AN SSC PUBLICATION

Speed Up Web App Development with Rails

LINUX JOURNAL

Since 1994: The Original Magazine of the Linux Community
NOVEMBER 2005

LINUX JOURNAL READERS' CHOICE 2005

2005 Readers' Choice Awards

Echo Cancellation for your VoIP PBX System

Improving Security with Two-Factor Authentication

PENGUIN-POWERED PINBALL

Dead electronics but solid mechanicals? Bring new life to old games with real-time Linux.

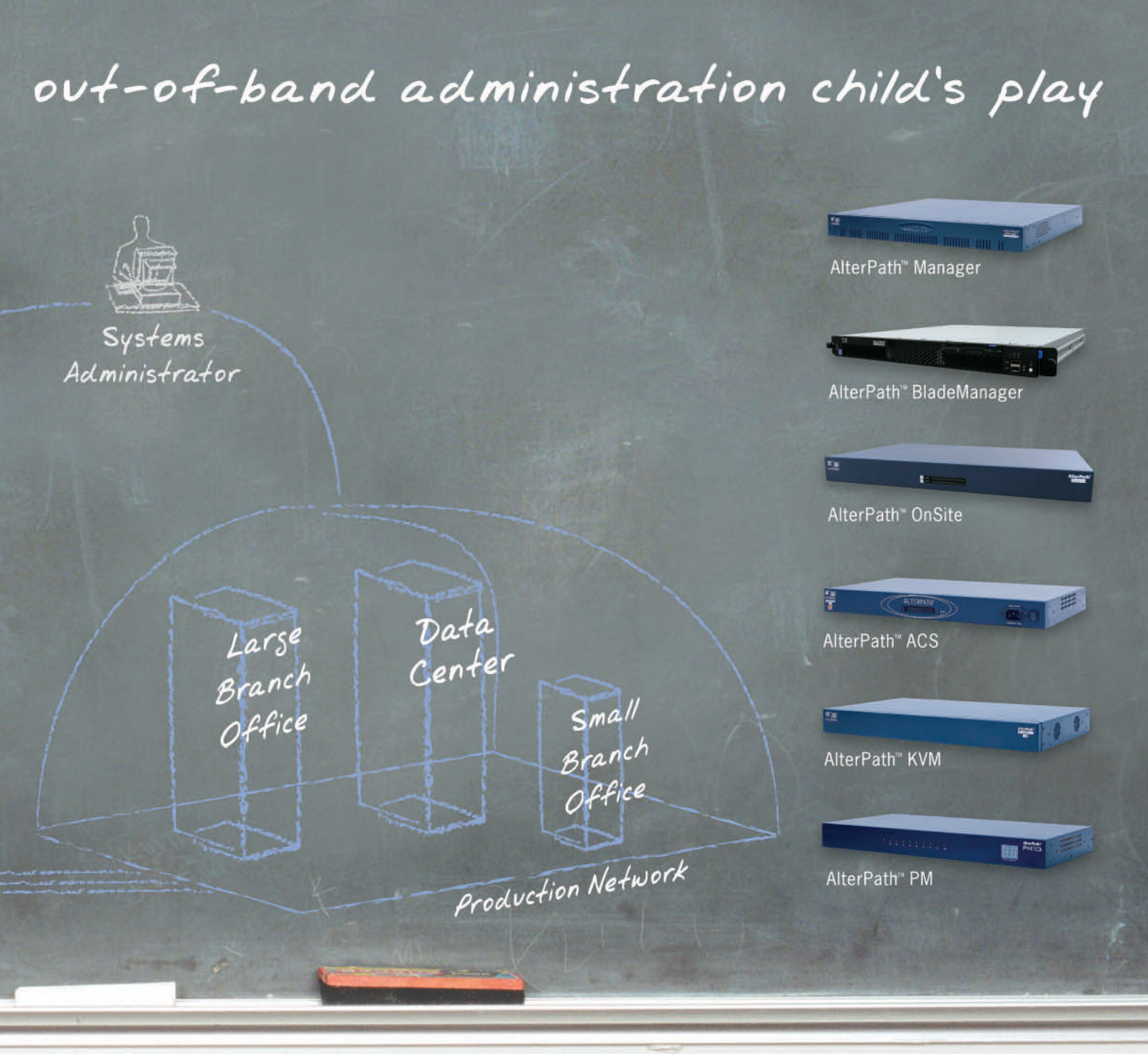
INTERNET RADIO YOUR WAY

16-NODE BEOWULF IN A TOOLBOX

Power Editing with the Happy Hacking Keyboard

USA $5.00 CAN $6.50
www.linuxjournal.com
11
0 71486 03102 4

LINUX JOURNAL · HACK ANYTHING · 2005 Readers' Choice Awards · Internet Radio · Lunchbox Cluster · NOVEMBER 2005 ISSUE 139



The Next-Generation IT Infrastructure

Cyclades AlterPath™ System is the industry's most comprehensive Out-of-Band Infrastructure (OOBI) system. The AlterPath System allows remote data center administration, eliminating the need for most time-consuming, remedial site visits. When fully deployed in your data center, Cyclades AlterPath System lowers the risks associated with outages, improves productivity and operational efficiency, and cuts costs.

Each component of the AlterPath System is designed to seamlessly integrate into the enterprise, able to scale in any direction. Whether you need serial console management of networking equipment, KVM for access to Windows® servers, branch management, IPMI or HP iLO for service processor management or advanced power management, the AlterPath System delivers. Cyclades brings it all together, making OOBI administration seem like child's play.

Over 85% of Fortune 100 choose Cyclades.

www.cyclades.com/lja

1.888.cyclades · sales@cyclades.com

cyclades®

Our salespeople don't like using our CRM system.
Our CRM system is slow and outdated.
Our CRM costs too much

SugarCRM™ solves all these problems that proprietary and hosted solutions created.

Get hooked now. Implement a Linux-based CRM system in under 15 minutes.

Start using SugarCRM today.

Open Source Customer Relationship Management

SUGARCRM™

www.sugarcrm.com/swap

1.877.842.7276
+1.408.454.6941

LINUX JOURNAL
NOVEMBER 2005 ISSUE 139

COVER STORY

FEATURES

INDEPTH

EMBEDDED

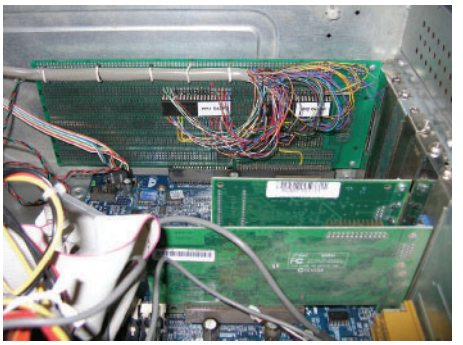
TOOLBOX

COLUMNS

REVIEWS

DEPARTMENTS

COVER PHOTO: JOHN R. BORK



This homemade 48-port I/O board easily handles the 11 inputs and 20 outputs needed to work the pinball machine (page 50).

NEXT MONTH

MULTIMEDIA

Interested in a full-featured Linux-based TiVo replacement? Well, MythTV is no Myth. It's a full-featured Digital Video Recorder similar to those provided by your cable provider, but without monthly fees and restrictions. Find out how to set it up, configure it and how to export the video you record to other formats.

Is your company's security infrastructure based on Linux? If so, don't miss Ti Leggett's piece on configuring a secure corporate directory. Ti will cover details on securing LDAP using OpenSSL and then replicating LDAP directories securely.

Learning Ruby and Rails? Reuven Lerner continues his great series on this topic by looking at how ActiveRecord makes implementing data integrity checks a snap.

dmarti:~$ logout

Do something for freedom every day, especially when you're building new systems.

BY DON MARTI

Since this is my last column as editor in chief, I get to give a bunch of advice, so I'll cover two great inventions that we should all take a fresh look at and come up with more things like them. First, the most important technology for the Internet isn't on the Internet. Want a hint? 12:00. 12:00. 12:00. The second most important technology has a symbol that you probably look at in a Web browser several times a day.

And I get to thank people for making the *Linux Journal* editor job the best job ever. Edsger Dijkstra once wrote, "Besides a mathematical inclination, an exceptionally good mastery of one's native tongue is the most vital asset of a competent programmer." By this measure, our authors are competent programmers, some even in a non-native human language. There has been no better way for me to get my Linux questions answered than to assign articles to these informed, helpful people.

Thanks to the editorial staff too. *Linux Journal* is fortunate to have Jill Franklin's managerial, editorial and XMLitorial skills; Heather Mead's quiet but effective powers that bring in links like few other Linux sites; and of course Garrick Antikajian's eye for good design, even when it includes hairy-looking code. Thank you all for not selling out to the Mainstream IT Media and sticking with your fans.

The humble VCR clock is the Internet's most important technology because it saved civilization in 1984. The big movie studios wanted to create a standard for copyright infringement that would crush any new communications technology. In a scarily close decision—5 to 4—the Supreme Court allowed the VCR to exist because you can use it for time-shifting.

The principle got a thorough test in the Grokster case decided this June, and although the new "affirmative steps to foster infringement" test will surely scare the venture capitalists away from media-oriented startups, the so-called *Sony* principle gives you the right to continue inventing.

The lesson here is that lawmakers and courts look at the wrappers of things and their real uses, not just at principles. If an invention is great for freedom, put a big obvious "clock" on it—a way for it to prove itself to society. How about a virus checker updater that uses a new P2P system? Inventing has always been part showmanship, and the features of an invention let it speak for itself in debates about laws and norms.

If you thought in the 1980s that you would be able to participate in global communication and commerce using freely licensed software and high-grade crypto on a cheap computer, you should probably tone your optimism down a little. Our other invention to appreciate is the little "lock" in the Web browser. The Internet doesn't work for business transactions without strong crypto. Every big company that wants to run a shopping site, share documents with traveling employees or run a remote backup had to join the side of freedom in the crypto debate. When inventing something that makes big business sense, build in a dependency on freedom and enroll powerful interests on freedom's side.

This is really our best issue yet. We have a brand-new feature of the latest kernel, possibly the most productive Web tool ever, a Beowulf cluster in a toolbox, freedom-enabled tools for designing electronics projects and of course a real-time Linux pinball machine. Stay free and enjoy the issue.

Don Marti is editor in chief of *Linux Journal*.

LINUX JOURNAL

NOVEMBER 2005
ISSUE 139

***LINUX JOURNAL* IS PUBLISHED BY, AND IS A REGISTERED TRADE NAME OF, SSC PUBLISHING, LTD.**

PO Box 55549, Seattle, WA 98155-0549 USA • linux@ssc.com

SUBSCRIPTIONS

E-MAIL: subs@ssc.com • URL: www.linuxjournal.com

PHONE: +1 206-297-7514 • FAX: +1 206-297-7515

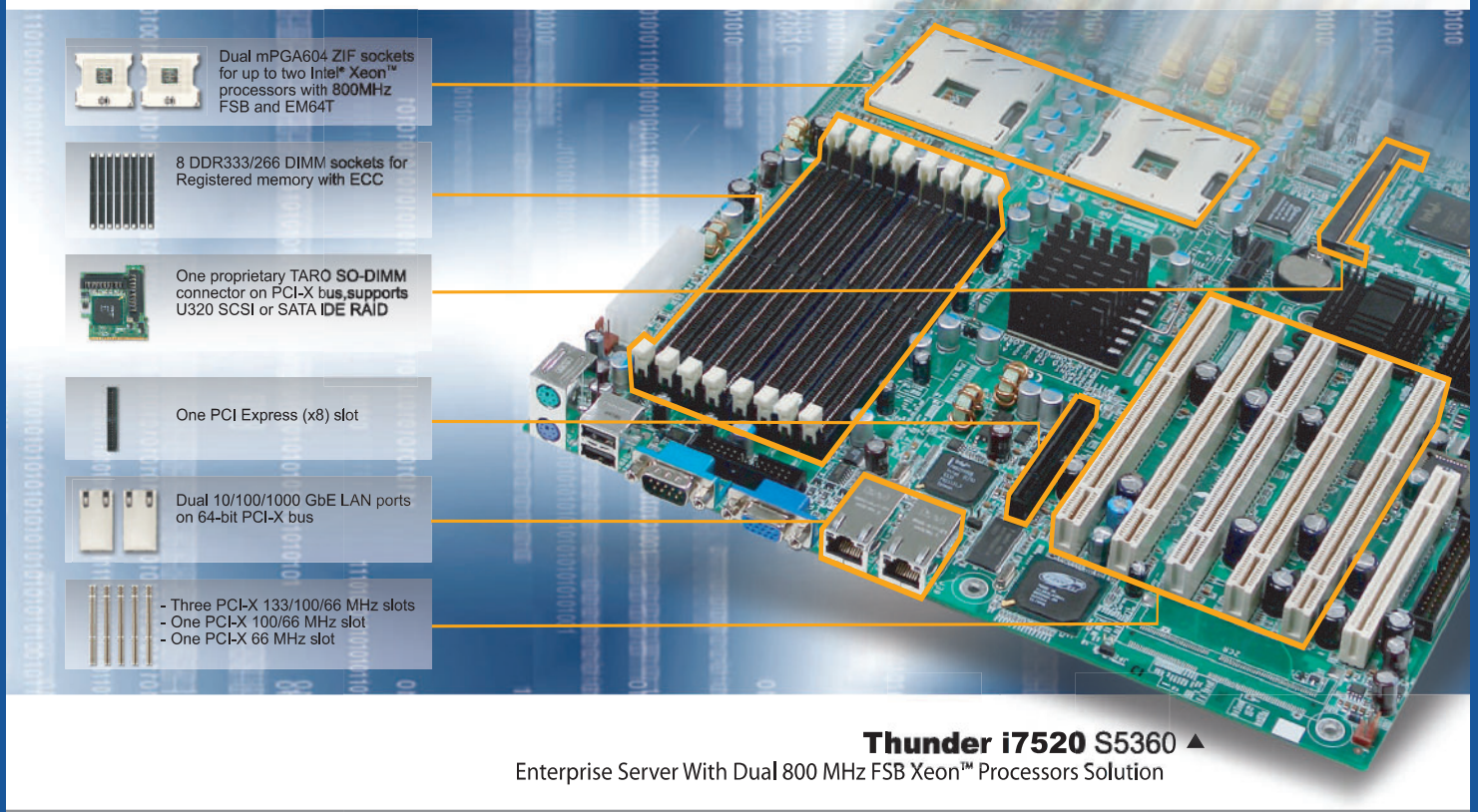
TOLL-FREE: 1-888-66-LINUX • MAIL: PO Box 55549, Seattle, WA 98155-0549 USA • Please allow 4–6 weeks for processing address changes and orders • PRINTED IN USA

USPS *LINUX JOURNAL* (ISSN 1075-3583) is published monthly by SSC Publishing, Ltd., 2825 NW Market Street #208, Seattle, WA 98107. Periodicals postage paid at Seattle, Washington and at additional mailing offices. Cover price is $5 US. Subscription rate is $25/year in the United States, $32 in Canada and Mexico, $62 elsewhere. POSTMASTER: Please send address changes to *Linux Journal*, PO Box 55549, Seattle, WA 98155-0549. Subscriptions start with the next issue. Back issues, if available, may be ordered from the Linux Journal Store: store.linuxjournal.com.

TYAN

Fast, Flexible, and Feature-Rich!

PCI Express and EM64T Servers Have Arrived



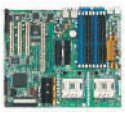
Thunder i7520 S5360 ▲
Enterprise Server With Dual 800 MHz FSB Xeon™ Processors Solution

intel inside XEON

Thunder i7520 S5360

- Supports two Intel® Xeon™ processors with 800 MHz FSB and EM64T
- 8 DIMMs for DDR266/333 memory
- Three PCI-X 133/100/66 MHz slots, one PCI-X 100/66 MHz slot, one PCI-X 66 MHz slot and one 33MHz PCI slot
- One PCI Express™ x8 slot
- One proprietary SO-DIMM connector on PCI-X bus, supports U320 SCSI or SATA
- Dual GbE LAN

Tiger i7320 S5350



- Supports two Intel® Xeon™ Processor with 800MHz FSB and EM64T
- 8 DIMMs for DDR266/333 memory
- Two PCI-X 64/66 MHz slots; three 32/33 PCI 2.3 slots
- One proprietary SO-DIMM connector on PCI-X bus, supports U320 SCSI or SATA
- Dual PCI Express GbE LAN

TYAN COMPUTER CORP.

Tyan Computer USA
3288 Laurelview Court
Fremont, CA 94538 USA
Tel: +1-510-651-8868 Fax: +1-510-651-7688
Pre-Sales Tel: +1-510-651-8868 x5120
Email: marketing@tyan.com

For more information about this and other Tyan products, please contact Tyan Pre-Sales at (510) 651-8868 x5120, or contact your local Tyan system integrator/reseller.

www.tyan.com

SSH Tip

I was very happy to see John Ouellette's article in the September 2005 issue. In particular, it was nice to see someone discuss limiting of remote command execution using the authorized _keys file. However, I would like to point out that with a little extra work, it is entirely possible to secure your ssh private key with a decent password and still use it in scripts and cron jobs without human intervention.

Keychain, when combined with ssh-agent, allows you to re-use an ssh-agent session between logins. Once you use keychain to launch ssh-agent, you need to enter the password for each of your private keys only once. Keychain then keeps your key decrypted until it is killed. We use this method on all of our production servers for secure remote backups. Since our servers are rarely rebooted, the keychain remains active for as long as I need it. Should I reboot the server, or should I be forced to kill the keychain or ssh-agent, then and only then will I have to retype my password. See www.gentoo.org/proj/en/keychain/index.xml.

--
Chris Poupart

Fresh Air for Reading

This weekend some friends and I were trekking in Jotunheimen, a popular national park in Norway. At the top of a mountain called Surtningssue (2,368 meters), I felt a sudden urge to read *LJ*.

--
Lars Strand

Enough with the Kid Pictures

I read *Linux Journal* because I run a Linux consultancy business, and *LJ* does an excellent job of helping me keep up to date with some of the developments in the Linux world, both commercial and technical.

Some of the letters published are amusing, particularly when people seem disproportionately upset by a particular advertisement. However, we all have our foibles and mine is an abhorrence of the pictures of readers' children month after month. How many readers really need or even want to see them?

--
Keith Edmunds

Baby with *LJ*

Our son Sam is excited about being a member of the Linux generation.

--
Bob Overberg

Happy Archos User

I happened to come across your review of the Archos PMA430 [September 2005] and would like to offer some counterpoint. The PMA430 is my third Archos unit; a 20GB MP3, the great AV340 and now the PMA. So I am happy with their products in general.

Now specifically to the PMA430. First of all, the SDK package, such as it is, has been released. Second, programs for Microsoft formats such as Excel, Word and PowerPoint are also available, and in fact work quite well. Third, there are some sync problems with films, but this is easily overcome with the correct software. Fourth, I use the PMA Wi-Fi quite often, and overall it seems fast enough for me.

Another complaint was that the PMA430 did not have enough software or functions to make it worthwhile. This seems rather not the point, since not many people would have use for a bare-bones PC. We all find apps outside of those that come with the PC. So it seems quite natural for PMA430 users to find more and better ways to use it.

I have had mine for about two months and find new uses for it every day. In this short time I feel lost when it isn't in my pocket. Last, I was able to purchase one for less than $700, and Archos had a special that threw in $150 worth of accessories.

My only complaint is that I don't have any Linux experience, so the learning curve is pretty steep. Luckily, there are lots of great people out there working hard to make this product even better.

--
Alan E. Kayser

You know this means we're going to bug you to write an article for LinuxJournal.com on apps for your Archos, right?—Ed.

Java Tool Recommendation

About the article titled "Developing GNOME Applications with Java" [July 2005]: the article is excellently written and provided some important insight, precisely as I'm integrating several legacy applications into a Linux/Java enterprise solution for a company in Italy.

I'd like to point out to *LJ* readers that Borland has released JBuilder Foundation free of charge, even for commercial use. I had been plugging several solutions in to a toolset for Java GUI development under Linux (including some mentioned in the article, such as the Glade XML GUI generator), but then I came upon JBuilder Foundation, and it solved all my needs in one powerful tool.

I'd like to suggest you contact your distributor in Brazil because they're charging us $13.60 US per issue here, or 31.95 Reais on today's exchange rate. That is a 272% increase from the US newsstand price. Brazil is one of the world's biggest Linux and open-source bases and still we pay a hefty price for valuable printed information.

--
Jose Melo de Assis Fonseca

Linux in 1856

I read with interest the article "First Beowulf Cluster in Space", in the September 2005 issue of *Linux Journal*. However, I was surprised to read, in his profile, that co-author Ian McLoughlin has been using Linux since 1856! What kernel was he using then and what CPU was he running it on?

--
James Knott

He borrowed the Linux-powered time machine we use to set the publication dates on our Web site.—Ed.

Why Split LinuxWorld Booths?

I visited LinuxWorld Expo here in the San Francisco Bay Area (Moscone Center) today. I was very disappointed, nay, irritated to find the glitzy, high-roller moneyed exhibitors on the first floor, with the .org exhibitors (for example, Free Software Foundation, Debian, Fedora, Gentoo, Mozilla, LTSP, Etherboot and so on) ghettoed onto the second floor. These .org organizations are the heart and lifeblood of Linux and deserve their places cheek by jowl (and, do I mean jowl!) with the commercial stuff they enable by their existence and the hard work and dedication of their supporters and developers.

--
Robert Lynch

Try working a show next to a vendor's loud T-shirt giveaway area, and you might start pining for the friendly "dot-org" area too.—Ed.

Networking Tip

Marcel Gagné's instructions for setting up ndiswrapper leave out an obscure adjustment that is needed at least on the Fedora Core 2 distribution I am using. The problem may not occur in other Linux distributions, but it is the source of frequent networking failures at boot-up on FC2.

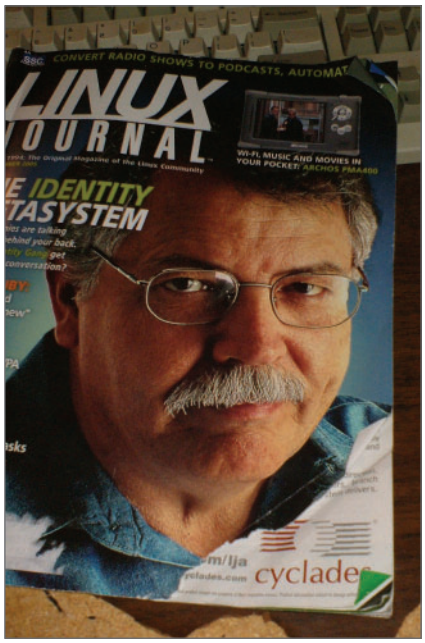
If your network card is a pcmcia device, the pcmcia driver has to be ready before the attempt to bring up wlan0. Unfortunately, in /etc/rc3.d, /etc/rc4.d and /etc/rc5.d, the pcmcia script has a much later sequence number (S24pcmcia) than the network script (S10network). Since these are merely symbolic links, the order can be changed with a minimum of risk. I moved the network link to S11network, and the pdmcia link to S10pcmcia in all three directories.

As Marcel would say, voilà!

--
Pierre MacKay

Dog-Eared *LJ*

I am having some problems with my *Linux Journal*. It has been showing up as if the US postal service has been reading my magazine. I receive every issue with dog-eared pages, front cover torn. It's a great magazine; I look forward to every issue!



--
Scott Wilson

We'll send you a replacement copy and ask the Postal Service to get their own.—Ed.

Split Off the Baby Section, Please

What is this, *Parenting Magazine*? You need to make a separate section for all the pictures of babies/kids/stuffed animals/pets and dedicate the Letters section to actual intelligent commentary. I, for one, am sick of wading through all the "my daughter sketched a penguin just...for...you!" nonsense so that I can <gasp> read actual technical letters about Linux. I subscribe to your magazine for Linux know-how and articles; if I wanted family-friendly piffle, I could send my dollars to *Family Circle* or *Parents*.

--
Chris

Awww, Look at the Baby!

Our six-month-old daughter Guen loves Linux, as you can clearly see from this picture. She writes "ggggg [d.ddss 4449dlddd", which I think means "Does this ultimate Linux box come with a baby-sized keyboard?"

--
Matthew and Karen Miller

Viv(a|e) Marcel!

To the everyday Canadian, French is probably as ubiquitous and familiar as Spanish is to us Southerners (I grew up in Florida where we learned "Cuban" in school—seven years of conversational Spanish—and I now live in Mexico II, aka Los Angeles). However, to the majority of English-speaking Linux enthusiasts who have any second-language experience at all, it is frequently Spanish, not French. This totally leaves us out of the joke.

I finally cracked this month and had another look; the mention of Damn Small Linux and a photo of a USB pen drive proved irresistible [August 2005], so I did a flyby and took one more look. I'm glad I did. Marcel seems to have lowered the language-barrier veil and made his excellent column accessible to us all, not just the French-speaking sector.

Con Mucho Respeto (your turn to look it up).

--
Jeff Jourard

MonarchComputer.com

AMD Solution Provider PLATINUM

Visit our website or Call 1-800-611-0875

PayPal

FREE! - Get Half-Life 2, a Patriot SD FlashCard Kit, and a LIVESTRONG™ Wristband when you upgrade to Athlon 64/FX/X2!

Purchase any AMD Athlon™ 64/FX/X2 bundle from Monarch and get Half-Life 2, a 128 MB Patriot SD Flash Card Kit w/USB Reader and a LIVESTRONG™ wristband for helping Monarch support the battle against cancer. Monarch Computer and you can make a difference.

PATRIOT MEMORY

FREE INSTALLATION SETUP & TESTING BY CERTIFIED TECHS

FREE TECH SUPPORT! Monarch makes it quick and easy to upgrade with FREE setup and testing on Motherboard Combos and $18.00 build fee on Barebones.

AMD Motherboard Combos

DFI K8T800PRO-ALF w/ AMD Sempron™ processor 2600+ (754 - 64 bit) — Only $129

Asus K8N-E Deluxe Mainboard with AMD Sempron™ processor 3400+ (754 - 64 bit) — Only $249

Asus A8N-E nForce4 Mainboard with AMD Athlon™ 64 processor 4000+ — Only $475

Abit AN8 SLi nForce4 with AMD Athlon™ 64 X2 processor 3800+ (Dual Core) — Only $539

Monarch A0110289 (H8DCE) Mainboard w/ AMD Opteron™ processor 246 — Only $585

Tyan S2895UA2NRF K8WE MB w/ AMD Opteron™ processor 275 (Dual Core) — Only $1629

Mainboard - Processors - Heatsink and Fan with Memory Options - FREE INSTALLATION AND TESTING
LASTEST BIOS loaded for easy upgrades - AMD Sempron™ Athlon™ 64, Athlon™64 FX, and Opteron™ Combos Available

AMD Barebone Systems

Go to www.monarchcomputer.com. select Barebones from the Menu. Choose AMD Sempron™, AMD Athlon™ 64 or AMD Opteron™. Then configure your barebones online or call 1-800-611-0875.

Enermax CS-10182-BA E-ATX Tower w/460W PS
Tyan S2875ANRF (Tiger K8W 8151) MB
AMD Opteron™ processor 242 1.6 GHz
Starting @ $589

Antec Sonata II Piano Black Quiet Mid Tower w/450W PS
Abit AN8 Ultra nF4
AMD Athlon™ 64 processor 3700+ (939 - 90nm)
Starting @ $509

Lian-Li PC-70 Aluminum Server Case w/550W PS
Tyan S2882G3NR-D (Thunder K8S)
AMD Opteron™ processor 270 (Dual Core)
Starting @ $1579

Apex, FoxConn Mid-Tower TU-150 Case w/400W PS
DFI K8T800PRO-ALF Motherboard
AMD Sempron™ processor 2600+ (754 - 64 bit)
Starting @ $217

***AMD Athlon 64 and Athlon 64 FX are the first Windows®-compatible 64-bit PC processor

www.monarchcomputer.com

AMD Price Drop!

AMD Athlon™ 64 Retail Box CPUs
AMD Athlon™ 64 3000+ (939) $146.00
AMD Athlon™ 64 3200+ (939) $190.00
AMD Athlon™ 64 3500+ (939) $219.00
AMD Athlon™ 64 3700+ (939) $267.00
AMD Athlon™ 64 3800+ (939) $323.00
AMD Athlon™ 64 4000+ (939) $368.00
AMD Athlon™ 64 FX-55 (939) $811.00
AMD Athlon™ 64 FX-57 (939) $1011.00

AMD Athlon™ 64 X2 Retail CPUs
AMD Athlon™ 64 X2 3800+ $347.00
AMD Athlon™ 64 X2 4200+ $473.00
AMD Athlon™ 64 X2 4400+ $526.00
AMD Athlon™ 64 X2 4600+ $690.00
AMD Athlon™ 64 X2 4800+ $884.00

The AMD Athlon™ 64 X2 dual-core processor provides the same level of system features customers have grown to expect with the AMD Athlon™ 64 product family: HyperTransport™ technology - Enhanced Virus Protection for Microsoft® Windows® XP- SP2 - Cool'n'Quiet™ technology

AMD Opteron™ OEM CPUs
AMD Opteron™ 146 2.0GHz $173.00
AMD Opteron™ 148 2.2GHz $211.00
AMD Opteron™ 150 2.4GHz $270.00
AMD Opteron™ 242 1.6GHz $158.00
AMD Opteron™ 244 1.8GHz $158.00
AMD Opteron™ 246 2.0GHz $203.00
AMD Opteron™ 248 2.2GHz $307.00
AMD Opteron™ 250 2.4GHz $441.00
AMD Opteron™ 252 2.6GHz $669.00
AMD Opteron™ 254 2.8GHz $825.00
AMD Opteron™ 846 2.0GHz $677.00
AMD Opteron™ 848 2.2GHz $677.00
AMD Opteron™ 850 2.4GHz $847.00
AMD Opteron™ 852 2.6GHz $1130.00
AMD Opteron™ 854 2.8GHz $1469.00

AMD Opteron™ Dual Core OEM CPUs
AMD Opteron™ 265 1.8GHz $825.00
AMD Opteron™ 270 2.0GHz $1019.00
AMD Opteron™ 275 2.2GHz $1260.00
AMD Opteron™ 865 1.8GHz $1469.00
AMD Opteron™ 870 2.0GHz $2085.00
AMD Opteron™ 875 2.2GHz $2570.00

Components and Upgrades
1000s of In-Stock Components

100988
Antec SLK1650B
Black Mini Tower w/350W PS
$70.00

150439
Western Digital
320GB SATA
8MB Cache 7200RPM
(WD3200JD)
$135.00

140270
1 GB DDR (266) PC-2100
REG ECC Corsair
(CM72SD1024RLP-2100)
$136.00

140226
2 GB DDR (2 pcs 1GB)
PC-3200 (400)
OCZ Dual Channel Platinum
(OCZ4002048ELDCPE-K)
$259.00

190522
ATI (Connect3D) Radeon 9250
256MB DDR/8x-AGP
TV-Out/DVI (Retail Box)
$56.00

150239
Western Digital
74 GB SATA 10K
Raptor (WD740GD)
$169.00

100298
PC Power & Cooling
Silencer 470W ATX
12V Power Supply (OEM)
$95.00

110192
Asus K8N-DL nForce4
Audio/GB-LAN/USB/IEEE
PCI-E/SATA/DDR/E-ATX Opteron
$284.00

Educational and Government POs Welcome.

GSA Schedule Contract GS-35F-0202P

We ship to the Continental U.S., Alaska, Hawaii, APOs, Puerto Rico, and Canada

Commercial leasing available for purchases as low as $1000.

Prices subject to change without notice. Monarch Computer not responsible for typographical errors.

Monarch Has The LOWEST PRICES
Custom 64-Bit Servers,
Workstations & Desktops
Available with AMD Dual-Core Technology!

AMD 64 Athlon X2 | AMD Sempron | AMD 64 Athlon FX | AMD 64 Athlon | AMD 64 Opteron

NEW! Monarch Server Overhaul!

Monarch's EMPRO™ line makes buying AMD Opteron™ Workstations and Servers Easier than ever before!

Take the Monarch Challenge: Give us 5 minutes and Save $100s if not $1000s! Come see our jaw-dropping prices! In most cases, we're cheaper than the competition BY AS MUCH AS 20%!

1. Visit our website and select "Servers".
2. Choose 1U, 2U, 3U or 4U and choose SATA or SCSI.
3. Custom configure your new server, or use one of our pre-configurations based on:

Price Point,	Application,	or General Use
Least Expensive	Fileserver	Max Storage RAID 5
Under $2000	Appserver	Standard RAID 5
Under $3000	Computational	Standard Mirror
Under $4000	AV Editing	Most Redundant
		Maximum Config.

4. Purchase your server or save your quote and lock in pricing for up to seven days. Only at MonarchComputer.com!

SAVE MONEY SACRIFICE NOTHING!

"BOTTOM LINE: MUST BUY"

LINUX MAGAZINE EMPEROR CLASS

"What's not to like? Monarch provides top parts, excellent customer service, and has earned the highest-level solutions provider status recognized by AMD and other key component vendors."

Jason Perlow
Linux Magazine
April 2005

For more information on the Empro™ Line Visit www.monarchcomputer.com/empro

"I bought the dream system with top components that were hard to find elsewhere. The guys at Monarch did a perfect job of building it, down to the smallest details. Wiring was perfect, position of hard drives within the chassis also perfect, etc. Even the packaging was superb. And every communication I've had with them was also perfect. Ordering from Monarch was a good move."

User: canbbb
ResellerRatings.com

The AMD Opteron processor—built upon forward-thinking AMD64 technology—provides flexibility with a 1–8-way scalable design.

LINUX FRIENDLY!
CUSTOM AND PREBUILT PCS - Linux Preinstalled!
Fedora - RedHat - SuSe

LOOK FOR THIS SYMBOL ON LINUX COMPATIBLE MONARCH SYSTEMS!

9 MAXIMUM PC KICKASS | PC WORLD | CPU COMPUTER POWER USER | LINUX MAGAZINE EMPEROR CLASS | COMPUTER SHOPPER EDITORS' RATING 8.0 | PRICEWATCH | RESELLERRATINGS.com | BizRate Overall Rating: | PriceGrabber.com | Shopping.com Store rating:

Photo of the Month: Dad's Ride

LJ's pages are full of smiling babies, but what about the readers' parents? Take a look at my father riding an armored car in 1947. Pingouin means penguin in French.

--
Le Glaude

Photo of the Month gets you a one-year extension to your subscription. Photos to ljeditor@ssc.com.

Archives, Patents

I am a subscriber to your excellent magazine. Can I suggest you make a service that at the end of each year it would be possible to order a DVD with all the contents from 1994 to present day? That would be nice, and I for one certainly would order it each year.

But my main reason to write to you is these damn patents. It was good they did not succeed in the EU (I am in Finland). It was only delayed—patents will be back on the agenda in a year, and we will have to live again through waiting for an axe to our neck. I have been thinking what a counterstrike would be.

There should be an organization that would take care of people's patents so that GPL software can use them gratis but others must pay. These moneys will be used to finance further patents, defend patents, buy patents to be used in GPL software and so on. I am sorry I don't have couple of millions to kick the show up.

Microsoft and others have been very keen to point out that Linux uses some patented algorithms. But this finger pointing has been—should I say—one-sided. All MS wares are closed source, so if the source code were combed, I might think a lot of patented things might be found.

--
Kari Laine

We have good news for you. Check out https://www.ssc.com/cgi-bin/lj/back_issue *for the archive CD and* osdl.org *for the Patent Commons Project.—Ed.*

Acer Laptop Refund Offer

I thought I'd share the following story that has some interesting angles and happened just over the past few weeks as I bought a new Acer laptop (Aspire 1674WLMi).

I bought the machine at a local (Dutch) consumer electronics reseller called MediaMarkt. I asked the salespeople if I could buy it without an OS, which, of course, was not possible. However, I could try contacting Acer themselves through the local importer, Acer Benelux, they said. I contacted Acer by e-mail, and indeed there was a restitution procedure. I couldn't believe my luck!

After supplying them with a serial number and a scanned copy of the receipt, I received the "agreement" in PDF. Unfortunately, this turned out to be a disappointment: the restitution would amount to EUR 30 (about the same in US$), but I would have to send the laptop to the Acer offices somewhere else in the Netherlands, where they would reformat the drive and send it back to me within five working days. Obviously, sending an expensive machine at my own risk and at my own cost would cost me far more than EUR 30, and during that time I could not use the machine. And, I'd have to reinstall again after getting it back. In fact, having installed Fedora Core 4 as soon as I arrived home after the purchase, already invalidated the agreement (how can you know beforehand?), although my contact at Acer did not specifically complain about it. So much for the restitution procedure.

The interesting angle is that Acer *does* have a procedure, but it is constructed in such a way that it is not profitable for the average consumer to exercise it. Furthermore, when in my final message to Acer I concluded that it was a financially uninteresting proposal and asked if I could simply return the Microsoft CDs and license (obviously, it's of no use to me), they said that the procedure was the only formal way, since Acer, being an OEM, was the owner of the license. Then I decided I would give away the Microsoft stuff to a friend and asked a befriended M$ employee how that works. He said that you can't!

--
Michel

The trick is to break up the "bundle" before you accept the license for the preinstalled software. Until you power up the machine, actually have a chance to read the license and click OK the license doesn't bind you (see www.linuxjournal.com/article/5628*).—Ed.*

Fan Mail

Once again, you have more than justified the subscription fee! This issue [September 2005] contains a bunch of pearls....

The Open and Free Software aficionado in me was overwhelmed by the social-economic revolution report in the "identity metasystem" article....I crave the day when I will be able to explain fully to my die-hard capitalist friends the practicality of the grass-roots economy!

The embedded developer in me rejoiced in reading the story and the specs of the "First Beowulf Cluster in Space".

And the average Linux user in me got up to speed on Syndication and Podcasting.

Keep the focus and motivation!

--
Vasco Névoa

We welcome your letters. Please submit "Letters to the Editor" to ljeditor@ssc.com or SSC/Editorial, PO Box 55549, Seattle, WA 98155-0549 USA.

THE UNSUNG HERO:

Network Administrator

9:42 am Singapore branches go offline, trouble ticket created

9:44 am Jeff uses diagnostics to isolate failure to core router – not responding

9:45 am Out-of-band access to core router established via the **AlterPath™ ACS**

9:47 am Router shows subnet mask set incorrectly during previous configuration

9:48 am Jeff resets subnet mask properly, reboots router

9:49 am Link to Singapore restored, Singapore comes back online

9:50 am Jeff is planning his next vacation

Advanced Console Server

AlterPath™ ACS

Download a **FREE** White Paper on Console Management
www.cyclades.com/wpcm

www.cyclades.com/ljb
1.888.cyclades · sales@cyclades.com

cyclades

On the WEB

When it comes to Linux adoption, educational institutes and government agencies around the globe continue to lead the way. We hear a lot about various countries mandating open-source and free software usage—China, Germany and Brazil are only a few. To learn more about some other international initiatives, check out these articles on LinuxJournal.com:

- Marco Fioretti is writing a Web series for us that outlines how various provinces in Italy are bringing Linux into their high schools. Part 1 (www.linuxjournal.com/article/8309) focuses on The Istituto Tecnico Commerciale De Sterlich of Chieti Scalo in Central Italy. Part 2 (www.linuxjournal.com/article/8507) looks at The Istituto Tecnico Commerciale (ITC, Commercial-Technical Institute), "F. Besta", in Ragusa. In Part 3 (www.linuxjournal.com/article/8508), Marco takes us to Abruzzo to learn about one elementary-school teacher's free software project for schoolkids.

