Problem Determination with Linux on System z

Holger Smolinski, Steffen Thoss
IBM

Apr 2nd 2008
Agenda

- Troubleshooting First aid-kit
- Remarks about customer incidents
- Customer reported incidents 2H2006 and 1H2007
  - Storage Controller caching strategies
  - TSM - Network connectivity breaks
  - Disk I/O bottlenecks
  - FCP disk configuration issues
  - More customer problems: in a nutshell
- Ideas to give relief
Trouble-Shooting First Aid kit

- **Install packages required for debugging**
  - s390-tools/s390-utils
  - sysstat
  - lkcdutils

- **Collect dbginfo.sh output**
  - Proactively in healthy system
  - When problems occur – then compare with healthy system

- **Collect system data**
  - Always archive syslog (/var/log/messages)
  - Start sadc (System Activity Data Collection) service when appropriate
  - Collect z/VM Monitor Data if running under z/VM when appropriate
  - Enable /proc/dasd/statistics (see Device Drivers book)
Trouble-Shooting First Aid kit (cont'd)

- **When System hangs**
  - Take a dump (see backup chart)

- **In case of a performance problem**
  - Enable sadc (System Activity Data Collection) service
  - Collect z/VM Monitor Data if running under z/VM
  - Enable DASD statistics:
    See /proc/dasd/statistics on how to enable

- **Function does not work as expected**
  - Enable extended tracing in /proc/s390dbf or /sys/s390dbf for subsystem
Trouble-Shooting First Aid kit (cont'd)

- **Attach comprehensive documentation to problem report:**
  - Output file of dbginfo.sh
  - z/VM monitor data
    - Binary format, make sure, record size settings are correct.
  - When opening a PMR upload documentation to directory associated to your PMR at
    - ftp://ecurep.mainz.ibm.com/, or
    - ftp://testcase.boulder.ibm.com/

- **When opening a Bugzilla at Distribution partner attach documentation to Bugzilla**
Introductory Remarks

- The incidents reported here are real customer incidents
  - Out of years 2006 and 2007
  - Red Hat Enterprise Linux, and Novell Linux Enterprise Server distributions
  - Linux running in LPAR and z/VM of different versions

- While problem analysis look rather straightforward on the charts, it might have taken weeks to get it done.

- The more information is available, the sooner the problem can be solved, because gathering and submitting additional information again and again usually introduces delays.
  - See First Aid Kit in the end of this presentation.

- This presentation focuses on how the tools have been used, comprehensive documentation on their capabilities is in the docs of the corresponding tool.
Performance: ‘disk cache bits settings’

- **Configuration:**
  - This customer was running database workloads on FICON attached storage
  - The problem applies to any Linux distribution and any runtime environment (z/VM and LPAR)
  - The problem also applies to other workloads with inhomogeneous I/O workload profile (sequential and random access)

- **Problem Description:**
  - Transaction database performance is within expectation
  - Warm-up basically consisting of database index scans, takes longer than expected.
Performance: ‘disk cache bits settings’

- Tools used for problem determination:
  - Linux SADC/SAR and IOSTAT
  - Linux DASD statistics
  - Storage Controller DASD statistics
  - Scripted testcase

- Problem Indicators:
  - Random Access I/O rates and throughput are as expected
  - Sequential IO throughput shows variable behaviour
    - always lower than expected
    - As expected for small files, lower than expected for large files
  - Test case showed even stronger performance degradation, when storage controller cache size was exceeded
Use and configure SADC/SAR and iostat:

- **Capture Linux performance data with** `sysstat` **package**
  - System Activity Data Collector (sadc)
  - System Activity Report (sar) command
  - `iostat` command

- **SADC example (for more see man sadc)**
  - `/usr/lib/sa/sadc <interval> <count> <binary outfile>`
  - `/usr/lib/sa/sadc 5 10 sadc outfile`
  - Should be started as a service during system start

- **SAR example (for more see man sar)**
  - `sar -A` --> Analyse data from current sadc data collection

- **IOSTAT example (for more see man iostat)**
  - `iostat -dkx` --> Analyse io related performance data for all disks

