

SCSI

James Bottomley
SteelEye Technology

22 July 2003

How Did We Fare in the Last Year?

- Barriers—implemented but not really used (error handling still a problem)
- Persistent naming—Tried it in the kernel, didn't like it.
- Hotplug—still evolving.
- Command Allocation—Now under control, commands come dynamically from a slab.
- Lazy Device Allocation—no.
- User Level Scanning via hotplug—no.
- Queueing—All pathological queueing cases solved.

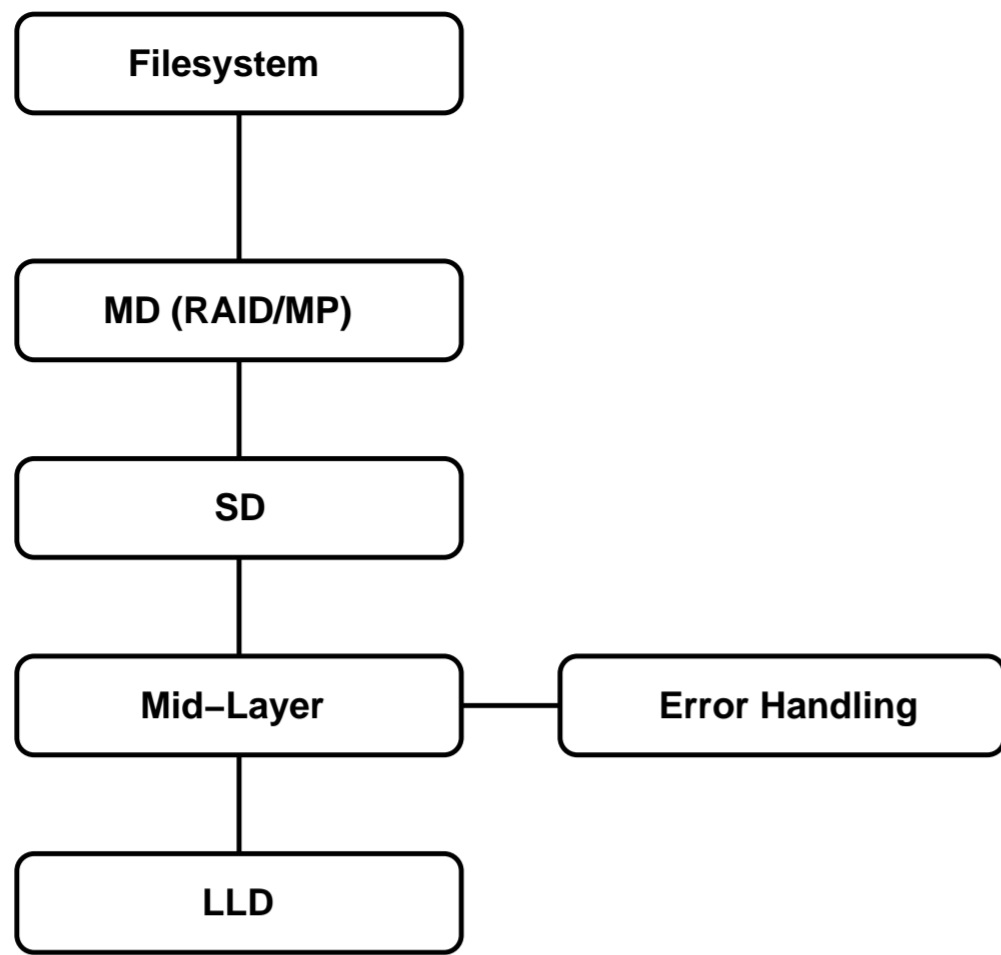
Hotplug

- Need to complete the recounting and get the actual model sorted out
- Main issue is stack can be deep: Unplug USB/Firewire device can give both SCSI host and device hot unplug block queue flush and filesystem unmount.
- Unsolved problem is who does what.
- Once we get hotplug sorted out, a lot of other solutions will fall into place.
 - parallel probing
 - spin up policy

Error Stacks

- This was mentioned last year
- For multi-path (and to a lesser extent, Software RAID and AIO) it is becoming critical
- Currently, bio contains no downward hints and return is essentially success/fail.
- Need
 - Downward indication of error handling required (from non, tell me every problem to don't bother me until the error becomes fatal).
 - Upwards. Need actual error plus translation indication (mainly fatal/retryable and whether the error is internal, transport or device).

Error Handler Stack



Migration to the Block Layer

- Already migrated quite a lot of SCSI code to the block layer.
- Most obvious is queueing (although, apparently no-one other than SCSI is using it).
- Have we gone as far as we can?

Barriers

- Originally designed to be used to protect transaction boundaries in the block layer
- These are supported by a **few** SCSI drivers.
- There's been no pressure to convert more
- Error handling problem for barriers still not solved
- Benchmarks still don't seem to show improvement for directly implemented barriers
- Are they still important?

Linked Commands

- Alternative view of transactions
- Instead of a long pipe with hard barriers, you link individual bios with a strict ordering directive.
- This can be coded directly into the SCSI command.
- Can also be simply abstracted for subsystems that don't do it (IDE).