- During a summer trip to the Middle East, Tom Adelstein learned about many projects in the region that are using OSS. In "Linux in Government: Building Bridges and Managing Water" (www.linuxjournal.com/article/8504), he talks with Ammar Ibrahim about Bulk Meter Flow and Operations (BMFO), a project started by the Ministry of Water and Irrigation in Jordan.

In other news, in case you missed it, we took your input to heart and created a searchable category system for the *Linux Journal* Web site. Check out the home page for a list of 16 categories to search—from Audio/Visual to Webmaster—for articles going back to Issue 1 and the early days of LinuxJournal.com.

diff -u

What's New in Kernel Development

Linus Torvalds has put together a **git** repository for the full 2.6 tree, going all the way back to the introduction of BitKeeper. Kernel development still takes place on a new tree, but the old tree now exists for reference or for any other purpose one might have. This is not the first time these patches have been incorporated into a git repository, but with all the work going into git during the last three months, this is the first time a git repository for these patches has been small enough to fit in a reasonable space. The progress git has made since its inception has been utterly amazing, and a 1.0 release is apparently imminent. Although people will want their favorite revision control feature in git before they'll start using it, these features can, for the most part, all be regarded as icing on the cake. The core functionality, the stuff that controls distributed development, exists in a robust, powerful form for the very first time in a free project.

It looks as though **RelayFS** will soon be going into the main kernel tree. It's had a long stint in **Andrew Morton's -mm tree**, and it has needed no major fixes in months. A good crop of users has found it useful for a variety of applications, and the only real objection to its ultimate inclusion has been the fact that **DebugFS** performs a similar function. But if for no other reason, a filesystem called DebugFS just doesn't seem to invite users to use it for anything other than debugging. Andrew has expressed a clear willingness to push the RelayFS code up to Linus Torvalds, especially as the RelayFS developers themselves feel the time is right.

Timothy R. Chavez and others have produced a patch to enhance the **Virtual Filesystem** (VFS) auditing support to be able to audit a filesystem object based on its location and name. In the current VFS implementation doing this is impossible. When Timothy first proposed the idea, there was a bit of resistance from kernel folks who pointed out that inotify existed and performed a quite similar function. But when Timothy's auditing project started, inotify was not very mature and existed only as an external patch, so it made more sense at that time to develop this auditing code as a separate feature entirely. Now that inotify is at least in the -mm tree, a better argument can be made to use inotify instead. But Timothy and the other developers of this patch, along with critics like **Greg Kroah-Hartman**, have hatched a plan to abstract the basic functionality common to both this auditing code and inotify and make these projects simply access the abstracted features directly to get what they need. When this actually will be done is still an open question.

Adrian Bunk has tagged a number of OSS sound drivers for removal. The decision as to which drivers to remove and which to keep is not an easy one. The goal is to preserve support for all existing hardware, and so before any driver can be removed, Adrian must determine whether an ALSA equivalent exists and works. This determination often can be made only by someone with very old sound hardware, and such users may be difficult to find. Each case must be confirmed individually, and Adrian does the legwork for each one, following up on e-mail and asking questions of users. Housekeeping patches like Adrian's are often thankless, if not downright unwanted by users afraid of losing support for their favorite hardware. It's nice every once in awhile to acknowledge the hard work of folks like Adrian, who put in many hours each week, just on kernel cleanups like this one.

Wireless Security Lock gadgets are finding support in Linux. These devices allow a wireless system to detect when it has traveled too far from a given location, in order then to perform some security function, like locking the monitor. **Brian Schau**, for his first kernel driver, coded up support for WSLs. In spite of the fact that a Bluetooth phone can provide similar functionality, and in spite of the fact that a user-space application might be better suited to the task than Brian's kernel driver, the project clearly has merit, because it supports an actual existing piece of hardware. One way or another, it seems, Linux will be supporting WSLs.

Andrew Morton has offered some clarification on whether users should prefer a swap file over a swap partition, and why. In 2.6, he says, the difference is virtually nil. Both in terms of performance and reliability, swap files and swap partitions are equally good, with one exception: if the swap file created is very fragmented, performance will suffer. But because swap file fragmentation does not increase over time, simply creating a nonfragmented swap file initially solves that problem completely. In 2.4, the situation favors swap partitions over swap files, because the partition can avoid certain memory allocations that swap files require.

—ZACK BROWN

MBX is the Industry Leader for Server Appliances

From design to delivery we are dedicated to building a partnership with you.

MBX RP-1013 Platform

- Intel® Pentium 4 630 Processor at 3.0 GHz
- 1U Rackmount Chassis 16.5" Deep
- 1GB PC4200 DDR2 Memory
- Maxtor 80GB Serial ATA Hard Drive
- Eight Gigabit NIC's, 4 ports with by-pass
- Optional 16x2 LCD with Keypad
- On-board Compact Flash Socket
- Branded With Your Color and Logo
- Custom OS and Software Install
- No Minimum Quantity Required
- 3 Year Warranty

$1,699 or lease for **$52**/mo.

MBX RP-2012 Platform

- Intel® Xeon Processor at 3.0D GHz
- 2U Rackmount Chassis
- 1GB PC3200 DDR Memory
- Maxtor 73GB SCSI Hard Drive
- Dual On-board Gigabit NIC's
- Custom OS and Software Install
- No Minimum Quantity Required
- 3 Year Warranty

$2,899 or lease for **$89**/mo.

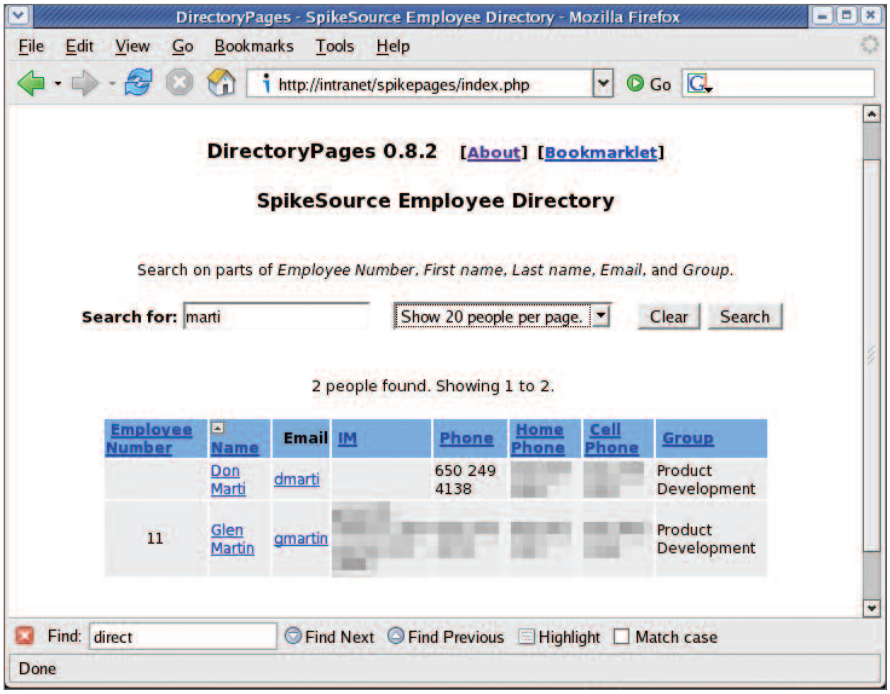
MBX is the leader in custom appliances. Many premier application developers have chosen MBX as their manufacturing partner because of our experience, flexibility and accessibility. Visit our website or better yet, give us a call. Our phones are personally answered by experts ready to serve you.

MBX systems

www.mbx.com

1.800.939.0971

DirectoryPages



As many of you already know, I have left *Linux Journal* to join SpikeSource. My first day at the new company, I looked up some people in the employee directory, which is a simple PHP interface on top of the LDAP server—no need to change someone's account info on the server and the person's employee directory entry separately. It all stays together, and employees can update their own info, such as cell-phone numbers.

DirectoryPages is not merely a handy way to keep everyone's info in one place and put an easy Web interface on it, it's also a good example of how to use LDAP data in a PHP script. A full article on how it works is bundled with it. Now all this thing needs is TeX integration to autogenerate business cards.

(Yes, the following is a link to my new employer, but I got *Linux Journal* executive editor Jill Franklin to approve it.)

Home page:
www.spikesource.com/info/search.php?c=DIRECTORY-PAGES

Support forum:

From the Christmas Penguin

One of the hottest home electronics products is the Sonos Digital Music System: a Linux-based wireless audio setup that works as a kind of iPod for the home.

Although other whole-home systems integrate with the TV and contain hard drives, the Sonos works strictly as a wireless distribution system. Your music and other audio files live on Linux, Mac or Windows PCs (or combinations of them—file sharing is through Samba) and are displayed in color on Sonos' wireless handheld controller. Each room has its own ZonePlayer—a small 50-Watt amplifier. You can choose the speakers or buy Sonos' own bookshelf units.

Writing in MadPenguin.org, Christian Einfeldt says, "It's the current state of the art for wirelessly controlling music in a large home or business where you need just the right music in the right room at the right time. And best of all, it's powered by GNU/Linux!" The *Wall Street Journal* calls it "...easily the best music-streaming product I have

EmperorLinux

...where Linux & laptops converge

You choose your laptop... from a wide selection of top tier laptops manufactured by Dell, IBM, Lenovo, Sharp, and Sony. They come in all sizes from two pound ultra-portables to eight pound desktop replacements; get exactly as much Linux laptop as you need. Need help deciding? Our experts will help you select a Linux laptop to meet your needs.

The Meteor: 3 lb Linux

- Sharp Actius MM20/MP30
- 10.4'' XGA screen
- 1.6 GHz Transmeta Efficeon
- 20-40 GB hard drive
- 512-1024 MB RAM
- CDRW/DVD (MP30)
- 802.11b/g wireless
- ACPI hibernate
- 1'' thin
- Ask about the 3D Molecule

The SilverComet: 4 lb Linux

- Sony VAIO S380
- 13.3" WXGA+ screen
- X@1280x800
- 1.6-2.13 GHz Pentium-M
- 40-100 GB hard drive
- 256-1024 MB RAM
- CDRW/DVD or DVD-RW
- 802.11b/g wireless
- ACPI hibernate
- Ask about the 17" Gazelle

You choose your distribution... from among the most popular Linux distributions available. We'll install the distribution you select, then we'll install our custom, laptop-specific kernel and configure your distribution for full hardware support, including: X at the native resolution, wireless ethernet, power management, 3-D graphics, optical drives, and more.

The Toucan: 5 lb Linux

- ThinkPad T series by Lenovo
- 14.1" SXGA+/15.0'' UXGA
- X@1400x1050/X@1600x1200
- ATI FireGL graphics
- 1.6-2.26 GHz Pentium-M 7xx
- 40-100 GB hard drive
- 512-2048 MB RAM
- CDRW/DVD or DVD-RW
- ACPI suspend/hibernate
- Ask about the Raven X41 Tablet

The Rhino: 7 lb Linux

- Dell Latitude D810/M70
- 15.4'' WUXGA screen
- X@1920x1200
- NVidia Quadro or ATI Radeon
- 1.73-2.26 GHz Pentium-M 7xx
- 30-100 GB hard drive (7200 rpm)
- 256-2048 MB RAM
- CDRW/DVD or DVD±RW
- 802.11a/b/g wireless, GigE
- Ask about the tiny Koala X1

Let EmperorLinux do the rest. Since 1999, EmperorLinux has provided pre-installed Linux laptop solutions to universities, corporations, and individual Linux enthusiasts. We specialize in the installation and configuration of the Linux operating system on a wide range of the finest laptop and notebook computers made by Dell, IBM, Lenovo, Sharp, and Sony. We offer a range of the latest Linux distributions, as well as Windows dual boot options. All systems come with one year of Linux technical support by both phone and email, and full manufacturers' warranties apply.

www.EmperorLinux.com 1-888-651-6686

Rails and Databases

After years of painful Web development, here's a development framework based on understanding how Web developers really use relational databases. Rails standardizes the tweaky parts for you to save time. BY REUVEN M. LERNER

Last month, we began looking at Ruby on Rails, a Web development framework that has captured a great deal of attention in only a short time. Much of the success of Rails is due to the ease with which Web/database developers can accomplish various tasks. Indeed, Rails fans often tout the fact that their applications have almost no configuration files, allowing programmers to concentrate on development, rather than logistics.

This month, we begin to look at how Rails works with relational databases. Even if you won't be using Rails in your own Web development work, the way Rails addresses many different issues is extremely elegant and may well influence future generations of object-relational technologies.

The Problem

The database side to Rails attempts to solve a seemingly simple problem. Where and how should a Web application store persistent information? Nearly any Web application we might want to build, from a shopping cart to a calendar/diary, needs to store its information somewhere. And because Web applications run on the server, rather than on the user's desktop, we need to keep track of data for many different users, rather than just one.

Back in the olden days of Web development, when applications were far less sophisticated, some of us used basic text files. But we quickly discovered that a relational database was an improvement on nearly every level. Relational databases are designed to provide fast, secure and flexible access to the data that we want—so long as we can represent our data as two-dimensional tables.

But as simple as that last sentence makes it sound, moving data from a program into a database is neither simple nor straightforward. Sure, the simple stuff is indeed pretty simple; it's not a big deal to keep track of customers' bank balances, or even the latest transactions in their checkbooks. But there are big differences between the objects that are increasingly at the center of the programming world and the tables that are at the center of the database world. Consider the contortions that database programmers go through in representing arbitrarily deep hierarchies, and you'll begin to understand how the mapping between objects and tables can be quite complex.

There are basically three ways to bridge this gap between objects and tables: handle it manually, replace the tables with objects and use an automatic mapping tool. The manual approach, which is probably the most common and popular, simply means that the programmers stick SQL queries into the code. To get the contents of a shopping cart, we do something like this Perl code:

```
# Send the shopping-cart query
my $sql = "SELECT item_id, item_name,
                  item_price, item_quantity
             FROM ShoppingCart
            WHERE user_id = ?";
my $sth = $dbh->prepare($sql);
$sth->execute($user_id);

my $total_cost;

print "<table>
               <th>Name</th>
               <th>Price</th>
               <th>Quantity</th>\n";

# Iterate over the elements of the shopping cart
while (my $rowref = $sth->fetchrow_arrayref())
{
    my ($item_id, $item_name, $item_price,
            $item_quantity) = @$rowref;

    $total_cost += $item_price * $item_quantity;

    print "<tr><td>$item_name</td>
                <td>$item_price</td>
                <td>$item_quantity</td></tr>\n";
}

print "<tr><td>Total cost:</td>
           <td>$total_cost</td></tr>
       </table>\n";
```

The first few times you write such code, it doesn't seem so bad. But after a while, it begins to grate on you. Why are you writing so much SQL, when all you want is the elements of your shopping cart? Even if you wrap the SQL inside of an object, you'll find yourself creating many such objects over the course of a project.

The people who wrote Zope, a Python-based Web application framework, decided that although relational databases have their place, the real solution to this problem is to avoid the object-table translation as much as possible, opting instead for an object database. ZODB (Zope Object Database) thus allows you to store and retrieve Python objects as part of a hierarchy. If you can represent data in a Python object, ZODB makes it easy to keep that data persistently.

But of course, ZODB has its problems as well. To begin with, you can use it only from Python; by contrast, relational databases typically can be accessed from any number of languages. And although ZODB now has multiversion concurrency control (MVCC), transactions and a host of other features, the fact that it simply stores a set of objects means that you can't easily sort, search or perform "joins", which are the cornerstone of the relational world.

Object-Relational Mappers

The third alternative, namely that of having an object-relational mapper, has become increasingly popular. The basic idea is pretty simple. Your program uses objects, and those objects are automatically transformed into rows, columns and tables in a relational database.

For many years, object-relational mappers have had all sorts of difficulties, particularly when working with sophisticated data sets. But they are now increasingly robust and impressive; and though I have not worked with either of them, Hibernate (for Java programmers) and SQLObject (for Python programmers) offer just these sorts of services, and Alzabo (described in this column several years ago) provides such services for Perl programmers. When implemented correctly, object-relational mappers provide the best of both worlds, including all of the speed, cross-language and maintenance benefits of a relational database along with the flexibility and consistency of working with objects from within the code.

When Rails burst onto the Web development scene about a year ago, its proponents touted the fact that Rails allows you to produce a Web/database application with almost no configuration and with very little code. And indeed, this is the case, thanks to several different features. One of the key features that makes this possible, however, is a sophisticated object-relational mapper known as ActiveRecord.

ActiveRecord is a Ruby class that is traditionally used as the parent of model classes within a Rails application. As you may recall, Rails uses the traditional model-view-controller (MVC) paradigm to build Web applications. Unlike some MVC application frameworks, Rails makes the differences between these explicit, creating models, views and controllers subdirectories within the application's app directory. A model class in Rails doesn't have to inherit from ActiveRecord, in which case it functions like any other data structure or class. But if it does inherit from ActiveRecord (or more precisely, from ActiveRecord::Base), the object knows how to store and retrieve its values from a table in a relational database.

At this point, you might be asking, "Wait a second—how is it possible that inheritance alone can provide an object-relational mapping? Don't I need to configure something?" The short answer, amazing as it might seem, is "no". There is, of course, a slight trade-off, one that might bruise your ego if you aren't careful. Rails is able to accomplish this magic by forcing all programs to adhere to a particular set of conventions. Indeed, one of the Rails mantras is "convention over configuration." If you are willing to name your tables, columns and objects according to the accepted convention, Rails will reward you handsomely. If you insist on using your own conventions, or if you want to connect Rails to an existing set of tables, you might find yourself struggling to implement even the simplest application.

Connecting

So, how do we connect Rails to our database? Much of the documentation I have seen uses the popular open-source MySQL database for its examples; I strongly prefer PostgreSQL, and thus use it in my examples instead. However, you will soon see that the choice of a back-end database is almost invisible when it comes to Rails.

If you haven't done so already, install the Ruby Gems package, and then use the gem command to install Rails, all of its dependent classes and postgres-pr:

```
$ gem install --remote rails
$ gem install --remote postgres-pr
```

Now we use the rails command to create a new Rails application. If you still don't have the Weblog application we began last month, you can create it by typing:

```
$ rails blog
```

In many Web/database frameworks, the individual page or

program must connect to the database each time. In Rails, the underlying system connects to the database for us, automatically tying the database connection to the ActiveRecord object class. The configuration is kept under the application directory in config/database.yml. No, that's not a typo; the extension is yml (YAML, or Yet Another Markup Language, or YAML Ain't a Markup Language), a simplified text format that is easier to read, write and parse than XML.

Traditionally, every Rails application uses three different databases, one each for development, testing and production. These three databases are created with a prefix that reflects the application name and a suffix that reflects its use (either development, test or production). For example, this is the database.yml file for the blog application:

```
development:
  adapter: postgresql
  database: blog_development
  host: localhost
  username: blog
  password:

test:
  adapter: postgresql
  database: blog_test
  host: localhost
  username: blog
  password:

production:
  adapter: postgresql
  database: blog_production
  host: localhost
  username: blog
  password:
```

Notice how the database adapter name is postgresql, even though I used the postgres-pr gem to connect to it. Also notice that the database is accessed by a user named blog. For this to work correctly, I now have to create the blog user in PostgreSQL (not as a Linux user):

```
$ /usr/local/pgsql/bin/createuser -U postgres blog
Shall the new user be allowed to create databases? (y/n)
y
Shall the new user be allowed to create more new users?
(y/n) n
CREATE USER
```

Now that we have created the blog user, we use it to create the three databases:

```
$ /usr/local/pgsql/bin/createdb -U blog blog_development
CREATE DATABASE
$ /usr/local/pgsql/bin/createdb -U blog blog_test
CREATE DATABASE
$ /usr/local/pgsql/bin/createdb -U blog blog_production
CREATE DATABASE
```

Finally, we should create a table in our database. We use only the development database for now, but we adhere to the convention of writing our table definitions in the blog/db directory, in a file named create.sql:

```
CREATE TABLE Blogs (
id            SERIAL    NOT NULL,
title         TEXT      NOT NULL,
contents      TEXT      NOT NULL,

 PRIMARY KEY(id)
);
```

I have already mentioned the importance of following Rails conventions when working with the ActiveRecord object-relational mapper, and the above table definition, as simple as it seems, already uncovers two of them. To begin with, every row has a unique ID field named id. (PostgreSQL, following SQL standards, has case-insensitive table and column names by default.) In PostgreSQL, we ensure that every row has a unique value of id by declaring it to be a SERIAL type. If you're like me, and have always used more explicit names (such as, blog_id) for the primary key, you'll need to change in order to work with Rails.

Another convention, and one that is a bit more subtle to notice, is that our table name is Blogs, a plural word. A class descended from ActiveRecord::Base is automatically mapped to a database table with the same name, but pluralized. So if we create a blog class that inherits from ActiveRecord::Base in models/blog.rb, it is automatically mapped to the blogs table in our database. As you can see, your choice of a name can affect the readability of your code; be sure to choose a name that makes sense in a number of different contexts, both singular and plural. (In this case, my choice of words was admittedly unfortunate, because each row of the Blogs table represents one posting, rather than one Weblog.)

But it gets better—we don't need to create blog.rb ourselves, at least not at first. We can ask Rails to create it for us, using script/generate. script/generate can be used to create a model, controller or view; in this case, we create our model:

```
ruby script/generate model blog
```

You will see some output that looks like this:

```
exists  app/models/
exists  test/unit/
exists  test/fixtures/
create  app/models/blog.rb
create  test/unit/blog_test.rb
create  test/fixtures/blogs.yml
```

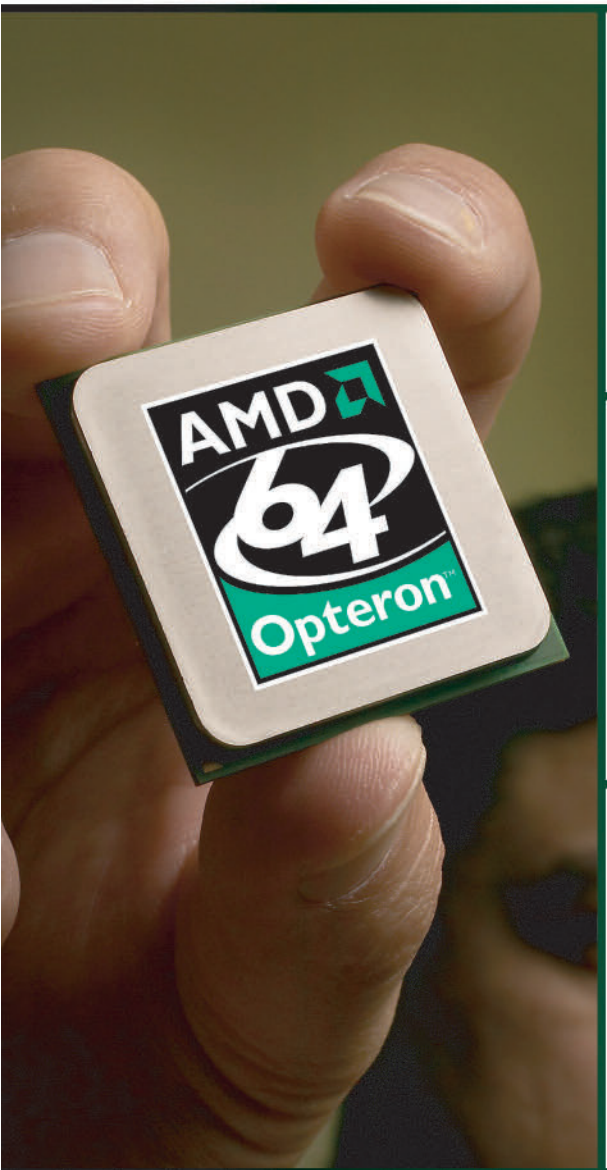
If we open up app/models/blog.rb, we see that it's nearly empty:

```
class Blog < ActiveRecord::Base
end
```

Although we can (and will) add new methods to our Blog class, we can actually leave it as it stands. That's because ActiveRecord provides our class with enough skeleton methods that we can get by without them.

8-Way Opteron Server

Introducing the Thinkmate 5U 8-Way AMD Opteron High Performance Server



Thinkmate 5U 8-Way Server

- Supports 8-Way AMD Opteron 800 series processors, including Dual-Core.
- Supports up to 128Gb DDR ECC/Registered Memory
- HTX Pro Technology
- 4x 10/100/1000 Ports, 1x 10/100 Port
- 4x PCI-X Expansion Slots

Thinkmate 4U 4-Way Server

- Supports 4-Way AMD Opteron 800 series processors, including Dual-Core.
- Supports up to 64Gb DDR ECC/Registered Memory
- 8x Hotswap Serial-ATA or SCSI Drives
- Up to 3.2 Terabytes of Storage
- 2x 10/100/1000 Ports
- 4x PCI-X Expansion Slots

Thinkmate 1U 4-Way Server

- Supports 4-Way AMD Opteron 800 series processors, including Dual-Core.
- Supports up to 64Gb DDR ECC/Registered Memory
- 2x 10/100/1000 Ports
- Serial-ATA or SCSI Drive
- Exceptional Density

THINKMATE

(800) 371-1212

www.Thinkmate.com

Although it's nice that we now have a Ruby class that is automatically mapped to our Blogs table in the database, we still have to access our table via the Web. This means we need to create a controller class, because controllers (the C in MVC) are the components in Rails that handle incoming HTTP requests. We can generate a controller automatically:

```
ruby script/generate controller blogadmin
```

Unfortunately, this controller isn't tied to our class at all. And although we could make such a connection ourselves, the fact that we're at the very beginning of our application definition means we can take a bit of a shortcut, asking Rails to generate an entire set of scaffolding, or bare-bones classes, that will do much of what we want. Creating such scaffolding is a great way to get jump-started with Rails development or even for working on a new project. At the same time, generating the scaffolding means blowing away class definitions you already have written. Because we have (so far) used only the default classes, this shouldn't be much of a problem.

We generate the scaffolded application with:

```
ruby script/generate scaffolding Blog Admin
```

(You should answer "Y" or "a" to replace one or all of the existing files, as appropriate.)

This creates a controller class named Admin that gives us basic access to a Blog class. The latter then connects to the Blogs table in the database.

With only the scaffolding in place, we can now start the server:

```
ruby script/server
```

Then, we point our browser to the application, at the /admin URL: http://localhost:3000/admin.

Sure enough, we see—nothing at all, aside from a few links that let us add a new entry into our Blogs table. If you click on add, you now will see a form that lets you create a new Weblog entry. These automatically generated pages are in the app/views subdirectory. In particular, look at new.rhtml and list.rhtml in app/views/admin. You can, of course, change these views—and in a production application, you will. But for getting your feet wet with Rails, or just trying out an application idea, this is indeed pretty useful.

Now, when you go to the add page, you might be surprised to discover that there is one field for each of the columns in the Blogs table, except for id. This is the result of some cleverness on the part of the automatically generated scaffolding code; it looked at the table definitions and decided what kind of input area to show. What happens if we add another column to our Blogs table that represents when the blog entry was added? (After all, a Weblog whose contents aren't sorted in date order isn't going to be very useful.)

To save time, we simply go in and modify our table definition, using the ALTER TABLE command:

```
$ psql -U blog blog
% ALTER TABLE Blogs ADD COLUMN posted_at
        TIMESTAMP NOT NULL DEFAULT NOW();
```

If you look at the table definition (with the \d command in the psql client program), you'll see that it now has a new column named posted_at. The naming conventions in Rails extend to the names of columns; columns of type DATE should be named xxx_on, and columns of type TIMESTAMP (that is, both date and time) should be named xxx_at.

We now need to regenerate our scaffolding code, blowing away any previous version that might have existed (which is okay in this particular case):

```
ruby script/generate scaffolding Blog Admin
```

Next, restart the server and go back to the new blog page. You will see that it has changed, so that it now includes a posted at field. Moreover, you can't enter arbitrary text there; a full-blown date-entry set of selection lists is in place. If you ever have written code to handle the entry of dates in a Web application, this alone should be a pleasant change.

Finally, take some time to explore both the application (using your Web browser) and the updates that take place in the database as you add, modify and delete rows. Without having written even a single line of Ruby code, you should find yourself able to use the Web-based forms to modify the database. If you want to be a bit adventurous, you can even modify list.rhtml, which shows you the current list of blog entries.

Conclusion

Many Web/database frameworks have struggled to offer a persistent storage layer that interfaces cleanly with the programming language itself. Embedded SQL code isn't too terrible on a small scale, but even a medium-size application can result in a great deal of SQL queries in the middle of an otherwise object-oriented application. The Rails solution strikes a balance that I find quite pleasing, forcing very small, logical changes on me in exchange for a great deal of time savings.

Of course, it's not very hard to create an object-relational mapper when all you need to worry about is column types and individual tables. Moreover, you'll quickly discover that as written, our simple blog application has several problems. To begin with, it has an administrative interface, but no method for displaying the blog to the world! Also, it doesn't display blog entries in any sort of chronological order. Next month, we will see how to solve these problems, as well as how Rails enforces data integrity with a few simple lines in our model definitions.

Resources for this article: www.linuxjournal.com/article/8526.

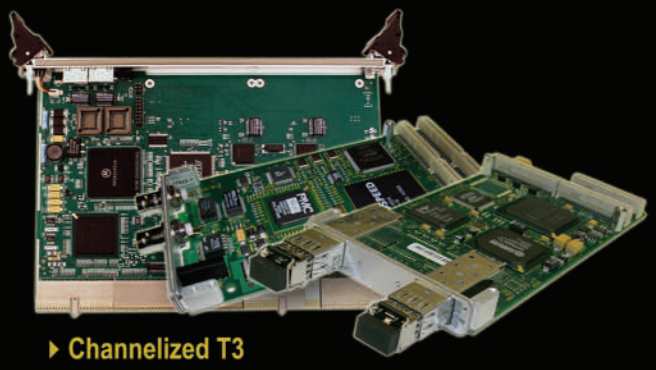
Reuven M. Lerner, a longtime Web/database consultant and developer, now is a graduate student in the Learning Sciences program at Northwestern University. His Weblog is at **altneuland.lerner.co.il**, and you can reach him at reuven@lerner.co.il.

The Power of Choice...

Command the game with your next I/O move.

Modularity. Scalability. Reliability. Cost-effectiveness. These represent the solid foundations that SBE delivers to OEMs for building innovative end solutions. Partnering with SBE for networking and communications I/O solutions allows you to take advantage of proven technology and field-tested products designed to optimize performance for your unique application needs.

SBE offers a full spectrum of interface cards, ranging from T1 and T3 to Gigabit Ethernet and IPsec/SSL acceleration. These boards are available in multiple form factors, including PCI, PMC, and PTMC. Customers have the choice of buying these boards individually or bundling any of the PMC/PTMC modules with our intelligent core processing platforms to create a flexible, cost-efficient blade solution ideal for serving demanding telecom applications. Full Linux support is available on every board.



- Channelized T3
- 24-port T1/E1
- LAN/Ethernet
- Storage
- IPsec/SSL Encryption
- Blade platforms
- I/O and beyond...

sbe®

Linux *On Demand*

flexibility on demand | 925-355-2000 | info@sbei.com | www.sbei.com

Intro to inotify

Applications that watch thousands of files for changes, or that need to know when a storage device gets disconnected, need a clean, fast solution to the file change notification problem. Here it is. BY ROBERT LOVE

John McCutchan and I had been working on inotify for about a year when it was finally merged into Linus' kernel tree and released with kernel version 2.6.13. Although a long struggle, the effort culminated in success and was ultimately worth every rewrite, bug and debate.

What Is inotify?

inotify is a file change notification system—a kernel feature that allows applications to request the monitoring of a set of files against a list of events. When the event occurs, the application is notified. To be useful, such a feature must be simple to use, lightweight with little overhead and flexible. It should be easy to add new watches and painless to receive notification of events.

To be sure, inotify is not the first of its kind. Every modern operating system provides some sort of file notification system; many network and desktop applications require such functionality—Linux too. For years, Linux has offered dnotify. The problem was, dnotify was not very good. In fact, it stank.

dnotify, which ostensibly stands for directory notify, was never considered easy to use. Sporting a cumbersome interface and several painful features that made life arduous, dnotify failed to meet the demands of the modern desktop, where asynchronous notification of events and a free flow of information rapidly are becoming the norm. dnotify has, in particular, several problems:

- dnotify can watch only directories.
- dnotify requires maintaining an open file descriptor to the directory that the user wants to watch. First, this open file descriptor pins the directory, disallowing the device on which it resides from being unmounted. Second, watching a large number of directories requires too many open file descriptors.
- dnotify's interface to user space is signals. Yes, seriously, signals!
- dnotify ignores the issue of hard links.

The goal, therefore, was twofold: design a first-class file notification system and ensure that all of the deficiencies of dnotify were addressed.

inotify is an inode-based file notification system that does not require a file ever be opened in order to watch it. inotify does not pin filesystem mounts—in fact, it has a clever event that notifies the user whenever a file's backing filesystem is unmounted. inotify is able to watch any filesystem object whatsoever, and when watching directories, it is able to tell the user the name of the file inside of the directory that changed. dnotify can report only that *something* changed, requiring applications to maintain an in-memory cache of stat() results and compare for any changes.

Finally, inotify is designed with an interface that user-space application developers would want to use, enjoy using and benefit from using. Instead of signals, inotify communicates with applications via a single file descriptor. This file descriptor is select-, poll-, epoll- and read-able. Simple and fast—the world is happy.

Getting Started with inotify

inotify is available in kernel 2.6.13-rc3 and later. Because some bugs were found and subsequently fixed right after that release, kernel 2.6.13 or later is recommended. The inotify system calls, being the new kids on the block, might not yet be supported in your system's version of the C library, in which case the header files listed in the on-line Resources will provide the necessary C declarations and system call stubs.

If your C library supports inotify, all you should need is the following:

```
#include <sys/inotify.h>
```

If not, grab the two header files, stick them in the same directory as your source files, and use the following:

```
#include "inotify.h"
#include "inotify-syscalls.h"
```

The following examples are in straight C. You can compile them the same as any other C application.

Initialize, inotify!

inotify is initialized via the inotify_init() system call, which instantiates an inotify instance inside the kernel and returns the associated file descriptor:

```
int inotify_init (void);
```

On failure, inotify_init() returns minus one and sets errno as appropriate. The most common errno values are EMFILE and ENFILE, which signify that the per-user and the system-wide open file limit was reached, respectively.

Usage is simple:

```
int fd;

fd = inotify_init ();
if (fd < 0)
        perror ("inotify_init");
```

routerOS: routerBOARD.com

User Management

- support more than 3000 PPPoE or HotSpot clients
  - full RADIUS support for user parameters
  - tx/rx speed, address, filter rules
  - supports RADIUS real time modification of parameters while users are online
- Peer to Peer protocol control (P2P)
  - per client P2P tx/rx rules with burst support
  - P2P pool
  - complete blocking of P2P

Wireless AP and Backbone

- Wireless monitoring
  - Frequency scanning with detailed report
  - Raw wireless packet sniffer
    - streaming option to Ethereal analyzer
    - option to save to a file format supported by Ethereal
  - Snooper packet inspection
    - analyzes all raw frames received for wireless parameters
    - monitors a single channel or all channels

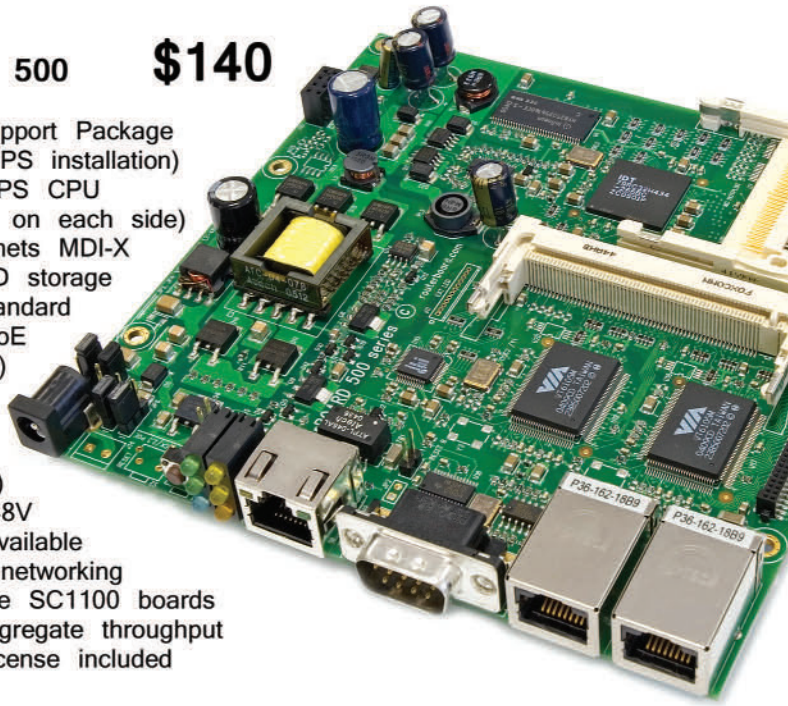
- Nstreme wireless polling protocol
  - no decrease in speed over long distances (as seen with the 802.11 ACK packet bottleneck)
  - polling improves speed and eliminates contention for access to the wireless bandwidth
  - access point control over Nstreme clients tx data to optimize use of the wireless medium
  - RADIUS support for the access control list including bandwidth settings for wireless clients

- Full 802.11a/b/g support

The above is a brief description of a few features, for more information and a fully featured 24 hour demo go to:

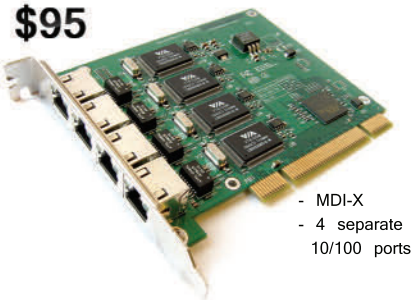
www.mikrotik.com

RouterBOARD 500 $140



- Linux Board Support Package (full Debian MIPS installation)
- 266-400MHz MIPS CPU
- 2 miniPCI (one on each side)
- 3 10/100 Ethernets MDI-X
- 64/128MB NAND storage
- PoE 802.3af standard and passive PoE (also 12V PoE)
- Compact Flash
- Low power
- 32MB DDR (64MB optional)
- 6-24V and 25-48V power mode available
- 2-3x faster for networking than the Geode SC1100 boards
- 200-300Mb/s aggregate throughput
- L3 RouterOS license included

$95



- MDI-X
- 4 separate 10/100 ports

RouterBOARD 44

For the Router Builder !

- rackmount servers and routers
- up to 24 Ethernet ports in a PC
- no more straight/cross cable problems
- server quality VIA VT6105 chips

$195

RouterBOARD 230

No feature left behind !

Integrated router with various interfaces. Use as an AP on a tower with up to 500ft PoE. Includes IDE/CF, miniPCI, USB, PCMCIA, UART, PCI, GPIO, LCD controller, Linux SDK, and more.

$120 $65

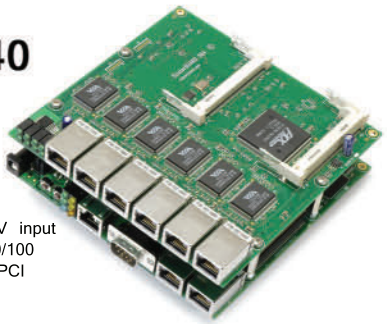
Eight ports Four ports

RouterBOARD 11/14/18

Multi radio tower !