- Please include the binary sadc data and sar -A output when submitting SADC information to IBM support
Sysstat Demo
lostat

- **lostat**: shows averaged performance data per device
  - More detailed decomposition than achieved with sadc
  - Especially watch queue size and await/svctm

![Iostat Example](image-url)
Linux DASD statistics

thoss-11:20:27/~temp#cat statistics
36092283 dasd I/O requests
with -1725707784 sectors (512B each)

<table>
<thead>
<tr>
<th>Sector Size</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;4B</td>
<td>1008619</td>
</tr>
<tr>
<td>8B</td>
<td>655629</td>
</tr>
<tr>
<td>16B</td>
<td>3360987</td>
</tr>
<tr>
<td>32B</td>
<td>2579503</td>
</tr>
<tr>
<td>64B</td>
<td>1098338</td>
</tr>
<tr>
<td>128B</td>
<td>215814</td>
</tr>
<tr>
<td>256B</td>
<td>86155</td>
</tr>
<tr>
<td>512B</td>
<td>18022</td>
</tr>
</tbody>
</table>

Histogram of I/O times (microseconds)

<table>
<thead>
<tr>
<th>Sector Size</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>204086</td>
<td>551833</td>
</tr>
<tr>
<td>376809</td>
<td>487413</td>
</tr>
<tr>
<td>760823</td>
<td>1020219</td>
</tr>
<tr>
<td>1447413</td>
<td>1752571</td>
</tr>
</tbody>
</table>

Histogram of I/O times per sector

<table>
<thead>
<tr>
<th>Sector Size</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1244</td>
<td>106729</td>
</tr>
<tr>
<td>462435</td>
<td>645039</td>
</tr>
<tr>
<td>687343</td>
<td>673292</td>
</tr>
<tr>
<td>1073946</td>
<td>1697563</td>
</tr>
<tr>
<td>1921045</td>
<td>1212557</td>
</tr>
<tr>
<td>429291</td>
<td>82078</td>
</tr>
<tr>
<td>23062</td>
<td>5681</td>
</tr>
<tr>
<td>1409</td>
<td></td>
</tr>
</tbody>
</table>

Histogram of I/O time till ssch

<table>
<thead>
<tr>
<th>Sector Size</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>4202149</td>
<td>97492</td>
</tr>
<tr>
<td>144502</td>
<td>41229</td>
</tr>
<tr>
<td>6349</td>
<td>6189</td>
</tr>
<tr>
<td>13122</td>
<td>30505</td>
</tr>
<tr>
<td>70775</td>
<td>112524</td>
</tr>
<tr>
<td>199203</td>
<td>337873</td>
</tr>
<tr>
<td>494914</td>
<td>624231</td>
</tr>
<tr>
<td>892960</td>
<td>961439</td>
</tr>
</tbody>
</table>

Histogram of I/O time between ssch and irq

<table>
<thead>
<tr>
<th>Sector Size</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>513787</td>
<td>173339</td>
</tr>
<tr>
<td>80344</td>
<td>19694</td>
</tr>
<tr>
<td>343</td>
<td></td>
</tr>
</tbody>
</table>

Histogram of I/O time between ssch and irq per sector

<table>
<thead>
<tr>
<th>Sector Size</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>315034</td>
<td>70795</td>
</tr>
<tr>
<td>21271</td>
<td>113</td>
</tr>
</tbody>
</table>

Histogram of I/O time between irq and end

<table>
<thead>
<tr>
<th>Sector Size</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>358030</td>
<td>1224909</td>
</tr>
<tr>
<td>2667755</td>
<td>970430</td>
</tr>
<tr>
<td>369618</td>
<td>185642</td>
</tr>
<tr>
<td>43442</td>
<td>14481</td>
</tr>
<tr>
<td>6120</td>
<td>1779</td>
</tr>
<tr>
<td>427</td>
<td>202</td>
</tr>
<tr>
<td>81</td>
<td>66</td>
</tr>
</tbody>
</table>

