**Internet2 AL2S**
**Triannual Report**
**8-1-13 through 11-30-13**

## Introduction

This is the third in a series of periodic reviews of the Internet2 AL2S infrastructure. The goal is to examine what is working and what isn't with an eye toward improvements and a focus on information important to the community for their decision-making process.

During this reporting period several major programs have been implemented, from additional documentation to more rigor in operations, and significant improvement in managing deliverables. On the hardware side, this reporting period is marked by several RMA's that caused significant issues as well as a required feature and a (non-forwarding impacting) SDN bug.

There will be some overlap between this report and the previous. The previous report came out late and therefore covered a major incident (Sept 3) that technically occurred outside of the reporting period. This report covers that incident again, updating information.

The raw data that was used to create this report is contained in the AL2S Weekly reports, compiled in to spreadsheet form. The data is available at: https://www.internet2.edu/products-services/advanced-networking/layer-2-services/al2s-reports/

## Report Overview:

*Upgrade Details*
*Unscheduled Outages/Incident Management*
*Scheduled Outages/Change Management*
*Post-Mortem Improvements*
*Availability Information*
*Bandwidth Utilization*
*OESS Utilization*
*Progress from Last Report Deliverables*
*Future Plans for Layer 2*
*Timeline of Installs/Upgrades*

**Upgrade Details**

Overall, during the reporting period, we saw a decrease in the number of vendor software loads and open source loads. This reflects the maturing of both the vendor code and the applications that support the services.

There were no vendor software loads (Brocade or Juniper) during the past four months. This is in contrast to the last reporting period in which there was a code load on the Brocades and Junipers and several enhancements of the OESS application.

There was one major upgrade to the OESS software during the reporting period. A community-focused webinar was held on 10/22/13 to discuss proposed changes to workgroup management, at which positive feedback on the approach was received. The OESS software was upgraded to version 1.1.0 on October 30, 2013, implementing the new workgroup functionality. There have been no requests for changes, reversions, or improvements to the new workgroup management functionality since. We introduced one new piece of production software during the reporting period: FOAM. This GENI aggregate manager acts as an application interfacing through the OESS API, facilitating the integration of AL2S resources into the GENI environment.

The OESS applications components may see additional upgrades to add minor features as feedback is received from users, however major development work of the core feature set is complete.

**Unscheduled Outages/Incident Management**

The largest impact to availability for AL2S remains maintenance or fiber cuts on the underlying L1 circuit infrastructure. The ability to offer resiliency, either at L2 or at higher layers, remains extremely important to mitigate particularly the uncontrollable external influence of long duration fiber cuts. There was a rise in undetermined issues this period, showing the "<600ms link flap" issue.

There was a major incident with a flapping backbone circuit between CLEV & STAR that was tracked back to a failing line card in a Brocade. This happened during the same time period that the 'mystery flaps' were occurring and the two problems were initially lumped together with them. Once the signature was recognized, the line card was replaced and the problem was repaired.

The September 3rd incident was the result of a failed fabric module in a Brocade. The error was initially identified to be an Amazon issue but once identified as belonging to the Internet2 switch the issue was quickly resolved.

A major data user in the Midwest encountered a pattern problem. Bit patterns of a certain type were being dropped by a Brocade. Enabling FEC on the Brocades scrambles the bit pattern on the backplane and resolves the issue. Note that this is the ONLY issue identified by Brocade as requiring FEC; it is otherwise unnecessary. Statements to the contrary have distracted from the successful resolution of several incidents.

There was a second major issue in Denver on November 29th, resulting in 11 hours of unavailability on a specific linecard. A reboot of the linecard and fabric module reset resolved the issue. The improved performance monitoring put in place after Sept 3rd immediately detected the issue (up to 50% packet loss.)

