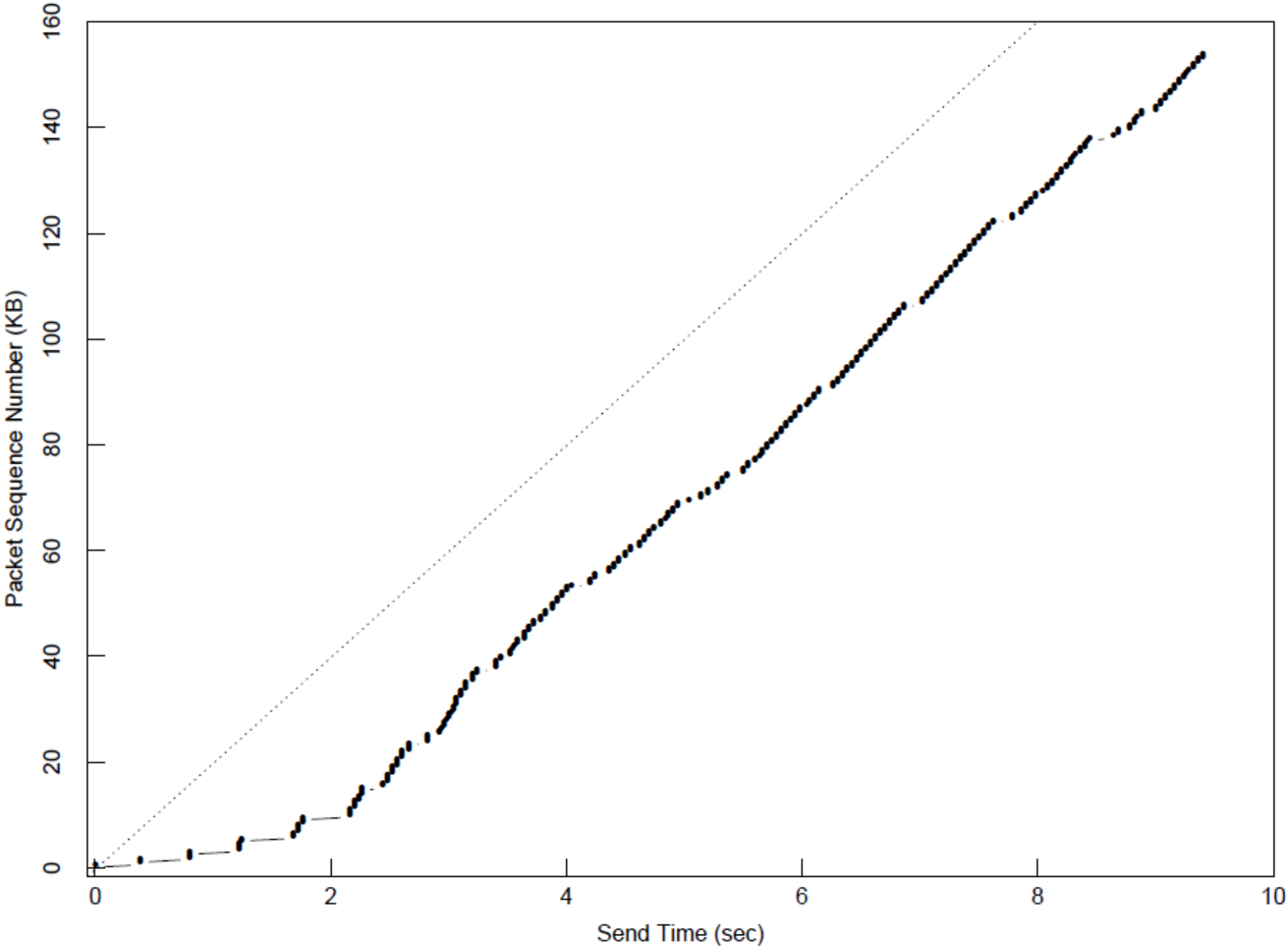


Figure 4: Startup behavior of TCP with Slow-start



June 1988

BSD4.3 Tahoe

November 1988

Congestion Avoidance and Control

Van Jacobson
Michael J Karels

http://ee.lbl.gov/papers/congavoid.pdf

November 1989

RFC 1122

"Recent work by Jacobson on Internet congestion and TCP retransmission stability has produced a transmission algorithm combining 'slow start' with 'congestion avoidance'. A TCP MUST implement this algorithm."

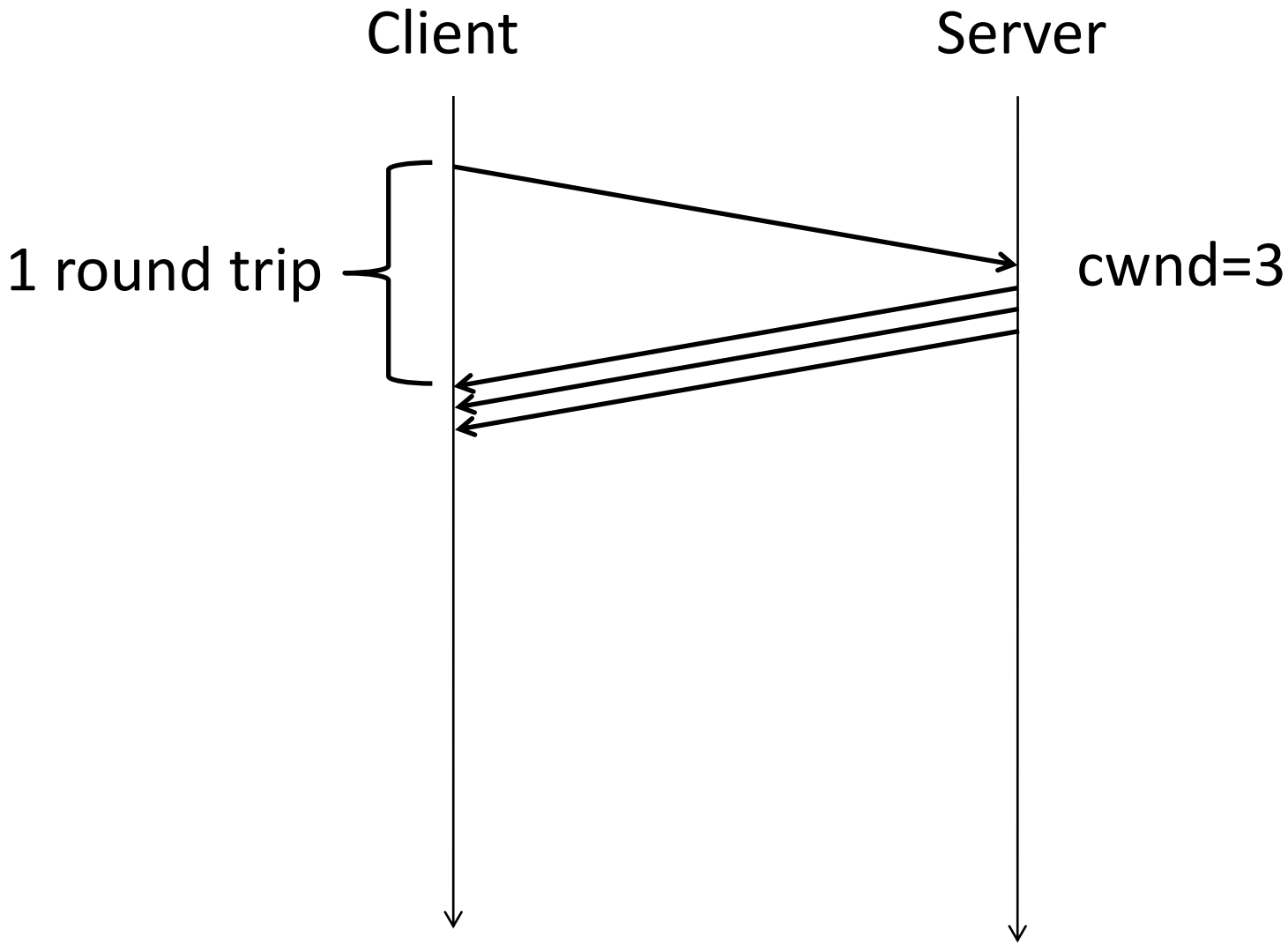
-RFC 1122

"Recent work by Jacobson on Internet congestion and TCP retransmission stability has produced a transmission algorithm combining 'slow start' with 'congestion avoidance'. **A TCP MUST implement this algorithm."**

-RFC 1122

TCP slow start

network latency strictly limits the throughput of new connections


Client
Server
1 round trip
cwnd=3

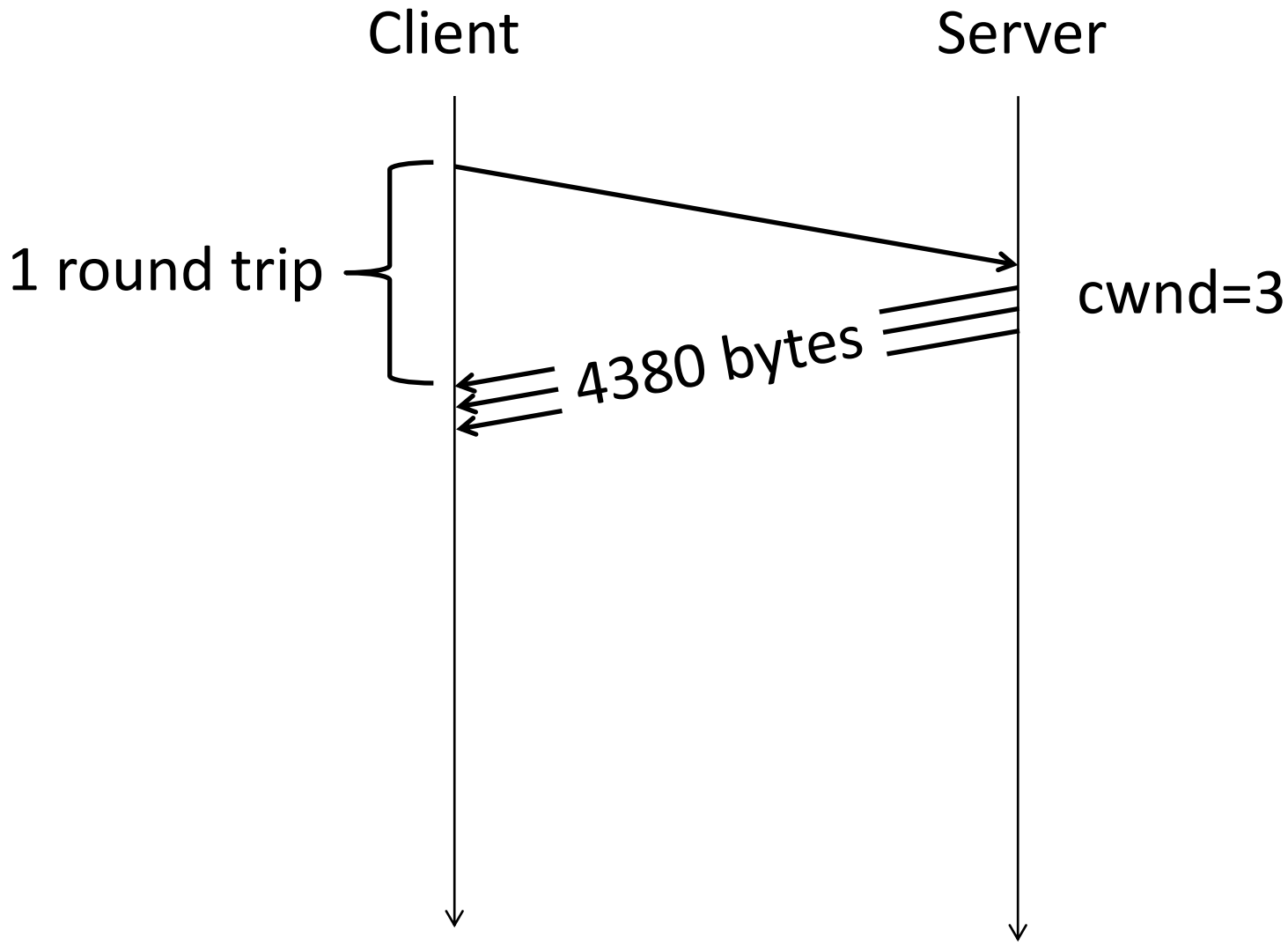
No problem!

Send bigger packets

“… the maximum length of an IP datagram sent over an Ethernet is 1500 octets”

-RFC 894

Assuming 1460 byte segments



No problem!

It’s only on the first hit

Yahoo 2007:

One hit in five is uncached

http://www.yuiblog.com/blog/2007/01/04/performance-research-part-2/

If average session length is N,

(and you assume equal probability of departure on each hit)

then 1 hit in N is a first hit

No problem!

