

---

# The 17th LISA Conference (2003)

---

by **Paul Anderson** <dcspaul@inf.ed.ac.uk>  
**Jeremy Olsen** <J.Olsen@ed.ac.uk>

School of Informatics  
University of Edinburgh

## General

This year's conference in San Diego was rather overshadowed by the very bad fires in southern California. The following picture was taken before the conference started, but shortly afterwards, ash was falling like snow from sky and the air was choking. The air was bad for the rest of the week, making it very unpleasant, even indoors.



**Before the smoke arrived**

However, the conference was bigger than last year with 1347 participants; 86% were from the USA, and 65% were from commercial organisations. There seemed to be a considerable growth of interest in configuration issues, which made the conference interesting for me. There was also quite a bit of talk about usability issues which I have been interested in recently. (*paul*)

This was my first conference and I found it really interesting, not least the opportunity to stand back and get a perspective on what we are doing compared to other people. As Iain said last year, there are probably papers to be done even if you don't think what you are doing is that exciting/ground-breaking. Even if you don't produce a paper I think it is valuable experience and we ought to be pushing for enough budget to get 2 funded people per year there.

For those not familiar with how the conference is organised see the Appendix. (*jeremy*)

This is personal (biased) view of the things we attended, and anything we found particularly interesting is marked  $\Rightarrow$ .

## Workshops

### Advanced Topics Workshop

The advanced topics workshop is a forum for a group of experienced systems administrators to talk about their current concerns, things they have discovered, and future predictions. This is always valuable for the collective view of current issues, and for the many small tips, but it is virtually impossible to document. However, here's some things I found interesting (*paul*)...

- It is now common in big, long meetings to see people working on laptops, but I hadn't noticed laptops being used before as a "back-channel" to the main meeting, by using IRC to make asides and communicate between subgroups during the meeting! This was explicitly acknowledged, and the IRC log published as part of the record of the meeting. There was talk of experimenting next year with some special software to allow voting, managing the "hand up" queue for people wanting to speak, and recording of comments during the meeting.
- People seemed to be more concerned this year with "soft" issues, including (independent of the configuration workshop discussions), the cultural gap necessary to manage systems in a higher-level way, rather than "hacking" individual machines. There were also related discussions about users who wanted to manage their own machines.
- There were long discussions about "charging" for sysadmin support which didn't seem relevant at first until I realised that it does have implications, even in our context; some users may get much more than their fair share of support, possibly without realising how much it really costs. This evolved into a number of related issues such as the "cost of not doing things" and the value of having responsive local support.

- ❑ There was a general agreement that people seems to be under more pressure and there was less opportunity for people to take on “fun” jobs.
- ❑ I initiated a discussion on directory services, and we polled to see who used what:
 

NIS	48%
NIS+	12%
LDAP	56%
AD	35%
Flat files	42%
Kerberos	27%

OpenLDAP was considered unreliable, and the IPlanet LDAP server somewhat better. Nobody was attempting anything like our degree of replication and hence hadn't had major problems with this. Most people using NIS wanted to get rid of it and many assumed that LDAP was “the accepted way to go”.

- ❑ *Nobody* else is really attempting to do managed laptops.
- ❑ Significant numbers of people were considering moving from Redhat to Debian (or Mandrake).

### Configuration Workshop

I organised this. Thirty-two people present (and quite a few turned away because we were full) with mixed experience. However several people had a good understanding of important configuration issues, and many interesting topics were discussed, notably “usability” of configuration systems, and distributed management of configuration specifications. A more detailed summary of this, including slides of the presentations, is available at <http://homepages.inf.ed.ac.uk/group/lssconf/config2003/>. (paul)



**Configuration workshop**

## Tutorials

### T8: Next Generation Backup and Restore

⇒ Half-day session given by Jacob Farmer of Cambridge Computer Services <http://www.cambridgecomputer.com>, a vendor-independent storage/backup company. He has been in the storage business around 15 years so knows his subject well. It was a fast-paced session (180 slides in 3 hours). I've listed a variety of interesting/useful points for general digestion. (I have hard copy of the slides but they aren't online.) (jeremy)

