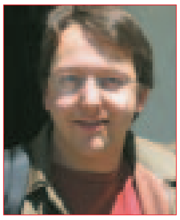
Zack's Kernel News

INFO

The Kernel Mailing List comprises the core of Linux development activities. Traffic volumes are immense and keeping up to date with the entire scope of development is a virtually impossible task for one person. One of the few brave souls that take on this impossible task is Zack Brown.



Our regular monthly column keeps you up to date on the latest decisions and discussions, selected and summarized by Zack. Zack has been publishing a weekly digest, the Kernel Traffic Mailing List for several years now, reading just the digest is a time consuming task.

Linux Magazine now provides you with the quintessence of Linux Kernel activities straight from the horse's mouth.

Replacement for TCP

Linux 2.5 will soon support SCTP (Stream Control Transmission Protocol), a general purpose networking protocol that attempts to solve some problems encountered with standard TCP (Transmission Control Protocol). David S. Miller recently promised to merge the existing SCTP patch into the main kernel source tree.

There are several reasons why users have been looking for a replacement for TCP. While TCP controls the order of data transmission, some applications that do not require strict data ordering must suffer unnecessary delays as TCP blocks to ensure proper ordering. In addition, TCP is more vulnerable to denial-of-service (DoS) style attacks. SCTP attempts to answer these and some other drawbacks.

SCTP was originally developed by the IETF (Internet Engineering Task Force) SIGTRAN (signal transport) working group to transport SS7 (Signalling System 7) over IP, but it may also be used as a general purpose protocol. ■

Hyperthreading

Recently, there has been a big push to support hyperthreading in 2.4 and 2.5 kernels. Hyperthreading is a bit like the opposite of Symmetric Multiprocessing. Instead of using multiple CPUs as one, hyperthreading treats a single CPU as many, with some interesting performance boosts.

Currently the only processor that officially supports hyperthreading is the Pentium 4 XEON, but it is rumored that other P4s can turn on hyperthreading in their BIOS.

Hyperthreading made its first appearance in the Linux kernel in November of 2001 in kernel 2.4.14, under the name SMT (Symmetric Multithreading). At the time, very few developers knew what to make of it, and there was much speculation. Intel claimed a 30% performance boost under their unreleased, proprietary benchmarks, but this was taken generally with a pinch of salt.

At the time, no available hardware supported hyperthreading. Only when the P4 XEON came out was there a possibility of wide-scale testing and development of this feature.

In recent weeks, many large patches have appeared, and seem to be making their way into the main kernel tree. Ingo Molnar made a big splash with his patch to integrate hyperthreading with his new scheduler code. One problem with SMP systems is that if more than one OOPS occurs simultaneously, they could overwrite each other, destroying the evidence needed to debug them. David Howells has been working on a patch to force all OOPS output to wait its turn before dumping to the screen. There is still some question as to whether his implementation is quite right; but it seems clear that he is on the right track, and that this code will be a welcome addition to the main tree.

OOPS reports contain essential information about what the system was doing just before a crash. When decoded by the ksymoops program, an OOPS can provide developers with a valuable clue in the hunting down and fixing of an elusive bug.

There are a number of problems with trying to capture OOPSes, and developers are always trying to expand their possible options. The main problem is that the system has crashed, and so there are only a limited number of behaviors that can be counted on. The OOPS code must do its best to generate a useful OOPS report in an environment in which much of the system may not be operational. There have previously been patches to dump OOPS output to a file, to send it over a serial port or even across the network; now there are patches to deal with multiple simultaneous OOPSes. ■

User mode linux in kernels

Jeff Dike's User-Mode Linux has finally made it into the official 2.5 kernel tree. UML is a patch to allow the Linux kernel to run as a user process, creating one or more virtual computers running simultaneously on a given system. UML recently became self-hosting, meaning that users may run UML from within a running UML process. Kernel version 2.5.35 is the first to contain the full incorporated UML patch.

