Intro to Kernel Debugging: Just make the crashing stop!

Welcome. Here's what's in store if you stick around.

We'll Introduce:

- Gathering debug information
- Kernel development processes
- Oops analysis
- Code inspection
- Git tricks for finding fixes
- Engaging the kernel community
- How to dive deeper into debugging

We'll also cover a case study of a real-life XFS filesystem corruption bug

Dave Chiluk, Linux Platform Engineer at Indeed

indeed

Intro to Kernel Debugging:
Just Make the Crashing Stop!

Dave Chiluk

Linux Platform Engineer, Indeed

Intro to Kernel Debugging:

Just Make the Crashing Stop!

We'll Introduce:

- Gathering debug information
- Kernel development processes
- Oops analysis
- Code inspection
- Git tricks for finding fixes
- Engaging the kernel community
- How to dive deeper into debugging

We'll walk through a real-life case study.


indeed

We help people get jobs.

Dave Chiluk

- Indeed.com: Linux Platform Engineer
  - Fix issues in the open-source code that Indeed uses
- Canonical: Ubuntu Sustaining Engineering
  - Supported Customers with Ubuntu Kernel problems
  - Ubuntu core developer

Pets vs. Cattle

Scale Up

- Servers are like pets
- You name them, and when they get sick, you nurse them back to health

Scale Out

- Servers are like cattle
- You number them, and when they get sick, you shoot them

Bill Baker, Distinguished Engineer, Microsoft

- Scaling SQL Server 2012 by Glenn Berry

Pets vs. Cattle… and Wolves

Scale Up

- Servers are like pets
- You name them, and when they get sick, you nurse them back to health

Scale Out

- Servers are like cattle
- You number them, and when they get sick, you shoot them
- **If wolves start eating too many cattle, shoot them too.**

Case Study: An xfs crash

- SYSENG-2163: Rekick dc1-srv3
  Description: "We need to rekick dc1-srv3 due to filesystem corruption. Currently this host is in downtime and removed from the mesos cluster."
- SYSENG-2336: Rekick dc2-srv11
- SYSENG-2342: Rekick dc2-srv13
- SYSENG-2398: replace dc1-srv3; /var is corrupt for the n'th time
- SYSENG-2624: Rekick dc3-srv7: /var corrupted
- SYSENG-2723: Rekick dc1-srv6
- SYSENG-2770: /var corrupted on dc1-srv16
- SYSENG-2802: Fix the corrupt disk issue on dc1-srv10 (kernel bug)
- SYSENG-2849: dc4-srv5/6 var corruption
- SYSENG-2850: Monitor on corrupt filesystems
- SYSENG-3056: dc2-srv28 /var corruption by xfs bug

Step 1: Gather Information

Step 1: Gather Information

- **Kernel Version**
- **Logs**
  - First Oops
  - /var/log
  - Console output
  - rsyslog if necessary
- crashdump
- `sosreport`
- `sar`

```
XFS (dm-4): Internal error XFS_WANT_CORRUPTED_GOTO at line 3505 of
file fs/xfs/libxfs/xfs_btree.c.  Caller xfs_free_ag_extent+0x35d/0x7a0
[xfs]
CPU: 18 PID: 9896 Comm: mesos-slave Not tainted
4.10.10-1.el7.elrepo.x86_64 #1
Hardware name: Supermicro PIO-618U-TR4T+-ST031/X10DRU-i+, BIOS 2.0
12/17/2015
Call Trace:
dump_stack+0x63/0x87
xfs_error_report+0x3b/0x40 [xfs]
? xfs_free_ag_extent+0x35d/0x7a0 [xfs]
xfs_btree_insert+0x1b0/0x1c0 [xfs]
xfs_free_ag_extent+0x35d/0x7a0 [xfs]
xfs_free_extent+0xbb/0x150 [xfs]
xfs_trans_free_extent+0x4f/0x110 [xfs]
? xfs_trans_add_item+0x5d/0x90 [xfs]
xfs_extent_free_finish_item+0x26/0x40 [xfs]
xfs_defer_finish+0x149/0x410 [xfs]
xfs_remove+0x281/0x330 [xfs]
xfs_vn_unlink+0x55/0xa0 [xfs]
vfs_rmdir+0xb6/0x130
do_rmdir+0x1b3/0x1d0
SyS_rmdir+0x16/0x20
do_syscall_64+0x67/0x180
entry_SYSCALL64_slow_path+0x25/0x25
RIP: 0033:0x7f85d8d92397
RSP: 002b:00007f85cef9b758 EFLAGS: 00000246 ORIG_RAX: 0000000000000054
RAX: ffffffffffffffda RBX: 00007f858c00b4c0 RCX: 00007f85d8d92397
RDX: 00007f858c09ad70 RSI: 0000000000000000 RDI: 00007f858c09ad70
RBP: 00007f85cef9bc30 R08: 0000000000000001 R09: 0000000000000002
R10: 0000006f74656c67 R11: 0000000000000246 R12: 00007f85cef9c640
R13: 00007f85cef9bc50 R14: 00007f85cef9bcc0 R15: 00007f85cef9bc40
XFS (dm-4): xfs_do_force_shutdown(0x8) called from line 236 of file
fs/xfs/libxfs/xfs_defer.c.  Return address = 0xffffffffa028f087
XFS (dm-4): Corruption of in-memory data detected.  Shutting down
filesystem
XFS (dm-4): Please umount the filesystem and rectify the problem(s)
```

