REVIEWED: grsecurity ... kernel security patch suite

O3:

The Open Source Enterprise Data Networking Magazine

Issue 3 / January 2006

http://www.o3magazine.com

Linux and Open Source on mainframes
A look at Linux on the IBM zSeries

Real Time Streaming Protocol RTSP uses Examined

Building a global Internet Presence

Deploying Localized Web Services

Intrusion Detection Load Balancing with SNORT and Nortel Application Switches

Networking on the IBM zSeries for developers

doing business in Europe ?

are your servers close enough to your target market ?

high performance Internet Services

colocation
dedicated servers
web hosting

Layer Two
http://www.layer-two.net

+44 870 141 7273
sales@layer-two.net

CONTENTS

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 2/28/06
Get $500 Off
Enter PROMO CODE: 03M1205

LEVANTA
www.levanta.com
1.877.LEVANTA

WINNER
LinuxWorld
Most Innovative Hardware Solution

Sometimes things..

JUST DON'T GO ACCORDING TO PLAN. THIS MONTHS ISSUE DEVIATES SLIGHTLY FROM OUR PLANNED CONTENT BUT IS IN NO WAY WORSE OFF FOR IT...

BY JOHN BUSWELL

Welcome to Issue 3 of o3 magazine. It has been an exciting first three months with Issue 2 downloads eventually exceeding the initial Issue 1 downloads by over 100,000 readers. We would like to thank everyone who took the time to contact us by mail, phone and email to praise the publication. We would also like to thank the brave individuals who took the time to provide public criticism regarding our first two issues in their blogs.

If you need to blame someone for having to wait an extra two weeks for Issue 3, look no further. The decision was made to give ourselves enough time to review all of the feedback, both direct and indirect, and make improvements to the presentation and delivery of o3.

Issue 3 was built using a custom environment built entirely from source under Mandriva 2006. We're now running Scribus 1.3.2 which, if you are unfamiliar with it, is an open source publishing application available from http://www.scribus.org.uk. This latest development release has improved PDF export capabilities, so hopefully that should address the page layout complaints we received about previous issues displaying 4 pages at a time in some PDF viewers.

Hopefully we have corrected the baseline technicalities that some professional publishers pointed out about o3 with the upgrade and some improved quality control measures. The overall success of o3 is something we hope we can continue to enjoy, and the concept of o3 seems to be far more threatening to the mainstream IT media than our free publication.

If you are a professional IT writer and you see us make a mistake, please let us know. The entire o3 publication is built by engineers, not English majors, so your help is appreciated.

Due to space constraints, the planned article for building secure appliances has been postponed until a later date, and replaced with an article that introduces grsecurity. The next issue of o3 (February 2006) is almost complete, and again, part of our reasoning behind the delay of this issue was to get ahead. That issue looks at building a secure Internet presence with a focus on Dynamic Network Protocols, QoS, rrdtool, deploying secure DNS and much more. If you're concerned your Internet presence isn't deployed properly, then don't miss our next issue.

Finally, we're pleased to announce the addition of our own dedicated European server and another 10Mbps of bandwidth in Florida.

If you happen to be near Ohio on Thursday March 23rd 2006, check out the ad on the back page. This will be the first event since Ohio LinuxFest 2005 that you'll be able to meet and greet the team behind o3 and AppOS in person. We'll be happy to discuss o3, scribus, open source and answer your questions.

O3 Magazine
January 2006
Issue 3

EDITOR IN CHIEF
JOHN BUSWELL
EDITOR@O3MAGAZINE.COM

EXECUTIVE EDITOR
JAMES HOLLINGSHEAD
JAMES@O3MAGAZINE.COM

ARTWORK
JOHN BUSWELL

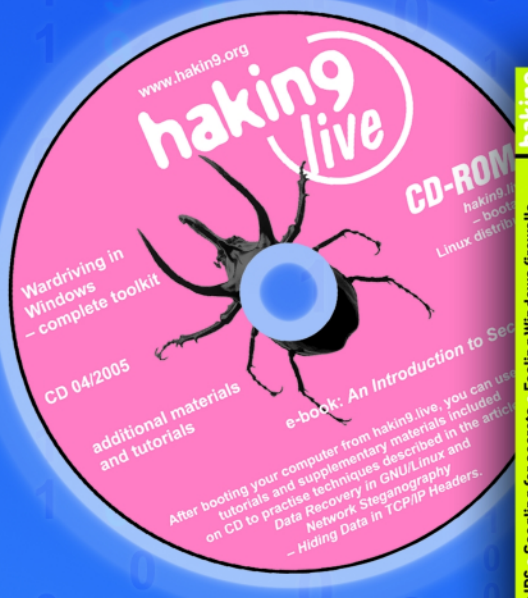
PROOF READERS
GREG JORDAN
SHAWN WILSON
FRANK BOYD
STEW BENEDICT

SALES AND MARKETING
GREG JORDAN
SALES@O3MAGAZINE.COM

SUBSCRIPTIONS
O3 MAGAZINE IS DISTRIBUTED ELECTRONICALLY FREE OF CHARGE BY SPLICED NETWORKS LLC. TO SUBSCRIBE VISIT WWW.O3MAGAZINE.COM.

SOFTWARE
SCRIBUS 1.3.2
GIMP 2.0.5
OPENOFFICE 1.1.2

We have knowledge. Want some?



www.hakin9.org/en

If you want to buy a magazine, please visit us at www.shop.software.com.

available at the beginning of July

UPCOMING EVENTS

Developer Relations Conference
February 6 - 7, 2006
San Francisco, California, USA
http://www.evansdata.com/drc2

Implementing IT Security - Strategy to Reality
February 8 - 9, 2006
Seattle, Washington, USA
http://www.ip3seminars.com/

Charting Your Course Through Open Source
February 11, 2006
Burlington, Massachusetts, USA
http://www.mitforumcambridge.org/WW06/OpenSource.html

Southern California Linux Expo
February 11 - 12, 2006
Los Angeles, California, USA
http://www.socallinuxexpo.org/

RSA Conference 06
February 13 - 17, 2006
San Jose, California, USA
http://2005.rsaconference.com/us/C4P06/

OSBC West
February 14 - 15, 2006
San Francisco, California, USA
http://www.osbc.com/live/13/events/13SFO06A

Embedded World 2006
February 14 - 16, 2006
Nuremberg, Germany
http://www.embedded-world-2006.de/

LinuxWorld Conference & Expo
February 14 - 17, 2006
Mexico City, Mexico
http://www.linuxworldexpo.com.mx/

UPCOMING EVENTS

Open Source World Conference
February 15 - 17, 2006
Malaga, Spain
http://www.opensourceworldconference.com

Sun Tech Days 06
February 22 - 23, 2006
Singapore, Singapore
http://developers.sun.com/events/techdays/

PyCon 2006
February 24 - 26, 2006
Addison, Texas, USA
http://www.python.org/pycon/2006/

Have an upcoming event? Tell us about it, send email to events@o3magazine.com with details.

FEATURED FUTURE EVENT

FOSDEM 2006

February 25 - 26 2006
Brussels, Belgium
http://www.fosdem.org/index

FOSDEM is a free and open source software developers' European meeting organized by volunteers. The event is a two day event to promote the use of Free and Open Source software. This year marks the sixth event, and is being held in the city of Brussels, Belgium. FOSDEM meetings are recognised as the best Free and Open Source events in Europe.

FOSDEM is a free event that relies upon donations to help organize and to keep the event free.

JANUARY OPEN SOURCE REPORT

Welcome to the Open Source Report. This is the section of O3 where we give a brief run-down of the major applications which made releases during the month.

SCRIBUS
http://www.scribus.org.uk/
Release: **1.3.2**

The latest release of Scribus resolved over 290 requests and bugs. A major code restructuring was undertaken in preparation for a new modular file load and save system. Support for EXIF and more TIFF and PSD file formats were added. Image tinting and sharpening effects were implemented. Section-based page numbering is now supported. Updates were made for Windows and Mac OS X compatibility. Significant updates were made to the documentation.

WINE
http://www.winehq.org/
Release: **0.9.6**

The latest release of Wine includes several OLE fixes and improvements, DirectSound improvements, including full duplex support, a fix for the Windows metafile vulnerability, many static control improvements, some fixes for copy protection support and many bugfixes.

FAST LOGGING PROJECT FOR SNORT
http://www.geschke-online.de/FLoP/
Release: **1.5.0**

The latest release of .FloP, a control thread was added so that some parameters can be chained during runtime. The restriction of one snort process per sensor was removed. This way it is also possible to encrypt the communication via stunnel or an SSH tunnel.

LIGHTTPD
http://www.lighttpd.net/
Release: **1.4.9**

The latest release of lighttpd fixed a critical source-file retrieval issue on case-insensitive file-systems like HPF+ on Mac OS X and an issue with endless logfile writing if the fastcgi backend is dead. A new statistics framework was added. mod_cml now has a power-magnet for rewriting requests with LUA.

THUNDERBIRD
http://www.mozilla.com/thunderbird/
Release: **1.5**

The latest release of Thunderbird includes new features including automatic updates, anti-phishing protection, inline spellchecking, saved search folders, podcasting, RSS improvements, the ability to delete attachments from messages, and a whole lot more .

FREERADIUS
http://www.freeradius.org/
Release: **1.1.0**

The new release of FreeRADIUS is focused on new features and bugfixes without sacrificing stability. New features include more vendor dictionaries, support for Lucent and Starent VSAs, support for Juniper encrypted VSAs, N-tier certificates, and load-balanced access to back-end databases. In addition, the Perl module is now stable, and a new "sql log" module may be used to lower the load on an SQL server.

POSTGRESQL
http://www.postgresql.org
Release: **8.1.1**

This release of Postgres improves concurrent access to the shared buffer cache and allows index scans to use an intermediate in-memory bitmap. Two-phase commit has been added. Creates a new role system that replaces users and groups. Automatically uses indexes for MIN() and MAX(). Adds shared row level locks using SELECT ... FOR SHARE. Adds dependencies on shared objects, specifically roles. Improves performance for partitioned tables.

LINUX ASIA 2006

Asia's Premier Open Source Conference & Expo

FEBRUARY 8-10, 2006 ■ NEW DELHI, INDIA

The Customer Speaks

- Industry Sessions
- E-Governance Forum
- Education Forum
- Technology Forum
- Technology Workshops
- LFY Awards
- Penguin Party
- .ORG Mela
- Industry Expo
- .GOV Pavilion

REGISTER TODAY

www.linuxasia.net

Grsecurity Reviewed

GRSECURITY IS A PATCH SUITE FOR THE LINUX KERNEL

THAT PROVIDES A WIDE RANGE OF INTERESTING SECURITY ENHANCEMENTS

BY JOHN BUSWELL

Grsecurity s a suite of patches distributed in a single patch file for the Linux kernel that provide a wide range of security enhancements. Earlier this month, news reports uncovered the United States government engaging in covert surveillance of its own citizens without a warrant. An effort spearheaded by the NSA. The same NSA responsible for the SELinux patches which are currently used in the mainstream kernel. While it is unlikely the NSA has tried to insert any suspect code into the SELinux patches, the possibility is there. The NSA released SELinux under the GPL in January 2001, approximately 11 months prior to the presidential order handing the NSA special powers to investigate within the United States. If you manage a foreign business or government network, then you might want an alternative to SELinux, or at least audit the SELinux source code with caution.

Armed with your distrust for the United States government, you want to run something that might be more politically correct in your country, or simply something that provides superior security. In that case, you'll like grsecurity.

Grsecurity physically alters the Linux kernel, providing tighter restrictions. Grsecurity features can be manipulated through the standard /proc interface (sysctl). Grsecurity is available from *http://www.grsecurity.net.* The latest release at the time of writing is patched against Linux 2.6.14.6.

[johnb@x2 tmp]$ cd linux-2.6.14.6

[johnb@x2 linux-2.6.14.6]$ patch -Np1 -i ../grsecurity-2.1.8-2.6.14.6-200601211647.patch

Several screens roll by, so check that nothing broke by scanning for .rej files :

[johnb@x2 linux-2.6.14.6]$ find ./ | grep rej

[johnb@x2 linux-2.6.14.6]$

Next, simply clean the environment and run the kernel configuration tool:

[johnb@x2 linux-2.6.14.6]$ make mrproper

[johnb@x2 linux-2.6.14.6]$ make menuconfig

The grsecurity features can be found under Security options. You will see two new menu options, PaX and Grsecurity. You need to enable Grsecurity under the Grsecurity menu option in order for the PaX options to appear.

