Kernel Dumping

GUUG Frühjahrsfachgespräch 2012, München

Karsten Keil <keil@b1-systems.de>
Stefan Seyfried <seyfried@b1-systems.de>
B1 Systems GmbH

http://www.b1-systems.de

March 2, 2012

Agenda

Agenda

- What's kernel dumping?
- Why kernel dumping?
- How does this work?
- Closer look on a kernel dump

What is kernel dumping?

What is kernel dumping?

What is kernel dumping?

A kernel dump

- is an entire dump/copy of your physical memory
- in the moment of a "crash"
- saved in a file

What is kernel dumping?

A kernel dump

- doesn’t have to be a real kernel crash
- can be also created manually
- can be done manually e. g. on a system hang

Why kernel dumping?

Why kernel dumping?

Why kernel dumping?

- postmortem analysis of the system
- ... in the moment the problem occured...
- ... without needing the real system

Why kernel dumping?

- Very helpful on "mission critical" environments
- productive platforms which don't allow maintenance outage for reproducing/debugging or time for capturing information about the crash
- analysed by support
- gives supporters/developers a lot of helpful information
- very, very helpful if the problem is very hard to reproduce

How does kernel dumping work?

How does kernel dumping work?

How does kernel dumping work?

- There are/were different techniques:
  - netdump
  - lkcd
  - **kdump**

How does kdump work?

- memory hole is reserved on boot
- kernel parameter: `crashkernel=256M-:128M@16M`
- a (special) kernel is booted via kexec on a "crash"
- kdump-kernel: `CONFIG_CRASH_DUMP=y`
- `makedumpfile` is called...
- ... and copies the physical memory content into a file
- (regular reboot)

What does a kernel crashdump look like?

What does a
kernel crashdump look like?

What does a kernel crashdump look like?

- ELF corefile / *kdump-compressed*-format (makedumpfile)
- similar to userspace core files
- known from `ulimit -c unlimited` or tools like `apport`
- might **contain sensitive information!**
- it's your entire physical memory!

crash - the application

crash - the application

- crash is based on gdb
- provides lots of helpful kernel debugging helper functions
- these are functions which allow easier access to known kernel structures
  - ringbuffer / dmesg
  - process list
  - easy access to task_struct
- allows debugging of kernel modules, too

crash vs. gdb

- kernel crashdump can be loaded like a regular coredump with gdb
- each CPU is a process thread in gdb
- with debuginfo it also allows browsing the source code
- (and assembler/source intermixing, helpful with basic assembler knowledge)

kdump testing

kdump testing

- test all planned ways to trigger a manual crashdump
- (e. g. on a potential system hang)
- SysRq via keyboard
- NMI button on the chassis
- NMI via remote management board
- SysRq via (virtual) serial console
- load the crashdump with `crash`, the debugging tool

kdump testing / troubleshooting

- does `kexec` work at all on the target box?
- SysRq activated?
- plan to get a crashdump of hanging system?
- enough space for crashdump? (problem especially on machines with lots of memory ← full copy of memory!)
- do you get notified once a crashdump has ben written? ("silent" reboot after dump ...)
- debuginfo of "crashed" kernel available?
- Have you got a trustworthy contact for dump analysis? support subscription?

Prepare for crashdump

Prepare for crashdump

Prepare for crashdump

- kdump setup
- reboot with `crashkernel=` in your kernel commandline
- `grep -q crashkernel /proc/cmdline`
- enable `sysrq`
- set `sysctl`:
  - `kernel.unknown_nmi_panic`
  - `kernel.panic_on_oops`
- get familiar with remote management board if available

Problems with crashdump / crashdump debugging

"But it does not work!"

Problems with crashdump - nothing happens

- Graphics drivers?
  - `vga=0`
  - `nomodeset`
- Verbose boot: `debug`
- Driver problems: hardware which was already initialized
- If it still fails: bugreport against the kernel

Problems with crashdump - kdumptool fails

- `kdumptool` is a compiled binary
  - "Because most functionality is needed in the initrd, design decision was to provide that functionality in binary without huge dependencies and without a scripting language." – kdumptool(8)
  - Problem: if it fails, it is hard to debug
- On failure, you get dropped into an emergency console
- Design problems can be worked around with `KDUMP_PRESCRIPT` in `/etc/sysconfig/kdump`

Dumping Virtual Machines

Dumping Virtual Machines

- Dump is done by the hypervisor
- No cooperation of the VM is necessary
- Not much preparation is necessary
- Works with KVM and Xen
- `virsh dump <domain> /path/to/core.file`

Thanks

Further reading:

- `/usr/src/linux/Documentation/kdump/`
- `http://www.dedoimedo.com/computers/crash-book.html`

Questions?

Comments and suggestions: `info@b1-systems.de`