Step 1: Gather Information

For Subsystem Issues

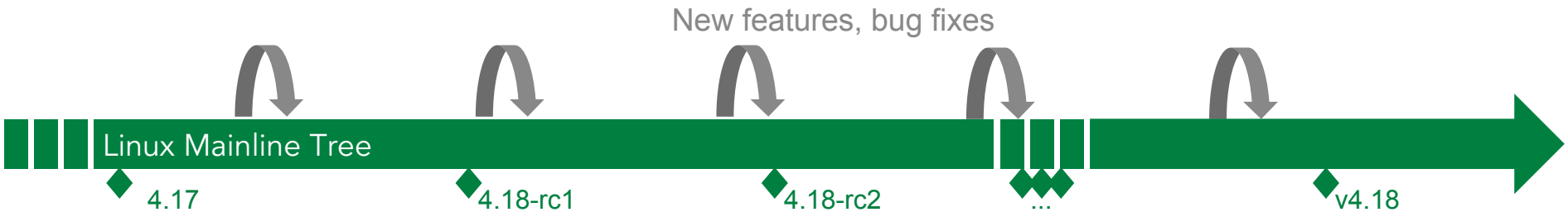
Device Driver Errors	Network Errors	Filesystem Errors
Firmware level?	`tcpdump`	Dump the filesystem - `dd`, physical removal
Module arguments?	`wireshark`	`xfs-metadump`
`modinfo` output		`xfs-restore`

Many others ...

Step 2: Get the Sources

Find the exact sources
used to build *your kernel*.

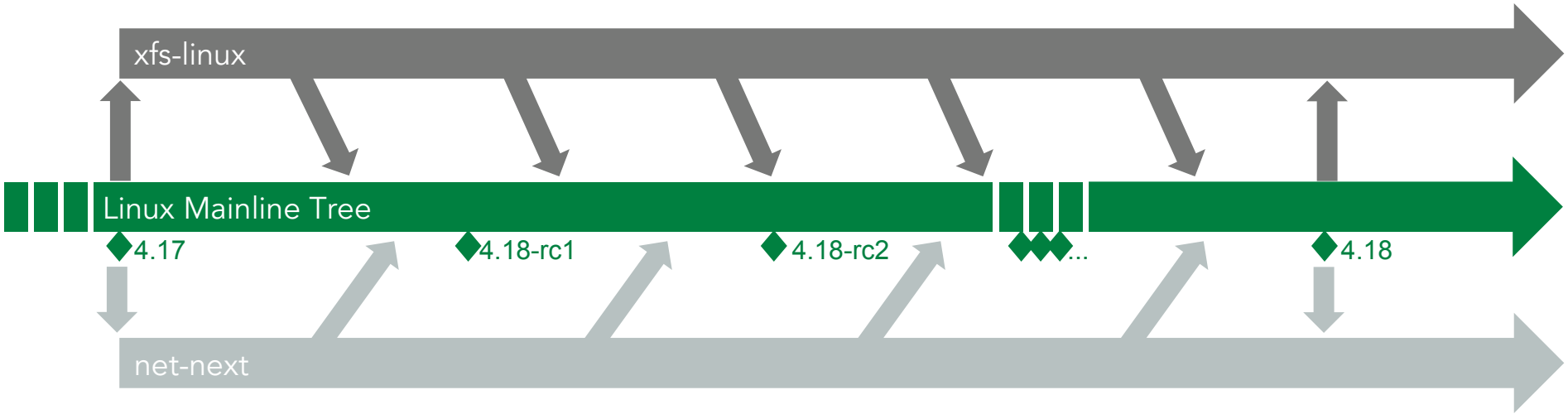
Get the Sources: Kernel Development



Mainline Kernel Development

- **All active kernel development eventually gets merged here**
- **git://git.kernel.org/pub/scm/linux/kernel/git/torvalds/linux.git**
- **Currently Maintained by Greg Kroah-Hartman (previously Linus Torvalds)**
- 14432 Patches from v4.17 (June 3, 2018) to v4.18 (Aug 12, 2018) **- 10 WEEKS!**

Get the Sources: Kernel Development



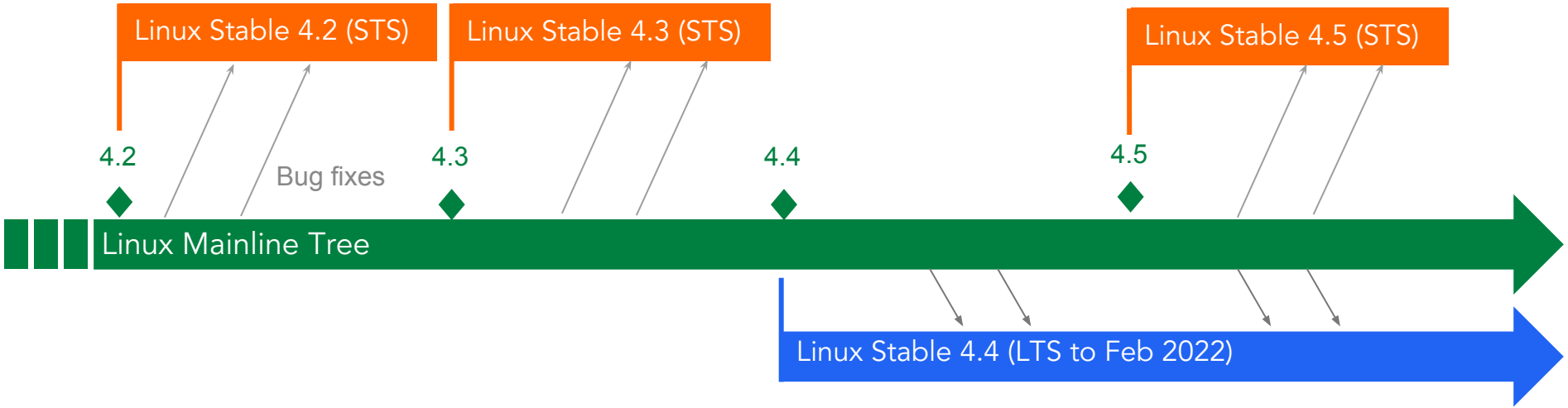
Subsystem Development trees

- Maintained by “Lieutenant” Subsystem Maintainers
- XFS https://git.kernel.org/pub/scm/fs/xfs/xfs-linux.git
- Networking git://git.kernel.org/pub/scm/linux/kernel/git/davem/net-next.git