[*] Grsecurity

Security Level (Custom) --->
Address Space Protection --->
Role Based Access Control Options --->
Filesystem Protections --->
Kernel Auditing --->
Executable Protections --->
Network Protections --->
Sysctl support --->
Logging Options --->

While you can run with the low, medium, high presets under Security Level and call it a day, we're going to run through the Custom options. Its likely that the presets will break some application you want to use, so its usually a good idea to walk through the custom options.

ADDRESS SPACE PROTECTION

If you don't need to use loadable modules, which is often the case on a dedicated server, you can close up shop by enabling Deny writing to /dev/kmem, etc. However, if you do need to use modules, perhaps a third party loadable module thats only available in executable format, then this option is simply

unavailable to you.

Disabling priviledged I/O, blocks access to all ioperm and iopl calls, which can be used to modify a running kernel. Some programs do need these calls to function properly, one is Xfree86, however since most servers don't run X, this is perfectly ok. The other notable application is hwclock, which can be fixed by enabling RTC support in the kernel.

The Deter exploit bruteforcing is a neat feature that helps deter brute force attacks against applications such as apache or sshd that fork child processes. With this feature enbled, the parent process is delayed 30 seconds upon every subsequent fork following an illegal instruction crash or having been killed by PaX.

Runtime module disabling enables kernel modules to be loaded at boot time, but once loaded, they cannot be unloaded. This is a useful protection against root kit installation by malicious users.

Hiding kernel symbols, provided you are not using a precompiled distribution kernel, and restricting access to the kernel files themselves, can help protect against local and remote kernel exploitation.

In our server configuration, it is safe to enable the last three options.

[ ] Deny writing to /dev/kmem, /dev/mem, and /dev/port

[ ] Disable privileged I/O (NEW)

[ ] Deter exploit bruteforcing (NEW)

[ ] Runtime module disabling (NEW)

[ ] Hide kernel symbols (NEW)

ROLE BASED ACCESS CONTROL OPTIONS

The RBAC menu enables you to hide kernel process information and sets limits to lock out password attempts. Once locked the menu has an option to wait a specific amount of time before unlocking. Control over grsecurity's RBAC system is done with a utility called gradm. The steps to manage RBAC are fairly straight forward.

First set a password:

gradm -P admin

Now you can login with the admin command:

gradm -a

Now you can setup RBAC by using grsecurity's learning mode. In this mode, grsecurity monitors the system looking for processes that run with root privileges, access the network or write to key files. Then grsecurity generates an access control list (ACL) to run these processes with minimal privileges.

gradm -F -L /etc/grsec/learning.log

During the learning process, you need to avoid any administration tasks involving root. As it runs for a few days, grsecurity will have had enough time to recognize normal system activity, then its time to disable learning mode and log into gradm in admin mode.

gradm -F -L /etc/grsec/learning.log -O /etc/grsec/acl

The above command will write out the ACL to disk. Finally start up the RBAC with the new ACL, at any time you can disable RBAC with gradm -D.

gradm -E

FILESYSTEM PROTECTION

One of the best features of Grsecurity is its capability to lock down chroot() within the kernel. If your not familiar with chroot(), the easiest way to describe it is that it manipulates the root directory (/) so that it becomes a specific directory for a running process. For example, lets say we start our application in /chroot/dns/, here we run the named process with / = /chroot/dns. The environment sees /chroot/dns as /, so anything below /chroot/dns is unavailable, normally (eg. You cannot cd ../ from /).

The first set of filesystem protections restict /proc to either a particular user (UID) or a particular (GID),

meaning you can run your applications under their own username , but only users in a special group can actually access /proc. If your named user for running bind (dns) for example is compromised, if they're not in the /proc group or are not the /proc UID, they cannot access /proc. If you select to restrict to a UID, you cannot use the group feature.

[*] Proc restrictions
[ ] Restrict /proc to user only
[ ] Allow special group (NEW)

The linking restrictions prevent users from following symlinks that are owned by another user in world-writable directories. This prevents /tmp race exploits. The FIFO restrictions, do the same thing for FIFOs.

[*] Linking restrictions
[*] FIFO restrictions

The chroot() restrictions in Grsecurity are very impressive. Most of the options are pretty straight forward, ones that might not be so apparent, we will discuss. The pivot_root capability in Linux allows you to switch the root directory. Its intended for booting from initial ramdisks, however it has possible uses to compromise security. The option can be enabled via sysctl, enabling the use of the command during boot in an initrd, then disabling it once that has completed. If you run all your applications within chroot() then the Deny sysctl writes, tidies up any potential reversal of the security options via sysctl.

The Deny double-chroot prevents chroot() from being used within an existing chroot() environment to break out of the chroot. This combined with the Enforce chdir("/") cleans up the shortfalls with the chroot command.

Unfortunately, many administrators rely too much on chroot, and many of the capabilities Grsecurity adds to chroot() are things that administrators think chroot() already provides when it does not. In our configuration, its safe and recommended to enable all of these.

The chroot part of the filesystem protection menu in the kernel configuration looks like this:

[*] Chroot jail restrictions
[ ] Deny mounts
[ ] Deny double-chroots
[ ] Deny pivot_root in chroot
[ ] Enforce chdir("/") on all chroots
[ ] Deny (f)chmod +s
[ ] Deny fchdir out of chroot
[ ] Deny mknod
[ ] Deny shmat() out of chroot
[ ] Deny access to abstract AF_UNIX sockets out of chroot
[ ] Protect outside processes
[ ] Restrict priority changes
[ ] Deny sysctl writes
[ ] Capability restrictions

NETWORK PROTECTIONS

Grsecurity cleans up a number of things with Linux Networking. The larger entropy pool enables better randomization. Randomized TCP source ports uses a random value based algorithm instead of simple incrementing source port numbers. The socket restrictions provides controls to block specific groups from creating sockets, client sockets and server sockets. If you run a daemon that doesn't need to create client sockets, then adding that user to the deny clients socket group prevents that service, if compromised from being used to send packets to systems that might trust that now compromised system more than an outside IP.

[*] Larger entropy pools
[*] Randomized TCP source ports
[*] Socket restrictions
[ ] Deny any sockets to group
[ ] Deny client sockets to group
[ ] Deny server sockets to group (NEW)

Kernel Auditing, Logging and Sysctl support

The kernel auditing menu provides a wide range of events to log. Grsecurity has the capability of only auditing a specific group, which enables an administrator to add users to a specific group that they want to watch, instead of having to watch trusted users on the system, if desired. The logging menu provides control to rate limit the amount of log

information that is logged to prevent flooding. The sysctl menu enables sysctl support, which I would highly recommend, and disable turning on all features as default. It is better to handle turning the features on during boot as to not over-restrict the system so that it cannot boot.

EXECUTABLE PROTECTION

Grsecurity offers a number of protections for protecting executables. It can check for resource limits during execve() instead of just during fork(). It can clear unused shared memory, it can limit access to dmesg, randomize process Ids and has support for Trusted Path Execution.

Further executable protection is offered under the PaX menu under Non-executable pages. Utilizing the Restrict mprotect() and Disallow ELF text relocations, if your operating system is compiled as PIE (position independent executables) with PIC (Position Independent code), its possible to add a high degree of security by enabling this option. For further information on PIE and this option, refer to previous o3 security articles in Issue #1.

PAX

PaX is controlled through the paxctl application. PaX is a patch that implements least privilege protections for memory pages in the Linux kernel. PaX is a substantial asset in securing a Linux system, for a detailed explanation of PaX and how PaX works, there is an excellent entry on Wikipedia (*http://en.wikipedia.org/wiki/PaX*)

While grsecurity may not have the high profile of SELinux, it is by far the more capable solution.

JOHN BUSWELL IS CTO OF SPLICED NETWORKS LLC. EMAIL JBUSWELL@SPLICEDNETWORKS.COM.

>THIS IS **THE WAY**

600 MILLION PEOPLE MOVE AROUND THE PLANET.

You'll find Nortel™ in every single one of the world's top twenty airlines. And wherever secure, reliable data and voice communications are most critical.

>THIS IS NORTEL™

www.nortel.com/commerce

Free Software MAGAZINE

The free magazine for the free software world

- Articles are released under a free license
- Available online as HTML or PDF
- Packed with amazing content
- Both technical and non-technical articles

GO AND SEE FOR YOURSELF!

WWW.FREESOFTWAREMAGAZINE.COM

Evaluating your Internet Infrastructure

TODAY IT IS NO LONGER JUST A MATTER OF HAVING AN INTERNET PRESENCE

THAT PRESENCE MUST BE HIGHLY AVAILABLE AND RESPONSIVE FOR YOUR TARGET MARKET

BY JAMES HOLLINGSHEAD

Evaluating your Internet infrastructure has a lot in common with the strategic planning and evaluation of most things. After the techno-speak and business-speak are stripped away, it really comes down to two basic concepts:

What you have and what you need.

WHAT DO I HAVE?

For the moment, put the issue of need aside. The first thing you want to look at are the resources you currently have.

Take a look at the way your network as it is right now. You need to ask yourself the following set of questions. Don't worry if you don't know all of the answers off the top of your head, but it is important to find them.

First, what public servers do you have on your network? Is it just a web or mail server or do you have half a dozen production boxes, a mail server, and the Robosapien that you got for Christmas? How many of them need to be at your physical location? (hint – the Robosapien probably isn't one of them.)

Which of the services that you have are business critical? If there is any doubt as to whether a service is business critical or not, think of what would happen if it suddenly went down. If the answer is that you and your employees would not be able to work, would be unable to make money, or would lose a great deal, then it is a business critical service.

Where, physically, are your servers? Are they in Chicago, somewhere on the east coast, California, or somewhere else?

Where are the majority of your customers? Are they in the eastern United States or in central Germany? If you have multiple large concentrations of customers, list all of them. The important thing for this point is to have a list of where the bulk of your traffic comes from.

How much traffic do you have? Do you experience consistent traffic or does it does it come in spikes with mild to moderate traffic at all other times? If you experience major traffic spikes, do they follow any certain pattern? Do they only happen for a few hours once a month or do they last for days?

Next, how much bandwidth do you pay for? Are you paying for a T1, fractional T1, business class DSL, or some other from of connection? If it's a T1 or T3, remember that there are usually two charges – one for the carrier (who supplies the physical line – generally the phone company) and one for the ISP (who provides your Internet connection).

How much downtime do you experience in a typical month? Is it just a few hours or can it be measured in days? Do you have a service level agreement (SLA) with your provider? If so, is it satisfactory for your needs?

How many physical connections to the outside world does your network have? Is there just one point of entry for your network, or are there multiple ones?

HOW DO I USE MY RESOURCES?

I know I said that it boils down to resources and needs (which it does), but before you can answer the question of what you need, you have to think about how you use the things that you already have.

What are you using your public servers for and how much of your bandwidth do they use? Are they mail or VoIP servers (which make sense to have on site) or are they production boxes which can generally be pretty much anywhere? If they are production boxes and they are using a lot of bandwidth, you might want to see if it makes sense to move them to a co-location site.

How much bandwidth do you use? You already looked at how much of your bandwidth your production boxes take up. Now it's time to look at the network as a whole (both incoming and outbound traffic). Also consider the amount of bandwidth you would be using if your production boxes were co-located somewhere else.

Are you paying bandwidth or transfer overage

charges? Some service providers offer connections that normally function at one speed but are burstable to a higher speed (sometimes charging a fee if you exceed the normal speed for a certain period of time). Likewise, some providers meter the amount of data that you can transfer per billing period and charge you extra for the amount that you exceed that limit by.

Are the services that you identified as business critical redundant? If the server you have it hosted on exploded, would you still be able to do work? Believe it or not, this is something which I have seen in the past. It was impressive and awe inspiring but it was also not the best way to start a day at work (especially since it was the middle of February and we had to open the windows in order to vent the resulting smoke).

You've answered the questions of where your servers are and where your customers are. Do these two answers compliment each other? If most of your customers are in the western portion of the United States and your server is in California, then the answer is yes. If most of your customers are in New York and your servers are in Germany, that's something you might want to change. Lastly, do you have heavy concentrations of customers in countries other than the one your home office is in?

WHAT DO I NEED?

Now that you've made a list of what you have and how you use it, you can start to look at how things need to be changed.

Are you paying for a lot of bandwidth that you aren't using or paying overage charges? This question is a little tricky because the objective is to have more bandwidth than you need, but not too much more. Having too much bandwidth is a bad thing, but not having enough is even worse.

