Nova-doc-8658
CS-doc-5012

NOνA Far Detector Networking Review Report

Kurt Biery, Andrey Bobyshev, Stu Fuess, Ruth Pordes, Tammy Whited and Stephan Lammel

Feb. 14$^{\text{th}}$, 2013
version 1.0

The large, 14,000 ton detector of the Fermilab NuMI Off-Axis ν_e Appearance Experiment, NOνA, is located over 500 miles outside the Fermilab campus. It will be operated and controlled from Fermilab and will also send its data there. Connection to and networking at this far detector are paramount for the success of the experiment.

The NOνA deputy project manager requested a review of the networking for the NOνA far detector. The Scientific Computing Division, SCD, NOνA liaison convened a review in January 2013. The committee reviewed and discussed requirements, design, implementation, support, documentation, procedures and evolution of the NOνA far detector networking over a several weeks period. The observations of the committee are documented in this report.

Executive Summary

The network setup at the NOνA far detector as described in [1] meets the documented requirements of the experiment. Additional requirements exists and should be specified. The committee is confident that the network setup can be adapted to fulfill those.

Roles and responsibilities in the support of the network at the far detector are not well defined. Support agreements with Fermilab, University of Minnesota and other parties are incomplete. As a result, lines of communication and approval are not clear and are thus not working well. This should be remedied at highest priority. In preparing/negotiating the support agreements we recommend to consider that NOνA needs will be shifting from construction/commissioning to operations/data taking during the next years.

Computer security of the far detector network should be reviewed. Procedures for rapid network isolation, operations during network isolation, access requirements to VLANs during commissioning and during data taking need to be developed. We recommend NOνA to write a minor application security plan and to submit it for review to computer security.

Procedures need to be worked out for several tasks: monitoring, fault detection, problem diagnosis, emergency repairs, power outages (including partial) and power up sequence. Those can only be prepared after roles in the support and operation of the network are agreed upon.

Working out a complete set of requirements, the roles of people and groups, the responsibilities of each, a computer security plan and the various procedures will require significant technical effort and we expect this to lead to design updates/revisions and changes and/or additions to the network setup (and potentially even other infrastructure at Ash River). The scope and depth of these modifications will be driven by how special the far detector networking requirements are and will become clearer as requirements are better understood.

1 Introduction

The Internet Protocol, IP, network at the NOνA far detector at Ash River, Minnesota consists of a pair of Cisco Nexus 5548 routers, Cisco 4948 and 2960 concentrator switches, wireless access points, device management appliances and cabling infrastructure provided by Fermilab and a network setup with 500 MBit/sec digital subscriber line, DSL, uplink provided by the University of Minnesota, UMN. The Fermilab network provides five subnets, for data acquisition (public/private), detector control systems (public/private) and general purpose use. The UMN network provides guest, voice over IP telephone service and wide area network, WAN, access [1].

The pair of Cisco Nexus 5548 routers builds the core of the IP network. Over 20 Cisco 4948 and 2960 switches with dual uplink (one to each Nexus) provide the network connections for the NOνA far detector equipment. One of the Nexus uplinks to the UMN network and thus provides the WAN connection to the main Fermilab campus while for LAN traffic the Nexus switches provide redundancy. Figure 1 shows the current network setup.