Get the Sources

Stable Kernels

- Release kernels + bug fixes
- Maintained by Greg Kroah-Hartman
- git://git.kernel.org/pub/scm/linux/kernel/git/stable/linux-stable.git
- One repository many branches.
- Short-term stable and long-term stable



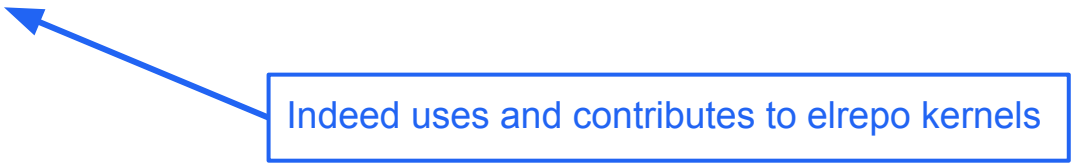
Step 2: Get the Sources

Mainline or Stable Kernels

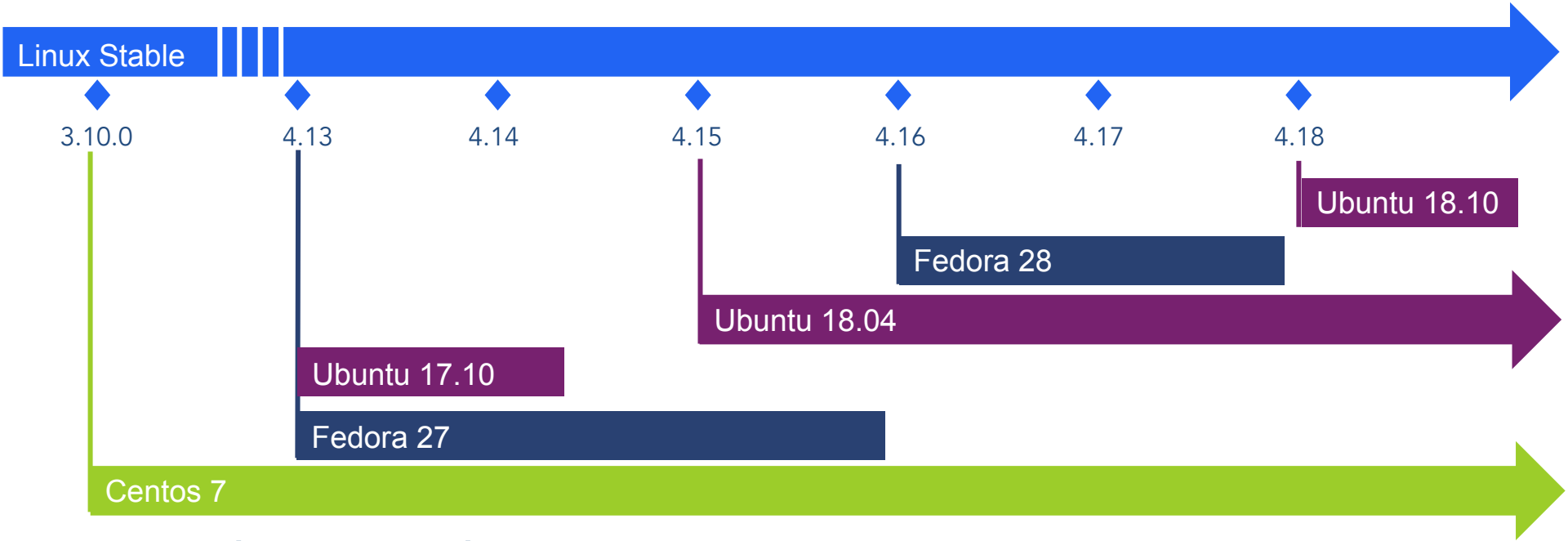
- `git clone git://git.kernel.org/pub/scm/linux/kernel/git/torvalds/linux.git`
  `make old config`
  `make; make modules_install; make install`
- `update-initramfs`
- `update-grub`

Prebuilt mainline kernels are available for many distributions

- https://wiki.ubuntu.com/Kernel/MainlineBuilds
- http://elrepo.org - Centos linux-stable kernels.



Get the Sources



Distribution Kernels

- Typically branched from Linux-stable kernels, but not necessarily LTS kernels
- Follow linux-stable process + feature work
- Own maintainers

Step 2: Get the Sources

Centos

- ```
  $ git clone https://git.centos.org/summary/rpms!kernel.git
  $ git clone https://git.centos.org/git/centos-git-common.git
  $ cd kernel && git checkout c7
  $ centos-git-common/get_sources.sh
  $ rpm-build -ba
  ```
- Provides an RPM-centric source tree, a source tarball, and a bunch of individual patches that are applied
- **This is not a “real” git repository**
  ```

Step 2: Get the Sources

Ubuntu / Debian

- apt-get source linux-image-$(uname-r)
- http://kernel.ubuntu.com/git/?q=ubuntu%2F
- git clone git://kernel.ubuntu.com/ubuntu/ubuntu-*
- rebuild

```
$ fakeroot debian/rules binary-generic
```

Step 2: Get the Sources

Debug Information

Debuginfo is unstripped versions of vmlinux

- This is needed if you want to run crash against a crashdump or do register analysis against the stack trace in your oops.

Centos/ RHEL

- yum --enablerepo=base-debuginfo install -y kernel-debuginfo-$(uname -r)

Ubuntu

- https://wiki.ubuntu.com/Debug%20Symbol%20Packages

```
$ sudo apt install linux-image-$(uname -r)-dbgsym
```

Step 2: Get the Sources

Kernel Structure

Start here!