First, consider the amount of bandwidth that your production servers (or any other server that it might make sense to co-locate) are using. If you have multiple T1s just to supply bandwidth to your production boxes and they don't necessarily have to be on site while the bulk of your remaining traffic is simple email or browsing, it might make more financial sense to have them hosted elsewhere and downgrade the service in your office to business DSL.

On the other hand, if you have a fractional T1 and actually use the bandwidth, it might make more sense to upgrade to a T1 because pricing for a full T1 is generally cheaper due to extra work required by the phone company for fractional T1.

If you have any questions as to whether you are getting a good price for your T1 or T3 line, *www.bandwidth.com* is a good place to check for price quotes. It should also be noted that installation fees might be waived and your fees lowered if you sign a long term service contract (much like the pricing difference between a regular cell phone plan and with a two year contract).

One thing to keep in mind if you want to change the type of connection you have is the SLA. T3 and T1 generally have very good SLAs while the SLA for business class cable tends to be much poorer and not really suited for running production servers. DSL tends to be somewhere in between cable and T1 in terms of SLA but access for business DSL is often bandwidth or distance limited.

Now that you know where your customers are and where your servers are, do you need servers in other locations to cater to them? If so, you might want to look at the web tech section in this issue.

If you have heavy concentrations of customers in foreign countries, it would also make sense to have country specific (i.e. .co.uk) domains for those countries. This will allow you to have language specific sites as well as save both you and your customers from trying to figure out how to deal with currency differences, tax, shipping, etc.

In the last section, you looked at whether or not your business critical services were redundant. If the answer is no, you have a few options.

- Simply add another server where you are that will run backup versions of those services in the event that the main server goes down. This is often the cheapest option. However, if something were to happen to your building (fire, flood, earthquake, etc), your business critical services would still go down.

- Open a remote office with redundant servers and staff. While this option works really well (providing you have a competent staff), it tends to be much more expensive than most companies are able (or willing) to deal with.

- Pay to have redundant servers co-located. While more expensive than simply adding another server to your rack, it is markedly less expensive than opening a whole new office. This option is discussed in more depth in the web technologies section of this issue.

If there is only one physical access point from your network to the outside world, you might want to consider adding a second one. The reason for this is that these links occasionally fail for various reasons. The most infamous of these is often referred to as the Backhoe of Doom which is when someone digging a trench doesn't make absolutely certain that there is nothing important buried where they want to dig before they bring in the earth moving equipment. If this happens to you and that link to the outside world is a single point of failure, you're going to be stuck playing solitaire and emailing to the other people in your building because you certainly won't be able to do much else.

Above all, you should remember to leave yourself room to grow as far as network requirements are concerned. If you've locked yourself into a certain plan for an extended period of time in order to get a good rate and your needs suddenly increase, you may find yourself stuck with either overage charges or a network which can not handle the amount of data being demanded of it if you stay with that provider or cancellation fees if you need to go with a different provider because your current one no longer fits your needs (it is rare for a provider to charge you for breaking a contract if you go for a larger contract with that provider) unless you can use co-location and load balancing to move enough traffic through a network other than the one at your location.

WHAT ELSE COULD I USE?

Covering what you need is vital. However, there are things which, while not necessary to your network and Internet infrastructure, can help improve it.

Among these possible non-vital improvements, if you decide to keep your production servers at your location, and want the connection they have to the outside world to be solid while it is alright for the rest of your traffic (such as email and browsing) to go down on rare occasions, you might want to consider keeping your production servers on a T1 and the rest of your traffic on business class DSL. To do this, you can use QoS (Quality of Service), which is now available in the Linux kernel, to route outbound traffic from your production servers over the T1 while routing all non-essential traffic over your DSL connection.

Your infrastructure isn't something that should only be evaluated once and then forgotten. Like the rest of your business, it may have to change over time. You should re-evaluate it periodically to see if you need to expand or shrink. It's all a matter of knowing what you have, what you need, and how to balance the risks and rewards of how you deal with your resources.

JAMES HOLLINGSHEAD IS EXECUTIVE EDITOR AT O3 MAGAZINE. JAMES CAN BE REACHED AT JAMES@O3MAGAZINE.COM. CONTACT JAMES IF YOU ARE INTERESTED IN HAVING BANDWIDTH.COM QUOTE YOU FOR LOWER COST BUSINESS DSL, T1 OR T3 PRICING.

Providing High Performance Web Services

DOING BUSINESS ON A GLOBAL SCALE REQUIRES A GLOBAL INTERNET PRESENCE

LOCALIZED CONTENT IS A START BUT LOCALIZED CONTENT DELIVERY IS A REQUIREMENT IN TODAYS WORLD

BY JAMES HOLLINGSHEAD

Back in the mid 90's, when the world wide web was relatively new, just having a website that had information about your company was sufficient. They were pretty expensive to maintain because of bandwidth and server costs involved, so most businesses only had one.

Things are a little different now. With the decrease in the cost of bandwidth and server grade computers, those websites are now used for everything from simply supplying information to e-commerce and the large-scale distribution of content (such as this magazine).

With the changes in the way the Internet is used, the way that websites were handed had to change as well. One set of servers in a single geographic location providing all the global services for your customers is no longer sufficient.

WHY SHOULD I BE CONCERNED?

Having multiple servers worldwide which hosts your business services has several advantages. The most obvious is redundancy. If one server goes down, customers can still access the remaining servers. This is very important if you use your server for e-commerce or critical customer services because downtime has a direct effect on the bottom line.

There was one very dramatic example of what can happen to a server in one location when Hurricane Katrina made landfall at New Orleans. Many businesses located there had great difficulties in relocating their servers and experienced severe amounts of downtime. One data center there was the notable exception, because they staffed their floor of the building through the entire ordeal, keeping the servers up with a diesel generator and the looters away through other means.

They even kept a weblog during their ordeal to show people what the city was like during the disaster. In fact, they were one of the main, unbiased sources of news for the city of New Orleans for a significant portion of the disaster. However, while their efforts were extremely impressive (and I admit that I read the blog daily), not every place which hosts servers will go to such lengths to ensure your uptime.

Having multiple servers in different geographic locations allows you to let customers in other parts of the world have faster access to your sites and allows you to better tailor your content to other cultures. This means that someone in Asia would be able to access your site on a server closer to where they are rather than having to deal with the latency of a connection between them and North America (for instance) and will be able to view a site with, to them, a more native feel than the one that your American customers would use. This is very important because we all want to feel comfortable when using on-line services, and most people will not wait more than a few seconds to begin seeing results when accessing your website.

Tied to these first two points is an important third point – sometimes it isn't the server that goes down. While it's a very good start, it's not always enough simply to have servers in different parts of the world. The unfortunate truth is that, on thankfully rare occasions, entire networks can lose connectivity to the outside world. The bad news is that it can sometimes take days to get service fully resorted if something catastrophic happens to the network.

The solution to this is to have at least one of your servers on a different network provider than the others. Don't worry too much about this point right now, because we'll discuss it later.

Having multiple servers also allows you to do something else that can be very useful – load balancing. If one of your servers is experiencing heavy traffic for whatever reason (like we tend to after releasing a new issue), traffic can be routed to one of your other servers which is less busy. This means that, instead of having to wait or make multiple attempts to connect to your site, a customer gets access to your site hosted on a different server

with little delay.

HOW DO I DO THIS?

There are a couple of ways to obtain co-location for your servers. One way is to set up another office at a separate location and hire staff for it in order to manage your servers. As you might expect, this option is pretty expensive.

The other option is to pay another company for dedicated hosting and co-location services. This saves you the cost of having full time staff in multiple locations, though you still need to buy your own servers. It does however come with a great deal of choice and a few pitfalls to overcome.

Granted, you could also go with shared hosting (where you share the server that you use at the hosting company with some of their other customers) to save even more money, but that comes with a whole other set of problems. The first of these is that it isn't your server. You only have so much space on it and the rest is reserved for the other customers. You also have to hope that what they are using the server for isn't too processor or memory intensive. The upshot of this is, unless you're really strapped for finances, it's usually better to get your own dedicated server.

First, you need to find a company to host your server. When checking out prospective hosting companies, the first thing you want to do is make sure that they're actually up and running. While it may be funny to find hosting companies who can't even keep their own servers on-line, you really don't want to use one for obvious reasons.

Next, try digging around a little on-line and see if you can find any reviews of them. Google is really your friend in this, because seeing the praise and complaints that other customers have had for the company will give you a better idea of what to expect. Remember to take all reviews with a grain of salt, because not all of them are honest, but the chances are if you see nothing but bad reviews for the company, it's best to look elsewhere.

After you've looked around for reviews of your potential hosting company, get in touch with them. Ask them for a test IP address and get it in writing that the address is on the same network as the service that you want. Perform a traceroute on the IP that you've been given to make sure that it doesn't take too many hops to get to.

Now it's time to get in touch with the people that will actually be running the server. Before we even get to the physical requirements, ask them what your hardware options are. Do they host both towers and rack mounted servers? If they host rack mounted servers, do they host (or charge more for hosting) 1U, 2U and 3U servers? Do they have deals with server vendors which may save you money on support in the event of hardware failure?

What operating systems do they support? If they're a windows only shop and you have things which have to run on Linux or BSD, then they're really not an option. Make sure that they do a minimal install on the server. One of the main causes of security breaches is having a server that is running services which you aren't using.

Ask them what their power situation is. See if they have adequate battery backups and an on site generator for extended power outages.

Make sure that their network is in good order. Remember that a network can only transmit data as quickly as its slowest part. If they're running a firewall on a 386 box which all traffic has to go through, you're going to have trouble. Also make sure that the number of hops from your server to the outside world isn't too high. If the traffic just goes through a couple of switches, you're fine. If, on the other hand, to paraphrase a colleague of mine, it goes through half a dozen hubs, a hamster running on a wheel, the aforementioned 386, a hot air balloon, and then another hub or two before reaching your server, there's going to be a problem.

Make sure that the fiber they use to connect to the outside world has multiple access points so their network won't become inaccessible in the event of a backhoe accidentally cutting the cable. I know it sounds unlikely, but it can and does happen, and when it happens, it takes a considerable amount of time to fix.

Be absolutely positive that their hosting facility is climate controlled and has appropriate fire suppression systems. Also make sure that it's secured – preferably guarded and in a building with closed circuit cameras and the servers in cages. Nothing can ruin your day faster than a hard drive that's died from the heat unless it's the server being engulfed in flames or someone just completely walking off with it.

It's also important to ask when they have staff there

in case of emergency reboot and how often they make backups of the server. In addition, be sure to ask if they keep a copy of the backups off site in case of fire or flood.

Finally, check for hidden costs. For example, if they offer the server as 5Mbit/sec but have it on a 100MBit/sec port with huge overage charges, it might be worth considering finding a different host.

WHY WOULD I NEED TO BE ON DIFFERENT NETWORKS?

No matter how many precautions are taken, sometimes backbone networks do become inaccessible to other backbone networks. It can be because a backhoe cuts a vital cable or any number of other reasons, but the truth is that it does happen albeit rarely.

The way to prevent this from affecting your business is to have at least one of your co-located servers on a different backbone network than the others. For instance, if all of your servers are on networks supplied by Qwest, it would be a good idea to have one hosted on a network that is supplied by another backbone network like MCI. This means that, even if Qwest disappears off the face of the earth for a few days, your site will still be accessible. This can help prevent a lot of headaches.

DO I NEED A SEPARATE WEBSITE FOR EVERY SERVER?

The short answer is no, you don't need a separate website for every server. You may, however, want to have separate websites for region or language specific content.

For example, commerce carried out in the United States has a whole different set of taxes, shipping requirements, etc compared to commerce that might be carried out in the European Union. In order to minimize the strain on your customers, it would make sense to set up a separate domain (say, .co.uk) for purchases made in that region.

HOW DOES ALL THIS WORK?

I realize that, at the moment, how all of this works together is probably a little fuzzy. After all, we're talking about having servers in different parts of the world that, as far as anyone accessing them is concerned, are all the same machine. They don't want to have to worry about choosing a server closer to them, and you don't see how to get around that.

Don't worry. There's help with that too.

Companies such as Nortel (*www.nortel.com*), F5 (*www.f5.com*), Radware (*www.radware.com*) and Coyotepoint (*www.coyotepoint.com*) sell networking devices that help take care of this problem for you. Basically, the devices they sell for this purpose are DNS servers that reply with a specific site based on the source of the client. This way, if someone from Asia accesses your site, they get sent to the server closest to them instead of the server in the United States or wherever it is you happen to be.

Their products are fairly simple and painless to get up and running and some of them can even handle the network load balancing that was discussed earlier.

