

Network Measurement and Monitoring Activities

OSG Site-admin meeting
Aug 6-7, 2009
Rich Carlson
rcarlson@internet2.edu

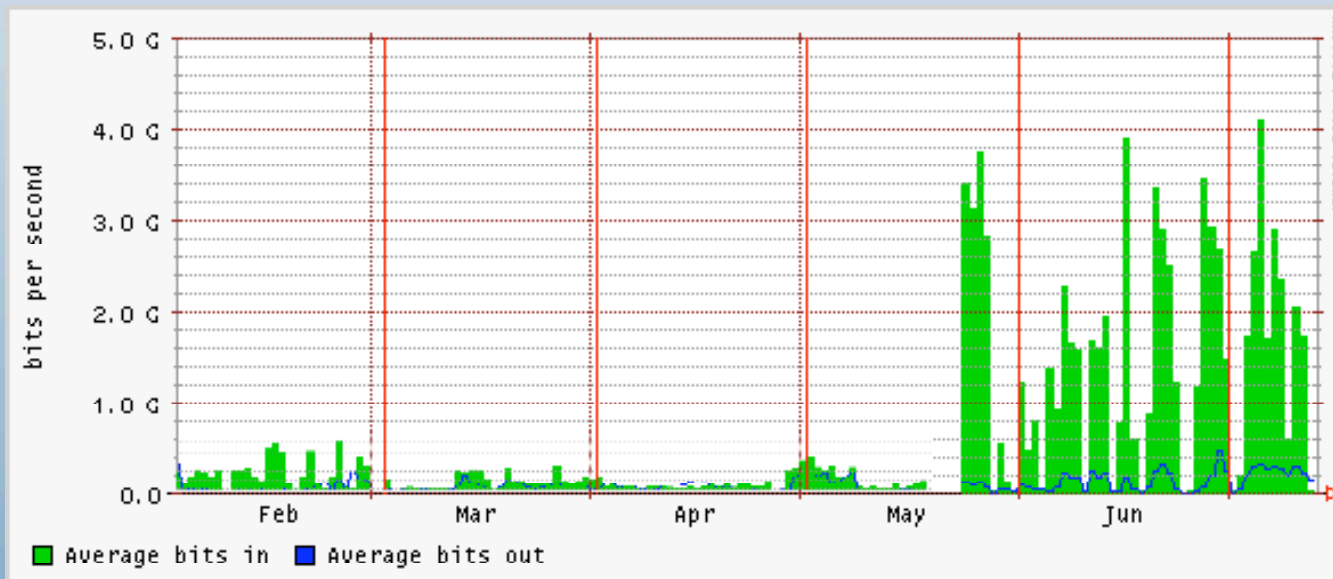


Network Management 101

- There are 3 major functions that every OSG site-admin needs
 - Monitoring – verifying the health and operation of the internal network
 - Measuring – verifying the capacity and utilization of the peer networks
 - Advanced diagnostics – identifying what has failed when a fault is found

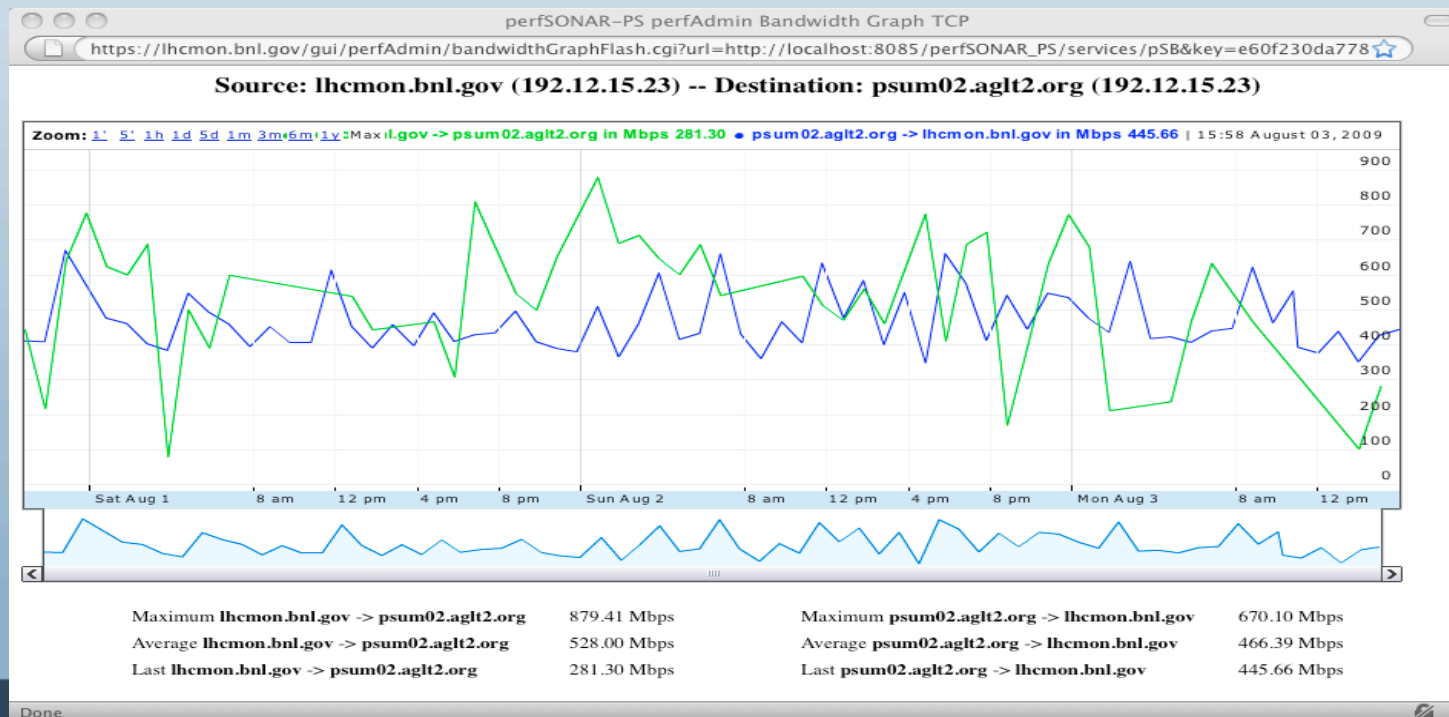
Network Monitoring

- Collect statistics from routers and switches
- Display results in graphical form



Network Measurement

- Verify performance levels between peer sites



Network Diagnostic

- Collect/Analyze data to respond to faults
- Detailed results can point to real problems

PNNL sending to:	Seattle	Salt Lake City	Washington
Throughput	966 Mbps	930 Mbps	328 Mbps
Min RTT	6 msec	23 msec	76 msec
Loss	0.0094%	0.0045%	.0049%
Reordering	6.04%	5.5%	5.15%

perfSONAR Overview

- A set of high level services for managing multi-domain measurement/monitoring infrastructures
- International community of developers
 - Implementing Open Grid Forum (OGF) Network Measurement (NM-WG) recommendations
- Multiple sets of interoperable software

perfSONARized tools

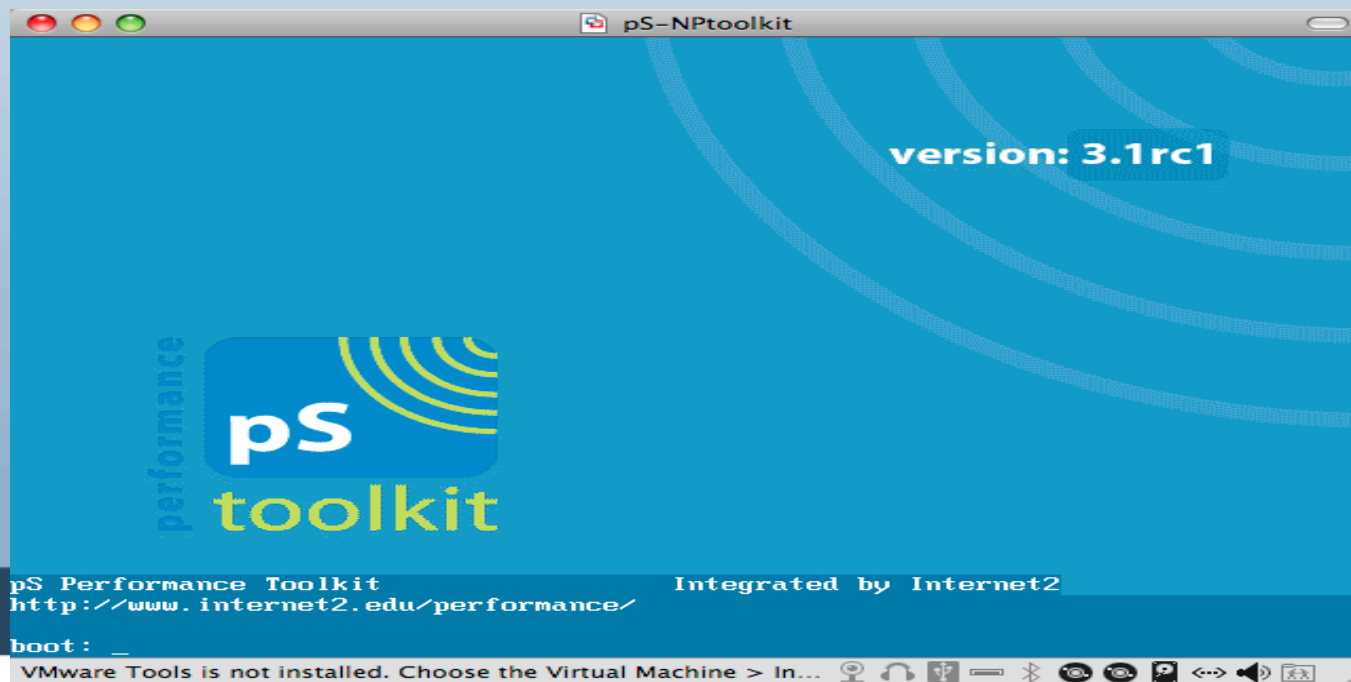
- Implement 1 or more perfSONAR services
 - Registration, Data Storage, ...
- Existing tools include
 - Cacti
 - BWCTL
 - NDT
 - NPAD
 - OWAMP
 - PingER

