SHARKFEST '13

Wireshark Developer and User Conference

Wireshark in the Large Enterprise

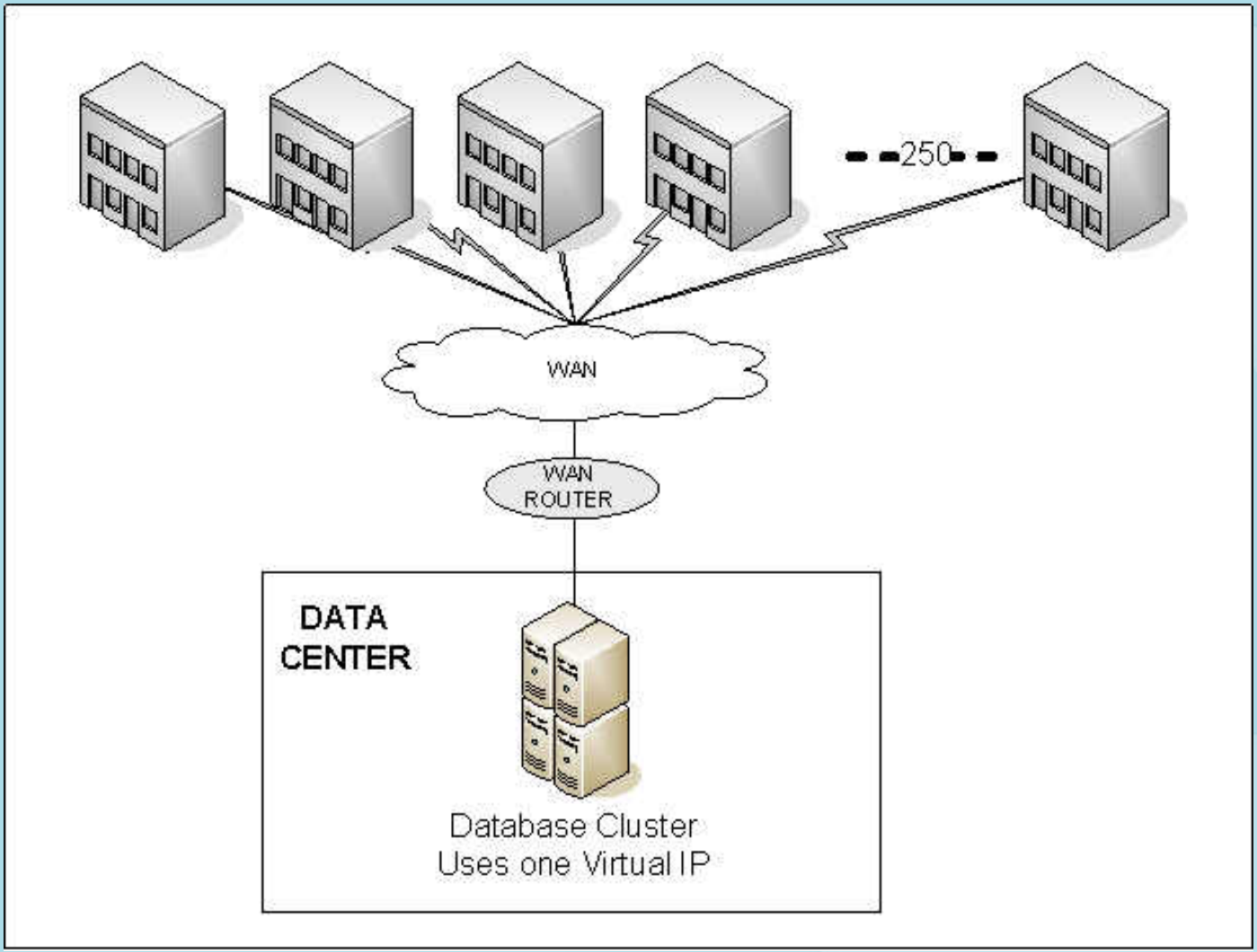
Hansang Bae, Director
Riverbed Performance Management, Architect

http://www.youtube.com/hansangb has the Camtasia recorded sessions.
https://www.box.com/Sharkfest2013 has the trace files used in this session
REFER to the TAKE-AWAY sheet at the end for some key notes/findings!