If you want to save yourself the cost of buying an appliance to do the job of making this all work, you can do it yourself with the language and country data in web browsers and open source tools like mod_rewrite.

Any way you look at it, unless you're just running a personal site or one that isn't very important to your business, it's a good idea to have it set up in such a way that there is no single point of failure and that it can be accessed quickly from any part of the world in which you might do business. You can easily save more than the cost of the co-location due to increased uptime.

While this hasn't been an overly in-depth look at the problem and its solutions (that would take entire books), I hope it has been a useful overview and has helped point out why and how you should distribute your web presence around the world. If you're still in doubt as to whether or not you should distribute your web presence, honestly consider the loss your company would experience in the event of the catastrophic destruction of your web server verses the fairly small cost of a dedicated co-located setup. You'll probably find that the second option is a lot cheaper.

James Hollingshead is Executive Editor of o3 magazine. James can be reached via email -- james@o3magazine.com. If you are interested in writing an article, providing feedback or have any comments or suggestions regarding o3 or this article. Please feel free to contact James. If you liked this article or discovered an error with the content, please let James know directly rather than commenting on it privately in your blog.

FREE ADVERTISING

This publication was started to help companies, consultants and other IT decision makers make informed decisions when it came to selecting Open Source solutions for Enterprise Data Networking problems.

Sometimes the most innovative technology doesn't make it due to a lesser technology having better marketing. To help innovative companies deliver their products to the hands of IT decision makers, we are please to announce a new advertising program to help innovative small businesses.

Whether your an established small business or simply a developer looking to promote their own project, you now have the opportunity to promote your products to over 500,000 readers in over 140 countries.

Each advertisement that we receive will be placed into a pool, each month we will randomly select at least three advertisers. Whenever there is space at the end of an article, we will also tap the free advertisment pool.

To submit your advertisments, please send in JPEG format at 300 dpi, no compression:

Column
3.35" (wide) x 9.00" (high)

Square
4.00" x 4.00"

Advertisements should be sent to sales@o3magazine.com with a Subject containing *o3 small ad pool submission.*

Requirements
All applicants must have an annual sales revenue under US$1,000,000 and have under 100 employees.

o3 magazine reserves the right to refuse submissions based on content and quality.

Performance Hosting
quality service

Whether you are promoting a new business venture or expanding your existing one, you need a solid and secure hosting partner.

Why settle for less?

BLACKNIGHT SOLUTIONS

www.blacknight.ie sales@blacknight.ie +353 (0)59 9137101

Linux on Big Iron

EVERYONE KNOWS THAT LINUX IS REVOLUTIONIZING IT DATA CENTER OPERATIONS

A WELL HIDDEN SECRET IS THAT LINUX IS REVOLUTIONIZING THE IBM MAINFRAME "BIG IRON"

BY DAVE JONES

By now, it is very well known that Linux is revolutionizing IT data center operations, capabilities and procedures in a wide range of environments, from SMB (small and medium business) settings to Fortune 100 organizations. What is not so widely known is that Linux is also revolutionizing the IBM mainframe ("big iron") environment as well.

There are many reasons for this, but three important ones are:

- Linux can take optimal advantage of the classical mainframe strengths
- Linux allows for significant server consolidation using mainframe virtualization
- Linux support of on-demand e-business initiatives that are growing in importance to large organizations

We'll take a look at each of these reasons in turn, starting with the classic mainframe strengths.

MAINFRAME STRENGTHS

The S/390, zSeries and now the newest z9 family of hardware have been IBM's flagship enterprise system offering for decades. This family of systems has an unparallel record for high-availability, reliability and security for supporting mission-critical systems and data across a wide variety of enterprises and applications. Some of the unrivalled strengths (strengths that are just now seeing a renewed focus of development effort in competitive offerings) of the zSeries are it's availability, scalability and manageability.

To support the high availability characteristics found in the zSeries design, all major system components are replicated as a standard feature, providing automatic recovery capability and automatic switchover to spare components without interrupting system operation. Most major components can be serviced concurrently with normal system operations, limiting the amount of time spent in unscheduled outages. The newest member of IBM's mainframe offerings, the z9-109, actually allows a complete "book" of four processors to be concurrently removed from the server and reinstalled during an upgrade or repair without affecting the operation of the other installed processor books or of the system itself. Redundant I/O interconnections between the processor books provides connectivity to I/O resources on other books during a processor's removal.

In today's large IT data centers, it's not uncommon to find a number of systems working together in tandem to support an organization's critical business data processing needs. Multiple zSeries systems, both physical and virtual, can be monitored, controlled, and maintained from a single central point. Since everything that runs on a server does not share equal priority to the business, zSeries systems allows a site to manage the relationships of various transaction types, the interdependencies and change management in a complex environment.

zSeries processors, I/O channels and devices, and communications interfaces are available in a number of configurations designed to support the requirements of a few tens of users to thousands of concurrent users processing data from the megabyte to the multi-terabyte range. Processors can be incrementally upgraded or replaced to meet growing demands, allowing the business to quickly and non disruptively adapt to changing business needs and requirements.

The On/Off Capacity on Demand capability of the IBM zSeries processors is designed to provide even greater flexibility by allowing IT data centers to turn on additional, temporary system resources to meet the demands of business cycles or unexpected demand throughout the year, and then turn them back off when they're no longer needed. This can help IT departments control costs while meeting

unpredictable, or transient capacity needs.

Recent events in the rapidly expanding on-demand e-business economy have highlighted the continuing importance of these design criteria – where IT is not only a peripheral component of the business but is the core business as is the case in established e-business originations such as eBay and amazon.com. These factors, coupled with IBM's substantial reduction in the cost of zSeries component systems over the last 20 years (for example, the copper on silicon chip technology) are making a significant impact into the much-touted "cost savings" of alternative platforms.

Because of Linux's modular and well designed structure, it can very easily be adapted and tuned to take advantage of the many strengths of the IBM zSeries. Architecture-specific open source patches to the Linux kernel, provided by IBM and commercial distributors like Novell SuSE and RedHat, now allows Linux to:

- Utilize the hardware cryptographic accelerators
- Share applications in memory via an "execute in place" file system
- Share parts of the kernel among different Linux virtual guests, reducing the total amount of real memory required
- Utilize z/VM support for virtual disks and DCSS memory areas as very high performance swap devices
- Produce performance data that can be processed and reported on by performance monitoring software

SERVER CONSOLIDATION

Server consolidation rests on the mainframe's ability to easily and quickly create virtual processors, communications, storage and I/O devices, thus helping to reduce the overhead of planning, purchasing and installing new hardware to support new workloads. Unlike many of the popular processors in use today, the zSeries processor and its instruction set have been designed from the outset to support efficient and fast virtualization. The virtualization technology for zSeries is composed of two elements, a hardware element – the processor, its memory and I/O subsystems, and a software element -- the z/VM operating system. Both of these elements are considered at the time the zSeries server is being designed; zSeries virtualization is not an afterthought, but is built in from the beginning, and virtualization is not a new concept to the zSeries systems.

Linux running on IBM's zSeries servers has brought new meaning to the term server consolidation. The zSeries platform has long been recognized for the ability to scale to support consolidation of diverse workloads onto zSeries servers. In the past this was typically "vertical" scaling, or put another way, getting a more powerful processor to contain the workloads of less powerful processors. Linux adds another dimension to that. Now, with IBM's premier virtualization operating system, z/VM running on a zSeries processor, "horizontal" consolidations are possible as well. By supporting the creation of many virtual application servers on single zSeries servers, IT sites now have the advantage of being able to deploy solutions using the familiar "single server / single application" model, while taking advantage of savings in floor space, power, maintenance effort, and networking complexity by consolidating onto a single zSeries server.

ON-DEMAND E-BUSINESS SUPPORT

On demand e-business can be defined as building responsiveness into every part of IT. This can be accomplished by building an IT infrastructure that can support rapid, but controlled, changes to central business objectives. Linux on the zSeries can be a central part of the foundation of an on demand e-business environment, because it combines the industrial strength scalability, security and reliability of the IBM zSeries with the flexibility and open standards of the Linux operating system. This can help to achieve objective and measurable results for the IT organization.

An on demand e-business infrastructure has these characteristics:

- It is integrated so applications and processes can interoperate across platforms
- It is open so IT organizations have the flexibility

to run applications on the platforms that make the most sense

- It is virtualized so it can help improve utilization rates, realize cost-efficiencies and leverage existing assets

- It is autonomic so less human intervention is needed to manage the system

Linux on the IBM zSeries helps provide a proper foundation for building an on-demand e-business structure. The Linux environment on the IBM zSeries mainframe is designed to provide the following capabilities:

- Infrastructure simplification through virtualization for rapid deployment, configuration and management of virtual Linux servers. Virtualization can be provided either by z/VM or by zSeries' LPAR capabilities. z/VM provides a way to support hundreds to thousands of Linux guests, while up to 60 Linux images can be supported on the new z9 system (up to 30 LPARS supported on older zSeries systems).

- Business integration through open and industry standards, fast data access, resource sharing, and system utilization efficiency to easily integrate large amounts of data and their applications.

- Robust and strong security features built-in from bottom to top: hardware, virtualization, operating system, applications

- Automatic systems management for rapid and dynamic responses to a wide variety of changing workloads, while providing hardware utilization efficiencies.

Running the Linux operating system on an IBM zSeries mainframe, either directly in an LPAR or as a guest of z/VM, is a smart choice. In today's intensely competitive on-demand, e-business world, putting Linux on the zSeries means that Linux can transparently take advantage of the strong IBM support for its mainframe hardware architecture and it's reliability, availability and serviceability (RAS) features. Coupled with IBM's premier virtualization engine, z/VM, Linux can provide an IT organization with the best of both worlds, the robustness, security and reliability of the IBM mainframe and the wealth of cutting edge applications and tools available in the open source environment., as well as applications from a number of leading software vendors.

DAVE JONES IS A LEADING IBM ZSERIES EXPERT. DAVE CURRENTLY WORKS FOR V/SOFT SOFTWARE BASED OUT OF HOUSTON, TEXAS. DAVE CAN BE REACHED BY SENDING EMAIL TO DAVE@VSOFT-SOFTWARE.COM.

Networking on the IBM zSeries

CONNECTING THE MAINFRAME TO THE NETWORK.....

IS FAR MORE COMPLEX THAN SIMPLY TURNING UP AN ETHERNET INTERFACE

BY DAVE JONES

With Linux now running ever more mission critical and business centric applications on IBM's new class of zSeries mainframes, getting the information from those applications out on the net and to the people that need it rapidly and reliably is important. This article will take a look at how the zSeries mainframes can be physically connected to networks and how industry standard network connectivity is implemented. Future articles will cover how these physical connectivity options can be virtualized and used by a large number of Linux systems running on the mainframe.

PHYSICAL NETWORK CONNECTIONS

There are four ways a zSeries mainframe can support physical connections to networks and other systems:

- Open Systems Adapter-Express (and it's antecedent, the Open Systems Adapter-2)
- Common Link Access to Workstation (CLAW) interface
- Channel-to-Channel Adapter (CTC)
- HiperSockets

Let's take a closer look at each of these, in reverse order.

HIPERSOCKETS

HiperSockets is IBM Licensed Internal Code which runs on both standard and Integrated Facility for Linux (IFL) processors in both 31-bit and 64-bit environments, as well as with the new zSeries Application Assist Processor (zAAP). It is part of z/Architecture technology including QDIO and advanced adapter interrupt handling to jump start message processing and minimize the frequency and overhead associated with I/O interrupts. The data transfer itself is handled much like a cross address space memory move, using the memory bus, not the Self-Timed Interface I/O bus. On z890, z990 and z9-109 processors, spanned channel support allows sharing of HiperSockets across multiple Logical Channel SubSystems (LCSS) and multiple LPARs. HiperSockets is designed to minimize contention with other system I/O activity; it does not use CPU cache resource, and thus has minimal effect with other activity in the zSeries server.

Currently LP-to-LP communication is typically done through some type of external TCP/IP network, such as ESCON-attached external devices or open systems adapter. HiperSockets provides “Network in the Box” functionality that allows high-speed any-to-any connectivity among different operating systems images within the zSeries mainframe server without requiring any physical cabling. This “Network in the Box” concept minimizes network latency and maximizes bandwidth capabilities between Z/VM, Linux for zSeries and Z/OS images (or among combinations of these) to enable optimized ebusiness and ERP solutions within a single server. These images can be first-level (such as, directly under a LPAR), or second-level images (such as, under VM or VIF). Up to four separate Cluster LANs can be configured within a server thereby allowing OS images to be grouped according to the function they provide. These groupings are independent of sysplex affiliation.