Measurement/Monitoring Tools

- Client / Server based tools
 - Measure throughput
 - Measure delay & loss
 - Analyze infrastructure, switch/router configuration, and host configuration

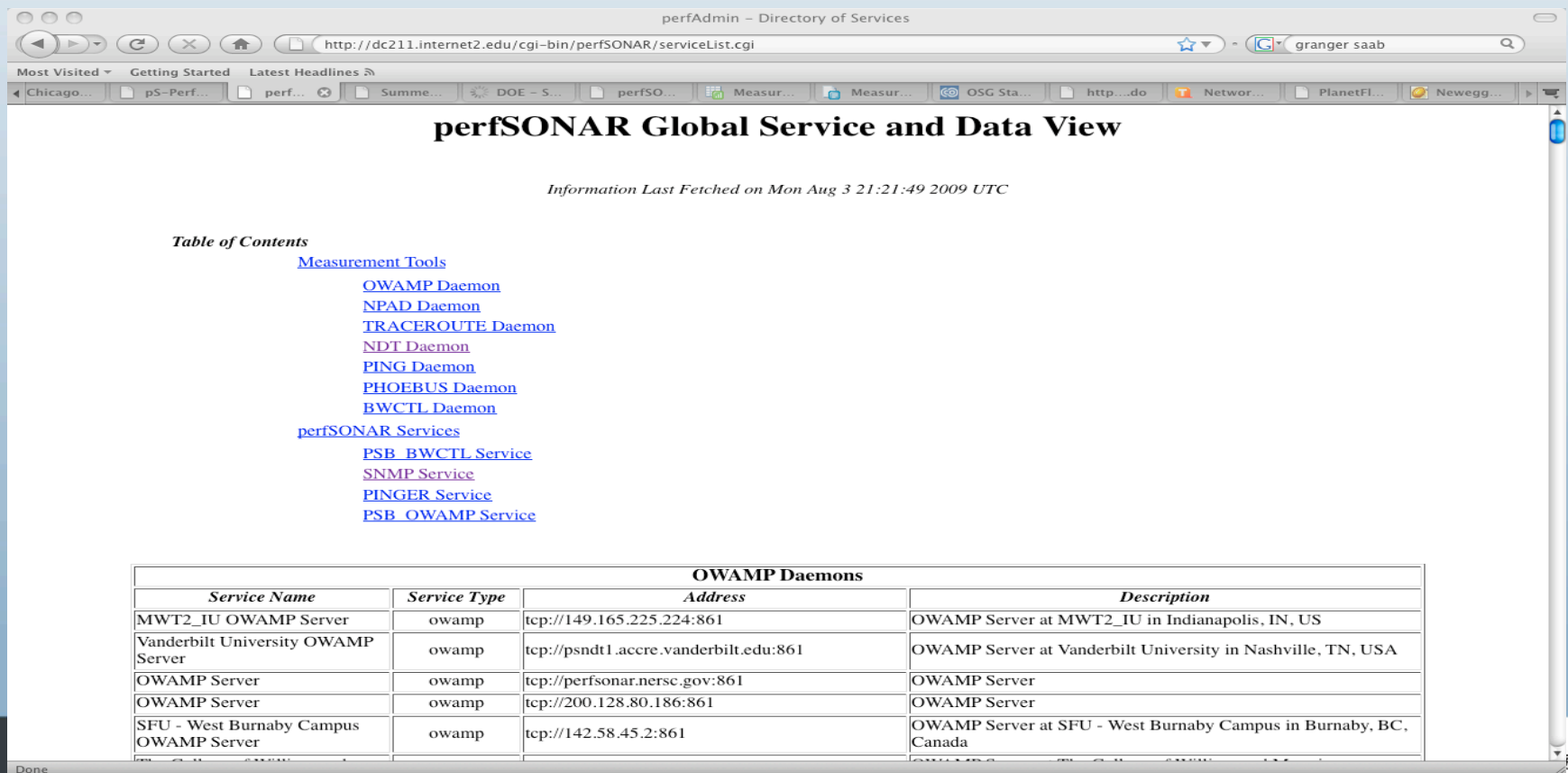
Server Deployment options

- Source tar files – build your own
- OS Distribution packages (yum, apt)
- Bootable CD



Finding a deployed server

- <http://www.perfsonar.net/activeServices>



perfAdmin - Directory of Services

http://dc211.internet2.edu/cgi-bin/perfSONAR/serviceList.cgi

granger saab

perfSONAR Global Service and Data View

Information Last Fetched on Mon Aug 3 21:21:49 2009 UTC

Table of Contents

- [Measurement Tools](#)
 - [OWAMP Daemon](#)
 - [NPAD Daemon](#)
 - [TRACEROUTE Daemon](#)
 - [NDT Daemon](#)
 - [PING Daemon](#)
 - [PHOEBUS Daemon](#)
 - [BWCTL Daemon](#)
- [perfSONAR Services](#)
 - [PSB_BWCTL Service](#)
 - [SNMP Service](#)
 - [PINGER Service](#)
 - [PSB_OWAMP Service](#)

Service Name	Service Type	Address	Description
MWT2_IU OWAMP Server	owamp	tcp://149.165.225.224:861	OWAMP Server at MWT2_IU in Indianapolis, IN, US
Vanderbilt University OWAMP Server	owamp	tcp://psndt1.accre.vanderbilt.edu:861	OWAMP Server at Vanderbilt University in Nashville, TN, USA
OWAMP Server	owamp	tcp://perfsonar.nersc.gov:861	OWAMP Server
OWAMP Server	owamp	tcp://200.128.80.186:861	OWAMP Server
SFU - West Burnaby Campus OWAMP Server	owamp	tcp://142.58.45.2:861	OWAMP Server at SFU - West Burnaby Campus in Burnaby, BC, Canada

Client Deployment options

- Source tar files – build your own
- OS distro packages
- Java based clients using web browser
- Client executables distributed via VDT
 - Allows testing from any CE/SE
 - Command syntax and output will be covered later in this talk

Monitoring Tools

- BWCTL – Bandwidth Control
 - Allows single person operation over wide area testing environment
 - Runs NLANR ‘iperf’ program
- OWAMP – One way Delay Measurement
 - Advanced ‘ping’ command
 - Allows single person operation over wide area testing environment