It’s only five round trips until the window is fully open

The window field in the TCP header is 16 bits

$$2^{16} = \frac{65536}{1460} = 44 \text{ segments}$$

Round Trip	Congestion window size
1	3
2	6
3	12
4	24
5	44

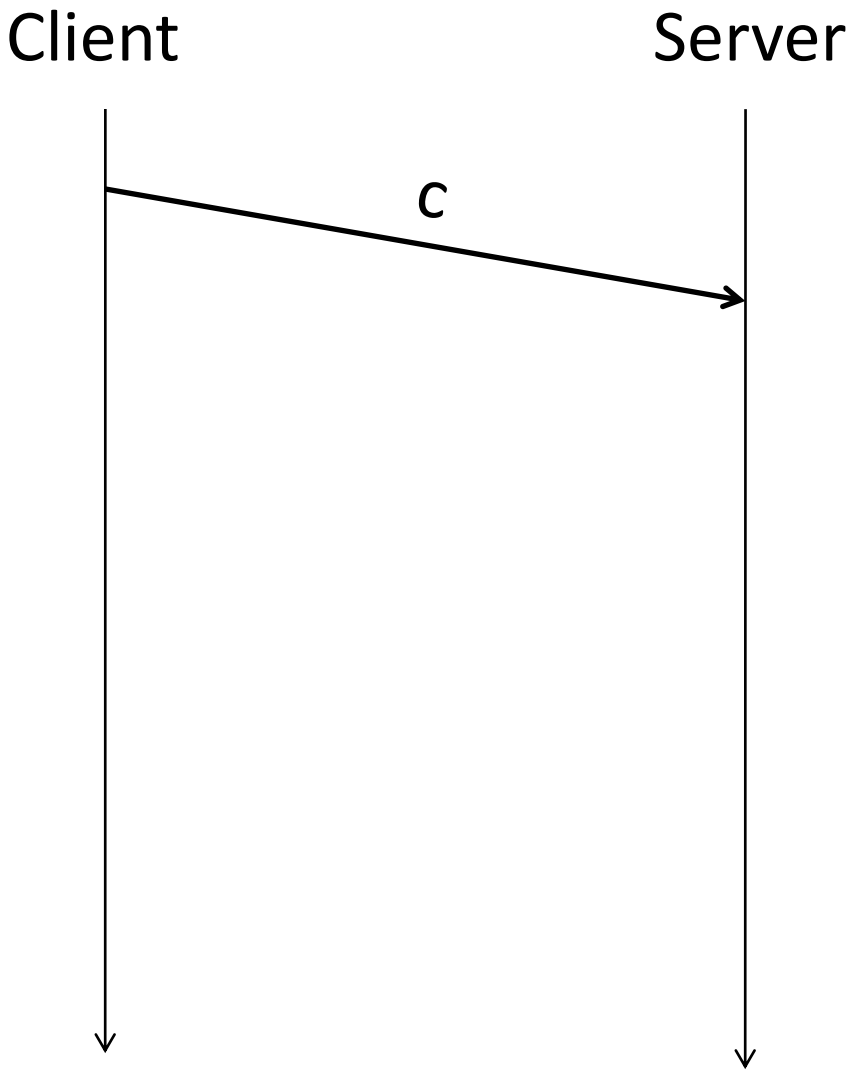
Round Trip	Congestion window size
1	3
2	6
3	12
4	24
5	44

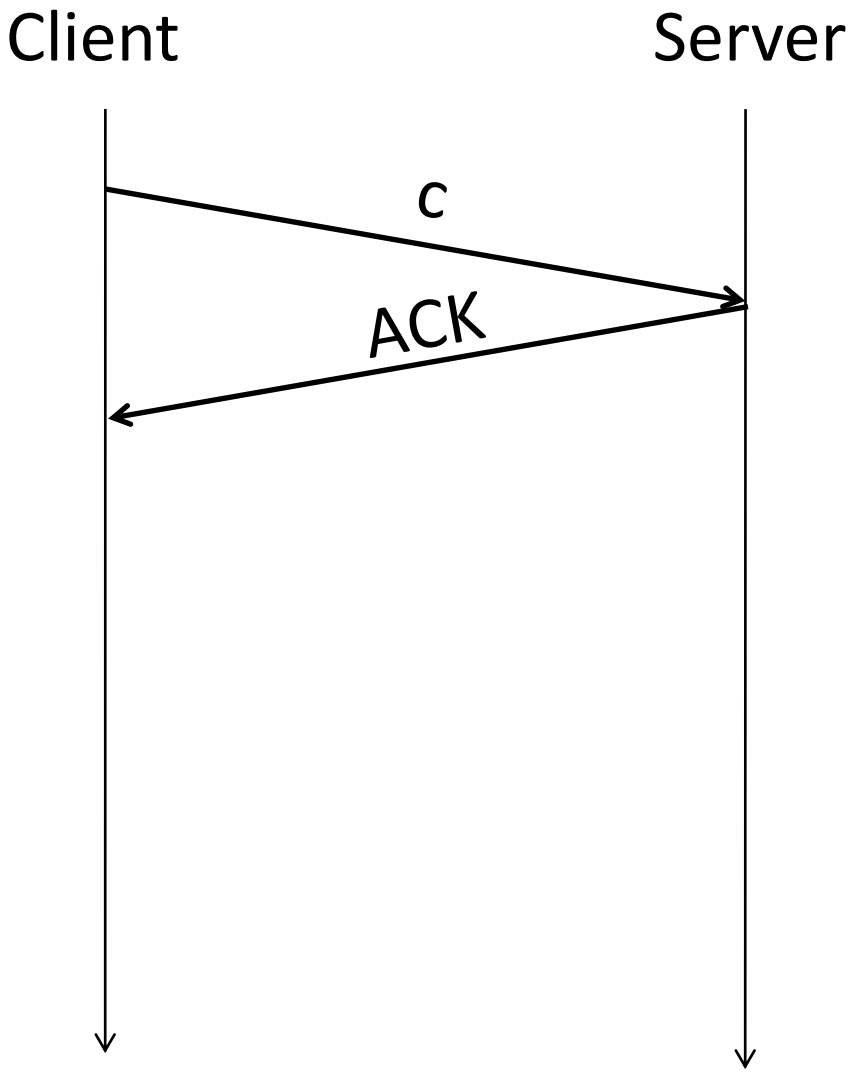
Delayed ACK

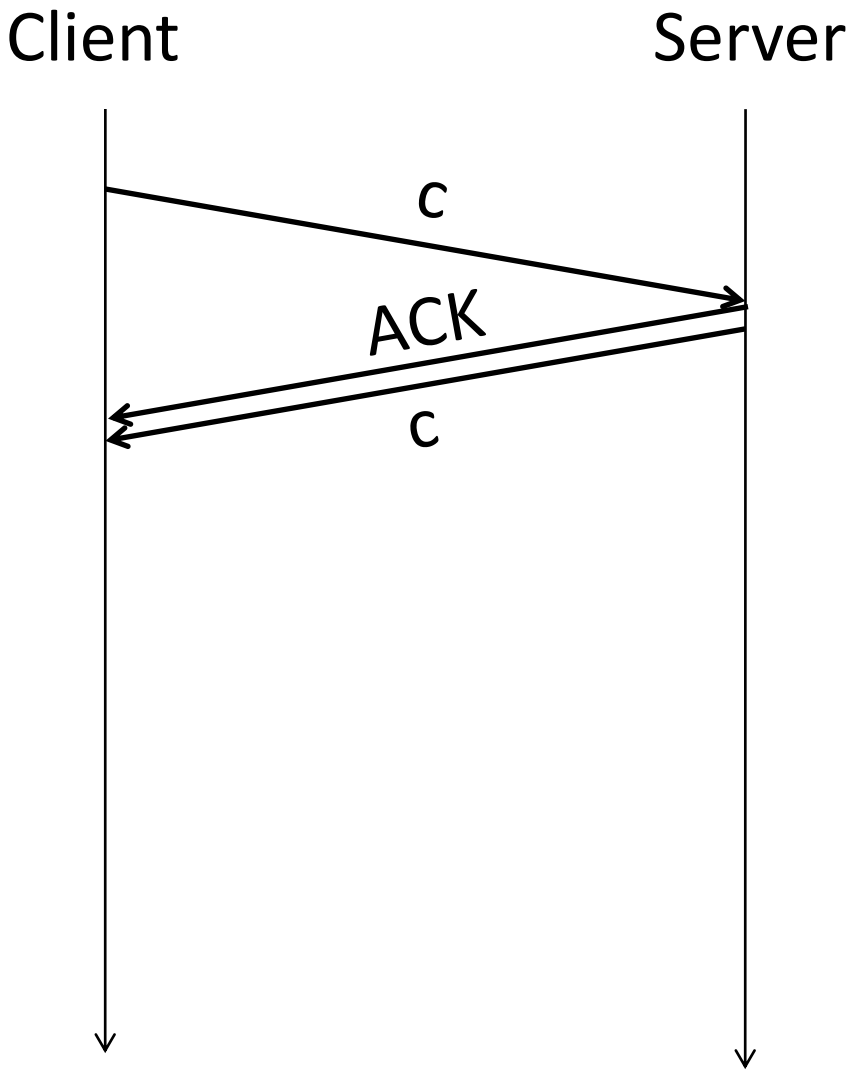
RFC 813, July 1982

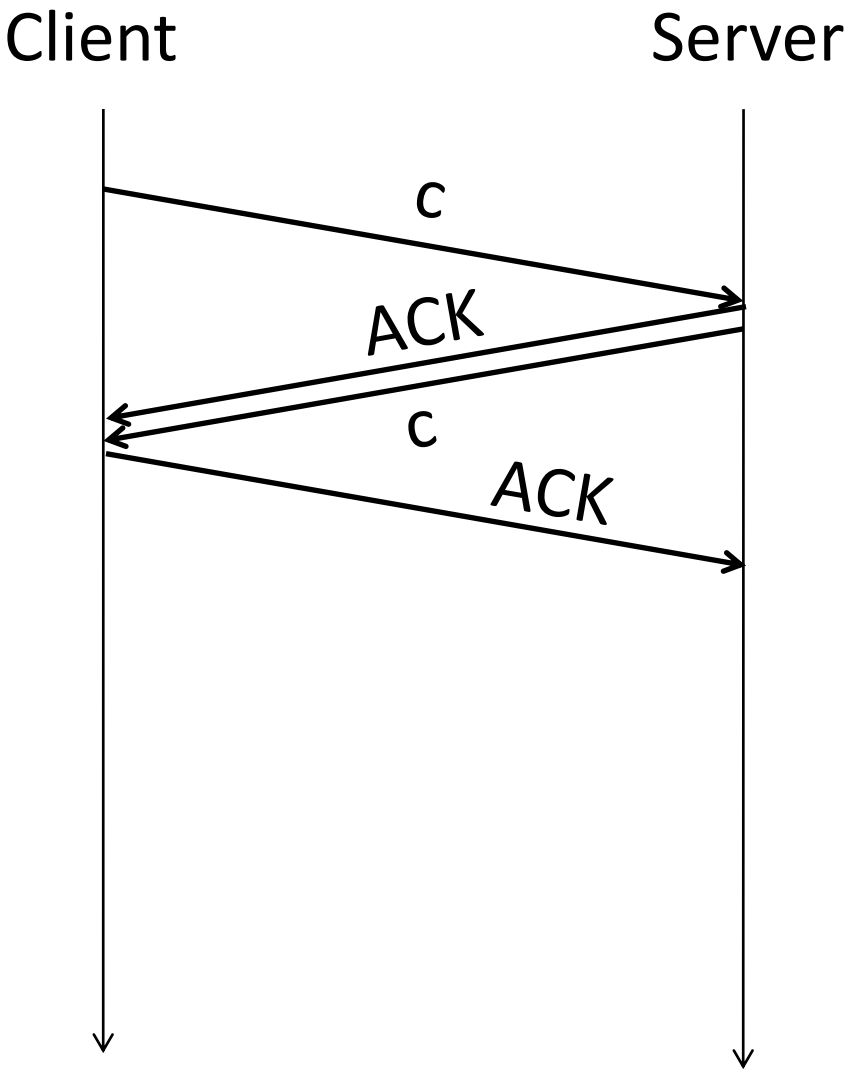
“…overly frequent acknowledgement …greatly increases the processing time at the sender's end.”

