Linux on System z

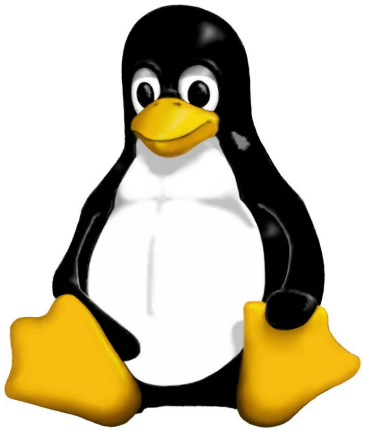
Linux on System z Problem Reporting and Analysis

Michael Holzheu (holzheu@de.ibm.com)
Linux on System z Development
IBM Lab Boeblingen, Germany

WAVV 2008

Agenda

- Error reporting tools
- Dump tools
- Dump analysis tools
- System Tap
- Demo

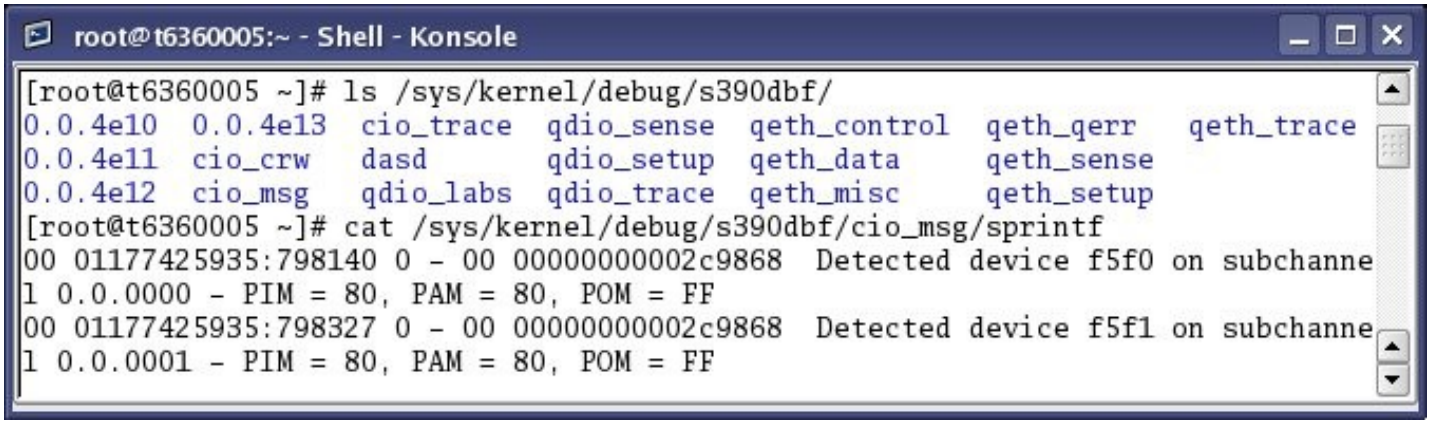


Collect technical setup information

- Run script dbginfo.sh (part of the s390-tools package)
- s390-tools package can be downloaded from developer works
- dbginfo.sh captures the following information:
  - /proc/[version, cpuinfo, meminfo, modules, partitions, devices ...]
  - System z specific device driver information: /proc/s390dbf
  - Kernel messages: /var/log/messages
  - Config files /etc/[ccwgroup.conf, chandev.conf, modules.conf, fstab]
  - Output of several commands: ps, dmesg
- Query setup scripts:
  - lscss, lsdasd, lsqeth, lszfcp, lstape

S390 debug feature (s390dbf)

- z/Linux specific driver tracing environment
- Uses wraparound memory buffers
- Available in live system and in system dumps
- Live System:



- Views: hex_ascii and sprintf
- Set Level: # `echo 6 > level`
- Flush dbf: # `echo - > flush`
- Increase buffer size: # `echo 10 > pages`

Dump Tools

Linux on System z Dump Tools

- DASD dump tool:
  - Writes dump directly on DASD partition
  - Uses s390 standalone dump format
  - ECKD and FBA DASDs supported
- Tape dump tool:
  - Writes dump directly on Escon/Ficon Tape device
  - Uses s390 standalone dump format
- SCSI dump tool
  - Writes dump into filesystem
  - Uses lkcd dump format
- VMDUMP
  - Writes dump to vm spool space (VM reader)
  - z/VM specific dump format

