Informatics Network Requirements

George D.M. Ross
gdmr@inf.ed.ac.uk

March 2, 2016

This document describes the "requirements" for the Informatics network. It was intended specifically as input to Information Services' 2016 review of EdLAN, but should also be useful standing on its own as a statement of what we expect of the network and the motivating factors behind the way it currently exists. It was written based on the current Informatics network configuration and operating procedures, and on experience[1] gained over quite a few years designing and operating the network for the School of Informatics, and the Department of Computer Science before that.

The document is divided into two parts. The first main part (sections 1–5 below) outlines the requirements which the School of Informatics would look for when considering the adoption of any replacement network management system. This is followed by two appendices, starting on page 9, which describe how the School's network is currently configured.

In the text below items marked as ***M*** are **mandatory** requirements, and we would be unlikely to adopt a centrally-managed network system which does not meet all of these; items marked as ***H*** are **highly desirable** requirements, which we would very much prefer to be available if at all possible; and items marked as ***D*** are **desirable** requirements, which would incline us towards use of such a system, all other things being equal.

Please note: in all cases, items below which are marked as "mandatory" or "highly desirable" correspond to features and procedures which are in active use in the current Informatics network. Their absence would prevent some functionality from being provided; or would result in significant additional management cost and consequently perhaps require additional staff; or introduce error-prone workflows which might impact on the service provided to users.

Many thanks to all those members of the Informatics Computing staff who commented on previous drafts of this document.

1 General

As a general comment, we suggest that it would be useful all round for the EdLAN charging model to be reviewed and clarified.

2 Logical Network

M Must fully support both IPv4 and IPv6

We are currently rolling out an IPv6 pilot service throughout Informatics, and expect it to become increasingly important in future.

M Multiple subnets

Informatics currently has over 50 subnets in use. Some of these are used to segregate different classes of machine, for security and performance reasons, and some are set up as private

[1] http://history.dcs.ed.ac.uk/archive/docs/dcs-inf-network-history.pdf

subnets for research groups. It is entirely normal for users to have several different subnets and VLANs configured for their use.

For example: we place centrally-managed and self-managed machines on different subnets with different per-VLAN security policies to allow the former class of machine to use fixed addresses to avoid boot-time dependencies on network servers, while blocking network-level attacks such as ARP-poisoning by the latter; we have have some IP microphones in their own subnet on a small number of switches to protect the rest of the network and end systems from being swamped by heavy broadcast traffic; we have noted that many desktops throttle back their interface speeds while sleeping to minimise energy use, and as a result we have to keep the background noise on subnets below 10Mbps to prevent errors as a result of their links or interfaces being overloaded.

Note that it would definitely *not* be acceptable to adopt the "one subnet per building" approach currently used in some other parts of the University.

M Sites (Informatics Forum, Appleton Tower, JCMB) able to function independently

As far as possible the Informatics sites must be able to operate autonomously, in the event of any of them becoming disconnected from the others or the global network. It would not be acceptable, for example, for students not to be able to use the teaching labs in Appleton Tower just because the Informatics Forum or JCMB or the global Internet were not accessible; and similarly for the research labs in the Informatics Forum.

H Sites (Informatics Forum, Appleton Tower, JCMB) provide backup services for each other

Orthogonal to the above requirement, it would be highly desirable if Informatics sites could provide backup cover for each other, in the event that some part of their infrastructure were unavailable.

M Informatics policies set by Informatics

It must be possible to tailor the network policies specifically to Informatics's needs, for academic or research reasons. For example:

- As a research department, we get not-infrequent requests to attach pieces of kit to the network in various ways. Experience with the (outsourced to Information Services) wireless network suggests that "commodity EdLAN" is aimed towards conventional human use, which does not always sit well with the style required by this research equipment.
- It must be possible to segregate machines based on at least their management and purpose, in contrast to the subnet-per-building approach which applies in much of the University at present.
- The enforcement or otherwise of IP-use, such as by DHCP/ARP checks, would have to take into account whether machines were managed by Computing staff or self-managed, and hence were or were not allowed to use locally-cached addresses to avoid boot-time dependencies. A blanket setting of either mode would not work for us.
- Student practical exercises have for a number of years required the attachment to the network of equipment designed, built and controlled by them.

(These examples are not meant to be exhaustive. We would have to look at the *actual* policies in place at the time, aware that these might not be in place at the beginning but rather develop over time.)

M Filter rules to be applied automatically from our system-configuration system (lcfg[2])

It must be possible to configure internal and edge traffic filter rules automatically from data fed from our system-configuration system. Error-prone time-wasting potentially-inconsistent

[2] `http://www.lcfg.org/`

manual transcription of data would not be acceptable. This requirement helps to ensure that holes are created only where actually required, and are removed again as soon as they are no longer needed. At the time of writing our existing system had well over 4000 IPv4 rules and over 200 IPv6 rules generated for nearly 400 separate hosts.

M Filter rule changes to take immediate effect

In order to implement a security lockdown for on-line exams, it is a requirement that any changes to filter rules take immediate effect everywhere they are to be applied. We would otherwise not be able to guarantee that the University's exam rules would hold.

D Date- and time-based filtering

It would be useful if filter rules could be applied or disabled based on day, date, time or some combination of these. This would allow for time-limited holes to be created, for example.

