Linux Symposium 2007

Ottawa, Canada
June 27–30, 2007
Summarized by Rick Leir (rickleir@leirtech.com)

OLS is the conference for Linux kernel programmers, across the spectrum from embedded to large SMP systems. It also attracts application programmers and systems admins. Last year it was co-located with the Kernel Summit, but not this year. Attendees came from around the world. For me, travel arrangements were simple: The express city bus is convenient to me!

There were three tracks concurrent with tutorials, and several times I wanted to be in two places at a time. For a complete schedule see www.linuxsymposium.org/2007/. For a more detailed summary see www.linux.com/feature/115608. The attendees included a few hobbyists and academics, but most people were from companies including IBM, Intel, Sony, Red Hat, and AMD.

Jon Corbet gave his yearly Kernel Report. The trend is toward faster major releases. Where these used to be spaced by years, now they are spaced by months. The release cycle is more predictable than before, with a merge window of 2 weeks followed by 6 weeks for stabilization. This quickly moves changes out to users. Also, distributors (e.g., Red Hat, Ubuntu) are closer to the mainline. There is excellent tracking and merging of patches considering the volume (though some say quality was horrific for 2.6.21). There is ongoing work on automated testing.

For kernel 2.6.22, there will be:

- A new mac80211 wireless stack
- UBI flash-aware volume management
- An IVTV video tuner driver
- A new CFQ (Complete Fair Queueing) IO scheduler
- A firewire stack
- A SLUB memory management allocator (http://lwn.net/Articles/229984/)

In terms of scalability, SMP with 512 CPUs works well, and work on locks and page management processes for larger systems.

For filesystems, Jon observed that disks are getting larger but not faster, so fsck time can be a problem. New filesystems such as chunkfs and tilefs are more scalable and subdivide a disk so that only the active part needs to be fsck'd. The btrfs filesystem is extent-based, with subvolumes. It supports snapshot checksumming, online fsck, and faster fsck. Jon talked about ext4 with its 48-bit block numbers and use of extents. [Editor's note: Also see the article in the June 2007 issue of *;login:* about ext4.]

Jon finds reiser4 interesting, but the project is unfortunately stalled because Hans Reiser is no longer able to work on it. It needs a new champion.

Jon talked about virtualization. It is getting more attention, as shown by the many related presentations at this conference (e.g., KVM, lguest, Vserver).

The kernel is unlikely to go to GPL version 3 even if that was desired, because it is currently licensed with GPL version 2 and thousands of contributors would have to be contacted to make the change.

Jon's article has summaries of the Symposium and a summary of the work going into 2.6.22 (see http://lwn.net/Articles/240402/).

Greg Kroah-Hartman (www.kroah.com) taped to the back wall a 40-foot-long chart that linked together the people who have contributed patches. Developers were invited to sign the chart and about 100, of the 900 people who contributed to the 2.6.22 kernel, did so.

Mike Mason presented SystemTap, a dynamic tracing tool based on kprobes. Its simple scripting language provides a safe and flexible way to instrument a Linux system without modifying source code or rebooting.

There was considerable interest in embedded Linux. Robin Getz (blackfin.uclinux.org/) presented a tutorial on how to program for embedded systems with no MMU. I can't do uclinux justice here, but it seems to be the way to go.

Tim Chen talked about keeping kernel performance from regressions. He does weekly performance tests on the latest snapshot, and he occasionally sees large regressions. There are about 7000 patches per week, so it is not surprising that there would be problems. The 14 benchmarks include OLTP, an industry-standard Java business benchmark, cpu-int, cpu-fp, netperf, volanomark, lmbench, dbench, iozone, interbench, and httperf. The project is at kernel-perf.sourceforge.net/.

Arnaldo Carvalho de Melo talked about tools to help optimize kernel data structures. By rearranging the fields in structures you can avoid "holes" and thereby pack them into less memory. At times when related fields are close enough to be in the same cache line, performance improves. The pahole tool analyzes a struct and suggests field reordering.

Intel sponsored the reception Wednesday evening, and they demonstrated Ultra Wide Band (UWB) wireless networking (480 Mbps) between two laptops. Each was screening a video from the other's disk. This is a low-power technology to conserve battery power and avoid radio interference while being effective to 30 feet. UWB will be appearing in products soon.

Leonid Grossman talked about the challenges of 10Gb Ethernet. The transition to this is turning out to be more complex than earlier technology cycles. Part of the TCP stack is offloaded to the NIC TCP Offload Engine (TOE), so the kernel networking code has to change. I hear from osdl.org that there are significant problems with this.

Christopher James Lahey talked about Miro, which is a podcast client. He argues that "culture" is currently expressed via video, and we need a desktop app to search for video, display it, and organize channel folders or playlists. Miro uses Python, Pyrex, Javascript, CSS, and DOM. It uses some interesting database concepts. See more at getmiro.com.

Arnd Bergman of IBM talked about the Cell Broadband Engine, which is in Sony PS3 and IBM blades. This processor has the PowerPC Architecture with an L2 cache of 512 KB and 8 SPUs. The SPU is a co-processor that does fast floating-point math (though not so fast for double precision). There is a high-bandwidth bus (25 GB) connecting these processors, using explicit DMA. Each SPU has limited local memory (in effect, it executes out of its own cache), and overlay programming is used. Gcc emits DMA requests. Arnd evaluated the pros and cons of this package; on balance, it comes out very well.

From other sources I hear that the Sony PS3's Linux support involves a hypervisor that permits Linux to see only 6 of the SPUs. The NVIDIA video hardware is partly off limits to Linux programmers. Sony wants to interest the Open Source community in its products and very generously gave away several PS3s.

Andrew Cagney talked about Frysk, which is a user-level debugging tool for C and C++. It appears (my impression) to be as useful as Eclipse while being of considerably lighter weight. He talked about test-driven development and described a kernel regression test suite that has been discovering recent kernel bugs.

Jordan H. Crouse talked about using LinuxBIOS to speed up boot times and provide a more friendly boot environment. Be careful loading your motherboard flash memory.

Rusty Russell (ozlabs.org) entertained us with lguest, his simple virtualization project. He talked about how he went about coding lguest. He requires the guest OS to be the same version of Linux as the host, and his system does not support many of the features touted by the other virtual server systems, thereby saving much effort. He has my support!

Marcel Holtmann (bluez.org) talked about the latest Bluetooth tools and integration with Linux D-Bus. There was lots to talk about here, whether you are interested in the desktop or embedded devices.

Peter Zijlstra talked about the pagecache lock, which is not scalable. He has a way to avoid using a lock here. He alters the radix tree in order to support concurrent modifications of the data structure.

Jon "Maddog" Hall's ending keynote covered the Linux Terminal Server Project (LTSP.org) and how it could benefit poorer communities in the developing world. He made a convincing argument that this project was more practical than the One Laptop per Child (OLPC) project, and he showed a photo of himself in a school in Brazil. The terminals are not the VT200 of yore, but diskless Linux systems.