DASD dump under z/VM

- Prepare dump device under Linux, if possible on 64Bit environment:
  `zipl -d /dev/<dasd>`
- After Linux crash issue these commands on 3270 console:
  ```
  #cp cpu all stop
  #cp store status
  #cp i <dasd_devno>
  ```
- Wait until dump is saved on device:
  ```
  00: zIPL v1.6.0 dump tool (64 bit)
  00: Dumping 64 bit OS
  00: 00000087 / 00000700 MB
  ...
  00: Dump successful
  ```
- Only disabled wait PSW on older Distributions

VMDUMP

- The only method to dump NSSes under z/VM
- Works nondisruptive
- Create dump:

```
#cp vmdump to cmsguest
```

- Receive dump:
  - From reader into CMS dump file: `dumpload`
  - NEW: vmur device driver: `vmur rec <spoolid> vmdump`
- Linux tool to convert vmdump:

```
vmconvert vmdump linux.dump
```

- Problem: Dump process relatively slow

Install SCSI dump disk

- Create partition with PCBIOS disk-layout (fdisk)
- Format partition with ext2 or ext3 filesystem
- Install dump tool:
  - mount disk:

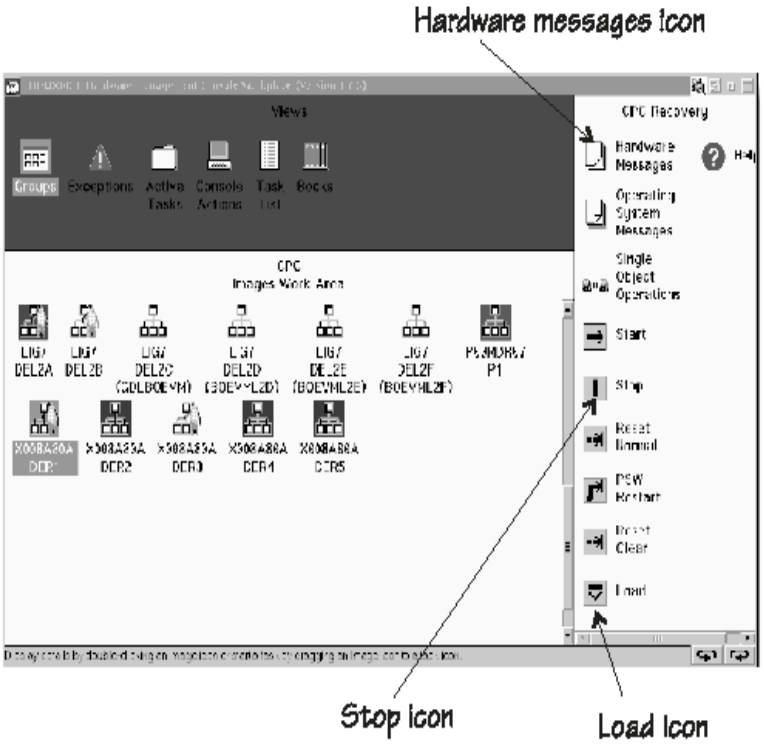
    ```
    mount /dev/sda1 /dumps
    ```
  - prepare disk:

    ```
    zipl -D /dev/sda1 -t /dumps
    ```
  - Optional: etc/zipl.conf:

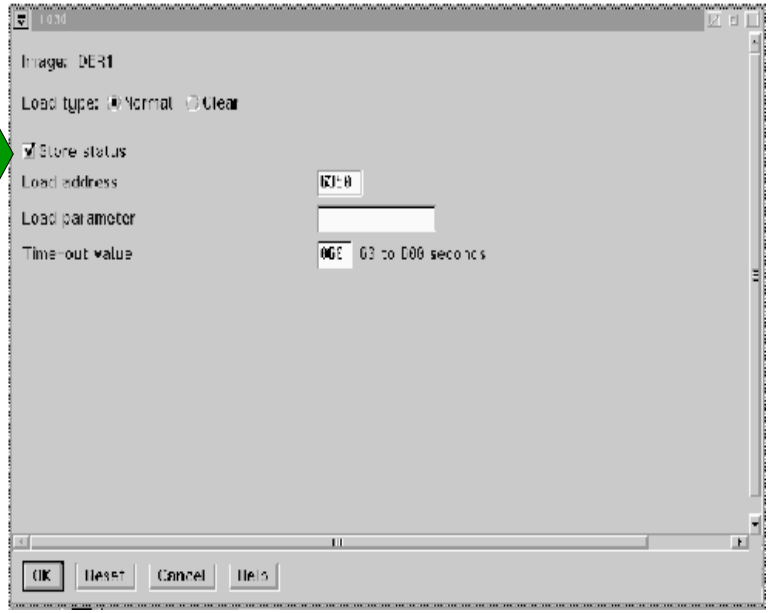
    ```
    dumptofs=/dev/sda1
    target=/dumps
    ```