Failure to Download (FailureToDownload.pcap)

- After implementing router based IPSec solution, random branches stopped receiving the nightly Oracle DB update. The failure is seemingly completely random.
  - The update job is kicked off from the mainframe via one Virtual IP address. All branches talk to the same VIP for the download.
  - Update jobs start simultaneously to as many branches as possible.
  - WAN aggregation routers terminate up to 250
  - The DB's VIP in this example is 192.168.121.231
  - Branch 172.16.152.106 is an example of a working, non-IPSec branch. This is the baseline branch of the working "before IPSec" world.
  - Branch 192.168.101.152 is an example of an IPSec enabled branch that failed to receive the Oracle update.

Failure to Download



SSH is Slow in One Direction (SlowXferChopped.pcapng)

- When transferring files, the direction of the transfer seems to determine the throughput.
  - In one direction, the transfer is fast: 10Mbps+ However, in the opposite direction, it's much slower: 3Mbps or less
  - What must you rule out immediately?
  - Why are there stop and start behavior? PSH bits are not set in between so it's not application buffer tearing issue….or is it?

File Transfer is slow in one direction

- Remember what's important in file transfer throughput issues?
  - You have to rule out Window size issues.
  - Retransmissions (not fast retransmissions) that cause slow start behavior – draining the pipe.
  - Is buffer tearing going on? More common than you think.

Odd UDP Packets

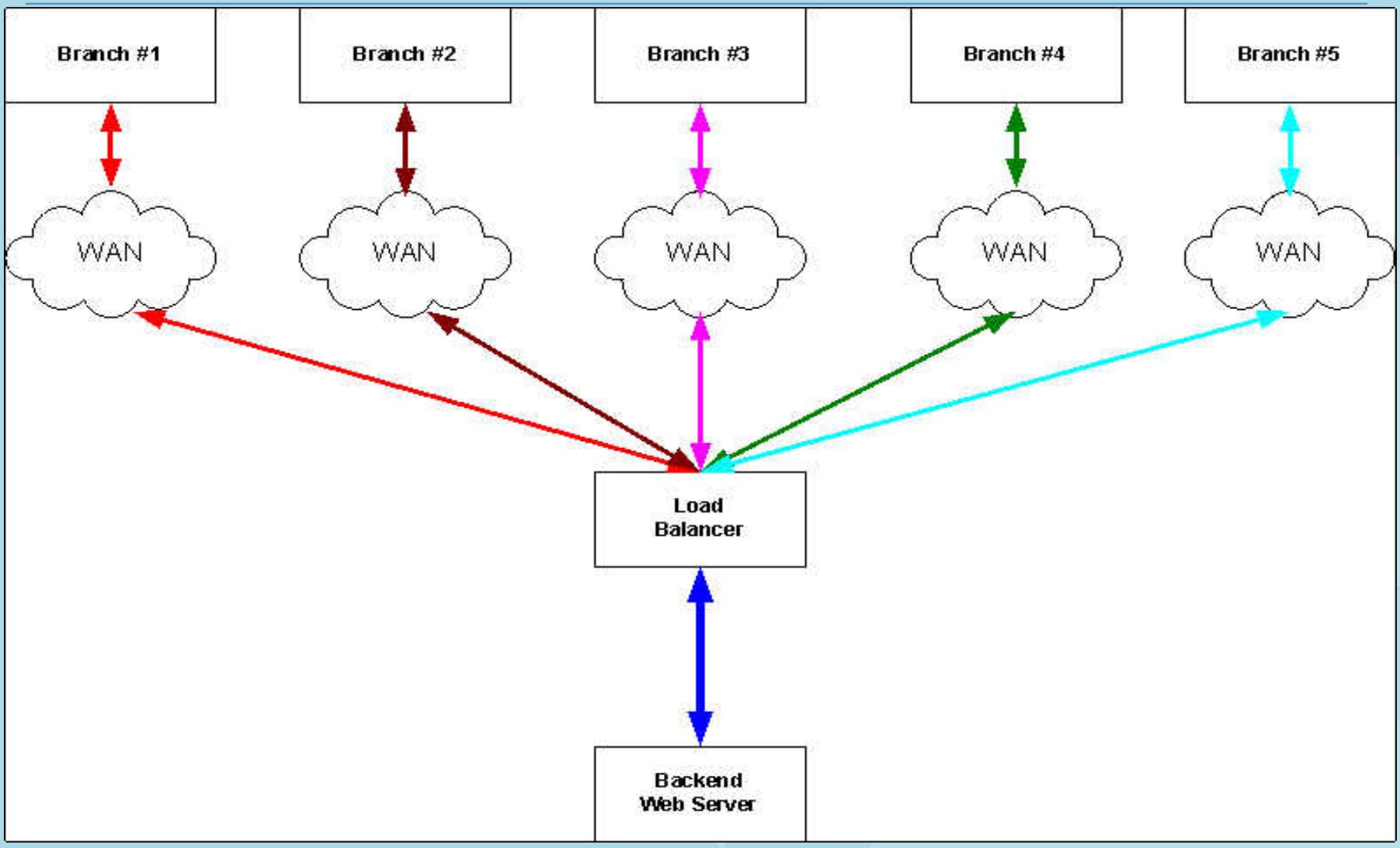
- Who uses UDP port 0?
  - Could be an attack…of some sort
  - It happens day and night, pretty much non-stop
  - What application uses fixed size packets?

