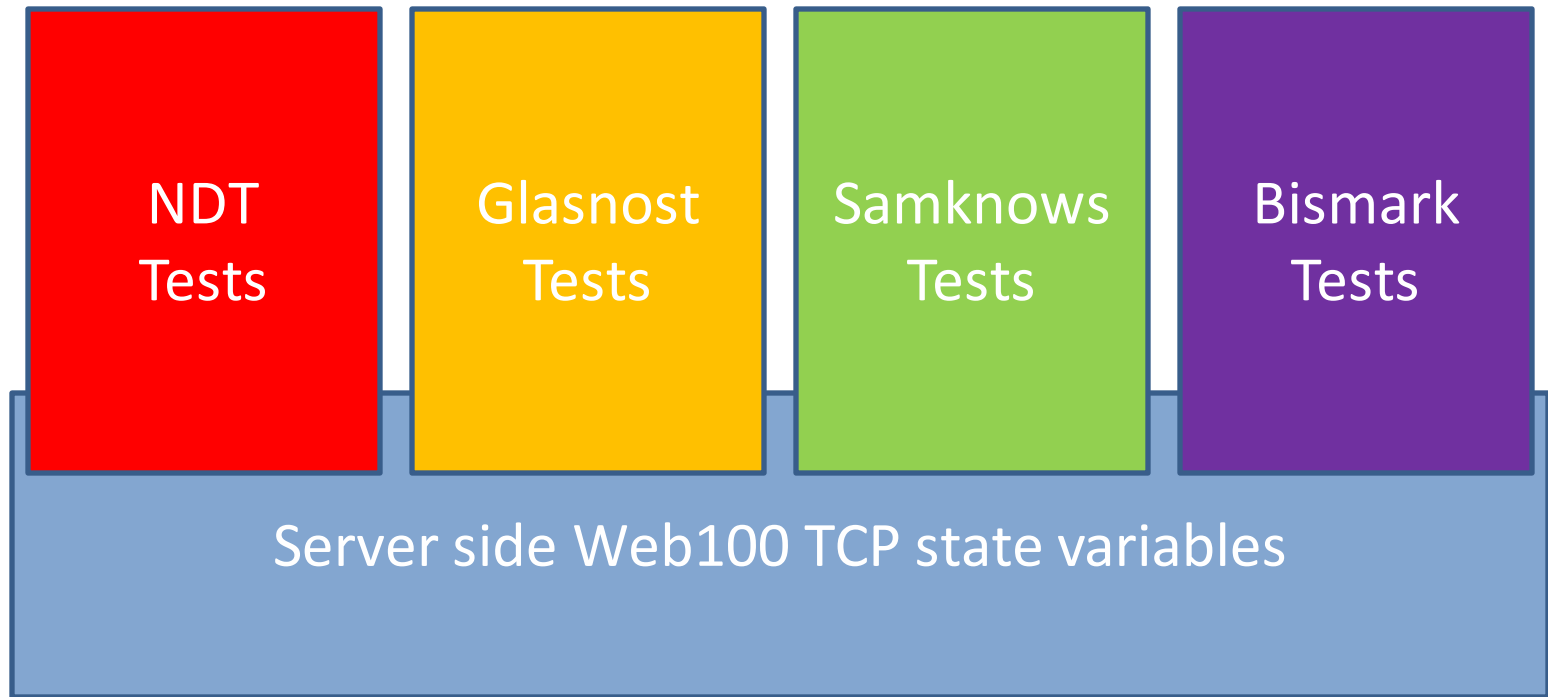


The State of TCP

Steve Bauer

MIT



NDT
Tests

The diagram consists of a wide, low blue rectangular base. Centered on this base is the text 'Server side Web100 TCP state variables'. On top of the base, there are two vertical rectangular blocks. The left block is red and contains the text 'NDT Tests'. The right block is green and contains the text 'Samknows Tests'.

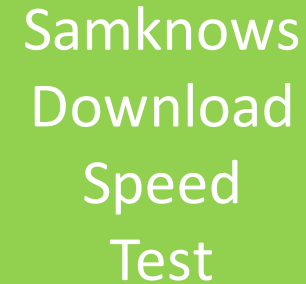
Samknows
Tests

Server side Web100 TCP state variables



NDT
Download
Speed
Test

- Single TCP connection
- 10 second duration



Samknows
Download
Speed
Test

- Three TCP connections
- 30 second duration

-
- Web100 snapshot every 5 msec
 - Packet trace

- Data logged every 5 seconds
- In the following analysis we used result from 10 second into test since that is most comparable with NDT.