- Dump tool creates dumps directly in filesystem
- Works only on LPAR – not under VM!

DASD dump on LPAR



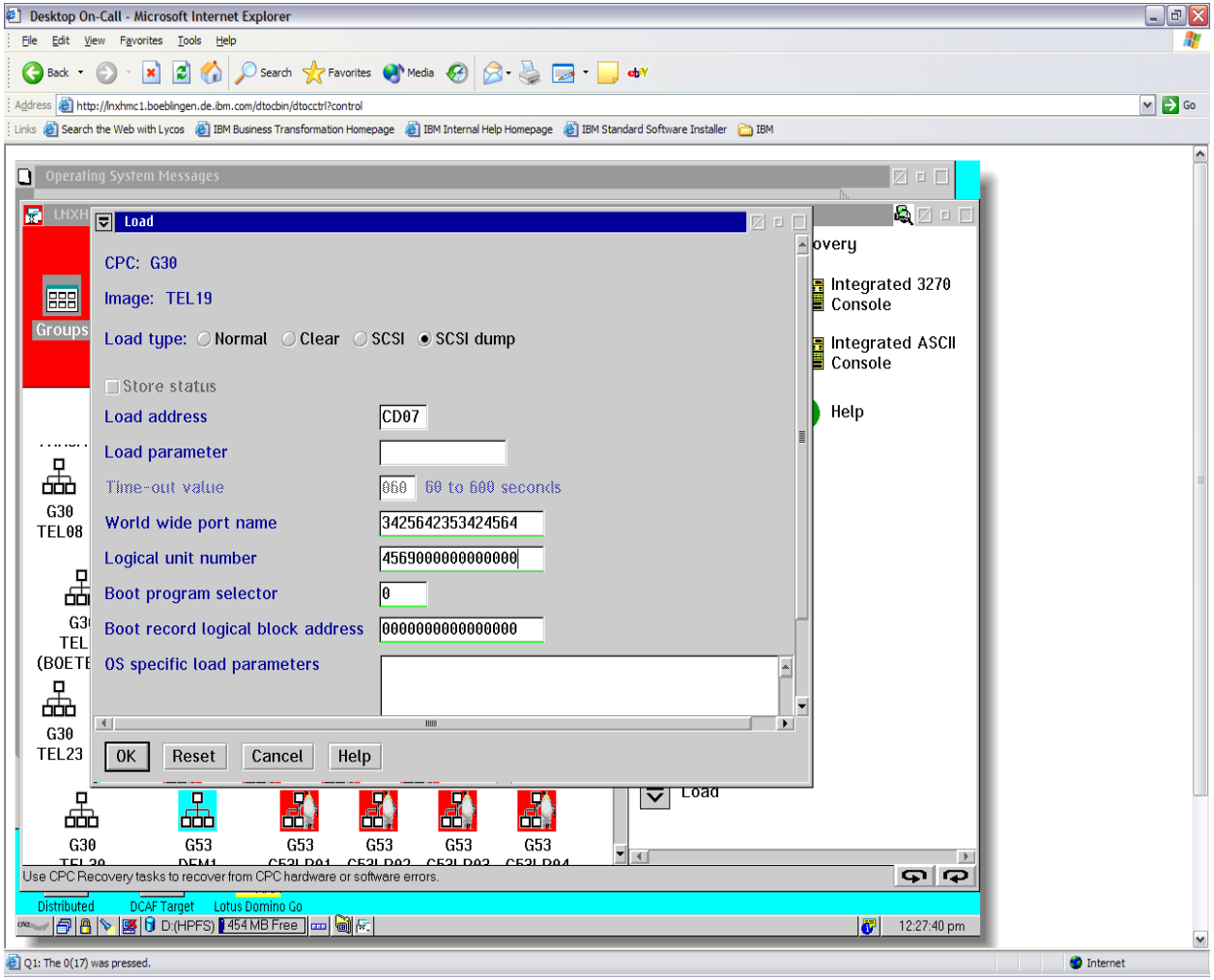
Store Status



Note: The appearance of the Stop and Start icons can vary. On some consoles they appear as green and red traffic lights.