Sample BWCTL results

```
rcarlson@triton:~/ndt — ssh — 116x33
-bash-3.2$ bwctl -i2 -t20 -s lhcmon.bnl.gov
bwctl: Unable to contact a local bwctld: Spawning local tool controller
bwctl: NuttcpAvailable(): We were unable to verify that nuttcp is working. Likely you do not have it installed. exit
status: 1: output: exec(nuttcp): No such file or directory
bwctl: Couldn't initialize tool "nuttcp". Disabling it.
bwctl: Using tool: iperf
bwctl: 28 seconds until test results available

RECEIVER START
bwctl: exec_line: iperf -B 207.75.164.104 -s -f a -m -p 5001 -t 20 -i 2
bwctl: start_tool: 3458324284.538605
-----
Server listening on TCP port 5001
Binding to local address 207.75.164.104
TCP window size: 85.3 KByte (default)
-----
[ 12] local 207.75.164.104 port 5001 connected with 192.12.15.23 port 5001
[ 12] 0.0- 2.0 sec 19.5 MBytes 81.8 Mbits/sec
[ 12] 2.0- 4.0 sec 22.4 MBytes 94.0 Mbits/sec
[ 12] 4.0- 6.0 sec 22.4 MBytes 94.0 Mbits/sec
[ 12] 6.0- 8.0 sec 22.4 MBytes 94.0 Mbits/sec
[ 12] 8.0-10.0 sec 21.4 MBytes 89.7 Mbits/sec
[ 12] 10.0-12.0 sec 19.0 MBytes 79.8 Mbits/sec
[ 12] 12.0-14.0 sec 21.6 MBytes 90.6 Mbits/sec
[ 12] 14.0-16.0 sec 22.4 MBytes 94.1 Mbits/sec
[ 12] 16.0-18.0 sec 19.6 MBytes 82.4 Mbits/sec
[ 12] 18.0-20.0 sec 20.2 MBytes 84.6 Mbits/sec
[ 12] 0.0-20.1 sec 213 MBytes 88.5 Mbits/sec
[ 12] MSS size 1448 bytes (MTU 1500 bytes, ethernet)
bwctl: stop_exec: 3458324308.780934

RECEIVER END
-bash-3.2$
```

Using BWCTL: commands

```
bwctl -i2 -t20 -c bwctl.losa.net.internet2.edu
```

```
bwctl -i2 -t20 -s bwctl.newy.net.internet2.edu
```

- bwctl = name of program
- i2 = report intermediate results every 2 seconds
- t20 – run test for 20 seconds
- s name = remote end will send data to you
- c name = you will send data to the remote host

3rd party testing: command

```
bwctl -i2 -t20 -c bwctl.salt.net.internet2.edu -s bwctl.atla.net.internet2.edu
```

- User can run a test between 2 remote hosts

OWping Results

```
rcarlson@nms-rlat:~ — ssh — 116x34
[rcarlson@nms-rlat ~]$ ping -c3 owamp.losa.net.internet2.edu
PING eth-1.nms-rlat.losa.net.internet2.edu (64.57.17.162) 56(84) bytes of data.
64 bytes from nms-rlat.losa.net.internet2.edu (64.57.17.162): icmp_seq=1 ttl=60 time=56.6 ms
64 bytes from nms-rlat.losa.net.internet2.edu (64.57.17.162): icmp_seq=2 ttl=60 time=56.7 ms
64 bytes from nms-rlat.losa.net.internet2.edu (64.57.17.162): icmp_seq=3 ttl=60 time=56.6 ms

--- eth-1.nms-rlat.losa.net.internet2.edu ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2005ms
rtt min/avg/max/mdev = 56.692/56.697/56.705/0.275 ms
[rcarlson@nms-rlat ~]$ owping owamp.losa.net.internet2.edu
Approximately 13.2 seconds until results available

--- owping statistics from [2001:468:2:12::17:34]:35621 to [owamp.losa.net.internet2.edu]:55664 ---
SID: 00170162ce21d9957fb85e80c7f6958a
first: 2009-08-03T21:48:06.786
last: 2009-08-03T21:48:15.918
100 sent, 0 lost (0.000%), 0 duplicates
one-way delay min/median/max = 28.1/28.2/28.3 ms, (err=3.13 ms)
one-way jitter = 0.1 ms (P95-P50)
TTL not reported
no reordering

--- owping statistics from [owamp.losa.net.internet2.edu]:34879 to [2001:468:2:12::17:34]:35622 ---
SID: 00170034ce21d9959d6ffclld84cfeac
first: 2009-08-03T21:48:06.907
last: 2009-08-03T21:48:17.704
100 sent, 0 lost (0.000%), 0 duplicates
one-way delay min/median/max = 28.5/28.6/28.6 ms, (err=3.13 ms)
one-way jitter = 0 ms (P95-P50)
TTL not reported
no reordering

[rcarlson@nms-rlat ~]$
```


Using OWPING

- `owping owamp.losa.net.internet2.edu`
 - `owping` = program name
 - `owamp...` = name of server
- Output results
 - Separate statistics for both directions
 - Number of packets sent and lost
 - One-way delay statistics min/median/max
 - Number of IP hops in path
 - Number of packets that arrives out-of-order

Advanced Diagnostic tools

- Existing NDT tool
 - Allows users to test network path for a limited number of common problems
- Existing NPAD tool
 - Allows users to test local network infrastructure while simulating a long path

Network Diagnostic Tool (NDT)

- Measure performance to users desktop
- Identify real problems for real users
 - Network infrastructure is the problem
 - Host tuning issues are the problem
- Make tool simple to use and understand
- Make tool useful for users and network administrators

NDT Client test

```
rcarlson@nms-rexp:~ — ssh — 116x35
[rcarlson@nms-rexp ~]$ web100clt -4 -n ndt.kans.net.internet2.edu -l
Testing network path for configuration and performance problems -- Using IPv4 address
Checking for Middleboxes . . . . . Done
checking for firewalls . . . . . Done
running 10s outbound test (client to server) . . . . . 983.18 Mb/s
running 10s inbound test (server to client) . . . . . 977.73 Mb/s
The slowest link in the end-to-end path is a 1.0 Gbps Gigabit Ethernet subnet
Information: Other network traffic is congesting the link
Server 'ndt.kans.net.internet2.edu' is not behind a firewall. [Connection to the ephemeral port was successful]
Client is not behind a firewall. [Connection to the ephemeral port was successful]

----- Web100 Detailed Analysis -----

Web100 reports the Round trip time = 47.32 msec; the Packet size = 8948 Bytes; and
No packet loss was observed.
This connection is receiver limited 1.63% of the time.
This connection is network limited 98.08% of the time.

Web100 reports TCP negotiated the optional Performance Settings to:
RFC 2018 Selective Acknowledgment: ON
RFC 896 Nagle Algorithm: ON
RFC 3168 Explicit Congestion Notification: OFF
RFC 1323 Time Stamping: ON
RFC 1323 Window Scaling: ON; Scaling Factors - Server=10, Client=10
The theoretical network limit is 154.82 Mbps
The NDT server has a 32768 KByte buffer which limits the throughput to 5410.20 Mbps
Your PC/Workstation has a 16673 KByte buffer which limits the throughput to 2752.82 Mbps
The network based flow control limits the throughput to 1732.74 Mbps

Client Data reports link is ' 7', Client Acks report link is ' 7'
Server Data reports link is ' 7', Server Acks report link is ' 7'
Packet size is preserved End-to-End
Server IP addresses are preserved End-to-End
Client IP addresses are preserved End-to-End
[rcarlson@nms-rexp ~]$
```



Finding Results of Interest

- Duplex Mismatch
 - This is a serious error and nothing will work right. Reported on *main* page, on *Statistics* page, and **mismatch:** on *More Details* page
- Packet Arrival Order
 - Inferred value based on TCP operation. Reported on *Statistics* page, (with loss statistics) and **order:** value on *More Details* page

Finding Results of Interest

- Packet Loss Rates
 - Calculated value based on TCP operation. Reported on *Statistics* page, (with out-of-order statistics) and **loss:** value on *More Details* page
- Path Bottleneck Capacity
 - Measured value based on TCP operation. Reported on *main* page