**NDT
Download
Speed
Test**

March 2011

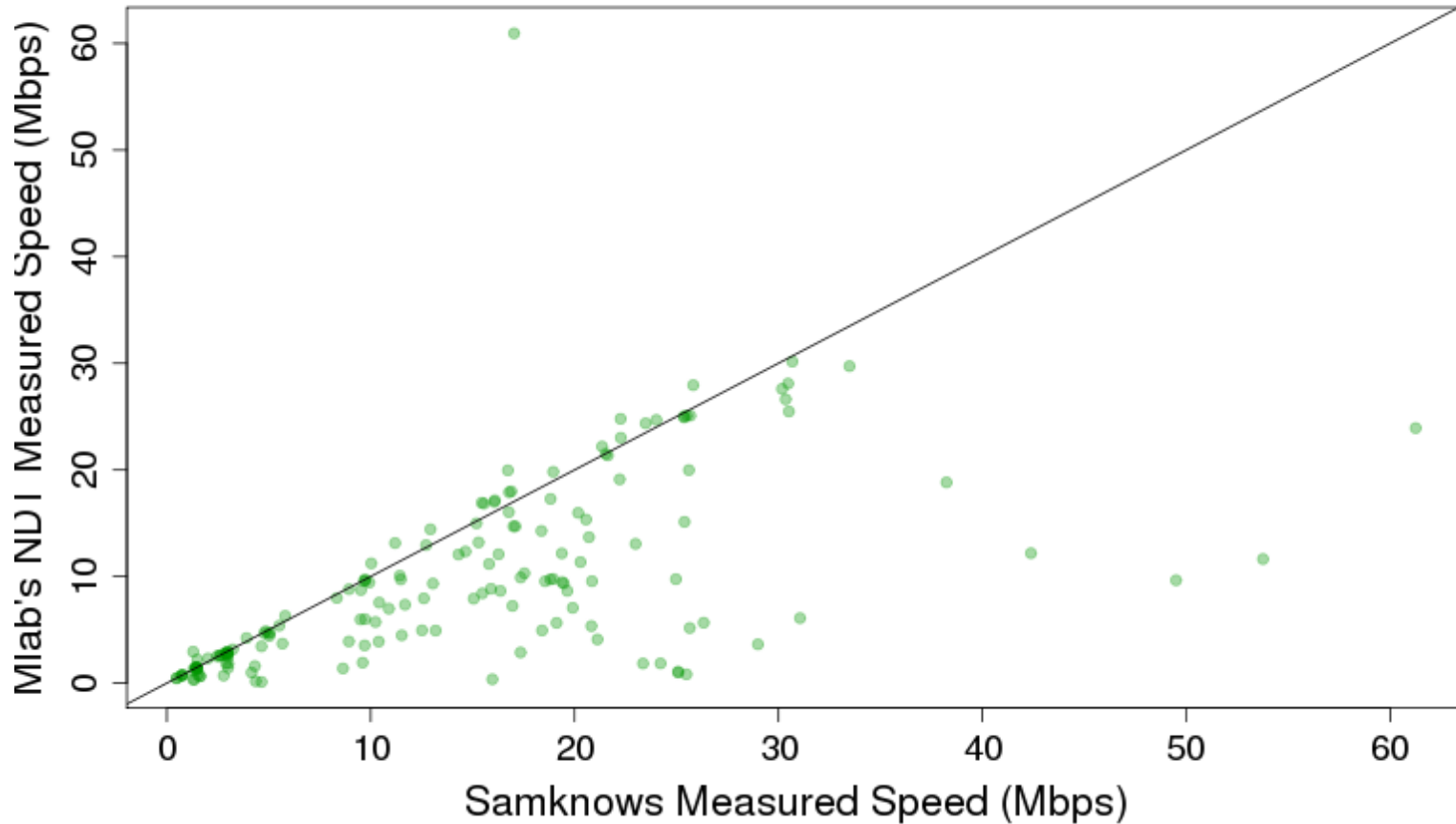
**Samknows
Download
Speed
Test**

- 4.7 million tests
- 2.5 million unique IPs
- 1.8 million tests
- 7300 units

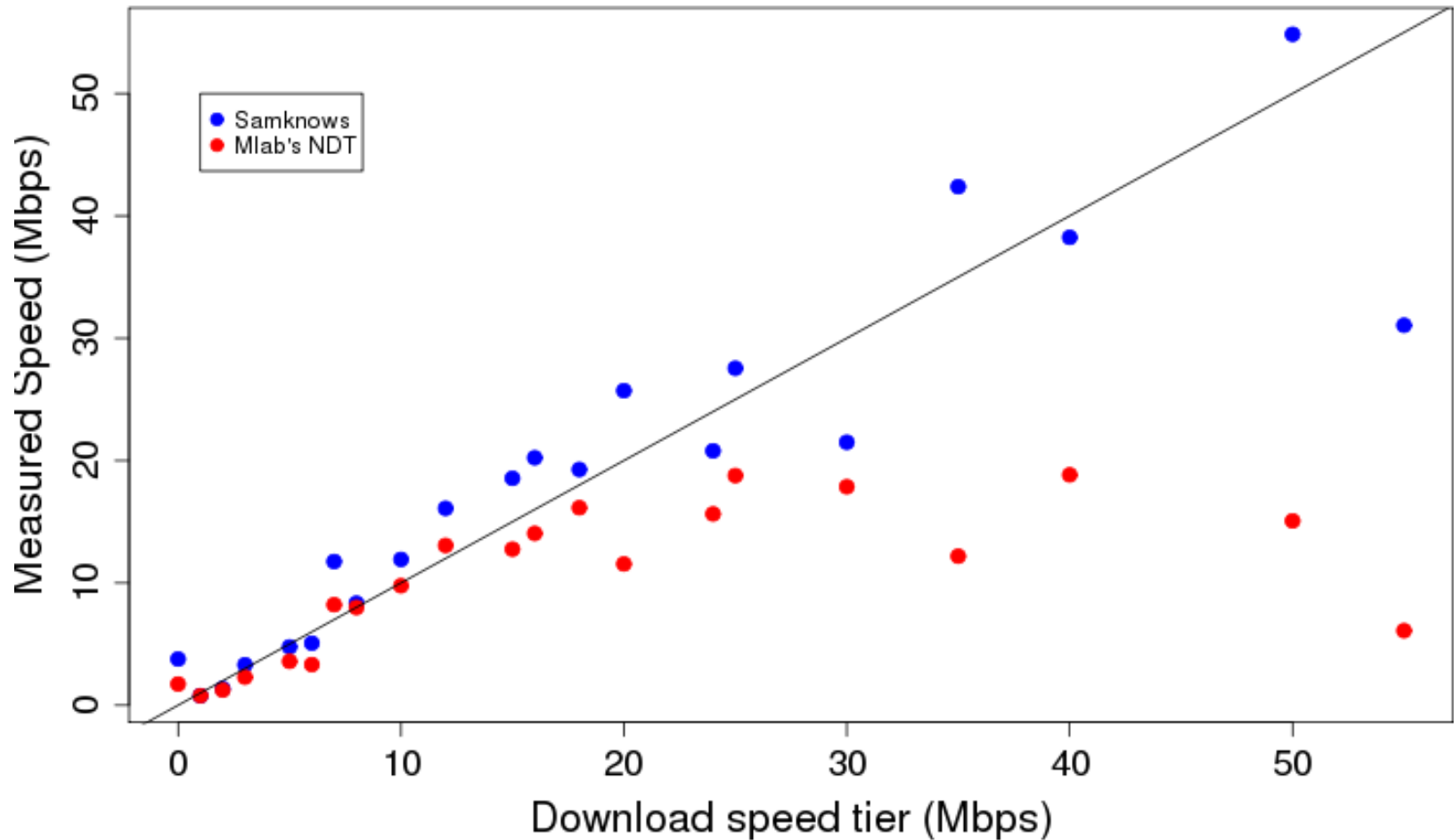
