

HPC Networking Performance

What's TCP got to do with it

Utah

- up041{tjpratt}21: traceroute www.utah.edu
- trying to get source for www.utah.edu
- source should be 134.9.39.61
- traceroute to www.utah.edu (155.97.155.10) from 134.9.39.61 (134.9.39.61), 30 hops max
- outgoing MTU = 1500
- 1 fpcf39-rtr.llnl.gov (134.9.39.254) 1 ms 0 ms 0 ms
- 2 lcdmzsw1-522.llnl.gov (134.9.5.22) 0 ms 0 ms 0 ms
- 3 bbrtr0-453-rtr.llnl.gov (134.9.3.254) 1 ms 1 ms 1 ms
- 4 foutnet0-rtr.llnl.gov (128.115.190.17) 16 ms 17 ms 8 ms
- 5 esnet0-rtr.llnl.gov (128.115.190.1) 9 ms 11 ms 16 ms
- 6 snllmr1-llnlrt1.es.net (134.55.219.21) 178 ms 200 ms 236 ms
- 7 snv1mr1-snllmr1.es.net (134.55.217.18) 3 ms 3 ms 3 ms
- 8 snvcr1-snv1mr1.es.net (134.55.218.21) 2 ms 3 ms 2 ms
- 9 abilene-pos-snvcr1.es.net (198.129.248.85) 3 ms 4 ms 4 ms
- 10 dnvrng-snvang.abilene.ucaid.edu (198.32.8.2) 30 ms 28 ms 50 ms
- 11 205.124.237.9 (205.124.237.9) 37 ms 37 ms 37 ms
- 12 205.124.244.113 (205.124.244.113) 65 ms 63 ms 62 ms
- 13 205.124.244.2 (205.124.244.2) 67 ms 62 ms 65 ms
- 14 205.124.249.118 (205.124.249.118) 74 ms 64 ms 63 ms
- 15 crpark-wr1ebc.net.utah.edu (155.99.132.102) 67 ms 64 ms 80 ms
- 16 ddrebc-crpark.net.utah.edu (155.99.132.121) 91 ms 90 ms 78 ms
- 17 www.utah.edu (155.97.155.10) 72 ms 66 ms 63 ms

UIUC

- up041{tjpratt}22: traceroute www.uiuc.edu
- trying to get source for www.uiuc.edu
- source should be 134.9.39.61
- traceroute to swan29.admin.uiuc.edu (128.174.254.29) from 134.9.39.61 (134.9.39.61), 30 hops max
- outgoing MTU = 1500
- 1 fpcf39-rtr.llnl.gov (134.9.39.254) 1 ms 1 ms 1 ms
- 2 lcdmzsw1-522.llnl.gov (134.9.5.22) 0 ms 0 ms 0 ms
- 3 bbrtr0-453-rtr.llnl.gov (134.9.3.254) 1 ms 1 ms 1 ms
- 4 foutnet0-rtr.llnl.gov (128.115.190.17) 7 ms 17 ms 5 ms
- 5 esnet0-rtr.llnl.gov (128.115.190.1) 1 ms 7 ms 4 ms
- 6 snllmr1-llnlrt1.es.net (134.55.219.21) 1 ms 3 ms 2 ms
- 7 snv1mr1-snllmr1.es.net (134.55.217.18) 6 ms 2 ms 8 ms
- 8 snv2mr1-snv1mr1.es.net (134.55.217.6) 3 ms 24 ms 8 ms
- 9 snv2sdn1-snv2mr1.es.net (134.55.207.37) 5 ms 3 ms 2 ms
- 10 chicr1-oc192-snv2sdn1.es.net (134.55.209.54) 52 ms 51 ms 56 ms
- 11 chislmr1-10ge-chicr1.es.net (134.55.217.54) 92 ms 74 ms 72 ms
- 12 198.125.140.94 (198.125.140.94) 116 ms 78 ms 78 ms
- 13 198.125.140.114 (198.125.140.114) 98 ms 84 ms 79 ms
- 14 t-exiteo.gw.uiuc.edu (130.126.0.65) 78 ms 83 ms 79 ms

UChicago

- traceroute www.uchicago.edu
- trying to get source for www.uchicago.edu
- source should be 134.9.39.61
- traceroute to krypton.uchicago.edu (128.135.12.53) from 134.9.39.61 (134.9.39.61), 30 hops max
- outgoing MTU = 1500
- 1 fpcf39-rtr.llnl.gov (134.9.39.254) 1 ms 1 ms 1 ms
- 2 lcdmzsw1-522.llnl.gov (134.9.5.22) 0 ms 0 ms 0 ms
- 3 bbrtr0-453-rtr.llnl.gov (134.9.3.254) 1 ms 1 ms 1 ms
- 4 foutnet0-rtr.llnl.gov (128.115.190.17) 1 ms 1 ms 1 ms
- 5 esnet0-rtr.llnl.gov (128.115.190.1) 3 ms 2 ms 1 ms
- 6 snllmr1-llnlrt1.es.net (134.55.219.21) 38 ms 2 ms 1 ms
- 7 snv1mr1-snllmr1.es.net (134.55.217.18) 2 ms 2 ms 3 ms
- 8 snv2mr1-snv1mr1.es.net (134.55.217.6) 112 ms 8 ms 3 ms
- 9 snv2sdn1-snv2mr1.es.net (134.55.207.37) 2 ms 3 ms 2 ms
- 10 chicr1-oc192-snv2sdn1.es.net (134.55.209.54) 50 ms 53 ms 51 ms
- 11 chislmr1-10ge-chicr1.es.net (134.55.217.54) 51 ms 50 ms 51 ms
- 12 198.125.140.94 (198.125.140.94) 61 ms 52 ms 54 ms
- 13 MREN-IWIRE-10G-local.uchicago.edu (128.135.247.121) 51 ms 51 ms 51 ms

What causes slowness

- Application Buffering
- OS Copies
- File System Speeds/contention
- TCP Default Settings being used
- Local Network Congestion/Competition
- Wide Area Network Congestion/ TCP congestion window Collapse
- IP Data Errors

Local Performance is good, WAN Performance Is bad

- Problem could be local
 - Small TCP window size
 - Application doesn't provide enough data to fill a longer network pipe
 - Application is too transaction based
 - Local congestion causing dropped packets
- Verify Packet drops are occurring
 - Netstat utility provides Retransmission data
 - ICMP PING
 - RTT metric
 - Use large packets, Co-located packet more likely to be discarded
 - Ethereal/snoop can show the actual packet data
 - No packet drops means network is not your problem

How big to set the TCP Window speed

- Common thinking
 - $RTT / (1 / \text{Network speed})$
 - $60\text{ms} / (1 / 1\text{gigabit/sec}) / 8\text{bit} = 7.5\text{MB}$
- Local Example
 - $1\text{ms} / (1 / \text{gigabit/sec}) / 8\text{bit} = 125,000\text{B}$
- Large Buffered Switches(NEW)
 - Large Buffer Factor(LBS)
 - 1.5 to 4
 - $2 * 60\text{ms} / (1 / 1\text{gigabit/sec}) / 8\text{bit} = 15\text{MB}$
- Multiple TCP streams strategy produces better results during congestion