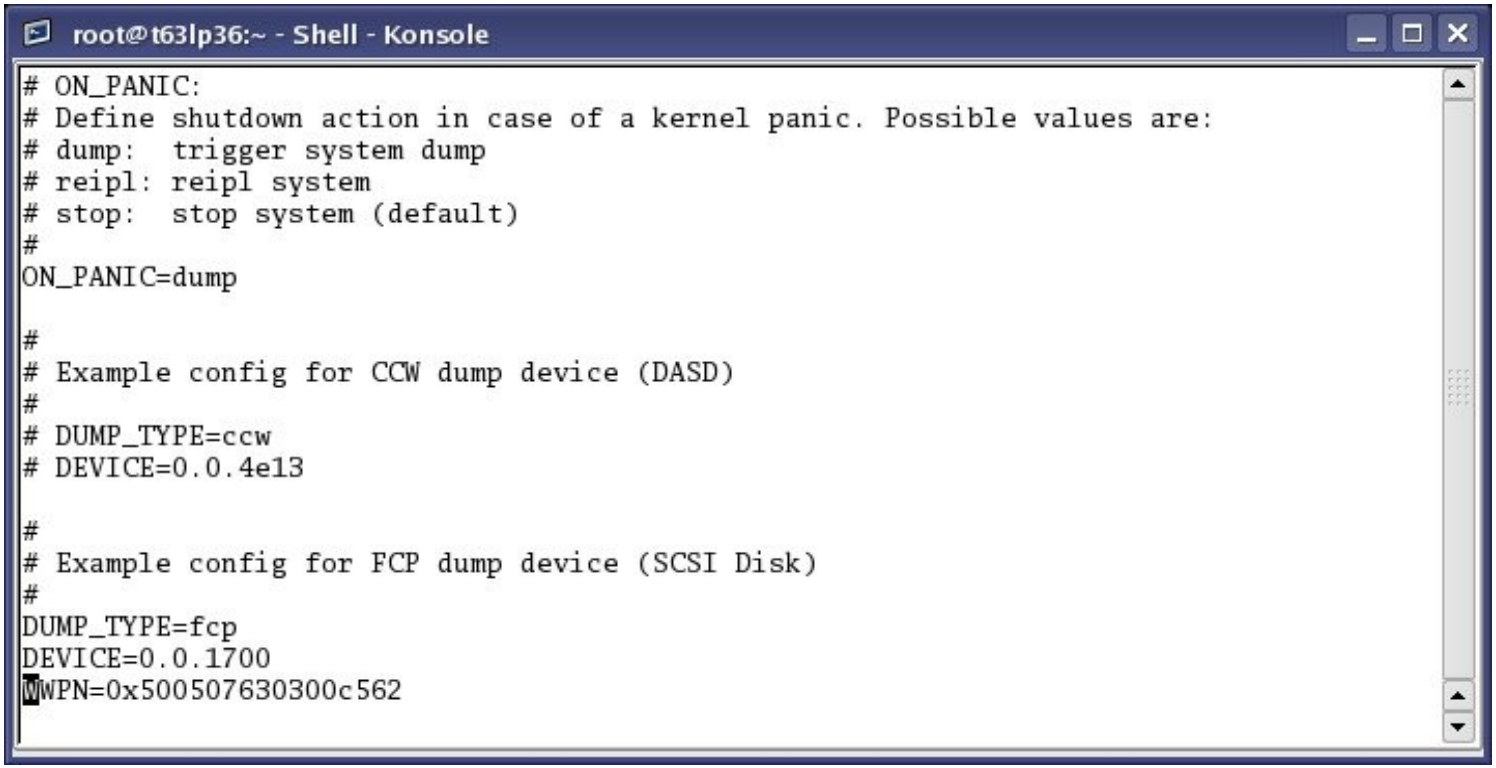
SCSI dump on LPAR

- Specify:
  - WWPN
  - DEVNO
  - LUN
  - BPS



Automatic dump on panic (SLES10 / RHEL5)