MiniPCI to PCI adapters for multi radio system. Tested with sixteen radios in one Router/AP.

$240



- PoE
- 10-56V input
- 9x 10/100
- 6x mPCI

RouterBOARD 500 & RouterBOARD 564

The Wireless Switchboard !

For a complete multi-radio tower system, the RouterBOARD 500 can carry a daughterboard (RouterBOARD 564) which adds six ethernets and four miniPCI.

contact sales@routerboard.com or go to www.routerboard.com

Watches

The heart of inotify is the watch, which consists of a pathname specifying what to watch and an event mask specifying what to watch for. inotify can watch for many different events: opens, closes, reads, writes, creates, deletes, moves, metadata changes and unmounts. Each inotify instance can have thousands of watches, each watch for a different list of events.

Adding Watches

Watches are added with the inotify_add_watch() system call:

```
int inotify_add_watch (int fd, const char *path, __u32 mask);
```

A call to inotify_add_watch() adds a watch for the one or more events given by the bitmask mask on the file path to the inotify instance associated with the file descriptor fd. On success, the call returns a watch descriptor, which is used to identify this particular watch uniquely. On failure, minus one is returned and errno is set as appropriate.

Usage is simple:

```
int wd;

wd = inotify_add_watch (fd,
                  "/home/rlove/Desktop",
                  IN_MODIFY | IN_CREATE | IN_DELETE);

if (wd < 0)
         perror ("inotify_add_watch");
```

This example adds a watch on the directory /home/rlove/Desktop for any modifications, file creations or file deletions.

Table 1 shows valid events.

Table 1. Valid Events

Event	Description
IN_ACCESS	File was read from.
IN_MODIFY	File was written to.
IN_ATTRIB	File's metadata (inode or xattr) was changed.
IN_CLOSE_WRITE	File was closed (and was open for writing).
IN_CLOSE_NOWRITE	File was closed (and was not open for writing).
IN_OPEN	File was opened.
IN_MOVED_FROM	File was moved away from watch.
IN_MOVED_TO	File was moved to watch.
IN_DELETE	File was deleted.
IN_DELETE_SELF	The watch itself was deleted.

Table 2 shows the provided helper events.

Table 2. Helper Events

Event	Description
IN_CLOSE	IN_CLOSE_WRITE | IN_CLOSE_NOWRITE
IN_MOVE	IN_MOVED_FROM | IN_MOVED_TO
IN_ALL_EVENTS	Bitwise OR of all events.

As an example, if an application wanted to know whenever the file safe_combination.txt was opened or closed, it could do the following:

```
int wd;

wd = inotify_add_watch (fd,
                  "safe_combination.txt",
                  IN_OPEN | IN_CLOSE);

if (wd < 0)
         perror ("inotify_add_watch");
```

Receiving Events

With inotify initialized and watches added, your application is now ready to receive events. Events are queued asynchronously, in real time as the events happen, but they are read synchronously via the read() system call. The call blocks until events are ready and then returns all available events once any event is queued.

Events are delivered in the form of an inotify_event structure, which is defined as:

```
struct inotify_event {
      __s32 wd;            /* watch descriptor */
      __u32 mask;          /* watch mask */
      __u32 cookie;        /* cookie to synchronize two events */
      __u32 len;           /* length (including nulls) of name */
      char name[0];        /* stub for possible name */
};
```

The wd field is the watch descriptor originally returned by inotify_add_watch(). The application is responsible for mapping this identifier back to the filename.

The mask field is a bitmask representing the event that occurred.

The cookie field is a unique identifier linking together two related but separate events. It is used to link together an IN_MOVED_FROM and an IN_MOVED_TO event. We will look at it later.

The len field is the length of the name field or nonzero if this event does not have a name. The length contains any potential padding—that is, the result of strlen() on the name field may be smaller than len.

The name field contains the name of the object to which the event occurred, relative to wd, if applicable. For example, if a watch for writes in /etc triggers an event on the writing to /etc/vimrc, the name field will contain vimrc, and the wd field will link back to the /etc watch. Conversely, if watching the file /etc/fstab for reads, a triggered read event will have a len of zero and no associated name whatsoever, because the watch descriptor associates directly with the affected file.

The size of name is dynamic. If the event has no associated filename, no name is sent at all and no space is consumed. If the event does have an associated filename, the name field is dynamically allocated and trails the structure for len bytes. This approach allows the name's length to vary in size and consume no space when not needed.

Because the name field is dynamic, the size of the buffer passed to read() is unknown. If the size is too small, the system call returns zero, alerting the application. inotify, however, allows user space to "slurp" multiple events at once. Consequently, most applications should pass in a large buffer, which inotify will fill with as many events as possible.

It sounds complicated, but usage is simple:

```
/* size of the event structure, not counting name */
#define EVENT_SIZE  (sizeof (struct inotify_event))

/* reasonable guess as to size of 1024 events */
#define BUF_LEN        (1024 * (EVENT_SIZE + 16)

char buf[BUF_LEN];
int len, i = 0;

len = read (fd, buf, BUF_LEN);
if (len < 0) {
         if (errno == EINTR)
                  /* need to reissue system call */
         else
                  perror ("read");
} else if (!len)
         /* BUF_LEN too small? */

while (i < len) {
         struct inotify_event *event;

         event = (struct inotify_event *) &buf[i];

         printf ("wd=%d mask=%u cookie=%u len=%u\n",
                  event->wd, event->mask,
                  event->cookie, event->len);

         if (event->len)
                  printf ("name=%s\n", event->name);

         i += EVENT_SIZE + event->len;
}
```

This approach is undertaken to allow many events to be read and processed in a single swoop and to deal with the dynamically sized name. Clever readers will immediately question whether the following code is safe with respect to alignment requirements:

```
while (i < len) {
         struct inotify_event *event;
```

```
        event = (struct inotify_event *) &buf[i];

        /* ... */

        i += EVENT_SIZE + event->len;
}
```

Indeed, it is. This is the reason that the len field may be longer than the string's length. Additional null characters may follow the string, padding it out to a size that ensures the following structure is properly aligned.

But I Don't Want to Read!

Having to sit blocked on a read() system call does not sound very appealing, unless your application is heavily threaded—in which case, hey, just one more thread! Thankfully, the inotify file descriptor can be polled or selected on, allowing inotify to be multiplexed along with other I/O and optionally integrated into an application's mainloop.

Here is an example of monitoring the inotify file descriptor with select():

```
struct timeval time;
fd_set rfds;
int ret;

/* timeout after five seconds */
time.tv_sec = 5;
time.tv_usec = 0;

/* zero-out the fd_set */
FD_ZERO (&rfds);

/*
 * add the inotify fd to the fd_set -- of course,
 * your application will probably want to add
 * other file descriptors here, too
 */
FD_SET (fd, &rfds);

ret = select (fd + 1, &rfds, NULL, NULL, &time);
if (ret < 0)
        perror ("select");
else if (!ret)
        /* timed out! */
else if (FD_ISSET (fd, &rfds)
        /* inotify events are available! */
```

You can follow a similar approach with pselect(), poll() or epoll()—take your pick.

Events

The mask field in the inotify_event structure describes the event that occurred. In addition to the events listed earlier, Table 3 shows events that are also sent, as applicable.

Additionally, the bit IN_ISDIR is set telling the application if the event occurred against a directory. This is more than just a convenience—consider the case of a deleted file.

Because flags such as IN_ISDIR are present in the bitmask, it never should be compared to a possible event directly.

Table 3. Events That Cover General Changes	
Name	Description
IN_UNMOUNT	The backing filesystem was unmounted.
IN_Q_OVERFLOW	The inotify queue overflowed.
IN_IGNORED	The watch was automatically removed, because the file was deleted or its filesystem was unmounted.

Instead, the bits should be tested individually. For example:

```
if (event->mask & IN_DELETE) {
        if (event->mask & IN_ISDIR)
                printf ("Directory deleted!\n");
        else
                printf ("File deleted!\n");
}
```

Modifying Watches

A watch is modified by calling inotify_add_watch() with an updated event mask. If the watch already exists, the mask is simply updated and the original watch descriptor is returned.

Removing Watches

Watches are removed with the inotify_rm_watch() system call:

```
int inotify_rm_watch (int fd, int wd);
```

A call to inotify_rm_watch() removes the watch associated with the watch descriptor wd from the inotify instance associated with the file descriptor fd. The call returns zero on success and negative one on failure, in which case errno is set as appropriate.

Usage, as usual, is simple:

```
int ret;

ret = inotify_rm_watch (fd, wd);
if (ret)
        perror ("inotify_rm_watch");
```

Shutting inotify Down

To destroy any existing watches, pending events and the inotify instance itself, invoke the close() system call on the inotify instance's file descriptor. For example:

```
int ret;

ret = close (fd);
if (ret)
        perror ("close");
```

One-Shot Support

If the IN_ONESHOT value is OR'ed into the event mask at watch addition, the watch is atomically removed during generation of the first event. Subsequent events will *not* be generated against the file until the watch is added back. This behavior is desired by some applications, for example, Samba, where one-shot support mimics the behavior of the file change notification system on Microsoft Windows.

Usage is, naturally, simple:

```
int wd;

wd = inotify_add_watch (fd,
                "/home/rlove/Desktop",
                IN_MODIFY | IN_ONESHOT);

if (wd < 0)
        perror ("inotify_add_watch");
```

On Unmount

One of the biggest issues with dnotify (aside from the signals and basically everything else) is that a dnotify watch on a directory requires that said directory remain open. Consequently, watching a directory on, say, a USB keychain drive prevents the drive from unmounting. inotify solves this problem by not requiring that any file be open.

inotify takes this one step further, though, and sends out the IN_UNMOUNT event when the filesystem on which a file resides is unmounted. It also automatically destroys the watch and cleanup.

Moves

Move events are complicated because inotify may be watching the directory that the file is moved to or from, but not the other. Because of this, it is not always possible to alert the user of the source *and* destination of a file involved in a move. inotify is able to alert the application to both only if the application is watching both directories.

In that case, inotify emits an IN_MOVED_FROM from the watch descriptor of the source directory, and it emits an IN_MOVED_TO from the watch descriptor of the destination directory. If watching only one or the other, only the one event will be sent.

To tie together two disparate moved to/from events, inotify sets the cookie field in the inotify_event structure to a unique nonzero value. Two events with matching cookies are thus related, one showing the source and one showing the destination of the move.

Obtaining the Size of the Queue

The size of the pending event queue can be obtained via FIONREAD:

```
unsigned int queue_len;
int ret;

ret = ioctl (fd, FIONREAD, &queue_len);
if (ret < 0)
        perror ("ioctl");
else
        printf ("%u bytes pending in queue\n",
queue_len);
```

This is useful to implement throttling: reading from the queue only when the number of events has grown sufficiently large.

Configuring inotify

inotify is configurable via procfs and sysctl.

/proc/sys/filesystem/inotify/max_queued_events is the maximum number of events that can be queued at once. If the queue reaches this size, new events are dropped, but the IN_Q_OVERFLOW event is always sent. With a significantly large queue, overflows are rare even if watching many objects. The default value is 16,384 events per queue.

/proc/sys/filesystem/inotify/max_user_instances is the maximum number of inotify instances that a given user can instantiate. The default value is 128 instances, per user.

/proc/sys/filesystem/inotify/max_user_watches is the maximum number of watches per instance. The default value is 8,192 watches, per instance.

These knobs exist because kernel memory is a precious resource. Although any user can read these files, only the system administrator can write to them.

Conclusion

inotify is a simple yet powerful file change notification system with an intuitive user interface, excellent performance, support for many different events and numerous features. inotify is currently in use in various projects, including Beagle, an advanced desktop indexing system, and Gamin, a FAM replacement.

What application will use inotify next?

Resources for this article: www.linuxjournal.com/article/8534.

Robert Love is a senior kernel hacker in Novell's Ximian Desktop group and the author of *Linux Kernel Development* (SAMS 2005), now in its second edition. He holds degrees in CS and Mathematics from the University of Florida. Robert lives in Cambridge, Massachusetts.

Hack the Net? No, *NetHack*.

One of the oldest games on your system has a convoluted history, deep, complicated dungeons and some spiffy new graphical front ends.

BY MARCEL GAGNÉ

François, although I am very impressed with your initiative in documenting your experience in network security, that document will need some changes. Of course, I have not read it yet, *mon ami*, but I still know it needs some changes. Well, the title, for starters—somehow, I don't think you can call it "The Guide to Net Hack". *NetHack* is a game, François, and it has nothing to do with network security. Well, not much, anyhow.

Quoi? You have never heard of *NetHack*? *Mon Dieu, mon ami!* This is something we must resolve immediately, if not sooner. Unfortunately, it is time to open the restaurant and our guests will be here momentarily, but perhaps...ah, too late, they are already here! Welcome, everyone, to *Chez Marcel*, home of the finest in Linux fare and, of course, the most extensive wine cellar in the Linux world. Please sit and make yourselves comfortable. François will fetch your wine *immédiatement*. François, please head down to the wine cellar and bring back the 1999 Catena Alta Cabernet Sauvignon from Argentina.

Just before you walked in, François made a rather humorous mistake, telling me he was writing a network security guide about Net Hack, not realizing that *NetHack* is a game. For those of you who may not know, *NetHack* is one of the most popular dungeon-crawling games of all time, and it has been around seemingly forever. Back when I first started playing *NetHack*, it was just called *Hack* (and before that, there was a game called *Rogue*). If you want the juicy details, a nice, concise history of the game is available from inside the game itself (press the question mark during game play). Over time, the game was transformed by a huge number of people scattered from one side of the planet to the other. The code also was ported to many different platforms and operating systems so you could play *Hack* or *NetHack* on just about any machine imaginable. *Hack* is gone, but *NetHack* lives and breathes to this day. This is a game that has captured the imaginations of scores of Netizens and continues to be a hugely popular game. Amazingly, *NetHack* in its pure form is a text-only adventure game (Figure 1), and it still often is played that way.

In text mode, and with scores of beautiful graphical games to pull from, *NetHack* may look too boring to keep anyone

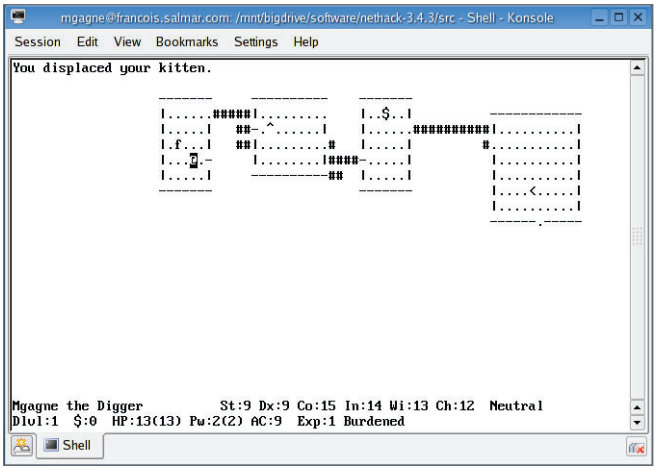


Figure 1. *NetHack* in Text-Only Mode

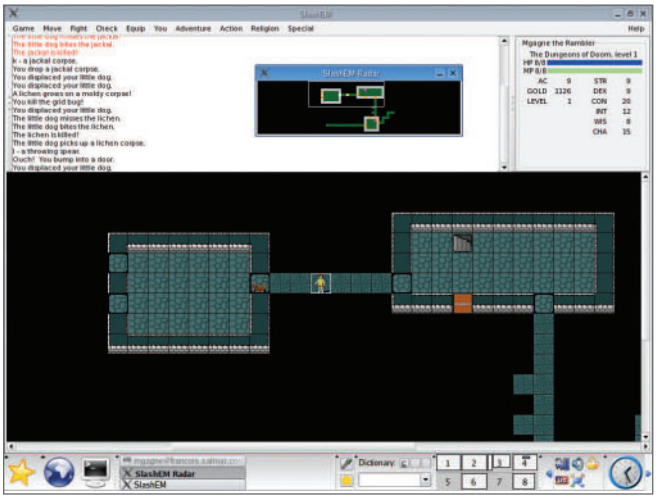


Figure 2. *gtk2hack* brings a clean graphical interface to *NetHack* along with a radar providing feedback on the explored areas.

interested, and yet it still does. After all, your character is an `@`, your dog companion a `d`, a gold piece is a `$` and so on. So why is a game like *NetHack* still so popular? It is because of the incredible richness and complexity of the game. The idea seems simple enough, but this is not an easy game and certainly not one you are likely to win in short order. Deep in the underground levels of the Mazes of Menace (or the Dungeons of Doom) lies the fabled Amulet of Yendor. To the one who finds the amulet, untold riches await along with the gift of immortality bestowed by the Gods. To gain the amulet, you must travel through the dungeons and mazes, encountering puzzles, strange objects, hidden pits from which there is no escape, demons, goblins, grid bugs and other monsters, including the simplest of dangers, hunger and thirst. You may be a barbarian, a monk, a knight, a wizard or merely a tourist. You may be human or not. At your side is a small animal companion, a dog or a cat.

I highly recommend that you check out the text version of the game at the *NetHack* Web site (see the on-line Resources), but make sure you visit Warren Cheung's *SLASH'EM* Web site, home of the "Super Lotsa Added Stuff Hack, Extended Magic" edition of *NetHack*. *SLASH'EM* is *NetHack* kept up to date

with new levels, new monsters, spells and so on. Getting and building *SLASH'EM* is also easier than navigating through the various cryptic instructions for building the official *NetHack*. *SLASH'EM* provides a simple configure script making this an easy extract-and-build five-step:

```
tar -xzvf se008e0.tar.gz
cd slashem-0.0.8E0
./configure
make
su -c "make install"
```

To play, run the command `slashem`. You'll be asked whether you want the program to pick your character's race, role, gender and alignment for you or whether you'd like to choose all of these yourself. I usually prefer to make that choice myself, but you can get some interesting combinations by being brave and going totally random. Once this is done, a small introduction tells you about your character and which god you serve, gives you a nice pat on the back and sends you off to your doom. It's great fun.

With time, and in keeping with *NetHack*'s evolution, graphical versions of the game came to be. By using graphic tiles and an easy-to-use menu-driven interface, the game took on a whole new dimension, all the while maintaining the same core functionality. One of these graphical incarnations is Mihael "miq" Vrbanec's *gtk2hack* (based on the *SLASH'EM* code), a great wrap-around of the latest version of *NetHack* that brings new life to the game (Figure 2).

As the name implies, *gtk2hack* is based on the GTK2 toolkit to provide the interface. It uses a two-dimensional overhead view with nice graphical tiles to display objects, monsters and so on. There's a small "radar" window that accompanies the main display that you can refer to during game play. Above the main graphical window, a game dialog is displayed along with the status of your possessions, your health, hit points, the level you are exploring and so on. If you have become familiar with text *NetHack*, you'll find this equally comfortable. Although you can navigate with mouse clicks, the same keystrokes apply.

Building *gtk2hack* is fairly straightforward—just another slightly modified extract-and-build five-step (skip the configure step). Because it comes with its own *NetHack*/*SLASH'EM* code bundled in, you don't need to download twice. Just remember that the executable is called gtkhack and not gtk2hack as you might logically expect.

One of the best graphical renditions of the game I have seen (and one of my favorite games) is Jaakko Peltonen's awesome *Falcon's Eye*. Although not as up to date as *SLASH'EM* in terms of story and development (it's based on *NetHack* 3.3.1, whereas *SLASH'EM* is based on 3.4.3), you have to try *Falcon's Eye*. It's that great, and if you still aren't hooked on *NetHack*, *Falcon's Eye* is sure to do the job. The dungeons enter the third dimension along with your character, your faithful companion dog and (of course) the monsters (Figure 3). The game is mouse-driven, and the graphics are high-resolution.

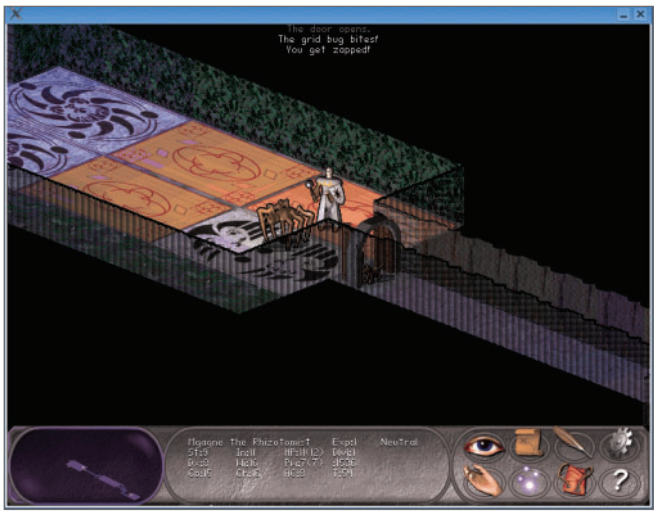


Figure 3. The beautiful *Falcon's Eye*, shown in windowed mode with transparent walls.

There's a slick panel at the bottom of the screen from which you can access your possessions, your spells and other information. Like *gtk2hack*, there's also a small "radar" screen on the lower left so you can get a better view of where you are and where you have been.

Falcon's Eye is available as a source download, but I've found binaries for a number of distributions (Fedora, SUSE, Debian, Mandriva and others), so check your distribution CDs and your distro's contrib sites first.

Falcon's Eye starts in full-screen mode by default, which although cool, isn't what I want when I'm pretending to work while slaying goblins. To change the screen resolution to windowed mode, you need to edit the game's configuration file. It is called jtp_opts.txt, and you'll find it in the game's config directory. Here's the section you are looking for:

```
screen_xsize=800
screen_ysize=600
fullscreen=0
```

In the above example, I've already changed the resolution to windowed mode by setting fullscreen to 0. To return to full-screen mode, change it back to 1. Have a look at the file, and you'll find other interesting changes you might want to make. One is to make the walls transparent, or at least not quite as opaque. The reason you might want to do this is to make it easier to spot objects that might be against the walls as you navigate the dungeons. You can also decide whether you want music or sound effects to accompany your journey.

Speaking of journeys, exploring dungeons is extremely thirsty work, I'd rather avoid those strange potions as long as possible. Luckily, we have a rather generous wine cellar here at *Chez Marcel*. François, if you would be so kind....

The only catch with *NetHack*—okay, there are several catches—the biggest catch is that it may start to take over every bit of free time you have. Should you find yourself so addicted that you need to have *NetHack* with you wherever you go, consider downloading a copy of NetHack Linux. This is a single-floppy Linux distribution that boots up directly into

To gain the amulet, you must travel through the dungeons and mazes, encountering puzzles, strange objects, hidden pits from which there is no escape, demons, goblins, grid bugs and other monsters, including the simplest of dangers, hunger and thirst.

a text-based game of *NetHack*. The most recent image contains *NetHack* version 3.4.3, the latest and greatest.

To get your copy of NetHack Linux, visit Benjamin Schieder's Web site (see Resources) and download the latest diskette image. Then, transfer the image to a diskette with the `dd` command:

```
dd if=nethacklinux_1.1.img of=/dev/fd0
```

To run NetHack Linux, simply pop the diskette in to any free PC's drive, reboot the system and a few seconds later, you are ready to go. A small menu appears from which you can edit the nethackrc file, show the current high scores or simply play the game. Select option one (Play NetHack), and you are ready to go.

I see by the clock on the wall that it is almost closing time. While François refills your glasses one final time this evening, let me direct you to a rather apropos, but strange little Web site. If, after crawling the Maze of Menaces for far too long, you start wondering what kind of *NetHack* monster you would be if you were a *NetHack* monster, I have just the Web site for you. Check out Kevan Davis' "Which NetHack Monster Are You?" site and answer the short questionnaire provided. The results can be entertaining or, in my case, embarrassing. Rather than embarrass myself by telling you, I'll merely point you to the on-line Resources for the address to the site. There's also the #nethack IRC channel on irc.freenode.net where dozens of people talk *NetHack* 24 hours a day. Finally, if you've had just enough wine (and if not, let François know) you may be ready for the *NetHack* theme song. Please raise your glasses, *mes amis*, and let us all drink to one another's health. *A votre santé! Bon appétit!*

Resources for this article: www.linuxjournal.com/article/8531.

Marcel Gagné is an award-winning writer living in Mississauga, Ontario. He is the author of *Moving to the Linux Business Desktop* (ISBN 0-131-42192-1), his third book from Addison-Wesley. He also makes regular television appearances as Call for Help's Linux guy. Marcel also is a pilot and a past Top-40 disc jockey. He writes science fiction and fantasy and folds a mean Origami T-Rex. He can be reached via e-mail at mggagne@salmar.com. You can discover a lot of other things (including great Wine links) from his Web site at www.marcelgagne.com.

Flexibility to power the enterprise.

From mail servers to databases, ZT servers powered by the 64-bit Intel® Xeon™ Processor can run the full range of 32-bit applications and offer extended flexibility for your 64-bit needs. So you can create powerful, all-purpose IT infrastructure that enhances business agility – and the bottom line.

ZT Optimum 1U Server X9519

Intel® Xeon™ Processor 3 GHz
(2MB L2 Cache, 3 GHz, 800MHz FSB)

- Intel® E7320 Chipset Server Board
- 1GB ECC Registered DDR 333 SDRAM (2 x 512MB)
- 2 x Seagate® 300GB 10,000rpm SCSI Hard Drive (Raid 1)
- 4 x 1" SCSI Hot-Swappable Drive Bays
- Slim CD-ROM
- 2 x Intel® 10/100/1000 Gigabit Network Controller
- 1U Rackmount Chassis w/ 420W Power Supply
- 3-Year Limited Warranty + First Year On-site Service

$3,899

ZT Optimum 2U Server X9520

Dual Intel® Xeon™ Processors 3.20 GHz
(2MB L2 Cache, 3.20 GHz, 800MHz FSB)

- Intel® E7320 Chipset Server Board
- 1GB ECC Registered DDR 333 SDRAM (Up to 8GB)
- 4 x Seagate® 500GB SATA2 Hard Drive (Total 2TB Storage)
- 8 x 1" Hot-Swap SATA2 Drive Bays
- 8 Channel High Performance SATA Controller (RAID 0,1,5,10,JBOD Support)
- Slim CD-ROM & Floppy Drive
- 2 x Intel® 10/100/1000 Gigabit Network Controller
- 2U Rackmount Chassis w/ 500W Redundant Power Supply
- 3-Year Limited Warranty + First Year On-site Service

$3,999

ZT Optimum 4U Server X9521

Dual Intel® Xeon™ Processors 3.20 GHz
(2MB L2 Cache, 3.20 GHz, 800MHz FSB)

- Intel® E7320 Chipset Server Board
- 1GB ECC Registered DDR 333 SDRAM (Up to 8GB)
- 4 x Seagate® 250GB SATA Hard Drive (Total 1TB Storage)
- 8 x 1" Hot-Swap SATA Drive Bays
- 4 Channel Raid Controller (RAID 0, 1, 1/0, 5, JBOD Support)
- DVD±RW Burner & Floppy Drive
- 2 x Intel® 10/100/1000 Gigabit Network Controller
- 4U Rackmount Chassis w/ 760Watt Triple-Redundant Power Supply
- 3-Year Limited Warranty + First Year On-site Service

$2,999

1. OEM Computer Manufacturer

- 3 year warranty with lifetime tech support
- Reseller and volume pricing available.

New Accounts Receive Free Gift

- Personal attention (Dedicated Technical Sales Team)
- Call now to customize using the latest technology

Find out how ZT Insider Program can help maximize your Business Solution

Go to **ztgroup.com/go/linuxjournal**

Call **866- ZTGROUP** (866-984-7687) promote code : LJ1105

Two-Factor Authentication

With faster cracking programs available, passwords alone are no longer enough to keep naughty people off of your system. Use a USB device as a second check. BY COREY STEELE

Two-factor authentication aims to solve the decades-old problem of password-based attacks, such as brute-force attacks and key-logging attacks. In Linux, two-factor authentication can be accomplished with pam_usb, a PAM module that provides a means by which you can authenticate against cryptographic tokens stored on removable media, such as a USB drive. Through the marvel of PAM's module chaining, this article walks you through configuring two-factor authentication.

PAM is short for pluggable authentication modules. According to the Linux-PAM home page:

> PAM provides a way to develop programs that are independent of authentication scheme. These programs need authentication modules to be attached to them at run time in order to work. Which authentication module is to be attached is dependent upon the local system setup and is at the discretion of the local system administrator.

pam_usb is a PAM module written by Andrea Luzzardi that facilitates authentication from removable media, such as USB devices, based on strong cryptographic key pairs stored on the drive and on the system itself. pam_usb is available in source form or in binary packages for a variety of distributions, including Debian, Gentoo, Fedora, Mandrake and SUSE. pam_usb lends itself quite nicely to accomplishing two-factor authentication, although it can be used as the sole authentication module.

The term two-factor authentication refers to authentication being achieved using two separate and distinct criteria to authenticate a user's identity: usually this is something the user knows and something the user has. The something the user knows, in the configuration we're building, is the user name and password pair, while the something the user has is the strong cryptographic tokens we are going to generate and store on the USB drive.

Strictly speaking, you should be able to accomplish everything discussed here with any flavor of Linux that has a working PAM configuration and a 2.4 or newer kernel on a system with a supported USB controller. You also need a supported USB drive, the pam_usb module source and a C compiler.

I achieved everything discussed here with a Lexar 128MB Impact USB 1.1 drive on an IBM NetVista with an Intel 82820 Camino USB controller. It is running Debian 3.0 stable with the stock bf kernel (2.4) and gcc-2.3.

You can check to see if your controller and USB drive are supported by attaching your USB drive and running `lsusb` as root. If your controller and drive are supported, you should see the drive listed in the output of `lsusb`. If it isn't, don't despair; your distribution may not have auto-loaded the necessary modules. Consult The USB Guide (see the on-line Resources) for help getting your USB environment set up. Your PAM install can be confirmed by checking to see if your login program is linked against libpam by running `ldd /bin/login | grep -i pam` and checking the output. If login is linked against libpam, your PAM configuration should be set.

The source for the pam_usb module can be downloaded from the project site (see Resources). Use any browser to navigate the Web site and download the latest source tarball. Remember where you save the download. When the download is complete, uncompress the tarball with `tar -zxvf pam_usb-X.Y.Z.tar.gz`, where X, Y and Z are the major, minor and build versions, respectively, of the particular version of pam_usb you downloaded. You now should have a pam_usb-X.Y.Z directory, so `cd` into the directory and take a quick peek to make sure you have some files in the directory.

pam_usb does not have any configure scripts, only a Makefile, so building is simply a matter of running `make` from within the pam_usb-X.Y.Z directory. If you encounter errors, as I did, you probably are missing libraries. On my Debian 3.0 stable system, I was missing the development packages for libncurses5, libpam0g and libreadline4. Once I installed the missing libraries, the make completed without errors. After pam_usb builds, you can install it with `make install` as root from within the pam_usb-X.Y.Z directory.

After the installation is complete, it's time to configure pam_usb. Configuring pam_usb is a relatively straightforward task that can be broken in to three broad steps: creating the pam_usb log file, backing up your existing PAM configuration and installing the new configuration.

Creating the pam_usb log file is a matter of choosing where to put it and what to call it, as well as creating the file. My personal preference is to keep all logs in /var/log, so that's where I set up my pam_usb log file and that is the location used throughout this article. Create the log file with `touch /var/log/pam_usb.log` as root. Next, set the ownership of the /var/log/pam_usb.log file to match the ownership of other files in /var/log, like this:

```
chown $USER:$GROUP /var/log/pam_usb.log
```

where $USER and $GROUP are the user and group that own the other files in /var/log. Once the file has been created and ownership has been set, simply change the permissions on the file to reflect those of the other files in /var/log by using this command:

```
chmod 0600 /var/log/pam_usb.log
```

More advanced users may want to configure a log rotation schedule for the pam_usb.log or even change the file to be append-only with `chattr`. Those options are left as exercises for the reader to explore.

Now that the log file has been set up, we need to back up the existing PAM configuration files. This is an important step, so do not skip it. On most systems, the PAM configuration files are stored in /etc/pam.d. As root, make a backup copy with:

```
cp -rfp /etc/pam.d ~/pam.d/
```

For testing sake, we are working with the PAM configuration for su, because it is the easiest PAM-aware application to test. As a precautionary method, you should keep a root shell open and accessible so that if a mistake is made in configuring pam_usb, you are able to rescue yourself by overwriting the edited configuration files with backups from your ~/pam.d. You also need to know what filesystem is used on the USB drive(s) you will be configuring. In an ideal world, we can use mount to do the work for us, provided /mnt/usb exists and your USB drive is on /dev/sda. Use:

```
mount /dev/sda1 /mnt/usb
```

and then run:

```
mount | grep usb
```

to see what filesystem is on the drive—the filesystem is listed in parentheses at the end of the line. Most USB drives use the vfat filesystem and do not have more than one partition. Thus, they are mountable with:

```
mount -t vfat /dev/sda1 /mnt/usb
```

Our first real step in configuring pam_usb is to alter the PAM-aware applications' PAM configuration file—this step is required for each application you want to use pam_usb to authenticate to. Because we're working with su for testing purposes, focus only on the /etc/pam.d/su file. Do not try to configure every PAM-aware application in a single mass-edit of your /etc/pam.d directory, or tears and sorrow surely will be your lot. The files in /etc/pam.d/ correspond to the applications they configure, so if you were to configure console logins or GNOME Display Manager logins, you would be concerned with /etc/pam.d/login and /etc/pam.d/gdm, respectively. The naming pattern for PAM's configuration files should be relatively self-evident. So, open /etc/pam.d/su in your favorite text editor and add the following line above the pam_unix line:

```
auth required pam_usb.so fs=vfat check_device=-1 \
  check_if_mounted=-1 force_device=/dev/sda \
  log_file=/var/log/pam_usb.log
```

If you do not include the above line before the pam_unix line, PAM never reaches the point of authenticating against the USB device. Instead, it is satisfied by the authentication that occurs through pam_unix, and it drops out of the authentication process.

A few options in the pam_usb configuration that need further explanation: the force_device option, the pam_usb mode, the filesystem of the device and the log file we're going to use.

pam_usb is capable of autodetecting which USB-attached

device houses the authentication keys. By not specifying the force_device directive, pam_usb walks through all of the attached devices and looks for keys matching the specified user name. This is helpful if the machine has multiple USB devices that are assigned device names according to the order in which they were attached—the first device is /dev/sda, the second is sdb and so on. If you specify the force_device directive, you are not able to authenticate unless your USB drive is assigned the device name specified in the PAM configuration.

pam_usb supports three modes of operation: unique, alternative and additional. With unique mode, you can log in using your USB drive, but if it's not present it isn't possible to log in. This is achieved by commenting out pam_unix in $PAMDIR/login and adding the configuration line above. The alternative mode allows you to log in simply by plugging in your USB key. If the key is not present, the system prompts for a password. This is accomplished by leaving pam_unix intact, adding the above configuration line to the PAM configuration file above the pam_unix entry and changing the `auth required` bits of the line to read `auth sufficient`. To achieve a true two-factor authentication, you need to require both the user name/password pair and the USB key, which is how the configuration above is set.

Andrea Luzzardi also points out an alternative two-factor authentication that involves encrypting the private key stored on the USB drive, after which the key requires a password to be decrypted and used for authentication. pam_usb is capable of passing the password provided to PAM through to decrypt the private key, thus accomplishing two-factor authentication off of a single user name and password pair. Furthermore, this is accomplished while not compromising any of the security benefits of having two-factor authentication. This method of authentication is contingent on using the same password for the user account that was used to encrypt the private key used by pam_usb. To encrypt the private key used by pam_usb, simply use the usbadm tool to create the cryptographic token:

```
usbadm cipher /path/to/usb/filesystem \
username algorithm
```

where the options have been specified according to the usbadm man page under cipher.

The fs= option tells pam_usb what filesystem to try to use to mount and read the USB drive. If your users have different filesystems on their USB drives, you'll have trouble with this. Simply specify whatever filesystem is used on your USB drives.

Once you've made the configuration changes to su's PAM configuration, it's time to set up a cryptographic key pair for each user using the system. Initially, this is done simply with:

```
usbadm keygen /path/to/mounted/usb/drive keysize
```

where keysize is the size (in bits) of the keys you want to generate and /path/to/mounted/usb/drive is the—you guessed it—path to the root of your mounted USB drive. For my setup, I chose a key size of 4,096 bits, which should be adequate to prevent even determined brute-force attempts against your key pair. RSA Labs recommends that DSA keys be no smaller than 2,048 bits, so at a minimum use a 2,048-bit key size. The usbadm program generates files in the root of your USB drive called .auth/$USER.$HOST, where $USER is the user name that executed the usbadm command and $HOST is the host-name of the machine on which the keys were generated. A corresponding set of keys in ~$USER/.auth must be present to authenticate with the USB token.

If a USB drive is lost, as is bound to happen, you can remove the user's ~/.auth/id_pub file and follow the instructions above to regenerate the key pair. Be certain you don't lose root's private keys or you'll have to boot to safe media, disable two-factor authentication and go through the whole setup process again to restore functionality.

Having freshly minted your key pair, you now are ready to test pam_usb and two-factor authentication with su. Insert your USB drive and try to `su` to a user who has a valid key pair; it's best to test this from a non-root account. You should be prompted for your user name as before, but instead of being prompted for your password immediately, you now should see a USB error as pam_usb tries to mount /dev/sda, or whatever base device you told it to try. Provided pam_usb was able to locate your USB drive, you should be prompted for the user's password, which if entered correctly, should result in a shell for that user account. You can make sure that the two-factor authentication worked by checking the pam_usb log file and verifying that somewhere near the last line is a line that reads `Access granted`. If you see that line in the pam_usb.log file, congratulations—su now is configured to use two-factor authentication.

Once you are satisfied with the functionality of pam_usb for su, you can duplicate the configuration for su with other applications that you want to set up with two-factor authentication. Be sure to issue all users the necessary keys and thoroughly test things before you log off the system and/or reboot.

As with any authentication system, two-factor authentication is not without its weaknesses. This particular implementation is vulnerable to private key theft, because it's easy to copy the contents of the USB drive. In the March 15, 2005, Crypto-Gram, Bruce Schneier writes a rather scathing article detailing why two-factor authentication is not the end-all-be-all of authentication—the crux of his point is that people are using two-factor authentication to achieve things it wasn't meant to achieve. With that in mind, remember that two-factor authentication is meant to address the age-old problems of password-based attacks. pam_usb achieves that end very well, and if properly configured, it can effectively improve the security of a given workstation.

Resources for this article: www.linuxjournal.com/article/8528.

Corey Steele is a security expert with six years of experience; he received CISSP certification in 2004. His primary interests in the security arena are access control and network security. He works in the financial sector for a company that makes core banking software. He has been an active member of the Free/Libre/Open Source Software community, having contributed to various projects, since 1995. In his spare time, he likes to write code and lecture on security topics.

Reclaim lost time

Levanta Intrepid™

The world's first Linux management appliance

Plug the Levanta Intrepid™ into your network and perform the most important Linux management tasks in a fraction of the time you spend now. And gain power and flexibility that you've never had before:

- **Fast & Portable:** Provision servers or workstations practically anywhere, anytime – in minutes. Swap them around, mix it up.
- **Flexible:** Supports commodity hardware, blades, virtual machines, and even mainframes.
- **Out of the Box:** Includes pre-defined templates for servers, workstations, & software stacks. Or create your own.
- **Total Control:** Track any file changes, by any means, at any time. And undo them at will.
- **Disaster Recovery:** Bring dead machines quickly back to life, even if they're unbootable.