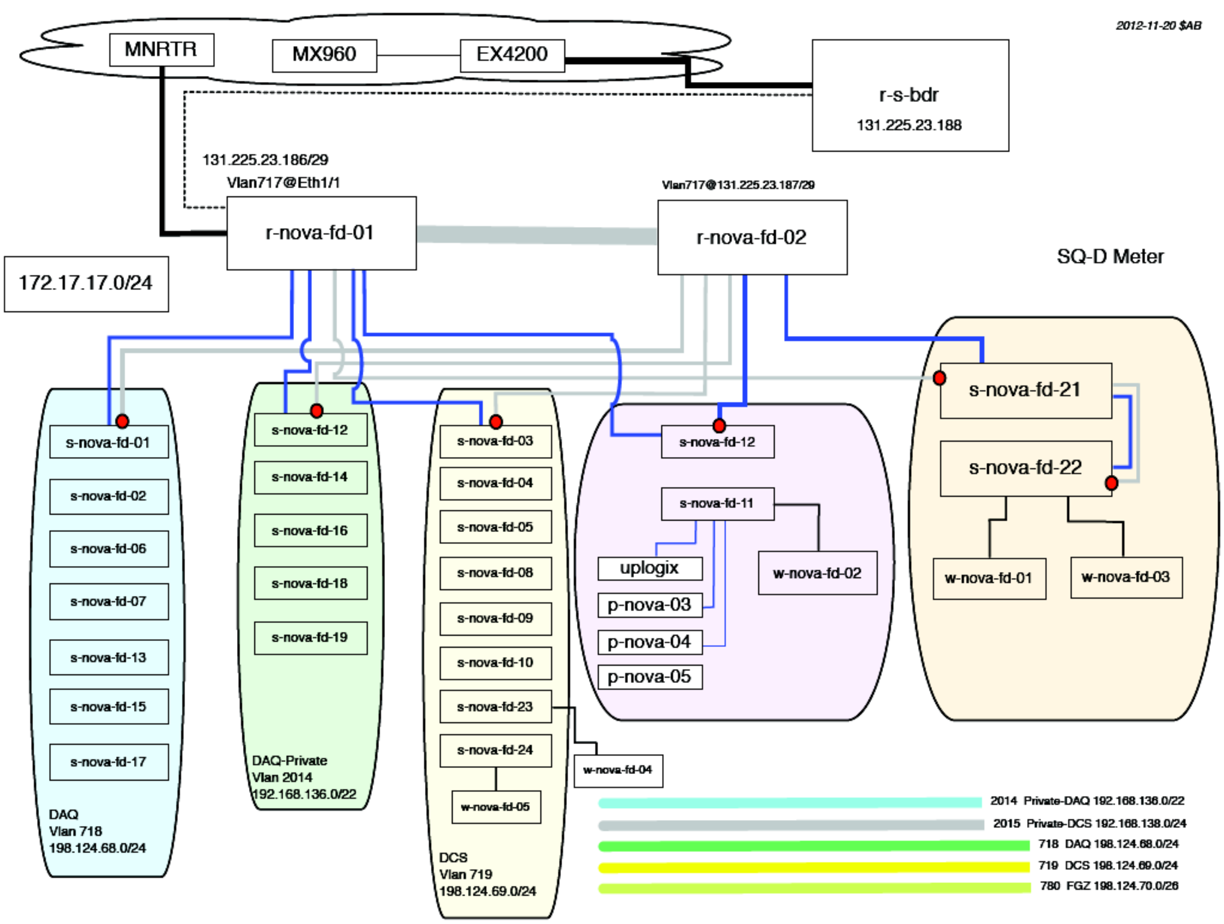

Figure 1: Current network topology at the far detector.

The far detector is currently being built. It is designed to have 385, 152 avalanche photodiodes, read out by 14, 000 front-end boards that are connected to 194 data concentrator modules, DCM. The design of the data acquisition system, DAQ, connects the DCMs via Gigabit Ethernet to a farm of 136 personal computers, PC, that buffer and process the data [2]. The DAQ will operate un-triggered, i.e. read-out, digitize, time-stamp and zero-suppress data continuous. The beam spill signal is expected to arrive over the WAN from Fermilab within about 10 seconds. NOνA estimates for the far detector to generate about 449 TBytes (after compression) of physics and calibration data a year [3, 4].

2 First Two Charges

The first charge to the committee asks:

> **Are the designs for each of the NOvA networks adequate to address the technical and operational requirements of the experiment?**

and the second charge:

> **Do the network hardware and configurations that have been deployed to Ash River satisfy the designs of each of the NOvA networks?**

The committee interprets "each of the NOvA networks" to mean the Fermilab provided virtual local area networks, VLAN, at the far detector site. Looking through the NOνA documentation the committee found estimates for WAN bandwidth [3, 5], network connections [6] and local area network, LAN, bandwidth [2].

The committee takes the WAN bandwidth requirement to be 204 MBit/sec with a potential desire of the experiment to double the 144 MBit/sec data rate in there (by lowering thresholds) in the future. The LAN bandwidth requirement is dominated by the DAQ data flow from the DCMs to PC farm of currently 43 GBit/sec.[1] Network connections are dominated by the number of DCMs and nodes in the PC farm.

The committee believes that the design and implementation of the IP network at the far detector meet the documented requirements listed above. The current WAN bandwidth exceeds the required bandwidth even in the case of lower signal thresholds. Available network ports and LAN bandwidth meet the requirements above. The Ciscso Nexus 5548 are switches/routers that provides 960 GBit/sec throughput[2]. Both DCMs and PC farm nodes are connected to Cisco 4948 switches with 10 GBit/sec uplinks. The number of concentrator switches could be doubled if additional bandwidth or ports are needed.

However, the committee considers the network requirements of the NOνA experiment listed above to be incomplete.

- **There is no uptime requirement for the network.** Although the pair of Nexus 5548 provides redundancy at the core, failure of a concentrator switch will cause outages of DCMs, buffer nodes and/or control systems. Due to the remote location repairs of the network may take hours or days.

[1] With 194 DCMs each capable of sending data at 40 MBit/sec the highest possible rate currently would be 76 GBit/sec.

[2] The DAQ DCM-to-PC farm traffic is inside a VLAN, so the layer 2 performance is the relevant number here.