**175 IP addresses ran
both tests**

- 293 tests
- 33,500 tests

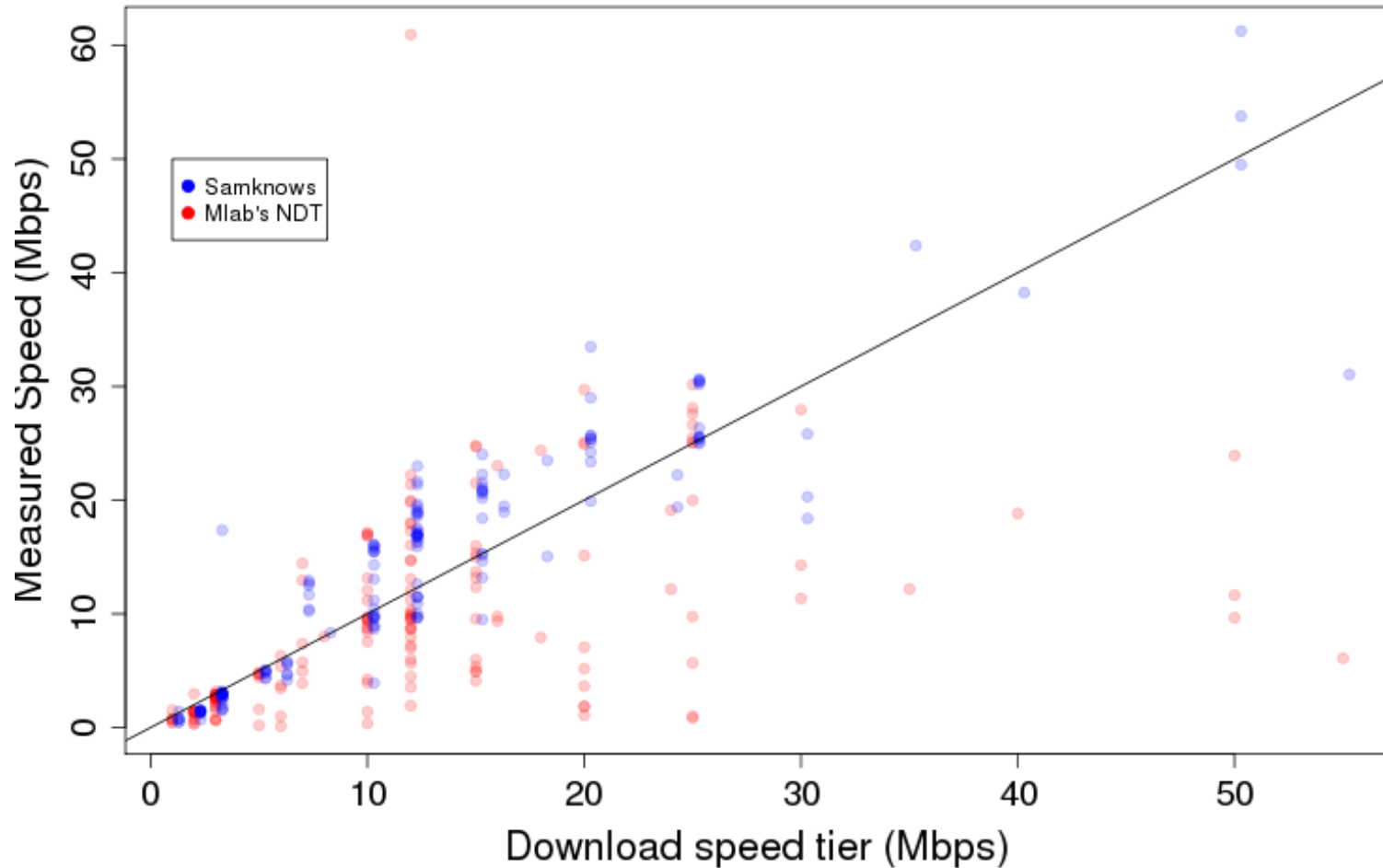
Comparison of measured speeds (March 2011)



Comparison of measured speeds (March 2011)