Based upon technology that's already been proven in Fortune 500 enterprise data centers. Now available in a box, priced for smaller environments. **Just plug it in and go.**

30-Day
Money-Back Guarantee

Order online by 11/30/05
Get $500 Off
Enter PROMO CODE: LJ1105

LEVANTA®
www.levanta.com
1.877.LEVANTA

WINNER
LinuxWorld Conference & Expo
Most Innovative Hardware Solution

Simple Linux IP Repeaters to Extend HomePlug Range

Simple Linux-based devices bring real networking features to a system that runs over power lines.

BY FRANCISCO J. GONZÁLEZ-CASTAÑO, PEDRO S. RODRÍGUEZ-HERNÁNDEZ, FELIPE J. GIL-CASTIÑEIRA, MIGUEL RODELGO-LACRUZ AND JOSÉ VALERO-ALONSO

Power line communication (PLC) technology allows you to transmit data by way of the electric grid's low- and medium-voltage power lines. Any device in a building thus may access a LAN to share resources. Figure 1 shows the Ovislink HomePlug Ethernet Bridges we currently are using.

Figure 1. HomePlug Ethernet Bridge

PLC offers obvious advantages, the main one being that it is unnecessary to lay cables as the network infrastructure already is deployed—the electrical grid. Yet, PLC also has strong limitations, such as:

- High attenuation, so it is efficient only across short distances.
- Impedance changes with power cycles, due to the presence of nonlinear devices such as diodes and transformers.
- Occasional impedance changes due to devices switching on and off.
- Reflections due to the home electrical grid topology.
- Power lines often lacking a ground connection.

To avoid these problems, HomePlug uses a robust orthogonal frequency division multiplexing (OFDM) scheme with 1,280 orthogonal quadrature amplitude modulation (QAM) carriers. Consequently, HomePlug's maximum point-to-point range is approximately 200 meters.

To extend the range further, we have developed a simple Linux IP repeater. We have implemented it on both desktops and an embedded microcontroller-based development card. The latter yields a small, low-consumption, low-cost device that could be installed easily in any building location.

Description of the Repeater

We divide the network into class C subnets (Figure 2), such that any two devices within the same subnet see each other. The devices in a subnet can communicate without a repeater, so we need it only when connecting devices in different subnets. A subset of the devices in any of the two subnets can see a subset of the devices in the other.

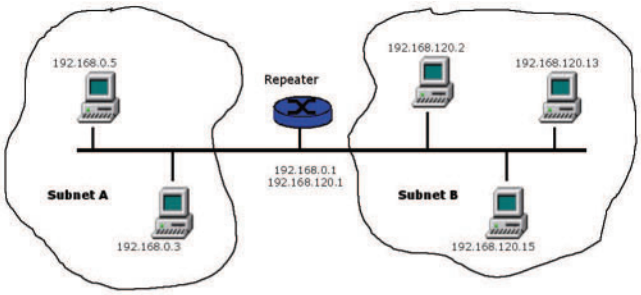


Figure 2. The IP repeater connects two subnets over HomePlug.

Let us assume the repeater initially is installed in parent subnet 192.168.0.X, with address 192.168.0.1 (it could be any address). For any new subnet 192.168.X.X, we reserve IP address 192.168.X.1 for the repeater gateway. When the destination IP address of a packet does not belong to the sender subnet, the repeater routes it. Actually, the repeater does no routing, as the same transmission line supports both packet ingress and egress. Thus, it needs no routing table, and it simply relays packets by using the same medium.

For the repeater to belong to different subnets, it must have several IP addresses. In other words, it is necessary to assign several network interfaces to its Ethernet card. In the example shown in Figure 2, the repeater card has two network interfaces, with respective IP addresses of 192.168.0.1 and 192.168.120.1. In Linux, this is done as follows:

```
# ifconfig eth0:0 192.168.0.1
# ifconfig eth0:1 192.168.120.1
```

The number of subnets is unknown beforehand, thus the repeater must autoconfigure itself. In our trials, we set its IP address to 192.168.0.1, as in typical commercial built-in DHCP servers.

We have implemented repeater self-configuration using a program called hprmanager, now available by e-mail from pedro@det.uvigo.es. This program sets the Ethernet card to promiscuous mode and looks for new subnets in order to register them.

The repeater discovers the subnets it interconnects by capturing every packet circulating in the network. In permanent state, even though the Ethernet card is in promiscuous mode, it does not receive all packets due to the PLC modem placed between the network card and the power line (Figure 2). This PLC modem blocks all packets except those whose destination address is a broadcast one, a multicast one or the repeater address itself. However, the repeater necessarily receives broadcast and multicast packets from unknown subnets. In any case, it also is possible to set network interfaces manually.

Each computer must select the gateway in its own subnet. Assuming we are configuring a computer in subnet 192.168.0.X, it must set 192.168.0.1 as the default gateway:

```
# route add default gw 192.168.0.1
```

To configure the repeater on a desktop Linux machine, it is necessary to do several things:

- Activate the packet forwarding module by adding, for example, the following line to /etc/sysctl.conf:

```
net.ipv4.ip_forward = 1
```

- Assign the default IP address; as previously stated, the repeater has the address 192.168.0.1.

- Start the repeater manager. Assuming it resides in /bin/, simply add this line to /etc/rc.d/rc.local:

```
/bin/hprmanager &
```

This procedure works for most Linux distributions. For those without the /etc/sysctl.conf file—such as Debian—it first is necessary to create a shell script file (beginning with `#! /bin/sh`) called /etc/init.d/local, which includes the line `/bin/hprmanager &`. Finally, one should add the script to the desired run levels, as in:

```
update-rc.d local start 80 2 3 4 5
```

µClinux Version

Because µClinux runs on embedded systems, the settings in the previous section must be active immediately after the load. The default installation of a µClinux operating system does not include the packet relaying module. Thus, we first must compile a kernel with packet relaying support, using the following four configuration steps:

- Enabling the IP: advanced router option in the Networking options section (Figure 3).
- Enabling the /proc filesystem support option in the Filesystems section.
- Enabling the Sysctl support option in the General Setup section.
- Using the board shown in Figure 5, we must disable the hardware byte-swapping support for CS89x0 Ethernet option in the Ethernet (10 or 100Mbit) section (Figure 4).

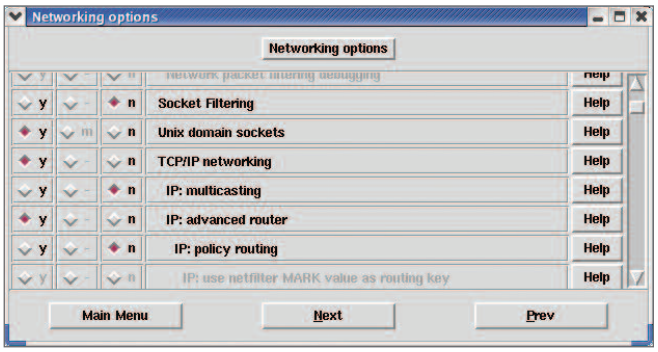


Figure 3. Enable advanced router functionality using the Networking options section of the kernel configuration menu.

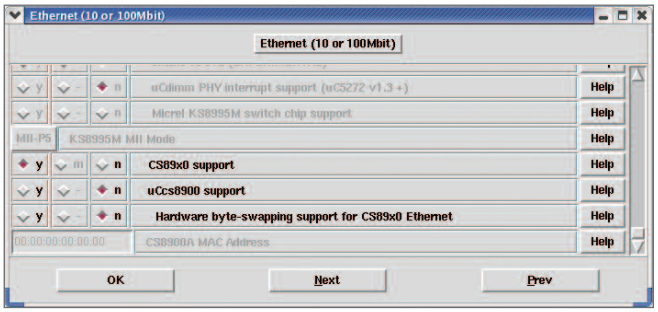


Figure 4. Ethernet Card Configuration

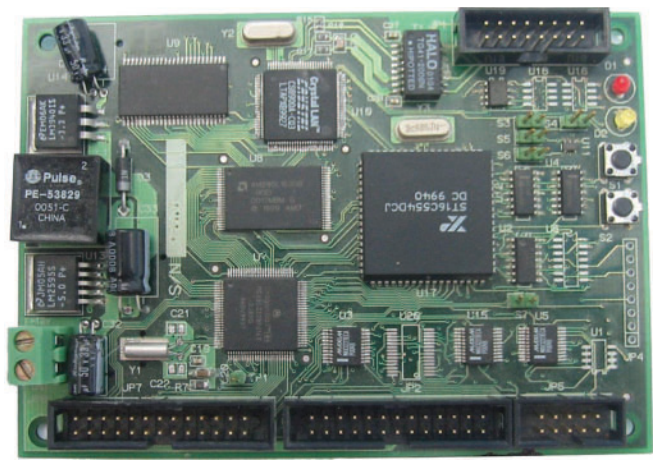


Figure 5. The Motorola development board used for µClinux is based on a DragonBall processor and includes an Ethernet interface.

Listing 1. Modifications to /etc/rc

```
1   hostname uCsimm
2   /bin/expand /etc/ramfs.img /dev/ram0
3   mount -t proc proc /proc
4   mount -t ext2 /dev/ram0 /var
5   mkdir /var/tmp
6   mkdir /var/log
7   mkdir /var/run
8   mkdir /var/lock
9   mkdir /var/empty
10
11  echo "1" > /proc/sys/net/ipv4/ip_forward
12
13  ifconfig lo 127.0.0.1
14  route add -net 127.0.0.0 netmask 255.0.0.0 lo
15  ifconfig eth0 192.168.0.1 promisc \
        netmask 255.255.255.0 broadcast 192.168.0.255
16
17  portmap &
18  cat /etc/motd
19  /bin/hprmanager &
```

Finally, we make three key steps of the repeater setup by modifying the initialization script /etc/rc. First, activate the packet forwarding module shown in line 11 of Listing 1. Second, assign the default IP address, as shown in line 15. Third, start the repeater manager, as shown in line 19.

We successfully tested these settings on a Motorola MC68EZ328 DragonBall microcontroller board (Figure 5) with 8MB of RAM, 2MB of Flash ROM, a 10Mbps Ethernet card and the µClinux v2.4.24 operating system.

Adding an Internet Connection

An extended HomePlug network may have an Internet connection through a modem router. Figure 6 represents this scenario.

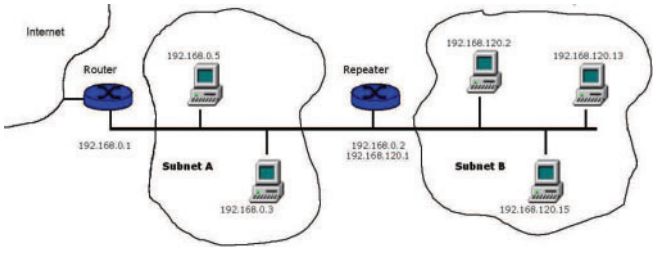


Figure 6. A Typical Scenario Featuring a Repeater and a Router with an Internet Connection

Let us consider the Linux desktop repeater to illustrate a solution to provide an Internet connection. If the router in the parent subnet has the address 192.168.0.1, it is necessary to assign a different address to the repeater. Moreover, the routing tables do change. However, the configuration of the computers in subnet B is the same. They simply route Internet-bound packets through the repeater by first issuing:

```
# route add default gw 192.168.120.1
```

The computers in subnet A route packets to subnet B through the repeater, and Internet-bound packets go right through the router. In them, we must execute the following commands:

```
# route add -net 192.168.120.0 netmask
 255.255.255.0 gw 192.168.0.2 dev eth0
# route add default gw 192.168.0.1
```

The repeater must route Internet-bound packets through the router by setting:

```
# route add default gw 192.168.0.1
```

Finally, the router sends packets to subnet B through the repeater. The configuration procedure depends on the router model. A typical and easy way is to log in to the Web-based configuration by going to the URL http://192.168.0.1 in any Web browser. Then, it is necessary to add route 192.168.120.0/24 through gateway 192.168.0.2.

Performance Evaluation

The most interesting result of our testing is, in addition to the repeater allowing communication beyond the HomePlug range, that it also enhances communications when two nodes barely can see each other. This is because the number of available HomePlug carriers increases.

For the sake of clarity, we assumed a configuration without an Internet connection in the parent subnet for our testing. First, we measured the response time and the throughput between two personal computers in a three-story building that could not see each other without the repeater in place. We tested both for UDP and TCP traffic. We used the Qcheck tool, a network-checking utility from Ixia. With a desktop-based repeater, we obtained response times for TCP and UDP traffic of approximately 100ms and throughput in the range of 2Mbps. This is realistic performance for medium-sized homes.

In a second test, we inserted the repeater between two computers that barely could see each other. The response time for both TCP and UDP doubled when inserting the repeater (50 to 100ms, approximately). However, the throughput grew from 1.5Mbps to 2Mbps.

We currently are testing the µClinux version on cards with a 100-BaseT Ethernet interface, such as the µCdimm ColdFire and the EV-S3C4530, both from Arcturus Networks.

Resources for this article: www.linuxjournal.com/article/8527.

Francisco J. González-Castaño is a professor with the GTI Group, Departamento de Ingeniería Telemática, Universidad de Vigo, Spain (www-gti.det.uvigo.es). He works in high-performance networking technologies and distributed computing, among other fields.

Pedro S. Rodríguez-Hernández is a professor with the GTI Group, Departamento de Ingeniería Telemática, Universidad de Vigo, Spain. He works with real-time and embedded systems.

Felipe J. Gil-Castiñeira is an assistant professor with the GTI Group, Departamento de Ingeniería Telemática, Universidad de Vigo, Spain. He works with wireless networking technologies and their applications.

Miguel Rodelgo-Lacruz is a researcher with the GTI Group, Departamento de Ingeniería Telemática, Universidad de Vigo, Spain. He works with high-performance networking technologies.

José Valero-Alonso recently received an Engineering degree from the GTI Group, Departamento de Ingeniería Telemática, Universidad de Vigo, Spain. He is interested in computer architecture and client-server systems.

Dialogue with Don

Departing Editor in Chief Don Marti talks with Doc about Linux as a better building material, durable free software principles, life beyond DRM, OpenLDAP, DIY, entrepreneurial IT and other ideas that grew during Don's tenure with the magazine.

BY DOC SEARLS

This issue is Don Marti's last one as Editor in Chief. I recruited Don to the magazine, and I hate to see him go. Don brought an ideal combination of know-how, commitment, integrity, insight, creativity and humor—all of which sustained him through a tough period for *Linux Journal*, the computer industry trade press and for the Linux community as well.

Don was a smart and tough editor. He suggested many of the topics at which I've become expert. He spiked (that's journal talk for rejected) more than a few of my pieces, always for good reasons. And he always pushed me to do better work. I wasn't always happy with that (few writers are), but I'll always be grateful.

The last time the editorial staff was together, at LinuxWorld Expo in August 2005, executive editor Jill Franklin gave me a fun assignment: interview Don. So, with the help of Steve Gillmor (impresario of the eponymous Gillmor Gang podcast, as well as a veteran producer of recordings, going back to his days with Firesign Theatre), we recorded what will surely also be a podcast, timed to come out along with this magazine.

DOC SEARLS: How long has it been?
DON MARTI: I've been at *Linux Journal* since 2000, and I've been Editor in Chief since 2002.

DOC SEARLS: When you came along, it was right when the bubble was bursting, and you came from VA Linux, which was the largest of the bubbles.
DON MARTI: Yes. I jumped off the dot-com bubble right as it was popping.

DOC SEARLS: [*laughing*] We're at LinuxWorld (Expo) now, and the whole show was on cocaine back then, in a way. I mean, it was very high; there was nothing but a weird kind of gassy optimism.
DON MARTI: Cocaine plus sushi and leather pants.

DOC SEARLS: So, I'm interested in your perspective on what's happened with Linux over the past four years. What did we understand well in the first place? What did we never quite understand?
DON MARTI: Well, Linux made a lot of big promises like every one of the technologies that touched the dot-com frenzy. Linux was better than most at delivering on them. And, in the years since the dot-com boom, I think people have had time to fill in the necessary gaps and move Linux into more and more niches. Things like logical volume management, for example. And real-time improvements in Linux, and cleaning up the desktop, and getting more hardware support—just checking off those to-do list items, one at a time.

DOC SEARLS: Last night we had this documentation BOF. One of the guys there said that we've reached the point when it's even possible to put Linux on a random laptop and there's a fair chance it's going to work out. A lot of the behind-the-scenes work has made that possible.
DON MARTI: One of the factors that helps account for that is the consolidation in the PC hardware market. Laptops used to have more weird bastard spawn hardware in them than they do today. With the introduction of USB hardware, you have a much smaller number of actual chips that your drivers have to talk to. Of course, through the same chips you're talking to everything in three aisles of the computer store, but the driver development for supporting all that can be saner and easier for more people to have a hand in.

When Greg Kroah-Hartman did an article for us on writing a driver for a multicolor LED blinky light device that plugs in to the USB port, he got a bunch of comments on that, includ-

Increase Your Application Performance up to 80%!

The revolutionary Ammasso 1100 is the world's first RDMA (Remote Direct Memory Access) enabled, Gigabit Ethernet Server Adapter. This *cost-effective* network adapter does not require specialized switches or cabling therefore, you can use your existing equipment.

Combine the Ammasso 1100 with the knowledge and expertise of *TeamHPC* for superior interconnect solutions to your server-to-server networking challenges.

- Increase CPU efficiency
- Increase network message capacity
- Increase speed of application data transfers
- Increase infrastructure flexibility

- Reduce application-level communication latency
- Reduce memory copies
- Reduce total cost of ownership by leveraging your existing investment in Ethernet

With the AMD Opteron™, faster throughput for network interconnects is achieved. The Ammasso 1100 RNIC is perfect for *TeamHPC's* AMD based Turnkey Cluster Systems which allow for simultaneous 32 and 64 bit computing capabilities.

SPEED. PERFORMANCE. MANAGEABILITY.

Get the advantage of <u>both</u> application performance and Ethernet simplicity, call *TeamHPC* today, 1-866-TeamHPC(832-6472).

Team HPC
HIGH PERFORMANCE COMPUTING
A division of M&A Technology, Inc.

ISO 9001
9001:2000 Certified

When it comes to price and performance, you can't beat the best team around . . . TeamHPC!

1040 OCL Parkway, Bldg. A • Eudora, KS 66025 • sales@teamhpc.com • www.teamhpc.com • GSA# GS-35F-4038G

ing one from a developer who, before the next article in the series came out, had written his own USB device driver and gotten it into the kernel tree.

DOC SEARLS: How much have people reading and writing in places like *Linux Journal*—especially *Linux Journal*—had an involvement with the development of Linux?

DON MARTI: Greg Kroah-Hartman again is a good example of that. He's now one of the top kernel people. Both through work and his own projects, he has become responsible for more and more of the kernel. He started off writing for *Linux Journal* in 2002. And, as he's gotten more responsibilities in the kernel, he's also written more articles for *Linux Journal*. Robert Love is another good example. And outside the kernel, many, many other contributors have both code that they maintain that's on the Linux CDs you get at the store, and also articles that they've written for *Linux Journal*.

DOC SEARLS: Yeah, it's always been interesting to me what role *Linux Journal* and journals in general have in a development ecosystem. What do you see as the future for *Linux Journal* and for magazines like that? At this point, it's a tough time for publications. We seem to have sustained a complete turnover of advertisers after the dot-com bubble—and managed to stay in business. But today so much more information is available freely on the Web. And we have a two- or three-month lead time. How can we stay current?

DON MARTI: On the Internet, every movement looks like a big argument, and one of the things a print publication can do is pick a side and stand by a considered opinion. So, when *Linux Journal* comes out against something like proprietary device drivers, or when *Linux Journal* comes out and says that the directory server is one of the most important pieces of software in your organization to commit to open source and open standards, then we can take a consistent position on something like that and put together a set of articles that helps people succeed if they agree with us either in whole or in part.

DOC SEARLS: You were involved in our *Embedded Linux Journal* effort. What's the story with that, and with embedded in general?

DON MARTI: *Embedded Linux Journal* was a controlled-circulation publication. And I think the idea of sending people a paper magazine for free, and that advertisers will pay to reach them, is sort of falling apart. I don't know how many of these controlled-circulation magazines you get, but it's something where the reader doesn't have a commitment in time or money to pay attention to this thing, and it ends up being one of the last things they get to. So, when *Linux Journal* has readers who are willing to pay for it and subscribe to it, I think that they're more likely to read it.

DOC SEARLS: I'm thinking also of the activity around Embedded Linux. Two years ago I had people telling me that the telephone OS market was going to come down to Java and Symbian. Now it's pretty clear Linux is going to be the big thing there, or one of the big things there.

DON MARTI: Java as an application environment is still thriving on the cell phones. When you get a Linux phone, one of the features of that is a Java virtual machine, with the ability to install and run Java applications. But Linux certainly has a huge advantage for full-featured cell phones in that it's the very first OS that most of the hardware vendors develop drivers for. So that shortens the development time for manufacturers who want to get that hardware into a phone.

DOC SEARLS: Most of the developers that we run into at a place like LinuxWorld, or the O'Reilly Open Source Convention, are doing applications for computers, not necessarily for phones. And phones, even if they have Linux in them, are still silos. They're still closed things to some degree. Whereas a server you can make into anything you want it to be.

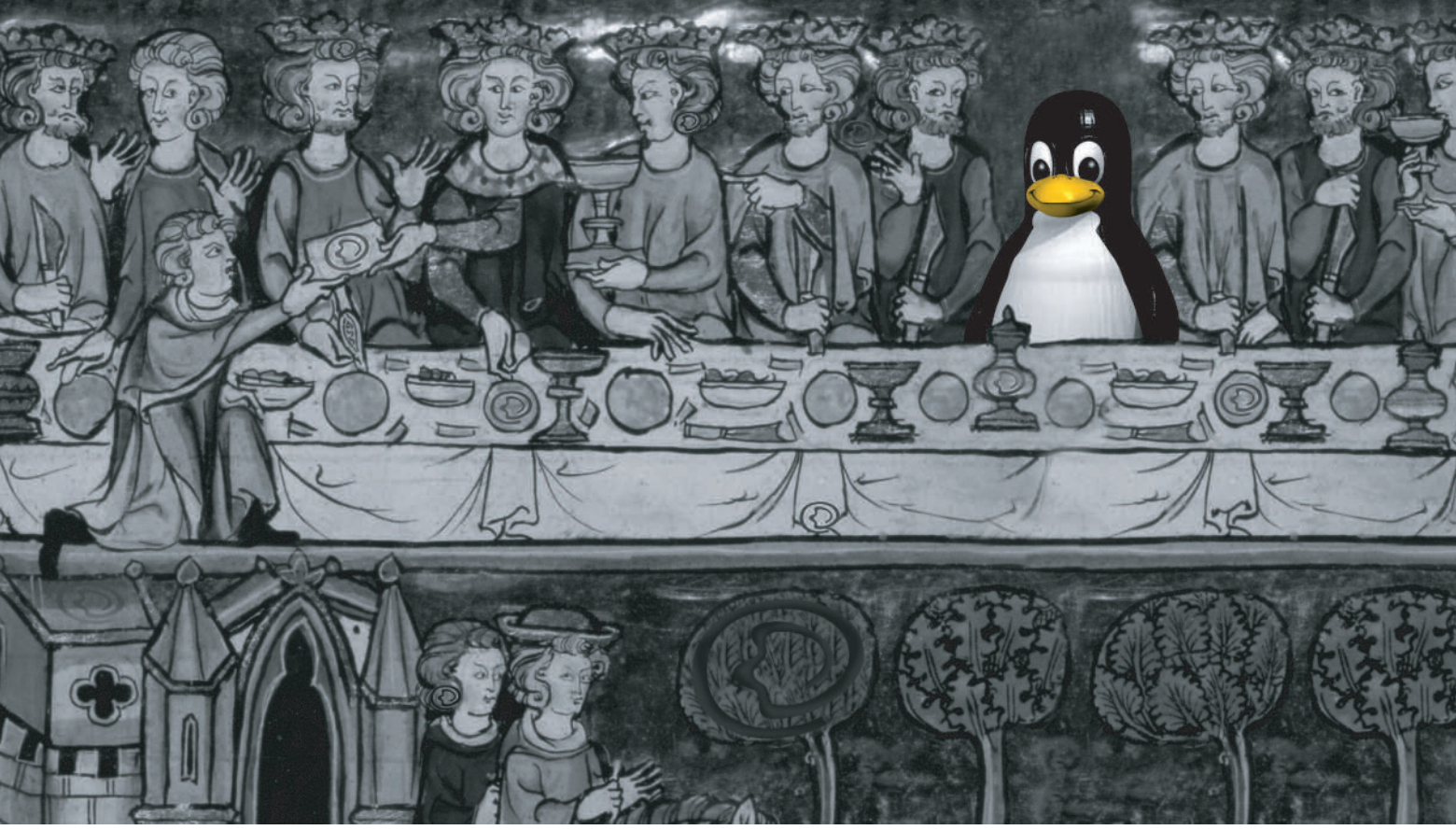
DON MARTI: When you get cell-phone service, they give you a phone. And free as in cell phones is not something that I think of as a bargain, because that phone is strictly controlled by the carrier, who determines what you can and can't run on it. Part of that is the carriers' need to conform to regulations. And part of that is their business model. They want you to buy applications through them, rather than being able to download and install your own.

DOC SEARLS: Yeah, they want to enforce behaviors. Like, if you accidentally took a picture where you have no choice to just discard it, you have to either send or save. That's what my phone wants me to do. They get money for that, I assume, or they wouldn't force me to do that. But there is a sense that there is, for me at least, a kind of a closed environment. Does it concern you that Linux is often used as the base operating system in things that are inherently closed, like a TiVo for example? I mean a TiVo is a sort of a closed environment, and TiVos run on Linux.

DON MARTI: A TiVo lawyer told me that the reason they have to be strict about video extraction is that they don't want to face a lawsuit from Hollywood. So, if you download and store a TV program in digital form on your TiVo, they do everything they can to make it difficult to get those exact bits off of that drive. You can record to a VHS tape, but you can't make a digital copy. And, like most of the other consumer electronics and IT vendors, I don't think TiVo is being 100% honest about big, bad Hollywood making them do this digital rights management. I think that there's a reason why IT vendors and consumer electronics vendors want to lock in their own customers and laying it all on Hollywood is not going to fly much longer. So, I'm concerned about devices that have lock-in built in to them, whatever OS they're on.

DOC SEARLS: You've said some interesting things about DRM in the past. For example, that all DRM is bad. You've gotten some push-back on that, but I'd like to hear what you mean by DRM being bad.

DON MARTI: Cory Doctorow made a great distinction between DRM and CA or conditional access. When you sign up for a service and they tell you, "You must log in to view this content", and you log in and then you can read and view, or cut and paste the information, that's conditional access. When you get a piece of content and it says, "Cut and paste are disabled", or "Print is disabled" or "Read aloud is disabled", then that's DRM. And DRM is deliberately micromanaging or removing the value from that information. It breaks some essential economic relationships that I think ultimately the authors of that information will be concerned about.



The Holy Grail

Easy, high-performance clustering. For years, many searched, but none could find it. Some said it didn't exist. But not the Penguin.

Penguin Computing® made easy, high-performance clustering a quest. Now you can find Linux hardware and software solutions, configured to order, driven by Scyld's commercially supported, industry-leading Linux clustering software.

For the turnkey clusters you need to run even your most important applications, come to Penguin Computing. Penguin Computing's dedicated experts, who are 100% focused on Linux, are waiting to serve.

Powerful, easy clustering. It's the once and future thing. Love what you do.

www.penguincomputing.com

Highly SCYLD

PENGUIN COMPUTING®

DOC SEARLS: If we had Hollywood executives sitting at this table today, saying they can't imagine any way other than DRM, what would you tell them about alternatives to DRM that would get them the same or similar economic benefit? That it's worth the trade-off?

DON MARTI: That's a really good question. I think that a lot of the understanding that Hollywood has built up over many years of trying to understand the Internet is based on sales pitches from vendors who are pushing DRM systems. So, when a DRM vendor goes to Hollywood and talks to them about, "We can control this, we can lock out this, this will enable you to make money", that really shapes the understanding of somebody who isn't in the technology business and who doesn't have the technical background. So, before I start spewing business ideas, I really want to listen to what the person understood to be the case about the technology and try to understand and fill in the gaps where the gaps are.

DOC SEARLS: This brings us to the cartelization of things. DVDs are encrypted, in their own way, because the cartel didn't want DVDs to run on any machine other than what they controlled or where they had a relationship. DVDs will run on Windows, on a Mac, but not on a Linux machine.

DON MARTI: And there were other business-model-related restrictions that were built into the DVD format. For example, region coding.

DOC SEARLS: I never understood why region encoding was there. I mean, it's a hassle that doesn't seem to have an upside to me.

DON MARTI: Well, imagine if a studio wants to release a movie on DVD in the US, when that move has not yet had its theatrical release in Europe. So, if they did not have the region coding system, then somebody might buy the DVD in the US and take it over to Europe and watch it and interfere with what has always been a classic Hollywood business model: show it in the theaters first, then wait a while, make it unavailable at all, and then release it on VHS and now DVD. And, interestingly enough, that model is being collapsed. Before the DVD format was decrypted, it was about a year from US theatrical release to DVD release, and within the past year or two, it's come down to about half a year. Hollywood wants to be able to play with business models, change who can see what when. So I think there is tremendous appeal that the DRM vendors are offering, saying, "We can control your audience, we can control the technology so that it fits with the business model that you want to try this year."

DOC SEARLS: I became familiar a few months ago with Lucene. Doug Cutting who used to work at Excite, felt that keyword search was a done science, essentially. The result is some open code that anybody could use. Now anybody can do keyword search. Lucene isn't even a full product. It's one piece of building material. Last night we talked about Struts, which is another one of those kind of things. It's been sitting out there. So, one concern that I have is that Linux, as it becomes more like a foundation stone, disappears. It turns into the building, it becomes rebar and cinder block. Does that concern you? Or is that just a natural course of things? Should we pay less attention to Linux after a certain point and to the general construction business that Linux is a part of?

DON MARTI: I think there are some lessons to be drawn from the history of the projects that are older than Linux and possibly more mature, as products, than Linux. And a good example would be GCC.

GCC for a long time was considered to be a good, stable compiler, capable of doing code for almost any processor out there. And, within the past few years, with a lot of the changes in the processor architectures and optimizations you can do for processors such as the Opteron, the need for ripping up and redoing parts of GCC has popped up. And, with things like the C++ standard template library, there's pressure on GCC on the language side as well. So, GCC is a piece of software that sits between the languages and the hardware. GCC was a stable, mature project, but as languages become more complex, and the number of languages people want to code in increases, and at the same time the hardware gets capable of doing hairier and faster things, then a mature piece of structure needs to have changes happen to it.

The same thing is going to happen with Linux, as hardware advances and the OS needs to be able to support more processors or processors in unusual configurations, such as the very many processors in a newer machine, or situations when you might have some processors on one die and some processors on another die, and the OS needs to be aware of which processors are where. As the hardware changes, the OS will need to advance, and as the applications that demand services from the OS change, the OS will need to advance. So, Linux won't entirely fade into the background unless hardware stops changing and the applications stop changing the way in which they use the kernel.

DOC SEARLS: Since we're on GCC, I know you're one of the folks who has a deep appreciation of Richard Stallman's role. I'm wondering....We've kind of gone back and forth on calling Linux "GNU/Linux" as Richard would like us to, and just Linux. Do you have a particular feeling about that?

DON MARTI: The official *Linux Journal* policy on it is, "Leave it the way the author wrote it." If someone wants to make clear in his or her article that the whole system is called GNU/Linux, then we leave that stand. If the author wants to say, "The name of my system is, say, Red Hat Linux", that doesn't have GNU in its name and so we leave that name as it stands in the original article.

Where GNU comes in as an absolutely key project is as a many-year development effort to bring together a system that lets people do what they need to do, to communicate, to get by in the world of computers. As Richard Stallman himself put it, "So that I can continue to use computers without dishonor."

And, the idea that when you click OK on that end-user license agreement, you say, "It is OK that I won't examine this piece of information that I have downloaded. It's OK that I agree not to change it or understand how it works, or explain it to someone else how it works." I've come to understand that I don't believe that. And, I've come to a lot of that understanding through what Richard has written about the subject.

DOC SEARLS: To me what's so interesting about Linux and about the Free Software movement—and to the understanding of computing and software that goes back to the earliest days of independent computing—is what Richard was saying about the nature of software in the first place: that it was inherently free and wanted to be free more or less the way the wood and the pine

Polywell High Performance Systems

Reliable Computing with Good Support

1U 4-way, 64GB DDR, 2TB RAID

- Dual AMD® Opteron™ Dual-Core Processors
- Upto 64GB 400MHz ECC DDR (16 Sockets)
- Upto 2TB 4 x 500G Swap Drive per 1U Rack
- 1 x PCI-X 133MHz, 1 x PCI-E Solt
- On-board Dual Gigabit Ethernet
- On-board ATI Graphics , 4x SATA-RAID
- Slim CD-ROM Drive, Optional DVD-RW or CDRW
- Optional Slim Floppy Drive
- 1U 24" Depth Rack Chassis with upto 600W PS
- 4 x Swappable Drive Bays (SATA or SCSI)
- Supports Linux, FreeBSD or Windows
- Custom Configuration Available
- Please call for other Options

1112ES-2200A, 242+, 2GB, 500G **$1,999**
1114AIS-2050M, 265+,16GB, 1TB **$6,499**
1114AIS-2050M, 275+,32GB, 2TB **$8,888**

Linux Appliance PCs

- Custom Made Odd Size Chassis
- AC or DC Power Supply
- Low Power Voltage AMD Processor
- or High Performance AMD® Athlon™ Processor
- Diskless or Flash OS Boot Drive
- Swapable Hard Drive, CD-ROM, FDD
- Integrated Graphics, Ethernet, USB
- Optional LCD LED Control Module
- IS2 Audio, MPEG2/4 Hardware Video
- Upto 4 Ethernet Ports or 4 Serial Ports
- We have over 18 years OEM Experience in Set top Box, Digital Media Player, POS Kiros, Thin Client, Networking Appliance,

SX2500SPLJ11B
OEM Appliance starts at **$299**

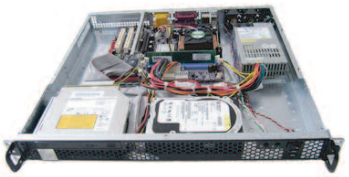
64GB RAM 4-way Workstation

- Quad AMD® Opteron™ Dual-Core 865+ Processors
- Upto 64GB 400MHz ECC DDR (16 Sockets)
- 2 x 133MHz PCI-X, 1 x PCI-E x16 Slots
- Dual Gigabit Ethernet, 4 x SATA-RAID Controller
- Quiet and Cool 12-Bay Tower + 600W 80Plus P/S
- 250G HD, DVD-RW, Floppy, Optional Card Reader
- 8-Layer Motherboard with Special Quiet Cooling
- On-board ATI Graphics, Sound Blaster 7.1 Sound
- Supports 64/32-bit Linux, FreeBSD or Windows
- Special for Large Memory Intensive Applications
- Built-to-Order or Configure-to-Order

16G RAM, 2x244+, QuadroFX1300 **$3,999**
32G RAM, 2x250+, QuadroFX3400 **$9,999**
64G RAM, 2x275+, QuadroFX4500 **$22,999**

1U Power Saving ISP Server

- AMD® Sempron™ or Opteron™ Processor
- 512M DDR 400MHz Memory
- 80GB Hard Drive
- 10/100Mbit Ethernet
- We provide Drive Image Service
- 1U 14" Short Rack, allow 2 x 1U per Rack
- Low Power Usage to save your Data Center Cost
- Perfect Entry Level ISP Server or Appliance System
- IDE Flash Drive is available
- Supports Linux, FreeBSD or Windows
- Custom Configuration is Available
- Please call us to discuss your specification



Order# VX2500SP1U-17LJ11A
starts at **$399**

2U 8-way, 5U 16-way Opteron

- 8 or 4 AMD® Opteron™ Dual-Core 865+ Processors with Hyper Transport Technology
- Upto 128GB DDR Memory for 16-way (32 sockets)
- Upto 64GB DDR Memory for 8-way (16 sockets)
- 4x Gigabit LAN, 8x SATA RAID-5 for 16-way
- 2x Gigabit LAN, 4x SATA, U320 SCSI for 8-way
- 4 x 133/100/66MHz PCI-X Slots for 16-way
- 2 x 133, 2 x 66MHz PCI-X, 1x PCI Slots for 8-way
- On-board ATI Graphics, USB 2.0
- 5U 26" Rack 1300W 3+1 Redundant P/S 16-way
- 2U 27" Rack 700W PFC P/S for 8-way
- Supports Linux, FreeBSD or Windows
- Custom Configuration Available
- Please call for other Options

2U 8-way 865+, 4GB, 750GB, 8422B **$10,999**
5U 16-way 865+, 32GB, 2TB, 8800U5 **$29,999**

SANNAS 4U 12TB Storage

- SAN Ready NAS Storage Server with upto 24 x 500GB Hard Drive
- Dual Opteron™ Processors NAS Appliance
- upto 4 x 2G Fibre Channel Ports for SAN
- upto 8 Gigabit Ethernet Ports for NAS
- Supports UNIX, Linux, FreeBSD, Windows
- 24 x Hot Swap Drive Bays for SATA or SCSI Drives
- 950W 3+1 Redundant Hot Swap Power Supply
- Custom Configuration is available
- Remote Support Available

6TB NAS starts at **$7,999**
12TB SANAS at **$14,999**

18 Years of Customer Satisfaction
5-Year Warranty, Industry's Longest
First Class Customer Service

AMD 64 Opteron™

888.765.9686

www.Polywell.com/us/LJ

POLY SYSTEMS

AMD64 architecture reduces I/O bottlenecks, increases bandwidth, and reduces memory latency. Critical information gets to those who need it quickly and efficiently.

Polywell Computers, Inc 1461 San Mateo Ave. South San Francisco, CA 94080 650.583.7222 Fax: 650.583.1974

tree wants to be free. He wasn't just talking about the economic uses of it; he was talking about the nature of the thing itself. And the feeling I have is that this is still not fully understood. Is that your sense as well?

DON MARTI: Well, my sense of software is that it's something that is both speech and a device, depending on how you define it. When you talk about software as speech, many good things tend to flow from that. When you use software as a device you can get into great benefits and also fairly scary issues. So, the challenge is to apply the best of what our culture has developed for the real world to the world of software.

On both sides of the software freedom debate, people try to make analogies comparing software to real-world items. So when Bob Young says, "You wouldn't buy a car with the hood welded shut", he's trying to make an analogy to a real-world object. When someone on the restrictive side of the debate says, "Well, you wouldn't walk into a store and walk out with a copy of the CD", this person is also trying to make an analogy to a real-world item. It's a huge issue to understand the best of what we value about real-world goods and translate those values to the software world and the on-line world.

DOC SEARLS: As you know, I've been fascinated by the parallels between the construction industry and computing in general, including the software industry. In construction there is a very mature understanding of how things work together. Now, we've been sitting in this building. I'm sure this floor is a synthetic material and there is clearly some kind of sedimentary rock that's a surface over there, and behind you there is the huge corpse of a trunk of what appears to be a eucalyptus tree. It's not structural; it just graces the place as an architectural element. There's steel and terrazzo over here. So one of the things that fascinates me about construction is that it's full of open source. I mean, there are no secrets to making terrazzo. Yet there's still what we call intellectual property in construction. But none of it is in position to take control over everything else. I'm looking at a door over there. It's probably a standard door, but the latch on it may have some patents in it, and it may have a lock in it and that lock may have some patents as well. But you can replace that lock, right? And, I'm wondering if you can see a path toward that. I don't think we're at that point in software yet, where we have that same sense of modularity. Do you see us getting toward something like that in software? What might Linux have to say about that, being something like a natural material?