- /etc/sysconfig/dumpconf



- Start service: # service dumpconf start

Get dump and send it to service organization

- DASD/Tape:
  - Store dump to Linux file system from dump device:
    ```
    # zgetdump /dev/<device node> > dump_file
    ```
  - Alternative: lcrash (Compression possible)
    ```
    # lcrash -d /dev/dasdxx -s <dir>
    ```
- SCSI:
  - Get dump from filesystem
- Additional files needed for dumpanalysis:
  - SUSE (lcrash tool): /boot/System.map-xxx and /boot/Kerntypes-xxx
  - Redhat & SUSE (crash tool): vmlinux file with debug info contained in debug kernel rpms:
    - Redhat: kernel-debuginfo-2.6.18-1.2910.el5.s390x.rpm
    - SUSE: kernel-s390x-debug-2.6.5-7.183.rpm

Dump Tools Summary

Tool	Stand alone tools			VMDUMP
	DASD	Tape	SCSI	
Environment	VM&LPAR		LPAR	VM
Preparation	Zipl -d /dev/<dump_dev>		Mkdir /dumps/mydumps zipl -D /dev/sda1 ...	---
Creation	Stop CPU & Store status ipl <dump_dev_CUU>			Vmdump
Dump medium	ECKD or FBA	Tape cartridges	LINUX file system on a SCSI disk	VM reader
Copy to filesystem	Zgetdump /dev/<dump_dev> > dump_file		---	Dumpload ftp ... vmconvert ...
Viewing	Lcrash or crash			

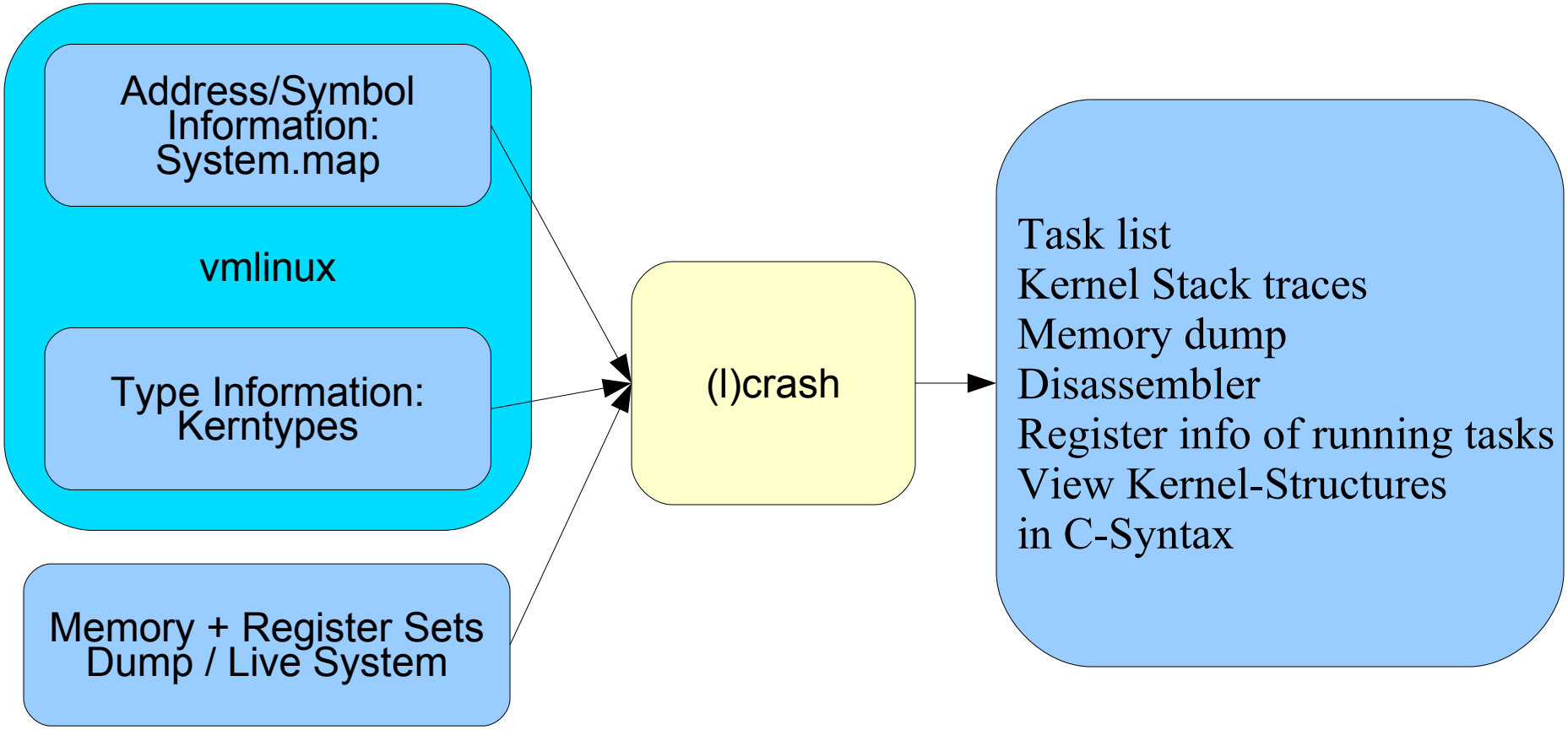
See “Using the dump tools” book on
http://www-128.ibm.com/developerworks/linux/linux390/index.html