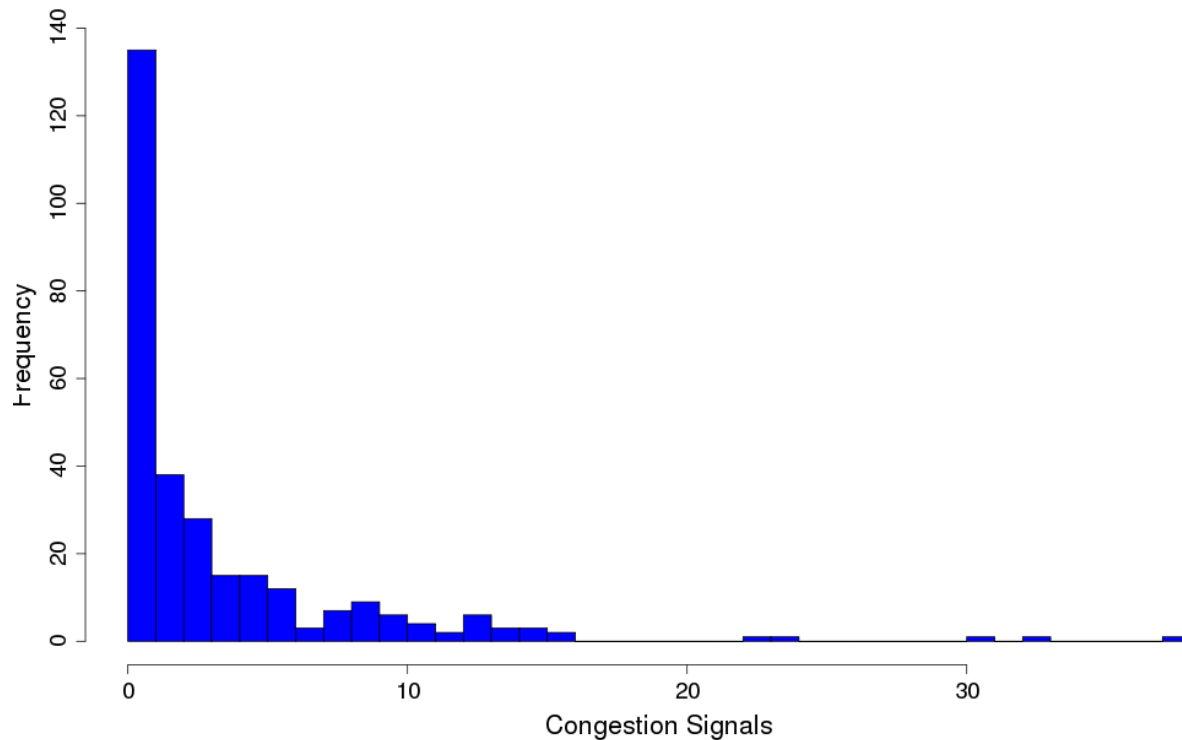


Comparison of measured speeds (March 2011)

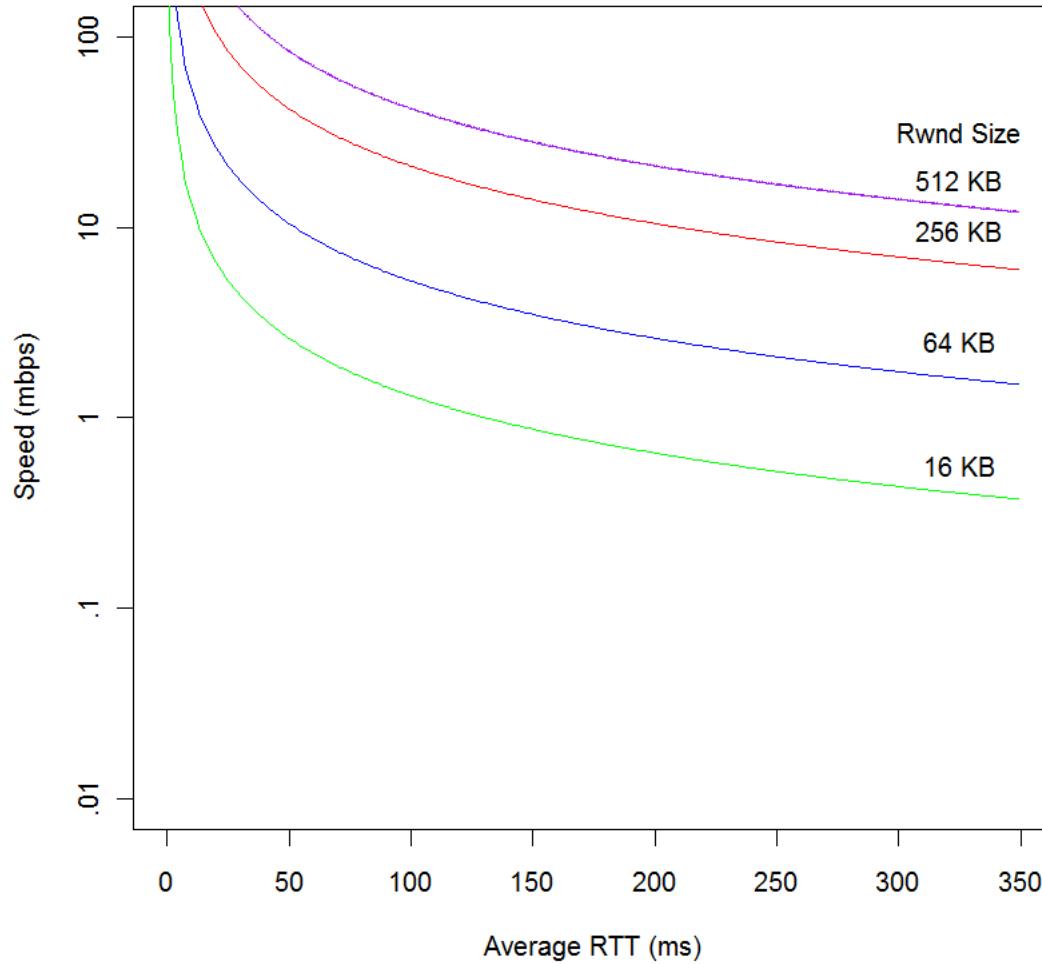


Why do these NDT tests measure slower speeds?

- Lots of potential reasons...
- 34% of tests don't fill the pipe enough to generate any congestion signal