H "Internal" filtering possible

We do have a number of subnets where it is desirable to apply additional filtering, both against the outside and indeed the rest of Informatics. In addition to the "exam lockdown" mentioned above, we have secure subnets for robot control for obvious health and safety reasons, some commercially-sensitive research projects, network management, and server management (where we are not confident as to the security of the on-board management systems).

H Inject routes for virtual subnets

We have two OpenVPN endpoints for staff and student use, tunnelling to within Informatics to allow access to protected resources. These, of course, require the ability to have routes for the subnets they control advertised globally.

M Rest of University to be "outside"

We have not "trusted" the rest of the University for many years. It would be a retrograde step to have to do so now. Greater separation all round would seem to be a more desirable approach, all the while taking into account such explicit cooperation between Schools as might be in place at the time.

M Private data paths between Informatics Forum and Appleton Tower

The bulk of the School's activity takes place in either the Informatics Forum or Appleton Tower (including Forrest Hill and Wilkie Building as "virtual AT" floors during the re-cladding decant). For efficiency and robustness the existing private paths between the two, bypassing the EdLAN core, must be retained.

M More than one (routed) path in to "central area" Informatics (i.e. Informatics Forum and Appleton Tower)

For resilience, there must be more than one path for routed traffic in to and out of the Informatics Forum and Appleton Tower. Traffic must fail over automatically should one of the links go down. (Bridged traffic need not, and indeed does not in the current Informatics network. However, given the automatic failover for routed traffic it is possible to connect in from outside to reconfigure links to re-enable bridged traffic.)

M Reconfiguration must be possible by either path

It must be possible to reconfigure the Informatics network in the central area using either of the above two paths in. It would not be acceptable from a robustness point of view to restrict management access to only one of these paths, in case that path then went down. (In the current Informatics network setup, network configuration servers are local to the site they manage, but are accessible to Computing staff through either of the paths in.)

H MAC addresses to come from lcfg (ultimately Informatics inventory database)

The Informatics inventory database holds MAC address information which is currently fed to the system-configuration system (lcfg) and ultimately used to drive various network configuration tools. It would not be attractive for time-wasting potentially-inconsistent error-prone manual transcription to be required.

M Multiple network servers to avoid single points of failure

Where possible we provide multiple network servers in the current Informatics network. Where this might not be possible (for example, servers for static vs. dynamic DHCP leased addresses) we segregate machines to different subnets so as to minimise the potential impact of any single points of failure.

M DHCP/ARP/BPDU/etc protection

To help maintain the security and stability of the Informatics network, it is a requirement that protection be available against potential network-based vulnerabilities, such as rogue DHCP servers, IP address misuse, and Spanning Tree Protocol perturbation. It must be possible to tailor this on a per-subnet or per-VLAN basis, as the variety of use within Informatics means that one single policy would not be suitable.

D Network login (802.1X)

We have a pending development project to set this up.

D Flow-based diagnostics and monitoring

We do not currently have the ability to perform flow-based monitoring of the Informatics network. This is a deficiency which we have been aware of for some time, but have not so far had development effort to address.

H Intrusion-detection tools available

We have run a `snort`[3]-based intrusion-detection system on all of our edge routers for some time now. We would expect a similar facility to be present, with flexible reporting being available to the School's Computing staff.

3 Physical Network

M Network as resilient as reasonably possible

H No single switch failure should affect users of any other

This applies to *all* core *and* edge switches, and reflects our current practice. It has proved to be extremely useful and effective.

M Adequate inter-switch bandwidth

H Inter-switch bandwidth sufficient to cover most peaks

We designed the current data network in the Informatics Forum with a target uplink contention ratio of around 8:1, with the ability to add additional uplinks easily as required. This has been an effective approach. For edge switches with 1Gbps "user" ports this implies 10Gbps or better uplinks; for switches with 100Mbps "user" ports this implies 1Gbps or better uplinks. For our server rooms the target contention ratio is 4:1, with direct connection to the core at up to 10Gbps available for any servers which really require it.

[3] `https://snort.org/`

H Spanning-tree isolated from outside influences

We have experienced network instability in the past when were were part of the EdLAN spanning-tree, so for some time now we have filtered BPDU packets where we interface to non-Informatics networks. This improved stability considerably.

M Sufficient edge switch port capacity, pre-patched

Standard Informatics user provision is 3 data ports and one VoIP phone port, all patched in and ready to use. (A number of our users have had additional ports patched in to supplement this standard provision.)

M Configuration must incorporate floor- and wall-port labels

It must be possible to search the configuration by floor- or wall-port label. It would not be acceptable for there to be a separate mapping between physical labels and switch port. Experience shows that this would inevitably result in errors.

M Soft configuration rather than physical re-patching

So far as possible, all outlets must be pre-patched to switch ports. Assignment of ports to VLANs/subnets must be by soft configuration of the associated switch(es), rather than physical re-patching. Past experience over a number of years has shown that incremental patching or later physical re-patching is a sure way to chaos.

H MAC-locking

For security and to prevent machines being moved around without due process, we configure many of our network ports so that they are locked to specific MAC addresses. This can either be by explicit address, or by name with the name-to-MAC mapping coming automatically from our system-configuration system (lcfg[4]).

M Multiple VLANs (including private)

Informatics currently has 75 VLANs in use. Some of these are mapped to equivalent subnets of various sizes, both public and private. Some are for private use by Informatics groups. Some are passed through from outside, for the benefit of third parties. Examples of the last include BEMS and metering for Estates & Buildings, and lock-controllers for Security.