- ❑ if storage doubles both backup capacity *and* throughput have to double in order to just stand still.
- ❑ TCP offload engines are becoming popular; a network interface card (“TNIC”) that handles the TCP processing rather than load the CPU
- ❑ backup host I/O processing is frequently the bottleneck
- ❑ SANs help significantly, allowing data to go direct to backup devices. However, not the complete answer; need a mixture of LAN/SAN for cost and throughput effectiveness
- ❑ iSCSI (SCSI-over-IP) is going to make things a lot easier (see next tutorial).
- ❑ tape and drive technology matters *significantly* and is very dependent on the particular situation.
  - ❑ helical scan/variable speed drives (eg, Exabyte Mammoth, Sony AIT) can perform significantly better than linear/fixed speed drives (eg, DLT, LTO), despite lower official transfer rates, due to lack of “shoe-shining”: the need for the tape to stop and re-position if data is not available.<sup>1</sup>
  - ❑ the problem is getting worse with faster drives: LTO-2 drives need 17.5MB/s and above to avoid shoe-shining. Allowing for compression of 40% or better you need at least 25MB/s of uninterrupted data.
  - ❑ LTO can work well in SAN environments where data is available at high-speed but you have to be careful about individual hosts grabbing a tape drive and using it slowly.

<sup>1</sup>Farmer cited the case of a large installation suffering from poor performance which was being blamed on the s/w. They simply turned off 26 of the 30 LTO drives and performance rocketed!

- ❑ through experience, they are in favour of AIT (AIT-3 is 100GB native, 12MB/s). SuperAIT (SAIT-1, 0.5TB, 30MB/s) is around and has moved to the same, one spool format as LT0/DLT (rather than AITDAT/Exabyte format of two spools). (The AIT roadmap goes up to SAIT-4: 4TB, 240MB/s!).
- ❑ Data provided by tape manufacturers isn't always that useful; eg, MTBF calculated for the human body is 2,078 years!
- ❑ media is getting better and more intelligent; some tapes are fitted with memory that allows re-positioning to be done without having to actually read the tape, preserving both tape and tape head longer. ("Dual tape path"; across the head for read/write; retracted for fast forward/rewind.)
- ❑ beware tape drives boasting a large data cache, usually a sign of it being needed to solve throughput problems.
- ❑ you can achieve better throughput by interleaving hosts (eg, 4 hosts in parallel) but be aware of slower restore (eg, only 25% of the data on a tape is for a particular host). Alternative is to have staging disks.
- ❑ due to cheapness of disk sub-systems hierarchical storage management (HSM) is making a comeback. Easier to deploy in an environment where the organisation dictates the data requirements.
- ❑ have enough tape drives so that you can restore without having to stop backups (your schedule may allow this anyway).
- ❑ consider backing up email separately for ease of restore. There's an increasing need to search for/restore email in litigation cases.
- ❑ expect to have a tape:disk capacity ratio of between 4:1 and 10:1 for a usable backup system.
- ❑ there is increasing use of block-level technologies (snapshots, replication, etc) as part of the data security model.
- ❑ the changes in storage topology by using SANs with a combination of FC and iSCSI means there are now *many* different ways to do things. That gives much greater flexibility but also adds complexity and a need to be aware of inter-dependencies.