DON MARTI: So far, the proprietary software vendors have really dropped the ball. On the free software side, Richard Stallman with the GPL has come out with a normative statement of a code of conduct for software developers and users. When someone releases software under the GPL, or chooses software under the GPL, the person is agreeing to those norms. If you want to talk about proprietary software becoming part of a mature market, or becoming a part of the useful structure, then there has to be some norm other than "all your base are belong to us".

When you look at Larry Ellison's licenses saying, "Thou shalt not publish benchmarks and you have to click on this to agree to that", that's not compatible with building a useful structure out of multiple materials or under multiple licenses. That's a trailer-park landlord's

idea of city planning. So, really, when the proprietary software license writers decide to put as much thought in their licenses as Richard Stallman and Eben Moglen and the rest of the free software side have put into theirs, then we have some potential for that kind of innovation and growth. Until that happens, I think those who want to treat software as a mature product and a responsible market are not going to have much choice except for the free software side. So show me a responsible, innovation-compatible and integration-compatible proprietary software license and we'll see what happens.

DOC SEARLS: In looking back over your five years or so with *Linux Journal*, what great articles or achievements stand out for you?

DON MARTI: I'm very proud that we did our 2.6 kernel preview very early in the 2.6 cycle, when it was still 2.5 development. That was when we let people who were doing Linux deployments and applications know, "Look, here's the great stuff coming along in the kernel." That issue [May 2003] with Robert Love wearing headphones and the headline, "Are You Ready to Rock?", that was the right issue at the right time to give 2.6 testing a nice kick. And, one article that I was so happy about that I had the authors do another version of essentially the same idea, was Craig Swanson and Matt Lung's "OpenLDAP Everywhere" [December 2002 and July 2005]. That company brought together the complete directory of services for all their clients, both Microsoft Windows and Linux, authenticating against it, sharing address books, using the file server and the intranet servers in a very compatible and customer-directed way. So, we, Doc, you and I talked about this and came up with the idea of DIY-IT—largely influenced by a small company.

DOC SEARLS: I get a lot of credit for that, but that really came from you. There's the notion of smart companies using Linux to make themselves smarter. That was an assignment that really became my mission with the magazine. The observation that everything that happens with Linux starts with smart individuals doing smart stuff, usually without big vendor assistance. I'm not knocking big vendors at all, it's just that DIY-IT acknowledges that they're part of the ecology, not the origin of the ecology.

DON MARTI: And when the vendor says, "there is no market for that yet", that's something the customers should hear as "your competitors aren't doing that yet". I think the next step, beyond DIY, is entrepreneurial IT. Where can you take those building blocks that are becoming large enough, stable enough, functional enough that you can get a lot of business value with very little integration work and staff time? How can you take those things and as an IT department create business value?

DOC SEARLS: I need to wrap this up by saying that I've been around *Linux Journal* from the beginning—and this is not a knock at any editors—but as far as I am concerned, you're the best editor we've ever had and it's been an honor to work with you.

DON MARTI: Thank you.

Doc Searls is Senior Editor of *Linux Journal*.

Controlling a Pinball Machine Using Linux

Create a master hack by bringing the power of Linux to the ultimate electronic toy.

BY JOHN R. BORK

An old electronic pinball machine is fascinating because it embodies complexity just within the grasp of a jack-of-all-trades hacker. You can learn how one works by visiting the open-source repository known as the US Patent and Trademark Office. The Bally Manufacturing Corporation used a system built around its AS2518 Microprocessor Unit (MPU) described by US Patent 4,198,051 in more than 350,000 units from 1977 to 1985. Maybe you remember playing *Evel Knievel*, *KISS*, *Mata Hari* or *Space Invaders*?

At the moment, you can buy most nonworking games for less than $250. Many come with original documentation that includes circuit schematics. Combined with what you can learn from the patents and other publications, plus your knowledge of PC hardware and free, open-source software, you can hack together something unique: a working, Web-enabled, classic pinball machine that plays by your rules, running your programs. You can do it legally, for less than the cost of a replacement MPU board, with an old PC and a stock Linux distribution like Fedora.

Reverse engineering the AS2518 MPU was the subject of my Master's thesis in Industrial Technology. Nonworking games often suffer the same tragic design flaw we see on old computer motherboards. Figure 1 shows the damage caused by a leaking Ni-Cad battery that was soldered directly onto the MPU. It ruins not only the electrical connections in IC sockets, but also corrodes the wiring harnesses joining the MPU to the rest of the system.

DON'T BE SQUARE! GET CUBED!

L Series Laptop - LS1250-L
Light &Thin Performance
Starting at $1,562.54

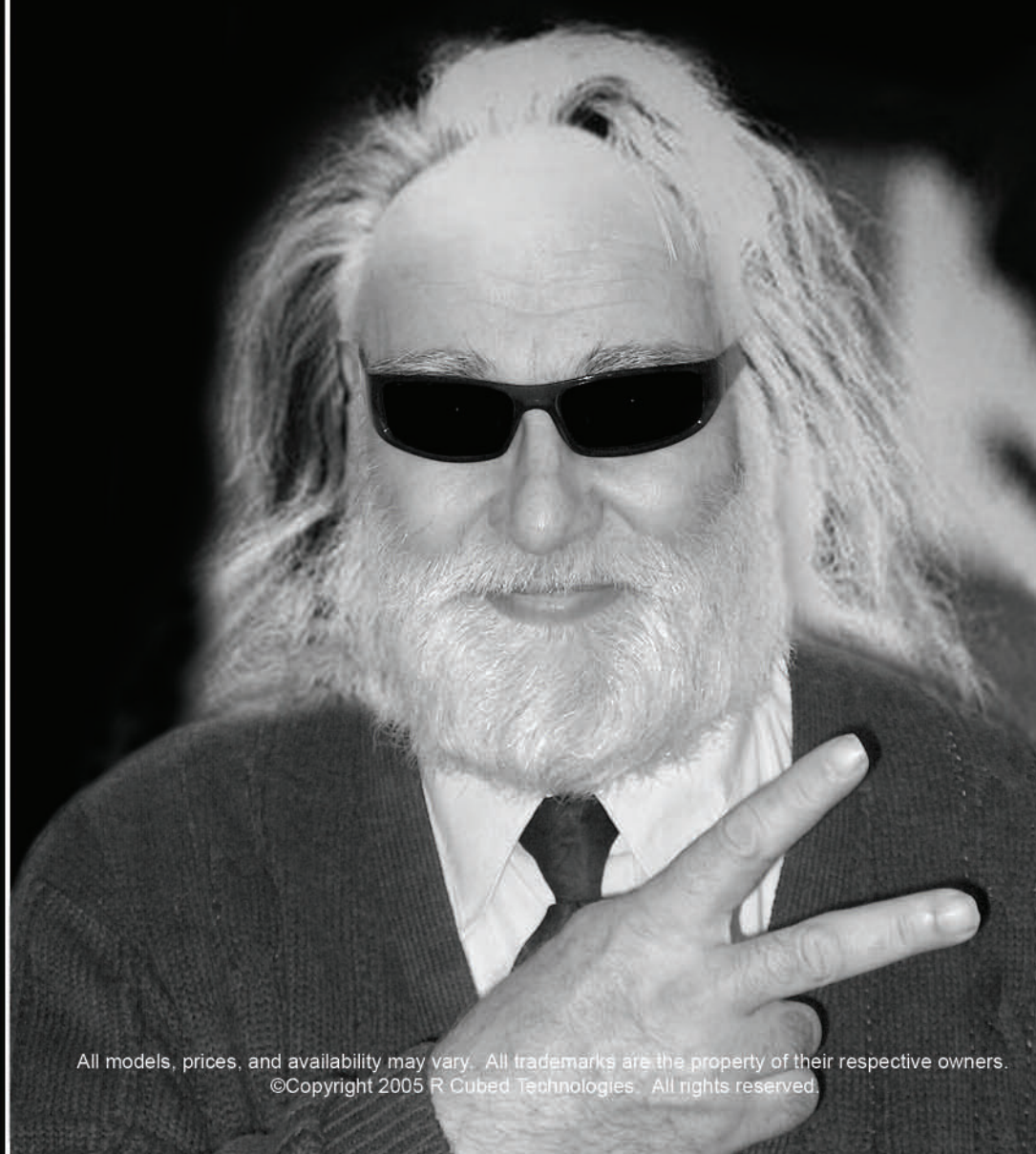
G Series Laptop - GW1550-L
Essential Technology on a Budget
Starting at $1,302.54

X Series Laptop - XW1550-L
Extreme Technology & Performance
Starting at $1,608.04

R Cubed Technologies has provided pre-installed Linux laptops without OS tax since 2003. We customize the Fedora Linux distribution for each laptops configuration providing support for: PCMCIA, USB, FireWire, X, CD/DVD/CDRW/DVDRW, Sound, Power Management, Ethernet, Modem, Wireless, and more. Our laptops are equipped with Intel Centrino Mobile Technology. We also offer Windows dual boot options. All of our laptops come with a one year parts and labor warranty. Visit us online at www.shoprcubed.com or call 309.34.CUBED for details.

R3 Technologies™
Working Hard to Bring Technologies to Life.

309.34.CUBED
www.shoprcubed.com



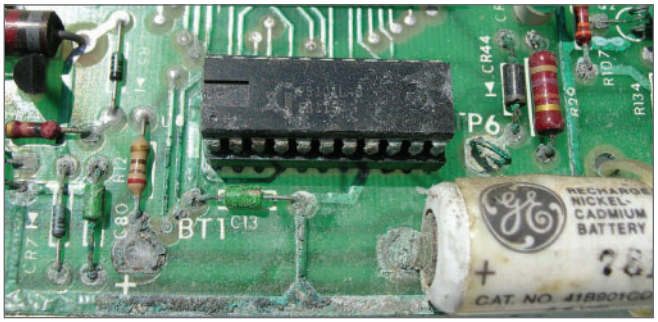


Figure 1. Corrosion on an AS2518 MPU Board

The other circuit boards are usually still intact. When you start working on your game, check the voltages at the test points to make sure. I chose to neuter the flaky +5 VDC circuit altogether and use the power supply from the PC. With the MPU removed, you are left with four wire harnesses holding a total of 66 wires. To connect your PC to the pinball machine, you will want to build an interface board with matching header pins. The design goal is to produce the same inputs and outputs on all of the wires that the original MPU has. This may seem like an overwhelming task, but remember, this is 1980s-era technology. I used an iterative, divide, design, build and test approach to reverse engineer one subsystem at a time.

What differentiates this project from the typical emulator is that no reference is made to the original programs encoded on the MPU firmware. Instead, I employed a black box, or clean room, method based on studying their function rather than their internal structure. For me, it made sense to interpret these 66 electrical connections in terms of their purpose in a closed-loop process control model. That is, each is either input, output, part of a feedback circuit or part of the power supply. The four main divisions of the pinball machine control system are the solenoids, switch matrix, feature lamps and digital displays. I intentionally left out the digital displays for the first prototype, which is why the apparatus uses the computer monitor to show the scores. The analysis yielded the process model shown in Figure 2.

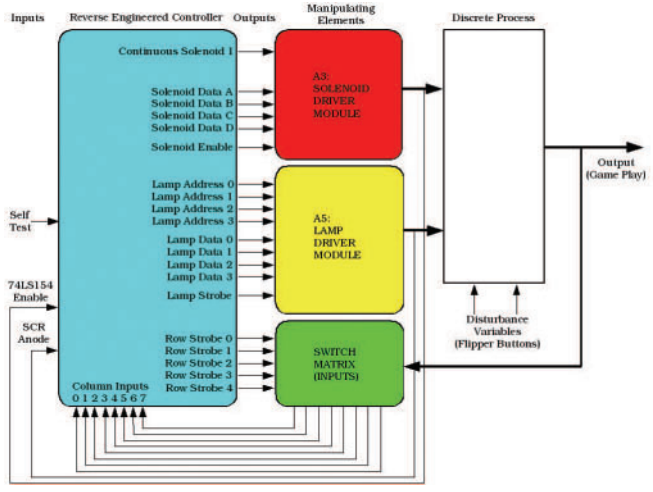


Figure 2. Reverse-Engineered Process Model

The Hardware, Part I: the I/O Board

Facing a total of 11 inputs and 20 outputs, and wanting room to grow, I decided to build a 48-port digital I/O board. Designs can be found with a little Web searching, and the components can be ordered from Jameco. The Intel 8255 Parallel Peripheral Interface (PPI) integrated circuit provides two 8-bit ports and two 4-bit ports, each configurable as either input or output. On my board, I hard-wired two of these ICs to addresses 0x280-0x283 and 0x2A0-0x2A3. The first three bytes of each are memory-mapped to the aforementioned ports. The fourth byte is used to control the port settings. I used a ten-foot piece of 25-pair twisted pair cable to connect it to the interface board via screw terminals. It's definitely a hack, as Figure 3 illustrates. You may want to use a 50-conductor SCSI cable and header pins.

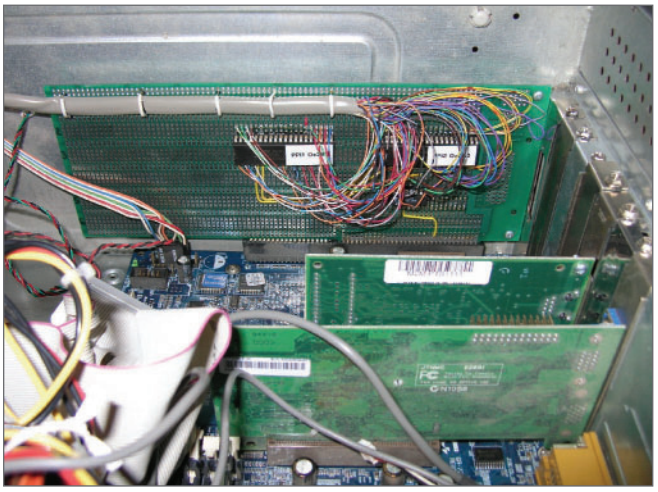


Figure 3. Homemade 48-Port ISA I/O Board

The Hardware, Part II: the Interface Board

The AS2518 MPU is based on the Motorola 6800 microprocessor. It uses two 6820 Peripheral Interface Adapters (PIAs) to provide I/O to the rest of the system. The Intel 8255s are functionally similar. What must be duplicated on the interface board are the circuit elements between the PIA I/O lines and the header pins. These are determined through direct inspection and study of the electrical schematics accompanying the patents and the operator manuals, and consist mainly of resistors and capacitors. A picture of the board I created is shown in Figure 4. A label maker works great for marking wires and connectors.

The Software, Part I: Basic Operation

First, I tried to make the control system work as an ordinary user-space program. Using the method of divide and conquer, the simplest subsystem of the pinball machine to hack is the continuous solenoids. They are either on or off for long periods of time. On my game, I implemented only the flipper relay, which is turned on during normal game play and off when the game is over or tilted so that the flipper buttons don't do anything. This operation was easily accomplished by a variation of a C program I wrote to test the I/O board. According to the schematic, the flipper relay is enabled by making its output low

Is your data center cramping your style?

Growth always seems like a good idea. An extra processor here—one more server there. Until, all the sudden your data center feels as crowded as a center seat in coach. Let **the Penguin** upgrade you. Penguin Computing introduces BladeRunner™ 4140 the industry's densest Linux blade server. It comes with the AMD Opteron™ HE processor, which offers simultaneous 32- and 64- bit computing. So now you can pack 48 cores into a miniscule 4U of rack space, and optimize your data center. And put that 8GB of PC3200 RAM per blade to work and run your 64-bit apps in a fraction of the space. So go ahead. Stretch your legs. Tilt your seat back. **Love what you do.**

Visit www.penguincomputing.com

Join us at the Supercomputing Conference 2005 in Seattle, Washington State and Convention Trade Center, Booth #6222

AMD Opteron is a trademark of Advanced Micro Devices, Inc
Other names are for information purposes only and may be trademarks of their respective owners.

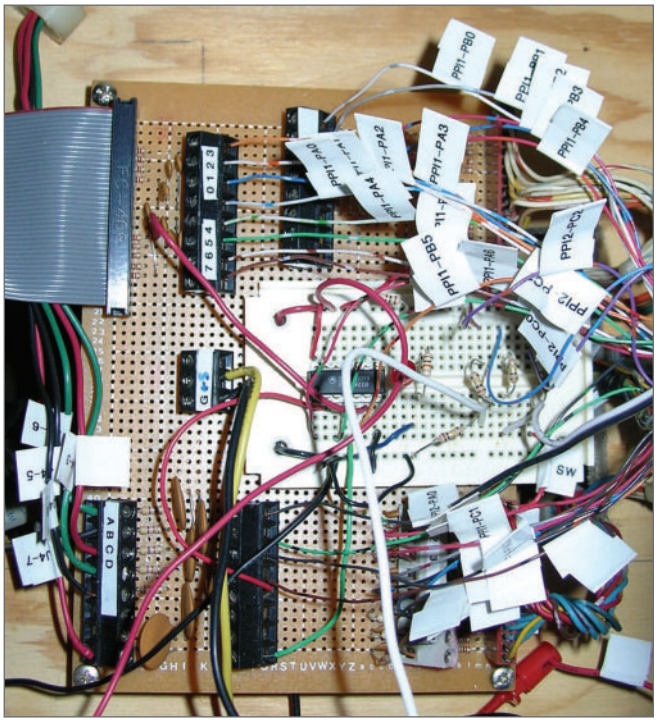


Figure 4. Interface Board

rather than high. This is known as negative logic. I quickly learned something about the PC architecture: even with a pull-up resistor, the port is in a low state from the moment the computer is powered up. This had the unintended result of turning on the flippers before the control program was even started. To work around it, I added a 7404 inverter to the interface board. Now the flippers are enabled when the output is set high.

Next, in order of complexity, comes control of the momentary solenoids. These are things like the pop bumpers, chimes, slingshots, saucers and the outhole kicker that are fired for brief bursts throughout the game. The Bally documentation states most are energized for a period of 26 milliseconds; some, like the drop target reset, for twice as long. To fire one of 16 possible solenoids, five output lines are used to drive a 74LS154 decoder on the solenoid driver board. Four lines provide the binary representation of the desired solenoid, and one line enables or disables the decoder outputs. Each output in turn drives one of the 16 momentary solenoids.

Like the continuous solenoids, the 74LS154 enable uses negative logic. Programming this action seems simple. Start with the enable high. Output the four-bit solenoid number, set the enable low for the desired duration, then set it high again. Actually, this creates a problem that challenges the ability of an ordinary Linux user process to behave in real time. You cannot depend on usleep(26000) to produce a 26-millisecond delay precisely; it may and often does yield a longer delay, as the man page warns. Leaving a solenoid enabled for much longer than 100 milliseconds can damage it and blow the fuse. One option discussed in the Port Programming HOWTO is using multiple outb() calls, because each one takes approximately a microsecond to execute. However, this amounts to a colossal waste of CPU time spent in a busy loop.

The prospects for a user-space control process diminished even more as I began to implement the switch matrix. The Bally documentation explains that once every 8.3 milliseconds a snapshot of the switch matrix is created and then analyzed for changes, such as when the pinball strikes one of the many switches on the play field. It is a matrix because 40 separate switches are wired into five rows of eight columns apiece. The rows are outputs and the columns are inputs. A logical high is output to the first row, also referred to as strobing the row. After a brief delay to allow the voltage to be detected at the other end of the circuit, an input operation reads the eight, single-bit columns as one byte of data. Then the process repeats for the next row, and so on.

Here is where the real-time requirements become critical for correct game operation. If an adequate delay is not created between the row strobe and the column input, you get garbage; the game's closed-loop feedback system fails. If too much time elapses between each sample, such as while the process is swapped out by the scheduler, a switch closure might be missed. The challenge of ensuring that the control process executes at a high frequency (120 Hertz) led me away from user space to the kernel.

The Software, Part II: the Kernel Module

The module I wrote is based on the examples given in the excellent tutorial *The Linux Kernel Module Programming Guide*. Every kernel module requires an initialization function that is called when the module is installed via insmod. This is where I write out the control words to the two 8255 PPIs defining which ports are for input and which are for output. Here is also a good place to register a character device file, which is a simple means to communicate between kernel space and user space. I created one called /dev/pmrek.

To turn this module into a periodic process, I declared a workqueue for it. Workqueues are a new feature of the 2.6 kernel. The function in my device driver I want to call with the workqueue is pmrek_process_io(). The workqueue is defined at the global level of the module code with the statements:

```
static struct workqueue_struct * pmrek_workqueue;
static struct work_struct pmrek_task;
static
DECLARE_WORK(pmrek_task, pmrek_process_io, NULL);
```

Then, in the module initialization function pmrek_init(), create the workqueue with:

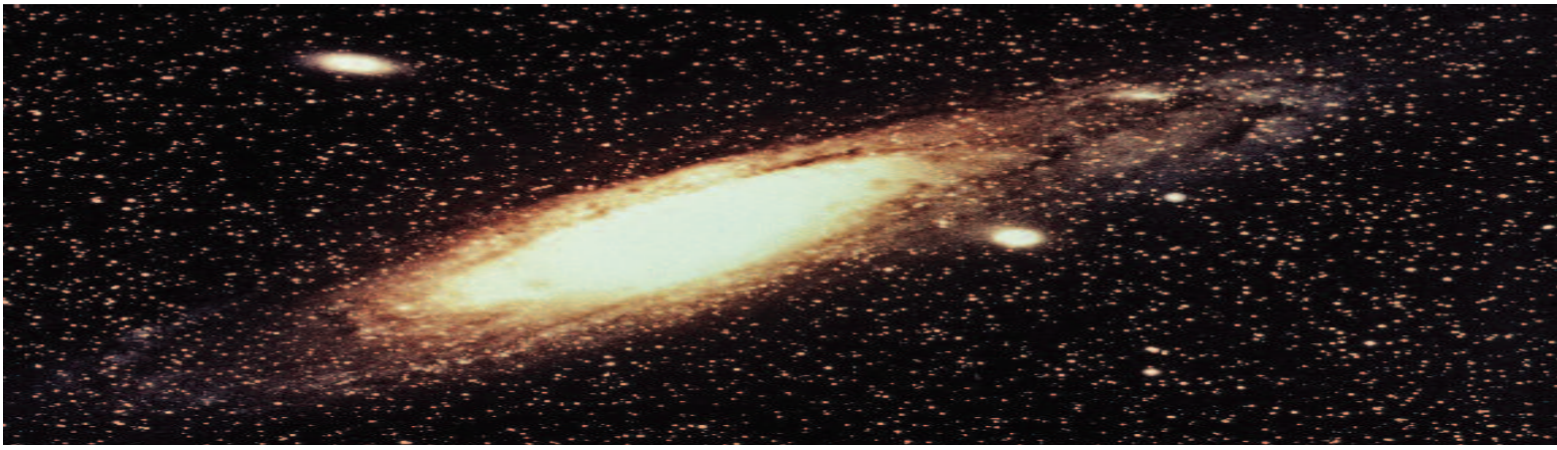
```
pmrek_workqueue = create_workqueue(pmrek_WORKQUEUE);
```

This does not actually schedule the workqueue yet. That happens when the supervisory program activates it. Figure 5 is a flowchart of the low-level hardware I/O operations performed by pmrek_process_io().

The first thing it does is read in the switch columns using inb(). If there are any valid switch detections, they are written to a log buffer. This log buffer is consumed by the supervisory process, and game play advances depending on the switches detected. Switch detections are stamped with the exact time they occurred by getting the CPU Real Time Stamp Counter (RTSC) via the inline assembly command:

```
__asm__ volatile (".byte 0x0f, 0x31" : "=A" (cpu_time));
```

MORE SPACE. LESS MONEY.



UNLIMITED AFFORDABLE NETWORK STORAGE

Everybody needs more space. And they need to spend less money. What if you can both have more space and spend less money?

What if you could put 7½ terabytes in only 3 rack units? What if that 7½ terabytes cost less than $10,000? Including the SATA disk drives. Imagine if you could glue it all together with a RAID appliance into one system. What if you could add as much storage as you wanted, one shelf at a time, and never have to `fork-lift' anything?

Coraid's new SATA EtherDrive Storage allows you to do just that. Using industry standard SATA disk drives, EtherDrive Storage connects disks directly to your Ethernet network. Each disk appears as a local drive to any Linux, FreeBSD or Solaris system using our open ATA-over-Ethernet (AoE) protocol. Since the disks just appear as local drives you already know how to use them.

The EtherDrive® SATA Storage Shelf is a 3U rack-mount network appliance that contains 15 SATA drive slots. Its triple redundant power supply protects you from your most likely failure. Its dual Gb Ethernet interfaces allow your data to go fast; 200MB per second. And at a very affordable price. List price for the EtherDrive Storage Shelf, without disks, is only $3,995.

Our companion product, the RAIDBlade RAID controller, allows a virtually unlimited number of Storage Shelves to be combined into a set of logical AoE storage devices.

Now you can have unlimited storage at a very affordable price. For complete information, visit our website at www.coraid.com, or call, toll-free, 1-877-548-7200. And we'll show how we've made network storage so affordable, you can have all the space you want.

www.coraid.com
info@coraid.com
1.706.548.7200

CORAID

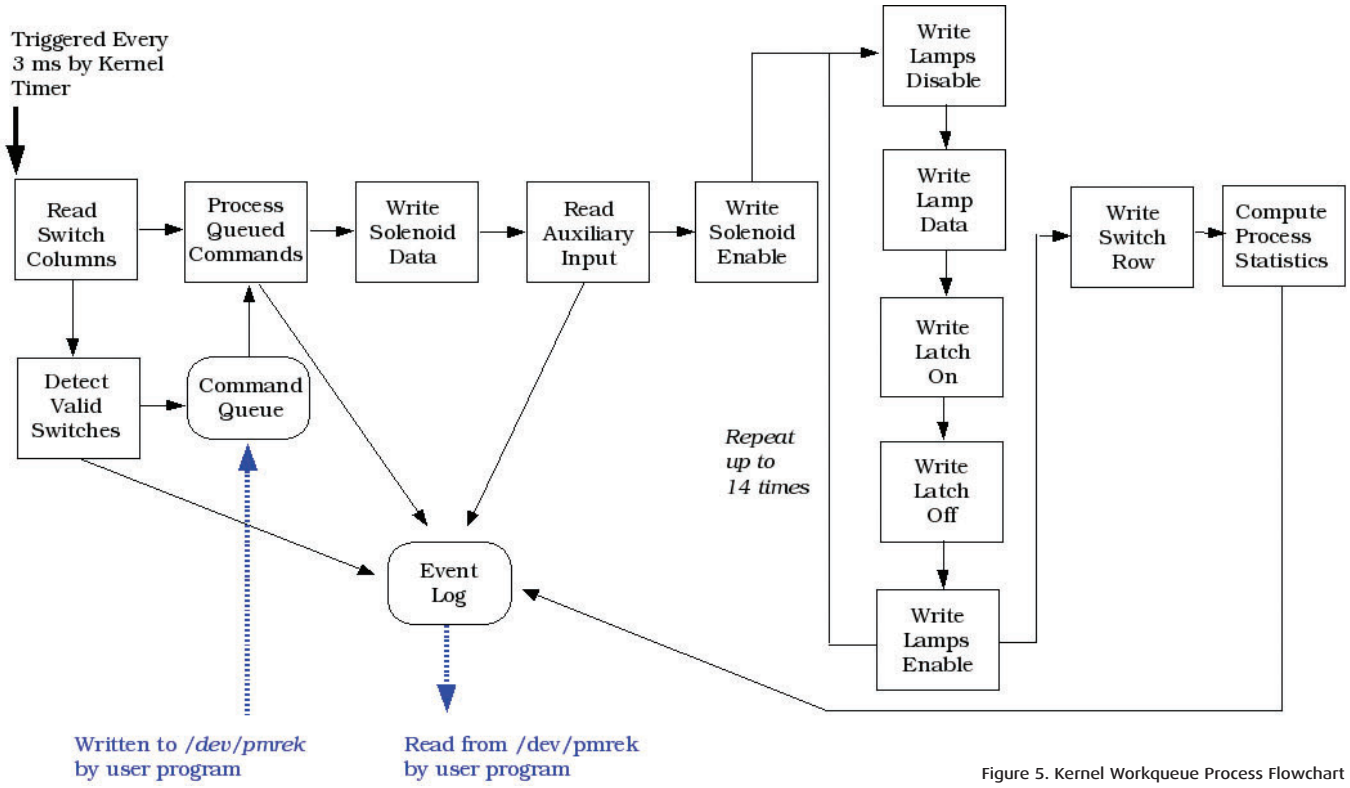


Figure 5. Kernel Workqueue Process Flowchart

Table 1. Source Code for the Pinball Machine Reverse-Engineering Kit

Source Code File	Purpose
analyze_testbed_output.php	Analyzes a game using the parsed text file output of user_pmrek.exe and the saved system activity records.
common_functions.php	Functions shared by PHP programs.
Makefile_pmrek	GNU Make command file to compile kernel module and executables.
pmrek_bash_profile	Appended to auto-login user's bash profile; calls start_testbed.
pmrek.c	Linux 2.6 kernel module for hardware control process.
pmrek.h	Header file containing definitions and data structures.
pmrek.sql	MySQL script to create database, tables and access permissions.
start_testbed	Shell script for running standalone testbed system; runs testbed.exe and restarts if terminated for upgrade.
testbed.c	Supervisory process for controlling kernel module, playing *Evel Knievel*, logging and analyzing process data; compiles into the executable testbed.exe.
testbed_performance.php	Creates summary statistics of all games analyzed.
user_pmrek.c	Utility program for parsing output of testbed.exe, displaying data structure sizes and simulating operation of the kernel module; compiles into the executable user_pmrek.exe.

This sets cpu_time to the number of CPU machine cycles that have occurred since booting. It is handy for precise timing measurements. Some switches, such as the pop bumpers and slingshots, require an immediate solenoid response.

Next, any enqueued commands are executed in order by calling the function pmrek_process_commands(). Commands can be sent from the supervisory program by writing to /dev/pmrek, or they can originate in the module itself. If a momentary solenoid is to be fired, the four-bit solenoid number is output using outb(). Then the enable output is set high to turn on the 74LS154 decoder output. The enable duration is kept by a counter that is decremented by the workqueue process delay, which is three milliseconds. Thus, a 26-millisecond solenoid pulse will take eight workqueue cycles before the enable bit is set low again to turn it off.

Next, the control process services the feature lamps. The AS2518 architecture includes a lamp driver board populated with 60 silicon controlled rectifiers (SCRs) to turn on or off individual light bulbs selectively on the play field and back box. Like the momentary solenoids, these SCRs are driven by decoders that take a four-bit input and turn on one of 16 outputs. To handle all 60 feature lamps, there are four decoders. The control program steps through the 16 positions and selectively turns on any of the four lamps associated with it. All of this must be done at the beginning of every cycle of the 120-Hertz, rectified DC power supply waveform. On the AS2518, this is accomplished using an interrupt triggered by a power supply zero-crossing detector. I decided not to use an interrupt. Instead, I employed a "shotgun" method by executing the control process at double this rate or faster, ensuring that the SCRs are triggered every cycle.

The last I/O operation performed by the workqueue process is to output the next row strobe for the next reading of the switch matrix. Then the process reschedules itself by issuing the command:

```
queue_delayed_work(pmrek_workqueue,
                   &pmrek_task,
                   pmrek_i.workqueue_delay);
```

The data structure pmrek_i contains all sorts of information about the pinball control system, including its workqueue delay, which has a value of 3. The kernel timer runs at 1,000Hz and is the heartbeat of the kernel. The workqueue delay is the number of beats before the delayed work is executed. Using this mechanism, frequencies much higher than what can be scheduled for ordinary user processes outside the kernel can be achieved, and they are more efficient in terms of the resources they use each time they execute.

The Software, Part III: Supervisory Control

Not everything in the pinball machine control system has to execute as frequently as the low-level hardware I/O operations. Game play itself—how the machine responds to switch detections, lighting different lamps and incrementing the player scores—operates just fine as an ordinary user process. In a sense, it is really a supervisory controller of the low-level I/O processing.

The kernel module should work for every game based on the AS2518 MPU. You can download the source code from the

Table 2. Supervisory Control Program Functions

Function Name	Purpose
game_add_player()	Called when the credit button is pressed (and there are credits) to start a new game or add more players.
game_ball_end()	Called when the outhole switch is detected while a ball is in play to initiate the bonus countdown, advance to the next ball, the next player or end the game.
game_collect_bonus()	Called after a ball ends to count down the current player's bonus.
game_segment_display()	Emulation of a seven-segment digital display on the computer screen for player scores, match count, credits and ball in play.
game_lamp_update()	Called after processing switch detections to update the disposition of all the feature lamps at once.
game_play_tune()	Plays various tunes by firing the chime momentary solenoids in predefined sequences.
game_switch_response()	Called for each valid switch detection retrieved from the kernel module; initiates all other events related to normal game operation.
game_watchdog()	Called every second to detect game faults, including missed switch detections, and either reprocesses the switch response or terminates the program.
process_output_file()	Called by the forked child process after a game is completed to analyze the log file recorded during the game play.
termination_handler()	Signal handler for cleanly ending the program; closes data log file and puts the kernel module into an idle state.
main()	Main program initializes kernel module data structures, computer screen and loops until a termination signal is caught; main loop processes user keyboard input, reads events from kernel module, calls game process functions, writes log file to disk and updates computer screen display.

Pinball Machine Reverse-Engineering Kit Project on SourceForge.net and compile it for your kernel. It will then be up to you to write the supervisory control software to play the particular game you are hacking. Table 1 lists other source code files in this package.

You are free to modify the C program testbed.c I wrote for *Evel Knievel*. It uses the ncurses screen handling package to provide a console color display and user input. A diagnostic display shows the disposition of the switch matrix, the lamps and the most recently fired solenoid. It also shows the player scores, as well as run-time statistics such as the average cycle frequency and execution time of the kernel workqueue process. Keyboard commands can be entered to turn the continuous solenoid on or off, fire momentary solenoids, turn feature lamps on or off and adjust the workqueue delay. Figure 6 shows a game in progress. Note the closed switches; these are drop targets that have been struck.

The supervisory program receives events passed from the kernel module by reading /dev/pmrek, which it has opened using the system call open(), just like any other file. Commands are then sent back to the module by writing to it. I tried to make the main functions correspond to my impression of the key events in a game of pinball. They are listed in Table 2.

You should be able to adapt this code to your particular game by tweaking the functions game_switch_response() and game_lamp_update(). How do you write the program without peeking at the original manufacturer's source code? There are plenty of clues painted on the play field itself, telling you what each switch scores and so on. Of course, you also can create your own rules, perhaps improving on weaknesses in the original design.

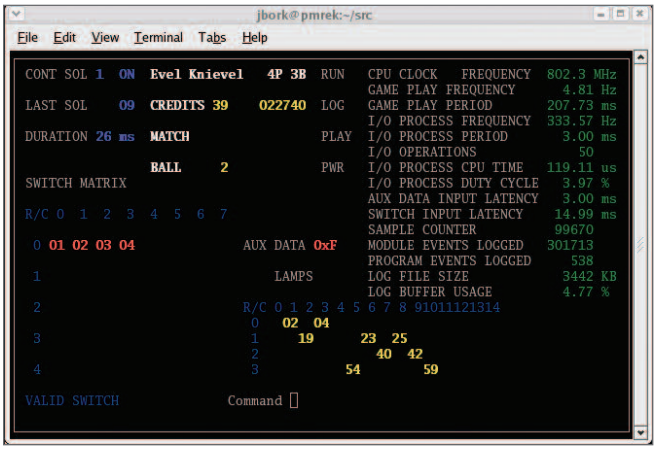


Figure 6. Supervisory Program Diagnostic Display

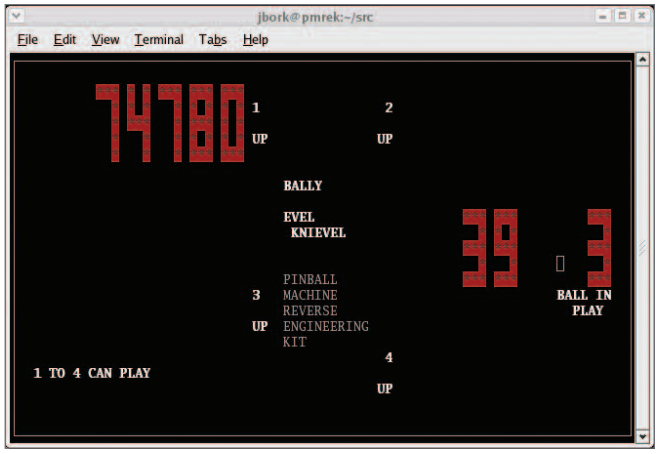


Figure 7. Back Board Score Display

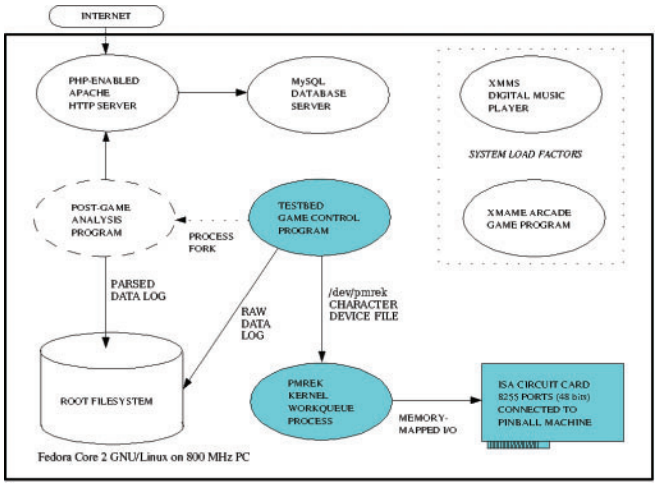


Figure 8. Computer System Block Diagram

Figure 9. Game in Action at Pinball at the Zoo

The diagnostic display is great for testing, but the player scores are too small. By default, the console simulates the large digital displays on the original back box, as shown in Figure 7. You can get to the diagnostic display by pressing the Self Test switch inside the pinball machine coin door.

We took the game to Pinball at the Zoo in Kalamazoo, Michigan in April 2005. Hundreds of people played the game, which collected statistical data that I used in my Master's thesis. After each game completes, a PHP program reads through the log file created by the game program. It generates an HTML document summarizing the event history of the game and statistics about its real-time performance. These results are then stored in a MySQL database to facilitate analysis of overall performance. Figure 8 is a block diagram of the setup. Figure 9 shows the game in action.

Conclusion

This project is a success story for the Linux 2.6 kernel. It demonstrates that a complex, real-time process control application can be created using a kernel workqueue instead of a complicated hardware interrupt or an additional, real-time package like RTLinux. Furthermore, through the choice of a pinball machine, a jack-of-all-trades hacker can produce something truly useful and fun to play.

Resources for this article: www.linuxjournal.com/article/8529.

John R. Bork is an IT System Integrator at Marathon Petroleum Company in Findlay, Ohio. He has been hacking Linux and pinball machines since 1999.

Radio's Next Generation: Radii

See how Linux can be used to prototype a sophisticated Internet appliance. BY DAN RASMUSSEN, PAUL NORTON AND JON MORGAN

Figure 1. Radii—a 1950s-style radio with Internet content.