- **Is there an uptime requirement for the WAN link?** The technical design report calls for disk space at the far detector to buffer data for weeks. However, a WAN outage will also interrupt access to manage and control the detector. Are there any critical, high-voltage, high-current systems that require un- or effectively uninterrupted network access? The proposed backup WAN link will probably share network infrastructure (outside Ash River and Fermilab) and should not be considered independent!

- **Networking services required at the far detector are not listed.** Does the experiment require domain name system, DNS, time, NTP, dynamic host configuration protocol, DHCP, Kerberos or other services to function? Are slave servers required at Ash River?

- **Access to the VLANs is not specified.** Fermilab network services group has forseen access control lists, ACL, for the far detector VLANs. The experiment needs to work out access requirements for the commissioning and operations phase of the experiment. A minor application security plan should probably be developed and submitted for reviewed to computer security, CS. (Fermilab Run II experiments had three levels of VLAN access: access from anywhere, access from only within Fermilab and access from only within CDF VLANs.)

- **Roles in the support and operation of the network need to be agreed upon and documented.** The Fermilab network services group installed and configured most of the network. However, several other groups are involved: UMN networking staff, members from the DAQ group and people located at Ash River commissioning/debugging detector components. Do people know their role and responsibilities and that of others and in the case of extraordinary/emergency circumstances?

- **Support agreements with Fermilab, UMN and even the DSL provider should be reviewed (or established) and documented.** Default support of Fermilab[3] is during the eight working hours on the five workdays a week, 8x5. This is rarely appropriate for an experiment during data taking. Network support and operation should be included in the memorandum of understanding, MoU, between NOνA and SCD [8] and the MoU maintained during the lifetime of the experiment.

- **Procedures for work on the network need to be established.** Most experiments demand zero changes during science data taking and DAQ/shift notification/hand-shake when the change is made. While Fermilab computing has implemented change management, a customization for NOνA may be desirable/needed. In case of an equipment failure are people at Ash River expected/available to perform (emergency) network repairs? Is there an agreement on cross-group assistance? Are people trained?

- **Procedures for fault detection and diagnostic.** What are the requirements on problem detection and troubleshooting? Are current tools sufficient, what is automated, who invokes them manaually, are they documented and known to DAQ and detector subgroups members? Who is responsible for detecting/preventing network loops (not unlikely with a UMN and several Fermilab networks)?

[3] The Fermilab Computing Sector developed a foundation service level agreement, SLA, [7] defining common terms, definitions and service needs/expectations to simplify drafting/negotiation of new SLAs.

- **What is the requirement on the network in case of a partial or full power outage?** The technical design report specifies backup electrical power for critical systems. Is networking considered a critical system? Is networking required to control equipment that is on uninterruptible power supplies, UPS?

- **How soon after a power on must networking be available?** Are devices set consistently to stay off, power up or fully boot?

- **What is the requirement on the network in case of a fire alarm or cooling failure?** Is an emergency power off system implemented at the Ash River facility? Is there a shutdown procedure in case of a cooling failure, is it automated, does it rely on networking? Are networking and the serial line servers required to power down gracefully high-voltage or high-current equipment?

- **What is the requirement for a rapid network isolation in case of an intrusion attempt or detection?** Is there a plan on how the NOνA far detector network would be isolated? Run II experiments tested DAQ and detector controls with an isolated online to verify mandatory controls and data logging are working properly.

- Does NOνA require jumbo frames in the future (to achive higher transfer rates) for DAQ or data logging to Fermilab?

- NOνA will be using IP multicast to send control messages. This should be documented and included in the network requirements.

- **Are there networking requirements in case of extreme environmental conditions?** The Ash River facility will be staffless and in a remote and very northern location. Networking equipment operates in a limited temperature and humidity range. Environmental conditions may deteriorate quickly in case of heating loss or breach of the building during the winter. In case monitoring of environmental conditions relies on the network, updating needs to be frequent relative to a worst case scenario.

- **Are there special requirements for the wiring infrastructure?** If the probability of rodent investation is significant then metal tubing may need to be considered.

- **NOνA may want to establish a procedure to turn off wireless networking around the detector.** The collaboration reviewed the frequencies used by wireless networking. New wireless networking standards and future NOνA electronic may make this desirable or even necessary.

3 Third Charge

The third charge to the committee asks:

> **Is the documentation that has been compiled by NOvA sufficient for support and maintenance of the network as the experiment noves into the build/operations phase of the far detector?**

The current support agreement between NOνA and the Fermilab computing sector, CS, is described in the *NOvA Construction Project Computing MOU* [8]. As the experiment moves from construction and commissioning into operations and data taking, needs change, revised support should be negotiated ahead of time and the MoU amended to document the new agreements.