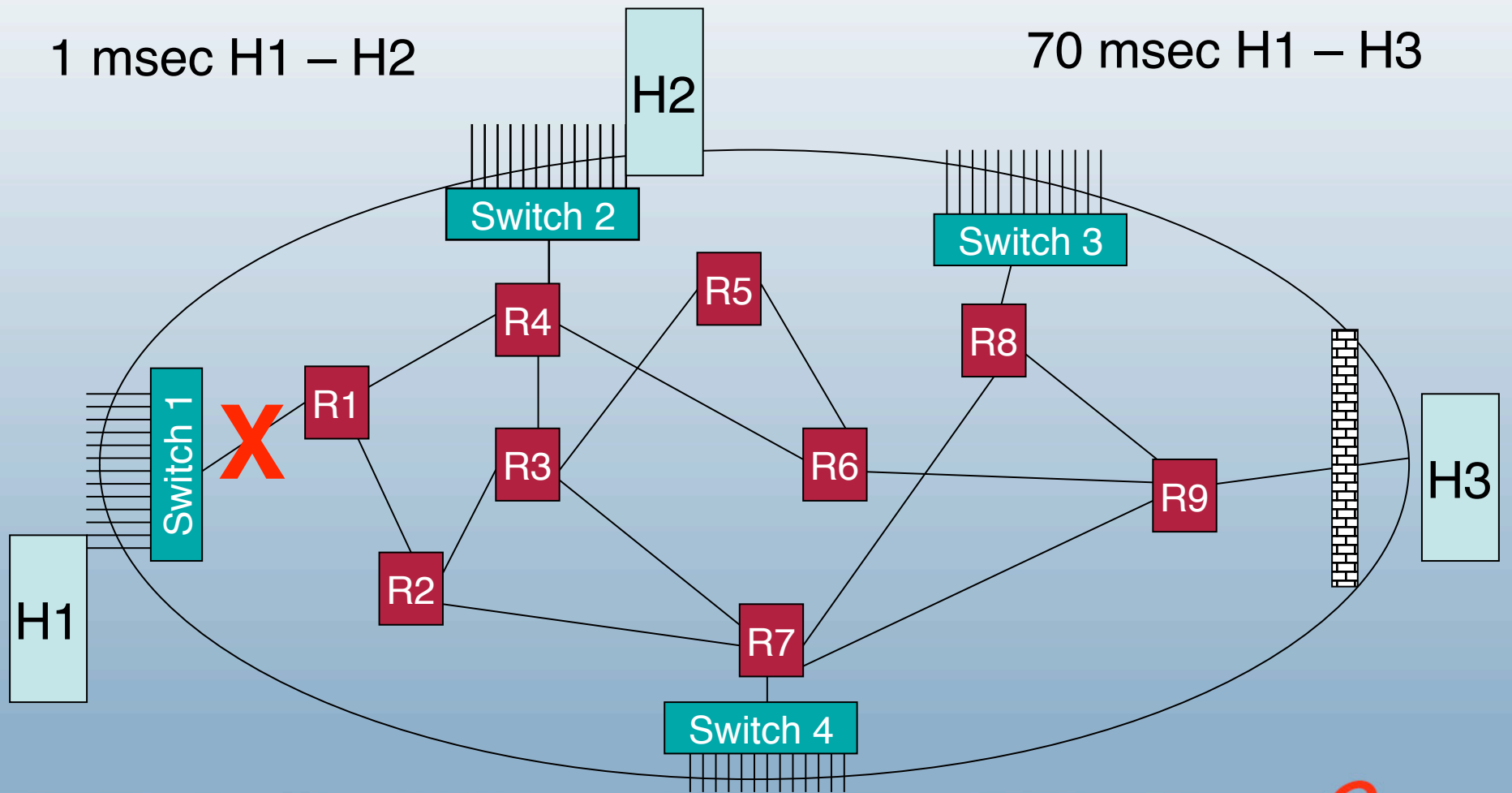
NPAD/pathdiag

- A new tool from researchers at Pittsburgh Supercomputer Center
- Finds problems that affect long network paths
- Uses Web100-enhanced Linux based server
- Web based Java client

Long Path Problem

1 msec H1 – H2

70 msec H1 – H3



Long Path Problem

- E2E application performance is dependant on distance between hosts
- Full size frame time at 1000 Mbps
 - Frame = 1500 Bytes
 - Time = 12 usec
 - In flight for 1 msec RTT = 83 packets
 - In flight for 70 msec RTT = 5833 packets

TCP Congestion Avoidance

- Cut number of packets by $\frac{1}{2}$
- Increase by 1 per RTT
 - LAN (RTT=1msec)
 - In flight changes to 41 packets
 - Time to increase back to 83 is 42 msec
 - WAN (RTT = 70 msec)
 - In flight changes to 2917 packets
 - Time to increase back to 5833 is 3.4 minutes!

NPAD Client test

```
rcarlson@nms-rexp:~ — ssh — 117x42
[rcarlson@nms-rexp ~]$ ping -c3 speedtest2.pnl.gov
PING speedtest2.pnl.gov (192.101.102.24) 56(84) bytes of data.
64 bytes from speedtest2.pnl.gov (192.101.102.24): icmp_seq=1 ttl=60 time=6.55 ms
64 bytes from speedtest2.pnl.gov (192.101.102.24): icmp_seq=2 ttl=60 time=6.52 ms
64 bytes from speedtest2.pnl.gov (192.101.102.24): icmp_seq=3 ttl=60 time=6.51 ms

--- speedtest2.pnl.gov ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2005ms
rtt min/avg/max/mdev = 6.518/6.532/6.555/0.067 ms
[rcarlson@nms-rexp ~]$ diag-client speedtest2.pnl.gov 8100 76 500
Using: rtt 76 ms and rate 500
Connected.
Control connection established.
port = 8003
Starting test.
Parameters based on 6 ms initial RTT
peakwin=4750117 minpackets=3 maxpackets=1283 stepsize=128
Target run length is 575094 packets (or a loss rate of 0.00017388%)
Test 1a (11 seconds): Coarse Scan
Test 2a (9 seconds): Search for the knee
Test 2b (9 seconds): ...
Test 2c (17 seconds): ...
Test 3a (9 seconds): Measure static queue space
Test 3b (9 seconds): ...
Test 3c (17 seconds): ...
Accumulate loss statistics, no more than 130 seconds:
Test 4a (10 seconds): Accumulate loss statistics
Test 4b (10 seconds): ...
Test 4c (10 seconds): ...
Test 4d (10 seconds): ...
Test 4e (10 seconds): ...
Test 4f (10 seconds): ...
Test 4g (10 seconds): ...
Test 4h (10 seconds): ...
Test 4i (10 seconds): ...
Test 4j (10 seconds): ...
Test 4k (10 seconds): ...
Test 4l (10 seconds): ...
Test 4m (10 seconds): ...
report url ServerData/nms-rexp.seat.net.internet2.edu:2009-08-03-22:34:23.html
[rcarlson@nms-rexp ~]$
```

NPAD Sample results

Test Results

http://speedtest2.pnl.gov:8200/ServerData/nms-rexp.seat.net.internet2.edu%3A2009-08-03-22%3A34%3A23.html

granger saab

Most Visited Getting Started Latest Headlines

Chicago... Chicago... pS-Perf... Test ... Summe... DOE - S... perISO... Measur... Measur... OSG Sta... http...do Networ... PlanetFI... Newegg...

Test conditions

Tester: (none) (192.101.102.24) [?]
Target: (none) (64.57.19.4) [?]
Logfile base name: nms-rexp.seat.net.internet2.edu:2009-08-03-22:34:23 [?]
This report is based on a 500 Mb/s target application data rate [?]
This report is based on a 76 ms Round-Trip-Time (RTT) to the target application [?]
The Round Trip Time for this path section is 6.000000 ms.
The Maximum Segment Size for this path section is 8948 Bytes. [?]

Target host TCP configuration test: Pass! [?]

TCP negotiated appropriate options: WSCALE=10, SACKok, and Timestamps. [?]
The target passed all tests! See tester caveats: [?]

Path measurements [?]

The path to the tester is too long for accurate measurements.
> **Test a shorter path section or reduce the target data rate and/or RTT.** [?]

Data rate test: Pass! [?]

Pass data rate check: maximum data rate was 989.456662 Mb/s [?]

Loss rate test: Fail! [?]

Fail: loss event rate: 0.000525% (190639 packets between loss events). [?]
Diagnosis: there is too much background (non-congested) packet loss. [?]
The events averaged 11.700000 losses each, for a total loss rate of 0.006137%. [?]
FYI: To get 500 Mb/s with a 8948 byte MSS on a 76 ms path the total end-to-end loss budget is 0.000174% (575094 packets between losses). [?]
> **Locate the excess packet loss in this section of the path.** [?]

Suggestions for alternate tests

FYI: This path may pass with a less strenuous application: [?]
Try rate=500 Mb/s, rtt=43 ms
Try rate=989 Mb/s, rtt=22 ms
Or if you can raise the MTU: [?]
Try rate=500 Mb/s, rtt=44 ms, mtu=9000 bytes
Try rate=989 Mb/s, rtt=22 ms, mtu=9000 bytes

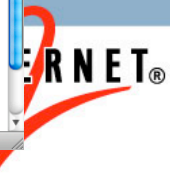
Network buffering test: Warning! [?]

This test did not complete due to other problems with the path, target or tester.
> **Correct other problems first, and then rerun this test.** [?]
Estimated queue size is at least: Pkts: 9 Bytes: 80532
This is probably an underestimate of the actual queue size. [?]
This corresponds to a 0.577737 ms drain time. [?]
To get 500 Mb/s with on a 76 ms path, you need 4750000 bytes of buffer space. [?]
> **Localize all path problems by testing progressively smaller sections of the full path.** [?]

Tester validation: Warning! [?]