M Informatics-specific physical network requirements

It must be possible to accommodate specific requirements, for research or teaching. For example: we have a robot lab which has over 100 power-over-ether cameras, for motion-capture; and one of our major student practicals requires the attachment of student-built equipment to the network.

M Redundant core

Our current network has multiple core switches to ensure resilience, both for switching and routing, and for uplink connectivity. The network is designed so that the failure of any core-switch card, or indeed of any entire core switch, does not result in any loss of service. We would expect any replacement core to be at least as resilient.

H Multiple redundant power for network core and server room switches

As well as simply giving extra robustness, we have found this very useful in the past for some infrastructure-management tasks. (We do have redundant power for some user-facing edge switches, but have not made this a priority and most do not have redundant power.)

[4] `http://www.lcfg.org/`

M Pre-configured hot-spare ports

Where there are unavoidable single points of failure, provision must be made for simple re-patching to pre-configured switch ports to work around them if required. Ideally this pre-configuration would automatically track that of the primary port.

M MIBS and documentation freely available

In the event that we continue to use our own management tools, it is an absolute requirement that the necessary documentation and MIBs are freely available to allow us to interface to any switches under our control.

4 Configuration and Management

H The configuration system must be available when required to all Informatics Computing staff and technicians

This is how the existing Informatics configuration tools have been set up for many years. Our procedures assume it and rely on it daily. Having to locate special designated "network managers" would be very disruptive to smooth working, as would any delayed access to the system.

H No special hardware or system required for network management

We run a variety of systems internally, and expect all of our management tools to be available from everywhere within Informatics. Having to use some specific designated management station would not be acceptable: as well as being simply inconvenient, it would prevent common straightforward operations such as cut-and-paste between tools.

M Flexible authentication and authorization

Our current network configuration tools are authenticated using Kerberos and use a flexible roles-based authorization system which is capable of quite fine-grained control. For the security and stability of the network we would expect similar robust access controls to be in place for any alternative configuration technology. (Our existing `rfe` tool is described in appendix B on page 17.)

M IPv4 and IPv6 equally well supported

We expect to IPv6 to be as well supported in future as IPv4. We have a development project under way at the moment which is intended to achieve this. Network management tools must be able to support this model.

M Tools must cover at least VLAN configuration

VLAN configuration is probably the most-used feature of our existing network management tools. We expect to see at least half-a-dozen changes per day across Informatics, with many more than that being entirely normal.

H Tools should also cover MAC locking, DHCP snooping, ARP protection, BPDU filtering and protection, IGMP and MLD, router-discovery and RA.

These are the other main areas which are under the control of our network management tools. Protections are automatically enabled as appropriate.

M Management coverage expandable in future

It must be possible for any management tool to be expanded in future to allow it to manage additional switch features as they become available or their use becomes desirable. (For the current Informatics tools we would just write a few more lines of `tcl`.)

H "include" and "variable" functionality

We make heavy use of these, to simplify configurations and help ensure consistency. It would be a retrograde step to have to try to apply identical settings in multiple places.

H Minimal training

Our current tools are easy to use and require little in the way of training, and the access methods are familiar as they are common to a number of our configuration systems. (Our existing `rfe` tool is described in appendix B on page 17.) Experience with this and with other tools and packages definitely points towards better productivity and fewer errors.

M Defined API to DNS/DHCP tools

It must be possible to set and query data in any DNS and/or DHCP management system through a well-defined API. Scripted changes and queries are commonly used in our current system. (Standard UNIX tools applied to textual configuration files, as we currently use, would qualify.)

M All DHCP options settable

In addition to the standard MAC and IP address fields, installation of our Linux systems requires setting at least `root-path` and `user-class`. Other options may be required in future, and there must be sufficient flexbility to allow these to be added as required.

H Incorporate data from our system-configuration system

MAC addresses and DNS SRV data are generated from our system-configuration system (lcfg). It should not be necssary to re-enter these into some other system.

M Defined API to network configuration tools

It must be possible to set and query data in any network configuration system through a well-defined API. Scripted changes and queries are commonly used in our current system. (Standard UNIX tools applied to textual configuration files, as we currently use, would qualify.)

M Tools not to require a graphical interface

Point-and-click graphical interfaces can be much slower at some tasks. While we would not be averse to the option being available where appropriate, we would require more powerful alternatives to be available too.

M Import existing configuration data

If we are to change to a new suite of tools, it must be possible to import our existing data from our current tools. At present our network configuration is spread across around 440 separate files.

M Export configuration data

It would be very useful to be able to export all configuration data (in a useful form) from any new tools. In particular, it must be straightforward to migrate to another set of tools.

H "default"

Being able to set a "default" value for network configuration items improves clarity, accuracy and consistency.

M Port location data, exported

MAC-use data along with port location information are essential for us to audit network usage and to respond to IRT reports. In addition, we feed the locations where MAC addresses have been seen into our inventory system to allow for the tracking of machines.

M Robust backup and recovery

Given the number of changes made daily to the Informatics network, any extended management downtime would not be acceptable. Ideally there should be hot-spare systems on standby with up-to-date data.

Our current backup strategy is to mirror all of the network configuration files at least twice per day to other on- and off-site network servers, as well as having them copied to tape nightly. These mirroring machines also carry a full set of management packages, so are able to take over the network management role quickly with minimal loss of changes. (The same mechanism is used during the rolling server-upgrade cycle, and is well-tested.)