# of req in chang at enqueueing (1..32)

<table>
<thead>
<tr>
<th>Sector Size</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

thoss-11:20:30/~temp
DASD statistics (cont’d)

- **DASD statistics decomposition**
  - Summarized histogram information available in `/proc/dasd/statistics`
  - Also accessible per device via `BIODASDPRRD` and `BIODASDPRRST` ioctls

```c
typedef struct dasd_profile_info_t {
    unsigned int dasd_io_reqs;       /* number of requests processed at all */
    unsigned int dasd_io_sects;      /* number of sectors processed at all */
    unsigned int dasd_io_secs[32];   /* histogram of request's sizes */
    unsigned int dasd_io_times[32];  /* histogram of requests's times */
    unsigned int dasd_io_timps[32];  /* histogram of requests's times per sector */
    unsigned int dasd_io_time1[32];  /* histogram of time from build to start */
    unsigned int dasd_io_time2[32];  /* histogram of time from start to irq */
    unsigned int dasd_io_time2ps[32]; /* histogram of time from start to irq */
    unsigned int dasd_io_time3[32];  /* histogram of time from irq to end */
    unsigned int dasd_io_nr_req[32]; /* histogram of # of requests in chanq */
} dasd_profile_info_t;
```
Storage Controller Cache Statistics

- **Available on selected distributions:**
  - ioctl BIODASDPSRD, returning:
    ```c
    typedef struct dasd_rssd_perf_stats_t {
        unsigned char invalid:1;
        unsigned char format:3;
        unsigned char data_format:4;
        unsigned char unit_address;
        unsigned short device_status;
        unsigned int nr_read_normal;
        unsigned int nr_write_normal;
        unsigned int nr_write_fast_normal_hits;
        unsigned int sr_read_seq;
        unsigned int nr_read_seq_bits;
        unsigned int nr_write_seq;
        unsigned int nr_write_fast_seq_bits;
        unsigned int nr_read_cache;
        unsigned int nr_read_cache_bits;
        unsigned int nr_write_cache;
        unsigned int nr_write_cache_bits;
        unsigned int nr_write_fast_cache_bits;
        unsigned int nr_inhibit_cache;
        unsigned int nr_bybass_cache;
        unsigned int nr_seq_dasd_to_cache;
        unsigned int nr_dasd_to_cache;
        unsigned int nr_cache_to_dasd;
        unsigned int nr_delayed_fast_write;
        unsigned int nr_normal_fast_write;
        unsigned int nr_seq_fast_write;
        unsigned int nr_cache_miss;
        unsigned char status2;
        unsigned int nr_quick_write_promotes;
        unsigned char reserved;
        unsigned short ssid;
        unsigned char reserved2[96];
    } __attribute__((packed)) dasd_rssd_perf_stats_t;
    ```

- **Shows details about storage controller cache utilization**
  - Nr or R/W requests and corresponding cache hits

- **Available through storage controller interface (Controller HMC) or Linux ECKD device driver as an ioctl.**
Performance: ‘disk cache bits settings’

- **Problem origin:**
  - Storage controller cache is utilized inefficiently
    - Sequential data not prestaged
    - Used data not discarded from cache

- **Solution:**
  - Configure volumes for sequential I/O different from ones for random I/O
  - And use the tunedasd tool to set appropriate cache-setting bits in CCWs for each device
Networking: 'TSM - breaking TCP connections' 

- **Configuration:**
  - Customer is running TSM backup over LAN with storage pool on minidisks provided by vendor supplied storage controller

- **Problem Description:**
  - During overnight backup runs the TSM clients report backup failure due to TCP/IP disconnect
Networking: ‘TSM - breaking TCP connections'

- Tools used for problem determination:
  - dbginfo.sh
  - Linux for System z Debug Feature
  - Linux SADC/SAR and IOSTAT
  - Linux DASD statistics
  - Storage Controller DASD statistics
Networking:
‘TSM - breaking TCP connections’

