UNIVERSITY OF CAMBRIDGE
Department of Computer Science and Technology

ITSC_AGENDA_02/12/2024

IT Strategy Committee, Department of Computer Science and Technology
2nd December 2024 at 14:00
SW00, William Gates Building

AGENDA

Membership

Richard Mortier, Chair [RM]
Mark Cresham, Secretary [MC]
Tim Jones, UTO Rep [TJ]
Rob Harle, Director UG Teaching [RH]
Helen Francis, PSS Rep [HF]
Nic Lane, GPU resourcing strategy lead [NL]
Thomas Sauerwald, Deputy HoD [TS]
Daniel Porter, IT Support Manager [DP]
Sam Nallaperuma, Research Staff Rep [SN]
Abraham Martin Campillo, UIS Rep [AM]
Malcolm Scott, IT Infrastructure Specialist [MS]

1. **Welcome & apologies [RM]**

2. **Approval of the Minutes of the previous meeting [RM]**

   Unconfirmed Minutes of the meeting held on 3rd October 2024 are attached for approval.

3. **Matters arising [RM]**

4. **Actions from the previous meeting [RM]**

(i) GPU Upgrades [MS, NL]
- MS and NL will discuss and decide whether to implement a trial for a full course or limit it to select projects.

(ii) Meeting Room Upgrades [DP]
- DP will pick up the conversation with UIS regarding the hybrid meeting room trial.

(iii) Legacy Services [DP,MS]
- DP and MS will compile a list of legacy services for review.

(iv) Network Upgrades [MS]

- MS will start planning of the firewall replacement.
- Further to this, approximately £2 million has been allocated for the future network upgrade budget. MS would like to discuss the options and strategies for utilising these funds effectively.

(v) Separation of Department Websites [RM]

- RM will inform mgk25 of the plan to separate the (cl) website from (cst) website.

5. Standing items

(i) * UIS update [AM]

- AM to provide a verbal or written update on any relevant developments from UIS, including updates on the door locking system.

(ii) * IT team update [DP, MS]

- DP and MS to provide written updates, summarising key developments since the last meeting and highlighting the progress of IT Services.

6. Main business

(i) Cisco to Teams Phone Migration [DP]

- DP to lead a discussion on the approach for the phone migration, including securing committee approval for a practical and proactive plan to ensure progress ahead of any official end-of-life announcement from UIS.

7. Any Other Business [RM]

8. Date of next meeting(s) [RM]

(i) Confirmation of the date, time and location of the next meeting(s).

Date: 21st January 2025
Time: 14:00 - 15:00
Location: SW00

Date: 29th April 2025
Time: 14:00 - 15:00
Location: SW00

Date: 17th June 2025
Time: 14:00 - 15:30
Location: SW00

ITSC report: Infrastructure December 2024

Malcolm Scott

Email forwarding

We now have a firm deadline from UIS for replacing our legacy email forwarding: they will be turning off their central mail routing and spam/malware filtering on 6th January 2025.

We are testing a replacement for this facility, outsourced to Forward Email but integrated with local data sources, and plan to switch our email domains across to Forward Email around Christmas.

One issue has arisen with a legacy format of email addresses (prefix-based tagging, whereby users can filter mail by placing an arbitrary tag *before* their username: prefix+username@cl.cam.ac.uk). Very few people are still using this, but at least one has hundreds or thousands of email address variants in use going back several decades. At present Forward Email do not support this. I am in discussion with them to try to find a way to support this, but it is quite likely that we will have to discontinue this addressing scheme at short notice.

Network upgrade

The department has agreed to provide funds to replace all of our network hardware over the next few years (and ongoing on a 10-year cycle). We will start by purchasing new hardware to be put into service as firewalls (taking that function off the core switches/routers “gatwick" and “heathrow”), and then expect to proceed to replacing the core switches/routers themselves.

New GPU servers

Three new GPU servers for ACS projects are now up and running. One is a shared development server with eight low-power NVIDIA L4 GPUs. This is an experiment to see whether low-power GPUs are useful in a development server. We have an alternative with four high-end NVIDIA L40S GPUs on standby ready to take over if this proves problematic as ACS projects progress.

We have just this week received confirmation of EPSRC funding for a shared departmental GPU server for research, equipped with four L40S GPUs, to replace the current shared GPU development server with one A100 GPU.

GPU cluster storage incident, 17 November

On Sunday 17th November during a scheduled overnight storage integrity scan, the storage server “tuffi” which holds user home directories and VM disks for the GPU cluster detected data corruption on several of its SSDs, affecting both mirrored copies of its data.

Whilst investigating the problem with tuffi, it became apparent that data was continuing to become corrupted, including data that had not been written to recently. With no clear cause for the corruption known, and not knowing whether the hardware or the software was at fault, I set about recovering the data to alternative servers. Due to cost constraints we do not have a dedicated spare server for tuffi, nor even a complete backup, so space was made on various alternative servers where we happened to have, or could add, spare SSD capacity and the data

was copied from the live server as far as possible, in combination with partial backups from another server.

Nearly simultaneously, server “hattie” which holds partial backups of tuffi (VM disks only) had also logged a hard drive fault, removing its resilience. This would have been routine, and it is not the first time recently that a disk had failed in hattie, but the coincidental failure reduced our options and slowed down the recovery of tuffi’s data as hattie’s performance was substantially lower whilst it rebuilt its RAID array onto a new disk.

The GPU cluster was heavily disrupted for five days, though for most of that a read-only snapshot of home directories was made available. The current solution is not permanent, with a large quantity of research data (tens of terabytes) occupying storage that was purchased for other uses. We have bought a new server to act as a replacement for tuffi (whose warranty has expired and the service will benefit from newer hardware) at a cost of £38,300 including just enough storage to accommodate the existing data, but no spare space. Additional SSDs may be needed at a further cost, depending on whether tuffi’s disks can safely be reused to provide additional capacity.

We hope that tuffi can remain in use as a backup server, subject to further tests – though we still do not know the cause of the data corruption and must proceed carefully.

Network disruption incident, 6 November

On the afternoon of 6th November, one of our core routers (“heathrow”) suffered a resource exhaustion issue which caused network traffic instability for a couple of hours, particularly as the connections to the University network from that router kept going down briefly.

This is not the first time that we have experienced such an issue. I believe that if the routers briefly have a big burst of work to do for any reason, they can enter a cascade failure state whereby protocols time out, causing more and more work for the router CPU to do to attempt to regain connectivity, which in turn causes other protocols to time out. The routers seem not to recover on their own; on this occasion I had to temporarily configure that router not to handle any routing and to allow its partner (“gatwick”) to take over all routing for long enough that heathrow’s processing backlog could be cleared. I believe this to be a design flaw of the current Cisco hardware and OS.

Moving firewall functionality onto new hardware as described above will mean that the routers’ CPUs have less work to do, which should reduce the frequency of this class of issue until we are able to replace the routers themselves.