**Scheduled Outages/Change Management**

During this reporting period there was an improvement in our ability to enact change: fewer changes were rescheduled. In addition there were fewer changes that delivered unexpected results. At the same time there were significantly fewer Software changes in this reporting period, with config changes to the nodes had the largest impact on change. FEC enabling required a reboot of the Brocades, was the cause a large number of these, resulting in a spike occurring in September.

Overall there is a general trend of L1, Hardware, and Software changes all trending downward, reflecting a more mature infrastructure and better active management during the last reporting period.

**Post-Mortem Improvements**

A number of incidents during this reporting period have resulted in post-mortems being conducted, a goal from the last quarter. The following is a list of improvements implemented in the AL2S cycle:

Formalize Post-Mortem Process

While an AL2S post-mortem exists no post-mortem process existed for the other services. A new post-mortem process now exists.

Formalize & Streamline Support Process for Net+

There has been confusion regarding communication and support responsibilities with regard to Net+ connectors and services. The support guidelines have been clarified and communication procedures updated.

Reexamine Change management process

Additional documentation has been created providing guidelines on what does and not require formal change management procedures to be engaged.

Update BC/DR Risk Assessment/Process

A critical communication path has been created for very high priority issues. This leverages the communication guidelines for BC/DR events.

Review & Update Escalation Matrix

The GlobalNOC & Internet2 escalation matrix has been updated with contact data and timelines refreshed.

Update Performance Assurance Infrastructure

Soft issues, such as loss, are as important as hard outages. Monitoring was put in place to immediately detect loss situations. This interim performance monitoring is being replaced by a permament perfSONAR infrastructure during Q1 2014.

Update Monitoring for Fabric Signature

Additional monitoring has been put in place to detect the specific signatures of the issues, including egress packet discards and fabric issues.

Drive Resolution with Vendors

Provide new documentation to engineers regarding matching NOC priority with vendor priority and riving issues to resolution with prompt escalation.

Move to Service Restoration Sooner

Explicit guidelines have been communicated with expectations set with engineering staff regarding the amount of time to spend troubleshooting and gathering data vs. restoring services.

Improved Categorization of Priorities

Initially the SmokePing alert was not categorized as an outage but rather an Inquiry. The documented process is being changed to ensure that the duty engineer reviews the ticket type and priority as a part of the initial set of duties, with the service desk ensuring follow-up

## Availability

Availability remains acceptable. While there are occasional drops in the 1-week availability figures the general trend is the improvement in 52-week availability as old issues aged out. A good example here is the ATLA-WASH circuit. CLEV-STAR is also a good example since we see 52-week availability increase as issues time out and then the 52-week availability drop off again (while still remaining well above 3 9's) as 1-week fiber issues occurs.

Nodal availability is more interesting since L1 impacts it less and there is less redundancy. With Scheduled Maintenance removed (IE: software loads) we have had 100% availability for quite some time. While there were no software loads there was a Scheduled reboot to enable FEC, showing up in the weekly figures, but the impact on the 52-week figures is so small that the impact is less than 4 9's. IE: We round to 100% because we are only showing 2 decimal places. Our overall 52-week nodal availability is 99.97 if we include Scheduled Maintenance. If we remove that, as is the industry standard, we improve to 99.989.

The Nov 29 Denver outage shows up in the node availability figures. The impact of this was a linecard, rather than a single circuit/port or a entire node. The system can't distinguish this so we've chosen to list the issue as node impacting, resulting in the lower figures this reporting period.

**Bandwidth Utilization**

95% bandwidth trends seem to indicate that we can expect several of the 100gig backbones to go over our “40%” headroom threshold around the March/April timeframe. The steepest slopes are where one would expect to see them, around Ashburn and on the busy EAST-WEST path through Chicago.

Trending the bandwidth four months out reveals that two circuits, ELPA-PHOE and PHOE-LOSA will be at or close to the 40% headroom threshold. This is primarily because of a couple of large spikes bumping up what would otherwise be a low slope value. Should the data spikes continue then those two circuits will need to be upgraded prior to April 1, 2014. Several other circuits should reach 30%-35% utilization by April 1, 2014 and 40% by the April-July 2014 reporting period.