```
documentation/
    process/ - how to interact with the community
    admin-guide/ - the manual
    admin-guide/bug-hunting.rst
mm/ - Memory Management
net/ - Network
fs/ - Filesystems
arch/ - Architecture Specific
drivers/ - Device Drivers
firmware/ - Binary Blobs
scripts/
    get_maintainer.pl
    checkpatch.pl
```

Step 3: Oops Analysis

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at
(null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc
videobuf2_memops videobuf2_v4l2 videobuf2_core hid_a4tech videodev
x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl
btintel bluetooth kvm_intel snd_hda_codec_hdmi kvm
snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw
ath i2c_algo_bit snd_hwdep mac80211 ghash_clmulni_intel
snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci
drm_kms_helper syscopyarea sysfillrect sysimgblt fb_sys_fops drm
xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169
rfkill mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video
soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206
08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>]
__mutex_lock_slowpath+0x98/0x120
RSP: 0018:ffffc900014cfce0  EFLAGS: 00010282
RAX: 0000000000000000 RBX: ffff8801a54315b0 RCX: 00000000c0000100
RDX: 0000000000000001 RSI: 0000000000000000 RDI: ffff8801a54315b4
RBP: ffffc900014cfd30 R08: 0000000000000000 R09: 0000000000000002
R10: 0000000000000000 R11: 0000000000000000 R12: ffff8801a54315b4
R13: ffff8801a639ba00 R14: 00000000ffffffff R15: ffff8801a54315b8
FS:  00007faa254fb700(0000) GS:ffff8801aef80000(0000)
knlGS:0000000000000000
CS:  0010 DS: 0000 ES: 0000 CR0: 0000000080050033
CR2: 0000000000000000 CR3: 00000001a3b1b000 CR4: 00000000001406e0
```

```
Stack:
ffff8801a54315b8 0000000000000000 ffffffff814733ae ffffc900014cfd28
ffffffff8146a28c ffff8801a54315b0 0000000000000000 ffff8801a54315b0
ffff8801a66f3820 0000000000000000 ffffc900014cfd48 ffffffff816c73e7
Call Trace:
[<ffffffff814733ae>] ? acpi_ut_release_mutex+0x5d/0x61
[<ffffffff8146a28c>] ? acpi_ns_get_node+0x49/0x52
[<ffffffff816c73e7>] mutex_lock+0x17/0x30
[<ffffffffa00a3bb4>] asus_rfkill_hotplug+0x24/0x1a0 [asus_wmi]
[<ffffffffa00a4421>] asus_wmi_rfkill_exit+0x61/0x150 [asus_wmi]
[<ffffffffa00a49f1>] asus_wmi_remove+0x61/0xb0 [asus_wmi]
[<ffffffff814a5128>] platform_drv_remove+0x28/0x40
[<ffffffff814a2901>] __device_release_driver+0xa1/0x160
[<ffffffff814a29e3>] device_release_driver+0x23/0x30
[<ffffffff814a1ffd>] bus_remove_device+0xfd/0x170
[<ffffffff8149e5a9>] device_del+0x139/0x270
[<ffffffff814a5028>] platform_device_del+0x28/0x90
[<ffffffff814a50a2>] platform_device_unregister+0x12/0x30
[<ffffffffa00a4209>] asus_wmi_unregister_driver+0x19/0x30 [asus_wmi]
[<ffffffffa00da0ea>] asus_nb_wmi_exit+0x10/0xf26 [asus_nb_wmi]
[<ffffffff8110c692>] SyS_delete_module+0x192/0x270
[<ffffffff810022b2>] ? exit_to_usermode_loop+0x92/0xa0
[<ffffffff816ca560>] entry_SYSCALL_64_fastpath+0x13/0x94
Code: e8 5e 30 00 00 8b 03 83 f8 01 0f 84 93 00 00 00 48 8b 43 10 4c
8d 7b 08 48 89 63 10 41 be ff ff ff ff 4c 89 3c 24 48 89 44 24 08
<48> 89 20 4c 89 6c 24 10 eb 1d 4c 89 e7 49 c7 45 08 02 00 00 00
RIP  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
RSP <ffffc900014cfce0>
CR2: 0000000000000000
---[ end trace 8d484233fa7cb512 ]---
note: modprobe[3275] exited with preempt_count 2
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
BUG: unable to handle kernel NULL pointer dereference at           (null)
IP: [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
PGD 1a3aa8067
PUD 1a3b3d067
PMD 0
Oops: 0002 [#1] PREEMPT SMP
Modules linked in: bnep ccm binfmt_misc uvcvideo videobuf2_vmalloc videobuf2_memops videobuf2_v4l2
videobuf2_core hid_a4tech videodev x86_pkg_temp_thermal intel_powerclamp coretemp ath3k btusb btrtl btintel
bluetooth kvm_intel snd_hda_codec_hdmi kvm snd_hda_codec_realtek snd_hda_codec_generic irqbypass crc32c_intel
arc4 i915 snd_hda_intel snd_hda_codec ath9k ath9k_common ath9k_hw ath i2c_algo_bit snd_hwdep mac80211
ghash_clmulni_intel snd_hda_core snd_pcm snd_timer cfg80211 ehci_pci xhci_pci drm_kms_helper syscopyarea
sysfillrect sysimgblt fb_sys_fops drm xhci_hcd ehci_hcd asus_nb_wmi(-) asus_wmi sparse_keymap r8169 rfkill
mxm_wmi serio_raw snd mii mei_me lpc_ich i2c_i801 video soundcore mei i2c_smbus wmi i2c_core mfd_core
CPU: 3 PID: 3275 Comm: modprobe Not tainted 4.9.34-gentoo #34
Hardware name: ASUSTeK COMPUTER INC. K56CM/K56CM, BIOS K56CM.206 08/21/2012
task: ffff8801a639ba00 task.stack: ffffc900014cc000
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
```