HiperSockets can be thought of as “internal LANs” for the zSeries. It is application transparent and appears as a typical TCP/IP device to the operating system software. HiperSockets provide very fast TCP/IP communications between servers running in different logical partitions (LPARs) on a zSeries machine. The z890, z990, and the newest z9-109 processors support up to 16 HiperSockets. The z800 and z900 processors support up to four HiperSockets.

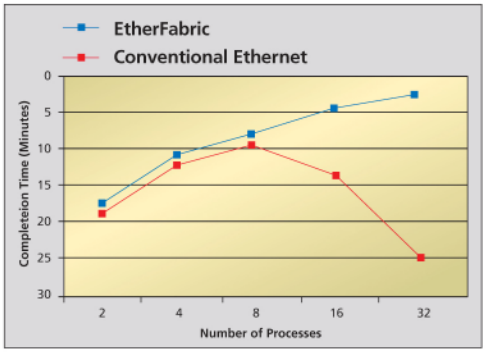
To communicate between servers running in the same zSeries Central Electronics Complex (CEC),

ACCELERATE APPLICATION PERFORMANCE!

>> HALF THE LATENCY

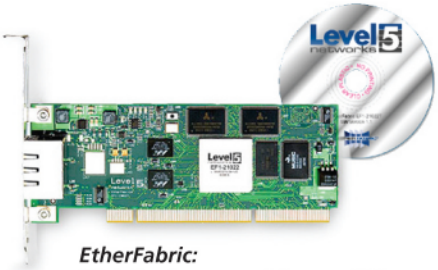
>> TWICE THE BANDWIDTH

>> 4X THE PERFORMANCE



Take EtherFabric for a ride today and experience the accelerated performance for yourself.

Visit **www.level5networks.com/landing/3.php** and take advantage of our limited time offer to ship you one extra EtherFabric NIC with your initial order.



EtherFabric:
High Performance Ethernet NIC

Level 5 networks

Level 5 Networks Inc. • 840 West California Avenue, Suite 240 • Sunnyvale, California 94086 • 408.245.9300

HiperSockets sets up I/O queues in the zSeries processor's memory. The packets are then transferred at memory speeds between the servers, thereby totally eliminating the I/O subsystem overhead and any external network latency.

HiperSockets implementation is based on the OSA-Express Queued Direct Input/Output (QDIO) protocol; therefore, HiperSockets is called internal QDIO (iQDIO). HiperSockets is implemented in microcode that emulates the Logical Link Control (LLC) layer of an OSA-Express QDIO interface. Although HiperSockets is a type of virtualization technology, it relies on zSeries microcode to run, and for the purpose of this article, it will be categorized as a physical networking option.

Typically, before a packet can be transported on an external LAN, a LAN frame has to be built, and the MAC address of the destination host or router on that LAN has to be inserted into the frame. HiperSockets does not use LAN frames, destination hosts, or routers. TCP/IP stacks are addressed by inbound data queue addresses instead of MAC addresses. The zSeries server microcode maintains a lookup table of IP addresses for each HiperSocket. This table represents an internal LAN. At the time a TCP/IP stack starts a HiperSockets device, the device is registered in the IP address lookup table with its IP address and its input and output data queue pointers. If a TCP/IP device is stopped, the entry for this device is deleted from the IP address lookup table.

HiperSockets copies data synchronously from the output queue of the sending TCP/IP device to the input queue of the receiving TCP/IP device by using the memory bus to copy the data through an I/O instruction. The controlling operating system that performs I/O processing is identical to OSA-Express in QDIO mode. The data transfer time is similar to a cross-address space memory move, with hardware latency close to zero.

HiperSockets operations are executed on the processor where the I/O request is initiated by the operating system. HiperSockets starts write operations; the completion of a data move is indicated by the sending side to the receiving side with the sending side executing a Signal Adapter (SIGA) instruction.

Optionally, the receiving side can use dispatcher polling instead of handling SIGA interrupts. The I/O processing is performed without using the System Assist Processor (SAP). This new implementation is also called a "thin interrupt". HiperSockets does not contend with other system I/O activity and it does not use CPU cache resources; therefore, it has no performance impact with other activity in the server.

The HiperSockets operational flow consists of five steps:

1. Each TCP/IP stack (image) registers its IP addresses into HiperSockets' server-wide Common Address Lookup table. There is one lookup table for each HiperSockets LAN.

2. The address of the TCP/IP stack's receive buffers are appended to the HiperSockets queues.

3. When data is being transferred, the send operation of HiperSockets performs a table lookup for the addresses of the sending and receiving TCP/IP stacks and their associated send and receive buffers.

4. The sending processor copies the data from its send buffers into the target processor's receive buffers (zSeries server memory).

5. The sending processor optionally delivers an interrupt to the target TCP/IP stack. This optional interrupt uses the "thin interrupt" support function of the zSeries server which means the receiving host will "look ahead," detecting and processing inbound data. This technique reduces the frequency of real I/O or external interrupts.

CHANNEL-TO-CHANNEL ADAPTER (CTC)

Channel-to-channel (CTC) is a point-to-point connection, using real hardware channels. CTC technology can be used to interconnect different physical servers, logical partitions, or both. Because all zSeries operating systems use the same link protocol, it is possible to connect a Linux server not only to another Linux image, but also to z/VM and z/OS TCP/IP stacks. CTC support exists for a number of channel IBM standard technologies including ESCON and FICON® channels.

ESCON CTC CONNECTIVITY

To connect two systems using ESCON, two separate channels are defined. The ESACON CTC

connections can either be point-to-point or switched point-to-point (that is, they can be connected to an ESCON director).

LPARs can share channel paths, and so optionally, they can share any control units and associated I/O devices configured to these shared channels. Sharing channel paths means that the number of physical connections between processor complexes can be reduced. This also helps reduce the amount of under-floor cable space needed.

FICON CTC CONNECTIVITY

Channel-to-channel communication in a FICON environment is provided between two FICON (FC) channel control units. There are several differences between the ESCON and FICON CTC implementations as shown here:

Number of required channels
ESCON: At least 2
FICON CTC: 1 or 2

Channel dedicated to CTC function
ESCON: Yes
FICON CTC: No

Number of unit addresses supported
ESCON: Up to 512
FICON CTC: Up to 16384

Data transfer bandwidth
ESCON: 12-17 MBps
FICON CTC: Up to 2 Gbps

Number of concurrent I/O operations
ESCON: 1
FICON CTC: Up to 32

Data transfer mode
ESCON: Half duplex
FICON CTC: Full duplex

It is not recommended to use ESCON or FICON CTCs as networking connectivity options for zSeries mainframe systems. For inter-LPAR communications, HiperSockets or OSA-Express are a much more robust choice.. For communications inside a single z/VM LPAR, virtual VSWITCHs should be considered.

Although CTC bandwidth is good (particularly FICON Express), CTC connectivity is less fault tolerant than other solutions. Often, if one side of the link has a problem, one or even both of the systems have to be re-IPLed in order to restart the CTC link. For communications between the zSeries machine and other systems in the network, OSA-Express Gigabit Ethernet or OSA-Express 1000BASE-T adapters should be used.

COMMON LINK ACCESS TO WORKSTATIONS (CLAW)

Common Link Access to Workstation (CLAW) is a point-to-point protocol. A CLAW device is an ESCON channel-attached device that supports CLAW protocol. These devices can be used to connect a Linux for zSeries, z/OS or z/VM system to another system, for example, a pSeries processor or a Cisco Channel Interface Processor (CIP) card.

CLAW devices are “old technology” and are not as efficient or reliable as some other solutions discussed in this article. Instead, for communications between Linux and other systems in the network, use OSA-Express Gigabit or 1000BASE-T adapters, if at all possible.

OPEN SYSTEMS ADAPTER (OSA-EXPRESS AND OSA-2)

The IBM Open Systems Adapter-Express adapter family consists of integrated hardware features that are designed to provide direct connection for zSeries and S/390 Parallel Enterprise Servers G5 and G6 to high speed routers and switches, to other high speed servers, and to clients on local area networks (LANs).

OSA-EXPRESS

The OSA-Express feature plugs into a zSeries or S/390 I/O slot just like a channel card, providing a direct, peer-to-peer network connection. OSA-Express consists of multiple different hardware adapter types supporting a variety of networks: Gigabit Ethernet,1000BASE-T Ethernet, Fast Ethernet, 155 Mbps ATM, and the 4/16/100 Mbps Token Ring.

All feature types can use IBM's Queued Direct I/O (QDIO) architecture to help eliminate the need for channel control words (CCWs) and interrupts, resulting in accelerated TCP/IP data packet transmission and more efficient TCP/IP stack processing. QDIO also enables dynamic configuration of the adapter TCP/IP addresses, and

offloading of functions like MAC handling, packet filtering and ARP function. QDIO supports TCP/IP only.

The 1000BASE-T Ethernet, Fast Ethernet, Token Ring, and ATM features also support the non-QDIO operating mode, which is designed to provide support for TCP/IP and SNA protocols similar to the support provided by the prior generation OSA-2, but at the higher performance levels of the faster OSA-Express hardware. For example, using non-QDIO mode, the new 1000BASE-T Ethernet can support both TCP/IP and native SNA traffic at up to Gigabit speeds.

Specific to the 1000BASE-T Ethernet feature, the new OSA-Express Integrated Console Controller (OSA-ICC) function is designed to provide up to 120 console session connections for z890 and z990 Initial Program Load and z/OS, z/OS.e, z/VM, VSE/ESA, and TPF operational consoles.

Each OSA-Express card has one port on G5 and G6 servers and two ports on zSeries servers and can be attached directly to a LAN or ATM network. These cards are recognized by the hardware I/O configuration as one of the following channel types:

- OSD (Queued Direct I/O)
- OSE (Non-Queued Direct I/O

The OSA-Express card on the zSeries 990 processor operating in QDIO mode can support up to 160 separate TCP/IP stacks and 480 devices per port.

QDIO MODE

Queued Direct I/O (QDIO) is a highly efficient data transfer mechanism. It reduces system overhead and improves throughput by using system memory queues and a signaling protocol to directly exchange data between the OSA-Express microprocessor and TCP/IP stack.

The QDIO-enabled OSA-Express adapter has a much shorter I/O instruction path length compared with the OSA-Express adapter in non-QDIO mode (which has the same I/O path length as the OSA-2 cards). Consequently, when running in QDIO mode, I/O interrupts and I/O path lengths are minimized. When running in QDIO mode, measurements have shown that there is a significant improvement in performance versus non-QDIO mode, in particular, a reduction of System Assist Processor (SAP) utilization and improved response time.

The TCP/IP stack(s) of each operating system (z/OS, z/VM, zLinux and z./TPF are all supported) that shares a port on an OSA-Express card in QDIO mode dynamically registers all of their IP addresses with the card. Whenever IP addresses are deleted from or added to a network stack, the device drivers download the resulting IP address list changes to the OSA-Express card.

The OSA-Express QDIO microcode assists in IP processing and offloads the TCP/IP stack functions in the following areas: a) multicast support, b) broadcast filtering, c) building MAC and LLC headers, and d) ARP processing. Offloading the processing of these functions to the PowerPC ® processors that make up the OSA-Express adapter means that CP cycles are freed up to do other work. In a single guest, the effect might not be significant, but in a z/VM LPAR with Linux guests generating a moderate-to-high volume of network traffic, there can be great savings.

Checksum processing calculates the TCP/UDP and IP header checksums to verify the integrity of data packets. This function is usually performed by a host system's TCP/IP stack. OSA-Express cards on the z990 and z890 processors have the ability to perform checksum processing on behalf of the upstream TCP/IP stack using a function called checksum offload. This function is only available for IPv4 packets. By moving the checksum calculations to an OSA-Express Gigabit or 1000BASE-T Ethernet card, host CPU cycles are reduced. This support is available with z/OS V1R5 and later and Linux for zSeries.

NON-QDIO MODE

When running in non-QDIO mode, a port on the OSA-Express card is defined as channel type OSE.

In non-QDIO mode, the data follows the same logical I/O path as an OSA-2 card. Linux uses the LCS device driver to communicate with the device when it is running in this mode. The non-QDIO mode requires the use of of the IBM OSA/SF program product for customization of the OSA-Express if the adapter is to be shared across multiple LPARs or Linux z/VM guests. The OSA-Express 1000BASE-T, FENET, and token-ring cards support both non-QDIO and QDIO modes. The OSA-Express Gigabit Ethernet card only supports QDIO mode.