**OESS Utilization**

OESS usage, the manner in which users configure VLANs across the infrastructure, has seen a decent uptick over the reporting period. We are up to 61 workgroups, about 50% more than the last report. Similarly, there has been a decent bump in the total number of users, the number of configured circuits has almost doubled to over 300, and many more ports are under management.

There were 2,229 circuit events in this reporting period, about 2000% more than the last period. Almost half came through the OSCARS IDC. Additional detail about those circuits would be useful. What is very interesting is that the other thousand did NOT come through the OSCARS IDC and are relatively spread out. That's an order of magnitude more usage. It will be interesting to see if that's just Supercomputing related or if the trend continues.

**Progress on previous report deliverables (From the August-November report)**

Good progress was made on the deliverables from the previous two reporting periods. While 10 or 11 items showed no progress during the pervious period, this period shows 11 complete, 5 with partial progress, 1 explicitly on hold, and 5 with no progress. This is a significant improvement in light of the reduced timeframe between the last report and this one. Of the tasks with No progress, several did not make it on to the daily tracking sheet. This is being solved through explicit review of the entire deliverable structure once a month.

- The reporting tools clearly need the ability to differentiate between Scheduled & Unscheduled outages to bring I2 in to line with best practices.
    - **Complete. The tool has been modified and now makes the distinction as of the Aug 11, 2013 report.**
- Availability reporting should shift from an individual elements/nod/circuit model to a Services based model.
    - **No deliverable during this reporting period although the modifications of the backend systems have begun to shift from a node/circuit model to a Services based model. In addition, a new Active Performance Monitoring infrastructure has been approved and is now in the process of being implemented. This will allow for better monitoring & reporting of internal & external availability of services. It is the first major deliverable in moving towards the shift to Service Based Availabilty.**
- Another tracking category Is needed. "Impairments", representing something other than a binary "Up/Down" status.
    - **No progress during the reporting period, although progress should be reported during the next report as a result of actions taken as part of the Performance Assurance portion of the "The Amazon/FEC" post-mortem.**
- More granularity is needed in the reporting system to differentiate the layer of the service impact. For example,L2 outages caused by L1 circuit outages.
    - **Complete. The Incident Management tracking and Change Management tracking have both had additional fields added in order to track these.**
- Further support training is needed for the engineering and systems support staff.
    - **Complete. Staff have received additional training in the use of the AL2S network. This goal will be modified to become a standing goal in each reporting period.**
- Better tools are needed to identify and localize problems. The Layer 2 environment is sufficiently different to require reexamination of the troubleshooting process.
    - **Partial Progress: A software development effort is underway to develop a L2 Ping/trace capability to enhance both internal and**