- Please note, PCAP file for this scenario is *not* included.

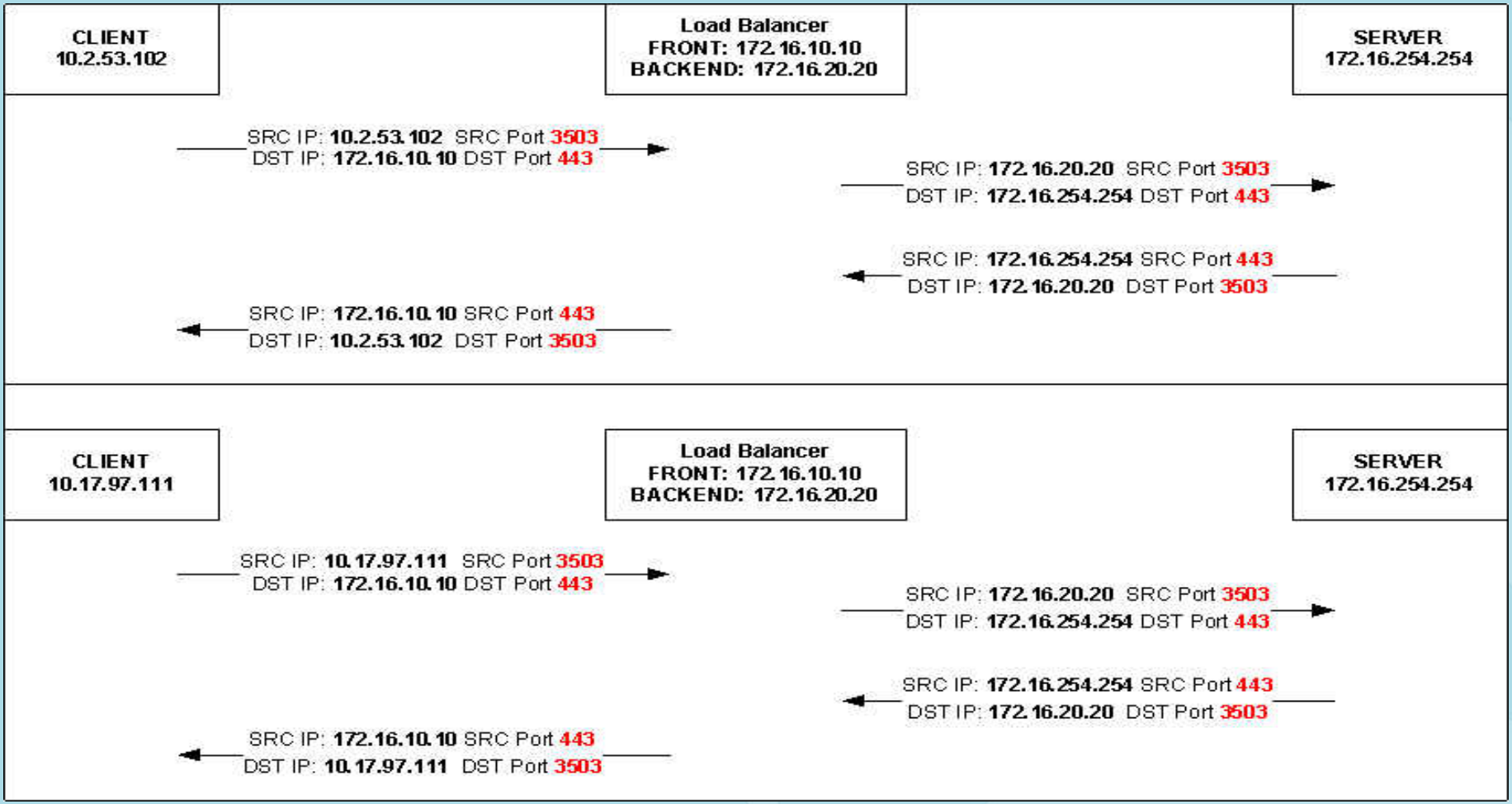
Slow Home Page (LBProbChopped.pcap and LBTCPHandshake.pcap)