The tester has a bottleneck. [?]
> **Please see the instructions on *tester flaws*.**
Tester version: Sid: pathdiag.py.v 1.42 2008/06/18 20:39:33 mathis Exp S

Done



Diagnostic Testing Results

- SC'04 – buffer problem on SGI server limited throughput to 20 Mbps
- PNNL – buffer limits in router causing TCP loss
- UMich – forwarding table overflowed some flows using CPU forwarding
- OU – Regional upgrade caused 100x slowdown
- MAX – Regional network uplink saturated
- I2 – Linux kernel mm penalizes applications that don't fill kernel buffers

Putting it all together – SCP file transfer

Bob and Carol are collaborating on a project. Bob needs to send a copy of the data (50 MB) to Carol every ½ hour. Bob and Carol are 2,000 miles apart. How long should each transfer take?

- 5 minutes?
- 1 minute?
- 5 seconds?

What should we expect?

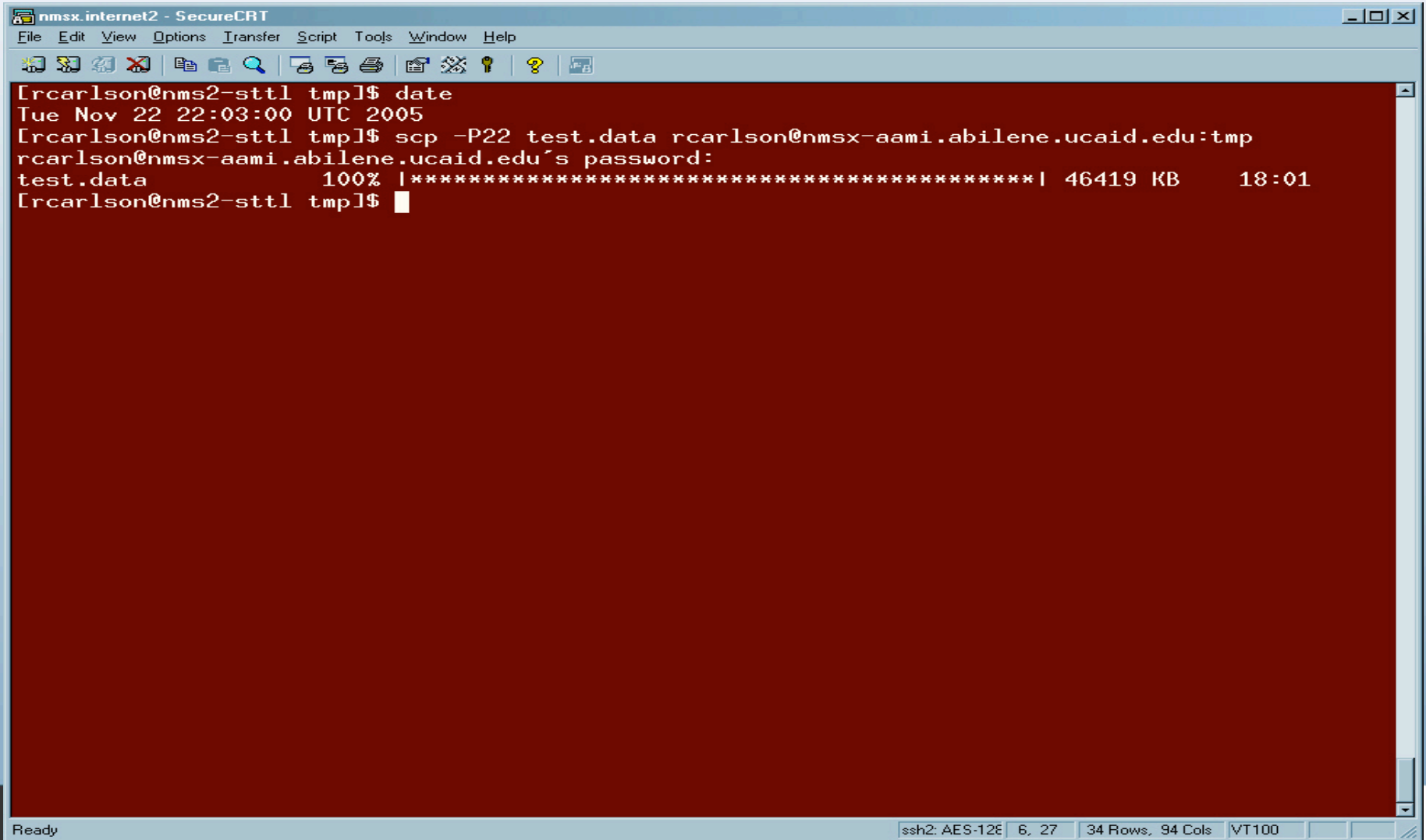
Assumptions:

- 100 Mbps Fast Ethernet is the slowest link
- 50 msec round trip time

Bob & Carol calculate:

- $50 \text{ MB} * 8 = 400 \text{ Mbits}$
- $400 \text{ Mb} / 100 \text{ Mb/sec} = 4 \text{ seconds}$

Initial SCP Test Results



The screenshot shows a SecureCRT terminal window titled "nmsx.internet2 - SecureCRT". The terminal output is as follows:

```
[rcarlson@nms2-sttl tmp]$ date
Tue Nov 22 22:03:00 UTC 2005
[rcarlson@nms2-sttl tmp]$ scp -P22 test.data rcarlson@nmsx-aami.abilene.ucaid.edu:tmp
rcarlson@nmsx-aami.abilene.ucaid.edu's password:
test.data      100% |*****| 46419 KB      18:01
[rcarlson@nms2-sttl tmp]$
```

The terminal window has a menu bar (File, Edit, View, Options, Transfer, Script, Tools, Window, Help) and a toolbar. The status bar at the bottom indicates "Ready", "ssh2: AES-128", "6, 27", "34 Rows, 94 Cols", and "VT100". A red checkmark is drawn in the bottom right corner of the image.

Initial NDT testing shows Duplex Mismatch at one end

File Edit View Go Bookmarks Tools Help

http://207.75.164.80:7123/

Getting Started Latest Headlines

Located at Seattle - WA; 1000 Mbps (Gigabit Ethernet) network connection

This java applet was developed to test the reliability and operational status of your desktop computer and network connection. It does this by sending data between your computer and this remote NDT server. These tests will determine:

- The slowest link in the end-to-end path (Dial-up modem to 10 Gbps Ethernet/OC-192)
- The Ethernet duplex setting (full or half);
- If congestion is limiting end-to-end throughput.

It can also identify 2 serious error conditions:

- Duplex Mismatch
- Excessive packet loss due to faulty cables.

A test takes about 20 seconds. Click on "start" to begin.

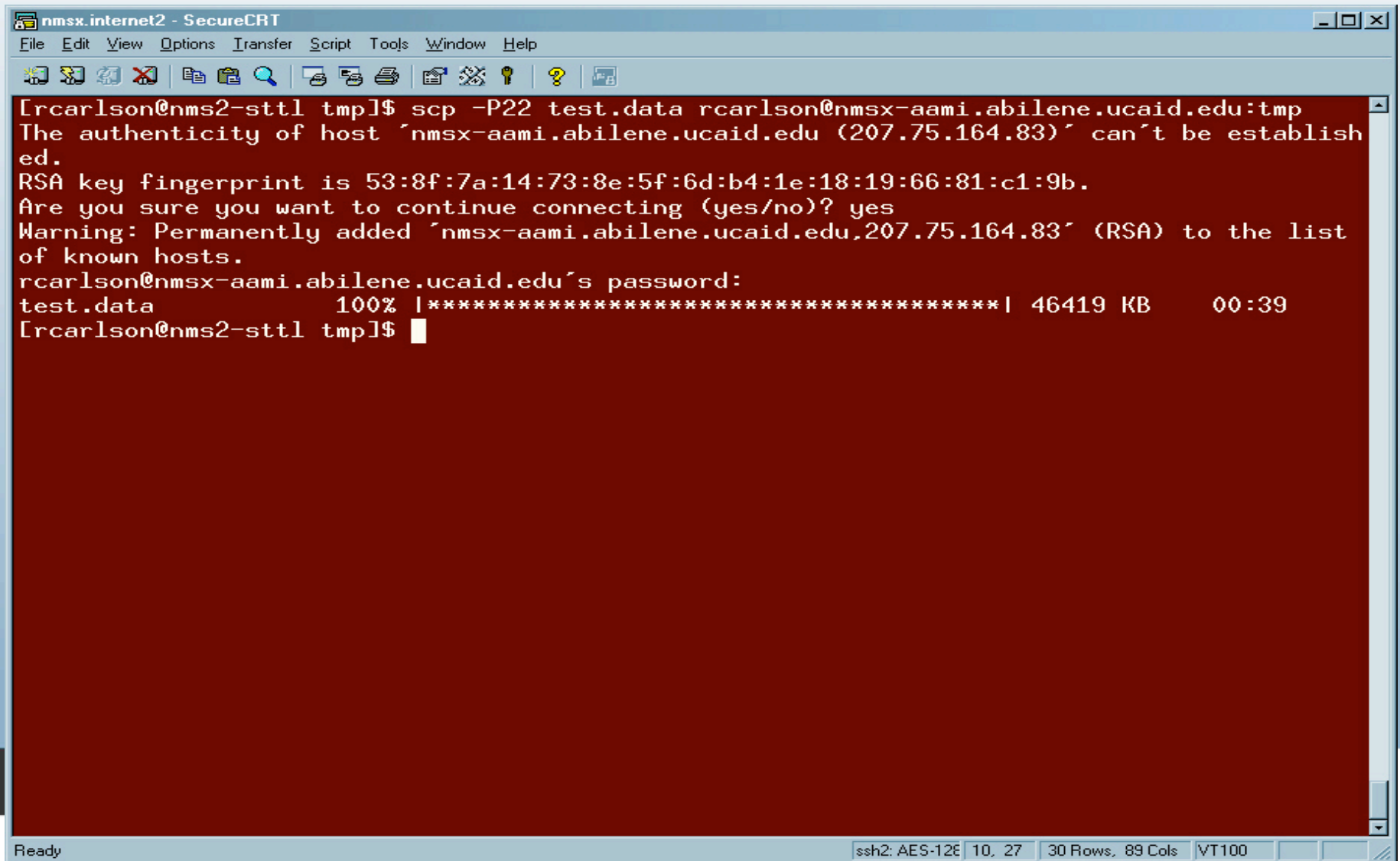
```
TCP/Web100 Network Diagnostic Tool v5.3.4e
click START to begin
Checking for Middleboxes . . . . . Done
running 10s outbound test (client to server) . . . . . 360.76Kb/s
running 10s inbound test (server to client) . . . . . 20.53Mb/s
Warning! Client time-out while reading data, possible duplex mismatch exists
The slowest link in the end-to-end path is a 100 Mbps Full duplex Fast Ethernet subnet
Alarm: Duplex Mismatch condition detected Switch=Full and Host=half

click START to re-test
```