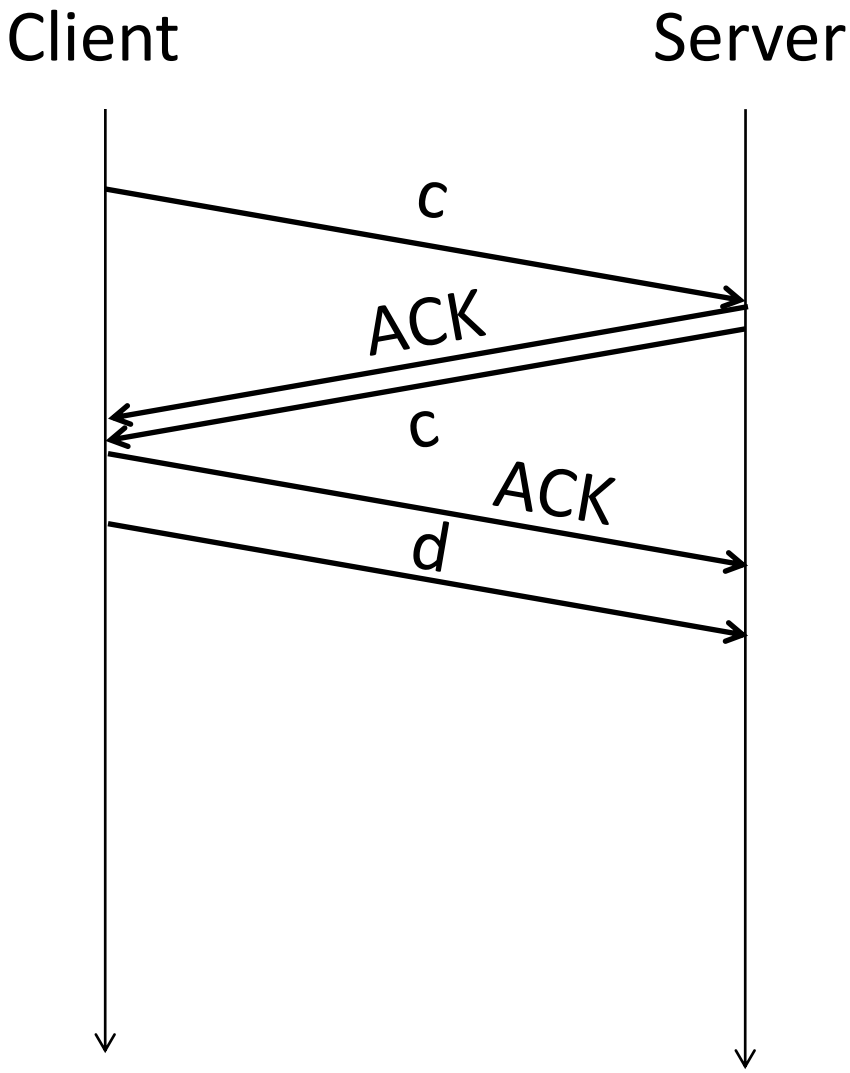
-RFC 813

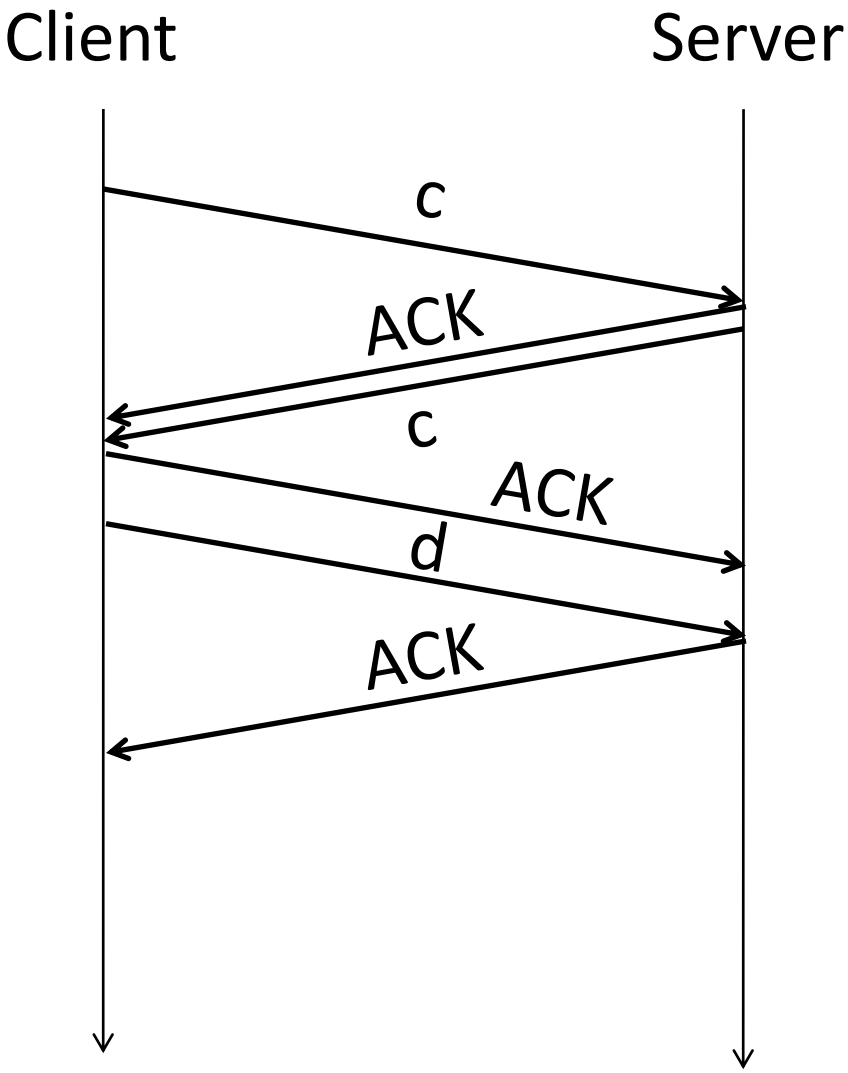

Client
Server
c

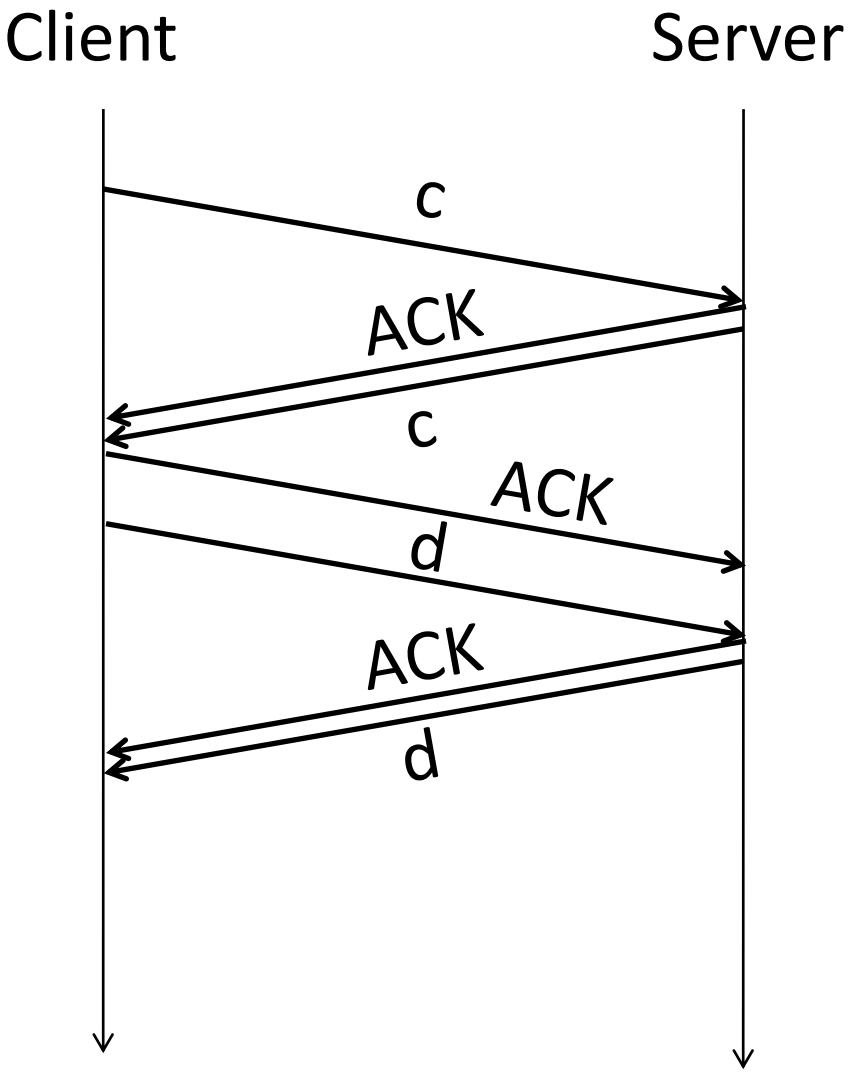

Client
Server
c
ACK

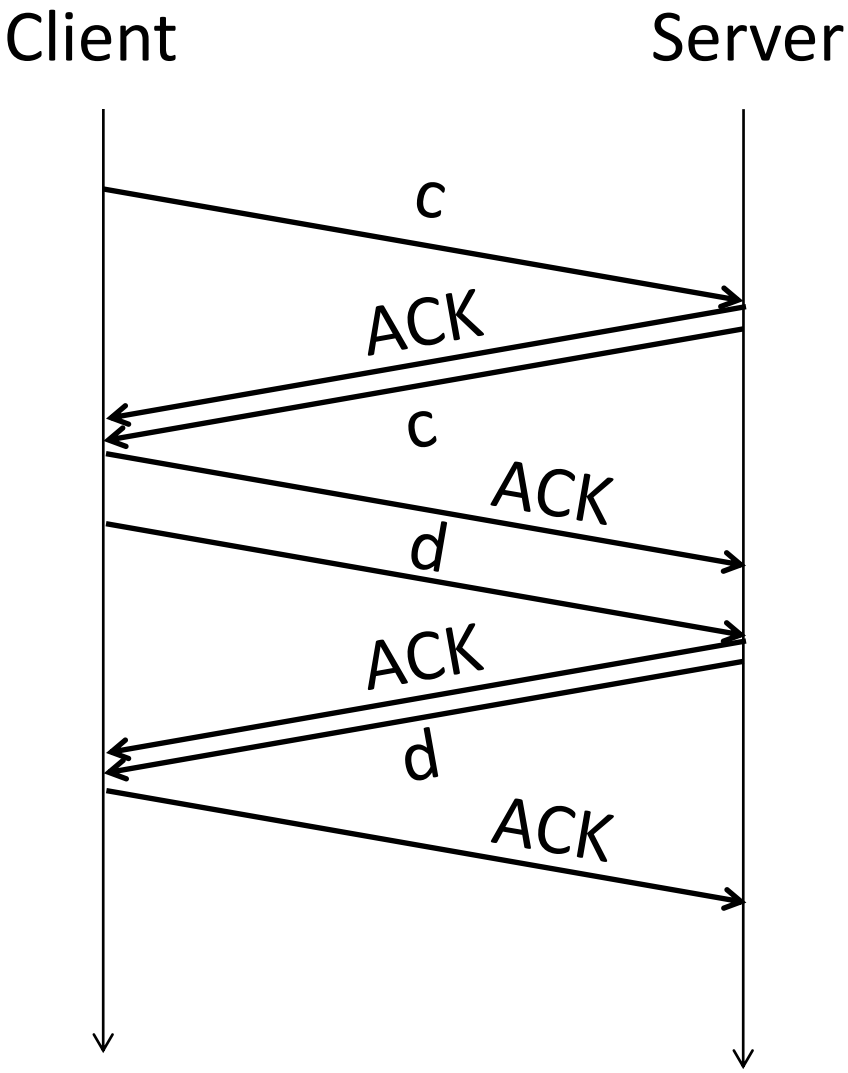

Client
Server
c
ACK
c

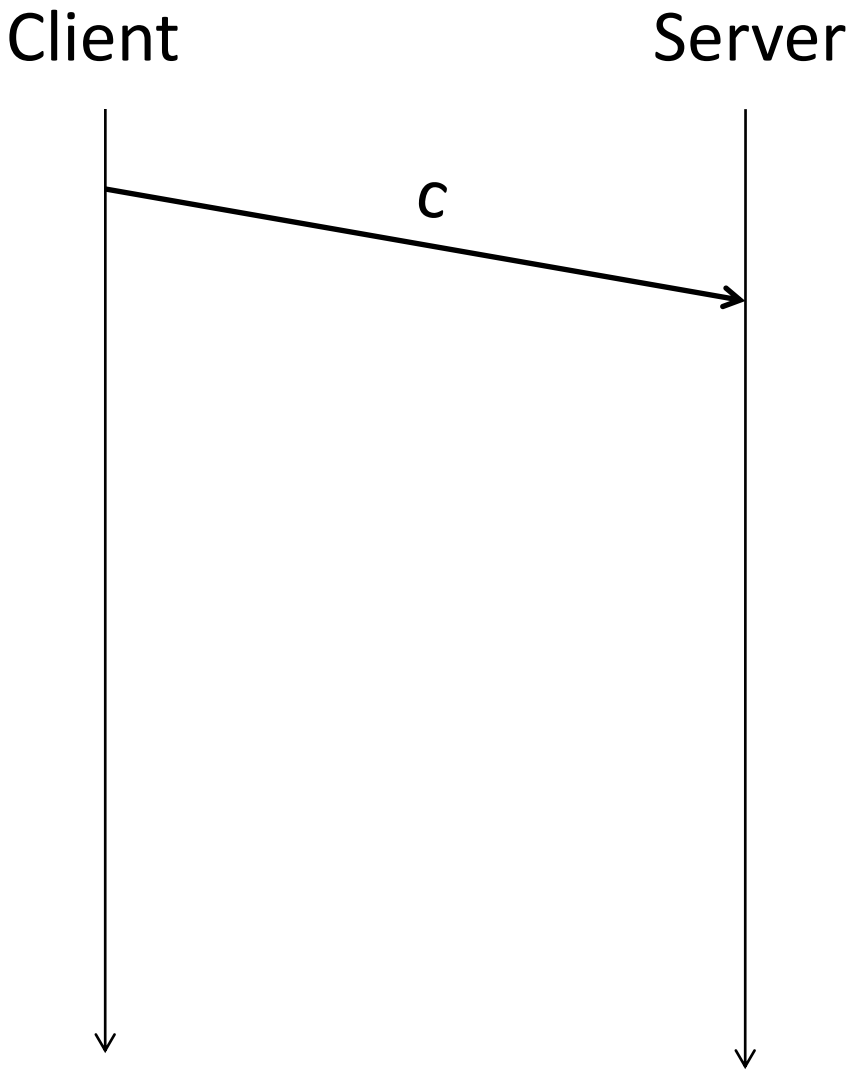

Client
Server
c
ACK
c
ACK

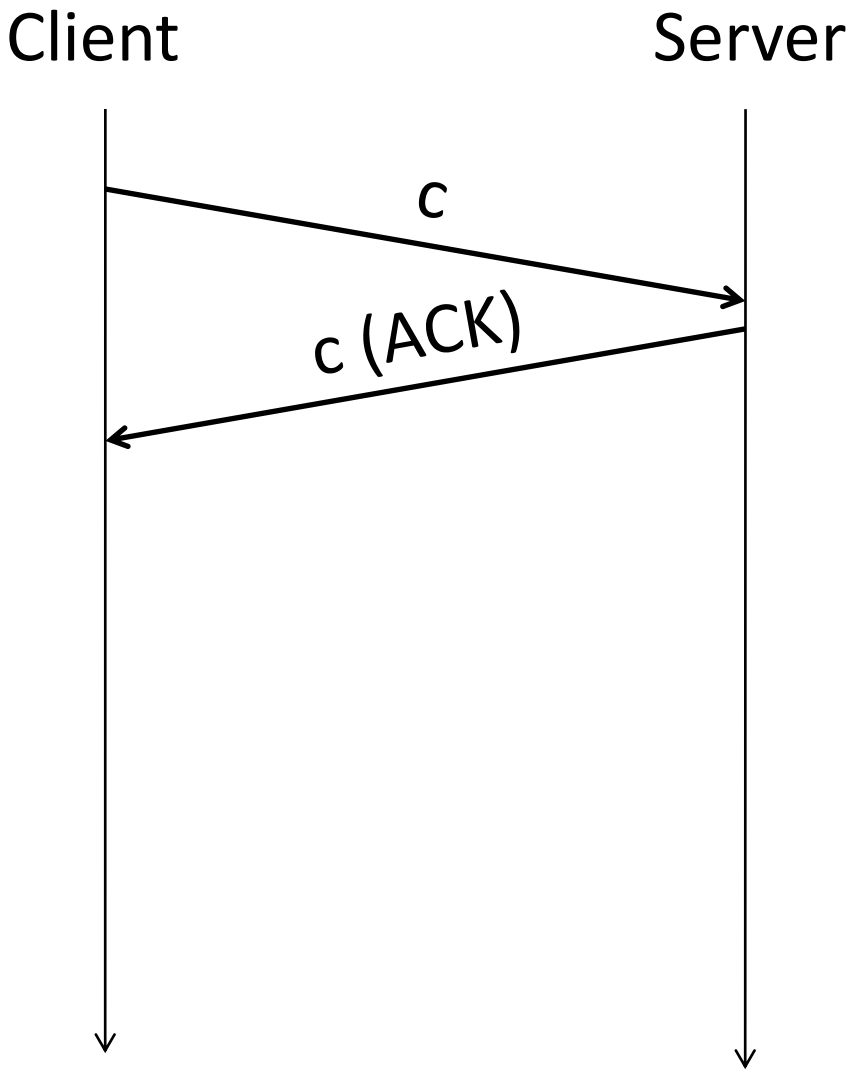

Client
Server
c
ACK
c
ACK
d

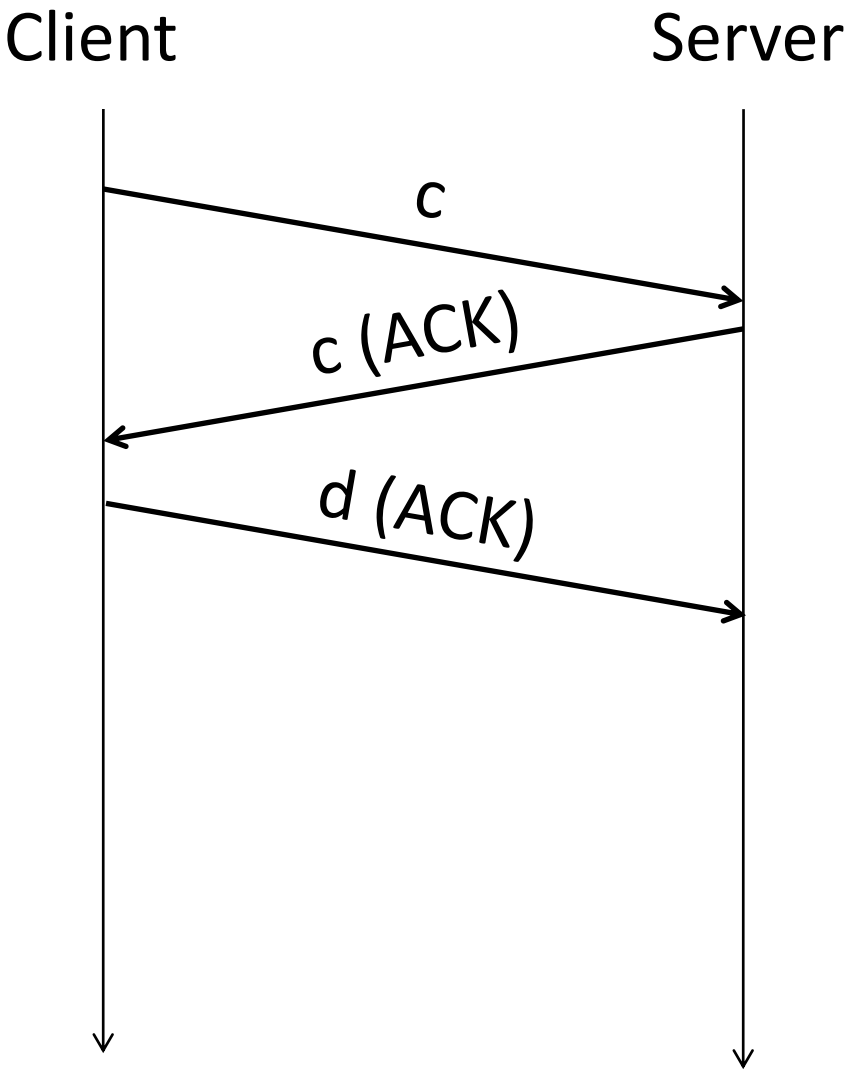