(We were also kept entertained in this session by frequent howls of laughter coming from the next room, a session called 'Time Management for System Administrators'. Draw your own conclusions :-)

### T11: iSCSI and IP Storage Networking

⇒ Also by Jacob Farmer. Salient points below. (I have hard copy of the slides but they aren't online.) (*jeremy*)

- ❑ conventional NAS (Network Attached Storage) is file-level data, accessed by NFS, SMB, etc over Ethernet, token ring, etc.
- ❑ by contrast, SAN (Storage Area Network) is block-level data, accessed by SCSI, FCP (SCSI over fibre), ATAPI over SCSI, FC and ATA interfaces.
- ❑ iSCSI: SCSI over IP either through dedicated host bus adaptor (HBA) or iSCSI router box (SCSI interface on one side, Gigabit or faster Ethernet interface on the other).
- ❑ FC topologies: point-to-point, arbitrated loop (FC-AL) and switched fabric (in principle just like ethernet).
- ❑ FC devices pretty much work out of the box; eg Vendor A's disk subsystem works fine with Vendor B's HBA.
- ❑ FC routers/switches usually need to be same vendor to avoid issues due to lack of standardisation with resolving device names (each device has a globally unique WWN, a vendor-assigned worldwide name, same principle as MAC addresses).
- ❑ unlike SCSI addressing using ID and LUNs, FC devices cannot be subdivided below device level.
- ❑ iSCSI is very attractive as it can make use of Ethernet infrastructure with all its benefits: mature, ubiquitous, cheap (relatively), good interoperability, etc.
- ❑ across Gigabit Ethernet can get around 125MB/s; 70-80MB/s more typical.
- ❑ there are FC configuration issues to know about such as zoning
- ❑ As with backups, there are now many more ways to organise the infrastructure so it can get quite complicated.

## M8: Mac OS X Security

By Leon Towns-von Stauber, <http://www.occam.com>, this was mainly an overview of the way OS X is using security in the services it offers, particularly the OS X Server product. Notable changes have occurred in Panther (10.3, which only officially came out 3 days before the tutorial). (Slides, updated for Panther, are available at <http://www.occam.com/osx>, as well as for Leon's OS X System Administration tutorial.) (*jeremy*)

- ❑ in general OS X seems to be doing things well according to the standards and incorporating/using them where possible (eg, relatively easy integration of GPG with the Apple Mail tool)
- ❑ Panther supports much better integration of authentication/authorisation offering single-signon via Kerberos for AFP (Apple File Protocol), SMB/CIFS and FTP.
- ❑ Panther ships with Samba-3 and can provide NT Domain controller services or integrate with AD.
- ❑ OS X can reshare NFS over AFP tunnelled in ssh, allowing Mac-based users to securely mount their home partition via AFP on a per-user authenticated/authorised basis.
- ❑ perhaps of most (theoretical) interest is the internal *Security.framework*, an infrastructure for security-related functionality for OSX applications, including privileged access, encryption certificate handling, password storage. Each application can have it's own self-contained framework and liaises with 'external' daemons to get passwords and private keys, keeping apps and security separate. The framework is based on an open standard called CDSA (Common Data Security Architecture) and Apple provide a variety of default plugins<sup>2</sup> and other vendors can supply their own. See <http://developer.apple.com/security/> for more information.

<sup>2</sup>Cryptographic Service Provider (CSP): random numbers, encryption/decryption, key generation, hashes, digital signatures; Data Library (DL): file-based storage for certificates, keys, etc; CSP/DL: manages Apple keychains; Certificate Library (X509CL): manage X509 certificates in memory; Trust Policy (X509TP): validity of X509 certificates

## Keynote, Invited Talks (InvT), Guru Sessions, WIP

### eBay Keynote

The keynote speech was by the CTO of eBay, Paul Kilmartin, who described the evolution of eBay's computing facilities. Some very impressive numbers; eBay definitely seems to be at the leading edge of many things in terms of scalability. However, nothing conceptually very interesting (although, unlike other dot-coms, to their credit they have always made a profit.).