There are many uses for this feature. Because UML is a user process, it now becomes possible to test each new kernel versions as UML invocations, without the risk of crashing the whole system. This saves the developers time that would otherwise be wasted by having to reboot the computer system after each failed test.

Another use for UML is in clustering. It has long been recognized by top developers that extending SMP to more than a few processors will result in tremendous complexity of the kernel's locking mechanisms. To avoid this, a number of alternatives have been actively pursued for some time.

One is the idea of SMP clusters, widely promoted by Larry McVoy; another is that UML may be a natural way to bridge multiple systems. Jeff has reported some success with his initial experiments, but the final direction of Linux clustering beyond SMP remains to be seen. ■

khttpd webserver is out

The controversial khttpd web server has been removed from the 2.5 kernel tree. Khttpd was written in response to the 1999 Mindcraft benchmarks that showed MS Windows serving web pages faster than Linux under certain conditions.

Although most Linux developers dismissed the benchmark as highly slanted, they were forced to admit that under the conditions of the test, MS Windows did out-perform Linux. To counter this, Linus Torvalds accepted the khttpd web server into the main kernel tree. This caused much violent protest, as a web server does not properly belong in kernel space.

Linus felt that it was important to beat the Mindcraft benchmark, however, and so the patch stayed. In recent months, however, a new user-space web server, Tux2, has consistantly out-performed khttpd, making the khttpd's presence in the kernel superfluous. Khttpd has also been unmaintained for some time, making the decision to eventually remove it somewhat easier.

Unfortunately, Tux2 is plagued by intellectual property disputes that no one seems inclined to fight over. Among other things, these disputes prevent the Tux2 web server from replacing khttpd in the kernel.

Some may argue that this is not a bad thing, but the fact remains that there are still many open questions surrounding a viable Linux webserver. One thing is certain: khttpd is gone. ■

XFS journaling filesystem

The journaled filesystem XFS has finally made it into the official 2.5 kernel tree. This has been a controversial project, with many folks arguing for XFS inclusion for a long time, and others saying the code was not ready yet. Kernel 2.5.36 is the first to contain XFS. SGI has been the main developer of XFS, and has been pushing for inclusion in the main tree for some time. Linux distributions such as Mandrake, SuSE and Slackware, have come bundled with the XFS patch for some time as well.

Linus Torvalds had refused to do the merge in the official tree because he felt there were certain implementation details that would have a negative impact on the rest of the system, and he wanted SGI to fix those details before he'd accept their patch. Ext3, ReiserFS, and JFS (from IBM), are examples of other journaling filesystems that have previously been accepted into the main kernel tree.

Journaling filesystems track all disk writes, and make sure that the filesystem is never in an inconsistant state. This means that a system crash will not require running fsck to bring the filesystem back into a usable state. Assuming all user data has been synchronized to disk, it is possible, with a journaling filesystem, to turn off the power, without fear of losing data.

While the ext2 filesystem remains the default on most Linux systems, it is only a matter of time before a journaling filesystem supplants it. ■

Netware filesystem sold to the Canopy Group

Timpanogas, a long-time contributor to the Linux kernel, has sold its intellectual property, including the Netware Filesystem, to the Canopy Group. Jeff V. Merkey, head of Timpanogas, would not specify which Timpanogas Linux project, if any, would continue under the new management.

Jeff and his company have been fairly controversial since they first became involved in Linux kernel development years ago. For a long time Jeff was regarded by many as something of a crackpot, but he managed to gain some measure of recognition for his technical skill and his ability to gain useful information from recalcitrant companies.

Andre Hedrick, the Linux kernel IDE maintainer, worked for him briefly at Timpanogas, but left after an apparent falling out between them.

The Canopy Group appears to be some sort of incubator of open source and Internet infrastructure companies. Their list of companies includes Linux Networx and Trolltech. It remains unclear how the Canopy Group intends to make use of the Timpanogas intellectual property. ■