IBM

IBM Enterprise2013
pOS58 – When bad things happen to good systems

Grover Davidson – grover@us.ibm.com

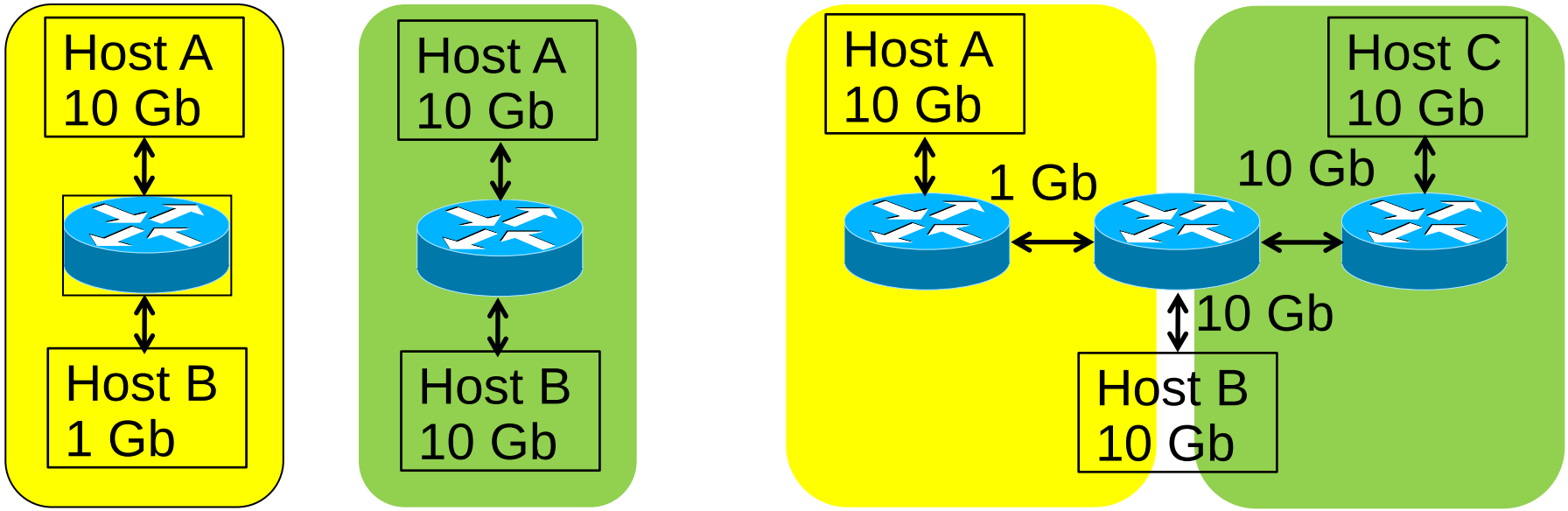
Enterprise2013

Agenda

- Issues with 1Gb and 10Gb Interconnectivity
- Entitlement
- Virtual Ethernet Tuning
- Enabling AME for Fun
- Enabling AMS for Fun
- Bad use of LDR_CNTRL

1Gb and 10GB Interconnectivity

- When performing the speed change between 1Gb and 10Gb networks there will always be an issue of buffering.
- Data coming in on a 10Gb adapter cannot be transmitted out at 10Gb and therefore must be buffered.
- Data from a 1Gb adapter going to a 10Gb network is arriving too slowly and must be buffered and then sent at a transmission rate of 10Gb.
- Also applies to 10Gb to 10Gb traffic if there is a 1Gb connection anyplace between the two end points.



1Gb and 10GB Interconnectivity Problem Symptoms

- At lower levels of network traffic there are no noticeable problems.
- As the network traffic increases there is usually a sudden increase in response times.
- Ping times do NOT increase.
- This makes it look like the problem is on the other node in processing data.
- Many switch vendors prioritize icmp traffic (ping uses icmp_echo requests) and as a result these packets go to the front of the line.
- Application level packets (like those used by SAP's niping program) will show longer latency.
- Problem is also seen if Host A has (1) 10 Gb adapter and Host B has (10) 1Gb adapter. This is (10) 10 Gb <->1 Gb connections!

Entitlement

- Proper entitlement is critical to getting good performance from an LPAR.
- This means avoiding over allocation of resources like more VPs than are needed.
- Allocating too low an entitlement results in resource contention and delays.
- This is especially important on VIO servers that can affect the performance of all it's clients!
- Proper sizing advise can be obtained by running Performance Advisors:
  - http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/p7hb1/iphb1_vios_perf_adv.htm
  - https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Power%20Systems/page/PowerVM%20Virtualization%20Performance%20Advisor
- Under entitlement is a major source of performance problems!
- Please run during peak workload periods for proper results.

Checking Entitlement

- Use ‘lparstat –h 10’ to monitor:

```
lparstat -h

System configuration: type=Shared mode=Uncapped smt=On lcpu=8 mem=12288MB
  psize=63 ent=0.40


%user  %sys  %wait  %idle physc %entc  lbusy  vcsw phint  %hypv hcalls
----- ----- ------ ------ ----- ----- ------ ----- ----- ------ ------
 19.8  18.8    0.1   61.3  0.17  42.8    3.6  1292     8   33.2  18016
  3.4   4.7    0.0   91.9  0.04  10.5    2.7   793     2   93.4   4774
  3.5   4.7    0.0   91.8  0.04  10.6    2.5   756     3   81.2   4722
```

- Check during peak workload.
- %entc should be below 100 *most* of the time.
- VIO Servers should almost never exceed 100% entitlement.
- Occasional spikes over 100 that are not sustained *may* be OK.
- Peak value of physc rounded up is a good rule of thumb for number of VPs in the LPAR.