- ❑ 85 million users
- ❑ 20 million items per day sold
- ❑ 235 million items on sale
- ❑ \$738 per second of transactions
- ❑ 5800 Mbits/sec bandwidth
- ❑ 500 million page views per day
- ❑ revenue related to eBay transactions makes it equal to 75th largest economy in the world, between Uzbekistan and Dominican Republic.
- ❑ eBay Motors has now sold 1 million cars<sup>3</sup>

I liked the following quote:

*If things seem like they are under control,  
you aren't moving fast enough.*

Clearly the site needs to be very high availability. This is implemented using about 1000 Sun servers and SAN storage, which are completely duplicated at two separate sites for disaster recovery (and data — about 6TB — is also replicated at each site, giving four copies in all). I noted that taking tape backups in this kind of environment is virtually pointless ("Please note: Veritas NetBackup is not called NetRestore!"). (*paul*)

Kilmartin had a lot of complaints about the way vendors don't look after customers and don't respond to their needs or issues. "You have a new firmware patch that no-one else is using? Sure, I'll try it out on my 85 million customers!". eBay are understandably in the position of always being at the bleeding edge, there isn't any other equivalent operation. He encouraged the big players to lean more heavily on vendors to investigate errors better, to test patches more fully and find out what customers actually want before building new hardware. (*jeremy*)

<sup>3</sup>As Kilmartin pointed out, that's why the techies don't get to write the business plan as well; he thought that no-one would be mad enough to buy a car online!

### A Study of System Administrators (InvT)

A very interesting session. A panel of HCI researchers (from HP, IBM, and Sun) presented their recent studies of system administrators at work (I liked the phrase “system administrators in the mist”). Some of the studies were oriented towards improving user interfaces, and others towards improving processes such as problem solving. The amount of (unacknowledged) time spend communicating, and attempting to communicate was a surprising outcome. I was very interested in usability from the point of view of configuration tools; I didn’t learn a great deal which was directly relevant, but it was noted that a distance between the user and “what is going on” creates a lack of trust which may put people off using a tool. I think this is an issue for high-level configuration tools. (*paul*)

(This point was also made at the configuration workshop: developers tend to expect complete trust or faith from people using their tools. Even with good documentation, there is still significant indirection, perhaps several layers, and it takes time to adjust to using tools particularly if it isn’t straightforward to find out the details of how they work or why they did what they did when they did.) (*jeremy*)

### Achieving Success and Happiness in Modern IT Environments (InvT)

A discussion of moving from “firefighting” mode, towards more control and predictability. This sounded promising, but it was biased towards a commercial environment and it didn’t have a great deal of interest. (*paul*)

### Security Lessons from ‘Best in Class’ Organisations (InvT)

⇒ Gene Kim, CTO of Tripwire Inc, co-wrote Tripwire with Gene Spafford at Purdue University in the early 90’s. He talked about having analysed those organisations that have ‘good kung-fu’: operating with predictable service levels and security, resolving issues quickly, well and with fewer resources than other organisations<sup>4</sup>. One of the key problems is that security and operations staff often perceive each other as the cause of their problems. Kim encourages them to realise that active cooperation is far better and much less time-consuming for both.

*Much of the critical knowledge on how*

<sup>4</sup>One organisation studied was the New York Stock Exchange. They work on a maximum permissible service disruption of 2 minutes, on the basis that every outage jeopardises the entire Western economies!

*things really work lives in a few very busy minds.*

Documentation is not a key strength of IT people. In problem resolution, 80% of the time is taken identifying the issue. Predictability and hence prompt fault resolution comes from (a) reducing the variation in configuration of systems by using scalable automated configuration and (b) *very* tightly controlled change management. If configuration is predictable this allows faults to be identified more readily by starting with change management logs and finding the differences between your current system and a ‘golden build’. (Obviously that doesn’t take account of dynamic issues, such as the state of the network, routing, DHCP, DNS, etc.)