A phrase we heard many times when we sought venture capital to develop the Internet appliance we call Radii was "If this were 1999, you would already have your money." Unfortunately, it was 2004 and there was no money for a risky consumer product such as Radii, despite our compelling prototype and a well-defined market. Rather than let our efforts go to waste, we decided to share the details of the prototype here with the Linux community that made its development possible. In this article, we explain how we quickly built our Radii prototype using low-cost hardware and Linux along with some of its companion software, including Perl and GCC.

Radii is a radio: a box with buttons and dials used to select bands and tune stations in a familiar way. Because this radio receives Internet radio, it provides hundreds of noise-free stations with a wide variety of listening options. The band selection dial, instead of AM and FM, is used to select genres such as News, Sports and Rock. The station selection dial scrolls through station names that can be tuned by clicking the select button.

At the beginning of this project, the three of us threw in $100 each and some spare time while continuing to work our day jobs. We never thought of this as an exercise in rapid prototyping; it was all about implementing our vision as quickly and inexpensively as possible. At every step of our development, we looked for the fastest way to get the task accomplished and balanced that against its cost.

The prototype is housed in a converted SW-54 radio made by the National Radio Company in the 1950s. The radio was in poor condition before the conversion. As admirers and collectors of old technology, we like to think we gave it a new lease on life.

Hardware Overview

The Radii core hardware platform is an old laptop running Linux. The operator interface consists of two rotary encoders, three momentary contact buttons, a 40x2 backlit LCD, a power supply and a retro radio cabinet. The encoders and buttons are connected to a PIC microcontroller development board that is, in turn, connected to the laptop's serial port. The LCD is connected to the laptop's parallel port.

On our budget of $300, cost was important. As such, eBay was our vendor of choice. Here is our hardware shopping list:

- PIC microcontroller dev board (OOPIC) ($70).
- One TTL to RS-232 chip (TI MAX232) and associated bits to interface the PIC to RS-232 ($5).
- Three momentary buttons for selection/special functions ($3).
- Two rotary encoders one for band selection, one for stations selection ($3).
- One 40x2 LED backlit LCD ($12 eBay).
- Gateway Solo 5150, 300MHz Pentium laptop, broken screen ($100 eBay).
- One National NC-54 vintage radio ($35 eBay).
- Power supply for PIC and LCD (3/$10 eBay).

The Straight Talk People℠
SINCE 1991

ABERDEEN

CAVERNOUS STORAGE BLINDING SPEED

INTRODUCING SATA II 3Gb/s

SERIAL ATA 3Gb/s

Aberdeen's new server and storage products featuring SATA II 3Gb/s technology offer these exceptional features:

- SATA II 3Gb/s enables the highest level of performance at low cost of ownership
- 300MBytes/sec per drive data transfer rates—twice as fast as regular SATA and three times as fast as Ultra ATA/100
- Single controller for up to 24 drives—single 12TB partion
- Intel RAID 6 Engine to support extreme performance RAID 6—can protect mission-critical data from two concurrent disk drive failures
- Status indication via HDD activity/fault LEDs and audible alarm
- Management port/firmware supports browser-based RAID manager, SMTP email notification, SNMP agent
- 500GB 8.5ms 7200rpm hot-swap SATA II 3Gb/s hard drives with 16MB cache featuring Rotational Vibration Safeguard and staggered (delayed) spin-up
- Native Command Queuing (NCQ) for extreme performance
- "Smooth Stream" technology for enhanced video streaming

ABERDEEN ABERNAS 128/213/233

2U Network Attached Storage with Hardware RAID 6

- **Up to 6TB of Network Attached Storage**
- Intel® Pentium 4® 3.0E GHz Processor with 800MHz FSB and 1MB Cache
- 1GB Dual Channel DDR400 SDRAM (Low Latency)
- 2 x 80GB SATA Operating System Drives Mirrored for Failover (one OS drive on AberNAS 128)
- Single controller for all drives with RAID level 0, 1 (10), 3, 5 and 6—single partion up to 6GB
- Up to 12 x 500GB 7200 RPM Hot-Swap SATA II 3Gb/s Hard Drives with 16MB cache
- Slim DVD Drive and Recovery DVD (no internal DVD drive on AberNAS 233)
- 460W Redundant Power Supply (300W non-redundant on AberNAS 128)
- Microsoft Windows® Storage Server 2003
- **5-Year Limited Warranty**

AberNAS 128 2TB 1U 4-bay	$4,995
AberNAS 213 4TB 2U 8-bay	$8,995
AberNAS 233 6TB 2U 12-bay	$11,995

"terrific for storage-intensive tasks" PC Magazine

ABERDEEN STIRLING X525 TERASTORUS

5U Storage Server—12 Terabytes in a Single Partition

- **Up to 12TB of Storage**
- Dual Intel® Xeon® Processors at 3.4GHz with 800MHz FSB and 1MB Cache
- Intel Extended® Memory 64 Technology (64-bit Support)
- Supermicro X6DHE-XG2 Motherboard with Intel® E7520 Chipset
- 2GB PC2-3200 ECC-Registered DDR2 SDRAM (Up to 16GB)
- 2 x 80GB 7200 RPM SATA Internal Hard Drives with 8MB cache
- Single controller for 24 drives with RAID level 0, 1 (10), 3, 5 and 6—single partion up to 12TB
- Up to 24 x 500GB 7200 RPM Hot-Swap SATA II 3Gb/s Hard Drives with 16MB cache
- Dual Intel 82546GB Gigabit NICs, ATI® Rage XL 8MB Graphics
- PCI-Express and PCI-X slots, 3.5" Floppy Drive and CD-ROM Drive
- LSI Logic 21320 Dual Channel Ultra320 SCSI Controller
- 950W 3+1 Redundant Hot-Swap Power Supply
- Easily scalable—up to three XDAS units with included SCSI controller
- Add more XDAS units with additional SCSI controllers
- **5-Year Limited Warranty**

Stirling X525 4TB 5U 24-bay	$9,275
Stirling X525 8TB 5U 24-bay	$13,105
Stirling X525 12TB 5U 24-bay	$16,945

ABERDEEN XDAS 500 SERIES

3U SCSI-to-SATA RAID Direct Attached Scalable Storage

- **Up to 8TB of Direct Attached Storage**
- Daisy Chain Units for Scalability
- OS and Host Independent
- Supports Dual Host for Server Failover Clustering
- 2 SCSI Host Channels
- Up to 16 x 500GB 7200 RPM Hot-Swap SATA II 3Gb/s Hard Drives with 16MB cache
- Intel® 80321 I/O Processor
- Modular/Cableless internal design
- RAID Controller Smart LCD Display
- RAID level 0, 1, 3, 5, 6, 10, JBOD
- HotSpare, HotSwap & Auto-Rebuild
- Multiple RAID arrays/LUNs
- Fast RAID Initialization and Fast Rebuild process
- Dual Flash ROM for Redundant Firmware Images
- Optional Fibre Channel Interface
- **5-Year Limited Warranty**

XDAS 501 4TB 2U 8-bay	$8,875
XDAS 502 6TB 2U 12-bay	$11,475
XDAS 503 8TB 3U 16-bay	$14,575

intel inside XEON

888-297-7409
www.aberdeeninc.com/linux

- Cables, connectors, bubble gum, baling wire and so on. ($25).
- Shipping, fees and taxes took up most of the remaining funds.

A PIC microcontroller is a single-chip computer produced by Microchip Technology, Inc. Although these tiny computers are capable of many useful things, we used it here simply to handle operator inputs. For prototyping with a PIC, a development board normally is used. PIC development boards provide an easy way to prototype a PIC application by allowing a range of input power options and easy access to the input and output pins for the chip. It is not necessary to use this, but it makes creating a prototype easier.

We used the OOPIC development board/system by Savage Innovations. It is inexpensive and provides a simple object interface for many input and output devices, including buttons, encoders and RS-232 serial communication. Unfortunately, there is no Linux development environment for OOPIC, although a SourceForge project is underway.

The hardware is rounded out with a Gateway Solo 5150 laptop that has a broken LCD. Similar laptops go for between $50 and $100 on eBay.

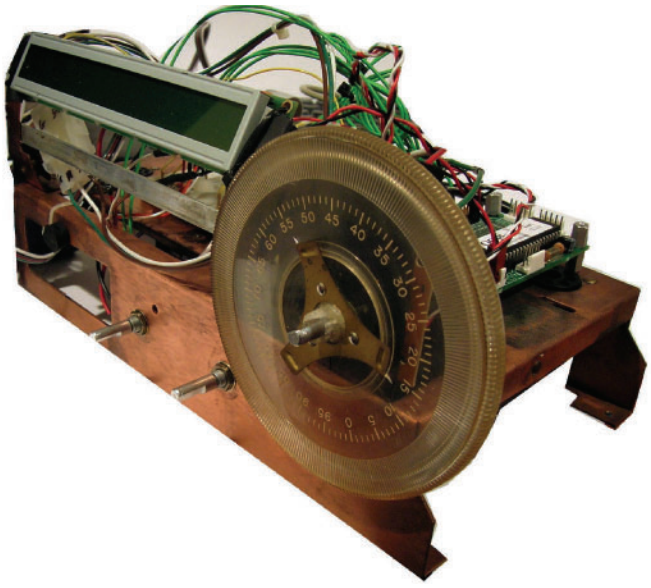


Figure 2. The original chassis is used to mount the controls, PIC development board, LCD and power supply.

The Operating System

We chose Linux from the start for many reasons. The primary reason is that most distributions are configured with many of the tools we thought we might use, such as mpg123, XMMS, Perl and compilers. It also helped us stay on budget because it's free. Linux makes prototyping easy, because many applications and utilities have retained their command-line interface, allowing their use from scripts, such as the one written for Radii and described below.

Installation and configuration of the OS was straightforward, except for audio support. Because our laptop was so old, most installers were not able to detect the audio hardware. In an unscientific way, we tried many different Linux distributions until we found one that installed easily on our machine. We wound up installing Fedora Core 2 with ALSA (Advanced Linux Sound Architecture) support.

To get sound working for your particular machine, it is most important to identify your sound hardware. In our case, we were able to determine the sound hardware by Googling on the model number for this laptop. Once we determined which sound hardware we had, we were able to locate and install the appropriate ALSA driver for our machine, the ES1879 ESS Audio Driver, from the ALSA Project site. You may need to tweak some of the default ALSA parameters by using the alsamixer utility.

Software Components

With the hardware in place and the OS working, it all came down to finding or creating the required software components. We had simple requirements:

- An audio stream player.
- An LCD controller.
- An application to process operator-induced signals from the serial port and interact with the stream player and LCD.

The Audio Stream Player

We needed a way to play streaming audio that we could control from our application. We initially dismissed XMMS because it is a GUI application, but we later re-examined it and discovered that XMMS can be manipulated from the command line.

The XMMS application provides many handy options that can be used to control an already-running instance of itself. It can be stopped by issuing the -s argument. The playlist can be updated by using `-p <playlist>` and the playlist argument can be the URL of a stream. Use `xmms -h` for complete details.

For example, you ask XMMS to switch from its current selection to the AM 1710 Antioch Internet station (old-time radio), by issuing the command:

```
xmms -p http://66.54.65.226:9022
```

To stop, use `xmms -s` and so on.

XMMS completely covered our needs for a player, but it introduced a problem as well. XMMS is a GUI application, so it requires a running X11 server. Rather than tax the available resources on our low-powered laptop, we used the X Virtual Frame Buffer, Xvfb. Xvfb provides a lightweight X11 server that can be used to provide X11 resources to applications that require them, but it does nothing else—it is invisible.

The LCD Controller

We required a CLI application that would display a string on our parallel port LCD. After Googling for this, we found a FOSS application called lcd-info. lcd-info displays system performance information on an HD44780-compatible LCD connected to the system parallel port. It was not quite what we needed, but after studying its source for a few minutes, we found that it could be adapted easily for our purpose.

Appro *Xtreme*Blade Solution

Clusters. High performance, high availability.

The Most Powerful InfiniBand™ Enabled Blade Solution

Fits up to 12 Blades in a 7U Height Subrack

Analyst Firm IDC examines the Appro *Xtreme*Blade based on AMD Opteron™ processors to address HPC and Enterprise needs. In addition, IDC provides good information on technologies that are experiencing growth in adoption such as blade servers. Get this IDC White Paper at http://www.appro.com.

- **Scalable** - 2-way and/or 4-way configuration
- **Flexible** - Ability to mix blade configurations in one rack cabinet
- **Connected** - Integrated Infiniband™ and Gigabit switches
- **Powerful** - AMD Opteron™ based server with leading 32-bit performance with 64-bit capabilities
- **Reliable** - Hot swappable blades, redundant switches, power supplies and cooling fans
- **Smart** - BladeDome II – Centralized remote blade server and system management
- **Balanced Architecture** - Memory, I/O and communication bandwidth match CPU bandwidth

AMD Opteron™ Processors
- Integrated AMD HyperTransport™ technology allows for concurrent multiple processors in a single system.
- Shorten run-time cycles and increase bandwidth for processing computing requests.
- 32 bit applications while you migrate to 64 bit computing for long-term investment protection.

APPRO
HPC Cluster Solutions

Appro has everything you need to create a network blade cluster-ready.
On-site maintenance and installations services are also available.

For more information, please visit www.appro.com
or call Appro Sales at 800.927.5464, 408.941.8100.

AMD 64 Opteron

lcd-info is written in C and compiles into a CLI application. We compile our simpler application with a trivial invocation of GCC:

```
% gcc -o setlcd setlcd.c iolcd.c
```

The low-level routines that control the LCD are in iolcd.c, which was borrowed without modification from the lcd-info Project. setlcd.c is the Radii-specific piece that uses functions found in iolcd.c. We called our binary setlcd, and it is run like so:

```
% setlcd <string to display>
```

Building the cable to interface the LCD to the parallel port was more time consuming than was adapting lcd-info. It seems that there should be an appropriate off-the-shelf cable, but the pinout on the LCD-side of the cable varies with the manufacturer/model. Rather than finding exactly the right cable/LCD pair, we elected to make our own cable for the LCD we had acquired based on price.

The Radii Application

We built the Radii application using Perl. We chose Perl because it's a language we know well, it has many supporting packages and the update/compile/debug cycle is fast.

The first thing to do is read the input from the PIC development board connected to the serial port. We used the Device::SerialPort package. Here is the beginning of our application, which shows how to initialize the serial port using the Device::SerialPort module:

```
#!/usr/bin/perl
use Device::SerialPort;

use strict;

# Set up the port.
# All port settings must match the PIC settings.
my $port = new Device::SerialPort("/dev/ttyS0");
$port->baudrate(9600);
$port->parity("none");
$port->databits(8);
$port->stopbits(1);
$port->handshake('none');
$port->write_settings;
```

Then we needed to handle the following messages sent from the PIC development board based on user input:

```
Msg Meaning
--- -------
U   The station encoder rotated one unit up
D   The station encoder rotated one unit down
s   The select button was pressed
u   The band encoder rotated one unit up
d   The band encoder rotated one unit down

while ( 1 )
{
  while (! ($code = $port->input))
  {
     select undef, undef, undef, 0.075;
  }
}
```

The outer while loop keeps the application running until it is killed or dies. The inner while loop attempts to read from the serial port. If there is nothing to read, it sleeps for a short time, 0.075 seconds, and then tries again. This sleep is important to keep the application from spinning too hard and consuming a lot of CPU time. Any messages that arrive while the loop is sleeping accumulate on the port and are available the next time we read.

When an input message is received, the application always should respond by updating the LCD. It sometimes should respond by changing the current station, that is, when the selection button is pressed.

When we get a Station Up (U) or Station Down (D) message, we need to display the next station on the LCD, but we don't want the station to change until the user sends a select signal. This brings us to the LCD message display. As previously noted, we use the setlcd command, but now we call it from the Perl script using the Perl system command:

```
system("setlcd",
       "Sel:$radiiStn{$curBand}{$choice}{name}");
```

where $radiiStn{$curBand}{$choice}{name} is a hash that is indexed by way of the band index and the choice index. It contains the necessary selection information: display name (used here), station URL and its band.

Once the operator clicks the select button, the PIC sends an s message. In response, the system updates the LCD to the new station name and signals XMMS to play the new stream, again using Perl's system command:

```
system("setlcd",
       $radiiStn{$curBand}{$choice}{name});
system("/usr/bin/xmms",
       "-p",
       $radiiStn{$curBand}{$choice}{station});
```

Configuration Using XML

The Radii application is configured using a simple XML input file:

```
<?xml version="1.0"?>
<Radii>
  <station url="http://66.54.65.226:9022">
    <band>OLD TIME RADIO</band>
    <name>AM 1710 Antioch</name>
  </station>
  <station url="http://205.188.234.38:8040">
    <band>Celtic</band>
    <name>CelticGrove.com 24/7 Celtic/Irish</name>
  </station>
  .
  .
  .
</Radii>
```

The XML configuration file can be read using the XML::Simple Perl module.

```
my @station;
my %radiiStn = ();
my %bands = ();

my $file = 'stations.xml';
my $xs1 = XML::Simple->new();
my $doc = $xs1->XMLin($file);

foreach my $key (keys (%{$doc->{station}}))
{
     $band = $doc->{station}{$key}{band};
     $url  = $doc->{station}{$key}{url};
     $name = $key;
     $bands{$band} += 1;
     $radiiStn{$band}{$bands{$band}}{name} =
        $bands{$band}.":$band: ".$key;
     $radiiStn{$band}{$bands{$band}}{station} = $url;
}
```

This code utilizes Perl hashes for the required band and station information. Band information, including name and number of stations, is kept in the bands hash. Station information, such as name, URL and band, is kept in radiiStn hash.

See the on-line Resources for the URL of a site with the complete script and other associated software, along with details on how to build the hardware.

Conclusion

Radii demonstrates how Linux can be used to prototype a complex consumer device quickly and cheaply. As the iPod revolution takes hold and satellite radio becomes more popular, Radii-like devices inevitably will change the way radio is broadcast and received all over the world.

Rapid prototyping does not require particular hardware, sets of tools or languages. It's not about finding the best solution; it's about getting it done quickly using the available resources. That pool of resources is vast when it is FOSS on Linux. Keep your eye on the goal while you sort through the potential building blocks. Tweak as necessary, and then glue it all together with your language of choice.

We configured our laptop to boot to run-level 3, full multiuser mode. After the laptop boots, we start Xvfb, set our DISPLAY variable, start XMMS and start the Radii application. The startup sequence is:

```
% Xvfb :1 &
% export DISPLAY=:1.0
% xmms &
% radii.pl
```

Then we hide the laptop and enjoy the radio that we call Radii.

Resources for this article: www.linuxjournal.com/article/8537.

Dan Rasmussen (dan@retro-tronics.com) is a Senior IT Specialist for IBM and holds a BS in Math from UMass/Amherst and an MSCS from RPI. He has been working as a software engineer and IT consultant for nearly 20 years. Dan is also an avid collector of vintage electronics.

Paul Norton (pddknorton@charter.net) spent his early career with large corporations, including Xerox, Litton Industries and Pearson. For the last 20 years he has worked exclusively with small companies, several of which were start-ups. His main business focus is establishing operations and strategic management of product and market development. He has a three-pronged education in technology, humanities and business, holding an MBA from Brunel University in West London. He has worked extensively in Europe and the United States.

President of Product Marketing at Tatara Systems, Jon Morgan has more than 18 years of marketing, technical and management experience in the telecom and data communications industries. Most recently, Jon was Director of Product Management/Marketing for Appian Communications, Inc. Prior to joining Appian, Jonathan held various management positions at Fujitsu Network Communications (FNC). Prior to Fujitsu, Jonathan spent seven years at Bellcore. Jon holds a BSEE from Washington University in St. Louis and an MSEE from Rutgers University.

The Ultimate Linux Lunchbox

For those of you with carry-on, high-performance computing clusters, please ensure that they are securely stowed underneath the seat in front of you. BY RON MINNICH

In this article, we describe the construction of the Ultimate Linux Lunchbox, a 16-node cluster that runs from a single IBM ThinkPad power supply but can, as well, run from an N-charge or similar battery. The lunchbox has an Ethernet switch built-in and has only three external connections: one AC plug, one battery connector and one Ethernet cable. To use the lunchbox with your laptop, you merely need to plug the Ethernet cable in to the laptop, supply appropriate power—even the power available in an airplane seat will do—and away you go, running your cluster at 39,000 feet. We've designed the lunchbox so that we can develop software on it, as a private in-office cluster or a travel cluster. The lunchbox is an example of a newer class of clusters called miniclusters.

Figure 1. Minicluster I used four Pentium-based single-board computers (courtesy Sandia National Labs).

Miniclusters

Miniclusters were first created by Mitch Williams of Sandia/Livermore Laboratory in 2000. Figure 1 shows a picture of his earliest cluster, Minicluster I. This cluster consisted of four Advanced Digital Logic boards, using 277MHz Pentium processors. These boards had connectors for the PC/104+ bus, which is a PC/104 bus with an extra connector for PCI.

As you can see, there are only four nodes in this cluster. The base of the cluster is the power supply, and the cluster requires 120 Volts AC to run. We also show a single CPU card on the right. The green pieces at each corner form the stack shown in the pictures. A system very much like this one is now sold as a product by Parvus Corporation.

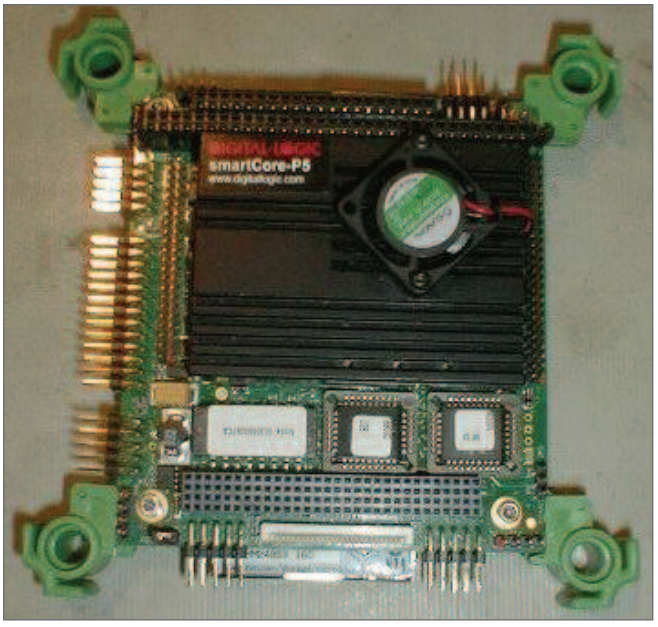


Figure 2. One Node of Minicluster I (courtesy Sandia National Labs)

HPC puts you in a win-win-WIN situation!

- Servers
- Workstations
- Storage Devices
- Cluster Solutions

WIN A FREE SERVER!

- 1U Dual Opteron Chassis
- 350W Power Supply
- Dual AMD Opteron 252 processors
- 2 x 120GB sATA hotswap hard drives
- 2GB PC3200 ECC Reg Memory

Enter the drawing by logging on to:

www.hpcsystems.com

Doing business with HPC Systems, Inc., has always been a win-win situation. We're an SDB / 8(A) certified manufacturer and integrator of computer solutions such as quad Opteron servers, supplying a broad range of customized server, storage systems, and cluster solutions to Government, University, Corporate, and High Performance Computing markets. HPC Systems, Inc., offers a complete range of consultancy, hardware integration, software debugging, cluster design, installation services, and system optimizations services.
Call today for a free no-obligation quote.

HPC SYSTEMS INCORPORATED

Experts in dual, quad, 8-way AMD Opteron Server

GSA GS-35F-0596R Contract Holder

SBA 8(A) CERTIFIED

Toll free: 888-725-3472 • Local: 408-943-8282 • Fax: 408-943-8341 • E-mail: sales@hpcsystems.com

The Bento Series

We were intrigued by this cluster and thought it would be an ideal platform for Clustermatic. In the summer of 2001, we ported LinuxBIOS to this card and got all the rest of the Clustermatic software running on it. When we were done, we had a card that booted to Linux in a few seconds, and that booted into full cluster mode in less than 20 seconds. Power and reset cycles ceased to be a concern.

We provided the LinuxBIOS and other software to Mitch, and he modified Minicluster I to use it. Mitch was able to remove three disks, reducing power and improving reliability. One node served as the cluster master node, and three other nodes served as slave nodes.

Inspired by Mitch's work, we built our first Bento cluster in 2002. In fact, the lunchbox used for that system is the one we use for the Ultimate Linux Lunchbox. This system had seven CPU cards. It needed two power supplies, made by Parvus, which generate the 5V needed for the CPU cards and can take 9–45 VDC input. It had a built-in Ethernet hub, which we created by disassembling a 3Com TP1200 hub and putting the main card into the lid. This cluster used three IBM ThinkPad power supplies. Two of the supplies are visible in the lid, on either side of the Ethernet hub. The third is visible at the back of the case. One supply drives the hub, the other two drive each of the two supplies. The supplies and fan board for each supply can be seen at the far right and left of the box; the seven CPU boards are in the middle.

Bento was great. We could develop on the road, in long and boring meetings and test on a seven-node cluster. Because the reboot time was only 15 seconds or so for a node at most, testing out modules was painless. In fact, on this system, compiling and testing new kernel modules was about as easy as compiling and testing new programs. Diskless systems, which reboot really quickly, forever change your ideas about the difficulty and pain of kernel debugging.

During one particularly trying meeting in California, we were able to revamp and rewrite the Supermon monitoring system completely, and use it to measure the impact of some test programs (Sweep3d and Sage) on the temperature of the CPUs as it ran. Interestingly enough, compute-intensive Fortran programs can ramp up the CPU temperature several degrees centigrade in a few seconds. The beauty of these systems is that if anyone suspects you are getting real work done, instead of paying attention to the meeting, you always can hide the lunchbox under your chair and keep hacking.

Bento used a hub, not a switch, and Erik Hendriks wanted to improve the design. The next system was called DQ. DQ was built in to an attractive metal CD case, suitable for carrying to any occasion, and especially suitable for long and boring meetings. As our Web page says, we'll let you figure out the meaning of the name. Hint: check out the beautiful pink boa carrying strap in the picture.

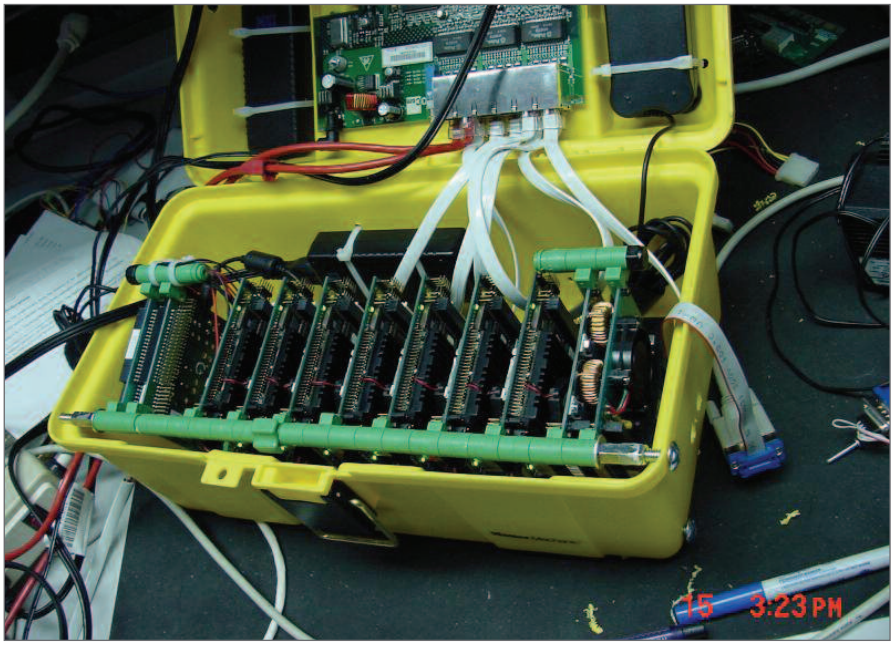


Figure 3. The First Lunchbox Cluster, Bento

DQ Cluster

We were able to get an awful lot of development work done on DQ at a meeting in Vegas. The switch improved the throughput of the system, and the package was bombproof (although we avoided using that particular phrase in airport security lines). The hardware was basically the same, although one thing we lost was the integrated ThinkPad power supplies—there was no lid on DQ in which to hide them. Nevertheless, this was quite a nice machine.

Figure 4. The DQ cluster featured an Ethernet switch and a colorful carrying strap.

Sandia was not asleep at the time. Mitch built Minicluster II, which used much more powerful PIII processors. The packaging was very similar to Minicluster I. Once again, we ported LinuxBIOS to this newer node, and the cluster was built to have one master with one disk and three slaves. The slave nodes booted in 12 seconds on this system. In a marathon effort, we got this system going at SC 2002 about the same time the lights started going out. Nevertheless, it worked.

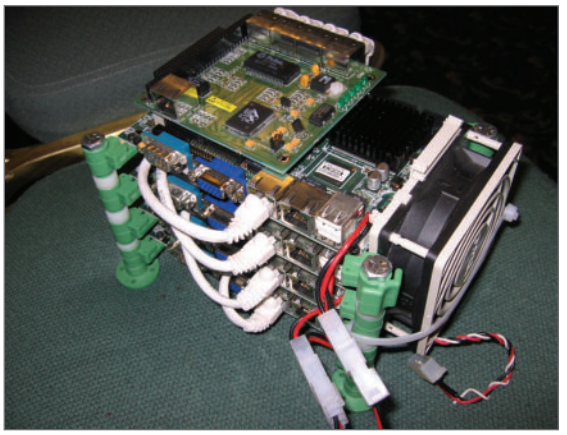


Figure 5. The Geode minicluster needed a full-size power supply to deal with the demands of Pentium III-based nodes.

One trend we noticed with the PIII nodes was increased power consumption. The nodes were faster, and the technology was newer, and the power needed was still higher. The improved fabrication technology of the newer chips did not provide a corresponding reduction in power demand—quite the contrary.

It was no longer possible to build DQ with the PIII nodes—they were just too power-hungry. We went down a different path for a while, using the Advantech PCM-5823 boards as shown in Figure 5. There are four CPU boards, and the top board is a 100Mbit switch from Parvus. This switch is handy—it has five ports, so you can connect it directly to your laptop. We needed a full-size PC power supply to run this cluster, but in many ways it was very nice. We preserved instant boot with LinuxBIOS and bproc, as in the earlier systems.

As of 2004, again working with Mitch Williams of Sandia, we decided to try one more Pentium iteration of the minicluster and set our hungry eyes on the new ADL855PC from Advanced Digital Logic. This time around, things did not work out as well.

First, the LinuxBIOS effort was made more or less impossible by Intel's decision to limit access to the information needed for a LinuxBIOS port to Intel chipsets. We had LinuxBIOS coming up to a point, and printing out messages, but we never could get the memory controller programmed correctly. If you read our earlier articles on LinuxBIOS (see the on-line Resources), you can guess that the romcc code was working fine, because it needs no memory, but the gcc code never worked. Vague hints in the available documents indicated that we needed more information, but we were unable to get it.

Second, the power demand of a Pentium M is astounding. We had expected these to be low-power CPUs, and they can be low power in the right circumstances, but not when they are in heavy use. When we first hooked up the ADL855PC with the supplied connector, which attaches to the hard drive power supply, it would not come up at all. It turned out we had to fabricate a connector and connect it directly to the motherboard power supply lines, not the disk power supply lines, and we had to keep the wires very short. The current inrush for this board is large enough that a longer power supply wire, coupled with the high inrush current, makes it impossible for the board to come up. We would not have believed it had we not seen it.

Instead of the 2A or so we were expecting from the Pentium M, the current needed was more on the order of 20A peak. A four-CPU minicluster would require 80A peak at 5 VDC. The power supply for such a system would dwarf the CPUs; the weight would be out of the question. We had passed a strange boundary and moved into a world where the power supply dominated the size and weight of the minicluster. The CPUs are small and light; the power supply is the mass of a bicycle.

The Pentium M was acceptable for a minicluster powered by AC, as long as we had large enough tires. It was not acceptable for our next minicluster. We at LANL had a real desire to build 16 nodes into the lunchbox and run it all on one ThinkPad power supply. PC/104 would allow it, in terms of space. The issues were heat and power.

What is the power available from a ThinkPad power supply? For the supplies we have available from recent ThinkPads, we can get about 4.5A at 16 VDC, or 72 Watts. The switches we use will need 18 Watts, so the nodes are left with about 54 Watts between them. This

is only 3W per node, leaving a little headroom for power supply inefficiencies. If the node is a 5V node, common on PC/104, then we would like .5A per node or less.

This power budget pretty much rules out most Pentium-compatible processors. Even the low-power SC520 CPUs need 1.5A at 5V, or 7.5 Watts—double our budget. We had to look further afield for our boards.

We settled on the Technologic TS7200 boards for this project. The choice of a non-Pentium architecture had many implications for our software stack, as we shall see.

The TS7200

The TS7200, offered by Technologic Systems, is a StrongARM-based single-board computer. It is, to use a colloquialism, built like a brick outhouse. All the components are soldered on. There are no heatsinks—you can run this board in a closed box with no ventilation. It has a serial port and Ethernet port built on, requiring no external dongles or modules for these connections. It runs on 5 VDC, and requires only .375A, or roughly 2W to operate. In short, this board meets all our requirements. Figure 6 is a picture of the board. Also shown in Figure 6 is a CompactFlash plugged in to the board, although we do not use one on our lunchbox nodes.

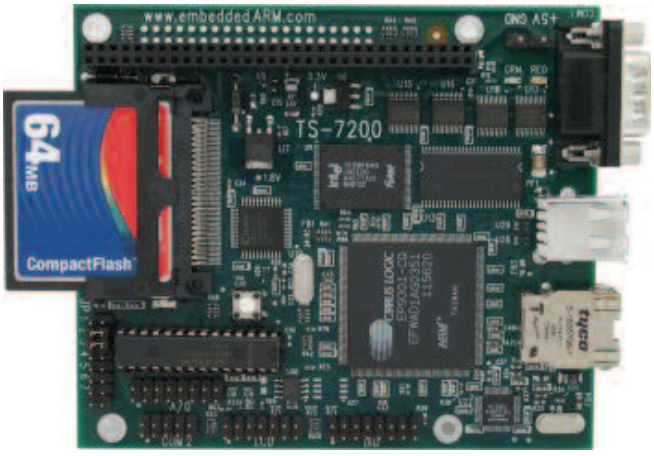


Figure 6. The TS7200, from Technologic Systems, is StrongARM-based, needs no heatsinks and draws only about two Watts (courtesy Technologic Systems).

One item we had to delay for now is putting LinuxBIOS on this board. The soldered-on Flash part makes development of LinuxBIOS difficult, and we were more concerned with getting the cluster working first. The board does have a custom BIOS with the eCos operating system, which, although not exactly fast, is not nearly as slow as a standard PC BIOS.

Building the Lunchbox

There are several factors that determine the shape of a minicluster: the box, the size and shape of the board and the board spacing, or distance between boards. The spacing tends to dominate all other factors and is complicated by the fact that PC/104 was not designed with multiprocessors in mind. All I/O boards in PC/104 stack just fine, as long as there is only one CPU board; we are breaking the rules when we stack CPU boards, and it gets us into trouble every time. On all the miniclusters shown, there was at least one empty board space between the boards. Nevertheless, the process of designing starts with the box, then the board shape and then the board spacing.

First, the box: it's the same box we used earlier. Also, we're going to use the same Parvus SnapStiks that we have been using for years to stack boards. We bought the professional set, part number PRV-0912-71. The SnapStik works well in the lunchbox format. One warning: just buy 1/4" threaded rod to tie the stack together. Do not use the supplied threaded plastic rod that comes with SnapStik kits. That plastic rod tends to, well, "snap" under load, and watching bits of your minicluster drop off is less than inspiring.

Second, the size and shape of the TS7200 nodes: there's a slight problem here. The boards are not quite PC/104: they're a little large. One way to tell is that two of the holes in the TS7200 are not at the corners. In Figure 7, the holes are in the right place, but the board extends out past them, leaving the holes too far in from the edge. The board is a bit bigger to accommodate the connectors shown on the right. These connectors caused two problems, which we will show below.

Third, the stack: the tight spacing was going to make the stack more challenging than previous miniclusters. We would have to find a way to make the SnapStiks work with a nonstandard board form factor and the close spacing.

To solve the SnapStik problem, we spent some time seeing how the supports could fit the board. The best we could find was a configuration in which three SnapStiks fit on three of the holes in the board, as shown in Figure 7. Notice the threaded metal rod, available in any hardware store.

For the fourth hole, we set up a spacer as shown in Figure 8.

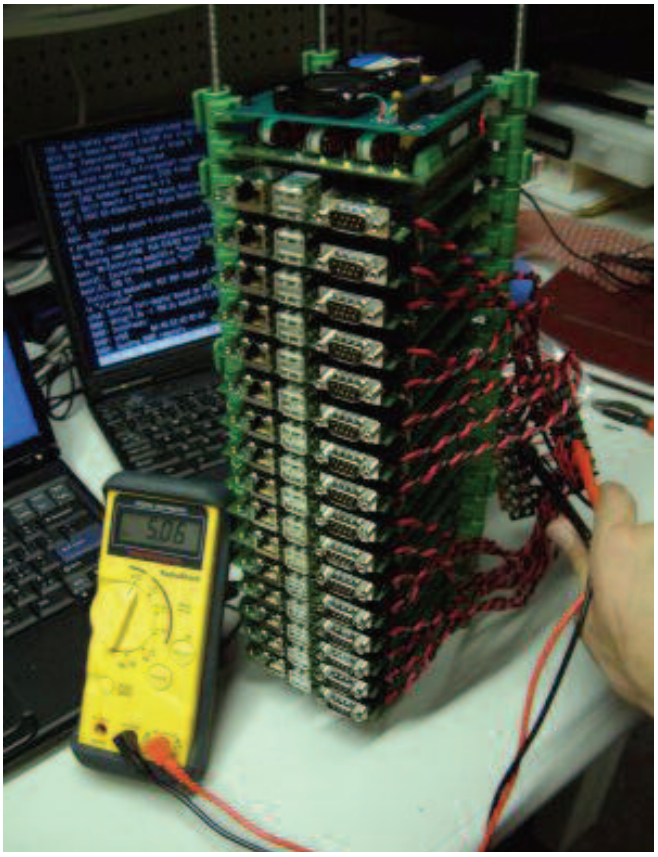
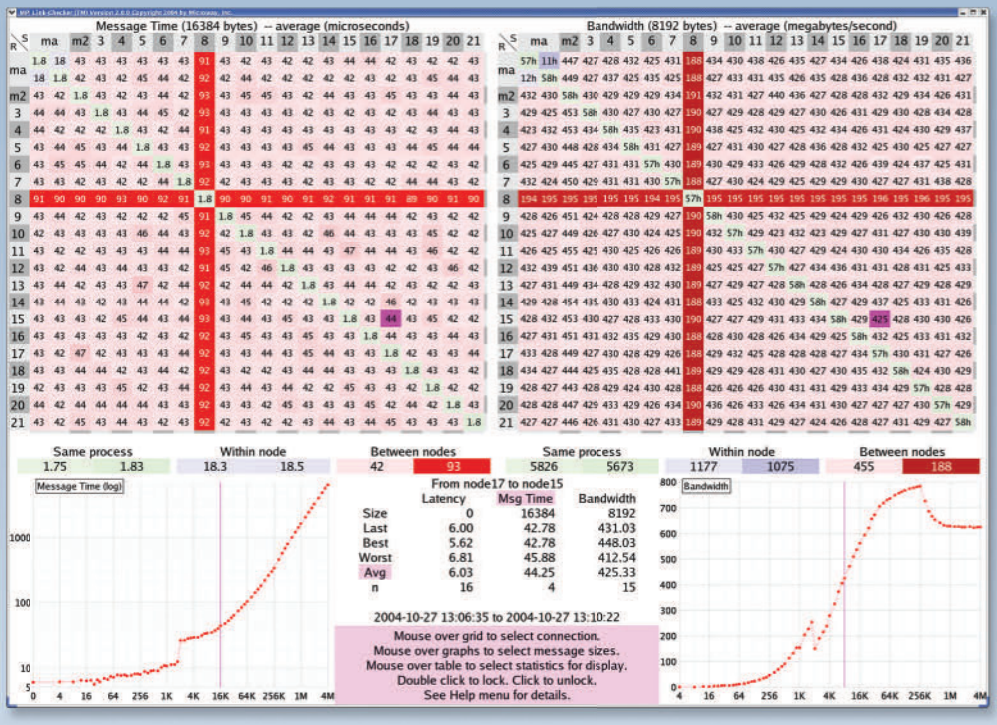


Figure 7. Stack Showing Three out of Four SnapStiks Connected

X Marks the Slow Node!