- dbginfo.sh collects /var/log/messages
  - At the time of the outages

```
Jan 17 22:40:55 zlinp03 last message repeated 6 times
Jan 17 22:40:55 zlinp03 kernel: qeth: no memory for packet from eth0
Jan 17 22:40:55 zlinp03 kernel: __alloc_pages: 0-order allocation failed (gfp=0x20/0)
Jan 17 22:40:55 zlinp03 kernel: qeth: no memory for packet from eth0
Jan 17 22:40:55 zlinp03 kernel: __alloc_pages: 0-order allocation failed (gfp=0x20/0)
Jan 17 22:40:55 zlinp03 kernel: __alloc_pages: 0-order allocation failed (gfp=0x20/0)
Jan 17 22:40:55 zlinp03 kernel: __alloc_pages: 0-order allocation failed (gfp=0x20/0)
Jan 17 22:40:55 zlinp03 kernel: qeth: no memory for packet from eth0
Jan 17 22:40:55 zlinp03 kernel: __alloc_pages: 0-order allocation failed (gfp=0x20/0)
Jan 17 22:40:55 zlinp03 kernel: __alloc_pages: 0-order allocation failed (gfp=0x20/0)
```
Networking: ‘TSM - breaking TCP connections'

- dbginfo.sh also collects contents of Debug Feature for Linux on System z

```bash
01132180673:456679 0 - 00 788606ba 4e 4f 4d 4d 20 20 20 38 | NOMM 8
01132180673:456810 0 - 00 788606ba 4e 4f 4d 4d 20 20 20 38 | NOMM 8
01132180673:456936 0 - 00 788606ba 4e 4f 4d 4d 20 20 20 38 | NOMM 8
```
Networking: ‘TSM - breaking TCP connections'

- **SADC data collection shows system low on memory at the time of the outages**

```

```

```
Networking:
‘TSM - breaking TCP connections'

- iostat shows long response times for disk I/O requests on certain devices
  - Good values would be between 8-15ms