Client
Server
c
ACK
c
ACK
d
ACK

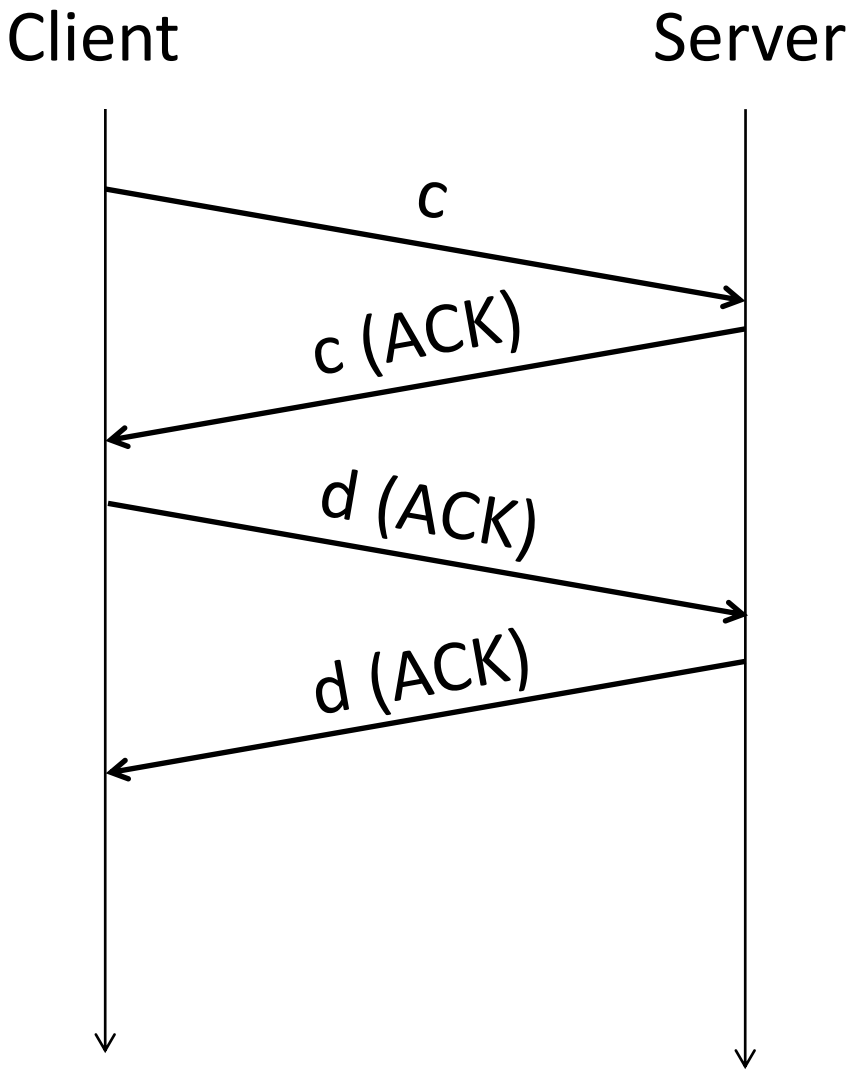

Client
Server
c
ACK
c
ACK
d
ACK
d

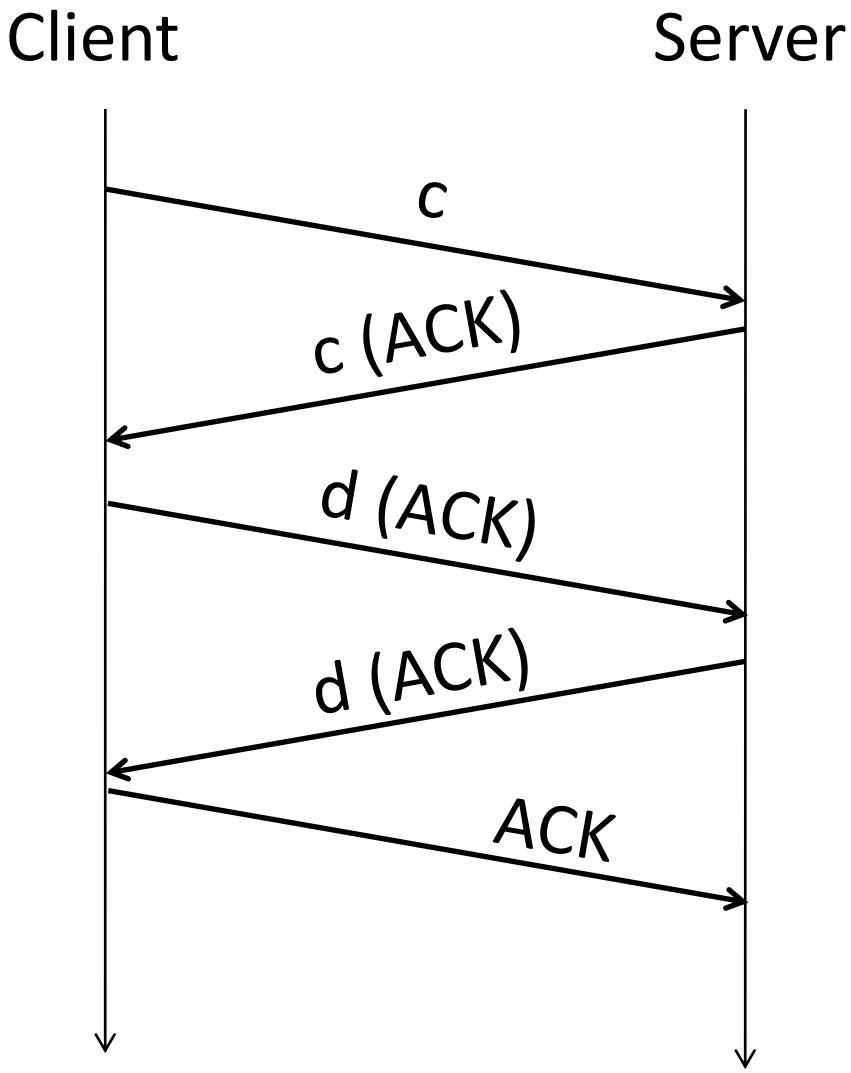

Client
Server
c
ACK
c
ACK
d
ACK
d
ACK


Client
Server
c


Client
Server
c
c (ACK)


Client
Server
c
c (ACK)
d (ACK)


Client
Server
c
c (ACK)
d (ACK)
d (ACK)


Client
Server
c
c (ACK)
d (ACK)
d (ACK)
ACK

“A TCP SHOULD implement a delayed ACK”