START Statistics More Details... Report Problem

Tcpbw100 done

SCP results after Duplex Mismatch Corrected



The screenshot shows a SecureCRT terminal window titled "nmsx.internet2 - SecureCRT". The terminal displays the following text:

```
[rcarlson@nms2-sttl tmp]$ scp -P22 test.data rcarlson@nmsx-aami.abilene.ucaid.edu:tmp
The authenticity of host 'nmsx-aami.abilene.ucaid.edu (207.75.164.83)' can't be establish
ed.
RSA key fingerprint is 53:8f:7a:14:73:8e:5f:6d:b4:1e:18:19:66:81:c1:9b.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'nmsx-aami.abilene.ucaid.edu,207.75.164.83' (RSA) to the list
of known hosts.
rcarlson@nmsx-aami.abilene.ucaid.edu's password:
test.data      100% |*****| 46419 KB      00:39
[rcarlson@nms2-sttl tmp]$
```

The terminal window has a menu bar (File, Edit, View, Options, Transfer, Script, Tools, Window, Help) and a toolbar. The status bar at the bottom indicates "Ready", "ssh2: AES-128", "10, 27", "30 Rows, 89 Cols", and "VT100".

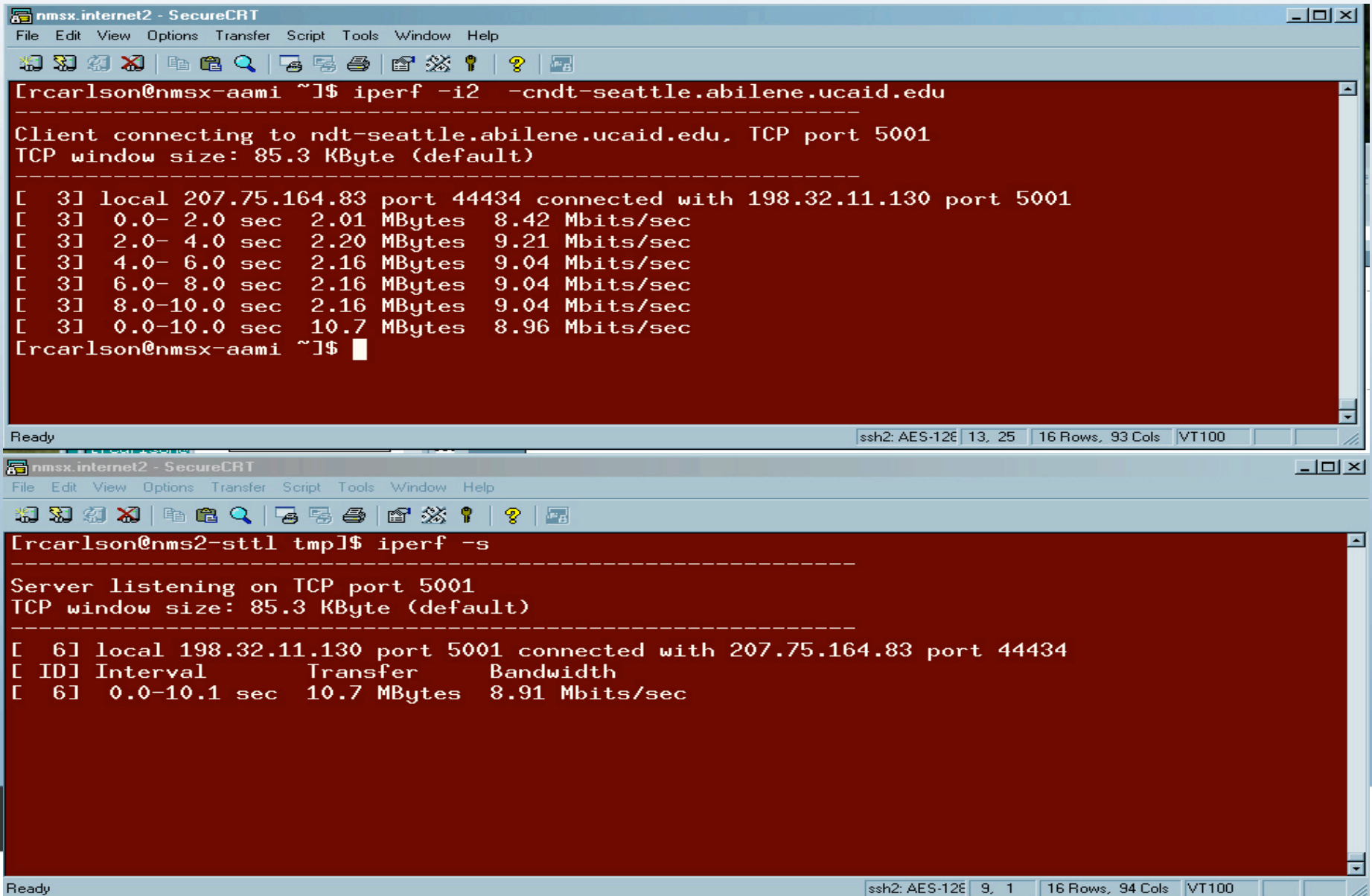
Intermediate Results

Time dropped from 18 minutes to 40 seconds.

But our calculations said it should take 4 seconds!

- $400 \text{ Mb} / 40 \text{ sec} = 10 \text{ Mbps}$
- Why are we limited to 10 Mbps?
- Are you satisfied with $1/10^{\text{th}}$ of the possible performance?

Default TCP window settings



The image displays two screenshots of the SecureCRT terminal application, showing the results of an iperf test. The top window shows a client connection from nmsx-aami to ndt-seattle.abilene.ucaid.edu, and the bottom window shows a server listening on nms2-sttl tmp.

```
nmsx.internet2 - SecureCRT
File Edit View Options Transfer Script Tools Window Help

[rcarlson@nmsx-aami ~]$ iperf -i2 -cndt-seattle.abilene.ucaid.edu
-----
Client connecting to ndt-seattle.abilene.ucaid.edu, TCP port 5001
TCP window size: 85.3 KByte (default)
-----
[ 3] local 207.75.164.83 port 44434 connected with 198.32.11.130 port 5001
[ 3] 0.0- 2.0 sec 2.01 MBytes 8.42 Mbits/sec
[ 3] 2.0- 4.0 sec 2.20 MBytes 9.21 Mbits/sec
[ 3] 4.0- 6.0 sec 2.16 MBytes 9.04 Mbits/sec
[ 3] 6.0- 8.0 sec 2.16 MBytes 9.04 Mbits/sec
[ 3] 8.0-10.0 sec 2.16 MBytes 9.04 Mbits/sec
[ 3] 0.0-10.0 sec 10.7 MBytes 8.96 Mbits/sec
[rcarlson@nmsx-aami ~]$
```