M "Disaster mode" reconfiguration possible

There must be no dependencies which would inhibit an easy recovery from a black-start situation. *In extremis* it must be possible to make configuration changes as necessary even with only very minimal surrounding infrastructure support.

H Changes pushed immediately

Ideally any configuration changes would be pushed to the switches immediately without any delay. This is how our current switch management tools operate.

M No more than a couple of minutes delay in pushing changes

In any case, it would be disruptive to work-flow, as well as inconvenient for there to be any significant delay in configuration changes reaching the switches or network servers.

M Change logs and audit trails

It must be clear where the definitive master version of the network configuration is held. It must be possible to determine after the fact what changes have been made to it, by whom, and why they were made. It must be possible to revert changes easily when required.

5 Diagnostics and Monitoring

M Full access to switches' own diagnostic logs and tools

We make quite extensive use of this ability to help diagnose more obscure network-related faults, in cases where summary logs do not provide sufficiently detailed information. (We have found that our lack of access to such data for the outsourced-to-IS wireless service has made it quite a bit slower and harder to help users with problems.)

D Full access to wireless diagnostic logs and tools

See above, and section A.3 on page 13 below.

M Traffic, error, and spanning-tree change graphs, all indexed by port label

M Nightly trap log summaries, with "important" items emailed and the rest available for later perusal

M other nightly reports: configuration changes, status reports

These three items are essential for network monitoring and for future planning.

M FDB available for correlation with other data sources

MAC-use data along with port-location data are essential for us to audit network usage and to respond to IRT reports. In addition, we feed the locations where MAC addresses have been seen into our inventory system to allow for the tracking of machines.

M All reports to be subject to a Data Retention Policy

All reports, statistics and data extracted from the switches must be saved in a searchable form for such time as corresponds to some well-defined Data Retention Policy and must then be automatically purged. Ideally this period would be defined by the School, and might be different for different classes of data.

M Search tools

In addition to the obvious uses by support staff, these also allow end-users to see how their network ports are configured, so that they can connect up their devices to the appriopriate subnets or VLANs.

H Link-layer discovery protocol (LLDP) results available

We use this as an alternative search technique to find particular machines or devices.

A Current Informatics Network

This appendix describes the current Informatics network configuration, the tools used to manage it, and the data flows to and from other system management components. Diagrams can be found linked from our "technical network documentation" site[5]. A description of the key `rfe` configuration-editing tool can be found in appendix B on page 17.

Note that the tools have evolved over many years as a result of requirements identified in the ongoing management of the Informatics network (and the Computer Science network before that). Although there are undoubtedly things which could be improved, and there are certainly things which we know are missing, there is almost nothing which is not in frequent active use.

It is our general philosophy to automate as much network and system management as possible, based on updating configuration databases, rather than making a number of repeated small changes on individual switches and machines. Experience over many years has shown that this is much less error-prone, as well as being considerably more efficient in staff time.

We do run a number of network services in addition to those mentioned in this document: for example, NTP and Kerberos. We would consider these out of scope here, however, and they are not discussed further.

A.1 Logical

The Informatics IPv4 network consists of over 50 subnets, both globally-routed and RFC1918, ranging in size from /26 to /22. As a general rule, subnets are confined to one of our three sites (Informatics Forum, "Appleton Tower" (including the Forrest Hill and Wilkie Building decant sites), and the College Server Room in JCMB), to avoid inter-site dependencies and unnecessary traffic on the inter-site links, and to minimise MAC-learning requirements by the switches. The three sites are logically interconnected using a private transit subnet. Most subnets are carried on corresponding VLANs, though not all our VLANs have corresponding managed Informatics IP subnets.

Our standard provision is that each user should have at least three data ports and one VoIP phone port. Data ports can be configured to use any of the subnets/VLANs available in the building to suit the users' requirements.

The School's primary DNS domain is `inf.ed.ac.uk`. In addition to this, the `dcs.ed.ac.uk`, `dai.ed.ac.uk`, `cogsci.ed.ac.uk` and `aiai.ed.ac.uk` domains are maintained for legacy purposes. These are all held in source form in simple text files, separately for IPv4 and IPv6, which can be edited by Computing staff using our "rfe" tool (see appendix B below), under the control of our School-wide authorization mechanism. Forward and reverse zones are automatically generated

[5] `http://www.dice.inf.ed.ac.uk/units/infrastructure/Documentation/Network/`

as required, and are immediately loaded by our master nameserver for propagation to the secondaries using the standard zone-transfer procedures. Where MAC addresses are available from our system-configuration system (lcfg[6]), the tools will also automatically add reverse-mappings for IPv6 SLAAC-style addresses, for user convenience. Due to the need for consistency with forward and reverse DNS lookups, and to be able to name individual interface addresses as required, the tools implement a simple convention around the use of '–' as a separator within names.

Sub-domain names are fully supported. For example, we have some service names (such as `computing.help`) which are named in this way, for clarity and to support future expansion; while network equipment would generally be placed in the `<thing>.<site>.net.inf.ed.ac.uk` sub-domain Some DNS data (e.g. SRV records) are generated automatically from our system-configuration system (lcfg), and for more complex cases it is also possible to inject verbatim records.

Most internal routing is done by the sites' core switch/routers, simply because of the sheer volume of traffic involved. External routing, and some internal routing, is done by PC network servers running the School's standard Linux system. There are currently five edge routers in the Forum, four in Appleton Tower, and three in JCMB. These are managed identically to all of the School's other Linux servers, which makes their configuration straightforward and consistent. Where required, these have been given 10Gbps connections to the Informatics core.