Step 3: Oops Analysis - The Toolkit

```
RIP: 0010:[<ffffffff816c7348>]  [<ffffffff816c7348>] __mutex_lock_slowpath+0x98/0x120
RSP: 0018:ffffc900014cfce0  EFLAGS: 00010282
RAX: 0000000000000000 RBX: ffff8801a54315b0 RCX: 00000000c0000100
RDX: 0000000000000001 RSI: 0000000000000000 RDI: ffff8801a54315b4
RBP: ffffc900014cfd30 R08: 0000000000000000 R09: 0000000000000002
R10: 0000000000000000 R11: 0000000000000000 R12: ffff8801a54315b4
R13: ffff8801a639ba00 R14: 00000000ffffffff R15: ffff8801a54315b8
FS:  00007faa254fb700(0000) GS:ffff8801aef80000(0000) knlGS:0000000000000000
CS:  0010 DS: 0000 ES: 0000 CR0: 0000000080050033
CR2: 0000000000000000 CR3: 00000001a3b1b000 CR4: 00000000001406e0
```

```
$ objdump -d -S -l ./vmlinux-4.15.0-22-generic > /tmp/objdump.out
```

```
ffffffff81992130 <__mutex_lock_slowpath>:
__mutex_lock_slowpath():
/build/linux-lZKWha/linux-4.15.0/kernel/locking/mutex.c:1130
ffffffff81992130:       e8 3b fb 06 00          callq  ffffffff81a01c70 <__fentry__>
ffffffff81992135:       55                      push   %rbp
/build/linux-lZKWha/linux-4.15.0/kernel/locking/mutex.c:1131
ffffffff81992136:       be 02 00 00 00          mov    $0x2,%esi
/build/linux-lZKWha/linux-4.15.0/kernel/locking/mutex.c:1130
ffffffff8199213b:       48 89 e5                mov    %rsp,%rbp
```

Step 3: Oops Analysis - The Toolkit

Register to argument mapping defined in:

[linux]/arch/x86/entry/calling.h

x86 function call convention, 64-bit:

```
arguments            |  callee-saved      | extra caller-saved | return
(callee-clobbered)   |                    | (callee-clobbered) |
-------------------- | -------------------| -------------------| -------------
rdi rsi rdx rcx r8-9 | rbx rbp [*] r12-15 | r10-11             | rax, rdx [**]
```

Step 3: Oops Analysis - The Toolkit

```
Stack:
ffff8801a54315b8 0000000000000000 ffffffff814733ae ffffc900014cfd28
ffffffff8146a28c ffff8801a54315b0 0000000000000000 ffff8801a54315b0
ffff8801a66f3820 0000000000000000 ffffc900014cfd48 ffffffff816c73e7
Call Trace:
[<ffffffff814733ae>] ? acpi_ut_release_mutex+0x5d/0x61
[<ffffffff8146a28c>] ? acpi_ns_get_node+0x49/0x52
[<ffffffff816c73e7>] mutex_lock+0x17/0x30
[<ffffffffa00a3bb4>] asus_rfkill_hotplug+0x24/0x1a0 [asus_wmi]
[<ffffffffa00a4421>] asus_wmi_rfkill_exit+0x61/0x150 [asus_wmi]
[<ffffffffa00a49f1>] asus_wmi_remove+0x61/0xb0 [asus_wmi]
[<ffffffff814a5128>] platform_drv_remove+0x28/0x40
[<ffffffff814a2901>] __device_release_driver+0xa1/0x160
[<ffffffff814a29e3>] device_release_driver+0x23/0x30
[<ffffffff814a1ffd>] bus_remove_device+0xfd/0x170
[<ffffffff8149e5a9>] device_del+0x139/0x270
[<ffffffff814a5028>] platform_device_del+0x28/0x90
[<ffffffff814a50a2>] platform_device_unregister+0x12/0x30
[<ffffffffa00a4209>] asus_wmi_unregister_driver+0x19/0x30 [asus_wmi]
[<ffffffffa00da0ea>] asus_nb_wmi_exit+0x10/0xf26 [asus_nb_wmi]
[<ffffffff8110c692>] SyS_delete_module+0x192/0x270
[<ffffffff810022b2>] ? exit_to_usermode_loop+0x92/0xa0
[<ffffffff816ca560>] entry_SYSCALL_64_fastpath+0x13/0x94
```