Virtual Ethernet Buffer Tuning

- Insufficient network buffers can result in serious network performance problems.
- Check the buffer allocations on both the VIO Server and Client:

```
entstat -d entX
Receive Information
  Receive Buffers
    Buffer Type              Tiny    Small   Medium    Large     Huge
    Min Buffers               512      512      128       24       24
    Max Buffers              2048     2048      256       64       64
    History
      Max Allocated           512     2048      128       24       24
```

- If Max Allocated equals Max Buffers then increase max buffers for that type.
- Small buffers need to be increased in the example above:

```
chdev -l entX -a max_buf_small=4096 -a min_buf_small=2048 -P
```

- Requires a reboot to take affect (experienced admins can configure and re-configure interface).
- For HEAVY traffic interfaces increase min to max value to eliminate dynamic buffer management.
- CAUTION: Setting these values too high with a large number of adapters can make the system fail to boot – Tune ONLY when needed.

Enabling AME for Fun and Problems

- AME stands for Automatic Memory Expansion.
- Intended for systems short on physical memory but with extra processor capacity.
- Requires a license key and boot of the LPAR to enable/disable.
- Enabling results in two types of memory pools:
  - Normal/uncompressed for real working memory.
  - Compressed pool to currently unused memory with data stored in a compressed format.
- Data is moved between the two pools based on the memory workload on the system.
- Pools are balanced by AIX.

Enabling AME for Fun and Problems

- Process for proper sizing indicates amepat should be run to determine memory compression factor.
- Amepat determines the compression factor by actually testing the compressibility of the data.
- Expansion factor of 1.0 *disables* compression but does NOT disable the two types of memory pools!
- 64KB pages are completely disabled on the LPAR due to the time needed to compress/decompress 64KB pages!
- 64KB are a key part of AIX performance!
- As a result, enabling AME with a compression factor of 1.0 introduces performance issues but does not give any benefits!
- See Developer Works paper on AME for more details:
  - https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Power%20Systems/page/Active%20Memory%20Expansion%20%28AME%29

Enabling AMS for Fun

- AMS stands for Automatic Memory Sharing.
- Allows a ‘pool’ of memory to be shared by multiple LPARs without using Dynamic LPAR operations to move the memory between the LPARs.
- Total memory ‘allocated’ to the LPARs using the memory pool may exceed the size of the memory pool.
- VIO Server interacts with the LPARs to ‘steal’ memory and keep in free pool.
- When an LPAR accesses the memory, it is moved from the pool to the LPAR.

AMS for Fun

- AMS does not work with 64KB pages.
- The act of enabling AMS disables 64KB pages for the LPARs that are using memory pools.
- As stated for AME, 64KB pages are a key part of AIX performance.
- LPARs not using the memory pools are unaffected.
- See Developer Works paper on AMS for more details:
  - https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Power%20Systems/page/Active%20Memory%20Sharing%20%28AMS%29

Bad Use of LDR_CNTRL

- LDR_CNTRL is an environment variable that changes how executables are loaded into memory.
- Most of the settings will result in more physical memory being used.
- Over-rides any options when the program was compiled/linked or changed with the ldedit command.
- Frequently used by:
  - Java to allow more VMMs segments to be used for heap/data.
  - Oracle sets page sizes.
  - Other applications set page sizes.

- Many programs function best when AIX automatically manages the page sizes.
- Setting LDR_CNTRL in /etc/environment file should be avoided!
- Exporting LDR_CNTRL in a user's environment is usually a bad thing and should only be done after very careful consideration! It affects ALL commands run by the user!

Bad Use of LDR_CNTRL

- Correct usage of LDR_CNTRL use to prepend it before the command being executed:
  - LDR_CNTRL=TEXPSIZE=64K /usr/local/bin/vi
- This adds the LDR_CNTRL to the environment for the command being run but does not export it to subcommands.
- Use SPECIAL CAUTION when using LARGE_PAGE_DATA!
  - Causes application to try and use 16MB pages.
  - There a limited numbers of these available and they can easily be exhausted.
  - If they become exhausted and the setting is for mandatory large pages (LARGE_PAGE_DATA=M) the application will receive and error that no memory is available.
  - Most applications will terminate as a result of this!

Additional References for POWER7

- Performance advisors :
- VIOS Advisor
  - https://www.ibm.com/developerworks/community/wikis/home?lang=en#/wiki/Power%20Systems/page/VIOS%20Advisor
- Java Performance Advisor (JPA)
  - https://www.ibm.com/developerworks/wikis/display/WikiPtype/Java+Performance+Advisor
- PowerVM Virtualization Performance LPAR Advisor:
  - https://www.ibm.com/developerworks/community/wikis/home?lang=en#/wiki/Power%20Systems/page/PowerVM%20Virtualization%20Performance%20Advisor

Additional References for POWER7

- References
- POWER7 Performance Best Practices checklist
  - http://www14.software.ibm.com/webapp/set2/sas/f/best/power7_performance_best_practices_v7.pdf
- Architecture of the IBM POWER7+
  - http://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/tips0972.html?Open