-RFC 1122

When a packet arrives, delay your ACK

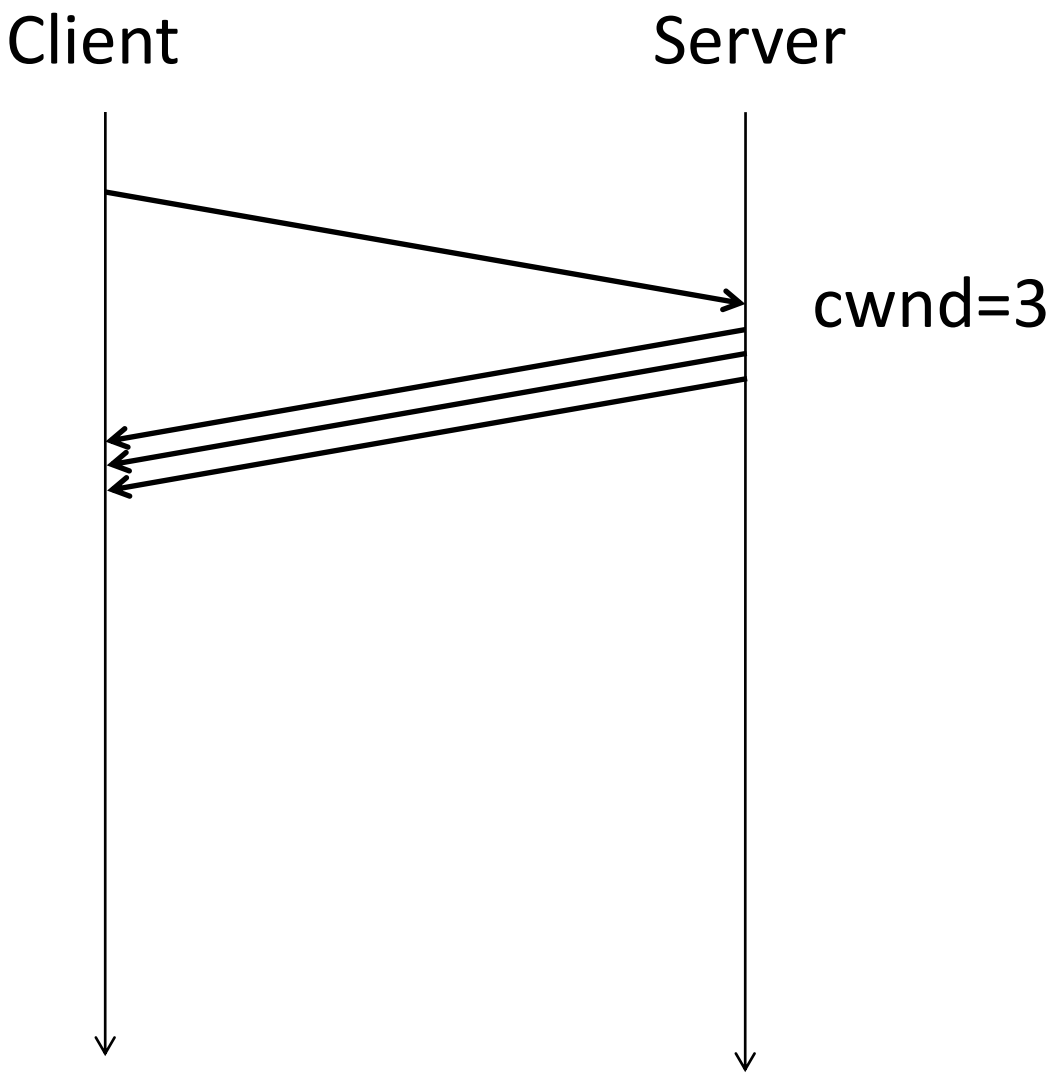
When a packet arrives, delay your ACK

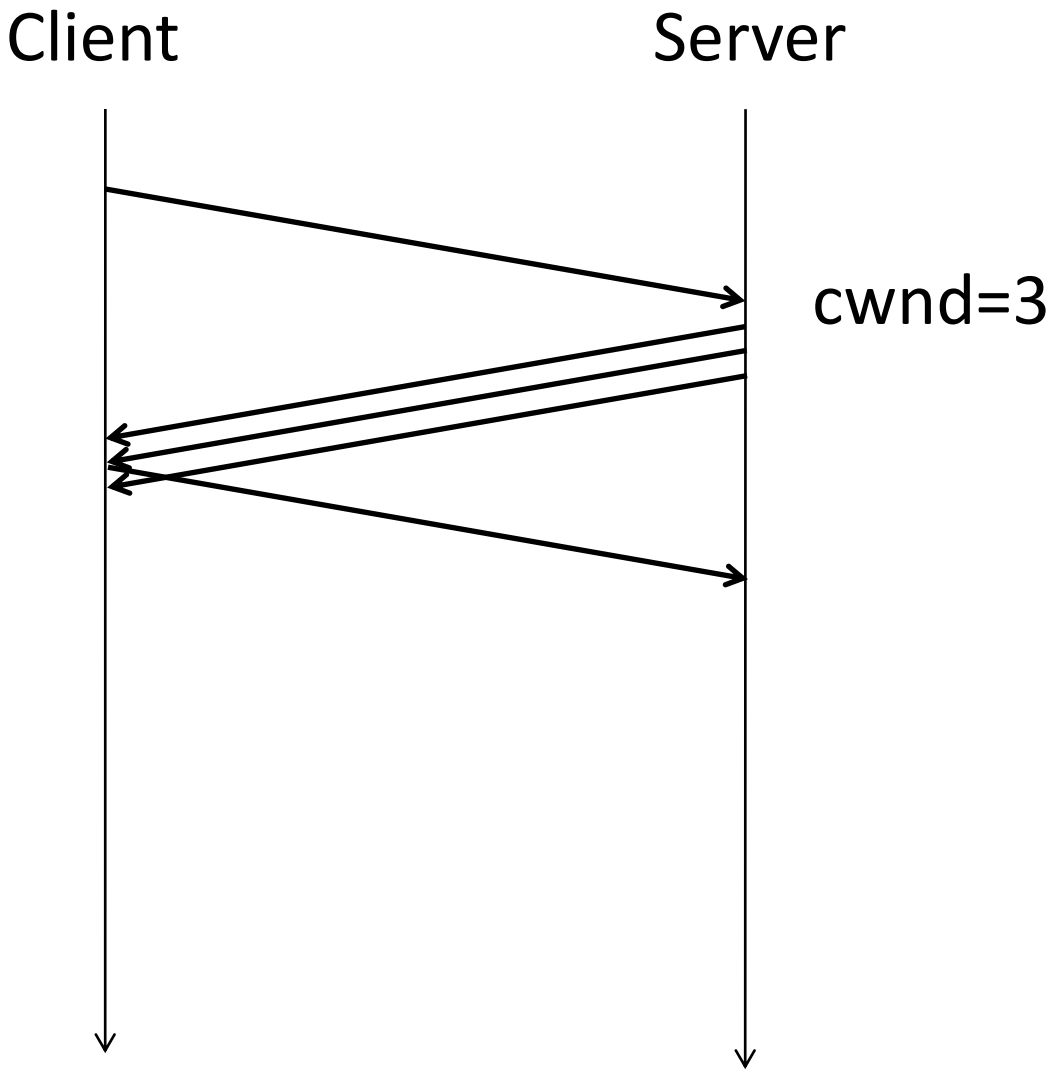
BUT

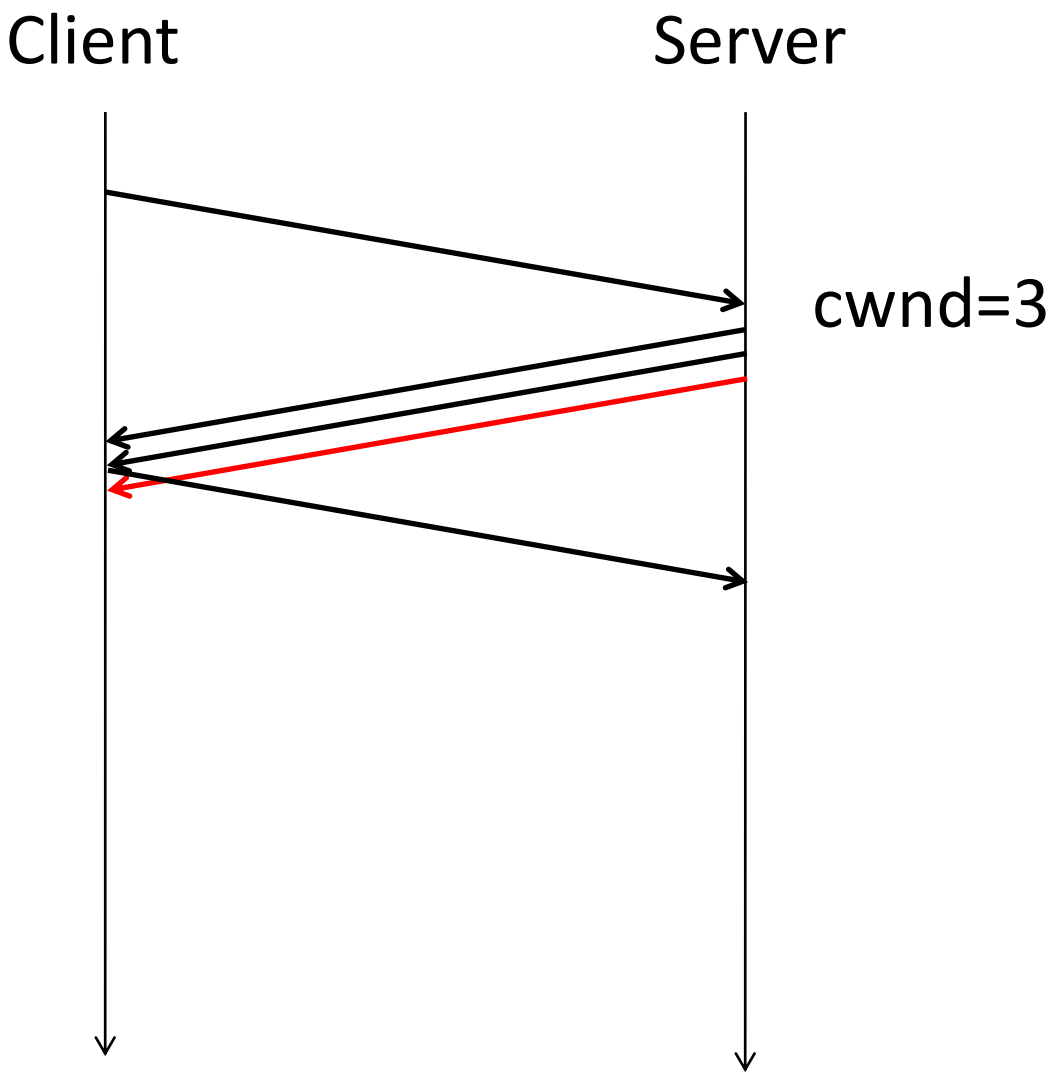
When a packet arrives, delay your ACK

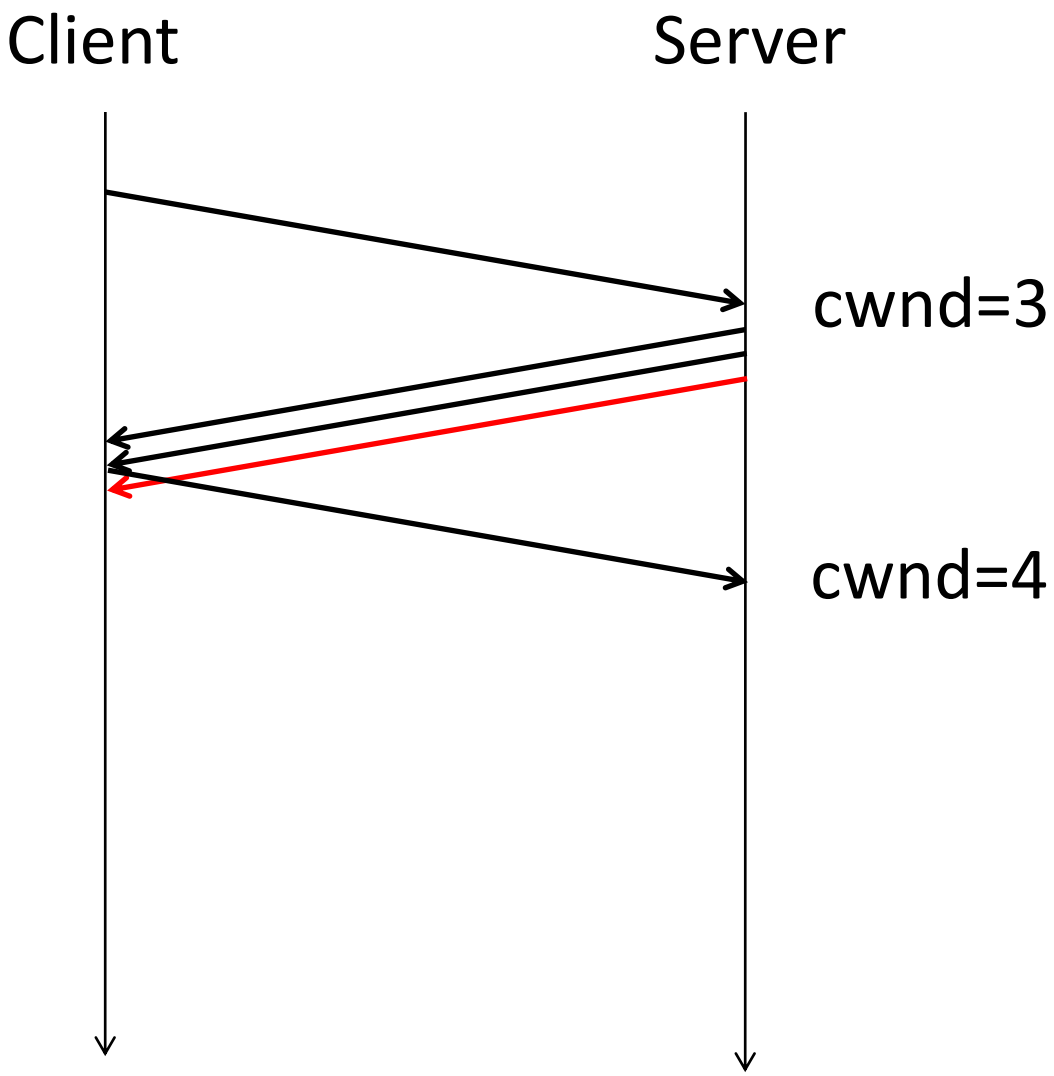
BUT

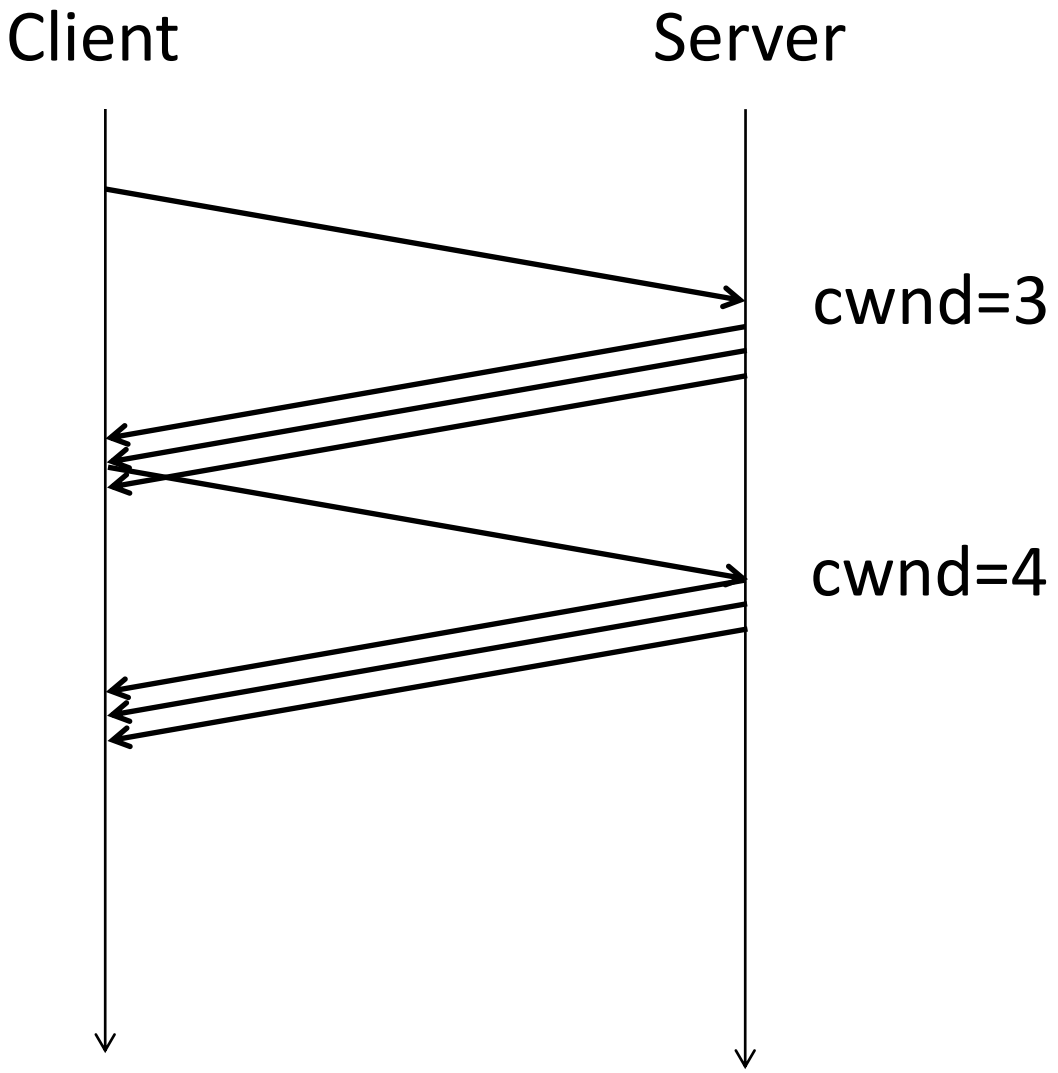
If *another* packet arrives while you're waiting, ACK both right away.