These routers all run stateful packet filters using `iptables`, with identical rulesets all automatically generated from data provided by our system-configuration system (lcfg). At the time of writing there were well over 4000 IPv4 rules and over 200 IPv6 rules generated for nearly 400 separate hosts. Traffic originating from outside Informatics (in which we include the rest of the University) is subject to a default-deny policy, with inbound filter holes being automatically applied based on the data provided by the Informatics system-configuration system (lcfg). Outbound traffic is generally unfiltered by default. Filter changes are automatically immediately emailed to (configurable) lists of Computing staff for review. We do also have some capability to add date- or time-limited rules, though this is not currently used as its implementation is a little less flexible than it really needs to be.

The edge routers run a `snort`[7]-based intrusion-detection system. Signature matches are logged to our central loghost, with the log entries being post-processed nightly into a more-easily-digestible format, with reports being emailed to a (configurable) list of Computing staff.

The OSPF costs for the edge routers are arranged such that each site's external traffic normally goes through a designated one of its local Linux routers, with the other site routers providing a failover capability. Should all of a site's external routers go down then traffic will fail over to be routed through one of our other sites. Having well-defined symmetric routes to and from outside Informatics at all times is essential to the correct working of `iptables`' connection-tracking mechanisms.

We have some internal filtering in place as well, though we apply this carefully so as not to load our Linux routers excessively. Examples include: exam lockdown, to enforce the University's rules; robot control, for health and safety reasons; network and server management; and some commercially-sensitive research projects. We have considered generating core-switch ACL entries from system-configuration data, but have so far not implemented anything as the ACL rules are less flexible than `iptables`' statefulness.

We run two OpenVPN[8] endpoints, one in the Informatics Forum and one in Appleton Tower. These allow both staff and students working away from the School to tunnel to inside Informatics and access protected resources in a secure way. (In contrast, the University's central VPN provision tunnels users to within EdLAN but *outside* Informatics.) These have proved to be extremely popular since they were introduced several years ago, so much so, in fact, that we recently had to double the allocated address space to two /24 subnets.

The system configuration system (lcfg) also feeds MAC addresses and other configuration details to the DHCP servers, which along with the DNS information allows them to build their

[6] `http://www.lcfg.org/`
[7] `https://snort.org/`
[8] `https://openvpn.net/index.php/open-source.html`

tables automatically. System installation requires that options other than the standard MAC and IP address can be set automatically from lcfg, including at least `root-path` and `user-class`.

Ports are assigned by Computing staff or technicians to one or other subnet (VLAN) depending on the type of machine that is expected to be plugged in to it: for example, managed Linux desktop, managed Linux server, managed Windows desktop, self-managed desktop or server, network management, server management, and so on. This helps to improve network security considerably, as well as to contain traffic to minimise interference. There are usually at least half-a-dozen port VLAN reconfigurations per day across Informatics, and often considerably more. Most users have ports configured onto several different VLANs. We can (and do):

- permit access to privileged resources, such as switch management interfaces, only to certain classes of machine
- apply different filtering rules to different classes of machine
- permit external-to-EdLAN access only to ports in offices, or where we can otherwise audit access to them
- limit the types of DHCP packets permitted, so as to prevent rogue servers disrupting service
- permit only DHCP-assigned IP addresses to be used, which improves security and stability by blocking the ability of machines to impersonate others, either accidentally or deliberately

IPv6 capability is under development, and is expected to be rolled out as a service during 2016. The core functionality (local routing, global routing, filtering, switch configuration, static address assignment) is operational now, and the pilot system has been opened to Computing staff for testing. The ultimate intention is that IPv4 and IPv6 should be equally well supported on the Informatics network.

We do not at present have any flow-based monitoring or diagnostic provision in place. There have been quite a few occasions when it would have been very useful, but so far we have not been able to resource its implementation.

802.1X is planned, but is currently queued waiting for resources to implement it. It remains to be seen how well some of our older switches support it.

A.2 Physical

The Informatics network has been designed so that so far as possible should any individual component (switch or card) fail, this will not result in loss of service for any other switch or port. This applies to *both* core *and* edge switches. The two exceptions are the "central area" bridged traffic and the College Server Room (JCMB), where manual intervention would be required to transfer the traffic to existing pre-configured hot-standby provision.

Redundant inter-switch links are used extensively throughout the Informatics network, managed automatically using rapid-spanning-tree protocol ("RSTP"). There is one tree covering the Forum and Appleton Tower, and a separate tree covering JCMB. We do not participate in any EdLAN-wide spanning-trees, having been affected by instability in the past, and we filter BPDUs on all our links to non-Informatics networks. This does of course mean that while we can have automatic failover for routed traffic, we can not do so for bridged traffic. This is a compromise which we have made in the interests of network stability.

All inter-switch links have been sized to provide adequate bandwidth for most peak loads, aiming for an edge-port to uplink contention-ratio of around 8:1 (though during the temporary decant from Appleton Tower to Wilkie Building and Forrest Hill we do have a few links with rather higher contention). For server-room switches, the target contention ratio is 4:1, with 10GbaseT connections being avilable for servers which really need the speed.

At the time of writing we have a total of 75 VLANs carried on the switches at our "three" (counting Forrest Hill and Wilkie Building as part of Appleton Tower) sites. Most of these correspond to routed IPv4 subnets, though we also have quite a few private VLANs set up for research

and commercialisation groups, as well as for University-wide infrastructure such as phones, wireless, locks and BEMS.