Kim is a strong advocate of a British Standard developed by the UK government due to a realisation in the late 1970s that many procedures and operational IT details about nuclear power plants were inadequately documented. These provided the basis for ITIL (IT Infrastructure Library, <http://www.itil.co.uk>) and the associated BS15000 Code of Practice for IT Service Management and BS7799 on Information Security Management. Further links are available via BSI (<http://www.bsi-global.com>) and ITPI, Information Technology Process Institute (<http://www.itpi.org>). ITPI use a benchmarking tool, IMCA, to assess an organisation’s IT stability.

I left shortly before the end to hear Paul present his paper on ‘SmartFrog meets LCFG’. I found this an interesting contrast: it deals with autonomous reconfiguration of systems, eg, in response to a change in resource availability, and is to some extent the antithesis, namely dynamic change management! That leads to an interesting discussion on the degree of predictability/traceability of such changes. Taking Kim’s perspective it seems like there will be a need to provide good feedback, reporting what changes have happened so that prompt problem analysis is still realistic. (*jeremy*)

### Security Without Firewalls (InvT)

Abe Singer manages systems at UCSD Supercomputer Center. They have thousands of systems, shifting 100s of terabytes of data over networks capable of 10s of Gigabits/s. This scenario, with significant external collaboration, largely precludes use of firewalls as the technology isn’t fast enough.

The core of their security model is, again, scalable automated configuration (using `cfengine`) to build Linux, Solaris, OS X and Windows systems. The con-

figuration is therefore tightly controlled, they don't allow root access to users, plaintext passwords are 'prohibited' and they patch early and often. They also run some 'shared fate' networks: you can do what you like and manage your machine how you like but so can anyone else on the same network so *caveat user*. (jeremy)

### Talking to the walls (again) (InvT)

Mark Burgess, author of `cfengine`, gave a thought-provoking talk about technical and sociological issues that may face us as we move towards a world of pervasive computing: powerful computing devices embedded in everyday objects and situations. Society and the economy will adapt in interesting ways to new use of technology and there will be some interesting interoperability challenges to developers. (jeremy)

### Spam mini-symposium

There was a session on spam with an initial overview of the general problems<sup>5</sup>. Adaptive filtering (using Bayesian principles) is a key approach but will not be the final answer. The implementation needs to be at organisational level and also at individual level as no ruleset will match all requirements. Don't use large data sets of spam to train your filtering as that will get a much higher number of false positives.

A number of tricks are being employed by spammers to trick adaptive filtering, mainly focused around the fact that 85% of all spam is HTML-based. Tricks include: (a) including largish blocks of normal text in `text/plain` ("It was really good to see you last night for dinner.") which many mail clients won't display in preference to HTML; (b) using 'invisible ink' (eg, white foreground and background) to hide the normal text and just have the important, eye-catching stuff visible; (c) using font colours as numbers instead of words so you can't obviously tell whether foreground and background colours are very similar (this can be dealt with by using Pythagoras to calculate the colour distance - "See, it was worth learning it at school!") and (d) breaking up known words with HTML comments so ordinary filters don't catch them (you then just need to check for instances of 'via' and 'gra').

One suggested solution to the problem is to allow all spam, we'll all get sick of email and go back to leading nice uninterrupted lives, getting on with more productive things. (jeremy)

<sup>5</sup>"Castration isn't a suitable punishment for spammers as it's not equal opportunity."

### Configuration "Guru" Session

I joined Steve Traugott on the panel for a Q&A session on configuration issues. Steve and I have quite different views on the problem, and the audience had a wide range of experience, but I thought that this was successful in getting over some fundamental issues. (paul)

### Works in Progress

A selection of 5 minute talks on works in progress. The only item I thought worth mentioning was: "Wigwam - A Package for Maintaining Projects". <http://www.wigwam-framework.org/> (paul)

### Literature

Not strictly part of the conference but there was an interesting flyer lying around referencing a paper called "CyberInsecurity: The Cost of Monopoly - How the Dominance of Microsoft's Products Poses a Risk to Security" (<http://www.ccianet.org/papers/cyberinsecurity.pdf>). It was written by a variety of security experts including Bruce Schneier (author of the Crypto-gram newsletter) with the support of the Computing and Communications Industry Association, published in late September and appears to have caused some controversy, including the firing of one of its authors, Dan Geer, from his position as CTO of a security company.

