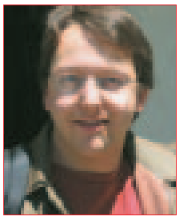
Zack’s Kernel News

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.

Gentle Hands for IDE coding

Last month I explained to you that the IDE code was being rewritten in 2.5 by Marcin Dalecki, amid heated controversy. Recently, Marcin decided to give up the fight, and all of his code changes have been removed from the 2.5 kernel tree.

A new set of changes by Andre Hedrick and others, that had been in development in the 2.4 kernel tree, have been forward-ported to 2.5; Alan Cox, although not the primary IDE developer, has agreed to take on the role of maintainer for the moment.

Andre, who would otherwise have been the obvious choice as the maintainer, demands such gentle personal handling that Linus Torvalds has found him impossible to work with. As a result of this Andre and everyone else working on IDE, will feed their changes to Alan, who will pass them along to Linus.

The IDE layer has been a problem in the kernel for quite awhile, and most developers agree it has been brought to an unmaintainable mess over the years, which may explain why there is so much contention around it.

One of the reasons Marcin came under such heavy fire was because of his uncompromising insistence on ripping out all of the broken code pieces, regardless of whether working replacements for the removed code were available or not.

The standards documents themselves may also be at fault; IDE discussions on the linux-kernel mailing list often examine the standards line by line in detail, and still lead to no clear agreement of what was intended by the standards body.

Andre, who has worked very closely with the standard bodies, claims to understand both the letter and the spirit of their documents; unfortunately he seems so far to be unable to share information without hurling insults at the people he is informing.

Speakers start to talk

The PC Speaker driver has been broken in the 2.5 kernel for a long time, and is finally receiving some attention from Stas Sergeev. But since the breakage was the result of correct changes to the Virtual Filesystem subsystem, as opposed to bad changes in the speaker driver itself, the new work has involved more than just bug fixing, and has been a long time coming.

So far, reports have come in that MP3s play well using the driver, but there have also been reports of other noise intruding on the proper sound. While Stas feels that this is almost certainly a problem with specific broken motherboards, and not a bug in his code, there are apparently other problems that may keep his driver out of the main kernel tree. For one thing, many modern motherboards come with soundcards already on them, making a speaker driver superfluous.

For another thing, the standards governing PC speaker hardware are weak, so that the great variety of possible configurations makes it difficult for any speaker driver to get the best performance out of the speaker. And finally, Stas’ driver in its current form uses a ton of CPU time. Stas feels this last is only a minor objection, since there is room to further improve his code. But he also points out that some motherboards are still made without sound cards, in which case the speaker would be the only source of sound on the computer. But he does agree that his code as it stands isn’t ready for inclusion in the main tree.

Version 4 this way comes

NFSv4 Is coming to the kernel. A number of developers have been working on this for awhile, and patches have begun cropping up for both the 2.4 and 2.5 trees. Now that some of the initial patches have laid the groundwork, Kendrick M. Smith has started to feed Linus and Marcelo patches that implement the actual server code.

NFSv4 seeks to answer some of the objections to earlier NFS versions, and to extend it further into new areas that were not taken into account when previous versions were designed. In particular, NFSv4 promises support for IPv6, strong security, good cross-platform interoperability, and in general, support for a range of extensions in other protocols. NFSv4 also promises to maintain the best features of earlier NFS versions, such as easy recovery and independence from particular transport protocols.

Mounting a networked filesystem is inherently tricky. It is difficult, for example, for the operating system to be certain not to reuse inodes. A duplicate inode can cause data loss or corruption, and the difficulties involved in reducing the risk of duplicate inodes in NFS has been the cause of much head shaking among kernel developers.

Latency issues have also plagued developers over the years, especially the question of how to be certain that rapid or nearly simultaneous changes at one end of the network connection are accurately represented to the user at the other end. Hopefully NFSv4 will address these issues as well.