Ready ssh2: AES-128 13, 25 16 Rows, 93 Cols VT100

```
nmsx.internet2 - SecureCRT
File Edit View Options Transfer Script Tools Window Help

[rcarlson@nms2-sttl tmp]$ iperf -s
-----
Server listening on TCP port 5001
TCP window size: 85.3 KByte (default)
-----
[ 6] local 198.32.11.130 port 5001 connected with 207.75.164.83 port 44434
[ ID] Interval Transfer Bandwidth
[ 6] 0.0-10.1 sec 10.7 MBytes 8.91 Mbits/sec
```

Ready ssh2: AES-128 9, 1 16 Rows, 94 Cols VT100

Calculating the Window Size

Remember Bob found the round-trip time was 50 msec

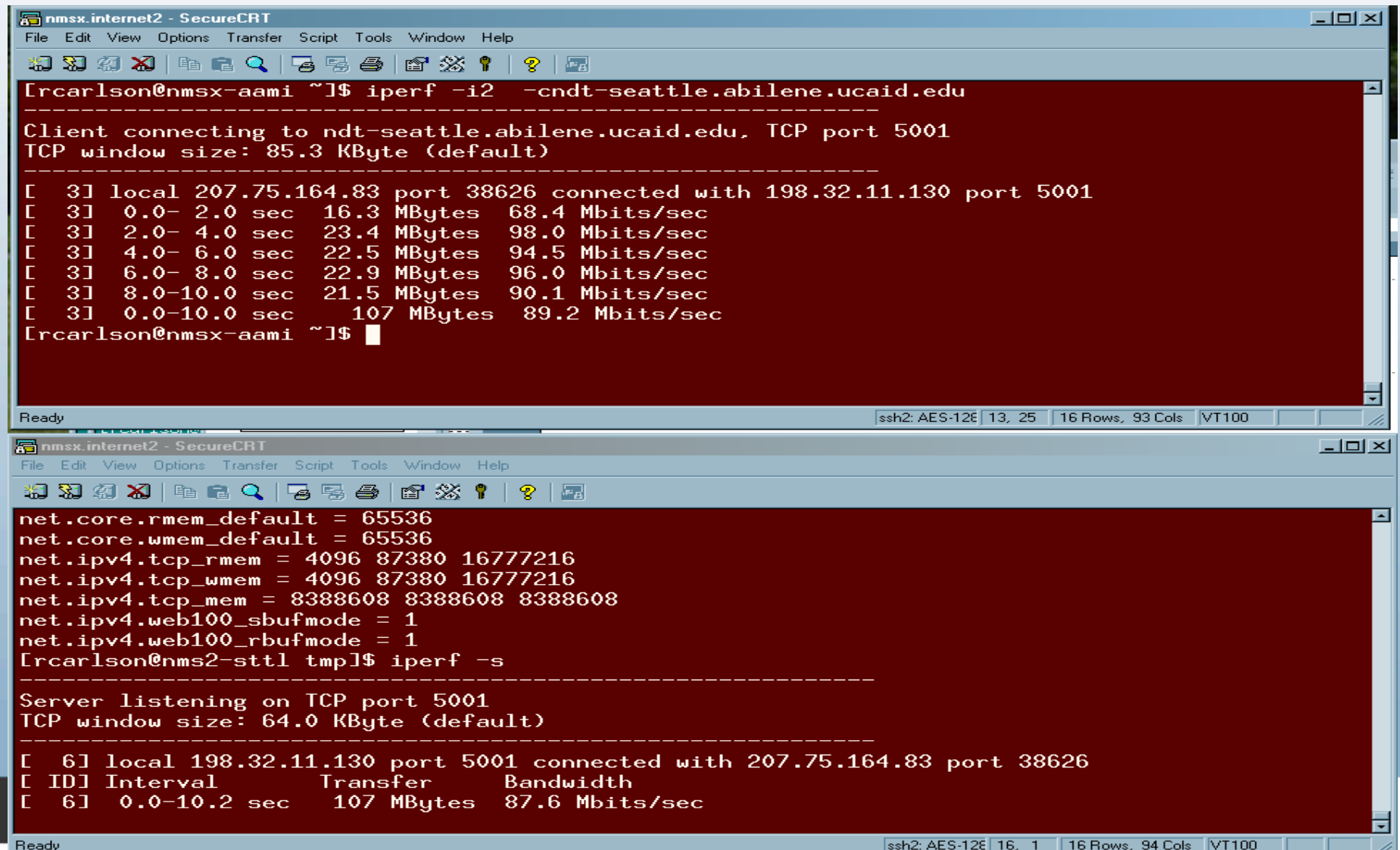
Calculate window size limit

- $85.3\text{KB} * 8 \text{ b/B} = 698777 \text{ b}$
- $698777 \text{ b} / .050 \text{ s} = 13.98 \text{ Mbps}$

Calculate new window size

- $(100 \text{ Mb/s} * .050 \text{ s}) / 8 \text{ b/B} = 610.3 \text{ KB}$
- Use 1MB as a minimum

With TCP windows tuned



The image displays two terminal windows from the SecureCRT application, showing the results of iperf performance tests. The top window shows a client test from nmsx-aami to ndt-seattle.abilene.ucaid.edu, and the bottom window shows a server test on nms2-sttl tmp.

```
nmsx.internet2 - SecureCRT
File Edit View Options Transfer Script Tools Window Help

[rcarlson@nmsx-aami ~]$ iperf -i2 -cndt-seattle.abilene.ucaid.edu
-----
Client connecting to ndt-seattle.abilene.ucaid.edu, TCP port 5001
TCP window size: 85.3 KByte (default)
-----
[ 3] local 207.75.164.83 port 38626 connected with 198.32.11.130 port 5001
[ 3] 0.0- 2.0 sec 16.3 MBytes 68.4 Mbits/sec
[ 3] 2.0- 4.0 sec 23.4 MBytes 98.0 Mbits/sec
[ 3] 4.0- 6.0 sec 22.5 MBytes 94.5 Mbits/sec
[ 3] 6.0- 8.0 sec 22.9 MBytes 96.0 Mbits/sec
[ 3] 8.0-10.0 sec 21.5 MBytes 90.1 Mbits/sec
[ 3] 0.0-10.0 sec 107 MBytes 89.2 Mbits/sec
[rcarlson@nmsx-aami ~]$
```

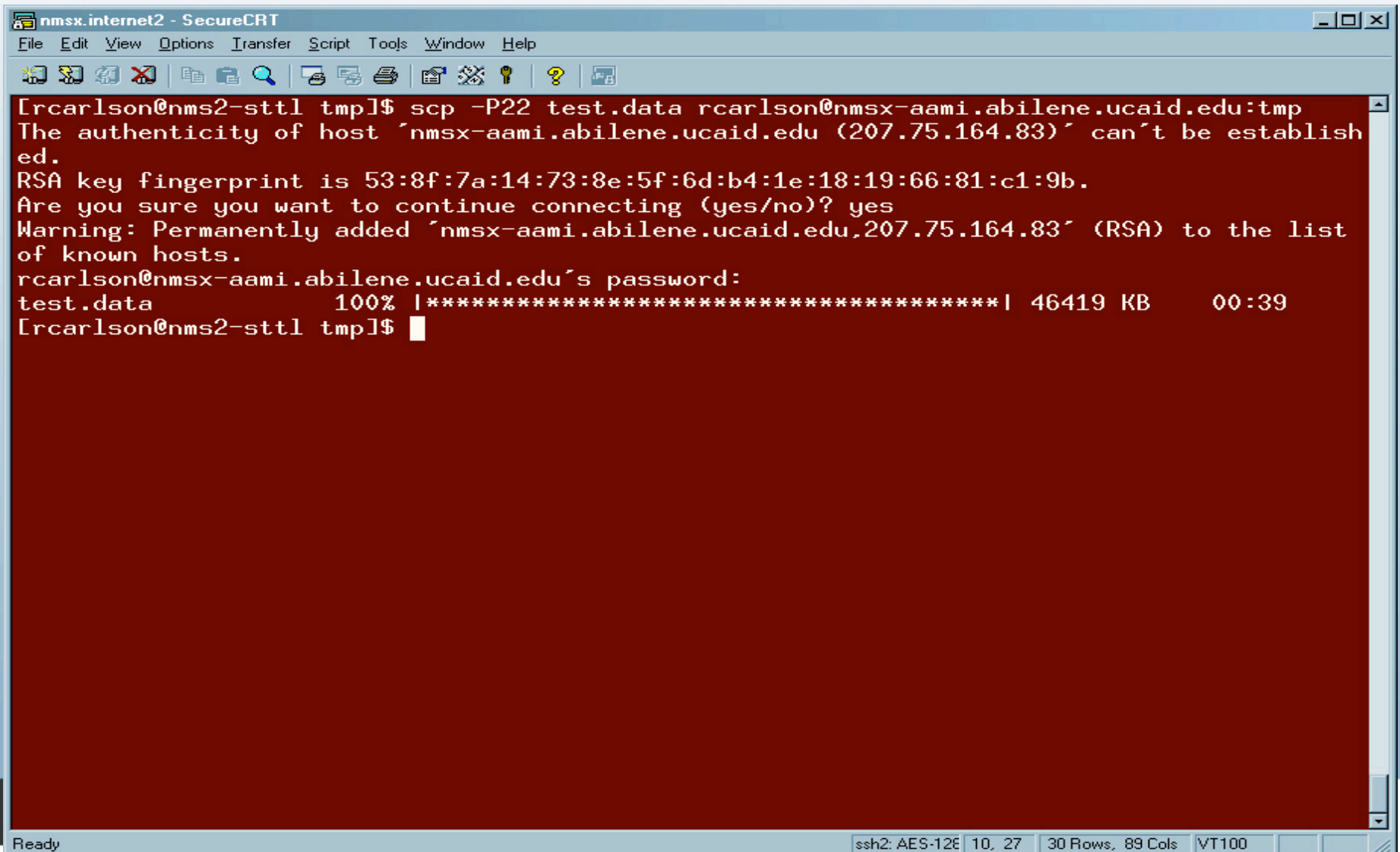
Ready ssh2: AES-128 13, 25 16 Rows, 93 Cols VT100

```
nmsx.internet2 - SecureCRT
File Edit View Options Transfer Script Tools Window Help

net.core.rmem_default = 65536
net.core.wmem_default = 65536
net.ipv4.tcp_rmem = 4096 87380 16777216
net.ipv4.tcp_wmem = 4096 87380 16777216
net.ipv4.tcp_mem = 8388608 8388608 8388608
net.ipv4.web100_sbufmode = 1
net.ipv4.web100_rbufmode = 1
[rcarlson@nms2-sttl tmp]$ iperf -s
-----
Server listening on TCP port 5001
TCP window size: 64.0 KByte (default)
-----
[ 6] local 198.32.11.130 port 5001 connected with 207.75.164.83 port 38626
[ ID] Interval Transfer Bandwidth
[ 6] 0.0-10.2 sec 107 MBytes 87.6 Mbits/sec
```