Microway® Quad Opteron™ Cluster with 36 Opteron 852s, redundant power and 45 hard drives in CoolRak™ cabinet.

MPI Link-Checker™ to the Rescue!

A single slow node or intermittent link can cut the speed of MPI applications by half. Whether you use GigE, Myrinet, Quadrix, InfiniBand or InfiniPath HTX, there is only one choice for monitoring and debugging your cluster of SMP nodes:

Microway's MPI Link-Checker™

Our unique diagnostic tool uses an end-to-end stress test to find problems with cables, processors, BIOS's, PCI buses, NIC's, switches, and even MPI itself! The newest release provides ancillary data on inter-process and intra-CPU latency which can vary by a factor of 10 between MPI versions. MPI Link-Checker is also useful for porting applications to new hardware. It provides instant details on how latency and bandwidth vary with packet size. It is available now for a free 30 day evaluation!

Wondering what's wrong with your cluster, or need help designing your next one? Call our HPC staff at 508-746-7341. Visit microway.com to learn about new low latency interconnects including the PathScale InfiniPath HTX Adapter, which delivers unmatched MPI latency of under 1.5 microseconds.

Microway has been an innovator in HPC since 1982. We have thousands of happy customers. Isn't it time you became one?

Call us first at 508-746-7341 for quotes and benchmarking services. Find technical information, testimonials, and newsletter at www.microway.com.

Excellence in CLUSTER Technology
64-bit Turnkey Solution
WINNER
ClusterWorld

Microway®
23 Years of Expertise Built In

PathScale™
Accelerating Cluster Performance™

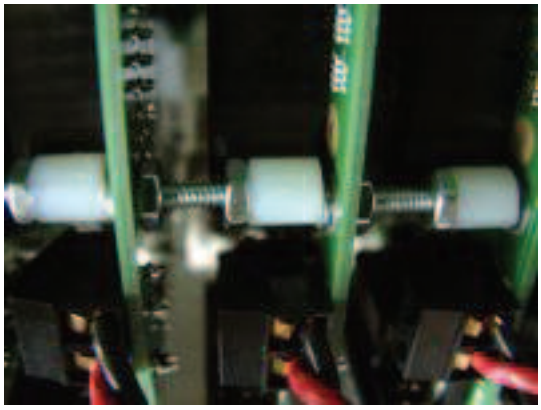


Figure 8. The Spacer in the Fourth Hole

The spacer is a simple nylon spacer from our local hardware store. The bolts and nuts allow us to create an exact spacing between the boards. We needed the exact spacing for the next problem we ran into.

The boards cannot be stacked at exactly a one-per-slot spacing. There is an Ethernet connector that needs just a bit more room than that—if the boards are stacked too closely, the Ethernet connector on the lower board shorts out the Ethernet connector pins on the higher board. The spacing could be adjusted easily with the nut-and-bolt assembly shown above, but how could we space the SnapStiks?

If you look at the Geode cluster shown in Figure 8, you can see some white nylon spacers between the green SnapStiks. That is one way to do it. But that spacing would have been too large to allow 16 nodes to fit into the lunchbox. We needed only about 1/32 of an inch in extra spacing.

Josiah England, who built this version of the lunchbox, had a good idea: small wire rings, which he says he learned how to build while making chainmail. The fabrication is shown in Figures 9–11. The wire rings add just enough space to create enough clearance between the boards, while still allowing us to put 16 boards in the lunchbox.

With this fix, we now had a stack that was spaced correctly. The stack shown above was finished off with a Parvus OnPower-90 power supply and a Parvus fan board, which you can see at the top. This supply can provide 18A at 5V, more than enough for our needs, as well as the 12V needed for the switch.

Our next step was the Ethernet switch. At first, we tried using several cheap eight-port switches in the lid, as shown in Figure 12. By the way, these miniclusters always include a bit

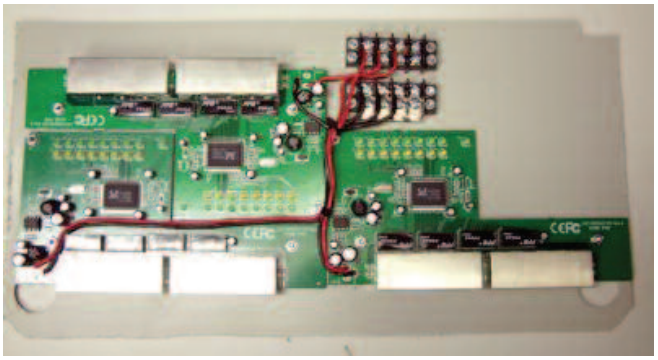


Figure 12. First try at switches: the gray panel is a mailbox shelf.

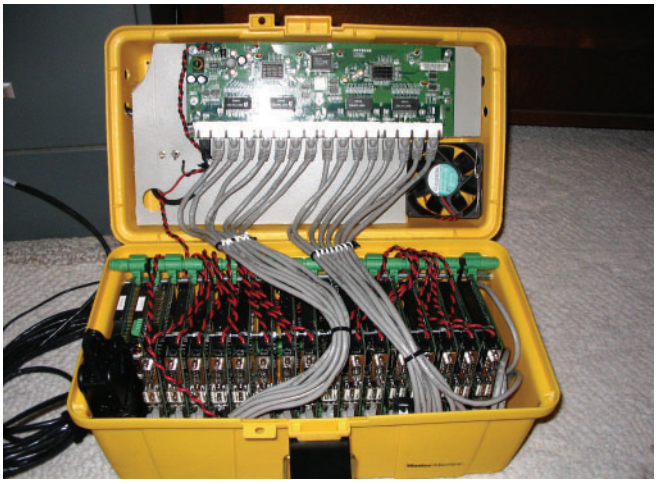


Figure 13. Final design: one of the switches on the gray metal panel, to the left of the Ethernet plugs, controls power to the nodes and the Ethernet switch, and the other one controls the fan.

of improvisation. The switches shown are bolted to a shelf from our departmental mailbox. The shelf is a nice, gray plastic and was ideal (once we trimmed it with a hacksaw) for our purposes. Notice the nice finger hole, which can be used for routing wires under the lid. We'd like to think we used the Erik Hendriks mailbox shelf, since Erik's bproc work was so important to our minicluster development. Erik is now at Google.

The cascaded switches worked very poorly. The nodes would not come up on the network reliably. It all looked great, with 48 LEDs, but it did not work at all. DHCP requests were dropped, and the nodes took forever to come up.

The second attempt was to get a Netgear 16-port switch,

Figures 9–11. Medieval solution to a 21st-century hardware problem: wire spacing rings constructed chainmail-style (courtesy Josiah England).

remove the switch from the case and put it into the lid. This required that we sacrifice another mailbox shelf, but we have plenty. This change worked fine. The nodes come up very quickly now, as packets are not getting lost.

You can see the final configuration in Figure 13. Notice the two switches: one switch controls power to the Ethernet switch and nodes, and the other controls power to the fan. We're not yet sure we need the fan but we're being careful.

Regarding Ethernet cables: always label them, and always make it so you can figure out, easily, which one goes into which network switch connector. Put them into the switch in some order, left to right or right to left. Just make sure you can tell, at a glance, which LED on the switch goes with which board. You'll be glad you did.

Lunchbox Software

Okay, we've built the hardware. Now, what is the software?

In years past, it would have been bproc, as found on the Clustermatic site (see Resources). bproc has a problem, however; it cannot support heterogeneous systems. The very nature of bproc, which requires that process migration works, makes the use of different architectures, in a single system, impossible. We're going to have to use something else. We want to continue using our ThinkPad laptop as the front end; there are no StrongARM laptops that we know of. It's clear that we are going to need new software for our minicluster.

Fortunately, the timing for this move is good. As of 2.6.13, there is now support for the Plan 9 protocol in the standard Linux kernel. This module, called 9p (formerly v9fs), supports the Plan 9 resource-sharing protocol, 9p2000. At the same time this code was being ported to the Linux kernel, Vic Zandy of Bell Labs was working with us on xcpu, a Plan 9 version of bproc. One of the key design goals of xcpu was to support heterogeneous systems. The combination, of 9p in the Linux kernel and xcpu servers ported to Linux, has allowed us to build a replacement system for bproc that supports architecture and operating system heterogeneity. Finally, the introduction of new features in 2.6.13 will allow us to remove some of our custom Clustermatic components and improve others. A key new feature is Eric Biederman's kexec system call, which replaces our kmonte system call.

Figure 14 shows a quick outline of the standard bproc boot sequence, as it works on our miniclusters and clusters with thousands of nodes.

The boot sequence, as shown, consists of LinuxBIOS, Linux, Linux network setup, Linux loading another kernel over the network and Linux using the kmonte system call (part of Clustermatic) to boot that second kernel as the working kernel. Why are there two kernels? In Clustermatic systems, we distinguish the OS we use to boot the system from the OS we run during normal operation. This differentiation allows us to move the working kernel forward, while maintaining the boot kernel in Flash.

The new boot sequence is shown in Figure 15. If it looks

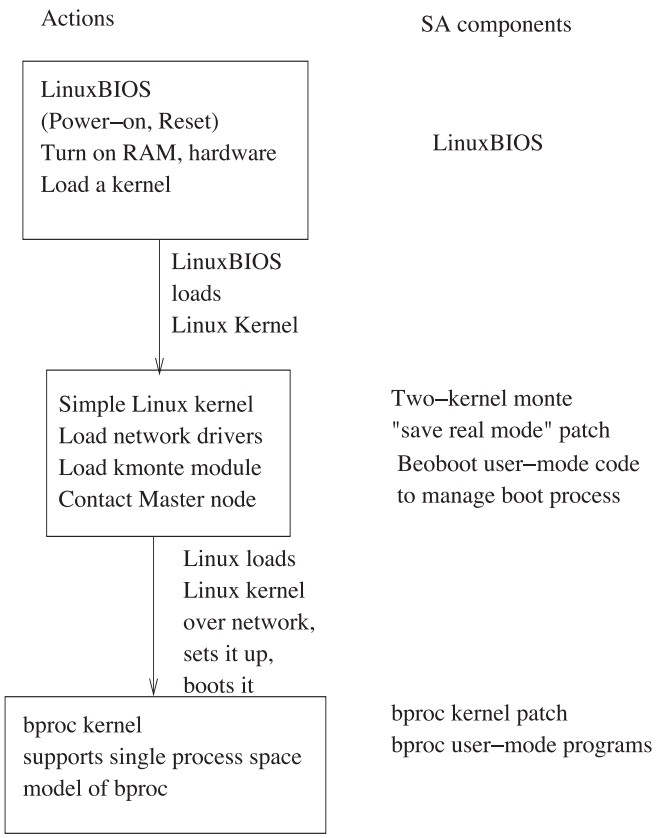


Figure 14. A View of SA Components

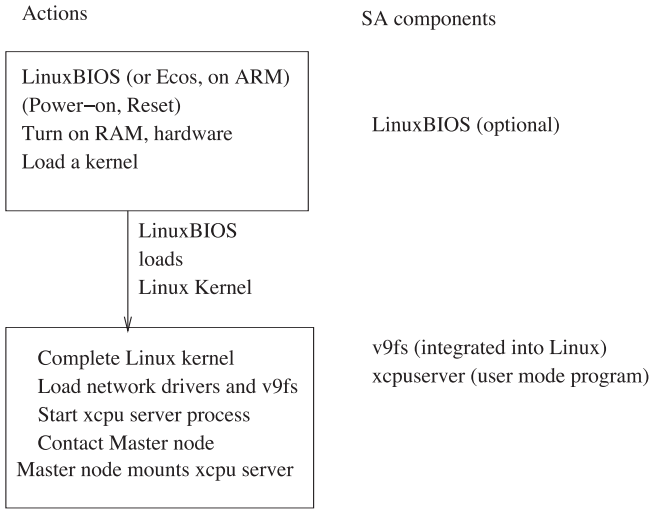


Figure 15. New Boot Sequence

simpler, well, it is. We no longer have a "boot kernel" and a "working kernel". The first kernel we boot will, in most cases, be sufficient. Experience shows that we change kernels on our clusters only every 3–6 months or so. There is no need to boot a new kernel each time. Because the 9p protocol and the xcpu service don't change, and the Master node kernel versions are not tightly tied together, we can separate the version requirements of the Master node and the worker node. We could not make this kind of separation with bproc.

The result is that we can weld the StrongARM boards and the Pentium front end (Master) into one tightly coupled cluster. In fact, we can easily mix 32- and 64-bit systems with xcpu. We can get the effect of a bproc cluster, with more modern kernel technology. Figure 16 shows how we are changing Clustermatic components for this new technology.

Conclusion

In this article, we showed how we built the Ultimate Linux Lunchbox, a 16-node cluster with integral Ethernet switch, in a small toolbox. The cluster is built of hardy PC/104 nodes and can easily survive a drop-kick test and possibly even an airport inspection. The system has only three connectors: one Ethernet, one AC plug and one battery connection.

We also introduced the new Clustermatic software, based around the Plan 9-inspired 9p filesystem, now available in 2.6.13. The new software reduces Clustermatic complexity, and the number of kernel modifications are reduced to zero.

Although there was not room to describe this new software in this article, you can watch for its appearance at clustermatic.org; or, alternatively, come see us at SC 2005 in November, where we will have a mixed G5/PowerPC/StrongARM/Pentium cluster running, demonstrating both the new software and the Ultimate Linux Lunchbox.

This research was funded in part by the Mathematical Information and Computer Sciences (MICS) Program of the DOE Office of Science and the Los Alamos Computer Science Institute (ASCI Institutes). Los Alamos National Laboratory is operated by the University of California for the National Nuclear Security Administration of the United States Department of Energy under contract W-7405-ENG-36. Los Alamos, NM 87545 LANL LA-UR-05-6053.

Resources for this article: www.linuxjournal.com/article/8533.

Ron Minnich is the team leader of the Cluster Research Team at Los Alamos National Laboratory. He has worked in cluster computing for longer than he would like to think about.

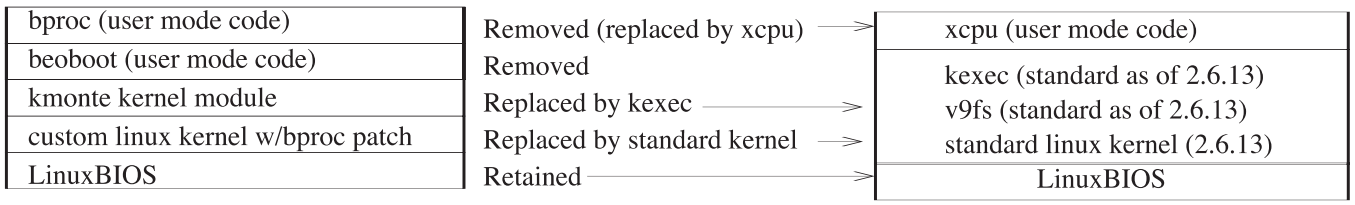
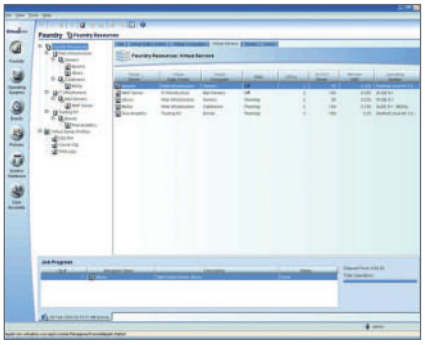


Figure 16. Clustermatic Component Changes

Virtual Iron (VFe)



Virtual Iron announced the general availability of its platform, VFe, which allows data centers to create virtual computing platforms that combine virtualization, clustering and provisioning technologies with policy-based system management in an integrated system. Virtual Iron works by seeing available hardware, disk I/O and network I/O devices as resources that can be allocated dynamically based on demand. VFe allows up to ten operating systems to run concurrently on a physical processor, a single operating system to span 16 processors or any combination in between, all sharing the same physical resources. These resources then can be provisioned automatically based on policies, thereby reducing latency and manual intervention. The VFe platform includes data center management capabilities that allow users to apply policy-based management toward provisioning and managing third-party virtual servers, including Xen. To this end, the Xen virtual machine monitor management module is included as a standard part of the Virtual Iron platform.

CONTACT Virtual Iron Software, Inc., 43 Nagog Park, Acton, MA 01720, 978-849-1200, info@virtualiron.com, www.virtualiron.com.

Scalix Community Edition

Scalix Corporation released Scalix Community Edition, a free, unlimited-use version of its e-mail and calendaring software. Community Edition includes a full version of Scalix's server and Scalix Web Access (SWA), a cross-browser, cross-platform Web client with integrated personal calendaring and address book capabilities. SWA works with IE, Mozilla or Firefox on Windows, Linux, Macintosh and UNIX desktops. Community Edition offers support for POP/IMAP e-mail clients, a GUI-installation wizard and Web-based administration console, a scripting environment as well as command-line access, complete documentation and community support through the Scalix Community Forum. Fee-based technical support is available from Scalix as well. Community Edition also comes with five free Scalix Enterprise Edition user licenses and is fully compatible with Enterprise Edition.

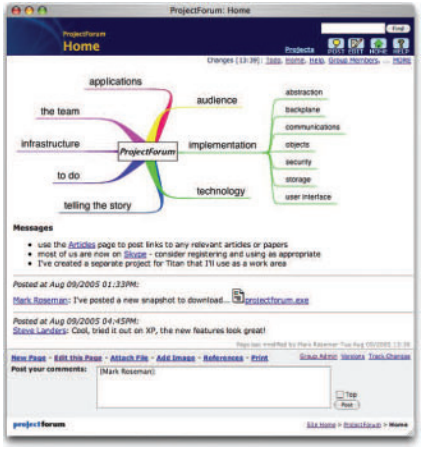
CONTACT Scalix Corporation, 1400 Fashion Island Boulevard, Suite 602, San Mateo, CA 94404, 650-931-9400, www.scalix.com.

Equilibrium MediaRich Server for Linux

Equilibrium MediaRich Server for Linux is server-based media templating software that automates image production and enables the dynamic delivery of digital media assets to the Web, mobile devices and print. For online retailers, MediaRich provides dynamic zoom and pan templates that generate product image derivatives from a single source image on the fly. MediaRich generates and displays crisp text and graphic elements onto an image or multiple images for dynamic product merchandising and text-graphics localization. Pre-press production houses can automate large amounts of CMYK conversions, dpi adjustments and scaling requests. MediaRich supports many popular file formats as well as loading, saving and merging IPTC, Exif and XMP metadata.

CONTACT Equilibrium, 3 Harbor Drive, Suite 100, Sausalito, CA 94965, www.equilibrium.com.

ProjectForum 4.5



ProjectForum offers shared Web-based work spaces that provide a central place to collect, manage and discuss topics and work relating to a shared project. ProjectForum offers full version control, group project support, multiple authentication options, image and file management, page templates, SSL, full branding support and multiple forums for meeting. ProjectForum is available either as a fully managed hosted service or as software that can be downloaded and run in-house. Versions are available for Windows, Mac OS X, Linux and other UNIXes, while users of the software need only a standard Web browser. New features for version 4.5 include RSS feeds for every page in the forum, which complements the existing per-forum RSS feeds. RSS feeds also can be directly included in ProjectForum pages. Also new is the option to allow forum changes to be broadcast by e-mail.

CONTACT CourseForum Technologies, 851 Birchmount Drive, Waterloo, Ontario, Canada N2V 2R7, info@courseforum.com, www.projectforum.com.

Intrepid M

Levanta recently introduced the Intrepid M management appliance, which combines Levanta's management and provisioning software with shared storage, preconfigured templates and open-source software in a single plug-and-play device. Intrepid M plugs in to the network and allows administrators to provision servers or workstations quickly with full Linux stacks and applications; to deploy software and patches simply and quickly to multiple machines without lengthy installation steps or file copying; to migrate all software and the entire OS from one piece of hardware to another at will; to allocate resources spontaneously using commodity components, with no vendor lock-in; and to track all changes made to a machine by any means. The appliance offers a full-color status LCD, 1.4TB of storage, hot-swap RAID-5 storage, six SATA hard drives in quick-change drive bays, shared storage functionality, dual hot-swap redundant power supplies, hot-swap fans and two 10/100/1000 Ethernet NICs.

CONTACT Levanta, Inc., 650 Townsend Street, Suite 225, San Francisco, California 94103, www.levanta.com.

Please send information about releases of Linux-related products to Heather Mead at newproducts@ssc.com or New Products c/o *Linux Journal*, PO Box 55549, Seattle, WA 98155-0549. Submissions are edited for length and content.

LISA '05

19TH LARGE INSTALLATION SYSTEM ADMINISTRATION CONFERENCE

DEC. 4–9, 2005

SAN DIEGO, CA

www.usenix.org/lisa05/lja

FIND THE MISSING PIECES
TO YOUR TOUGHEST PUZZLES.

6 DAYS OF TRAINING
by industry experts, including:

- Rik Farrow on Hands-on Linux Security
- Don Bailey on 802.11 Wireless Network Penetration Testing
- Richard Bejtlich on Network Incident Response
- Jacob Farmer on Disk-to-Disk Backup

LISA '05 offers the most in-depth, real-world system administration training available!

3-DAY TECHNICAL PROGRAM

KEYNOTE: Qi Lu, Vice President of Engineering, Yahoo! Inc., on "Scaling Search Beyond the Public Web"

20+ INVITED TALKS INCLUDING:

- Matt Blaze, University of Pennsylvania: "Picking Locks with Cryptology"
- Kevin Bankston, EFF: "How Sysadmins Can Protect Free Speech and Privacy on the Electronic Frontier"

NEW! Hit the Ground Running Track

Refereed Papers, Guru Is In Sessions, Vendor Exhibition, BoFs, WiPs, and more!

Real World
System Administration

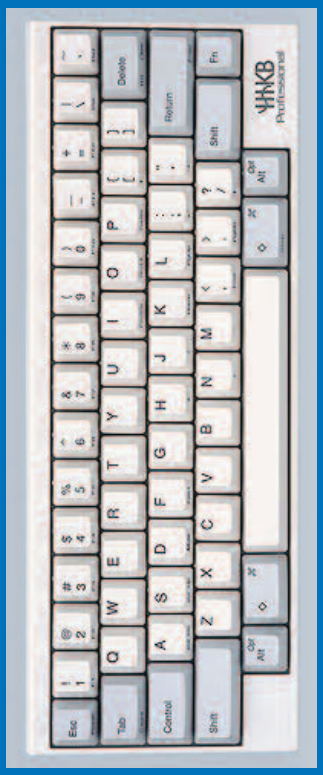
SPONSORED BY

USENIX THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

SAGE The People Who Make IT Work

Register by November 18 and save!
www.usenix.org/lisa05/lja

Happy Hacking Keyboard Professional Review



PRODUCT INFORMATION

Vendor: PFU

URL: www.pfu.fujitsu.com/en/hhkeyboard/index.html

Price: HHKB Pro, $269 US; Happy Hacking Keyboard Lite 2, $69 US

THE GOOD

- Excellent keyboard feel and large keys provide smooth typing.
- Super compact.
- DIP switches provide multiple configuration options.

THE BAD

- Lack of dedicated keys means common operations need Fn-<key> combinations.

REVIEWED BY STEVE R. HASTINGS

The Happy Hacking Keyboard Professional (HHKB Pro) is a compact USB keyboard with an excellent feel, some intriguing features and a hefty price tag. It's made by PFU, part of the Fujitsu Corporation.

The most important thing about any keyboard is this: how well does it work for typing? Although the HHKB Pro has fewer keys than a normal keyboard has, the keys it does have are full size and are mostly where your fingers expect to find them. The keys have an excellent feel too, clicking gently when you type but not clacking loudly. I find that I can touch-type at full speed with this keyboard. In fact, I wish my full-size keyboard had keys this nice.

Earlier keyboards in the Happy Hacking keyboard line have membrane keys with rubber caps. The HHKB Pro, however, has a circular cone spring system. According to the Happy Hacking Web site, this system provides softer keystrokes and a longer keyboard life.

As with many laptop keyboards, the HHKB Pro has a Fn key (for Function) that can combine with other keys to make a keystroke that is not otherwise available. The HHKB Pro, with only 60 keys, doesn't have dedicated function keys; but you can get an F1 keystroke with Fn-1, F12 with Fn-= and so on. This keyboard doesn't even have dedicated arrow keys; up, down, left and right are, respectively, Fn-[, Fn-/, Fn-; and Fn-'.

The HHKB Pro has the Esc and Ctrl keys in the traditional places. The most common keyboard layout today is the 104-key layout, based on the 101-key layout that IBM introduced in 1986. 104-key keyboards have a Caps Lock key to the left of the ASDF home row of keys and have two Ctrl keys, on opposite sides of the keyboard. The HHKB Pro has a single Ctrl key instead of a Caps Lock key; Fn-Tab serves as the Caps Lock key. A 104-key layout keyboard has the Esc key widely separated from the rest of the keyboard, at the extreme upper left. The HHKB Pro places the Esc key immediately above the Tab key and to the left of the 1 key.

The HHKB Pro also has a set of DIP switches that can be used to customize the way the keyboard works. These are located behind a small cover on the back side of the keyboard.

The SW1 and SW2 DIP switches select among three modes: default or HHK mode, HHK Lite mode and Macintosh mode. The only difference between the default mode and HHK Lite mode is some additional key combinations become available in HHK Lite mode. For example, you cannot use the Fn-Tab combination for Caps Lock in default mode; HHK Lite mode enables it. I can see no reason why anyone would prefer the default mode to the HHK Lite mode, and I recommend you use HHK Lite mode if you use an HHKB Pro keyboard.

Immediately above the Return key is a key labeled Delete. The SW3 DIP switch, when on, changes this to make it work as a Backspace key. Whether or not SW3 is on, Fn-Delete always works as a Backspace key, and Fn-` always works as a Delete key.

Two Alt keys are present, to the left and right of the spacebar. There also are two keys labeled with diamonds; these can be used as the logo keys from a 104-key keyboard. The SW5 DIP switches can be used to swap the functions of Alt and diamond keys. If you frequently use Alt keys—for example, if you use Emacs and Alt is your meta key—you probably will prefer this. The diamond keys are bigger and easier to press.

The SW4 DIP switch controls whether the left diamond key works as a logo key or as a second Fn key. If SW5 is enabled, making the left Alt key work as a logo key, the left Alt key becomes the second Fn key.

The last DIP switch, SW6, controls

whether the keyboard goes to sleep when the computer does. Fn-Esc makes a keystroke called Power that can be used to control a PC's sleep mode. I didn't test this feature, though.

The HHKB Pro also has a few multimedia key combinations: volume down, volume up, mute and eject are, respectively, Fn-A, Fn-S, Fn-D and Fn-F. However, these are supported only when the HHKB Pro is in Macintosh mode. In the other two modes, holding down the Fn key does not change the keystrokes these keys make. If you want the multimedia keys to work, you could try setting the keyboard to Macintosh mode, and in your desktop environment's keyboard preferences set your keyboard type to Macintosh. I tried this and it worked for me. The HHKB Pro even generated the same multimedia keystrokes as my other keyboard, so both keyboards could be used to adjust the volume of my speakers.

Daily Use

When you first use the HHKB Pro, the first thing you notice is the lack of dedicated arrow keys. Anytime you need an arrow key, you have to press a Fn-<key> combination. What's worse is the arrow keys are not immediately obvious; you need to take your hand off the keyboard, look at it, press the combination and then put your hand back for more typing. If you use the HHKB Pro long enough, though, you probably can learn to press the Fn combinations for the arrow keys without looking. But this simply is not as convenient as having dedicated arrow keys.

However, Linux builds on a long UNIX tradition, and UNIX was developed on many different terminals that had many different keyboards. As a result, both Emacs and vi are designed to be usable with only standard ASCII keys. In my college days, I used to write Pascal programs on ADM3A terminals that didn't even have a dedicated Backspace key; you had to press Ctrl-H when you wanted a backspace. If you can learn to use Emacs or vi keystrokes, you can get by fine without using arrow keys, and there are many programs in Linux that use these keystrokes.

I configured my bash shell to use vi keystrokes for command-line editing and quickly became comfortable with it. See the sidebar for notes on using vi or Emacs mode in the shell.

Actually, I'm kicking myself now that I didn't set my shell for vi mode long ago. Because I'm expert with vi, I can edit command lines much better in vi mode, without taking my hands from the home row keys. If you have spent time mastering either vi or Emacs, try them in the shell!

If you have a small laptop or a tablet PC, the HHKB Pro makes an excellent carry-along keyboard. If you pack the HHKB Pro into a bag, I recommend you fully unplug the USB cable. The HHKB Pro's cable is a standard USB cable with an A connector on one end and a mini-B connector on the other.

Price

Unfortunately, the HHKB Pro is rather expensive. The Web site lists the regular price as $269. I searched the Web and was able to find the HHKB Pro for as little as $249, which is still much more than I am willing to pay for a keyboard.

The Happy Hacking Keyboard Lite 2 model, in USB or in PS/2, is available for a regular price of $69.

Conclusion

If it were not for the price, I wholeheartedly would recommend the HHKB Pro. It's everything you could ask for in such a compact keyboard. Of course I'm using it to type this article, and I'm enjoying the smooth feel of the keys. It is nicer than my usual keyboard, but alas it costs more than six times as much.

vi or Emacs Mode in the Shell

By default, the bash shell already should be in Emacs mode. You can use Ctrl-P and Ctrl-N instead of the up and down arrow keys to scroll through the command history. You can use other Emacs keystrokes to edit command lines. To make bash use vi keys, edit a file called .inputrc in your home directory and insert these lines:

```
set editing-mode vi
set keymap vi-insert
```

Then, start up a fresh bash shell and try it out. If you press the Esc key, you enable editing mode, where hjkl keys work as left, down, up and right arrow keys. Other vi commands, including ^ for jump to start of line and $ for jump to end of line, also work.

If your system defaults to vi and you want Emacs mode, insert these lines in your .inputrc file:

```
set editing-mode emacs
set keymap emacs
```

These features come courtesy of the GNU Readline Library. For more information on Readline and its features, run `man 3 readline` or check the Readline Web site (cnswww.cns.cwru.edu/php/chet/readline/rltop.html).

Not only bash but any program that uses the GNU Readline Library can be customized by making changes to your .inputrc file. For example, the GDB debugger uses Readline.

If you use the tcsh shell, again Emacs mode is available by default. You can set vi editing mode by placing this line in your .tcshrc file:

```
bindkey -v
```

Read the tcsh man page for more information.

If you use the zsh shell, all you have to do is set the EDITOR or VISUAL environment variable to your favorite editor. If your choice contains the string "vi", zsh sets vi mode; otherwise it defaults to Emacs mode. You also directly can manage the editing mode with zsh's `bindkey` command. See the zsh man page for more information.

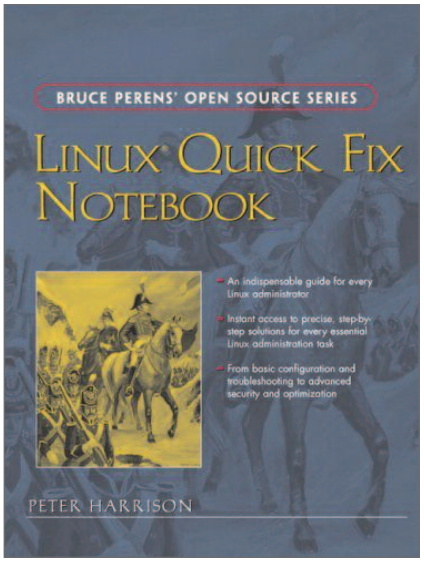
Even the Midnight Commander (mc) file manager supports Emacs-style command-line editing as well as Emacs-like and vi-like key bindings in its file viewer.

Steve R. Hastings first used UNIX on actual paper teletypes. He enjoys bicycling with his wife, listening to music, petting his cat and making his Linux computers do new things.

Linux Quick Fix Notebook

by Peter Harrison

Prentice Hall PTR, 2005 | ISBN: 0131861506 | $39.99 US



Peter Harrison's new *Linux Quick Fix Notebook* is the kind of book that all Linux professionals should have handy for times when they need immediate results. Harrison doesn't waste time explaining theory or concepts. Instead, he works off the assumption that if you need to build a DNS server, you already know what DNS is and how it operates.

The book covers topics ranging from configuring the boot process to building DHCP servers. Within each topic, Harrison jumps directly to what you need to do to get the application running right away. Although the directions and configurations are not always sophisticated, they are fully functional and technically correct. This approach of providing a starting point for a service and leaving the rest to the reader to configure is probably for the best, as each user has individual requirements.

The layout of the book is almost that of a FAQ. Each topic is covered within a few pages. Of all the computer books I own, this is the most direct and to the point when it comes to Linux configurations.

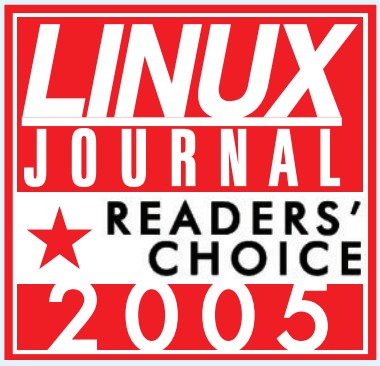
Harrison's writing style is clear and easy to understand. He manages to provide adequate detail on each step of a procedure without going overboard on details. *Linux Quick Fix Notebook* is suitable for all levels of Linux users. Novice Linux users will appreciate the ability to dive right in and begin setting up services. On the other hand, this book makes an excellent quick reference for the experienced Linux administrator who needs a little help remembering the proper steps to configure a particular service.

All in all, *Linux Quick Fix Notebook* has become one of my new favorite books on Linux administration. I've used it on several occasions at work, and it has yet to let me down.

—BRIAN WARSHAWSKY

2005 *Linux Journal* Readers' Choice Awards

Some of your old favorites dropped off the Readers' Choice results this year. Has the Linux scene changed for good? BY *LJ* STAFF



We overhauled the voting process for this year's Readers' Choice Awards in the hope of creating a fairer system that voters were involved in every step of the way. As such, we accepted nominations from readers in 31 categories and then held two rounds of voting to get this final list of your favorites.

Some readers were surprised by the list of candidates that made it to the final round. For instance, the big-name distributions, such as Debian, Red Hat and SUSE, were nowhere to be found. Although these absences may seem odd, we call these the Readers' Choice awards because they are exactly that—these are the products and tools our readers are using and loving this year.

Here we present the top two vote-getters in each category. In categories where vote totals were particularly close, we have listed the top three finishers.

FAVORITE AUDIO TOOL

1. XMMS
2. amaroK

For the sixth year in a row, XMMS is the first-place finisher in the audio tool category. So you know XMMS plays MP3, OGG, WAV and CD audio file formats. You also probably know that it supports a whole bunch of third-party input plugins. But do you know about its equalizer and playlist capabilities? Do you know about its advanced plugins for file I/O, special effects and visualization? If not, you must have missed Dave Phillips' "Getting the Most from XMMS with Plugins" (see the on-line Resources for links to articles), which covered some of

FAVORITE BACKUP SYSTEM

1. Amanda
2. Bacula

We split backups into two categories this year to differentiate between simple tools that can back up a single system (see Favorite Backup Utility below) and more complex programs administered centrally to back up multiple machines. Although not as flashy as some other backup systems, Amanda (advanced Maryland automatic disk archiver) offers "a reliable platform for many Linux and UNIX users who are comfortable with a command-line interface", according to Phil Moses, who wrote about it for us in "Open-Source Backups Using Amanda". Apparently, many of our readers

FAVORITE BACKUP UTILITY

1. tar
2. rsync

Even though many backup tools are available from vendors, we know that our readers often prefer to stick with the basics. Thus, your favorite backup utilities, tar and rsync, are basic command-line tools that were separated by less than a hundred votes in this year's competition. You can do a lot with tar, from building basic single-file archives to creating multivolume backups. Sometimes, though, the most tried-and-true tools are the ones we take for granted, so to learn more about what you can do with tar and rsync, take a look at these past *LJ* articles: "The Skinny on Backups and Data Recover, Part 3", "LVM and Removable IDE Drives Backup System" and "rsync, Part I and Part II".

FAVORITE DATABASE

1. MySQL
2. PostgreSQL

Celebrating its tenth anniversary this year, MySQL once again scores the top place in this year's voting. Besides offering more features than ever, MySQL also is being included in more big-name vendor products, thanks to the ever-increasing popularity of LAMP applications. In "An Open Letter to the Community from MySQL Founders David Axmark & Michael 'Monty' Widenius", the founders offered these impressive stats: "over 100 million copies of MySQL have been distributed" through the Web site and operating system distributions; approximately 40,000 new downloads every day; more than 1,500 projects on SourceForge.net are using MySQL; and current users include Craigslist, Slashdot, Wikipedia, Bugzilla, Technorati and the Human

FAVORITE DESKTOP ENVIRONMENT

1. KDE
2. GNOME

The dot.kde.org site carried a link to the Readers' Choice voting page this year—did the extra promotion to KDE fans make the difference? As detailed below, this year's favorite distribution is GNOME-based while the favorite language is the base language for KDE. People seem to be using the

FAVORITE DEVELOPMENT TOOL

1. GCC
2. KDevelop
3. Eclipse

Wait a second before skipping to the next category—this result isn't as boring as you might think. Yes, GCC won again, but it's a whole new GCC world out there. Earlier this year, Tom Tromey wrote that GCC "has undergone many changes in the last few years. One change in particular, the merging of the tree-ssa branch, has made it much simpler to write a new GCC front end."

FAVORITE DESKTOP WORKSTATION

1. Dell
2. Apple
3. Monarch

People like Dell's boxes, but it's still confusing to buy anything but a top-of-the-line workstation from them if you want to run Linux. And even then, according to the Dell Linux Engineering page, "all Dell N-Series Precision Workstation desktops are available and supported with Red Hat Linux. For help running other Linux distributions on your Workstation, you might consider posting to or viewing the linux-precision mailing list." Still, they sure do look nice—they'll even

Dell's XPS Gen 5 workstation (or should we say gamestation) comes with your choice of seven colors for the tower's chassis light—ruby, emerald, sapphire, amber, topaz, amethyst and diamond.