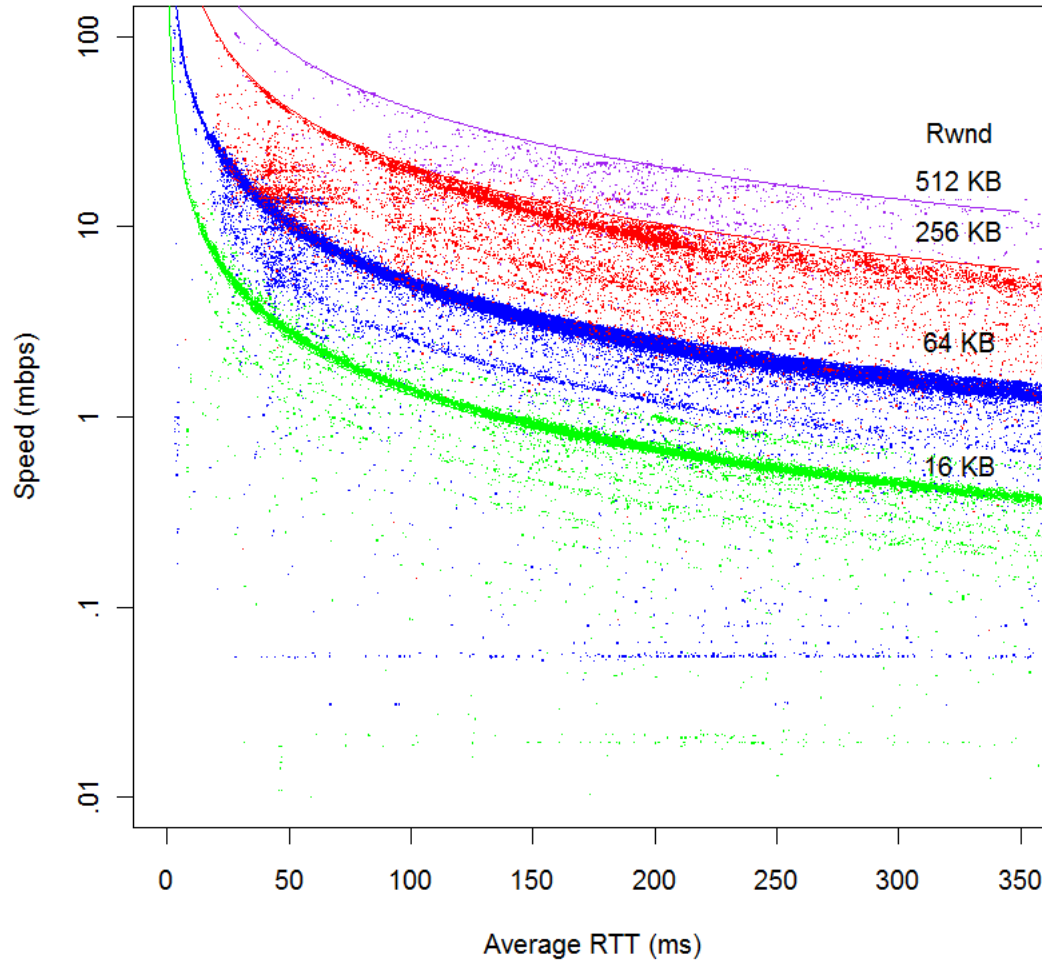
Predicted speeds if connection is receiver window limited



$$\text{Speed} = \frac{\text{Receive window}}{\text{Round trip time}}$$

Measurement Lab NDT data

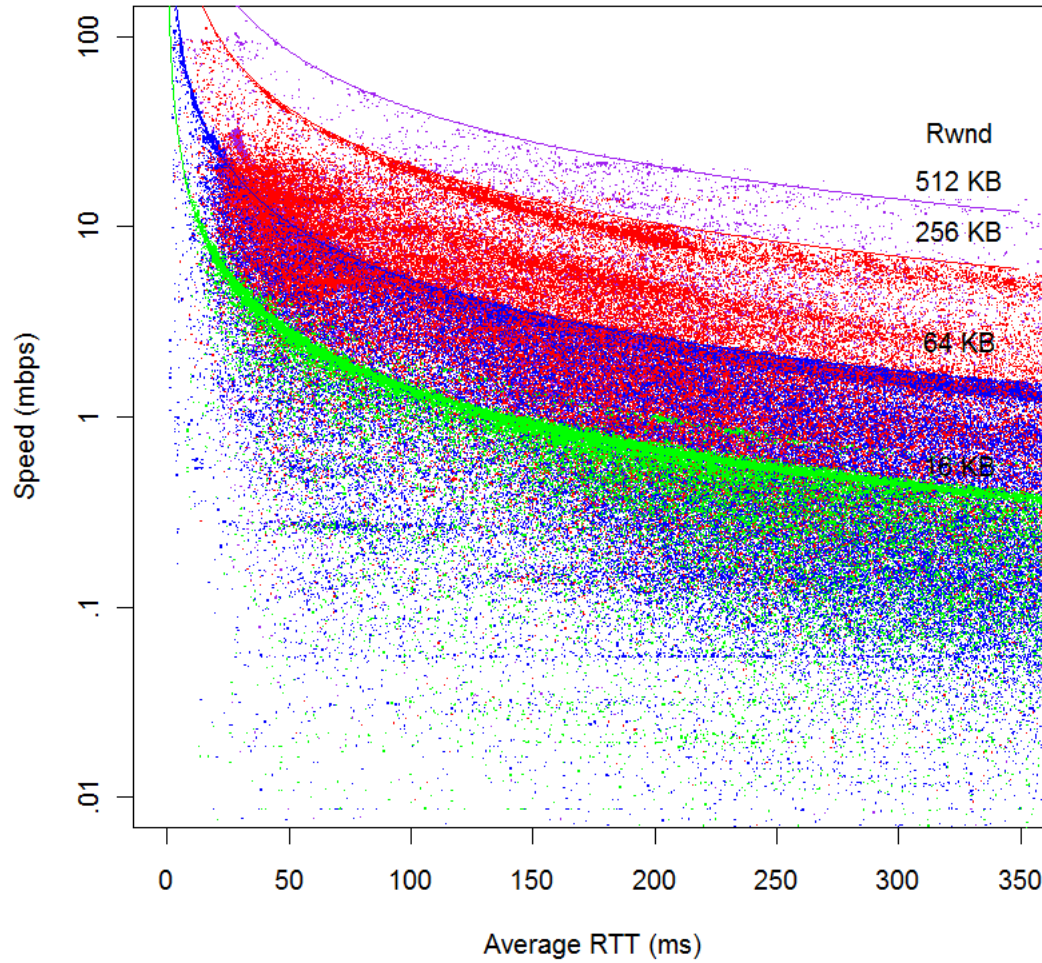
Download Speed Tests (no congestion)



$$\text{Speed} = \frac{\text{Receive window}}{\text{Round trip time}}$$

Measurement Lab NDT data

Download Speed Tests



What is the bottleneck for a TCP download on your phone?