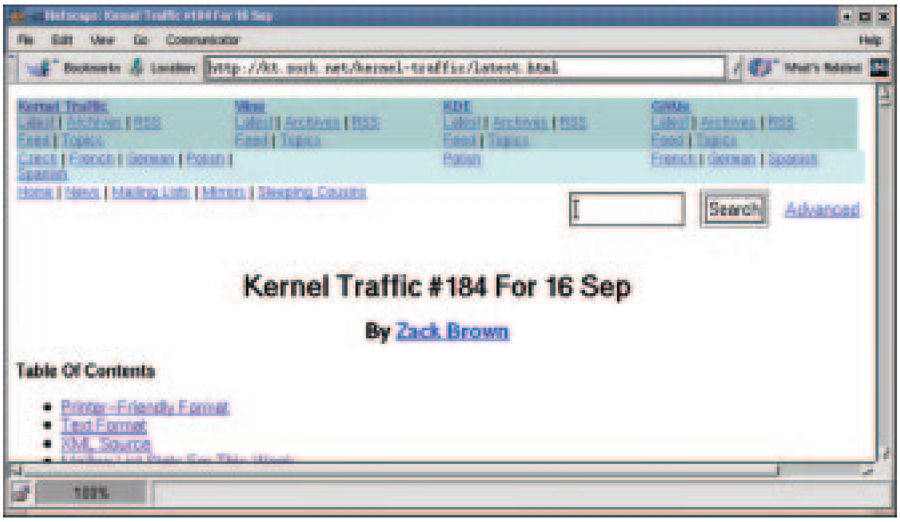
Lack of standards

POSIX compliance has always been a problem, mainly because Linus and the rest of the kernel developers never hesitate to abandon a standard if they feel it makes no sense. This was illustrated long ago in the clone wars, in which Linus eventually compromised by implementing POSIX thread-creation on top of semantics that he believed made much better sense.

Linus recently characterized POSIX compliance in these words: “POSIX is a hobbled standard, and does not matter. We’re not making a ‘POSIX-compliant OS’. People have done that before: see all the RT-OS’s out there, and see even the NT POSIX subsystem. They are uninteresting. Linux is a _real_ OS, not some ‘we filled in the paperwork and it is now standards compliant’.” The question is then, what does constitute a standard to which Linux adheres?

This is important for systems that wish to be Linux-compatible. If an OS wishing to be Linux-compatible must rely only on the current state of the Linux code, it will be difficult to guarantee that compatibility won’t be broken in the next kernel release. But it seems that any notion of a true “Linux standard” has not yet solidified.

Certainly POSIX and other legacy semantics play a large part. But it is not the final word. It may be one of the great strengths of Linux that it is unwilling to bow to tradition; but non-compliance to standards is also an accusation frequently made by free-software proponents against proprietary software companies like Microsoft. ■


File Edit View Go Communicator Help
Bookmarks Location: http://kt.zork.net/kernel-traffic/latest.html
What's Related
Kernel Traffic
Latest | Archives | RSS
Feed | Topics
Wine
KDE
GNUe
Home | News | Mailing Lists | Mirrors | Sleeping Cousins
Search
Advanced
100%

Kernel Traffic #184 For 16 Sep

By Zack Brown

Table Of Contents

- Printer-Friendly Format
- Text Format
- XML Source

Figure 1: Zack’s Kernel-Traffic web site

Raising the disk limit

It should soon be possible to support 128 or 256 SCSI disks on a single system, in both 2.4 and 2.5 kernels. Kurt Garloff posted some patches to do this in 2.4; and while Alexander Viro felt that these patches were not going to slide neatly into 2.5, Kurt felt that certain parts of the patch would not be too much trouble to forward-port.

EVMS could probably bring the full functionality to 2.5 without a problem, but Linus Torvalds has made it clear at the recent Kernel Summit, that he did not want EVMS to continue encroaching on the block layer’s domain. A number of key developers seem to be in favour of pursuing Kurt’s work with an eye toward acceptance into the 2.4 and 2.5 trees.

Raising the maximum number of SCSI disks is a long-standing problem. Solutions were being proposed as far back as 1992, when Linux was barely a year old. Richard Gooch offered patches in late 2001 to raise the maximum number to over 2000 disks, but his patches were not accepted. This latest attempt by Kurt shows the most promise of actually being accepted, though of course, the task will then be to raise the limit still further.

The quest to support bigger, taller systems is ongoing. Large memory, many processors, large files, large filesystem, large disks, large numbers of disks; at every level, developers struggle to support big systems, while still continuing to support smaller desktops and older hardware. ■

Taking the guess out of benchmarks

A new tool for benchmarking the Virtual Memory subsystem has emerged: VM Regress, by Mel Gorman. It is still in the early stages of development, but it’s already useful.

VM Regress has the ambitious goal of “eventually eliminating guesswork in development.” Although developed for 2.4 kernels, it compiles under the 2.5 kernel as well. The tool is not intended to benchmark real-world scenarios, but instead performs ‘micro-benchmarks’ of particular subsystems, on the assumption that if each individual subsystem or component performs well, then the whole system will perform well. This is not necessarily a safe assumption, however, as VM development has shown in the past.

Often an idealized benchmark has shown one VM version to be ‘better’ than another, while users report subjective impressions that are the exact opposite of the test results. At the same time, restricting benchmarks to only real-world loads will never provide specific, fine-grained numbers about particular areas of the VM.

The quest for the perfect VM benchmark is ongoing. This may in part account for the tremendous divisiveness surrounding VM development; Rik van Riel and Andrea Arcangeli have been proposing competing implementations of VM for years, with both sides having their egos bruised.

Linus’ decision to uproot Rik’s VM and replace it with Andrea’s in the midst of the 2.4 “stable” series, was met with tremendous criticism; though eventually most of the critics did come to believe Linus made the right choice. ■