### BOFs

#### Sun N1 BOF

Very difficult to read between the marketing hype and work out exactly what this is. I think it consists of:

- Network visualisation. Using VLAN technology, the network topology can be defined using a GUI (or API), and hosts can be assigned to the various networks.
- Service provisioning. A software packaging system allows a complete (multi-tier) application to be bundled as a single entity with the meta-data necessary to deploy it. Currently, the assignments of processes to nodes (or groups of nodes) is a manual process. The packaging system allows for updates (apply changes only) and machine-specific variation (by using simple variables).
- The future is to do automatic, policy-based provisioning.

This seems to be targeted at the application layer, and it is not clear if it addresses the problems of OS-level services, such as DNS, DHCP, NFS, etc. (although it does manage patches). The ability to specify configurations seems to be very primitive (per-node/group variables applied at deployment time). But I might well have missed the point. (*paul*)

### cfengine BOF

Mostly a social event for cfengine's 10th birthday (including cake). A summary of the new features in the latest version, including "methods" which provide some control over the sequencing of operations. (*paul*)

### Configuration BOF

A well-attended BOF dominated by the usual suspects from the configuration workshop. Rather disconcerting to see the general low-level of discussions, and the primitive tools being used by many people. Ultimately valuable though, as an education for everyone. I came away with some ideas of things that I would like to investigate. (*paul*)

A straw poll showed there were a large number of different methods being used for configuration (*cfengine*, *radmind*, *rdist*, *LFCG*, and at least 10–15 more) with most being relatively simple file manipulators, rather than having any notion of what a system should be doing. There is a suggestion of producing a paper summarising all these methods, as one way to clarify the differences. (*jeremy*)

### Usability BOF

This was mainly focused around large packages/systems (such as Oracle, Tivoli) and their use of GUI-driven administration tools. The standard complaint is that although GUIs are great for those 'once every few months' tasks, they are very poor/positively restrictive for repetitive tasks. Some suppliers are building in the ability to view the actual commands a GUI would generate so a sysadmin can easily include them in scripts. (*jeremy*)

## Technical Papers

### Radmind: The Integration of Filesystem Integrity Checking with Filesystem Management

A program rather like *lfu* which compares a filesystem with some master copy and reports differences as well as selectively repairing them. This is proposed as a way of managing clients, but I think this is extremely

naive. The authors defend this approach against more modern tools on the basis that it doesn't have such a steep learning curve. (*paul*)

### Further torture: more Testing of Backup and Archive Programs

Elizabeth Zwicky (co-author of Building Internet Firewalls) tested various programs (*dump/restore*, *tar*, *cpio*, etc) over 10 years ago under 'extreme' conditions and on different platforms. These latest tests were more than a rerun of those tests, having been made more comprehensive. The tests included file size, block device access, strange names, long names, access permissions and holes. She found that lots of utilities still have problems, particularly with long pathnames and maintaining holes in files (eg, for databases). Overall things have improved but notable advice is that (a) you can't draw conclusions about utilities given their name (eg, Sun's *tar* vs. RedHat's) and (b) the same utility (eg, GNU *cpio*) runs differently on different OS and filesystem types. The best advice was to test the program in the environment you are using. (*jeremy*)

### An Analysis of Database-Driven Mail Servers

Email is nicely structured and lends itself well to storing in a DB. The comparison was from a client (user) rather than server (sysadmin) perspective. They looked at UW-IMAP/mbox, Courier IMAP/maildirs, Cyrus IMAP/Berkeley DBM and a home-grown IMAP front-end to MySQL. Each users' email was in a different table. All the tests were done on already populated DBs (small, medium, large) so didn't assess insert/delete/indexing issues. For the majority of operations on larger individual mailboxes DBs performed significantly better, particularly for searching.