NDT Mobile Client

- WiFi?
- Cellular?



TCP receiver window limitations are not just a legacy system problem

- “Low” memory systems
- Take away messages:
 - TCP tuning is important for valid test measurements
 - Devices and OSes will eventually will be fixed. This will result in higher but potentially more variable performance.

TCP Slow Start

Latency of transfers ending during slow start =

$$\left[\log_{\gamma} \left(\frac{s(\gamma - 1)}{init_wnd} \right) + 1 \right] * RTT + \frac{S}{C}$$

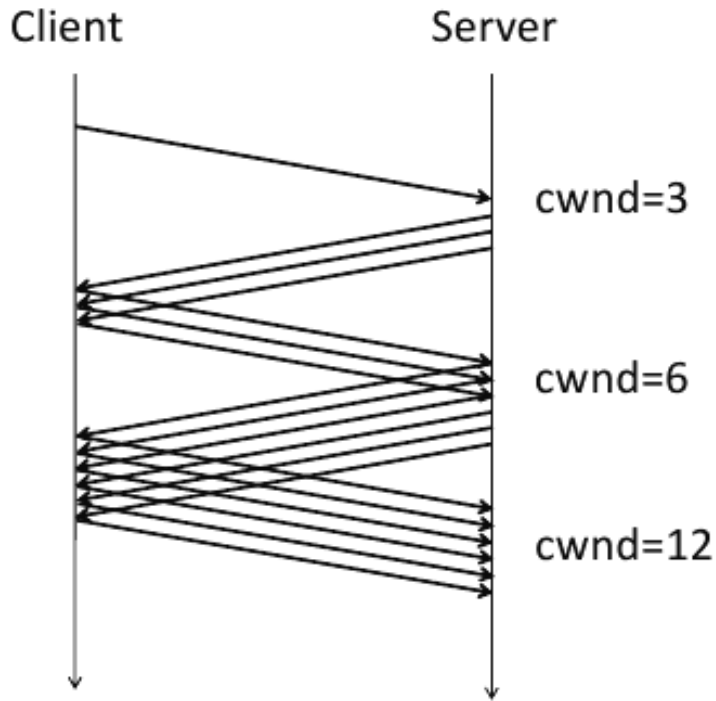
Transfer size

Capacity

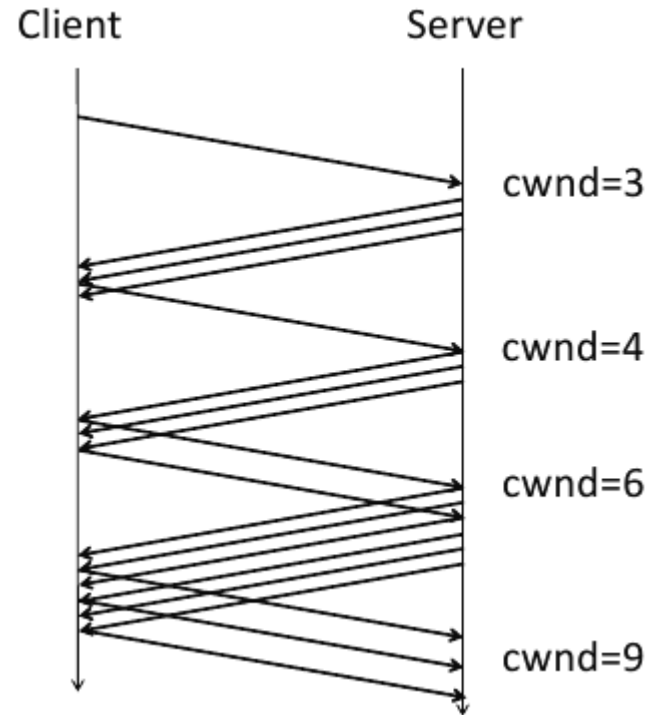
1.5 or 2 depending on whether delayed ACKs are enabled