```
Linux 2.4.21-251-default
Time: 15:23:02
Device: rrqm/s wrqm/s r/s w/s rsec/s wsec/s rkB/s wkB/s avgrq-sz avgqu-sz await svctm %util
/dev/dasdal 0.05 0.15 0.02 0.01 0.58 1.30 0.29 0.65 54.83 0.01 189.33 108.00 0.04
/dev/dasdb1 0.82 0.59 0.50 0.32 10.50 7.30 5.25 3.65 21.67 0.07 87.47 46.99 0.39
/dev/dasdc1 2.62 1.87 0.29 0.25 23.30 17.42 11.65 8.71 75.71 0.93 1722.87 82.23 0.41
thoss-13:16:24-#
```
Networking: ‘TSM - breaking TCP connections'

- z/VM Monitor data shows high service times in disconnected state while FICON channel utilization is rather low
Networking: ‘TSM - breaking TCP connections'

- Problem Indicators:
  - Network connections break, because buffers for inbound packets cannot be allocated due to insufficient memory
  - Disk I/O shows high service time on the storage controller
  - z/VM monitor data show long disconnect times while FICON channels still have capacity.
  - Disks with poor performance are configured as non-full-pack z/VM minidisks
  - Storage Controller statistics data shows large number of cache misses for write operations
  - Observed here, but not relevant: Paging space almost unused, because all memory is used for TSM I/O buffers, which are not pageable.
Networking: ‘TSM - breaking TCP connections'

Problem origin:

- Disk Storage Controller (this one was provided by an independent storage vendor) treated write requests to non-full-pack z/VM minidisks as cache miss and performed a write through operation instead of fast write to NVS cache.

Solution:

- Use fullpack minidisk or dedicated disk as storage pool
- For optimal disk configuration see http://www.ibm.com/developerworks/linux/linux390/perf/tuning_rec_dasd
Performance: 'disk I/O bottlenecks'

- Configuration:
  - Customer has distributed I/O workload to multiple volumes using VM minidisk and LVM striping
  - This problem also applies to non-LVM and non minidisk configurations

- Problem Description:
  - I/O performance is worse than expected by projecting single disk benchmark to more complex solution
Performance: 'disk I/O bottlenecks'

- Tools used for problem determination:
  - dbginfo.sh
  - Linux for System z Debug Feature
  - Linux SADC/SAR and IOSTAT
  - Linux DASD statistics
  - z/VM monitor data
  - Storage Controller DASD statistics

- Problem Indicators:
  - Multi-disk performance is worse than projected single-disk performance.
Performance: 'disk I/O bottlenecks'

- **Problem origin:**
  - bottleneck other than the device – e.g.:
    - z/VM minidisks are associated to same physical disk
    - SAN bandwidth not sufficient
    - Storage controller HBA bandwidth not sufficient
    - Multiple disks used are in the same rank of storage controller

- **Solution:**
  - Check your disk configuration and configure for best performance
    - Make sure, minidisks used in parallel are not on the same physical disk (e.g. for swapspace!)
FCP disk: ‘multipath configuration’

- **Configuration:**
  - Customer is running Samba server on Linux with FCP attached disk managed by Linux LVM.
  - This problem also applies to any configuration with FCP attached disk storage

- **Problem Description:**
  - Accessing *some files* through samba causes the system to hang while accessing other files works fine
  - Local access to the same file cause a hanging shell as well
    - Indicates: this is not a network problem!
FCP disk: ‘multipath configuration’

- Tools used for problem determination:
  - dbginfo.sh

- Problem Indicators:
  - Intermittent outages of disk connectivity
FCP disk: ‘multipath configuration’
FCP disk: ‘multipath configuration’
FCP disk: ‘multipath configuration’

- **Solutions**
  - Configure multipathing correctly:
    - Establish independent paths to each volume
    - Group the paths using the device-mapper-multipath package
    - Base LVM configuration on top of mpath devices instead of `sd<#>`
  - For a more detailed description how to use FCP attached storage appropriately with Linux on System z, see [http://download.boulder.ibm.com/ibmdl/pub/software/dw/linux390/c](http://download.boulder.ibm.com/ibmdl/pub/software/dw/linux390/c)
More customer problems:
In a nutshell
Performance: 'aio (POSIX asynchronous I/O) not used'

- **Configuration:**
  - Customer is running DB2 on Linux

- **Problem Description:**
  - Bad write performance is observed, while read performance is okay

- **Tools used for problem determination:**
  - DB/2 internal tracing

- **Problem Origin:**
  - libaio is not installed on the system

- **Solution:**
  - Install libaio package on the system to allow DB2 using it.
Memory: ‘higher order allocation failure’

- Configuration:
  - Customer is running CICS transaction gateway in 31 bit emulation mode

- Problem Description:
  - After several days of uptime, the system runs out of memory

- Tools used for problem determination:
  - Dbginfo.sh

- Problem Indicators:
  - Syslog contains messages about failing 4th-order allocations
    - Caused by compat_ipc calls in 31bit emulation, which request 4th-order memory chunks

- Problem Origin:
  -Compat_ipc code makes order-4 memory allocations

- Solution:
  - Switch to 31 bit system to avoid compat_ipc
  - Upgrade to SLES10
  - Request a fix from distributor or IBM
Memory: '31bit address space exhausted'

- **Configuration:**
  - Customer is migrating database contents to different host in a 31bit system.

- **Problem Description:**
  - Database reports system caused out-of-memory condition:
    'SQL1225N The request failed because an operating system process, thread, or swap space limit was reached.' indicating that a syscall returned -1 and set errno to ENOMEM

- **Tools used for problem determination:**
  - DB/2 internal tracing

- **Problem Origin:**
  - System out of resources due to 31bit kernel address space

- **Solution:**
  - Try to reduce memory footprint of workload (nr of threads, buffer sizes...)
  - Run migration in 31bit compatibility environment of 64 bit system
System stalls: 'PFAULT loop'

- **Configuration:**
  - Customer is running 35 Linux guests (SLES 8) in z/VM with significant memory overcommit ratio.

- **Problem Description:**
  - After a couple of days of uptime, the systems hang.

- **Tools used for problem determination:**
  - System dump

- **Problem Origin:**
  - CPU loop in the pfault handler caused by
    - Linux acquiring a lock in pfault handler although not needed

- **Solution:**
  - Request a fix for Linux from SUSE and/or IBM
System stalls: 'reboot hangs'

- **Configuration:**
  - Customer is running Linux and issuing 'reboot'-command to re-IPL

- **Problem Description:**
  - 'reboot' shuts down the system but hangs.

- **Tools used for problem determination:**
  - System dump

- **Problem Indicators:**
  - 'reboot' hangs, but LOAD-IPL works file

- **Problem Origin:**
  - Root cause: CHPIDs are not reset properly during 'reboot'

- **Solution:**
  - Apply Service to Linux, ask SUSE/IBM for appropriate kernel level.
Cryptography: 'HW not used for AES-256'

- Configuration:
  - Customer wants to use Crypto card acceleration for AES-encryption

- Problem Description:
  - HW acceleration is not used – system falls back to SW implementation

- Tools used for problem determination:
  - SADC/SAR

- Problem Indicators:
  - CPU load higher than expected for AES-256 encryption

- Problem Origin:
  - System z Hardware does not support AES-256 for acceleration.

- Solution:
  - Switch to AES 128 to deploy HW acceleration
  - Expect IBM provided Whitepapers on how to use cryptography appropriately
Cryptography: 'glibc error in openssl'

- Configuration:
  - Customer is performing openssl speed test to check whether crypto HW functions are used in SLES10

- Problem Description:
  - Openssl speed test fails with an error in glibc: "glibc detected openssl: free(): invalid next size (normal)"

- Solution:
  - Upgrade Linux to SLES10 SP1 or above
Storage: 'zipl fails in EAL4 environment'

- **Configuration:**
  - Customer installs an EAL4 compliant environment with ReiserFS

- **Problem Description:**
  - Zipl refuses to write boot records due to an ioctl blocked by the auditing SW

- **Problem Indicators:**
  - Zipl on ext3-FS works well

- **Solution:**
  - Use ext3-FS at least for /boot
Storage: 'DASD unaccessible'

- **Configuration:**
  - Customer is running SLES9 with LVM configuration

- **Problem Description:**
  - DASDs become not accessible after boot

- **Problem Indicators:**
  - Intermitting errors due to race between LVM and device recognition

- **Solution:**
  - Apply service to Linux
  - Race fixed, due to which partition detection couldn’t complete, because LVM had devices already in use.
Storage: 'non-persistent tape device nodes'

- **Configuration:**
  - Customer uses many FCP attached tapes

- **Problem Description:**
  - Device nodes for tape drives are named differently after reboot

- **Solution:**
  - Create UDEV-rule to establish persistent naming
  - Wait for IBMtape device driver to support persistent naming
Storage: 'tape device unaccessible'

- **Configuration:**
  - Customer has FCP attached tape

- **Problem Description:**
  - Device becomes unaccessible

- **Problem Indicators:**
  - ELS messages in syslog, or
  - Device can be enabled manually, but using hwup-script it fails