Dumpanalysis Tools

Dumpanalysis Tools

- Lcrash:
  - Lkcdutils project: http://sourceforge.net/projects/lkcd
  - Only supported by SUSE
  - Dump: `# lcrash System.map <dump> Kerntypes`
  - Live System: `# lcrash System.map /dev/mem Kerntypes`
- Crash:
  - http://people.redhat.com/anderson/
  - Supported by SUSE and Redhat
  - Dump: `# crash vmlinux <dump>`
  - Live System: `# crash vmlinux`

What you need and what you get



SystemTap

SystemTap features

- Infrastructure for dynamic kernel traces
- Provides C-like scripting language
- Trace points:
  - function entry / exit
  - Absolute address
  - Source code Line number
  - Timer
  - Start/End
- Access, modify and print variables
- Predefined tap sets for process, scsi, etc..
- Predefined functions like pid(), execname(), etc.
- Allows “embedded” kernel C-functions

Example

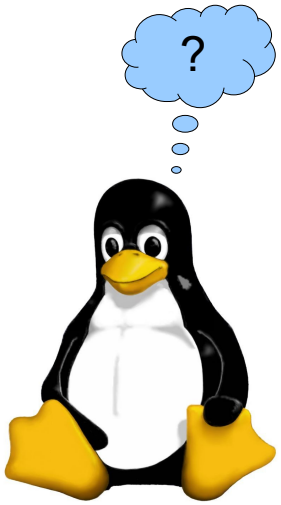
```
probe kernel.function("debug_set_level")
{
    old_level = $id->level;
    printf("%d: old = %i, new = %i\n",
           pid(),  old_level, $new_level)
}


void
debug_set_level(debug_info_t* id, int new_level)
{
    ...



# stap test.stp
# echo 5 > /sys/kernel/debug/s390dbf/dasd/level
# 25614: old = 1, new = 5
```

Demo

Demo Scenarios

- Spinlock deadlocks
- Mutex deadlocks
- Memory consumption problems
- Trigger kernel panic and create dump
- Dumpanalysis with lcrash and crash

Questions?



Trademarks

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries. For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml: AS/400, DBE, e-business logo, ESCO, eServer, FICON, IBM, IBM Logo, iSeries, MVS, OS/390, pSeries, RS/6000, S/30, VM/ESA, VSE/ESA, Websphere, xSeries, z/OS, zSeries, z/VM

The following are trademarks or registered trademarks of other companies

Lotus, Notes, and Domino are trademarks or registered trademarks of Lotus Development Corporation
Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries
LINUX is a registered trademark of Linux Torvalds
UNIX is a registered trademark of The Open Group in the United States and other countries.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.
Intel is a registered trademark of Intel Corporation
* All other products may be trademarks or registered trademarks of their respective companies.

NOTES:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Any proposed use of claims in this presentation outside of the United States must be reviewed by local IBM country counsel prior to such use.

The information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.