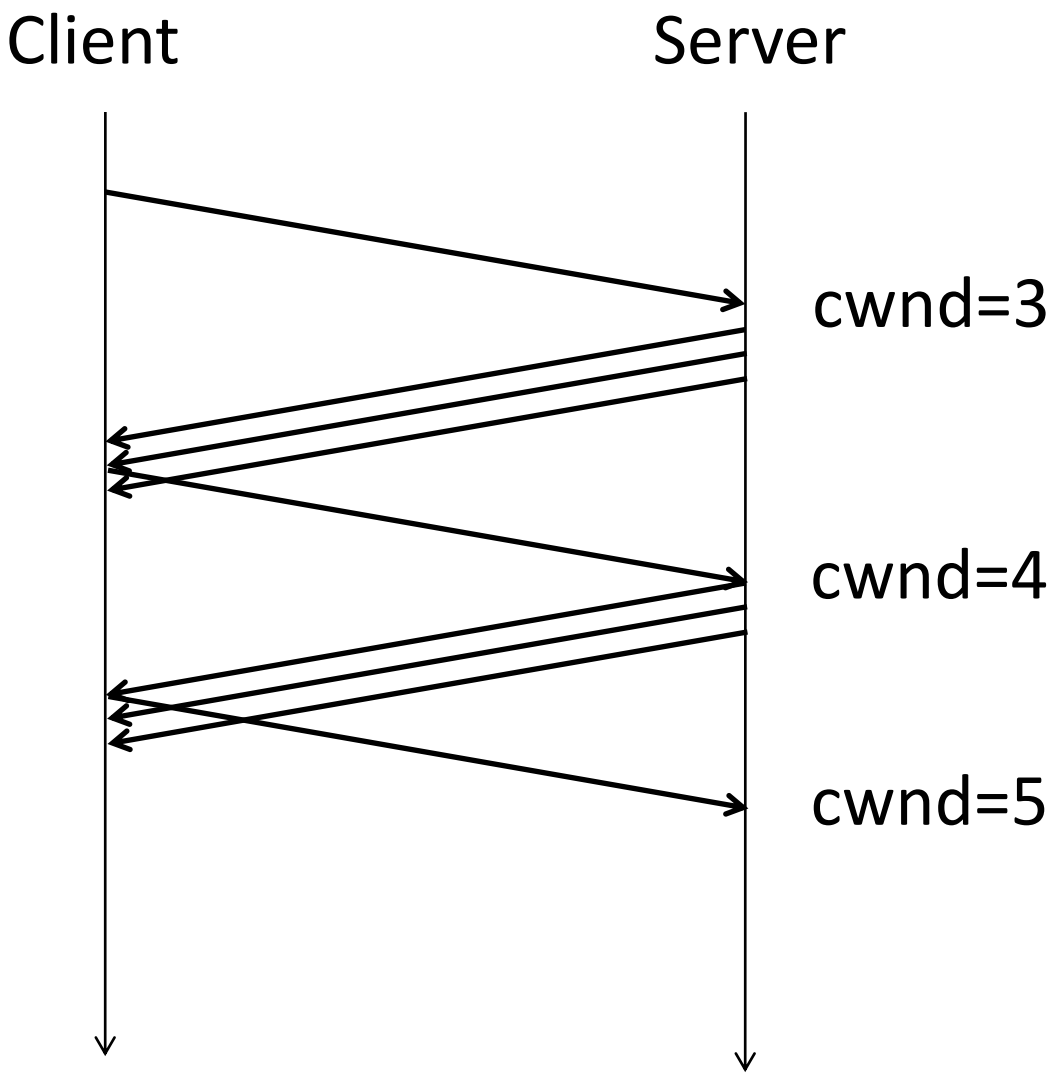

Client
Server
cwnd=3

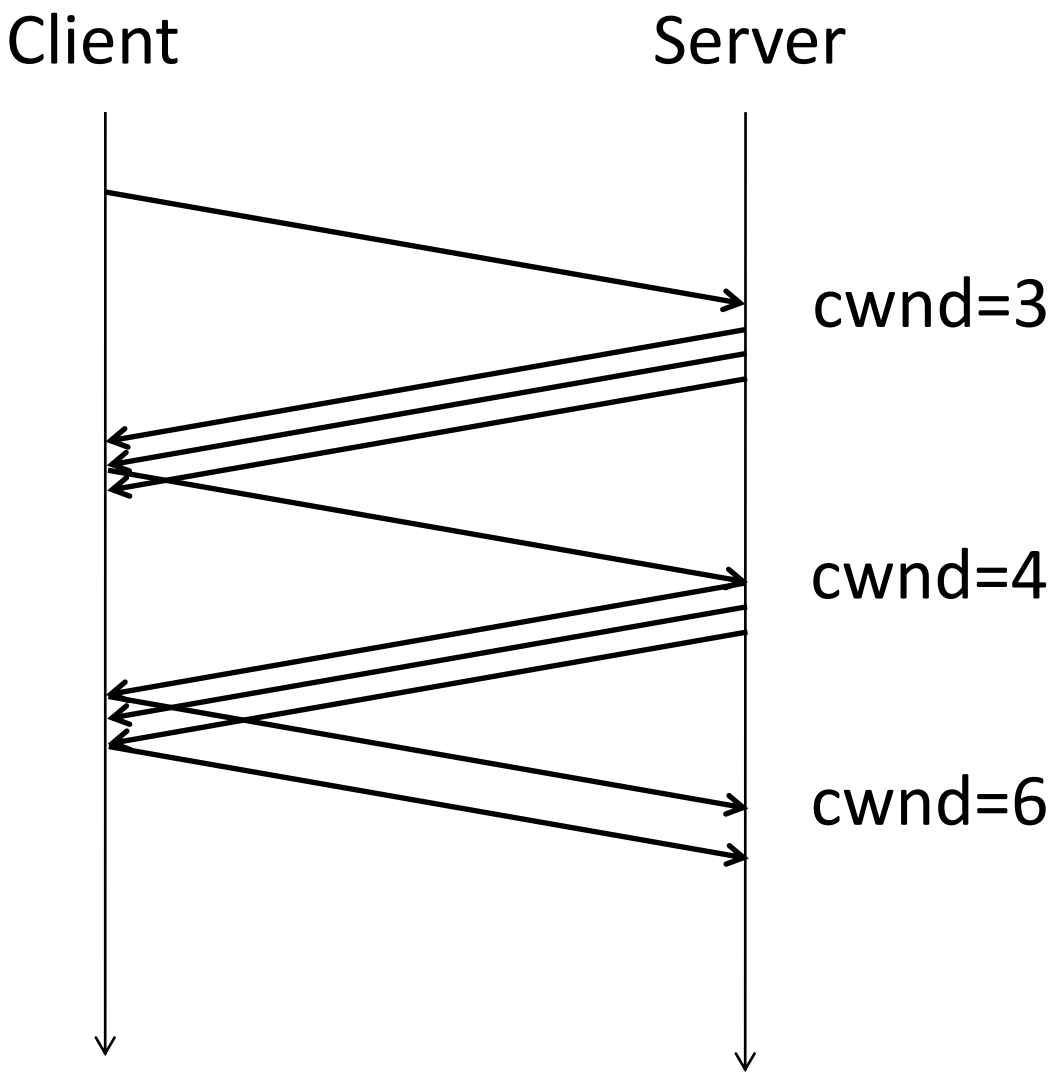

Client
Server
cwnd=3

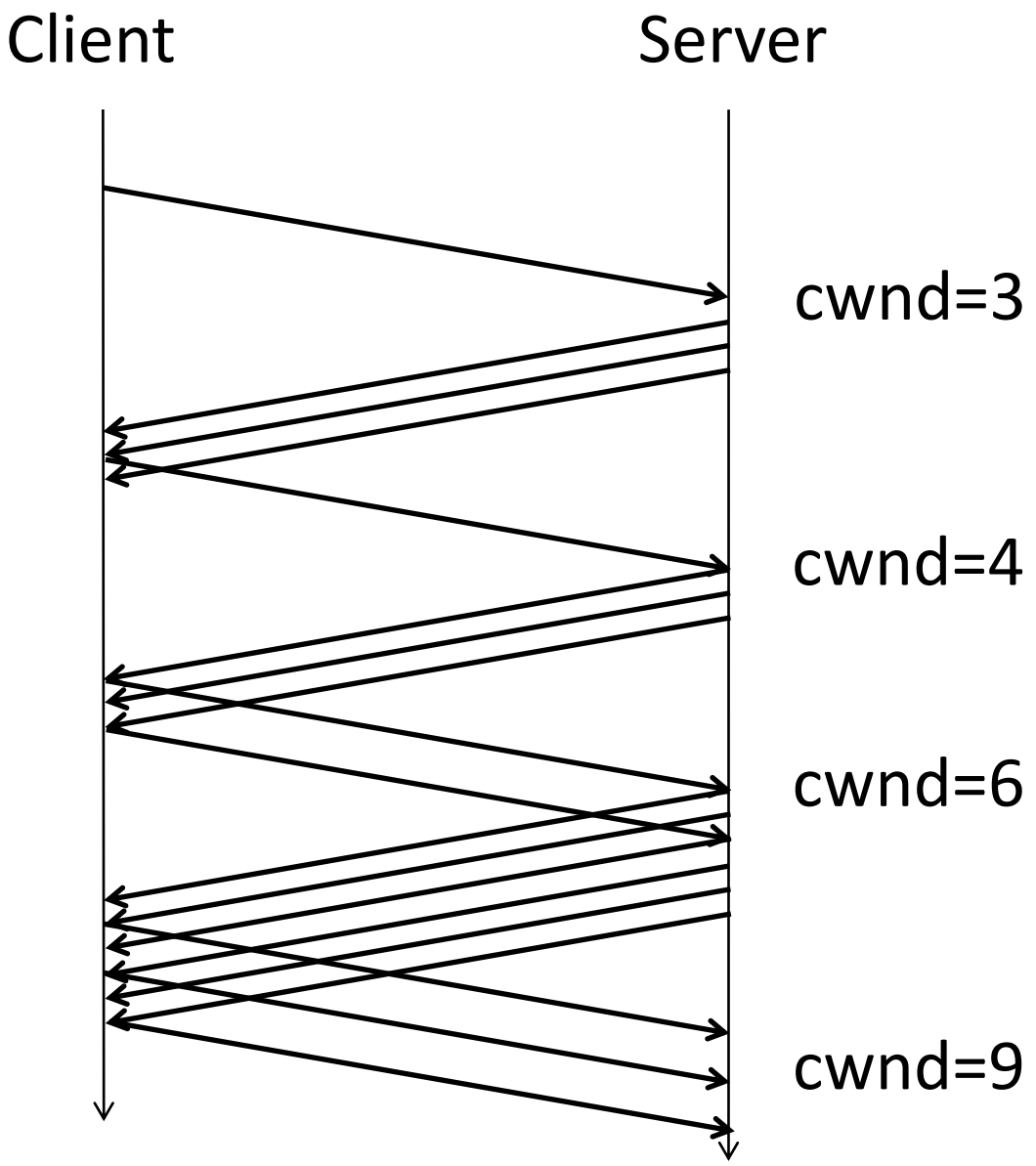

Client
Server
cwnd=3


Client
Server
cwnd=3
cwnd=4


Client
Server
cwnd=3
cwnd=4


Client
Server
cwnd=3
cwnd=4
cwnd=5

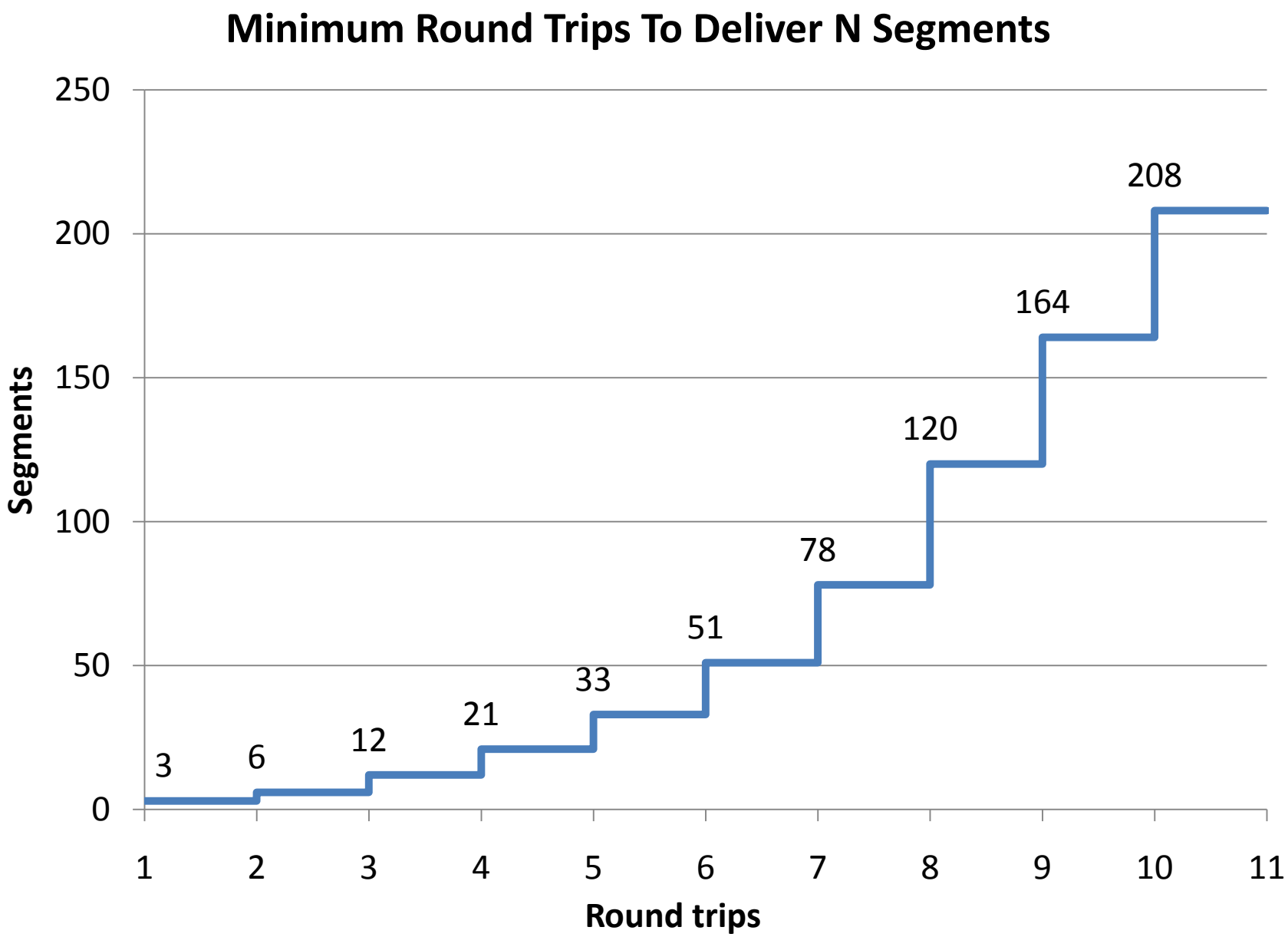

Client
Server
cwnd=3
cwnd=4
cwnd=6


Client
Server
cwnd=3
cwnd=4
cwnd=6
cwnd=9

Round Trip	Congestion window size
1	3
2	4
3	6
4	9
5	13
6	19
7	28
8	42
9	44

TCP slow start
and
delayed ACK

network latency strictly limits the throughput of new connections


Minimum Round Trips To Deliver N Segments
Segments
250
200
150
100
50
0
3
6
12
21
33
51
78
120
164
208
1
2
3
4
5
6
7
8
9
10
11
Round trips

the WHITE HOUSE PRESIDENT BARACK OBAMA

Get Email Updates | Contact Us

BLOG | PHOTOS & VIDEO | BRIEFING ROOM | ISSUES | the ADMINISTRATION | the WHITE HOUSE | our GOVERNMENT

Your WEEKLY ADDRESS

With doctors facing deep cuts in their reimbursements from Medicare unless Congress acts to correct long-standing problems, the President calls on Senate Republicans to stop blocking the remedy and pledges to work toward a permanent solution.