Unless there is a specific site requirement for non-QDIO mode (such as supporting SNA traffic), OSA-Express adapters should always be run in QDIO mode.

OSA-2

The Open Systems Adapter-2 (OSA-2) card was designed to provide direct, industry-standard network connectivity for the S/390® server. The OSA-2 card supports ATM, Ethernet, FDDI, and token ring.

OSA-2 cards, introduced in 1995, use the Interconnect Controller architecture and thus are not as efficient as OSA-Express cards running in QDIO mode. The OSA-2 Fast Ethernet card, under the best conditions, has a maximum bandwidth of 100 Mbps. The OSA-Express Gigabit Ethernet and 1000BASE-T cards have, by comparison, a maximum bandwidth of 1000 Mbps. With that in mind, our recommendation is that you move to OSA-Express Gigabit Ethernet or 1000BASE-T.

If the hardware environment supports only OSA-2 technology, directly connecting the Linux LPARs or Linux z/VM virtual guests to the OSA-2 adapters(s) will be the most efficient means of accessing the network. The alternative approach is to provide connectivity to the Linux guests through z/OS, z/VM, or a Linux guest acting as an intermediate router. That router machine "owns" the OSA-2 interface and the "back-end" Linux systems are connected to that router through a virtualization technology (virtual CTC, IUCV, or Guest LAN). Although this router can provide added functionality such as packet filtering and VPN support, it also adds latency and extra CPU overheard to the environment.

One point to note, however, is that OSA-2 cards can only support 16 IP addresses per port. This limits the number of Linux guests or LPARs that can share the port.

Dave Jones is a leading IBM zSeries expert. Dave currently works for V/Soft Software based out of Houston, Texas. Dave can be reached by sending email to dave@vsoft-software.com.

(IN)SECURE

Open. Informative. To the point. (IN)SECURE Magazine is a free digital security magazine discussing some of the hottest information security topics.

// www.insecuremag.com //

Real Time Streaming Protocol (RTSP)

THE REAL TIME STREAMING PROTOCOL (RTSP) IS UTILIZED BY A NUMBER OF INTERNET MEDIA TECHNOLOGIES FROM VOIP TO MEDIA STREAMING

BY RAJA HAMMAD

Internet video streaming is a demanding and challenging task that has evolved tremendously in the last couple of years. Streaming video over the Internet involves six key areas: video compression, application-layer QoS control, continuous media distribution services, streaming servers, media synchronization mechanisms and streaming protocols. Moreover, video streaming applications have stringent requirements in terms of end to end delay and require high bandwidth, low packet loss, video-cassette (VCR) like functionalities and low decoding complexity.

Protocols for video streaming can be categorized into network layer protocol, transport protocol and session control protocol. Network layer protocol, such as IP, provides basic network services whereas transport protocol, such as UDP, TCP, real-time transport protocol (RTP) and real-time control protocol (RTCP) provides end to end network transport services. Session control protocol, on the other hand, defines methods to control delivery of multimedia data during an established connection.

RTSP is a session control protocol, at the application layer, to stream multimedia data over the Internet. While it does not deliver multimedia data itself, it provides a framework to support VCR-like operations such as stop, pause, play etc, and supports tunneling RTP traffic to work around with firewalls.

RTP provides media packetization to deliver data over the Internet and RTCP provides QoS management statistics to RTP whereas RTSP is responsible for controlling the delivery of data. RTSP is designed to be independent of the transport mechanism and thus does not rely specifically on RTP. The protocol can control multiple data delivery sessions, provides means to choose delivery channels (such as UDP, multicast UDP or TCP) and delivery mechanisms based upon RTP and is designed to work for multicast as well as unicast.

The protocol is similar in syntax and operations to HTTP but differs in many ways such as:

- RTSP is a stateful protocol as compared to HTTP
- Both client and server can issue requests as opposed to HTTP where client issues requests and the server responds.

One of the main functions of RTSP is to establish and control either a single or multiple time-synchronized continuous audio and video media streams between the server and client. Following operations are supported:

- Media retrieval: The client can request a presentation description to setup an RTSP session to send the requested media.
- Invitation of a media server: A media server can be invited to join an existing conference to play back a presentation or to record the presentation.
- Addition of media to an existing presentation: The server can notify the client about any additional media that may become available during an established session.

A client sets up a RTSP session with the server to retrieve media. A session, maintained by the server, typically consists of setting up a transport mechanism for the media stream, streaming the media and closing the stream. During a session, many transport connections, either TCP or UDP, can be opened up by the client to issue RTSP requests.

RTSP streams are defined by a presentation description. The protocol defines a notion of presentation which is a complete media package, presented to the client, and may consist of one or more streams. In most of the cases, this means controlling multiple streams using a single timeline by the server (e.g. audio and video streams controlled by a single operation such as pause). The information about the streams within a presentation is specified

presentation description file. This information may include set of encodings, network address and information about the content. The presentation description file can be obtained by the client using HTTP, email or other means. Each presentation (set of media streams) or a media stream is identified by an RTSP URL. For example

rtsp://example.host.com/matrix/audio

defines the audio stream within the presentation "matrix" whereas the RTSP URL

rtsp://example.host.com/matrix

defines the presentation "matrix", which may be composed of audio and video data.

RTSP may use a different protocol such as TCP, compared to the one used for data delivery (such as UDP), to control a media stream. This means that data delivery is independent and continues even if there is no RTSP communication between the server and client.

RTSP is now an industry standard protocol. Work is in progress to revise and correct the flaws in the current standard, and for this an updated version RTSP 2.0 has been proposed.

SETTING UP AN RTSP SERVER AND CLIENT

For the purpose of this article, I used Apple Darwin Streaming server and openRTSP client to demonstrate RTSP streaming over RTP.

APPLE DARWIN STREAMING SERVER

Apple Darwin Streaming Server (DSS) is the open source version of Apple's QuickTime Streaming Server (available under Apple Public Source License (APSL)) that allows streaming media over the Internet using RTP and RTSP. DSS supports streaming QuickTime and MPEG-4 media and can deliver live as well as on demand media. It works for both multicast and unicast network transport to stream media, and it can also be configured to act as a relay where it listens to an incoming stream and forwards it to one or more destinations.

Installation requirements can be found in the official documentation or in the "readme" file bundled with the source code. I installed DSS on Fedora Core 3. In the following section, I will be writing the exact steps that are required to install and set up DSS for streaming. [The scripts that come with the source code may or may not work on your system or require little tweaks to fix broken paths].

1. Download DSS source code from *http://developer.apple.com/darwin/projects/streaming/*. This step requires registration.

2. Untar the source file and cd to the newly created source directory.

3. Run the script buildtarball. This should compile the package. The script will take a few minutes, depending on the machine. You should see "Success!" at the end unless you get errors.

4. Move to newly created "DarwinStreamingSrvr -Linux".

5. Run the Install script. There are two versions of the Install script, one under the root directory of source code and the other one under "DarwinStreamingSrvr-Linux". Make sure you are in "DarwinStreamingSrvr-Linux" or you might get errors. The Install script will ask for the administrator username and password. You should see "Setup Complete!" at the end of a successful installation.

6. Under the directory "DarwinStreamingSrvr -Linux", there is streamingadminserver.pl script that can be used to start the server. Moreover, it also initiates the streaming server administration interface. Run this script.

7. The administration of the server can be managed by entering the URL http://hostname:1220 into your browser. The screen will prompt you for the username and password entered earlier.

8. Next, another screen will appear that will ask for the MP3 password. This password will be used for streaming MP3.

9. The next screen will provide an option to enable SSL support. For the purpose of this setup, you can safely ignore it.

10. The next screen will present you with the default path of the streaming media folder. You can change it to whatever you want.

11. The next screen will provide an option to bypass firewalls. You can ignore it for the current scenario.

After finishing the whole process, you will see the server administration interface. At the top of the screen, you should see a message "Server is running". You can browse through other admin option available.

Now the streaming server is configured and ready for streaming the media. DSS comes up with sample streaming media, including hinted(1) mp4 files that can be found in the source directory. These files are also copied, while installing the server, to /usr/local/movies. You should be able to stream sample videos by using any media client, which supports the sample files format, by entering the RTSP URL e.g. rtsp://hostname/sample_100kbit.mp4.

(1). Hint tracks contain information to stream media properly by the streaming servers. Refer to DSS official guide for more information.

OPENRTSP

Live555 Streaming Media (*http://www.live555.com*) is a set of C++ libraries for streaming media over the Internet. The libraries are based on RTP, RTCP, RTSP and SIP and can be compiled for different platforms. Some popular applications such as liveCaster, Mplayer, and VLC are already using these libraries. The source code is available under LGPL and includes some sample programs to test the streaming setup. OpenRTSP is a command line utility, bundled with the libraries, that can be used to open, stream, receive and (optionally) record media streams. The program itself does not play the streams but can be used to forward them to another application.

The libraries can be compiled easily by following the instructions on the website. I compiled and installed it on both Red Hat Enterprise Linux Server 3.0 and Fedora Core 3. Following are the steps required to be working with OpenRTSP:

1. Download the libraries from *http://www.live555.com/liveMedia/public/live.2006.01.05.tar.gz* and uncompress and untar it. This will create a directory "live".

2. Under the directory "live", enter the command ./genMakefiles linux. This will generate necessary Makefiles.

3. Run make to compile the source code.

That's it. To test the library and sample programs, the binary executables can be run directly from their respective directory. These test programs can be found under "testProgs" directory.

For the next part of this article, I going to assume that DSS is not running right now. To see statistics while streaming, I would recommend not using streamingadminserver.pl script and run the server manually. To do this, enter the following command under the "DarwinStreamingSrvr-Linux".

```
# ./DarwinStreamingServer -d -S 10
```

The -d switch will force the server to run in the foreground whereas -S will update the statistics every 10 seconds.

I am also going to assume that sample movie files which comes with DSS are placed under the /usr/local/movies. Now it's time to test the setup using the OpenRTSP client. Move to "/root/rtsp/live/testProgs" and enter the following command.

```
# ./openRTSP \
      rtsp://server_host_name/sample_100kbit.mp4
```

You will see a lot of messages on the screen. These are basically what the client is communicating with the server over RTSP. At the same time, you should see streaming traffic by looking at the statistics running on server. OpenRTSP will not play anything on the screen, since it's not a media player; instead it will write the received data into an output file. The openRTSP client will retrieve all the subsessions (audio/video) and write them into a separate file. So in the above scenario, we should have "video-MP4V-ES-1" and "audio-MPEG4-GENERIC-2".

The output files can be played independently by a media player or can be combined to regenerate the complete presentation. Alternatively, any media player, that supports the mp4 format can stream and play. For instance, MPlayer can be compiled to play the media streams by enabling the Live media support. This can be done by passing --enable-live and –with-livelibdir=<path to live media library> option. In order, to stream your own mp4 files, you may have to hint them before streaming. Please refer to DSS official document for further details.

Live Media library comes with another useful tool, onDemandRTSPServer, under the "testProgs" directory that can be used as a test server to stream media file. All you need to do is rename the media file (for instance, "test.mpg" as the server expects) and put it under the same directory from where the server is running. Using openRTSP we can now stream the media files.

OPEN SOURCE RTSP IMPLEMENTATIONS

RTSP is a standard protocol for streaming media and has been a part of many streaming applications. Following is a brief overview of some popular streaming solutions:

MPEG4IP is an open source package, licensed under Mozilla Public License, and provides a streaming server and client. It integrates a bunch of other open source packages and can work well with DSS. The project is intended for developers to take advantage of the video/audio streaming and is not meant for an end user. It is primarily targeted toward Linux but has been tested for freeBSD, Solaris and Windows. MPEG4IP installation installs an additional handful tools that can be used for encoding, creating hint tracks, etc.

Helix platform, known as Helix DNA , is a digital media platform supporting media streaming over the Internet using open standard protocols such RTP and RTSP. The platform is developed by the Helix community and sponsored by RealNetworks and is available under both open source and commercial licenses. The solution can be deployed to a diverse set of platforms such as desktops, mobiles and set-top boxes and supports many open and proprietary formats. It offers both a server and a media player.

VideoLAN is another open source, GPL Licensed, product for streaming video over high bandwidth networks and supports a large number of multimedia formats. It is available for multiple platforms including Linux, Windows, Solaris, freeBSD, NetBSD etc and offers two different flavors of softwares: VideoLAN Server (VLS) which can stream multimedia data and VideoLAN Client (VLC) which can be used as a streaming server and a client.