Ready ssh2: AES-128 16, 1 16 Rows, 94 Cols VT100

SCP results with tuned windows



```
nmsx.internet2 - SecureCRT
File Edit View Options Transfer Script Tools Window Help

[rcarlson@nms2-sttl tmp]$ scp -P22 test.data rcarlson@nmsx-aami.abilene.ucaid.edu:tmp
The authenticity of host 'nmsx-aami.abilene.ucaid.edu (207.75.164.83)' can't be established.
RSA key fingerprint is 53:8f:7a:14:73:8e:5f:6d:b4:1e:18:19:66:81:c1:9b.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'nmsx-aami.abilene.ucaid.edu,207.75.164.83' (RSA) to the list
of known hosts.
rcarlson@nmsx-aami.abilene.ucaid.edu's password:
test.data      100% |*****| 46419 KB    00:39
[rcarlson@nms2-sttl tmp]$
```

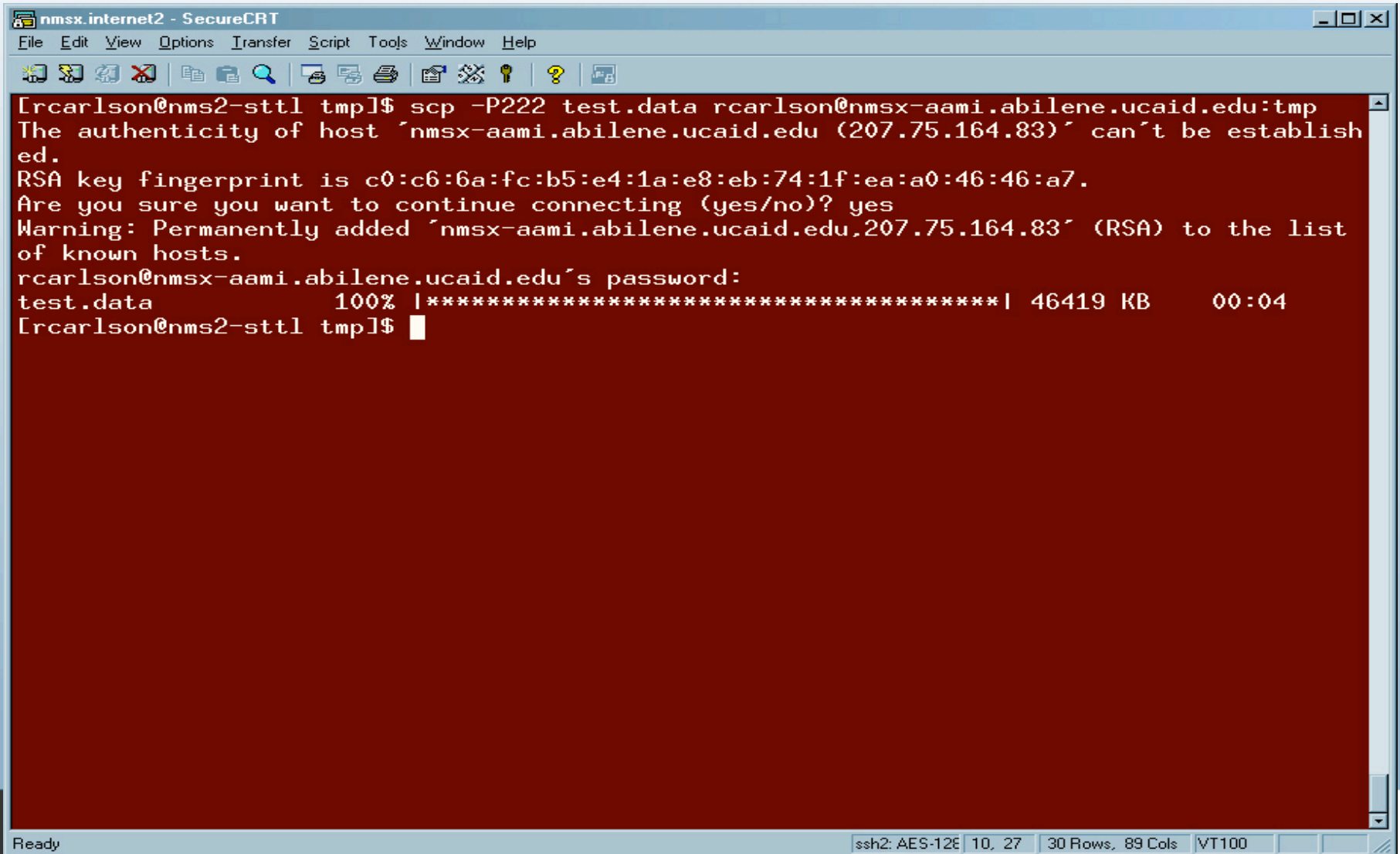
Ready ssh2: AES-128 10, 27 30 Rows, 89 Cols VT100

Intermediate Results

SCP still runs slower than expected

- Hint: SCP uses internal buffers
- Patch available from PSC

SCP Results with tuned SCP



The screenshot shows a SecureCRT terminal window titled "nmsx.internet2 - SecureCRT". The terminal displays the execution of an SCP command to transfer a file to a remote host. The output shows a warning about the host's authenticity, a confirmation to continue, and a successful transfer of a 46419 KB file in 00:04 seconds. A red arrow points to the status bar at the bottom right.

```
nmsx.internet2 - SecureCRT
File Edit View Options Transfer Script Tools Window Help

[rcarlson@nms2-sttl tmp]$ scp -P222 test.data rcarlson@nmsx-aami.abilene.ucaid.edu:tmp
The authenticity of host 'nmsx-aami.abilene.ucaid.edu (207.75.164.83)' can't be establish
ed.
RSA key fingerprint is c0:c6:6a:fc:b5:e4:1a:e8:eb:74:1f:ea:a0:46:46:a7.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'nmsx-aami.abilene.ucaid.edu,207.75.164.83' (RSA) to the list
of known hosts.
rcarlson@nmsx-aami.abilene.ucaid.edu's password:
test.data          100% |*****| 46419 KB      00:04
[rcarlson@nms2-sttl tmp]$
```

Ready ssh2: AES-128 10, 27 30 Rows, 89 Cols VT100

Final Results

Fixed infrastructure problem

Fixed host configuration problem

Fixed Application configuration problem

- Achieved target time of 4 seconds to transfer 50 MB file over 2000 miles

Conclusions

- Primary tools still useful
- Advanced tools are being developed
- Developing tools will make things even easier
- Demand 10 MB/s as the minimum acceptable throughput rate