As a general rule, we patch in all floor- and wall-ports systematically when we occupy a new (part of a) building. We have found that this is much more efficient and less error-prone than piecemeal patching later. Our standard provision is that each user should have at least three data ports, pre-patched and ready to use once they are configured, and one VoIP phone port. At the time of writing there are over 3500 data ports in the Forum, around 1200 in "Appleton Tower" (including Wilkie Building and Forrest Hill), and nearly 100 in the JCMB College Server Room; and around 500 phones in the Forum and over 120 in "Appleton Tower."

We have the ability to lock ports to specific MAC addresses for security and to prevent machines being moved around without due process. This can either be by explicit address, or by name with the name-to-MAC mapping coming automatically from our system-configuration system (lcfg[9]).

Each site has its own dedicated network infrastructure servers, each connected by multiple bonded links to the site's core switches. Again, this is for resilience, as it provides management, monitoring and routing connectivity even if one or more of the core switches is down for some reason.

We keep a number of hot-spare switches, so that we can arrange an immediate swap as required. Replacement for the faulty switches is then arranged under their next-day lifetime warranty, with the replacement being used as the new hot-spare.

A.2.1 Informatics Forum

The network in the Informatics Forum consists of three core switch/routers (ProCurve 5412zl), interconnected in a triangle by 40Gbps trunked links. Each of the four parts of these trunked links terminates on a separate switch card, for resilience. The switches have multiple redundant power supplies, fed from multiple local UPSes. These core switches also have multiple connections to our Appleton Tower core using private fibre running under Crichton Street between the buildings.

There is a fourth core switch, with a 1000baseLX link to the EdLAN-OC core router and local 1000baseT links to several Forum network servers. This provides automatic failover for routed traffic should our main link to the EdLAN-AT router go down for whatever reason. This switch also has normally-off 1000baseT links to the main Forum core, which allow bridged traffic to be manually switched over as required, with due care being taken to avoid creating forwarding loops as a result of not participating in any EdLAN-wide Spanning Tree system.

Having two independent ways in for routed traffic also improves resilience, as it makes it possible for Computing staff to connect remotely as required even in the event of a failure of the other link, in order to perform remedial action from off-site.

The server-room switches[10] are arranged pairwise in the racks, with servers expected to connect bonded links to both switches in the pairs for resilience. These pairs are connected back to the core by 10GbaseT links, with each link being connected to a different core switch, and taking into account the 5412zl's "top-half/bottom-half" power architecture. Each pair is also connected to another pair by a 1Gbps link as a protection against both of "its" core switches failing. Each pair of switches has its own redundant-power PSU shelf.

The IT closets covering the Forum offices[11] are arranged more or less one above the other, with cableways between, and so the edge "1Gbps" switches are connected together into vertical "stacks" with 10GbaseT connections between them and 10GbaseSR uplinks tapped off no more than two switches apart to maintain the resilience requirement. Adjacent uplinks go to different core switches, again for resilience. "100Mbps" switches are then connected locally to those 1Gbps switches, with two connections each to different 1Gbps switches for resilience.

[9] `http://www.lcfg.org/`

[10] `http://www.dice.inf.ed.ac.uk/units/infrastructure/Documentation/Network/Forum/SR.png`

[11] `http://www.dice.inf.ed.ac.uk/units/infrastructure/Documentation/Network/Forum/edge.png`

A.2.2 "Greater Appleton Tower"

(For simplicity during what is expected to be a temporary decant while the Appleton Tower recladding takes place, both Forrest Hill and Wilkie Building are being operated as virtual floors of Appleton Tower.)

There are two core switch/routers (ProCurve 5406zl) in our Appleton Tower basement comms area. These are interconnected by a pair of 10Gbps direct-attach trunked links connected to separate cards in the switches. The switches have two redundant power supplies, and are fed from a pair of local UPSes. One of these core switches has a 10GbaseSR link to the EdLAN-AT core router. This link normally carries most of our central-area and external routed traffic, as well as bridged (i.e. non-routed) traffic for all of our central-area sites.

The other core switch has a pre-configured hot-spare "EdLAN" port. Physical re-patching would be required to bring this into service, though routed traffic does fail over automatically to use the backup link from the Forum to the EdLAN-OC router. These two switches have multiple 10Gbps and 20Gbps links to the Forum core.

The server racks have a pair of switches, each connected back to a separate core switch over 10GbaseSR, and interconnected by a 10Gbps direct-attach link. There is a shared redundant-power shelf. The "cluster" racks are similar, though without the redundant power.

Each floor (including Wilkie Building and Forrest Hill as virtual Appleton Tower floors) has a pair of "1Gbps" switches, each with its own link back to one or other of our Appleton Tower core switches. Within each floor there are additional switches chained from these pairs, with multiple redundant connections arranged so that one single failure of a link or a switch does not result in any loss of service to any other switch or port.

A.2.3 JCMB College Server Room

Our main provision in the College Server Room consists of a pair of 48-port ProCurve 3800-48G-SFP+ switch/routers, interconnected by a 10Gbps direct-attach link. There is a third "100Mbps" switch connected to both of the main pair, giving additional connectivity mainly for infrastructure components (power-bars and UPS management cards), plus some power-over-ether provision. Servers are expected to connect bonded links to both of the main pair, for resilience. There is a 1000baseSX link from one of the main switches to the EdLAN-KB core router, with a pre-configured hot-spare link in the other main switch. Physical intervention would be required to re-patch this should there be a switch failure. The availability of two identically-configured interfaces has however proved to be extremely useful during network maintenance work. An upgrade to 10Gbps for this link is under consideration, as it is very heavily loaded at times. The two core switches were procured with this possibility in mind.