Step 3: Oops Analysis - The Toolkit

```
Stack:
ffff8801a54315b8 0000000000000000 ffffffff814733ae ffffc900014cfd28
ffffffff8146a28c ffff8801a54315b0 0000000000000000 ffff8801a54315b0
ffff8801a66f3820 0000000000000000 ffffc900014cfd48 ffffffff816c73e7
Call Trace:
[<ffffffff814733ae>] ? acpi_ut_release_mutex+0x5d/0x61
[<ffffffff8146a28c>] ? acpi_ns_get_node+0x49/0x52
[<ffffffff816c73e7>] mutex_lock+0x17/0x30
[<ffffffffa00a3bb4>] asus_rfkill_hotplug+0x24/0x1a0 [asus_wmi]
[<ffffffffa00a4421>] asus_wmi_rfkill_exit+0x61/0x150 [asus_wmi]
[<ffffffffa00a49f1>] asus_wmi_remove+0x61/0xb0 [asus_wmi]
[<ffffffff814a5128>] platform_drv_remove+0x28/0x40
[<ffffffff814a2901>] __device_release_driver+0xa1/0x160
[<ffffffff814a29e3>] device_release_driver+0x23/0x30
[<ffffffff814a1ffd>] bus_remove_device+0xfd/0x170
[<ffffffff8149e5a9>] device_del+0x139/0x270
[<ffffffff814a5028>] platform_device_del+0x28/0x90
[<ffffffff814a50a2>] platform_device_unregister+0x12/0x30
[<ffffffffa00a4209>] asus_wmi_unregister_driver+0x19/0x30 [asus_wmi]
[<ffffffffa00da0ea>] asus_nb_wmi_exit+0x10/0xf26 [asus_nb_wmi]
[<ffffffff8110c692>] SyS_delete_module+0x192/0x270
[<ffffffff810022b2>] ? exit_to_usermode_loop+0x92/0xa0
[<ffffffff816ca560>] entry_SYSCALL_64_fastpath+0x13/0x94
```

Step 3: Oops Analysis - The Toolkit

```
Stack:
ffff8801a54315b8 0000000000000000 ffffffff814733ae ffffc900014cfd28
ffffffff8146a28c ffff8801a54315b0 0000000000000000 ffff8801a54315b0
ffff8801a66f3820 0000000000000000 ffffc900014cfd48 ffffffff816c73e7
Call Trace:
[<ffffffff814733ae>] ? acpi_ut_release_mutex+0x5d/0x61
[<ffffffff8146a28c>] ? acpi_ns_get_node+0x49/0x52
[<ffffffff816c73e7>] mutex_lock+0x17/0x30
[<ffffffffa00a3bb4>] asus_rfkill_hotplug+0x24/0x1a0 [asus_wmi]
[<ffffffffa00a4421>] asus_wmi_rfkill_exit+0x61/0x150 [asus_wmi]
[<ffffffffa00a49f1>] asus_wmi_remove+0x61/0xb0 [asus_wmi]
[<ffffffff814a5128>] platform_drv_remove+0x28/0x40
[<ffffffff814a2901>] __device_release_driver+0xa1/0x160
[<ffffffff814a29e3>] device_release_driver+0x23/0x30
[<ffffffff814a1ffd>] bus_remove_device+0xfd/0x170
[<ffffffff8149e5a9>] device_del+0x139/0x270
[<ffffffff814a5028>] platform_device_del+0x28/0x90
[<ffffffff814a50a2>] platform_device_unregister+0x12/0x30
[<ffffffffa00a4209>] asus_wmi_unregister_driver+0x19/0x30 [asus_wmi]
[<ffffffffa00da0ea>] asus_nb_wmi_exit+0x10/0xf26 [asus_nb_wmi]
[<ffffffff8110c692>] SyS_delete_module+0x192/0x270
[<ffffffff810022b2>] ? exit_to_usermode_loop+0x92/0xa0
[<ffffffff816ca560>] entry_SYSCALL_64_fastpath+0x13/0x94
```


Top / most recent

Bottom / oldest

Step 3: Oops Analysis - The Toolkit

```
Stack:
ffff8801a54315b8 0000000000000000 ffffffff814733ae ffffc900014cfd28
ffffffff8146a28c ffff8801a54315b0 0000000000000000 ffff8801a54315b0
ffff8801a66f3820 0000000000000000 ffffc900014cfd48 ffffffff816c73e7
Call Trace:
[<ffffffff814733ae>] ? acpi_ut_release_mutex+0x5d/0x61
[<ffffffff8146a28c>] ? acpi_ns_get_node+0x49/0x52
[<ffffffff816c73e7>] mutex_lock+0x17/0x30
[<ffffffffa00a3bb4>] asus_rfkill_hotplug+0x24/0x1a0 [asus_wmi]
[<ffffffffa00a4421>] asus_wmi_rfkill_exit+0x61/0x150 [asus_wmi]
[<ffffffffa00a49f1>] asus_wmi_remove+0x61/0xb0 [asus_wmi]
[<ffffffff814a5128>] platform_drv_remove+0x28/0x40
[<ffffffff814a2901>] __device_release_driver+0xa1/0x160
[<ffffffff814a29e3>] device_release_driver+0x23/0x30
[<ffffffff814a1ffd>] bus_remove_device+0xfd/0x170
[<ffffffff8149e5a9>] device_del+0x139/0x270
[<ffffffff814a5028>] platform_device_del+0x28/0x90
[<ffffffff814a50a2>] platform_device_unregister+0x12/0x30
[<ffffffffa00a4209>] asus_wmi_unregister_driver+0x19/0x30 [asus_wmi]
[<ffffffffa00da0ea>] asus_nb_wmi_exit+0x10/0xf26 [asus_nb_wmi]
[<ffffffff8110c692>] SyS_delete_module+0x192/0x270
[<ffffffff810022b2>] ? exit_to_usermode_loop+0x92/0xa0
[<ffffffff816ca560>] entry_SYSCALL_64_fastpath+0x13/0x94
```

Step 3: Oops Analysis - The Toolkit

```
Code: e8 5e 30 00 00 8b 03 83 f8 01 0f 84 93 00 00 00 48 8b 43 10 4c 8d 7b 08 48 89 63 10 41 be ff ff ff ff
4c 89 3c 24 48 89 44 24 08 <48> 89 20 4c 89 6c 24 10 eb 1d 4c 89 e7 49 c7 45 08 02 00 00 00
```