TCP Slow Start

ACK every packet

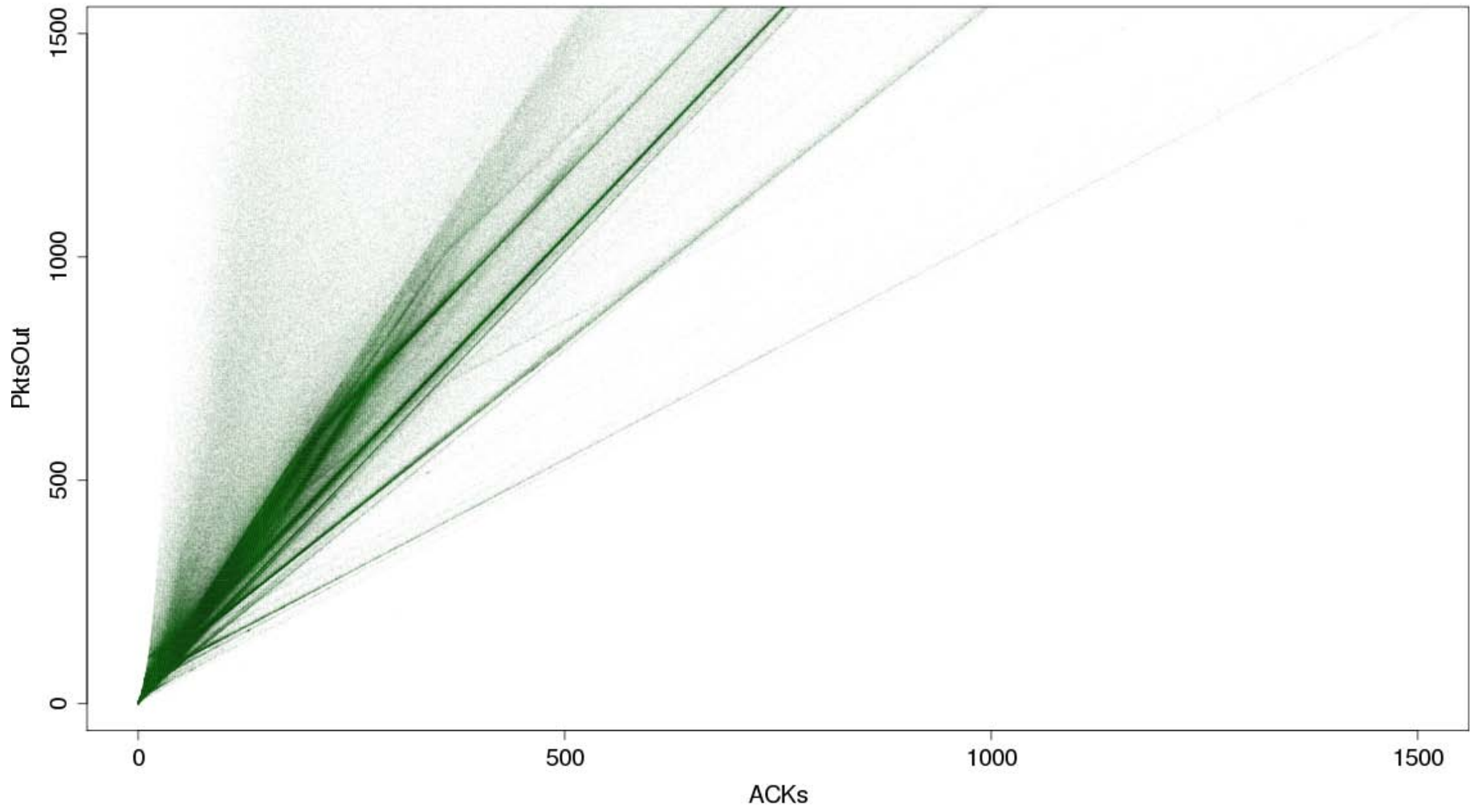


Delayed ACKs

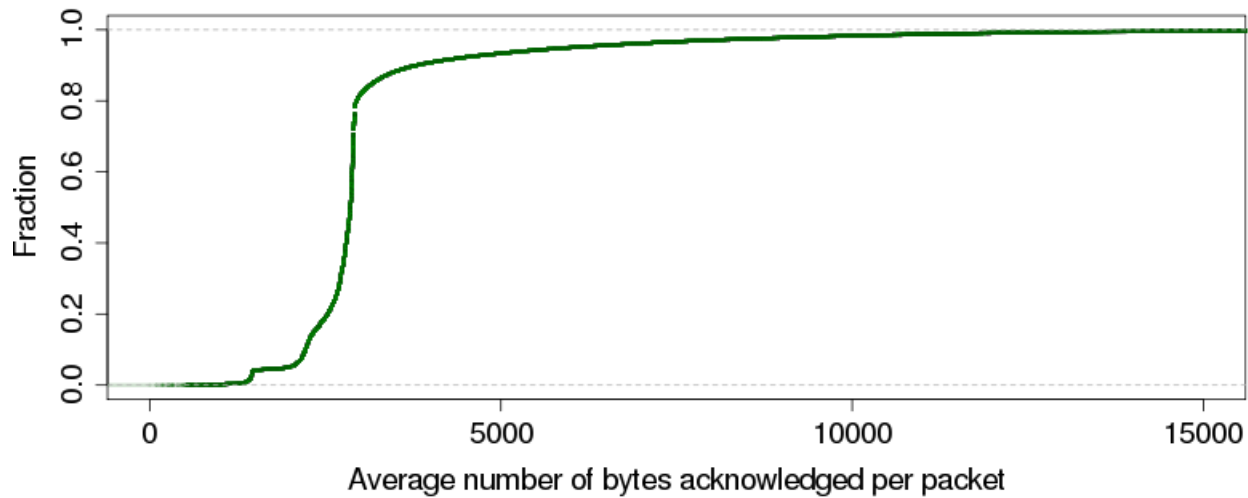


Description of data

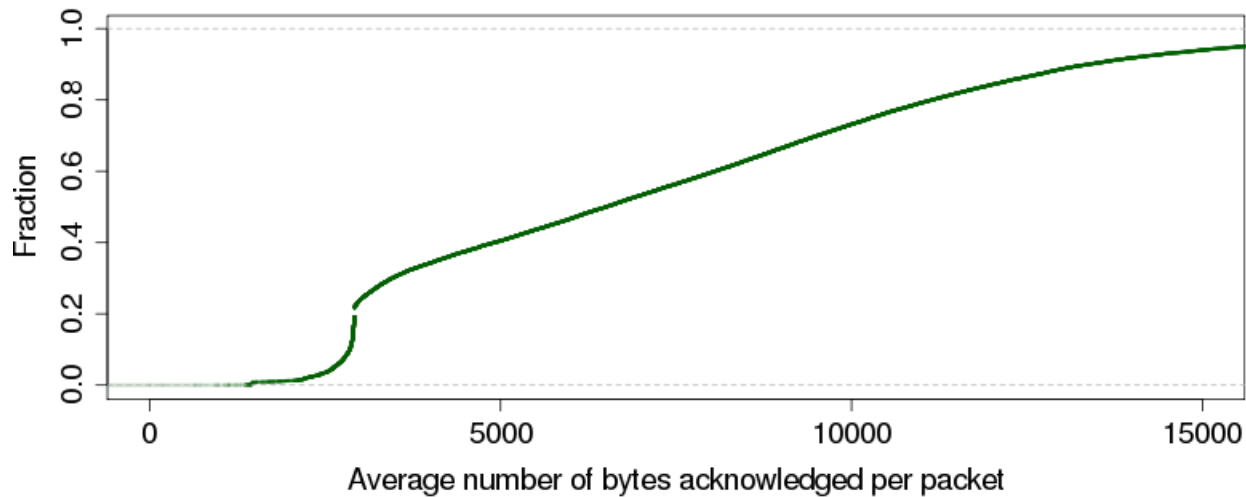
- March 2011 NDT Tests
 - 4.7 million tests
 - 2.5 million unique IPs
- Look at 3.1 million (67%) that experienced congestion
- TCP state in first web100 log line where Congestion Signals > 0