Connectivity to the rest of Informatics is by means of a private "transit" subnet/VLAN which is carried for us across EdLAN between JCMB and Appleton Tower. (This same transit subnet is also used for traffic between the Informatics Forum and Appleton Tower.)

A.3 Wireless

For many years the School of Informatics (and before that the separate Departments from which it was formed) operated its own wireless network. Around ten years ago we decided instead to outsource wireless provision to Information Services. There were a couple of reasons at the time why we did this:

1. As we had moved from JCMB, Forrest Hill and Buccleuch Place to the Informatics Forum, which is a much more densely occupied area of the University, we were aware that there would likely be overlapping cells, both from outside into our building and from our building spilling out into others. We anticipated users (both our own, and those in neighbouring buildings) roaming between these cells, and concluded that there would be a "smoother" experience all round if the wireless networks were more unified.

2. It allowed us to outsource the infrastructure costs and some of the support costs. However there is a recurrent per-WAP maintenance charge which is intended to cover these, so over time there may not actually be much of a saving for the School, if any.

So far as Informatics is concerned, our involvement now is limited to mounting the WAPs which are provided for us; ensuring that they receive power from our PoE+ switches; and passing through any VLANs that Information Services ask us to.

There have, however, been some definite disadvantages which have become apparent since the change, resulting from the loss of access to the logs and configuration details for both the access points themselves and their back-end controllers and RADIUS authentication servers:

- We no longer have a good handle on the use of wireless in our buildings. We are not able to see how the cells interact with each other, and with surrounding cells from other University buildings, shops and cafes. When users have connectivity problems, it is not immediately clear to us what the effects of signal strength and interference might be.
- When new users, including visitors, are unable to connect at all, we have no way to determine what the problem is. All we can do is gather as much local information as we can and then pass the problem over to Information Services. This inevitably introduces delays for the users, while at the same time still incurring a support cost at our end.
- We still have some "private" wireless provision, for robot control, and for management of the equipment in the Instrumented Meeting Room. We have implemented these using old WAPs in a way which is most likely less than ideal, but we have no way to determine how these deployments affect the "commodity" wireless systems.
- We have researchers who specialise in wireless communications. We can advise them as best we can as to how to avoid interfering with the commodity wireless users, but we have no way in practice to see what effect they really have.
- We are aware that some items of equipment brought into the building by staff and students have wireless capability, but we are not notified when potentially-interfering things start up and are therefore not able to take proactive action to contact the users and have their kit reconfigured to minimise effects on others.

A.4 Configuration and Management

The current Informatics network configuration is almost entirely done through locally-written tools, developed over a number of years as new requirements were identified.

- Definitive VLAN and port configuration details are held in simple text files on the sites' local network infrastructure servers, all of which run our standard managed Linux system. These are under change-control, using `rcs`, which allows for *post hoc* auditing of changes and for rollback as necessary.
- The configurations can all be edited by Computing staff and technicians using our "rfe" tool (described in appendix B below), under the control of our standard authorization mechanism. Configurations are error-checked and then immediately pushed to all affected switches as required when any of the files are changed. There are typically around half-a-dozen configuration changes per day, often many more than that, and very rarely none at all.
- Configuration files are all cross-mirrored between our network servers, and are backed up to tape as following standard procedures.
- The configuration language was deliberately kept as simple as possible, while retaining as much power as necessary, so that the amount of training required to use the system to perform routine actions is very low.

- The configuration language was designed so as not to be specific to any particular switch make or model. Although we currently use ProCurve switches, this has not always been the case, and in the past we have used other models. Adding support for some other switch make would be a straightforward if rather tedious process, provided that the necessary documentation and MIBs were freely available.
- In addition to VLAN and port settings, the configuration files contain floor- and wall-port location data which are set when ports are first patched in, making it easier to find which ports to configure later. These are also fed to our inventory, reporting and auditing tools.
- Configuration files may "include" other files, and the language makes provision for variables and templating. This makes it easier to configure ports, such as the hot-spares and the inter-site links, in a consistent manner.
- The tools can currently configure: port location, VLAN (untagged and tagged-list), speed, edge-port, BPDU filter, BPDU protection, MAC lock (by both name and address); router-discovery parameters, RA parameters; VLAN names and tags; DHCP snooping, ARP protection, CoS, IGMP and MLD by VLAN; manager address-list. The name-to-MAC mappings for MAC locking come automatically from our system-configuration system (lcfg[12]).
- Many settings can have default values. This simplifies the configuration files, and improves configuration correctness and consistency.
- The configuration files are backed up using our usual mechanisms, as well as being mirrored several times per day to a selection of other network servers ready for use in case of a machine failure.
- The use of plain text files allows the straightforward use of the full range of standard UNIX tools, including scripted queries and changes. *In extremis* it is possible for a user at the network management server's console with a root shell to be able to make network configuration changes using the standard tools, even with much of the rest of the Informatics infrastructure unavailable. This is an essential disaster-recovery feature.

Only the most fundamental network configuration needs to be done by hand (e.g. switch address, SNMP communities, firmware updates), minimising the number of people who need to know the manager passwords and be trained in the vagaries of the various CLIs and menu systems involved.