**external troubleshooting. While the tool hasn't been delivered yet, it is in the dev cycle progress.**

- A better tracking system for major project events (upgrades, major bugs, etc) is under way.
    - **Complete. Major event tracking is now available and updated weekly during the AL2S weekly management meeting.**
- The community groups are working on a set of metrics to better measure & compare the Layer 2 system.
    - **No progress during this reporting period. While the NTAC face-to-face meeting was a success, the meeting notes, and thus action items & deliverables, are only partially transcribed. The Metrics subgroup draft has not been completed yet.**
- The backbone upgrade policy is vague. The only current policy is based on a Layer 3 IP network running at 10gigabit speeds and handling generalized R&E traffic.
    - **No progress during this reporting period. While the NTAC face-to-face meeting was a success, the meeting notes, and thus action items & deliverables, are only partially transcribed. The Metrics subgroup draft has not been completed yet.**
- The Layer 3 backbone will begin using Layer 2 (via vlans and SDN-signalled circuits) for 100g transport.
    - **Partial Progress. The transition effort was begun but then halted when the <600ms flaps were noticed on the circuits. Development of a remediation plan is underway. A Problem Resolution Plan is being developed to ensure that the issue is resolved.**
- Openflow 1.3 support is heavily examined on OE-SS, FlowSpaceFirewall, and the backbone switching nodes, with a desire to move to it quickly to support several gaps in the 1.0 standard.
    - **A draft recommendation is complete. It is being updated with additional use cases and user feature requests enabled by 1.3.**
- A Service owner for AL2S to drive the service from more than a technical viewpoint. While this report focuses on the technical it is clear that's not the goal for AL2S. AL2S exists to enable research and higher-layer services. More focus on that can be brought through quarterly goal setting. We need to develop targets for inclusion on the next report that are more than just technical.
    - **Complete. A service owner is now assigned to AL2S and is driving forward on non-technical deliverables, such as better documentation and metrics outlining AL2S usage.**
- More assertive management of the deliverables. The deliverables must get on to the development roadmap faster and be fleshed out with dates assigned. It is critical that we be able to deliver on commitments.
    - **Partial Progress. A new system based on weekly sprints is now in place, with more flexibility in the hands of the staff. It is still being refined, however there has been a *significant* increase in deliverable**

**success during the last four months.**

- Better management of Brocade. We must be much more focused in our dealings with Brocade. Cases must be opened and aggressively engaged on every incident and we must strictly follow their escalation policies for each one. In addition, a call should be arranged between the community and Brocade to help explore the issues around the boxes.
    - **Complete. Guidelines have been drafted and reinforced with the staff about the severity of the issues and the escalation process within Brocade. In addition, Brocade has been notified that Internet2 will be formally reviewing their progress towards better support. This review to take place in early February 2014. Brocade & Internet2 are now performing an explicit joint post-mortem on each issue to ensure gaps are resolved. We are working cooperatively to ensure a successful medium-term outcome.**
- As indicated in the Unscheduled Outage section, we should explore the impact of SDN circuit failover churn on L2. This should probably start with measurement.
    - **Complete. A variety of new 'churn' testing scenarios are being integrated into the labs suite of tests that are performed before each new code release. The larger academic examination has been shelved due to other priorities.**
- In the Change reports differentiate between switch software and non-switch software loads.
    - **Complete.**
- Emergency Change was high in the last two moths (outside of this triannual report but in the next one) because of the reboots to enable FEC. It was a risk management decision: do we perform an emergency change with the risk that a shortened time window implies … and risk the 'silent frame drop' issue reappearing or do we perform a normal Change after a longer notification window? Better understanding of what causes us to perform an Emergency Change would help us make those decisions quicker and better.
    - **No progress during this reporting period, although it was brought up during the recent I2-IU Quarterly Operations Review as a topic to be addressed.**
- We should expand deployment of the active performance-monitoring infrastructure. This acknowledges the gap in our systems and the real goal not being "all links/nodes up" but rather "the community can transfer data error free." This would mirror the shift from thinking of AL2S-as-a-network to AL2S-as-a-service.
    - **Partial Progress. No deliverable during this reporting period, however a new Active Performance Monitoring infrastructure has been approved and is now in the process of being implemented. This will allow for better monitoring & reporting of internal & external availability of services. Of particular import is how this relates to Internet2 supporting Data Intensive Sciences.**
- Similarly, we need to look at the way we measure Service Availability on AL2S.