CDF of all connections



CDF of single selected provider



-50,000 connections
- Normal MSS sizes

Factors which reduce ACK stream

1. Delayed ACKs
2. Large receive offload
3. **ACK suppression**

TCP ACK Suppression

- TCP-aware link-layer technique that reduces the number of ACKs sent on the upstream link.
- When an ACK from the receiver is about to be enqueued at a upstream bottleneck link interface, the router checks the transmit queues for older ACKs belonging to the same TCP connection. If ACKs are found, some (or all of them) are removed from the queue, reducing the number of ACKs.

ACK Congestion

- RFC 3449 TCP Performance Implications of Network Path Asymmetry
- Normalized bandwidth ratio (k) is the ratio of the raw bandwidths divided by the ratio of the packet sizes used in the two directions
 - **if the receiver acknowledges more frequently than one ACK every k data packets, the reverse bottleneck link will get saturated before the forward bottleneck link does, limiting the throughput in the forward direction.**

Mitigating server side factor

- RFC 3465: TCP Congestion Control with Appropriate Byte Counting (ABC)
 - Limits byte counting to two x maximum segment size
 - Not generally turned on by default
 - Some large companies do enable it
 - Some large companies have patched TCP ABC to remove two x maximum segment size limit

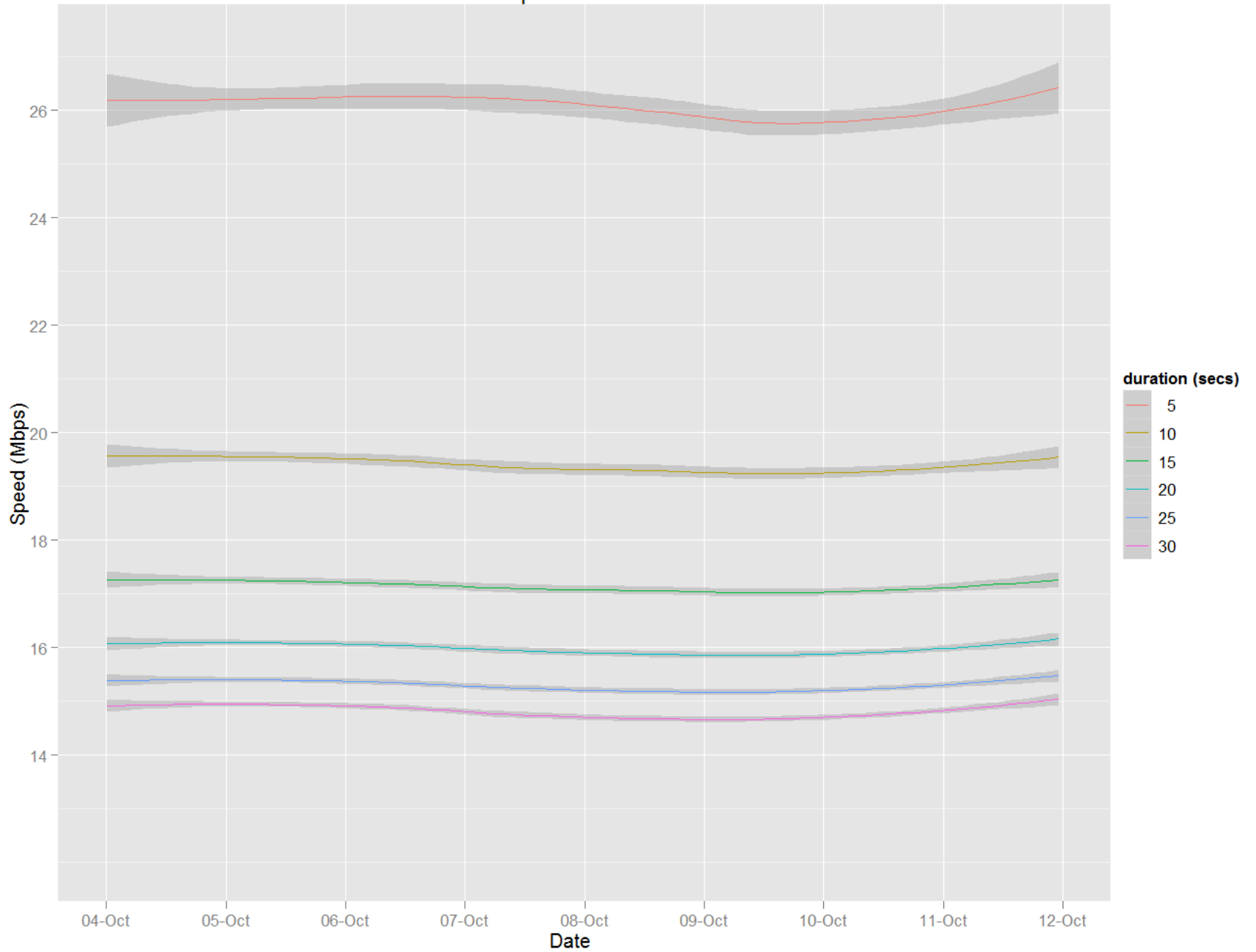
Conclusions

1. Inform regulatory processes
 2. Suggests research directions
 3. Suggests possible changes to IETF standards
 4. Improve networking stacks
- Samknows/NDT comparison
 - ACK suppression
 - Rwin limitations are **not** a legacy system only problem
 - Update RFC 3465 to remove two x maximum segment size limit
 - Lots more in the works...

Download speeds (~10 second tests)



Samknows download speeds for different test durations



Samknobs download speeds for different test durations