Watch the Video

1 2 3 4

White House Photo, Lawrence Jackson, 6/11/10

A NEW FOUNDATION

Health Reform: What It Means for You

Health Reform as signed by the President will put American families and small businesses in control of their health care,

SEARCH the SITE

Search WhiteHouse.gov

Search

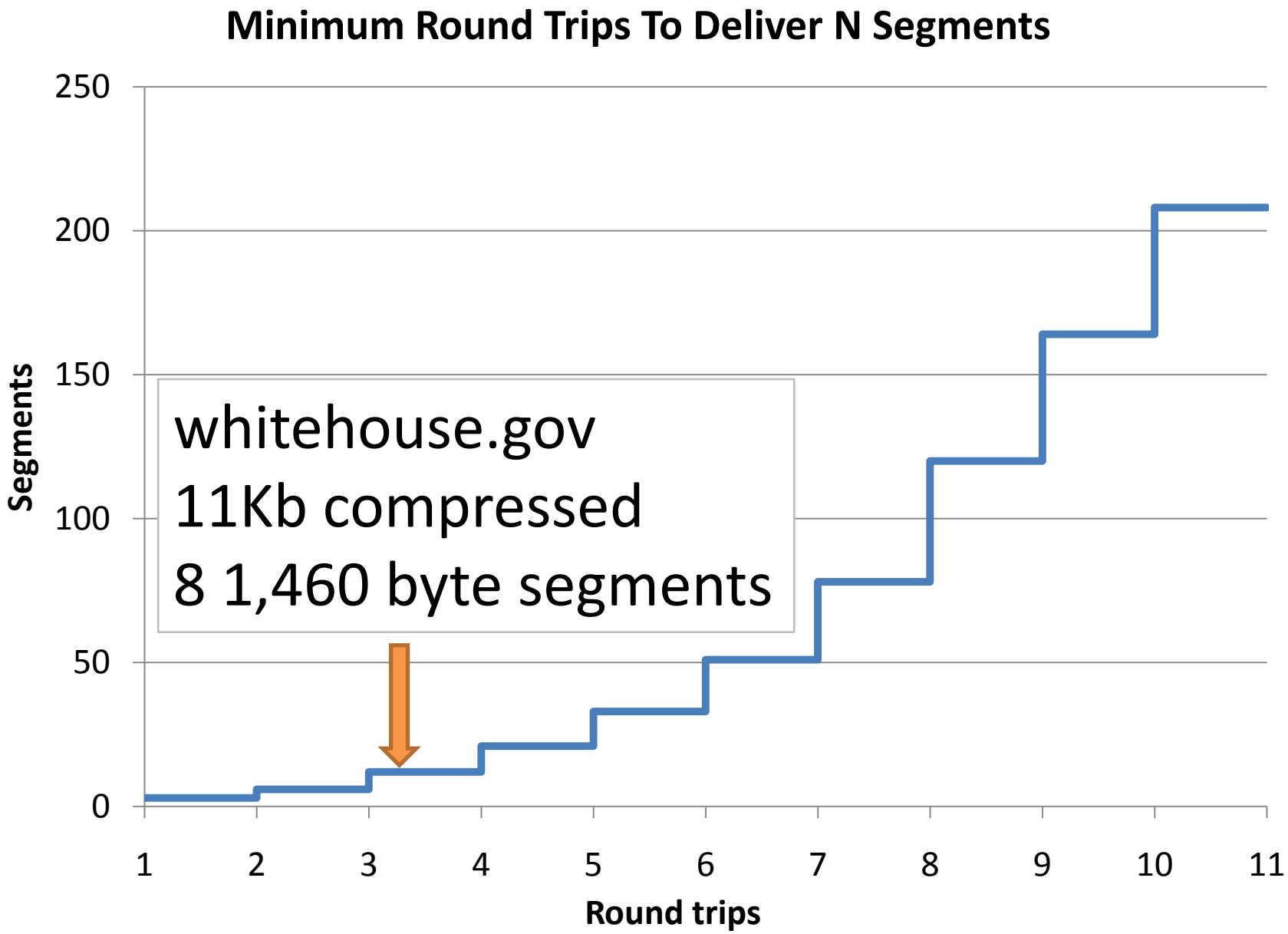
51 Kb

(just the HTML)

Accept-Encoding: gzip,deflate

Raw HTML: 51Kb

Gzipped: 11Kb


Minimum Round Trips To Deliver N Segments
250
200
150
100
50
0
Segments
whitehouse.gov
11Kb compressed
8 1,460 byte segments
1
2
3
4
5
6
7
8
9
10
11
Round trips

White House

From Wikipedia, the free encyclopedia
(Redirected from White house)

Coordinates: 38.89767°N 77.03655°W

For other uses, see White House (disambiguation).

See also: Executive Office of the President of the United States

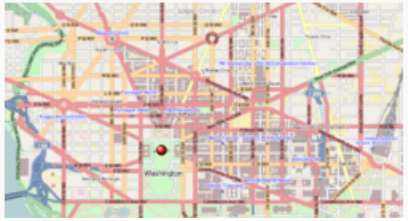
The **White House** is the official residence and principal workplace of the President of the United States. Located at 1600 Pennsylvania Avenue NW in Washington, D.C., it was designed by Irish-born James Hoban,[1] and built between 1792 and 1800 of white-painted Aquia sandstone in the late Georgian style. It has been the residence of every U.S. President since John Adams. When Thomas Jefferson moved into the home in 1801, he (with architect Benjamin Henry Latrobe) expanded the building outward, creating two colonnades that were meant to conceal stables and storage.[2]

In 1814, during the War of 1812, the mansion was set ablaze by the British Army in the Burning of Washington, destroying the interior and charring much of the exterior. Reconstruction began almost immediately, and President James Monroe moved into the partially reconstructed house in October 1817. Construction continued with the addition of the South Portico in 1824 and the North in 1829. Because of crowding within the executive mansion itself, President Theodore Roosevelt had nearly all work offices relocated to the newly constructed West Wing in 1901. Eight years later, President William Howard Taft expanded the West Wing and created the first Oval Office which was eventually moved as the section was expanded. The third-floor attic was converted to living quarters in 1927 by augmenting the existing hip roof with long shed dormers. A newly constructed East Wing was used as a reception area for social events; Jefferson's colonnades connected the new wings. East Wing alterations were completed in 1946, creating additional office space. By 1948, the house's load-bearing exterior walls and internal wood beams were found to be close to failure. Under Harry S. Truman, the interior rooms were completely dismantled and a new internal load-bearing steel frame constructed inside the walls. Once this work was completed, the interior rooms were rebuilt.

Today, the White House Complex includes the Executive Residence, West Wing, Cabinet Room, Roosevelt Room, East Wing, and the Old Executive Office Building, which houses the executive offices of the President and Vice President.

White House

South façade of the White House



Main page
Contents
Featured content
Current events
Random article

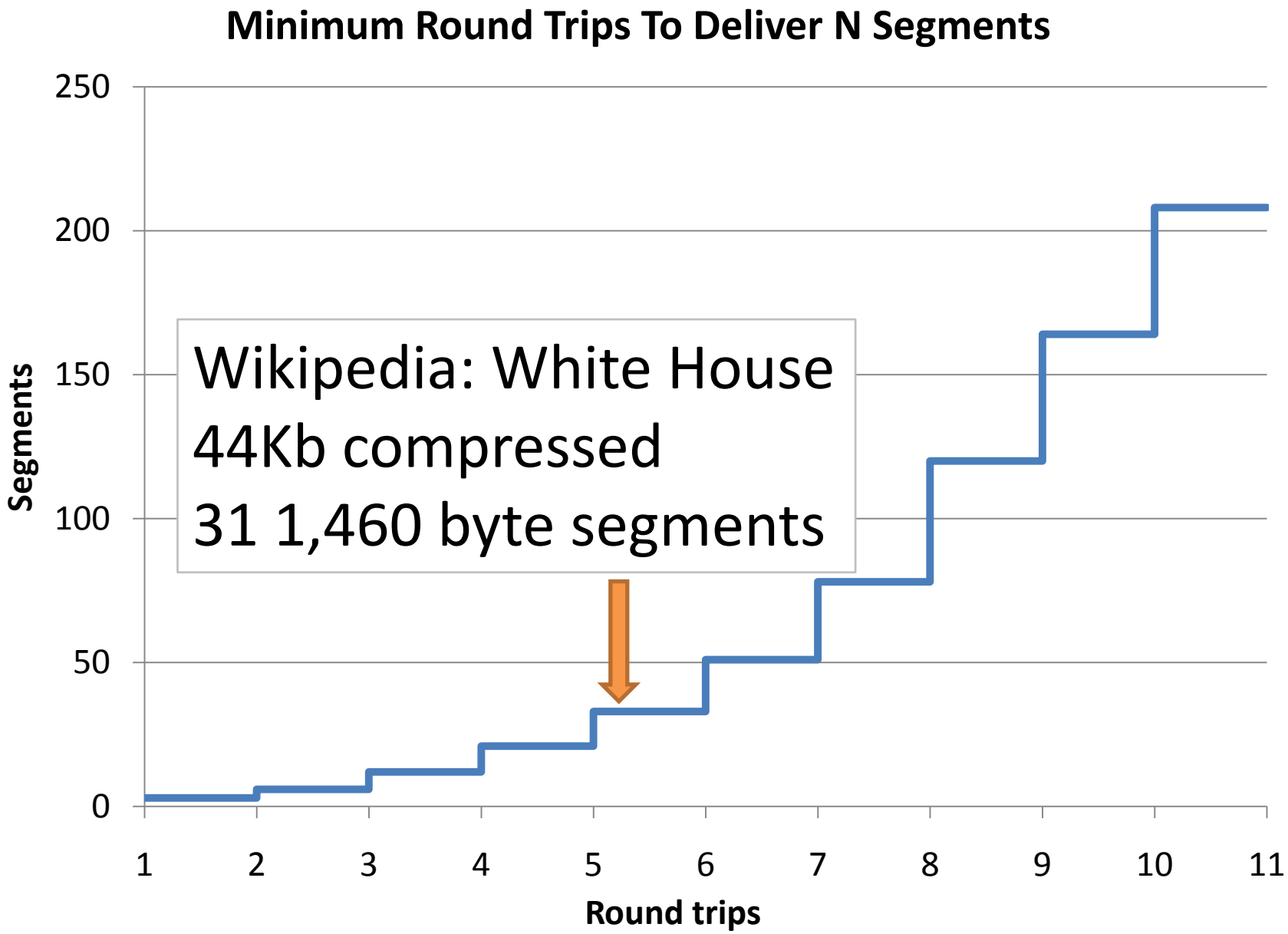
Interaction
About Wikipedia
Community portal
Recent changes
Contact Wikipedia
Donate to Wikipedia
Help