While link up/down may play a part in network availability the service availability should probably transition to something dealing with active performance monitoring. We should figure out a position on this in the next reporting period, at a minimum.
    - **On Hold. This is being delayed until after the Active Performance Monitoring deployment in order to gain additional experience.**
- While not directly related to AL2S, it is worth mentioning that we need to better understand what we do in relation to our connectors having 100% availability. At a minimum we should examine all of the L2 & L3 connectors and determine, in conjunction with them, if they have the physical infrastructure/connections in place to support 100% availability of services and what we can do to help beyond current incentives for dual homing and regional collaboration. Perhaps a goal for the next reporting period could be to examine our processes to ensure we don't step on backups during Scheduled/Unscheduled events.
    - **Partial Progress. A project description and timeline were developed and progress was made toward identifying duel-home connectors and working with them to verify backup capabilities. This is likely to be rescoped as a major 2014 initiative for Internet2.**
- We must finalize the headroom policy at L2. The existence of large research flows must be taken in to account.
    - **No progress during this reporting period. While the NTAC face-to-face meeting was a success, the meeting notes, and thus action items & deliverables, are only partially transcribed. The Metrics subgroup draft has not been completed yet.**
- We must protect higher-layer services on L2, be they TR-CPS, member backhaul to L3, or L3 backbones. We need to come up with a plan and deploy it in the next ternary.
    - **Complete. The Layer3 backbone traffic now has priority over native Layer2 traffic on the AL2S backbone.**
- We should embed an engineer in to the weekly/bi-weekly technical calls of our largest Big Data users. This should allow us to provide a higher level of service to these critical users and also to get ahead of problems that tend to show themselves first in this cohort.
    - **Complete. An engineer has been involved with one of the Big Science projects. We now have a process around adding him to additional Big Science/Data projects on a "1 per month" basis until the major projects are covered.**

**Goals for the Next Reporting Period (December 2013-March 2014)**

- We plan to improve our capacity planning in the next four months. Bandwidth trending should be a routine part of reporting and we should have explicit reports noting low port & slot capacity on the AL2S infrastructure.
- We plan to improve the tracking of our users usage of the infrastructure. The goal

here to identify sites that could use more outreach and education, and ultimately reduce the barriers to usage.

- We plan to improve the descriptions of our services. Descriptions are lacking for AL2S support of OESS, dynamic, circuits, static circuits, inter-domain circuits, etc. Further the bootstrap infrastructure, NDDI, is not well documented, nor is its capabilities.
- We should do a better job institutionalizing our data. The current process generally revolves around sending e-mail to various groups. Proposals and plans should be archived and available through the Internet2 website on a routine basis.

**Other Improvements made in the last ternary:**

- Vendor QBR's are now routinely occurring, with actions items being tracked.
- A scorecard for service improvement now exists. This is tracking our ability to define, fulfill, and improve on the various aspects of our service delivery. Examples include Incident Management, Change management, Post-Mortem process, Customer Requests, and so on.

**AL2S Service Roadmap, next Ternary**

Upgrades

It is likely we will to move to a GA release of the Juniper code in Q1 2014 and to the 5.6a version of the Brocade code in Q1 2014. Promised drivers include sFlow, converged layer2 and 3 matching, bigger table sizes, and improved flow mod performance.

Virtualization

We expect to begin to deploy components to support Virtualization in Q4 2013 and Q1 2014. This will include the introduction of FlowSpace Firewall between the switches and OESS. A community webinar about this plan is scheduled for December 19th.

Multipoint VLAN Support

We expect to upgrade the OESS software to 1.1.1 in Q4 2013 to support static multipoint VLANS.

**AL2S Timeline**

2013-08-01 – Change, Availability, and Post-Mortem tracking improvements
2013-08-11 – Tracking of Scheduled vs. Unscheduled outages implemented
2013-08-18 – Micro-flaps noticed
2013-08-26 – CLEV-CHIC flapping fixed/hardware RMA'd
2013-08-30 - Installation of the Pittsburgh AL2S node.
2013-09-02 – "The Amazon Issue"
2013-09-04 – Installation of the Portland AL2S node.
2013-09-15 – FEC enabled on Brocades
2013-10-22 – Community Webinar (Workgroup Administration)
2013-10-30 – OESS 1.1.10 deployed.
2013-10-24 – FOAM deployed.
2011-11-8 – NIH up in Ashburn
2013-11-14 – 3ROX up in PITT
2013-11-29 – Denver Packet Loss