Other popular open source solutions for media streaming include Peercast, Icecast, liveCaster, and Vovida RTSP Stack.

REFERENCES

http://www.ietf.org/rfc/rfc2326.txt

http://sourceforge.net/projects/rtspspec

http://tools.ietf.org/wg/mmusic/draft-ietf-mmusic-rfc2326bis/

http://tools.ietf.org/wg/mmusic/draft-ietf-mmusic-rfc2326bis/draft-ietf-mmusic-rfc2326bis-11.txt

http://developer.apple.com/darwin/projects/streaming

http://www.live555.com

http://mpeg4ip.sourceforge.net

http://www.videolan.org/

http://www.peercast.org/

RAJA HAMMAD IS THE GENERAL MANAGER OF ADVANCED DATA NETWORKING SOLUTIONS AT SPLICED NETWORKS LLC. HE IS BASED OUT OF PAKISTAN.

Mandriva

Businesses need rock-solid IT solutions

**Mandriva Linux Corporate Server & Corporate Desktop offer outstanding robustness, scalabilty, and reliability.
All with the ease of use specific to Mandriva products.**

- Full IT solution for server and desktop deployments
- Open standards
- Both x86-32 and x86-64 architectures are supported
- 5-year product maintenance
- 24/7 support
- Mandriva Online update service - Professional Level
- Incredible price

http://www.mandriva.com/business/corporate-server

http://www.mandriva.com/business/corporate-desktop

www.mandriva.com

Introducing dNMS

DNMS IS AN UPCOMING OPEN SOURCE PROJECT DESIGNED TO PROVIDE A DISTRIBUTED NETWORK MANAGEMENT SYSTEM (NMS) BASED ON RUBY, SNMP AND POSTGRESQL

BY JOHN BUSWELL

This issue we are looking at dNMS. Unlike past issues, we are focusing on an emerging open source technology that has great potential rather than an existing project. The dNMS project is a distributed Network Management System designed to provide a top down web based management and monitoring system for global IP networks and the devices contained within those networks. The project is led by a special engineering team at Ohio based Spliced Networks LLC. Scheduled for an initial beta release later this quarter, dNMS looks to be an exciting new project.

WHAT IS DNMS ?

The short answer to this question is that dNMS enables you to configure, monitor and manage anything on your network that can understand SNMP and has an IP address regardless of which part of the network it resides on. The system is also capable of monitoring non-SNMP speaking devices as well. Advanced capabilities such as high-performance real-time statistics and status information are possible with certain devices.

The system utilizes a combination of existing technologies - SNMP, Syslog and SQL. Similar to the thinking behind new technologies such as AJAX, dNMS takes these existing technologies and ties them together in an interesting manner. The idea behind dNMS provides a highly customized role based view of your network. What you see depends on your access privileges and your administrative role on the network. This customized view of the network gives you access to the logs, events, status and configuration options that are relevant to your job.

HOW DOES IT WORK?

The technology behind dNMS is relatively straightforward. SNMP, Simple Network Management Protocol, is an industry standard management protocol used for sending and retrieving management information from devices on an IP network. In dNMS, SNMP is used to configure and poll devices over a secure private management network, which most organizations have deployed on their networks in some form or another. SNMP v1 and v2 are currently supported, with SNMP v3 support planned for later this year.

On a network which has deployed dNMS, each system would send SYSLOG information to the dNMS server or pool of dNMS servers (depending on the size of the network). This information is taken by the dNMS server, parsed and stored into an SQL database. How data is logged (whether it was something useful or simply background noise) is configured by the administrator. Nothing is discarded; the information tagged as less than useful is logged to a database where data is stored on a short-term basis, instead of being intended for archival.

The dNMS system adds IP based monitoring into the mix. This is a highly customized solution that enables an administrator to use a wide variety of health checks as well as customized application tests to insure their systems are working correctly. Not only can dNMS check to see if your DNS server is responding correctly, it can check for changes in returned data, and cross reference those changes with configuration changes made on the network. If data has been changed without a verified configuration change through dNMS, the administrator is alerted.

The dNMS project is a collection of applications, which are controlled by a web based application written in Ruby. Currently the project supports a number of Linux based products including AppOS based hardware appliances from Spliced Networks. Support for more devices is being added on a daily basis.

FEATURES

The dNMS project provides a distributed network management system, while it has been designed to be

highly scalable and enable hundreds of management servers to work in unison, it is also versatile enough to run on a single management server. The following is a list of its features.

- Network Device / Server Configuration via SNMP
- Network Device / Server Statistics collection via SNMP
- Network Device / Server Monitoring
- Advanced Application Health Monitoring
- Report Generation
- Log parsing and archival
- Event generation
- Report generation
- Autonomous Event Handling
- Role based views and tasks
- Network Wide Traffic Assessment
- Configuration Archival

THE GLOBAL VIEW

At the very top, we have the global view. This view is only significant to enterprises with more than one physical site such as Spliced Networks (which has three sites based in Ohio, sites in California and Florida, and one site in the UK). This view shows the bandwidth and service utilization at each site, the status of links and services, as well as that information over time, and any critical security issues. All of this information is displayed in the global view, which is a type of web based dashboard. From this dashboard, the top level administrator has the capability of zooming in to the finest degree of detail about a single server at one location, as they step down through the views. Critical issues are hot-linked from this global view.

SITE VIEW

This is the top level view for single-site networks and provides a more detailed, site based look at that particular portion of the network. Anything done at the site level is specific to that site and doesn't effect the other sites. The site view contains more detailed information, access to switching and VLAN information, as well as site-centric stats and monitoring information.

NETWORK VIEW

The network view provides a detailed look at a particular VLAN or group of VLANs that the administrator has grouped into a “Network”. Typically, a site consists of at least one network, but often has many. Here the administrator has access to network level firewalls, the ability to track trouble spots on the network, and the ability to view individual systems on that network.

SYSTEM VIEW

The system view is specific to a particular device and it is often the lowest level view configured within dNMS. Here, particular system level details such as memory, cpu and network interface utilization are available. This information is often broken down into more detailed segments such as per application utilization. The system view is also where an administrator can make all the necessary changes to a particular system.

APPLICATION VIEW

A special view called an Application view which provides more fine grain control than the System View is also available. The application view focuses on a particular service such as FTP running on the system. However, the application view is special because while it belongs to a System View, the same application on a number of systems can be used to provide Network, Site and Global application views, which allow service based configuration changes and monitoring to occur across the enterprise.

TASK QUEUING

In a large organization, tasks which might be performed by the same person at smaller businesses are often broken down into tasks performed by different groups of individuals belonging to different technical groups within a company. The goal with dNMS task queuing is to permit each individual to complete their task without relying on another task being completed first. Task Queuing allows the administrator to do their job, run sanity checks, then place the task into the queue.

Once the task is in the queue, a manager determines the order in which the tasks are completed in a special project view and oversees the automated deployment of those tasks. The goal with this particular feature is to eliminate the often time-consuming and costly practice at many major corporations of holding day-long conference calls to complete relatively simple tasks which, due to the

wide variety of skill levels within the organization, become bogged down in red-tape and pointless tests.

TASK ESCALATION

Similar to Task Queuing, Task Escalation is intended for a lower level administrator investigating a particular network problem to have the ability to prioritize and send their work upstream for further investigation once it involves systems or networks they do not have access to.

SAFETY LOCK-OUTS

While dNMS offers a type of buffer zone between a device and its administrator, the safety lock-outs provide an advanced feature within the management system. The safety lock-outs enable a senior administrator to define specific boundaries which may prevent a junior administrator from performing tasks that they would normally have the capability of carrying out. If used correctly, the senior administrator can use the safety lockouts to prevent human error from resulting in costly downtime.

An example of a safety lock out might be that a junior administrator has the capability of restarting the production web services between the times of 9pm and 6am EST. The company has started to do business in Europe, so the administrator doesn't want the junior administrator to restart the services outside of a maintenance window or above a specific traffic threshold. With dNMS, the administrator can define the threshold, and enforce it on individual administrators or collective groups of administrators.

STATISTICS AND REPORTING

The bottom line of any management system is to obtain statistics from the network and build reports so that management can justify specific IT spending such as increasing bandwidth, purchasing additional servers or adding an additional site. The dNMS project collects stats and identifies problem areas on the network automatically, based on either factory preferences or customized thresholds placed by the administrator. These stats and logs are then compiled from the database and used to generate a wide range of reports which can be created automatically on a scheduled basis and emailed to whomever in the company needs to see them.

In addition to the usual stats and reporting capabilities, dNMS has a feature called DST (Direct Stats Transfer). The DST solution is a separate application which is integrated into AppOS 2.0 and will be available under the GPL. DST accepts SNMP commands to group specific sets of stats or logs, or dump all stats / logs, directly to a database.

The concept is fairly simple - there are often a large number of stats and logs that need to be transferred. To poll this information via SNMP is inefficient since it can result in hundreds of SNMP requests and responses. Instead, DST works by having the management system send a couple of encoded SNMP requests to the system. These SNMP requests instruct the remote system to create a results set, where to send it, what to send, whether its reoccurring or not, and the format to send it in (typically the SQL table format for the data).

These sets are typically setup once, and then one SNMP command is used to trigger the DST dump. The remote system then talks directly to the database. Once the data is dumped to the database, the management system is free to poll the database to produce stats. In a traditional system, the data is polled via SNMP, dumped to a database and then reloaded from a database to produce results.

AVAILABILITY

The current version of dNMS is 0.3.0 and a publicly available release of dNMS is planned for later this quarter. The dNMS project will be available from http://www.splicednetworks.com later this quarter and its availability will be announced in an upcoming issue of o3.

DEVELOPERS

If you are familiar with SNMP, Ruby and Postgresql, and are interested in becoming part of the development team working on the GPL version of this project, please contact John Buswell via email at jbuswell@splicednetworks.com

JOHN BUSWELL IS CO-FOUNDER AND CTO OF SPLICED NETWORKS LLC. JOHN IS BASED A FEW MILES OUTSIDE OF ATHENS, OHIO. HE CAN BE REACHED VIA EMAIL, JBUSWELL@SPLICEDNETWORKS.COM.

CANADA ON RAILS

YVR06

DAVID HANSSON
THOMAS FUCHS
DAVE ASTELS
DAVID BLACK
JOE O'BRIEN
JAMES ADAM
STEVEN BAKER
MICHAEL BUFFINGTON
ROBBY RUSSELL
GEOFFREY GROSENBACH
KYLE SHANK
JEREMY VOORHIS
ALEX BUNARDZIC
SEBASTIAN KANTHAK
AMY HOY

VANCOUVER, BRITISH COLUMBIA

ARPIL 13-14, 2006

TICKETS ON SALE NOW
www.canadaonrails.com

Intrusion Detection Server Load Balancing

OFTEN A SINGLE IDS SERVER IS SIMPLY NOT ENOUGH TO HANDLE THE BANDWIDTH UTILIZATION ON A NETWORK. INTRUSION DETECTION SERVER LOAD BALANCING OR IDSLB OFFERS A SOLUTION TO THIS PROBLEM

BY JOHN BUSWELL

We've spent the last two issues looking at the Open Source Intrusion Detection System, Snort. This month we're using Snort in unison with a commercial enterprise product to provide an advanced Intrusion Detection solution.

Sometimes the technology is simply not there to achieve the desired solution exclusively with Open Source. Sometimes the Open Source alternative is simply not mature enough or the solution requires specialized hardware to achieve the desired capacities which are only available through commercial solutions. Through a combination of Open Source software and commercial products, its possible to achieve a superior solution at a substantially lower cost than using commercial products exclusively.

Intrusion Detection Systems work by analyzing network traffic and using techniques such as pattern matching to predict events that are alerts to potential security breaches or activity that is a prelude to an attempted security breach. Intrusion Prevention Systems work by identifying the attack in progress and updating other security systems such as firewalls and access lists to block the attack in real time.

These systems often play an important first line of defense role within a network, and it is desirable to provide a degree of redundancy. While basic redundancy can be provided by simply adding a second Intrusion Detection System, that does not resolve the problems larger networks face where a single system might not have the physical bandwidth or the CPU capacity to process the large amounts of data involved. This is where IDS Load Balancing comes into play.

The Nortel Application Switch line of products can provide a wide range of Layer 2-7 intelligent switching capabilities for a network. For the purpose of this article, we will be using a Nortel Application Switch 2424 with two Linux servers running Snort. We will look at two solutions, one using IDS Load Balancing on the 2424 with Snort to provide an advanced IDS solution and the other using Firewall Load Balancing on two 2424s with Snort to provide an Intrusion Prevention System.