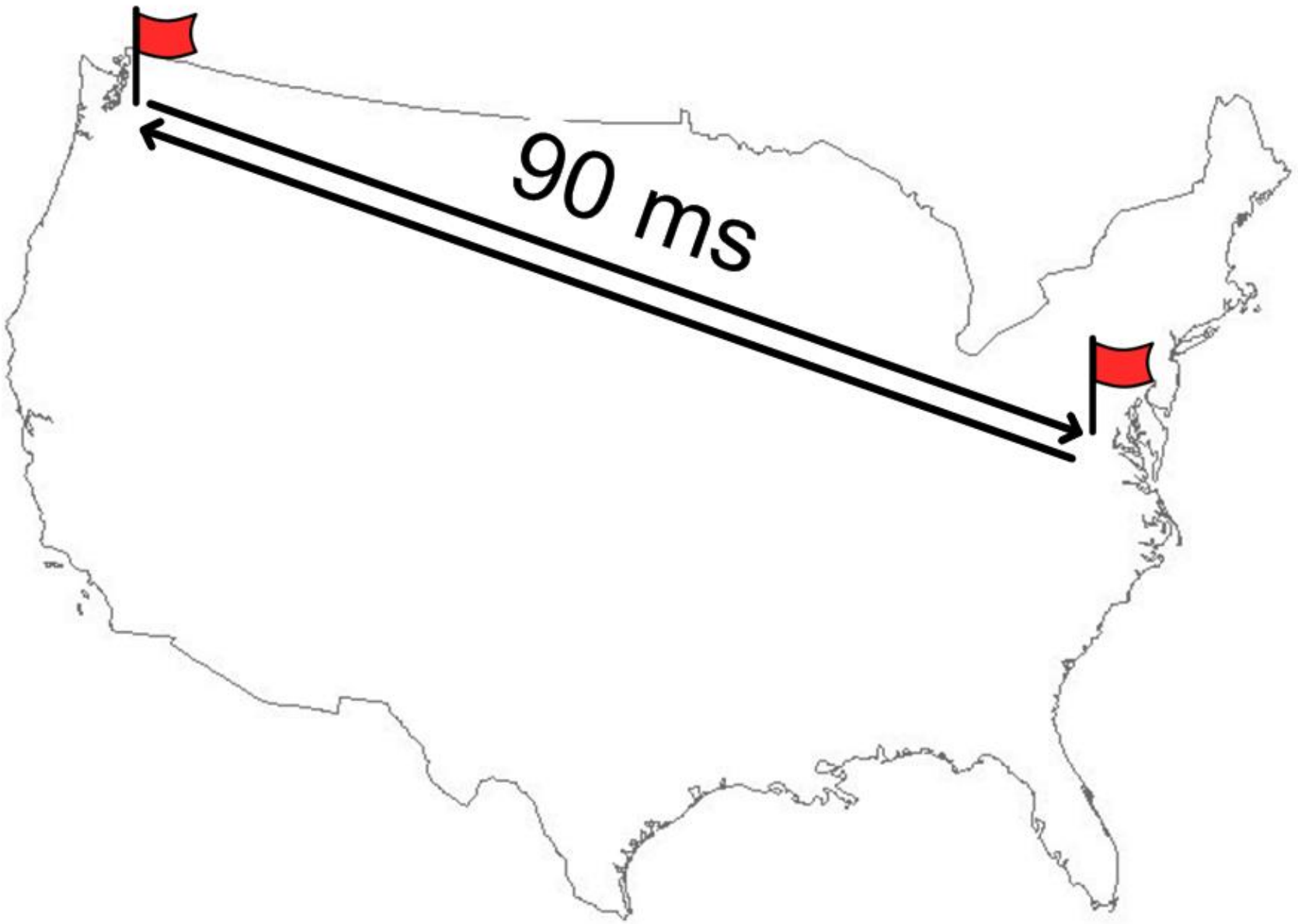
Toolbox

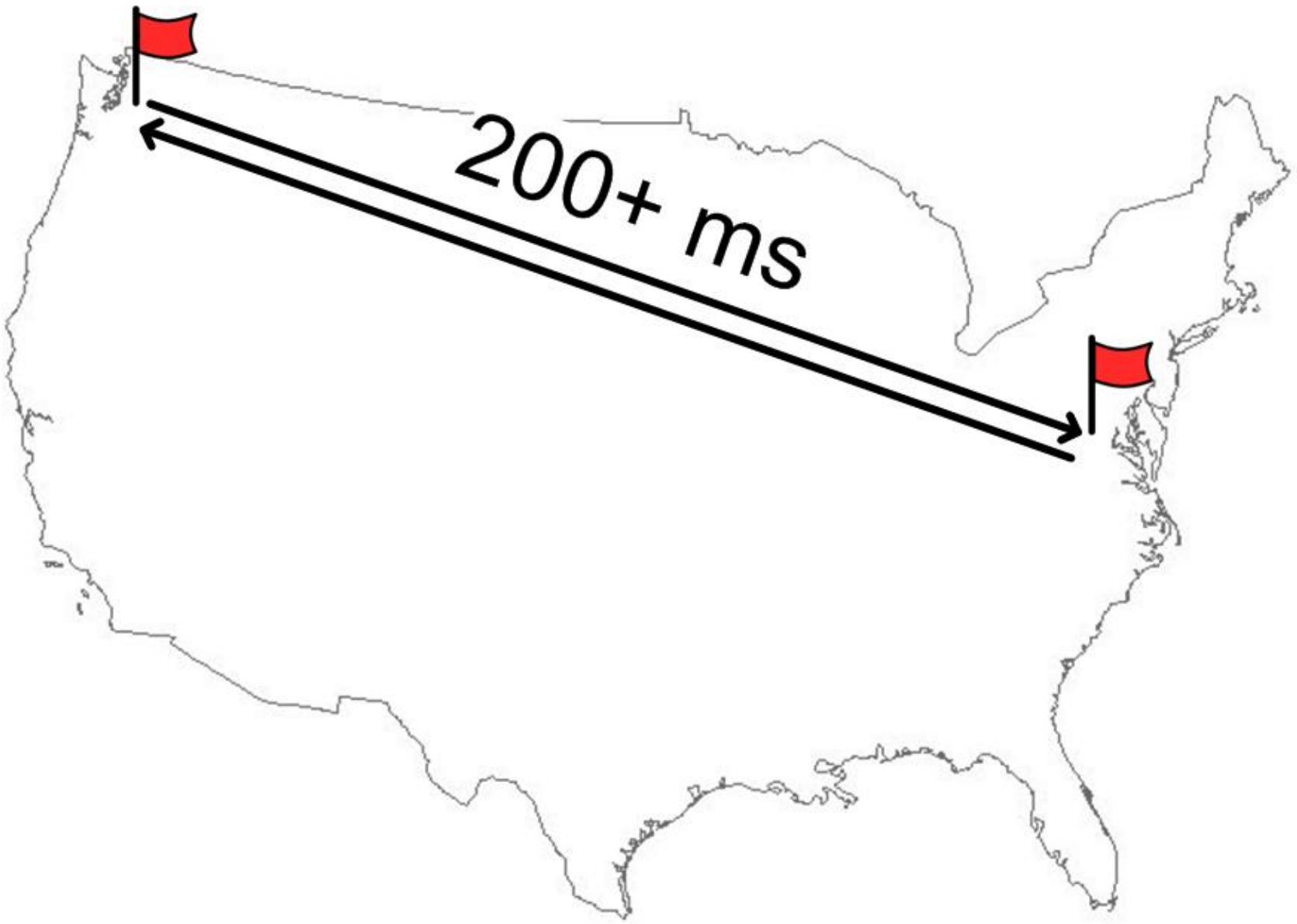
Print/export

Languages
Afrikaans
العربية
Azərbaycan
বাংলা
Bosanski

210 Kb
(just the HTML)

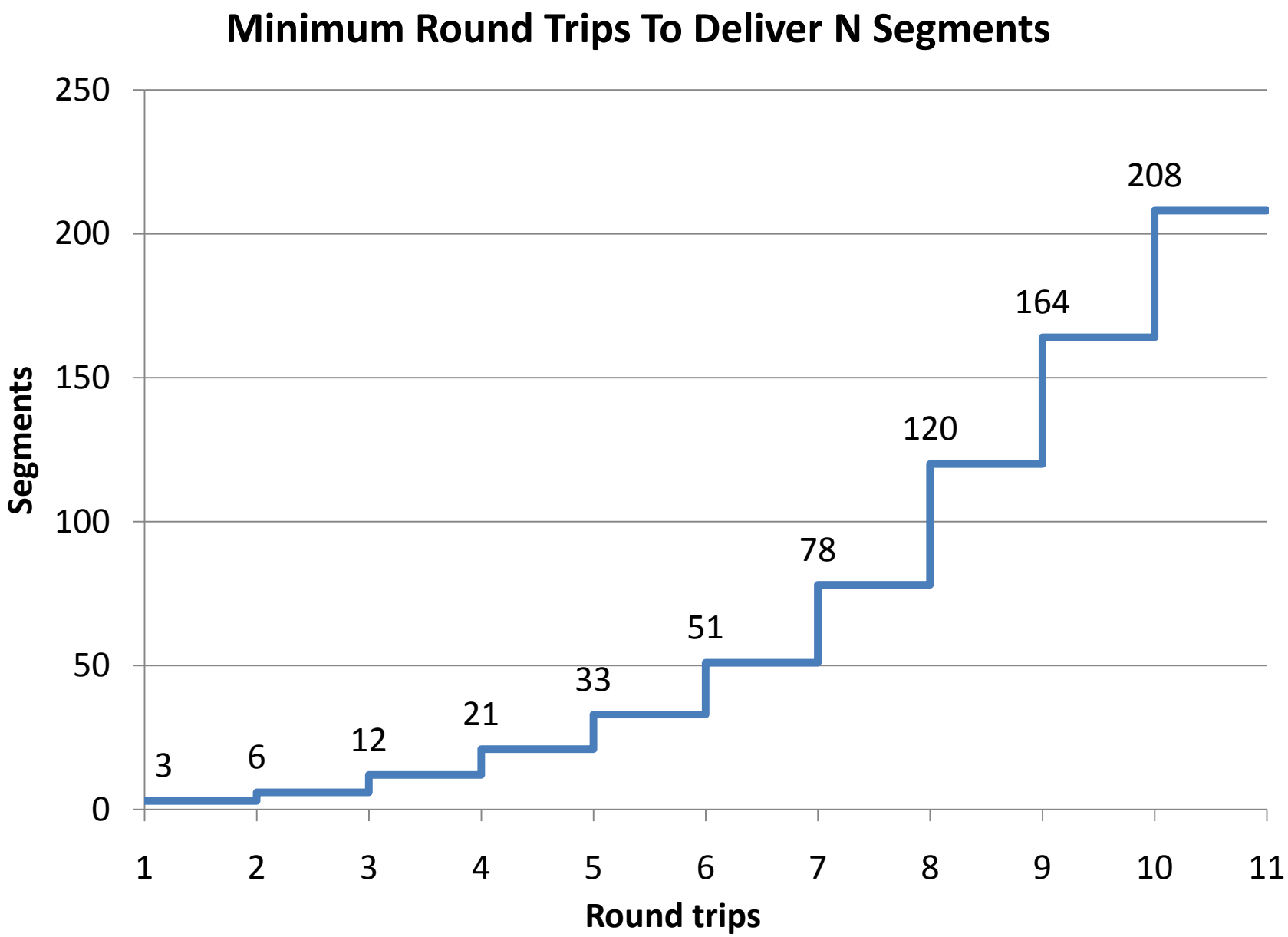

Minimum Round Trips To Deliver N Segments
250
200
150
100
50
0
Segments
Wikipedia: White House
44Kb compressed
31 1,460 byte segments
1
2
3
4
5
6
7
8
9
10
11
Round trips


90 ms


200+ ms

What to do about it?

1) Carefully consider every byte of content


Minimum Round Trips To Deliver N Segments
Segments
250
200
150
100
50
0
3
6
12
21
33
51
78
120
164
208
1
2
3
4
5
6
7
8
9
10
11
Round trips

```
          <li id="nav-the-admin" class="first-level">
  <a href="/administration" class="first-level"><span>The Administration</span></a>
            <div class="topnav-dd-outer"><div class="topnav-dd-inner clearfix">
                    <ul class="two-col clearfix">
                        <li><a href="/administration/president-obama">President Barack Obama</a></li>
                        <li><a href="/administration/vice-president-biden">Vice President Joe Biden</a></li>
                        <li><a href="/administration/first-lady-michelle-obama">First Lady Michelle Obama</a></li>
                        <li><a href="/administration/jill-biden">Dr. Jill Biden</a></li>
                      </ul>
                    <ul class="two-col clearfix">
                        <li><a href="/administration/cabinet">The Cabinet</a></li>
                        <li><a href="/administration/staff">White House Staff</a></li>
                        <li><a href="/administration/eop">Executive Office of the President</a></li>
                        <li><a href="/administration/other-advisory-boards">Other Advisory Boards</a></li>
                      </ul>
                </div></div><!--/ .topnav-dd-outer, .topnav-dd-inner -->

</li>
          <li id="nav-the-wh" class="first-level">
  <a href="/about" class="first-level"><span>About the White House</span></a>
            <div class="topnav-dd-outer"><div class="topnav-dd-inner clearfix">
                    <ul class="two-col clearfix">
                        <li><a href="/about/history">History</a></li>
                        <li><a href="/about/presidents">Presidents</a></li>
                        <li><a href="/about/first-ladies">First Ladies</a></li>
                        <li><a href="/about/oval-office">The Oval Office</a></li>
                        <li><a href="/about/vp-residence">The Vice President&#039;s Residence &amp;amp; Office</a></li>
                        <li><a href="/about/eeob">Eisenhower Executive Office Building</a></li>
                      </ul>
                    <ul class="two-col clearfix">
                        <li><a href="/about/camp-david">Camp David</a></li>
                        <li><a href="/about/air-force-one">Air Force One</a></li>
                        <li><a href="/about/fellows">White House Fellows</a></li>
                        <li><a href="/about/internships">White House Internships</a></li>
                        <li><a href="/about/white-house-101">White House 101</a></li>
                        <li><a href="/about/tours-and-events">Tours &amp;amp; Events</a></li>
                      </ul>
                </div></div><!--/ .topnav-dd-outer, .topnav-dd-inner -->

</li>
          <li id="nav-the-gov" class="first-level">
  <a href="/our-government" class="first-level"><span>Our Government</span></a>
            <div class="topnav-dd-outer"><div class="topnav-dd-inner clearfix">
```

Accept-Encoding: gzip,deflate

Tony Gentilcore:

~15% of users don’t do gzip

http://en.oreilly.com/velocity2009/public/schedule/detail/9072

Software	Accept-Encoding modification
Ad Muncher	Stripped
CA Internet Security Suite	Accept-EncodXng: gzip, deflate
CEQURUX	Stripped
Citrix Application Firewall	Stripped
ISA 2006	Stripped
McAfee Internet Security 6.0	XXXXXXXXXXXXXX: ++++++++++++
Norton Internet Security 2005	----------: ----------
Novell iChain 2.3	Stripped
Novell Client Firewall	Stripped
WebWasher	Stripped
ZoneAlarm Pro 5.5	XXXXXXXXXXXXXX: XXXXXXXXXXXX

Source: http://en.oreilly.com/velocity2009/public/schedule/detail/9072

1) Carefully consider every byte of content

2) Think about what goes into those first few packets

2.1) Keep your cookies small

2.2) Open connections for assets in the first three packets

2.3) Download small assets first

3) Accept the speed of light

Meta lessons

1) If your application is delivered on the web, you need to understand how the network functions

2) Humility

end