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
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
\[
\frac{1}{1.5} = 0.66
\]
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
From my house: 90 ms
Theoretical fiber: 37 ms ~ 2
It’s been this way for over a decade.
“Once you have bad latency you're stuck with it.”
Fascinating!
Alice captured a Neutral Base
Alice captured a Neutral Pillbox
Network latency matters for web applications
History of the Internet
September 1981
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
GET index.html
Client → Server

GET index.html

Response
Client

GET index.html

Response

ACK

Server
Client

GET index.html

Server

Response
GET index.html

Response
Client

GET index.html

Response

Response

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

Client ——— GET index.html ——— Server
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
August 1983

BSD 4.2
Growth in Internet hosts 1981-1991

Data from RFC 1296
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

Client → Server

Request
Window: 64kb
Max Segment: 1kb

Client

Response

Server
Client

GET index.html

Server

Response

Response
Window: 64kb
Max Segment: 1kb
“…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

- Polo (sun 3/50)
- Hot (sun 3/50)
- Surf (sun 3/50)
- Vs (sun 3/50)
- csam
- cartan
- Renoir (vax 750)
- VanGogh (vax 8600)
- Monet (vax 750)
- Okeeffe (CCI)

Connections:
- Polo to csam: 230.4 Kbs Microwave
- csam to cartan: 10 Mbs Ethemets
- cartan to Renaor: 10 Mbs Ethemets
- cartan to VanGogh: 10 Mbs Ethemets
- cartan to Monet: 10 Mbs Ethemets
- cartan to Okeeffe: 10 Mbs Ethemets
Figure 7: Multiple conversation test setup

- Polo (sun 3/50)
- Hot (sun 3/50)
- Surf (sun 3/50)
- Vs (sun 3/50)
- csam 230.4 Kbs Microwave cartan

10 Mbs Ethemets

- Renoir (vax 750)
- VanGogh (vax 8600)
- Monet (vax 750)
- Okeeffe (CCI)
Figure 7: Multiple conversation test setup

- Polo (sun 3/50)
- Hot (sun 3/50)
- Surf (sun 3/50)
- Vs (sun 3/50)
- csam (230.4 Kbs Microwave)
- cartan
- Renoir (vax 750)
- VanGogh (vax 8600)
- Monet (vax 750)
- Okeeffe (CCI)

10 Mbs Ethemets
Figure 7: Multiple conversation test setup
Figure 3: Startup behavior of TCP without Slow-start

- Y-axis: Packet Sequence Number (KB)
- X-axis: Send Time (sec)
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

Packet Sequence Number (KB)

Send Time (sec)
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.
\( \min(\text{window, cwnd}) = \) 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

cwnd=6

cwnd=12
Figure 3: Startup behavior of TCP without Slow-start
Figure 4: Startup behavior of TCP with Slow-start

Packet Sequence Number (KB) vs. Send Time (sec)
June 1988

BSD4.3 Tahoe
November 1988

Congestion Avoidance and Control

Van Jacobson
Michael J Karels

“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
1 round trip

Client

Server
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

Client

Server
cwnd=3

1 round trip

4380 bytes
No problem!

It’s only on the first hit
Yahoo 2007:

One hit in five is uncached

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.
$2^{**16} = \frac{65536}{1460} = 44 \text{ segments}$
<table>
<thead>
<tr>
<th>Round Trip</th>
<th>Congestion window size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
</tr>
<tr>
<td>Round Trip</td>
<td>Congestion window size</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
</tr>
</tbody>
</table>
Delayed ACK

RFC 813, July 1982
“...overly frequent acknowledgement ...greatly increases the processing time at the sender's end.”

-RFC 813
Client \rightarrow Server

\text{c} \rightarrow \text{c (ACK)}

\text{d (ACK)} \rightarrow \text{d (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

cwnd=4
cwnd = 3

Client

(Server)
cwnd = 4
cwnd = 5
Client

Server

cwnd=3

cwnd=4

cwnd=6
Client

Server

cwnd=3


cwnd=4


cwnd=6


cwnd=9
<table>
<thead>
<tr>
<th>Round Trip</th>
<th>Congestion window size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>42</td>
</tr>
<tr>
<td>9</td>
<td>44</td>
</tr>
</tbody>
</table>
TCP slow start and delayed ACK
network latency strictly limits the throughput of new connections
Minimum Round Trips To Deliver N Segments
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.

Search the SITE
Search WhiteHouse.gov
51 Kb
(just the HTML)
Accept-Encoding: gzip, deflate
Raw HTML: 51Kb

Gzipped: 11Kb
Minimum Round Trips To Deliver N Segments

whitehouse.gov
11Kb compressed
8 1,460 byte segments
White House

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 charing 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 1946, 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.
210 Kb
(just the HTML)
Minimum Round Trips To Deliver N Segments

Wikipedia: White House
44Kb compressed
31 1,460 byte segments
What to do about it?
1) Carefully consider every byte of content
Minimum Round Trips To Deliver N Segments

Segments

Round trips

1  2  3  4  5  6  7  8  9  10  11

3  6  12  21  33  51  78  120  164  208  250
Accept-Encoding: gzip, deflate
Tony Gentilcore:

~15% of users don’t do gzip

<table>
<thead>
<tr>
<th>Software</th>
<th>Accept-Encoding modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad Muncher</td>
<td>Stripped</td>
</tr>
<tr>
<td>CA Internet Security Suite</td>
<td>Accept-EncodXng: gzip, deflate</td>
</tr>
<tr>
<td>CEQURUX</td>
<td>Stripped</td>
</tr>
<tr>
<td>Citrix Application Firewall</td>
<td>Stripped</td>
</tr>
<tr>
<td>ISA 2006</td>
<td>Stripped</td>
</tr>
<tr>
<td>McAfee Internet Security 6.0</td>
<td>XXXXXXXXXXXXXXXXXXXXXXX: +++++++++++++++</td>
</tr>
<tr>
<td>Norton Internet Security 2005</td>
<td>---------------: --------:</td>
</tr>
<tr>
<td>Novell iChain 2.3</td>
<td>Stripped</td>
</tr>
<tr>
<td>Novell Client Firewall</td>
<td>Stripped</td>
</tr>
<tr>
<td>WebWasher</td>
<td>Stripped</td>
</tr>
<tr>
<td>ZoneAlarm Pro 5.5</td>
<td>XXXXXXXXXXXXXXXXXXXXXXX: XXXXXXXXXXXXXXXXXXXXXX</td>
</tr>
</tbody>
</table>

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