Code: A hex-dump of the section of machine code that was being run at the time the Oops occurred.

```
$ echo "Code: e8 5e 30 00 00 8b 03 83 f8 01 0f 84 93 00 00 00 48 8b 43 10 4c 8d 7b 08 48 89 63 10 41 be ff ff
ff ff 4c 89 3c 24 48 89 44 24 08 <48> 89 20 4c 89 6c 24 10 eb 1d 4c 89 e7 49 c7 45 08 02 00 00 00
" | ~/src/linux/scripts/decodecode

Code: e8 5e 30 00 00 8b 03 83 f8 01 0f 84 93 00 00 00 48 8b 43 10 4c 8d 7b 08 48 89 63 10 41 be ff ff ff ff
4c 89 3c 24 48 89 44 24 08 <48> 89 20 4c 89 6c 24 10 eb 1d 4c 89 e7 49 c7 45 08 02 00 00 00
All code
========
   0:    e8 5e 30 00 00          callq  0x3063
   5:    8b 03                   mov    (%rbx),%eax
   7:    83 f8 01                cmp    $0x1,%eax
   a:    0f 84 93 00 00 00       je     0xa3
  10:    48 8b 43 10             mov    0x10(%rbx),%rax
  14:    4c 8d 7b 08             lea    0x8(%rbx),%r15
  18:    48 89 63 10             mov    %rsp,0x10(%rbx)
  1c:    41 be ff ff ff ff       mov    $0xffffffff,%r14d
  22:    4c 89 3c 24             mov    %r15,(%rsp)
  26:    48 89 44 24 08          mov    %rax,0x8(%rsp)
  2b:*   48 89 20                mov    %rsp,(%rax)        <-- trapping instruction
```

Case Study: Log Output

```
XFS (dm-4): Internal error XFS_WANT_CORRUPTED_GOTO at line 3505 of file fs/xfs/libxfs/xfs_btree.c.  Caller
xfs_free_ag_extent+0x35d/0x7a0 [xfs]
CPU: 18 PID: 9896 Comm: mesos-slave Not tainted 4.10.10-1.el7.elrepo.x86_64 #1
Hardware name: Supermicro PIO-618U-TR4T+-ST031/X10DRU-i+, BIOS 2.0 12/17/2015
Call Trace:
dump_stack+0x63/0x87
xfs_error_report+0x3b/0x40 [xfs]
? xfs_free_ag_extent+0x35d/0x7a0 [xfs]
xfs_btree_insert+0x1b0/0x1c0 [xfs]
xfs_free_ag_extent+0x35d/0x7a0 [xfs]
xfs_free_extent+0xbb/0x150 [xfs]
xfs_trans_free_extent+0x4f/0x110 [xfs]
? xfs_trans_add_item+0x5d/0x90 [xfs]
xfs_extent_free_finish_item+0x26/0x40 [xfs]
xfs_defer_finish+0x149/0x410 [xfs]
xfs_remove+0x281/0x330 [xfs]
xfs_vn_unlink+0x55/0xa0 [xfs]
vfs_rmdir+0xb6/0x130
do_rmdir+0x1b3/0x1d0
SyS_rmdir+0x16/0x20
do_syscall_64+0x67/0x180
entry_SYSCALL64_slow_path+0x25/0x25
RIP: 0033:0x7f85d8d92397
RSP: 002b:00007f85cef9b758 EFLAGS: 00000246 ORIG_RAX: 0000000000000054
RAX: ffffffffffffffda RBX: 00007f858c00b4c0 RCX: 00007f85d8d92397
RDX: 00007f858c09ad70 RSI: 0000000000000000 RDI: 00007f858c09ad70
RBP: 00007f85cef9bc30 R08: 0000000000000001 R09: 0000000000000002
R10: 0000006f74656c67 R11: 0000000000000246 R12: 00007f85cef9c640
R13: 00007f85cef9bc50 R14: 00007f85cef9bcc0 R15: 00007f85cef9bc40
XFS (dm-4): xfs_do_force_shutdown(0x8) called from line 236 of file fs/xfs/libxfs/xfs_defer.c.  Return address =
0xffffffffa028f087
XFS (dm-4): Corruption of in-memory data detected.  Shutting down filesystem
XFS (dm-4): Please umount the filesystem and rectify the problem(s)
```

Case Study: The Oops

```
XFS (dm-4): Internal error XFS_WANT_CORRUPTED_GOTO at line 3505 of file
 fs/xfs/libxfs/xfs_btree.c. Caller xfs_free_ag_extent+0x35d/0x7a0 [xfs]
CPU: 18 PID: 9896 Comm: mesos-slave Not tainted 4.10.10-1.el7.elrepo.x86_64 #1
```

Provides exact Kernel Version + file + line number !

```
Call Trace:
dump_stack+0x63/0x87
xfs_error_report+0x3b/0x40 [xfs]
? xfs_free_ag_extent+0x35d/0x7a0 [xfs]
xfs_btree_insert+0x1b0/0x1c0 [xfs]
xfs_free_ag_extent+0x35d/0x7a0 [xfs]
xfs_free_extent+0xbb/0x150 [xfs]
xfs_trans_free_extent+0x4f/0x110 [xfs]
? xfs_trans_add_item+0x5d/0x90 [xfs]
xfs_extent_free_finish_item+0x26/0x40 [xfs]
xfs_defer_finish+0x149/0x410 [xfs]
xfs_remove+0x281/0x330 [xfs]
xfs_vn_unlink+0x55/0xa0 [xfs]
vfs_rmdir+0xb6/0x130
do_rmdir+0x1b3/0x1d0
SyS_rmdir+0x16/0x20
do_syscall_64+0x67/0x180
entry_SYSCALL64_slow_path+0x25/0x25
```