Comments from the floor were that Sun's iPlanet ended up using maildirs, Cyrus' system is in fact a combination of maildirs and DBs and that as an ISP the *only* thing you are really interested in is the insert/delete/indexing performance. The authors agreed it would be useful to analyse the server side issues. (*jeremy*)

### A Secure and Transparent Firewall Web Proxy

⇒ An "authenticating proxy" providing access for external users to internal web services which are normally restricted by IP access control, or firewalls. Very interesting implementation – the router is configured to send all port 80 traffic to the proxy which then decides where to forward the request, avoiding the need

for any special configuration of the browser. In addition, the proxy returns a redirect to SSL port for any plain HTTP connections, thus forcing all external traffic to use SSL. The authentication mechanism used is a cookie-based scheme already in use at the authors site, but it seems that this could easily be replaced with something like kx509. (paul)

### Designing, Developing, and Implementing a Document Repository

Like the title says. The paper seems to focus on the process of the development. (paul)

### DryDock: A Document Firewall

A “staging” system for web data, where data is published to an internal test server and only released onto the public server when it has been “authorised”. CVS is used for revision control. (paul)

### Run-time Detection of heap-Based Overflows

Like the title says. Interesting, but I’m not sure where the system administration comes in. (paul)

### Designing a Configuration and Monitoring Reporting Environment

A system to retrieve and store configurations from a very large number of machines for the Deutsche Bank. I think that the idea is to gather the information on the machines which have not been under any kind of central management control, with a view to analysing the configurations and later building a management system. (paul)

### New NFS Tracing Tools and Techniques for System Analysis

⇒ New tools (two, plus helpers) for capturing and analysing NFS traffic. Includes some good examples of the tools being used to track down problems such as “something is generating lots of NFS traffic, but we don’t know why”. (paul)

The main difference over `nfstrace` or `tcpdump` is these only look at NFS protocol traffic and can also get effective UID/GID to help identify the source. `nfs-dump` collects the data, `nfs-scan` converts it into distinct tables for easier analysis either by conventional database or spreadsheet, or using the helper applications, including `ns-quickview` which uses `gnuplot` to create appropriate plots.

The package is available at <http://www.eecs.harvard.edu/sos/software>. (jeremy)

### EasyVPN: IPSec Remote Access Made Easy

A web-based interface for users to set up VPN clients. (paul)

### The Yearly Review, or How to Evaluate Your Sys Admin

Like the title says. I found this an interesting read, but almost certainly too formal to be directly applicable. (paul)

### Peer Certification: Techniques and Tools for Reducing Admin Support Burdens ..

A process for (internal) “certification” of junior administrators to allow them to reliably take on some of the tasks normally escalated to specialists. An interesting read again, but probably not directly applicable. (paul)

### ISConf: Theory, Practice, and Beyond

A (well-needed :-)) critique of the “ISconf” approach to configuration tools, and particularly the “theory” presented in the infamous “why order matters” paper from last year. (paul)

### Seeking Closure in an Open World: A Behavioural Agent Approach to Configuration Management

⇒ An important, if rather long and detailed, paper from Alva Couch which provides some theoretical results relevant to the design of configuration tools. The talk provided a useful summary which was more accessible than the paper. This also provided a more rigorous criticism of the “order matters” paper. (paul)

### Archipelago: A Network Security Analysis Tool

Research using graph theory to analyse the connections between systems, and the associated social connections, in an attempt to identify nodes or people who might be particularly vulnerable to security attacks. (paul)

### **STRIDER: A Black-box, State-based Approach to Change and Configuration Management and Support**

⇒ A diagnostic tool for debugging problems in the Windows registry. A tracing tool identifies the registry entries read by the broken application, and these are compared with the same registry entries from a working machine (or a previous working state). There is normally a small number of entries which differ (after filtering things which are known to change) and these are looked up in a database which holds known problems and describes the function of each of the entries. The authors suggest that similar techniques could be applied to Unix systems; this is possibly worthwhile for detecting problems due to, say, configuration files changed by post-install scripts, but I think that the original problem is really a consequence of the way that the registry is used, and I don't think the problem is anything like as serious under Unix. (*paul*)