FAVORITE DISTRIBUTION

1. Ubuntu
2. CentOS
3. Fedora Core

Ubuntu: Linux for Human Beings. Linux users around the world have surged to Ubuntu this year. You have to feel good about using a distribution whose name means "humanity to others".

Judging by the comments posted on the *LJ* Web site during the voting process, a lot of voters were "shocked" and "flabbergasted" that the brand-new Ubuntu made it to the final round, while Red Hat, Debian, SUSE and other big names were absent. Maybe it's a passing phase of Ubuntu mania, but as Steve Hastings wrote in his *LJ* review, "Ubuntu Linux is an excellent choice for anyone who wants to run Linux on a desktop system. It's easy to install and to administer. Everyone from beginners to experts can use and

FAVORITE EMBEDDED DISTRIBUTION

1. Qtopia
2. LFS
3. OpenZaurus

Nitpickers might say that Qtopia isn't a distribution because it doesn't include the kernel, but it's a full-featured embedded development environment. Qtopia is built on Qt/Embedded, the C++ GUI and platform development tool for Linux-based embedded development. You get all the source code and can do whatever customization you want. Everyone from Samsung to Motorola and Phillips is using Qtopia for PDAs, cell phones and other cool new gadgets.

FAVORITE GRAPHICS PROGRAM

1. The GIMP
2. Inkscape

Everyone knows The GIMP rules this category and has for practically the past decade. But wow, there are a lot of votes for Inkscape this year. Our editors selected it for an Editors' Choice Award earlier this year as well. So maybe it's time the rest of you take a look at Inkscape, especially if you're concerned about making your graphics look good at a variety of screen sizes by using a vector format.

FAVORITE INSTANT MESSAGING CLIENT

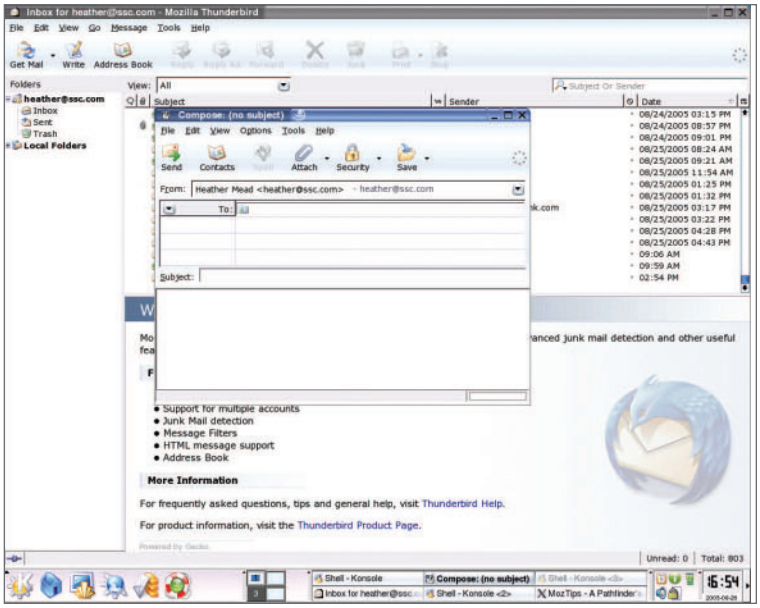
1. Gaim
2. Kopete

Gaim integrates with both GNOME and KDE, thereby setting a desktop application paradigm for the future—an application that plays standards, not desktop wars. Besides that, the selection of smiley-face icons is great for adding a touch of sarcasm with a well-placed smiley-face wearing a

FAVORITE E-MAIL CLIENT

1. Mozilla Thunderbird
2. Evolution

In the early days of the Readers' Choice Awards, the top finishers in this category always were mutt, pine and other text-based programs. The last couple of years, though, the majority of readers—at least the voting ones—have given up the basics for one of the smooth new GUI-based clients. And Thunderbird seems to be responsible for a lot of these conversions.



Thunderbird's interface will look familiar to users of other GUI-based and Web e-mail programs. But it's better.

FAVORITE LINUX BOOK

1. *Running Linux*, 4th Edition
2. *Gentoo Handbook*
3. *A Quarter Century of UNIX*

Here's a fun project for a cold fall evening: compare the table of contents in the first edition of *Running Linux* to the one in the fourth edition, and see how much more you can do now and how much less time you need to spend tweaking low-level stuff. Much space in the first edition, for example, was used to explain things such as kermit and elm—it even brought up troff (shudder). The fourth edition, however, talks about KDE and GNOME, not to mention the final section on Web development with LAMP.

FAVORITE LINUX TRAINING

1. IBM
2. lintraining.com
3. Novell CLP

Yes, we know training is important and the horrors of what can happen when a poorly trained sysadmin is set loose in a server room. But we don't know why IBM won; in the ads, that kid who looks like Eminem's little brother seems pretty bored. Maybe a Mick Bauer live security

FAVORITE LINUX GAME

1. *Frozen Bubble*
2. *Unreal Tournament 2004*

We know it's not your fault that you keep playing *Frozen Bubble*. We can't stand the pitiful little noise the penguin makes when we lose

You guys are suckers for cute animation.

intrusion demo would hold his attention a little better.

FAVORITE LINUX WEB SITE

1. Slashdot.org
2. Distrowatch.com
3. LinuxJournal.com
4. LWN.net

Readers always will have a special place in their hearts for a Web site that, on one page worth of headlines, offers updates on PSP 2.0, marketing strategies for *Firefly* (Joss Whedon's canceled TV show that made it to the big screen), Google's new IM client and house-sitting robots in Japan.

FAVORITE *LJ* COLUMN

1. Cooking with Linux
2. Paranoid Penguin
3. At the Forge

Oh, François, the readers, they love you still. *Un affair de cœur, c'est très beau, non?* 2005 was pretty significant for the second- and third-place finishers, as regular Paranoid Penguin columnist Mick Bauer turned it over to a rotating author list and Reuven Lerner celebrated his 100th At the Forge.

FAVORITE NETWORK OR SERVER APPLIANCE

1. Astaro Security Gateway
2. Cyclades AlterPath ACS
3. thinklogical Sentinel32

Besides the fact that Astaro works well, our readers appreciate that the Astaro box isn't just a "firewall" in the ordinary packet-filtering sense. It also comes with antispam, antivirus, intrusion detection and a Web proxy—features that would be expensive add-ons for

FAVORITE OFFICE PROGRAM

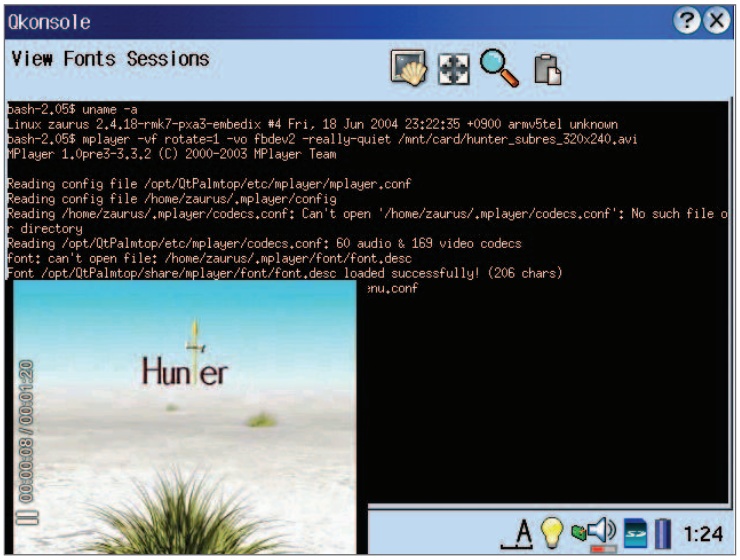
1. OpenOffice.org
2. KDE Kontact
3. LaTeX

Garnering over a thousand votes more than the second-place finisher, OOo has built a strong following in the Linux and Open Source community, thanks to its compatability and usability—not to mention our monthly Web column by Bruce Byfield, OOo Off the Wall. Check

FAVORITE MEDIA PLAYER

1. MPlayer
2. Xine
3. Kaffeine

We know our US readers aren't actually running MPlayer because of the software patent situation,
but it's nice to see what people in countries with more sensible patent systems can do.



For our readers living outside the US, MPlayer really does run on anything—even your Zaurus.

FAVORITE PORTABLE WORKSTATION

1. IBM ThinkPad
2. Apple PowerBook
3. Dell Latitude

We're all in suspense about what the new ThinkPad company, Lenovo, is going to do Linux-wise. Although ThinkPads are a common sight at Linux conferences, every one has to be tweaked or ordered through a company, such as EmperorLinux, that does a custom install for you. Do a Google search for ThinkPad, and right after thinkpad.com comes a Linux site, and six of the top ten results are Linux-related. HP's Linux laptop mysteriously vanished from the company's Web site without a trace, but maybe Lenovo will listen to their Linux-using fans instead of falling prey to mysterious marketing conspiracies.

out his past columns on the LinuxJournal.com site for

FAVORITE PROCESSOR ARCHITECTURE

1. x86-64
2. POWER
3. IA-64

Readers were waiting for it, they needed it, coveted it, and once the 64-bit next generation of x86 became available, first from AMD, then from Intel, things just haven't been the same here. It's not even close anymore. We shouldn't talk, though; we've featured x86-64's 64-bit processing power in the last three Ultimate Linux Box
articles. More power is good.

great documentation on using fields, creating templates

FAVORITE SERVER

1. HP ProLiant
2. Monarch Empro Custom Rack Server
3. Unisys ES7000 Family

Last year, the HP ProLiant BL20p G2 won the Editors' Choice Award for Server Hardware. Now the readers are singing the ProLiant's praises.

A note to HP: please take this first-place win here, where second-place votes were less than half of what you received, as proof that we like your boxes, so you can cut out the pointless marketing poo-flinging at

FAVORITE PROGRAMMING LANGUAGE

1. C++
2. Python
3. PHP

Back in early 2003, Don Marti asked the following question regarding C++: "Now that we have GCC 3.2.x...and an increasing collection of interesting free software using C++, is it time to take a second look at this perhaps unfairly maligned language?" He didn't expect that a mere two years later, C++ would win here. A lot of that has to be the rapid growth of Linux to include the world's C++ coders—

FAVORITE PROGRAMMING BEVERAGE

1. Coffee
2. Tea
3. Water

Mmmm, coffee, that sounds great. Can you get me a triple-shot Americano, please? #c0ffee is even a valid hex color to try on your Web site.

FAVORITE SYSTEM ADMINISTRATION TOOL

1. OpenSSH
2. Webmin
3. YaST

Looking back at past *LJ* articles on OpenSSH, we found titles such as "Doing It All with OpenSSH 1", "Doing It All with OpenSSH, Part 2" and "The 101 Uses of OpenSSH". So combining that with its big win here, it looks like you can do a whole lot with

and organizing work flows. And don't miss the reader comments, where questions are asked, answered, debated, clarified and argued some more.

FAVORITE TEXT EDITOR

1. Vim
2. Kate
3. Emacs

What, use something besides Vim? What do you have against *orphans*? Don't you know that "Vim is Charityware. You can use and copy it as much as you like, but you are encouraged to make a donation for needy children in Uganda. Please visit the ICCF Web site"; URLs available in the on-line Resources.

FAVORITE VERSION CONTROL SYSTEM

1. Subversion
2. CVS
3. GNU Arch

The LinuxJournal.com editor would like to point out that the site published "Setting Up Subversion for One or Multiple Projects" back in 2004. Print was snoozing and covering Arch while the Web site was doing the Subversion stuff that was a hit with readers used to CVS-style development. Yay Web!

FAVORITE VIRTUALIZATION SOLUTION

1. VMware
2. Xen

Virtualization is becoming bigger news these days. VMware lets you run an unmodified guest OS and has been around for longer than the rest, so one or both of these factors matters to voters. If you're new to VMware, we suggest you start by reading "VMware 5 Workstation Edition Reviewed" to get an overview of what it can do. Meanwhile, Xen is a solution that's easy to get started with for Linux-on-Linux setups.

FAVORITE WEB HOSTING SERVICE

1. Rackspace Managed Hosting
2. 1&1 Internet

Rackspace won here, although this category didn't collect a ton of votes. It did, however, manage to start a comment debate about a host's responsibilities when its clients are the subject of secret

FAVORITE WEB BROWSER

1. Mozilla Firefox
2. Konqueror

Firefox, so good everyone from our editors to the government recommends you use it. For more under-the-hood stuff, check out Nigel McFarlane's article "Fixing Web Sites with GreaseMonkey" from the October 2005 issue.

When everyone, including the United States Computer Emergency Readiness Team, recommends users switch to your browser, you have to know you're going to win the Readers' Choice Favorite Browser award.

Statement of Ownership, Management, and Circulation

1. Publication Title: *Linux Journal*
2. Publication Number: 1075-3583
3. Filing Date: October 1, 2005
4. Issue Frequency: Monthly
5. Number of Issues Published Annually: 12
6. Annual Subscription Price: $25
7. Complete Mailing Address of Known Office of Publication: PO Box 55549. Seattle, WA 98155-0549
   Contact Person: Jill Franklin 206-782-7733 x112
8. Complete Mailing Address of Headquarters of General Business Office of Publisher: PO Box 55549 Seattle, WA 98155-0549
9. Full Names and Complete Addresses of Publisher, Editor, and Managing Editor:
   Publisher: Phil Hughes PO Box 55549-0549 Seattle, WA 98155
   Editor: Don Marti PO Box 55549-0549 Seattle, WA 98155
   Managing Editor: Jill Franklin PO Box 55549-0549 Seattle, WA 98155-0549
10. Owner(s): Phil Hughes PO Box 55549 Seattle, WA 98155-0549 Joyce Searls PO Box 55549 Seattle, WA 98155-0549 Adele Soffa PO Box 55549 Seattle, WA 98155-0549
11. Known Bondholders, Mortagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortages, or Other Securities: None
12. Tax Status: Has not Changed During Preceding 12 Months

13. Publication Title: *Linux Journal* — 14. Issue Date: October 2005

15. Extent and Nature of Circulation	Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies: (Net press run)	73,319	67,849
b. Paid and/or Requsted Circulation		
(1) Paid/Requested Outside-County Mail Subscriptions on Form 3541.	24,876	23,844
(2) Paid In-County Subscriptions Stated on Form 3541	0	0
(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution	23,019	38,790
c. Total Paid and/or Requested Circulation	47,895	62,634
d. Free Distribution Outside the Mail		
(1) Outside-County as Stated on Form 3541	1,381	1,112
(2) In-County as Stated on Form 3541	0	0
(3) Other Classes Mailed Through the USPS	2	2
e. Free Distribution Outside the Mail	3,274	3,289
f. Total Free Distribution	4,657	4,403
g. Total Distribution	52,552	67,037
h. Copies Not Distributed	20,767	812
i. Total	73,319	67,849
j. Percent Paid and/or Requested Circulation	91%	93%

PS Form 3526

Echo and Soft VoIP PBX Systems

The new world of Internet telephony is facing one of the same challenges that early long-distance calling did. Here's one of the techniques for doing a high-quality call over VoIP.

BY DAVID MANDELSTAM

Most of us have experienced telephone calls with disturbing echoes on the line. Low echo volumes together with discernible delay can make a line completely unusable, with the call being terminated after the exchange of a few halting sentences. Traditionally, problems with echo have been experienced on long-distance or international calls, particularly those involving satellite connections.

For many people new to software-based VoIP telephony systems, such as Asterisk, the phenomenon of voice echo comes as an unpleasant surprise. This is true even for those who come to the business after working with traditional PBX systems or proprietary VoIP equipment. Suddenly echo is a problem on local calls, and the traditionally troublesome long-distance and satellite calls are completely echo-free.

In this article, we discuss the origins of echo and how it manifests itself in the VoIP world with particular reference to Asterisk and other software-based telephony systems.

Where Does Echo Come from and Why Is It a Problem?

Echo in telephony systems is caused by two main phenomena: the first is electrical echo due to imperfect impedance matching, and the second is acoustic echo due to microphone pickup of audio output. Both these sources produce similar effects and have to be treated similarly. The major difference is electrical echo is a property of the line connection and remains mostly constant throughout the call, while acoustic echo varies in strength and delay depending on the changing acoustic environment of the echo source. For instance, on a hands-free cell-phone call, the echo characteristics change as the speaker moves around.

Electrical signals of all types always are reflected at line terminations, except when the load at the line end exactly matches the impedance rating of the line itself. In fact, the meaning of, say, "75-ohm cabling" is precisely that in order to have no signal reflections, the cable must be terminated by a 75-ohm load. Line impedance is a property of the cable that is affected only by the cable geometry. As no cables are geometrically perfect over their length and no load impedance is perfectly accurate, there always is some reflection at a line termination.

Where digital signals are concerned, as long as the reflections are a small enough fraction of the data transmission, the reflections do not cause errors in reading the bit values. Thus, digital systems can tolerate considerable echo.

The human ear has quite different characteristics, however; it is an incredibly sensitive instrument. The softest sound that can be heard has an acoustic power about a hundred thousand billion times smaller than the power at the threshold of pain. As long as sounds vary by *only* about a factor of 100 or so, the ear hears a similar level of sound. So even what electrically looks like a small reflection can sound about the same volume as the original signal to the human ear.

And, the traditional telephone circuits are far from perfect. Two-wire circuits from analog lines terminate at devices called hybrids that convert the two-wire analog signal to four-wire signals before digitization. The loads at the hybrids vary quite widely, as does the impedance of the low-cost subscriber loop wiring. The result is almost every call that involves an analog telephone anywhere in the circuit has electrical reflections that can be interpreted by the ear as troublesome echoes.

If this is so, why is echo not a problem on every call? The answer is, if the echo is heard at the same time as the caller is speaking, it is heard as part of the side tone and goes unnoticed. Echo becomes noticeable only when there is a delay between speaking and hearing your voice echoed. This is why echo is a problem only for traditional telephony over long distances. The round-trip delay on a coast-to-coast US call is more than 30ms, which is enough for echo to cause irritation. Satellite delays are much longer still.

VoIP intrinsically has packetization, depacketization and processing delays built into its protocols. That is why, from the point of view of echo, every VoIP call is like a very long-distance call.

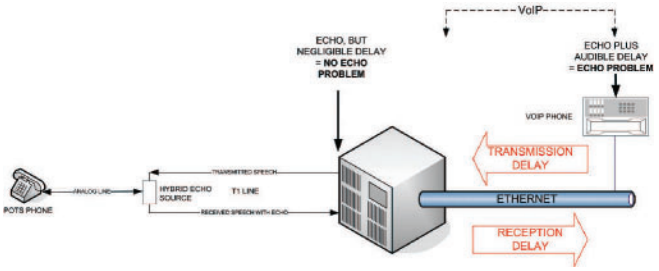


Figure 1. How VoIP and Analog Telephone Systems Interact to Cause Troublesome Echo

Figure 1 shows a typical VoIP scenario. The echo is heard on the VoIP phone: the caller on the analog line hears only a normal side tone, because there are no signal delays. Because delay is a necessary component of perceived echo, traditional PBXes that switch analog or T1/E1 traffic have no perceived echo problems, as their intrinsic end-to-end delay is low. It is

linux phone systems 101

lesson one

... choose the right team

The first step in any project is assembling the right team for the job. Let the experts at Telephonyware guide your Linux™ phone system project by helping you select the best hardware and software, and by providing the very best in service and support.

Take the guess work out of VoIP, choose a partner you can trust — Telephonyware.

lesson two

... get the right gear

For your VoIP project to be successful, you need the right gear! Let Telephonyware take the worry out of selecting the right hardware and software for the job.

We sell and support a full range of IP phones, analog and digital telephony cards, analog telephone adapters (ATAs), power over ethernet midspans and switches, and many more quality products. Our range is hand picked from the best manufacturers, and our helpful staff have used every product we sell.

lesson three

... put it all together

When it's time to turn plans into reality, Telephonyware is the right partner to take you from idea to completion. Our network of service partners, and excellent in–house support, give you the confidence you need, at a price you can afford.

Whether you're an experienced consultant deploying VoIP for your customers, a business replacing a phone system, or just looking for an IP phone or an ATA, Telephonyware will help you put it all together.

SANGOMA™

Introducing Sangoma's FXO/FXS Analog Cards

Sangoma has just launched a new FXO/FXS solution that takes care of the most demanding echo cancellation problems... and brings new levels of voice quality, value and serviceability to Asterisk™

Telephonyware sells, supports and recommends the full range of Sangoma analog and digital telephony cards.

Sangoma's AA series analog cards have the following benefits:

- They use the same PCI interface, architecture and digital path as Sangoma's T1/E1 cards, meaning no motherboard or compatibility issues and ultra-reliable interrupt handling.
- They have full line protection, making them legal to connect to the telephone network — this includes FCC Part 15, FCC Part 68 and CE certification with other certifications to follow.
- Sangoma's AA architecture supports up to 24 analog interfaces, both FXO and FXS, all operating through one FPGA and one PCI slot using one IRQ, using an external backplane card connector, avoiding the problems of multiple asynchronous DMA and interrupts that would occur with multiple PCI cards.

... visit www.telephonyware.com/sangoma for more info

TELEPHONY WARE

For online orders or more info, please visit us at **www.telephonyware.com/lj**

Call us on **(866) 864-2304** or write to **sales@telephonyware.com**

the packetization and processing delays inherent in VoIP that cause existing echo to become a problem.

What to Do about Echo

Those of you who have watched old black-and-white movies depicting long-distance conversations may remember the callers shouting into the mouthpieces in order for the other party to repeat what was said. The reason the callers had to shout was low receiver volume. The attenuated volume was the way echo was dealt with before powerful digital processing was available. The signal heard by a listener was attenuated considerably by the equipment. The echo passed through the attenuator twice—once on the way out and once on the way back—and this provided a measure of echo reduction. The use of attenuation to eliminate echo was not a satisfactory solution, and this method was abandoned when digital echo cancellation became available. However, the technique still is valuable in the soft PBX world as a mechanism for getting rid of the echo that remains after the somewhat limited software echo cancellers have done their job.

Digital echo cancellation is based on subtracting from the received signal a correction based on the response of the system to a short spike of sound, called the finite impulse response (FIR). The FIR is simply the echo you would hear from a short ping.

Figure 2 shows 128 digital sound samples or taps taken at a rate of 8,000 times per second, covering 128/8 = 16 milliseconds. The impulse occurred at time zero. The dots represent the individual sample values that have been normalized to an impulse size of 1.

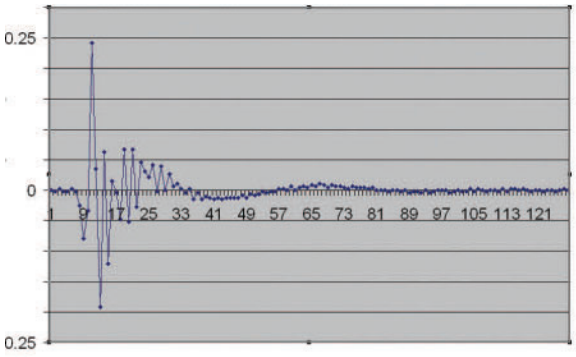


Figure 2. The Response of a Typical System to a Unit Impulse

The first thing to notice is the echo does not appear to be very strong. The impulse had a value of 1, and the highest peak in the response is less than 0.25, falling rapidly to tiny values. But because of the sensitivity of the ear, the echo produced by this system sounds almost as loud as the spoken voice, resulting in a completely intolerable echo on a VoIP system.

The echo from the impulse has an effect that lasts about 10ms (80 taps). To cancel out the echo properly, the input from all the nonzero taps needs to be taken into account. This is why the number of taps in an echo canceller is important. The number of taps is always a power of 2: 32, 64, 128, 256 and so on. Naturally, the higher the number of taps, the higher the computing load and memory requirement.

This echo starts at tap 7, or about 1ms after the impulse. The delay is due to switching and transmission delays on the digital and analog lines. You can see why it is important that echo cancellation takes place close to the echo source. If this echo were being cancelled at the far end of a transatlantic call, there would be many more leading idle taps, so the true echo would be shifted back, perhaps right out of the tap sample. When echo is heard on a system with good echo cancellation, it usually is because an unexpectedly complex system has switching and transmission delays that have shifted the FIR backwards out of the tap sample.

For this call, beyond about 70 taps, the echo tail is small. In practice, this echo canceller would be about as effective at 64 taps, particularly if the leading 8 taps were eliminated by better buffering. That would cut the echo cancellation computation load by half.

The FIR is used to calculate a series of correction factors

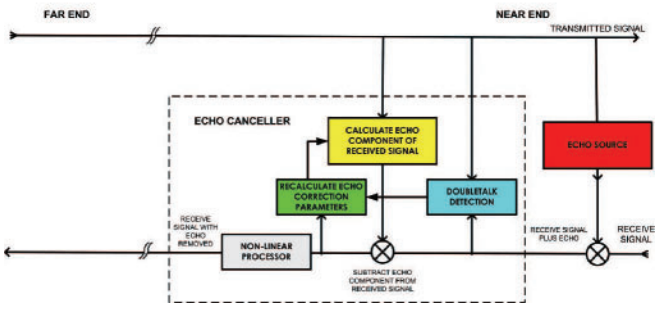


Figure 3. A Typical Echo Canceller

that represent the echo component of the received signal. Mathematically, the echo to be subtracted for each voice sample is given by the dot product of two vectors of dimension equal to the number of taps. On a 128-tap echo canceller, for example, it would look like this:

Echo = (128 values of FIR) • (128 previous tap samples of transmission)

By subtracting this "echo" from the signal as received, a substantially echo-free receive signal is obtained. However, because of rounding errors and non-linearities, some of the echo remains. The nonlinear processor cuts out the remaining received signal if the signal is small enough. In higher-performance echo cancellers, the nonlinear processor then substitutes "comfort noise", background noise so the line does not sound dead.

Obtaining the FIR is an iterative training process based on measuring the residual signal after the calculated echo has been subtracted and changing the FIR estimate. This process requires silence on the other end of the line—there is no doubletalk. The doubletalk detector detects when both parties are speaking at the same time and disables the FIR optimization process until the doubletalk condition has ceased. The iterative FIR optimization converges quite slowly, but as the calculations are done 8,000 times per second, within a second or two of the start of a call, a good echo canceller will be fully trained.

Echo Cancellation in Soft PBX Environments

Echo cancellation is a hugely CPU-intensive process. A complete echo canceller for 92 simultaneous calls, or four PRI T1 lines, consumes on the order of one GIPS. The calculations involve mainly 8-bit operations, and in other ways are not optimum for the PC architecture or CPU cache. Thus, software echo cancellation is one of the major factors limiting the performance of soft PBX systems.

In an effort to improve overall system performance, software echo cancellers are usually highly optimized to reduce the PC load. One compromise made in the interest of saving CPU cycles is that the "learning" algorithms that update the FIR estimate are not run every time a voice sample is processed, but much less frequently. So the system trains slowly. You often hear quite considerable echo well into the conversation until the echo canceller trains and the echo decreases.

Another of the trade-offs is the absence of a nonlinear processor, which often is eliminated completely in soft echo cancellers. This is why there is usually some residual echo on systems such as Asterisk, even after training.

The goal under Asterisk was to provide software echo cancellation for a full quad E1 card (120 channels) with current PC technology and still be able to do other useful voice and data processing. This is indeed possible, but as discussed, the echo canceller trains slowly and after training there is still usually some remaining echo.

You can use the old-fashioned attenuation method to reduce residual echo. The transmit and receive gain settings in Asterisk (txgain and rxgain) can be set to negative values that reduce the sound volumes, but also produce acceptable final echo performance. One limitation is the txgain and rxgain settings in Asterisk are global, meaning the gain settings are compounded for any system with bridging. For bridged TDM systems, it is hard to get the balance between voice volume and residual echo right. But for simpler systems, setting txgain = –10 or thereabouts usually produces acceptable call volume with little perceived echo after about 10 seconds.

The remaining problem under Asterisk is the slow convergence of the FIR estimation. An ingenious mechanism for dramatically improving the convergence time of the echo canceller is Asterisk's echo training option. Transmitted voice is disabled for a short time during ringing and a spike of sound is transmitted to measure the FIR directly instead of learning it iteratively over many samples. The echo training option eliminates most of the echo at the beginning of the call in many cases. But its use is restricted to simple systems where ringing can be detected. It does not function on PRI T1 or E1 lines.

Optimization of Echo Cancellation

Today, all long-distance calls over 600km routinely are echo-cancelled at each end. Cell-phone calls to the PSTN always are echo-cancelled. Calls originating from digital end points, such as ISDN or VoIP, should have no echo. Thus, only analog calls over distances less than 600km actually need any echo cancellation. Even local calls often are echo-cancelled by the PSTN, simply because the capacity is there.

The result is that on most VoIP-PSTN gateways, including Asterisk, a great deal of echo cancellation goes on that is unnecessary and, in fact, detrimental to voice quality. For example, a VoIP-based call center may handle mostly 1-800 calls, the majority being long-distance ones that require no echo cancellation.

Although it is complicated and computationally intensive to cancel echo, it turns out that it is quite easy to measure whether echo is present on a call (Figure 4). A simple algorithm built into a Field Programmable Gate Array can measure within a second or two of speech whether echo cancellation is required for the call. If the call has no echo, echo cancellation can be disabled. Thus, for a system using hardware echo cancellation in DSPs, it is possible to allocate DSP resources dynamically to the calls that need them. But the really dramatic improvements are seen in systems with software echo cancellation.

In software echo cancellers, the considerable CPU load that can be freed by echo detection is always immediately available to other processes, which in turn can increase the quality and capacity of the system significantly. More important, echo detection changes the optimization point of the echo canceller design. If only a fraction of calls will require any echo cancellation, the canceller itself can afford to be designed to include the additional features, such as nonlinear processing and fast convergence, that will make the audio truly toll-quality.

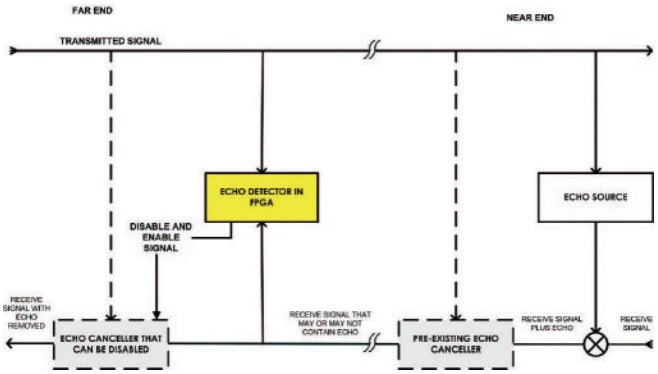


Figure 4. Echo cancellation isn't necessary for incoming calls that already are echo-cancelled. An echo detector can be used to switch off echo cancellation for these calls.

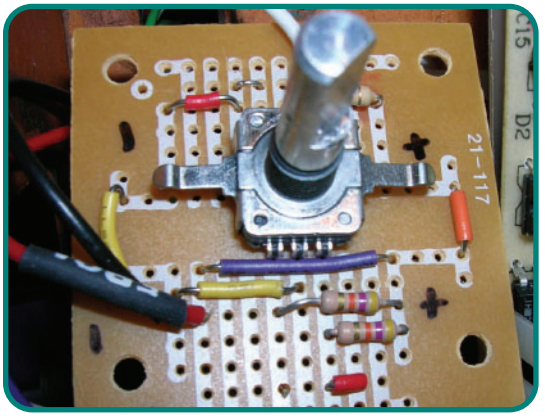
Conclusion

Echo on a telephone call is an annoying phenomenon that has been mostly under control in the classic telephony system, but it is rearing its head again as VoIP proliferates. Its effective control is vitally important for the eventual success of VoIP technologies in general, because of the effect of echo on perceived quality. For open-source VoIP PBX/IVR technologies to become truly mainstream, toll-quality audio must be a given, and this requires reliable, high-performance echo cancellation.

David Mandelstam is the President and CEO of Sangoma Technologies. Before founding Sangoma, David ran a private engineering company, was engineering VP of Solartech, an energy conservation company and was responsible for pricing at Spar Aerospace. Prior to immigrating to Canada, David was in charge of aircraft engine maintenance for South African Airways. David holds a BSc in mechanical engineering from the University of Witwatersrand in South Africa, an MSc in aerodynamics from the Cranfield Institute of Technology in the United Kingdom and a BComm from the University of South Africa.

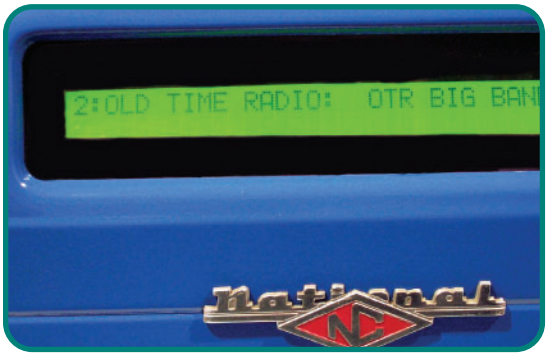
The Hardware Hacking behind the Software Radio

You can turn an old radio into a new Linux-based appliance that can catch a diverse collection of shows that would never get on the air in your hometown. The project needs both hardware and software work, but Linux ties it all together. Get all the details on page 60. **BY DAN RASMUSSEN, PAUL NORTON AND JON MORGAN**



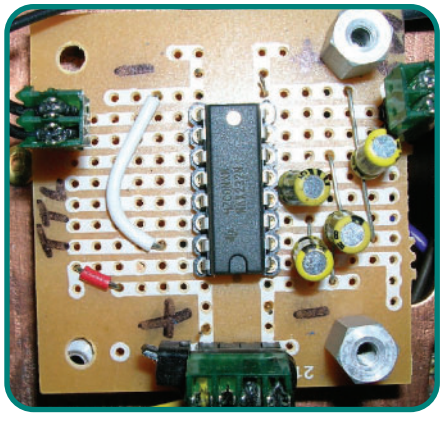
ROTARY ENCODER

A rotary encoder is a digital input device used to measure angular rotation and direction. It does this by sending two out-of-phase pulse trains. Direction is determined by which pulse arrives first. The pulses then can be counted to determine magnitude of rotation. There are many manufacturers and grades of rotary encoders. We used a unit by Bourns, part number PEC11-4225F-S0024. See the Radii home page for details on how to interface this encoder with a PIC.



SHOPPING FOR AN LCD

When shopping for an LCD, first make sure it is HD44780-compatible. This is the most widely supported interface; anything else could slow down your efforts. The backlight type for the display is also important. Electro Luminescence—think Timex Indiglo—looks great but has unusual power requirements. The fastest and easiest way to go for backlighting is to use an LED backlit display. An LED backlight generally requires standard 5 VDC power. When shopping for an LCD with backlight, be sure to verify the type of



INTERFACING A PIC TO RS-232

The PIC interface levels are TTL-level outputs (that's transistor-transistor logic). With TTL, about 5V is on and about 0V is off. Interfacing this to RS-232/serial port (12V on/0V off) requires the use of a TI MAX232 dual-driver/receiver chip and a handful of resistors/capacitors. The chip does most of the work for you, but some assembly is required for the interface board and the serial cable used.

Rackspace — Managed Hosting backed by Fanatical Support.™

Servers, data centers and bandwidth are not the key to hosting enterprise class Web sites and Web applications. At Rackspace, we believe hosting is a service, not just technology.

Fanatical Support is our philosophy, our credo. It reflects our desire to bring responsiveness and value to everything we do for our customers. You will experience Fanatical Support from the moment we answer the phone and you begin to interact with our employees.

Fanatical Support has made Rackspace the fastest-growing hosting company in the world. Call today to experience the difference with Fanatical Support at Rackspace.

LINUX JOURNAL READERS' CHOICE 2004

Thanks for honoring us with the 2004 Linux Journal Readers' Choice Award for **"Favorite Web-Hosting Service"**

rackspace™ MANAGED HOSTING

1.888.571.8976 or visit us at **www.rackspace.com**

The World's Fastest InfiniBand Switch

From a Company You've Trusted for 23 Years

▲ 72 Port FasTree™ Configuration

Microway's **FasTree**™ DDR InfiniBand switches run at 5GHz, twice as fast as the competition's SDR models. FasTree's non-blocking, flow-through architecture makes it possible to create 24 to 72 port modular fabrics which have lower latency than monolithic switches. They aggregate data modulo 24 instead of 12, improving nearest neighbor latency in fine grain problems and doubling the size of the largest three hop fat tree that can be built, from 288 to 576 ports. Larger fabrics can be created linking 576 port domains together.

Working with PathScale's InfiniPath HTX Adapters, the number of hops required to move MPI messages between nodes is reduced, improving latency. The modular design makes them useful for SDR, DDR and future QDR InfiniBand fabrics, greatly extending their useful life. Please send email to fastree@microway.com to request our white paper entitled *Low Latency Modular Switches for InfiniBand.*

Microway's **QuadPuter**® includes four AMD single or dual core Opteron™ processors, 1350 Watt redundant power supply, and up to 5 redundant, hot swap hard drives-all in 4U. One of the most powerful processing platforms in the HPC industry, QuadPuter can serve as a cluster node or a standalone supercomputer. Constructed with stainless steel, its RuggedRack™ architecture is designed to keep the processors and memory running cool and efficiently. The power supply exhaust does not mix with air in the motherboard chamber. Hard drives are cooled with external air and are front-mounted along with the power supply for easy access and removal. The RuggedRack™ is available with an 8-way motherboard, dual-core Opterons and up to 128 GB of memory for power- and memory-hungry SMP applications.

◀ QuadPuter® Navion™ with Hot Swap, Redundant Power & Hard Drives

Call us first at 508-746-7341 for quotes and benchmarking services. Find technical information, testimonials, and newsletter at microway.com.

Visit us at SC2005 Seattle

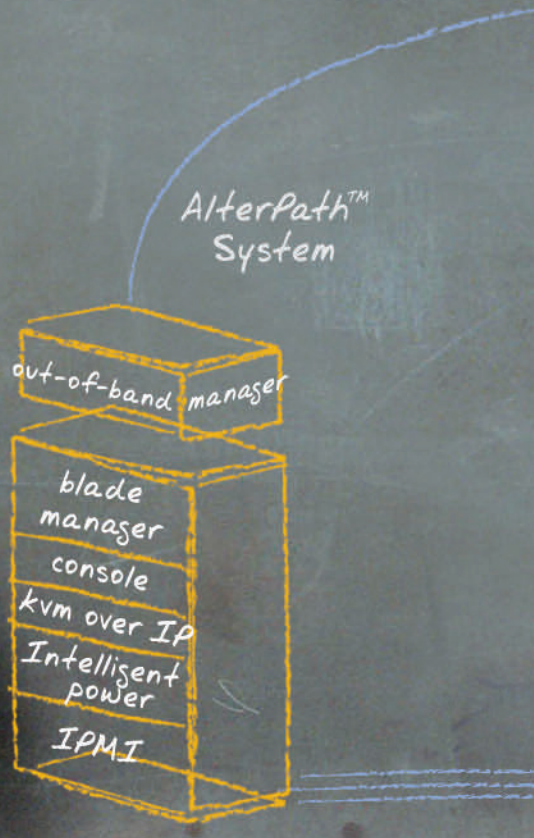
AMD 64 Opteron™

PathScale™ Accelerating Cluster Performance™

Microway®
Technology you can count on℠

508.746.7341 microway.com

Cyclades AlterPath™ System makes



A+

For a **FREE** copy of Enterprise Management Associates' white paper **Increased IT Operations Effectiveness With Cyclades AlterPath™ System for the Out-of-Band Infrastructure,** please visit us at www.cyclades.com/ema