



LPC 2013

PCI Microconference

PCI Locking and Reference Counting

Bug Reports

Panic on concurrent sysfs remove of parent/child devices

bz 54411

Deadlock on concurrent sysfs remove/rescan

bz 60672

Lockdep warning on concurrent sysfs remove/rescan

bz 60695

Panic Scenario (bz 54411)

```
00:1c.0 Root Port          PCI bridge to [bus 01-03]
01:00.0 Upstream Port     PCI bridge to [bus 02-03]
02:00.0 Downstream Port   PCI bridge to [bus 03]
03:00.0 Endpoint
```

```
# echo 1 > 00:1c.0/remove; \  
  echo 1 > 03:00.0/remove
```

```
...
```

```
general protection fault:
```

```
RIP: pci_remove_bus_device+0x86/0x100
```

```
RAX: dead000000200200  RBX: 0000000000000000  RCX: dead000000100100  
RDX: dead000000100100  RSI: dead000000200200  RDI: ffffffff81c56640
```

sysfs Attribute Store Function

Schedule callback to avoid issues deleting current sysfs node:

```
remove_store(struct device *dev, ...)  
    device_schedule_callback(dev, remove_callback)  
    kobject_get  
    queue_work
```

Callback Performs Removal

Serialized by mutex:

```
remove_callback(struct device *dev, ...)  
{  
    mutex_lock(&pci_remove_rescan_mutex);  
    pci_stop_and_remove_bus_device(pdev);  
    mutex_unlock(&pci_remove_rescan_mutex);  
}
```

Removal and Deallocation

```
pci_stop_and_remove_bus_device
pci_stop_bus_device
...
pci_remove_bus_device
bus = dev->subordinate
list_for_each_entry(child, bus->devices, ...)
    pci_remove_bus_device(child)
pci_destroy_dev
    list_del(dev->bus_list)
    put_device
```

remove_store

device_schedule_callback(00:1c.0, remove_callback)

kobject_get(00:1c.0)

queue_work

remove_store

device_schedule_callback(03:00.0, remove_callback)

kobject_get(03:00.0)

queue_work

remove_callback(00:1c.0)

mutex_lock(...)

pci_stop_and_remove_bus_device(00:1c.0)

mutex_unlock(...)

remove_callback(03:00.0)

mutex_lock(...)

pci_stop_and_remove_bus_device(03:00.0)

mutex_unlock(...)