Linux kernel v2.6

- Improvements relative to 2.4.x
- Regressions relative to 2.4.x
- New Features
- Filesystem-related
- Disk I/O
- Core kernel
- Drivers
- Networking
- Other
- As-yet unmerged new features
- Concluding overview

Andrew Morton <akpm@osdl.org> October 2003

Improvements relative to 2.4.x

- Multiprocessor scalability in the CPU scheduler, MM, FS and IO layers
  - 2.4 scales to 4 or 8 CPUS. 2.6 should scale to 16 or 32.
  - Scalability across large numbers of disks
  - Support for large amounts of memory (more than 8GB) on ia32 machines
  - Improved kernel build system
  - Improved kernel debugging support
  - Improved VM memory reclaim balancing
  - More robust handling of swapfiles
  - Reduced stalling in the memory allocator when under disk load
  - Enhanced direct IO (O_DIRECT) handling
  - New CPU scheduler (the “O(1) scheduler”)
  - Altered inode allocation algorithms in ext2 and ext3 (the “Orlov allocator”
  - Improved SMP scalability in ext2 and ext3
  - Rewritten “input” layer
  - 32-bit device numbering (was 16-bit). For 1000's of disks

Regressions relative to 2.4.x

- Worsened throughput under heavy swapping
- Global 1% slowdown due to (HZ) being increased from 100Hz to 1000 Hz
- Somewhat larger memory footprint

New features: filesystem-related

- SGI's XFS merged
- NTFS driver (read-only)
- IBM's JFS merged
- CIFS: new SMB client merged
- Hugetblfs: mmap-based access to large pages
- Basic support for version 4 of the NFS protocol
- NFS server over TCP
- Read-only AFS (Andrew File System) merged
- POSIX access control and extended attributes

New features: disk I/O

- Deadline I/O scheduler
- Anticipatory I/O scheduler
- Comprehensive disk I/O accounting
- 64-bit block numbers on 32-bit CPUs
- New “BIO” code: bypass the buffer_head layer for pagecache and direct IO

New features: core kernel

- Preemptible kernel
- Process <-> CPUset binding system calls
- Fast userspace mutexes (Futexes)
- Event poll system calls (epoll)
- Improved threading support
- Improved power management, including suspend-to-disk
- Sysfs: exposes the new driver, device and bus models
- POSIX timer support

New features: drivers

- ALSA sound drivers merged
- Improved CD recording, DMA support for ATAPI devices
- Intel 10 gigE “ixgb” driver merged

New features: networking

- Bridging firewall code merged
- New IPSec implementation
- RFC3173 IP Payload Compression
- IPVS (IP Virtual Server)
- SCTP (Stream Control Transport Protocol)
- NAPI

None of these are strictly “new” in 2.6: they have been or will be backported to 2.4.x

New features: other

- POSIX AIO infrastructure
- Oprofile kernel-based profiling support
- Logical Volume Manager rewritten, now called “device mapper”
- NSA security framework merged
- SELinux (Security Enhanced Linux) merged
- Generic cryptographic API library merged.
- Encrypted block loopback device driver merged
- User Mode Linux merged
- uCLinux for MMU-less CPUs: m68k, h8300, v850, SuperH
- x86 subarch support
- New x86 platforms: voyager, pc9800, x440, improved NUMAQ, etc.
- The separate MIPS32 and MIPS64 trees were merged together

New features: not yet merged

- CFQ (Complete Fair Queueing) disk scheduler
- AIO support for buffered disk I/O using internal retry model
- aio_poll, aio_fsync, etc...
- NFS O_DIRECT support
- Support for many more group Ids
- ia32 4G/4G address space split.

So what are the most important changes in 2.6?

- For servers: scalability
  - Much greater SMP scalability in filesystem, VM and scheduler
  - Support for attachment of more disks
  - Larger disks
  - Larger amounts of memory
  - Improved concurrency accessing multiple disks
  - Improved disk read-vs-read and read-vs-write performance
  - Faster threading support, more efficient NPTL locking

- For desktops: responsiveness
  - Dynamic priority adjustment in the CPU scheduler
  - 1000Hz system tick
  - Preemptible kernel
  - Reduced stalling under disk load (I/O scheduler, memory reclaim)