HOW IDS LOAD BALANCING WORKS

The Nortel Application Switch achieves IDS load balancing by forwarding a copy of the IP packets to IDS servers. The IDS servers are placed in what is called a Real Server Group. A server group is simply a collection of servers, with a metric assigned to that group. A metric is a special algorithm that tells the switch how it should select a server when a packet is received. The IDS SLB feature is enabled on inbound ports on the switch and enabled for the server group that the IDS servers are assigned to.

WITH OR WITHOUT IP ADDRESSES

The Nortel Application Switch supports two methods of sending data to a group of IDS servers. The preferred method is without an IP address where the servers imply connect to the switch and monitor all traffic that is passed down to that server interface. The other method is to provide each interface on the server with a dummy IP. The latter offers the option of saving ports on the Nortel Application Switch, as the servers maybe connected through another Layer 2 switch to the Nortel Application Switch. For the rest of this article, we assume the use of the IP-less method.

HEALTH STATUS

The Nortel Application Switch performs what is called a health check. If you're familiar with the keepalived project (*http://keepalived.sourceforge.net*), it is a similar concept. The switch will check periodically that the server is still active, and if the server fails a health check, it is removed from the server group until it has successfully passed a number of health checks, at which time it is added back into the group.

The 2424 offers a wide range of health checks, however for IP-less IDS load balancing, you have two

options. If you have each server plugged directly into a dedicated port on the 2424, then the Link health check method can be used. The Real server ID must be within the first 26 as well in order for that to work. The second method is to use the SNMP health check to test the status of a port on a remote switch. Here you must have the server plugged into a switch capable of SNMP, alternatively you could setup net-snmp and write a small agent (in Ruby or Python for example), to respond back with the appropriate SNMP response monitoring the status of the snort process on the server. This is a good example of how Open Source can be used to extend the functionality, because without it, only the link up/down status would be monitored, leaving the possibility of packets being forwarded to a system where the IDS process has crashed.

CONFIGURING SNORT

You configure Snort in the normal manner. There is no special configuration, you simply provide a dedicated interface. In our configuration, we used eth1 directly connected to the Nortel Application Switch.

CONFIGURING SINGLE GROUP IDS LOAD BALANCING

Configuring IDS load balancing on the Nortel Application Switch is relatively simple. While the switch can be configured via SNMP and Browser Based Interface as well as a separate management application from Nortel, we're looking at the command line interface for this article. The system uses a menu / prompt style CLI. In this example, we're going to configure IDS load balancing to send all inbound data to a single pair of IDS servers running Snort.

First we configure the IDS servers as real servers. The CLI requires that we give the configuration dummy IP addresses - this is a safe-guard to make sure the real servers haven't been accidentally configured without an IP address. We have selected an unused, un-routable network of 10.255.255.0/24 for this purpose. These IP addresses are not configured on the Snort servers themselves unless you were using the alternative IP based health checking method.

```
>> # /cfg/slb/real 10/rip 10.255.255.10/ena
>> # /cfg/slb/real 11/rip 10.255.255.11/ena
```

Next we add the IDS servers to a new group which must be within the first 63 groups. Here we've used group 20, adding real server 10 and 11 that we just created above:

```
>> # /cfg/slb/group 20
>> Real Server Group 20# add 10
>> Real Server Group 20# add 11
>> Real Server Group 20# metric hash
```

The last command tells the switch we want to use the hash metric. The hash metric uses a hash algorithm to force traffic from the same source IP to always go to the same server. Thus, the hash algorithm ensures continuity. IDS Load Balancing only works with this hash metric. Next we define the health check as link, as we have real 10 and real 11 plugged into ports 10 and 11 respectively on our 2424 switch.

```
>> Real Server Group 20# health link
```

Next, the important commands. We enable IDS load balancing and select all traffic to be sent to this group:

```
>> Real Server Group 20# ids ena
>> Real Server Group 20# idsrprt any
```

Now our inbound ports are 1 and 24 respectively. Port 1 has our router to the Internet, and port 24 is connected to our upstream LAN switch. On the Nortel Application Switch, we must enable IDS load balancing on these potential ingress ports :

```
>> # /cfg/slb/port 1/ids ena
>> # /cfg/slb/port 24/ids ena
```

Finally, we must create a filter to redirect the traffic to the IDS. The filter provides a mechanism to provide a finer grain of control in more complex configurations involving multiple groups of IDS servers. The commands below set the source and destination IP to any, sets the filter action to allow, and enables the filter. The last command enables idshashing on both the source and destination IP address for a packet to make sure it always goes to the correct server.

```
>> # /cfg/slb/filt 100
>> Filter 100# sip any
>> Filter 100# dip any
>> Filter 100# action allow
>> Filter 100# ena
>> Filter 100# adv/idshash both
```

Then the filter is applied to the ingress ports 1 and 24 respectively:

```
>> # /cfg/slb/port 1
>> SLB Port 1# add 100
>> SLB Port 1# filt ena
>> SLB Port 1# /cfg/slb/port 24
>> SLB Port 24# add 100
>> SLB Port 24# filt ena
>> SLB Port 24# apply
>> SLB Port 24# save
```

The last two commands commit the configuration and makes it active.

TESTING IDS LOAD BALANCING

Now that everything is configured, before we roll it into production we need to make sure that it is working correctly. To do this, simply run tcpdump against the IDS interface on both of your IDS/Snort servers. We simply ran tcpdump -i eth1 on both servers, as eth1 was our interface connected to the Application Switch. You can use ip link eth1 up to make sure the link is up as well before starting the test. Next, we connected a Dell 6350 running Gentoo to port 1 (10.1.2.3) on the switch to simulate Internet traffic, and another to port 24 to act as the local LAN (192.168.1.100). For the purposes of testing, we started lighttpd on the LAN side and ran Scapy on the Internet side to simulate traffic.

```
Welcome to Scapy (1.0.2.34beta)
>>> a =
IP(dst="192.168.1.100")/TCP(seq=0,sport=2100,dp
ort=80)
>>> a.src = "10.1.2.3"
>>> send(a)
.
Sent 1 packets.
>>>
```

We then monitored the tcpdump on the IDS server side for the packet, as well as tcpdump on the LAN side (192.168.1.100) to make sure the switch was also only sending a copy of the packet to the IDS group.

The switch has a command /info/slb/idshash which is a useful tool to determine what source and destination IP address pair will result in a hit on a specific IDS server. Using that tool, we quickly put together a second pair in Scapy to make sure that traffic would hash to both servers.

To roll the system into production we simply start Snort on the servers and plug the production Internet router and LAN switch into ports 1 and 24 respectively.

ADVANCED IDS LOAD BALANCING CONFIGURATIONS

So far we've just looked at sending all ingress traffic into the switch to a single set of IDS servers. However, The Nortel Application Switch allows for a much finer grain of control. Lets say a large organization has a specific Web Security Group, and a Network Security Group. Here it might be desirable to send all HTTP traffic to a set of IDS servers managed by the Web Security Group while other traffic is routed to the Network Security Group's IDS pool. To do this, we simply add another set of real servers and a new group to our original configuration:

```
>># /cfg/slb/real 15/rip 10.255.255.15/ena
>># /cfg/slb/real 16/rip 10.255.255.16/ena
>># /cfg/slb/group 30/add 15
>># /cfg/slb/group 30/add 16
>># /cfg/slb/group 30/metric hash
>># /cfg/slb/group 30/health link
>># /cfg/slb/group 30/idslb ena
>># /cfg/slb/group 30/idsrprt http
```

Here you will see the configuration process is almost identical except for the idsrprt http command where, instead of any, we are specifying http (port 80).

To test this configuration, we took our first snort server and plugged eth2 into port 11, and used eth1 and eth2 on the second snort server to ports 15 and 16. Essentially, our second snort server was now the Web Security Groups IDS pool. Next we simply passed port 80 traffic from Scapy, and checked tcpdump on eth1 and eth2 on the second snort server. The port 80 traffic only shows up on that server. Next

we test some SMTP (port 25) traffic by changing the port number in Scapy, and starting postfix on 192.168.1.100. The forwarded copy of the SMTP traffic only shows up on the first IDS server simulating real 10 and 11, so things are working as expected.

FURTHER ADVANCED CONFIGURATION

A wide range of configurations is possible through manipulation of the filter attributes and the real group idsrprt command. The filter options also contain an idsgrp option which enables you to specify a filter to a specific IDS group, allowing traffic from or to a specific subnet, for example, to be routed to a specific server group. By setting the idsgrp option in the advanced filter configuration, you can also do per VLAN load balancing of IDS traffic.

INTRUSION PREVENTION LOAD BALANCING

Snort in-line uses iptables rather than libpcap, allowing snort to manipulate traffic in real time as it is passed through the system. In this case, Snort in-line acts more like a firewall than an IDS server. So IDS Load Balancing won't work as its dealing with a copy of the forwarded packets rather than the real traffic. The Nortel Application Switch has a capability called Firewall Load Balancing. This advanced feature set can be used with Snort in-line to achieve Intrusion Prevention Load Balancing.

Basic Firewall Load Balancing (FWLB) works by having a “dirty” (Internet side) and “clean” (LAN side) of the Firewalls. A Nortel Application switch is placed on both sides, so this advanced configuration requires at least two Nortel Application switches. Each firewall is connected to the dirty side, and a second interface connected to the clean side.

Packets traverse the dirty side switch, are passed to the firewall, and may or may not make it to the clean side depending on the firewall. In our IPS solution, we utilize Snort in-line instead of a regular firewall, allowing for deeper packet inspection, and prevention of intrusions in real time.

CONCLUSION

This article demonstrates how Open Source software in combination with commercial networking solutions can provide a cost-effective solution for large scale network problems.

Nortel Application Switch 2424

WHICH SWITCH?

The current Nortel Application Switch line of products includes the 2208, 2216, 2424, 2424-SSL and 3408.

The 3408 is an 8 port Application Switch with GBIC slots on ports 3 thru 6. These GBIC slots provide dual media (Fast Ethernet and Gigabit) on ports 3 thru 6. The 3408 also has standalone GBIC ports on 9 thru 12.

The 2424-SSL is a 24 port Application Switch, it is similar to the 2424 but has the addition of a built-in SSL accelerator module.

The 2216 and 2208 are 16 and 8 port application switches respectively. The 2216 and 2208 have reduced capacities compared to the 2424 and 3408. The 2208 and 2216 are excellent products for small and medium enterprises who do not have the available budget for a 2424 or 3408.

For more information on Nortel Application Switches, contact your nearest Nortel partner or visit *http://www.nortel.com*.

Nortel Application Switch 2208

spring
 conference 2006

SEOMUG SOUTHEAST OHIO MACROMEDIA USER GROUP

ITAAO INFORMATION TECHNOLOGY ALLIANCE OF APPALACHIAN OHIO — A Member of Ohio's IT Alliance

OHIO UNIVERSITY 1804

Adobe

Thursday, March 23, 2006

The Midwest's Premier One Day Designers & Developers Event!

"Spring Break is one of the better user group regional events in existence. The speakers and sessions are second to none and the conference price of $25 is unmatched in the industry."

Ed Sullivan, Program Manager, Developer Relations Adobe Systems

Produced by the Southeast Ohio Macromedia User Group
In partnership with the IT Alliance of Appalachian Ohio, Ohio University and Adobe Systems

Admission is $25, which includes a Lunchtime session by Tim Buntel in which we provide your lunch!

Online Registration Beginning Feb 1st or Tickets At The Door **Website: www.seomug.org**

Presenters Include: (name / topic)

- **Phillip Kerman : Flash**
- **Rob Gonda : AJAX**
- **Carolyn Snyder : User Experience**
- **Dan Dura : ActionScript**
- **Glenda Vigoreaux : Dreamweaver (Hands On Sessions)**
- **John Cummings : ColdFusion**
- **Neil Ross : Development Patterns**
- **Joe Lowery : Dreamweaver**
- **Gary Kraeger : Enterprise Email**

Many More

Keynote by Tim Buntel
Product Manager
Adobe Systems

Check the Website beginning Feb 1
www.seomug.org

Tracks Include:

- **Design & User Experience**
- **Programming & Development**
- **3D Gaming & Simulations**
- **Adobe & Related Products**
- **Conference bookstore with author signings**
- **SEOUG Community Suite**
- **Wireless Access Provided**
- **Email Cafe**

Prizes - Prizes - Prizes - Prizes

Watch the website:
Prizes for early bird registration:

www.seomug.org