### **CDSS: Secure Distribution of Software Installation Media Images in a Heterogeneous Environment**

⇒ A web interface is used to select authorised images for download, and firewall rules are automatically set up to allow access to these images for the particular "user" (source address) using several alternative protocols (NFS, HTTP). This replaces the normal authorisation mechanisms of the file transfer protocols (if any) with a single mechanism. (*paul*)

### **Virtual Appliances for Deploying and Maintaining Software**

⇒ A very interesting paper which describes a system to deploy virtual machines, containing specific environments, onto real hardware. Of particular interest is the "CVL" configuration language used to describe the configurations of the virtual machines; this includes the ability to specify simple relationships between the virtual machines in terms of "provides" and "requires". (*paul*)

### **Generating Configuration Files: The Director's Cut**

⇒ An update of an older system for generating configuration files from data in an Oracle database. This update was interesting for the use of XSLT to generate different configuration file formats from a common XML representation of the data. (*paul*)

### **Preventing Wheel Reinvention: The psgconf System Configuration Framework**

A configuration system intended to be highly modular. Modules similar to LCFG components generate configuration files, and other modules fetch configuration data from different sources. The intention is to be able to share modules between sites. This would have been a very good idea if it had remained at this low level, but it also includes some code modules which implement "policy" in a very imperative (and suspect) way. (*paul*)

### **SmartFrog Meets LCFG: Autonomous Reconfiguration with Central Policy Control**

My paper on the GridWeaver project. (*paul*)

### **Distributed Tarpitting: Impeding Spam Across Multiple Servers**

⇒ Best paper award. Very nice. (*paul*)

### **Using Service Grammar to Diagnose BGP Configuration Errors**

Like the title says. Some of the references on service grammars look worth investigating for those of us interested in configuration languages. (*paul*)

### **Splat: A Network Switch/Port Configuration Management Tool**

Configuring switches from data stored in an SQL database. (*paul*)

## **Appendix: Conference Structure**

The conference consists of several aspects: the workshops, the tutorial sessions, the technical sessions (refereed papers, invited talks, guru sessions) and BOFs. (*jeremy*)

- **Workshops:** typically full day events, they are organised by a key player in a particular field and depending on space or practical discussion considerations, attendance may be restricted. The aim is to discuss issues in technical detail. The workshops were on the Sunday and Monday.
- **Tutorials:** half or full day sessions on the Monday and Tuesday presented by individuals in their own right or occasionally companies dealing with a particular subject. These are accompanied by full tutorial notes, most of which are available on the LISA03 delegates CD.

- **Technical Sessions:** running from Wednesday through to Friday, these included usually 4 parallel streams, organised in four 1.5 hour chunks (2/am, 2/pm). It was quite normal to move between sessions if you were interested in more than one parallel session.
  - **Refereed papers:** papers approved by the program committee and included in the proceedings; delegates also get an electronic copy but only for personal use. Each 1.5 hour session had 3 papers with time given to questions. A wide range of subjects and complexity, theoretical or practical, covered.
  - **Invited Talks:** given by known experts in the field usually on a subject of their expertise/interest. Usually 1 hour presentation followed by questions.
  - **Guru Sessions:** aka, 'The Guru is in' are informal sessions with no specific talk but an opportunity for attendees to ask questions about general or specific issues. These sessions worked to varying degree, depending on the extent to which the guru was a good chairperson and/or the nature of the questions.
- **BOFs** 'Birds of a Feather' sessions, usually a 1-hour slot, organised informally in the evenings after the technical sessions. A busy evening could involve 3 different BoFs (from 7-10pm).