Our current network management and monitoring scripts consist of around 14k lines of `bash`. around 18k lines of `tcl`, around 1k lines of `C` and a few dozen lines of `perl`. The configuration language is essentially unchanged since the late 1990s, when it was initially developed for 3Com SuperStack-II switches. It was later ported to ProCurve 4000m switches, and has had additional features added ever since to allow control of the facilities offered by newer-generation ProCurve switches. It should be straightforward, if rather tedious, to port to another range of switches. The oldest code still in everyday use actually dates back to the early 1990s, and is the `C`-language tool written to manage IPv4 DNS entries. (Unfortunately its ChangeLog dates back only to mid-1997 when it was merged into our central CVS repository.) As it still does the job it was written to do we have seen no particular reason to replace it.

A.5 Diagnostics and Monitoring

The Informatics network is extensively instrumented throughout, for auditing, debugging and resource-planning:

- Traffic and error counts are obtained for every port on every switch, whether configured at the time or not, every five minutes. These are fed into `rrdtool`, where they are used

[12] `http://www.lcfg.org/`

to produce "daily,", "weekly," "monthly" and "yearly" graphs. These are all automatically indexed several times per day for easy access.

- Per-switch CPU-use statistics and spanning-tree changes are similarly obtained and graphed.
- LLDP data are extracted from the switches and included in the traffic index pages.
- Per-site traphosts collect diagnostic messages from all the switches. These are analysed nightly, and email reports generated containing important messages. All analysed messages are retained for 120 days and then automatically purged, in accordance with the School's Data Retention Policy.
- Other nightly email reports include: switch configuration changes, DNS changes, switch status reports, and edge filter reports.
- Weekly reports include: firmware and reboot status; and hardware configuration details as reported directly by the switches, for comparison with our inventory database.
- IPv4 ARP use is logged for all subnets. Again, this is kept for 120 days. New and changed IP-MAC mappings are emailed to a (configurable) list of Computing staff, for immediate attention if required. Our IPv6 projects will extend this to ND in due course.
- The switches' forwarding databases are fetched every 10 minutes, merged, and kept for 120 days. Together with the location information in the switch configuration files and the ARP logs, this allows us to audit connection use. Personal Data are handled in accordance with the School's Data Protection statement[13].
- The per-site network infrastructure servers provide a web interface[14] to all of the logs and monitoring data, including graphs, and search tools for ports and addresses.
- We have full access to the switches' own internal logs and diagnostic counters when required, for detailed debugging.

A.6 Data Flow into the Informatics Network Tools

Edge and internal filter rules are automatically generated based on values supplied by the Informatics system-configuration system (lcfg[15]). This improves correctness and consistency, as the requirements are kept clearly in machine profiles where they can also be more easily subject to regular audit. Holes are automatically updated when machines' configurations are changed and are removed when machines are decommissioned, thus avoiding the potential for security breaches as IP addresses are recycled.

MAC addresses fed from the Informatics system-configuration system are used in several ways:

- Updated DHCP tables are automatically generated immediately a host's MAC address is configured.
- The switch status reports and nightly trap summaries can translate known MAC addresses back to the names of the machine profiles they are registered for.
- Ports can be locked to particular MAC addresses by machine name as well as by literal address.
- IPv6 SLAAC-style reverse DNS entries are generated.

[13] `http://computing.help.inf.ed.ac.uk/DP-statement-managed-systems`
[14] For example, `http://netmon.inf.ed.ac.uk/`
[15] `http://www.lcfg.org/`

A.7 Data Flow out of the Informatics Network Tools

The forwarding database and port-location data, including imported MAC data, are used to keep the Informatics inventory system up-to-date. In addition, they provide essential input to our network auditing processes.

B rfe: Configuration File Editing

The School of Informatics uses structured text files in many contexts to hold configuration data. In addition to the network configuration, we have *inter alia* system configuration and authorization capability data held in this form. In order to ensure a uniform secure access mechanism for editing all of these we have developed a "remote file editing" tool: `rfe`. There are several key features which this tool provides:

1. All use is authenticated using Kerberos, with access then granted (or not) using our role-based authorization mechanism. This access control can be as coarse- or fine-grained as necessary, with the possibility of access being granted or denied to Computing or Technical staff right down to the level of single configuration files as required.

2. *In extremis* it is possible to perform all configuration tasks from a root-login on a network configuration server's console. This provides a way to avoid interdependency deadlocks during a disaster-recovery situation or a black-start problem.

3. The system automatically applies exclusive locks on any files while they are being edited, to avoid the possibility of corruption due to simultaneous edits.

4. All changes are automatically checked into `rcs`, along with a change-message, which provides a full audit trail as well as a rollback mechanism.

5. A server storing `rfe`-controlled files has the ability to run scripts, as configured, once the editing session has closed and the files checked back in. In particular, our network tools use this hook to call the standard UNIX `make` command to evaluate dependencies and to push changes to any switches which have been affected. In practice this may range from just one switch to all of the switches at a site, depending on what has been changed in the configuration files. In no case is it actually necessary for the person changing configuration files to know which switches it might affect, as this is all determined automatically, thereby minimising errors and improving consistency.

6. The system is editor-agnostic, in the usual UNIX fashion, so users are not forced to learn new tools just in order to make configuration changes.

At the time of writing there were 440 separate files related to network configuration under `rfe` control.