Case Study: Code Inspection

```
XFS (dm-4): Internal error XFS_WANT_CORRUPTED_GOTO at line 3505 of file
 fs/xfs/libxfs/xfs_btree.c. Caller xfs_free_ag_extent+0x35d/0x7a0 [xfs]
CPU: 18 PID: 9896 Comm: mesos-slave Not tainted 4.10.10-1.el7.elrepo.x86_64 #1
```

```
3462 xfs_btree_insert(
...
3493         /*
3494          * Insert nrec/nptr into this level of the tree.
3495          * Note if we fail, nptr will be null.
3496          */
3497         error = xfs_btree_insrec(pcur, level, &nptr, &rec, key,
3498                 &ncur, &i);
3499         if (error) {
3500             if (pcur != cur)
3501                 xfs_btree_del_cursor(pcur, XFS_BTREE_ERROR);
3502             goto error0;
3503         }
3504
3505         XFS_WANT_CORRUPTED_GOTO(cur->bc_mp, i == 1, error0);
```

Case Study: Code Inspection

```
3241 xfs_btree_insrec(
3242     struct xfs_btree_cur    *cur,    /* btree cursor */
         ...
3248     int             *stat)  /* success/failure */
3249 {
         ...
3271     /*
3272      * If we have an external root pointer, and we've made it to the
3273      * root level, allocate a new root block and we're done.
3274      */
3275     if (!(cur->bc_flags & XFS_BTREE_ROOT_IN_INODE) &&
3276         (level >= cur->bc_nlevels)) {
3277         error = xfs_btree_new_root(cur, stat);
```

Case Study: Code Inspection

```
xfs_btree_insrec
|-> xfs_btree_new_root
    |-> cur->bc_ops->alloc_block(cur, &rptr, &lptr, stat);
        |-> xfs_allocbt_alloc_block
            |-> xfs_alloc_get_freelist
                |-> agfl_bno = XFS_BUF_TO_AGFL_BNO(mp, agflbp);
            if (bno == NULLAGBLOCK) {
                XFS_BTREE_TRACE_CURSOR(cur, XBT_EXIT);
                *stat = 0;
                return 0;
            }
```

Step 4: Check for Fixes

Step 4: Check for Fixes

Google

- Stack trace
- Mailing list archives
- Related terms given your understanding

Step 4: Check for Fixes

Check the git commit logs for fixes

```
$ git log -r v4.10.. -i --grep "crash" fs/xfs
$ git log -r v4.10.. -i --grep "agfl" fs/xfs
$ git log -r v4.10.. fs/xfs/libxfs/xfs_btree.c
```

Use git bisect to identify fixes already exist on mainline

```
$ git bisect start
$ git bisect bad <fixed version>
$ git bisect good <bad version>
```

Step 5: Gather Even More Info

Step 5: Gather Even More Info

- Enable crashdump
- Instrument the kernel code (kprintf) / build a custom kernel
- `systemtap`, ftrace, kprobes, ~~jprobes~~
- eBPF
- perf
- Analyze filesystem metadata offline

Step 6: Engage the Community

Step 6: Engage the Community

Mailing lists:

- ~~lkml~~
- Subsystem specific lists!
- patchwork.kernel.org - Mailing list hub

IRC

- Freenode
  - #xfs
  - ##kernel
  - #ubuntu-kernel
- OFTC
  - #kernelnewbies

Step 6: Engage the Community

E-mail sent to linux-xfs mailing list:

- Include full Oops output.
- Include any analysis you may have been able to do.

"My best guess given code analysis is that we are unable to allocate a new node in the allocation group free-list btree."

Response:

"Without xfs_repair output, … we have no idea whether this was caused by corruption or some other problem... If I had a dollar for every time I've seen this sort of error report, I'd have retired years ago."

- Dave Chinner

Step 6: Engage the Community

- **Gathered the requested information**
- **Responded via IRC freenode.net #xfs**

"This is the problematic issue: commit 96f859d52bcb ("libxfs: pack the agfl header structure so XFS_AGFL_SIZE is correct")... [I] need to resurrect the old patches [I] had that automatically detected this condition and fixed it."

- Dave Chinner

```
--- a/fs/xfs/libxfs/xfs_format.h
+++ b/fs/xfs/libxfs/xfs_format.h
@@ -786,7 +786,7 @@ typedef struct xfs_agfl {
        __be64          agfl_lsn;
        __be32          agfl_crc;
        __be32          agfl_bno[];     /* actually XFS_AGFL_SIZE(mp) */
-} xfs_agfl_t;
+} __attribute__((packed)) xfs_agfl_t;
```

Step 6.5: The Temporary Fix

Step 6.5: The Temporary Fix

- Explicitly removed problematic patch
- Submitted patch removal to the elrepo kernel
  - http://elrepo.org/bugs/view.php?id=829
  - http://elrepo.org/bugs/view.php?id=833
- Considered running xfs-repair on every /var volume in our cluster
- Proceeded to attempt to create a "fixup" patchset

Step 7: The Real Fix

Step 7: The Real Fix

4 patch rewrites on the `linux-xfs` mailing list

3 months later `a27ba2607` is accepted into the `xfs development` tree

Step 7: The Real Fix

Mainline

- A month later Linus merges patches from xfs-devel into 4.17rc1 during the 4.17 merge window.

Linux Stable

- Linux stable backport and submission to linux-stable mailing list.
- Greg KH accepts the patches and adds them to the stable-queue git tree for review.
- Eventually merged to stable branches such as 4.4.

Distribution Kernels

- Most distros follow Linux Stable guidelines.
- Check your distro just to make sure.

Don't forget to upgrade your cluster!

Questions?

OBJ