- **Solution:**
  - Apply service to get fixed version of hwup scripts
  - Apply service to Linux and µCode and disable QIOASSIST if appropriate
  - If tape devices remain reserved by SCSI 3rd party reserve use the ibmtape_util tool from the IBM Tape device driver package to break the reservation
Storage: 'QIOASSIST'

- **Configuration:**
  - Customer is running SLES10 or RHEL 5 under z/VM with QIOASSIST enabled

- **Problem Description:**
  - System hangs

- **Problem Indicators:**
  - System stops operation because all tasks are in I/O wait state
  - System runs out of memory, because I/O stalls
  - When switching QIOASIST OFF, the problems vanish

- **Solution:**
  - **Apply service to Linux, z/VM and System z µCode**
Networking: 'firewall cuts TCP connections'

- Configuration:
  - Customer is running eRMM in a firewalled environment

- Problem Description:
  - After certain period of inactivity eRMM server loses connectivity to clients

- Problem Indicators:
  - Disconnect occurs after fixed period of inactivity
  - Period counter appears to be reset when activity occurs

- Solution:
  - Tune TCP_KEEPALIVE timeout to be shorter than firewall setting, which cuts inactive connections
Networking: 'Channel Bonding'

- **Configuration:**
  - Customer is trying to configure channel bonding on SLES 10 system

- **Problem Description (Various problems):**
  - Interfaces refuse to get enslaved
  - Failover/failback does not work
  - Kernel Panic when issuing 'ifenslave -d' command

- **Solution:**
  - Apply Service to Linux, System z HW and z/VM
    - ask SUSE/IBM for appropriate kernel and μCode levels.
Networking: 'tcpdump fails'

- **Configuration:**
  - Customer is trying to sniff the network using tcpdump

- **Problem Description (Various problems):**
  - tcpdump does not interpret contents of packets or frames
  - tcpdump does not see network traffic for other guests on GuestLAN/HiperSockets network

- **Problem Indicators:**
  - OSA card is running in Layer 3 mode
  - HiperSocket/Guest LAN do not support promiscuous mode

- **Solution:**
  - Use the layer-2 mode of your OSA card to add Link Level header
  - Use the tcpdump-wrap.pl script to add fake LL-headers to frames
  - Use the fake-ll feature of the qeth device driver
  - Wait for Linux distribution containing support for promiscuous mode
Networking: 'dhcp fails'

- **Configuration:**
  - Customer is configuring Linux guests with dhcp and using VLAN

- **Problem Description (Various problems):**
  - Dhcp configuration does not work on VLAN because
    - Dhcp user space tools do not support VLAN packets

- **Problem Indicators:**
  - When VLAN is off, dhcp configuration works fine.

- **Workaround:**
  - Apply service to Linux to hide VLAN information from dhcp tools
    - Ask Distributor/IBM for appropriate kernel levels

- **Solution:**
  - Request VLAN aware dhcp tools from your distributor
Ideas for relief

- Provide more webpages like shown on the ‘Links’-chart
- Create regular Linux on System z Newsletter
  - Provide information about current hot topics
  - Hints & tips about system configuration
- Establish Linux & VM Health Check Offerings for the GEOs
  - Proactive check of system configuration
  - Risk assessment
  - Recommendations to optimize configuration
- Your ideas are welcome!
Your feedback and questions:

- Raise it right now!
- Write it on the feedback sheets!
- Submit it by email to
  - Steffen Thoss (thoss@de.ibm.com)
  - Holger Smolinski (smolinski@de.ibm.com)
  - linux390@de.ibm.com
- Please refer to this presentation
Backup
Links

- **Linux on System z project at IBM DeveloperWorks:** http://www.ibm.com/developerworks/linux/linux390/

- **HW and SW level requirements for QIOASSIST:** http://www.vm.ibm.com/perf/aip.html

- **Fixed I/O buffers with z/VM 5.1:**

- **Optimize disk configuration for performance:**

- **DASD cache bit tuning:**
Dump Tools Summary

<table>
<thead>
<tr>
<th>Tool</th>
<th>Stand alone tools</th>
<th>VMDUMP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DASD</td>
<td>Tape</td>
</tr>
<tr>
<td>Environment</td>
<td>VM&amp;LPAR</td>
<td>LPAR</td>
</tr>
<tr>
<td>Preparation</td>
<td>Zipl -d /dev/&lt;dump_dev&gt;</td>
<td>Mkdir /dumps/mydumps zipl -D /dev/sda1 ...</td>
</tr>
<tr>
<td>Creation</td>
<td>Stop CPU &amp; Store status ipl &lt;dump_dev_CUU&gt;</td>
<td>Vmdump</td>
</tr>
<tr>
<td>Dump medium</td>
<td>ECKD or FBA</td>
<td>Tape cartridges</td>
</tr>
<tr>
<td>Copy to filesystem</td>
<td>Zgetdump /dev/&lt;dump_dev&gt; &gt; dump_file</td>
<td></td>
</tr>
<tr>
<td>Viewing</td>
<td>Lcrash or crash</td>
<td></td>
</tr>
</tbody>
</table>