- Users are complaining of slow home page load times. There are approximately 20,000 users for this home page.
  - The problem is wide-spread, not easily reproducible….where do you start? What do you do? "Who you gonna call?"
  - What's common in the problem? Home page; use of load balancer; common backend servers; affecting many users.
- What's the job of a load balancer?
- Where should we take the trace?
- What "bad things" can happen if you are using a load balancer with Source NAT configured?

Slow Home Page



Slow Home Page



THANK YOU!!!

- I think I'd be remiss if I didn't thank Gerald, Janice, all the core developers and fellow presenters.
- Special thanks to Rich Siefert and Charles Kaplan for the outstanding key note speeches.
- It boggles my mind how much I learn every time I attend Sharkfest.
- If you consider yourself a protocol analyst, performance engineer, level 3 operations, or network troubleshooter, than you really must attend Sharkfest.
- Finally, any and all feedback (good or bad, but especially bad) is welcome. If you don't tell me what you didn't like, I can't fix it!!! See you at Sharkfest 2014!

Takeaway Sheet

- When using any type of tunneling protocol, you must insure that Path MTU Discovery is functioning properly. Remember, *any* router in your network will send out the ICMP Type 3, Code 4 message. So you must account for this in your firewall rules. The source IP of the ICMP 3/4 message will be the interface closest to the IP address of the receiver.
- For internal facing WAN routers with any type of tunneling protocol, you may want to open up the default rate-limiting behavior. The command is "`ip icmp rate-limit unreachable DF 10`" This example will allow the routers to send out ICMP Type 3 Code 4 messages every 10 milliseconds.
- In the Oracle download example, why do we see Full MSS followed by a smaller packet? This is because default SDU size for Oracle is 2048 bytes. If you are doing a bulk transfer or log file transfers, it behooves you to increase the SDU size. Refer to Oracle Note:274483.1 for more information.
- Remember, you can only transfer one full TCP window size per round trip. Keep this in mind for any throughput related troubleshooting.
- OpenSSH has its own application level windowing.
  - Here is a link (http://grid.ncsa.illinois.edu/ssh/dl/patch/openssh-5.9p1.patch) to a diff for 5.9p1.patch showing the change of application window size from 4 to 64 (CHAN_SES_WINDOW_DEFAULT) in the channels.h file.
  - In OpenSSH_5.8p1 this value is set to 4*CHAN_SES_PACKET_DEFAULT where CHAN_SES_PACKET_DEFAULT is 32*1024 bytes

Takeaway Sheet

- Remember that sequence numbers are nothing more than the number of bytes transferred. Acknowledgement is nothing more than an indication of how much of the data you received. You receive something outside of what's expected, something went horribly wrong!
- When you have a 22,000 user base, having a ephemeral port range of 1024-5000 can be exhausted quickly.
- Sometimes, you have to resort to turning off "relative sequence numbers" for analysis. This is especially true when load balancers – or any device that NATs – are in the data path. As a rule of thumb, when troubleshooting a potential LB issue, turn off relative sequence numbers.
- For the following, refer to the LBTCPHandshake.pcap file. This file only contains the TCP handshake information. Having a hard copy of slide 9 will help greatly! Print it now. No really, print it. I'll wait for you. You really want a copy of the diagram as you go through the following.
  - Frames 1-8 contain the orderly close of a connections.
  - Frame 9 which occurs approx. 14 seconds later is an attempt of a 'new' client to open a connection to the LB. (Frame 10 is the LB translated request to the web server).
  - Frame 11 is an acknowledgement for the prior connection (LB SNAT to Server). This occurs, because the Web server still has this socket in FIN-WAIT. (Frame 12 is the translated request – LB's VIP to client).
  - Frames 13 and 14 is the RST generated by the client (client to VIP), and the translated request (SNAT to server), respectively. Remember, the client sent a SYN with SEQ# of 2582982628. The ACK coming back should be 2582982629. But here, the ACK coming back is 4010391066. Does that ACK# look familiar? Look at packet number 6. Notice the ACK? That's the *prior* TCP connection to a completely different client.
  - Frames 15-18 contain a connection creation. This is allowed to occur because of the RST. However, the client pays a 3 second penalty.