The NOνA collaboration has been collecting information about the network at the far detector as it is being setup. It keeps copies of technical notes, status presentations and list of network connections in its document database.

The committee feels that this documentation is insufficient for problem diagnostic, troubleshooting and emergency intervention/repair by NOνA collaborators. There may be no plan for this. Please see the items about roles, support agreement and procedures for fault detection in the previous section. We like to point out that:

- Fermilab computing sector people maintaining and operating the far detector network will need to access, write and update documentation for NOνA and documents required by CS. Setting up both the NOνA and CS DocDB with proper groups (a CS group in the NOνA DocDB and NOνA group in the CS DocDB) and properly authorizing documents is recommended.
- While the document database of NOνA is an excellent tool for collecting and organizing documents, a web page (both at the far detector and Fermilab campus) may provide better and faster information and automate/aid in monitoring, problem diagnosis and repair.
- In case NOνA collaborators or UMN people assist in the support and maintenance of the far detector network they need access to monitoring and diagnostic tools.
- In case NOνA collaborators or UMN people assist in the support and maintenance of the far detector network training (and material for the training) will be required.
- In case NOνA collaborators or UMN people assist in the support and maintenance of the far detector network the committee recommends complete and easy visible labeling of network devices and up-to-date pictures of racks and devices.
- Preconfigured spare ports for the various VLANs have proven to be valuable in the Run II online environment in case of port/switch problems.

4 Fourth Charge

The forth charge to the committee asks:

> **Are there appropriate procedures in place for affecting changes to the network configuration at Ash River?**

The committee believes that appropriate procedures to manage changes to networking infrastructure are in place but the contacts in NOνA to inform, review and/or approve are not known. (See the item on roles in the first section above.)

Fermilab networking services are under change management [9], i.e. use an established process and procedures to efficiently handle changes to the network infrastructure that minimize risk and disruption to networking services. The process and its procedures are not only considered appropriate for the main campus and core IT networking infrastructure but also found appropriate by other running experiments. NOνA is encouraged to review, subscribe and/or customize the procedures involving the far detector network.

5 Fifth Charge

The fifth charge to the committee asks:

> **Are there appropriate procedures in place for incident reporting and for tracking incident resolution?**

The committee believes that appropriate procedures to report and track incidents are in place but that SLAs for networking at the NOνA far detector are incomplete. (See item on support agreements in the first section above.)

Fermilab operates a Service Desk to record, classify, route and and track requests and incidents as a single point of contact for users. The ServiceNow tool used by the Service Desk provides views of open and past requests and incidents and customizable metrics. The Service Desk is reachable via email and phone. ServiceNow is accessible via web browser and Fermilab Service account/password. The web based self-serve access is the prefered way to report and check on incidents. We note that in case of network problems at Ash River incident reporting from the site may be limited to the landline phone. For the Fermilab Service Desk to route and resolve incidents effectively the required support agreements need to be in place.

6 Other Observations

While reviewing networking the committee made a few observations that it likes to share:

- The WAN uplink serves both UMN network and Fermilab provided networks. Thus network traffic on the UMN network may adversely affect data logging to the Fermilab campus.
- Similarly, with the small and finite WAN bandwidth, users transfering large amount of data to/from the far detector for backup, monitoring or diagnostic may adversely affect data logging to the Fermilab campus. The experiment may want to take a proactive approach and establish a procedure/registration for automated transfers (at/cron jobs etc.).
- Computing and networking equipment seems to share the same power distribution units, PDU. What is the procedure for accessing the PDUs, how will the experiment prevent networking from accidentally power cycling computing equipment and vice versa?
- Is the spill server signal an accept or reject? In case of a WAN outage, a lost reject message will result in more data but a lost accept may result in data loss.
- Network traffic not associated with DAQ or detector control could easily increase such that it would cause saturation of the much smaller backup WAN bandwidth. NOνA should actively monitor the network traffic and have procedures in place to reduce, if needed, nonessential traffic beyond data logging to the Fermilab campus in case the WAN uplink switches from primary to backup.
- There is no out-of-band access to networking equipment at the far detector. That means that networking is required to access networking equipment for management, i.e. update, shutdown, power-cycle, reboot, and thus the potential to loose remote management access in case of a network problem/failure. This also leads easily to circular dependencies. The committee advises NOνA to have a sequenced power up procedure (from a well defined state of

drained UPS/site power off) that includes verification/monitoring that the components (network switches/services, serial line servers, computers, etc.) are indeed up and functional after each step.

7 Summary

The network setup at the NOνA far detector as described in [1] meets the documented requirements of the experiment. Additional requirements exists and should be specified. The committee is confident that the network setup can be adapted to fulfill those.

Roles and responsibilities in the support of the network at the far detector are not well defined. Support agreements with Fermilab, UMN and other parties are incomplete. As a result, lines of communication and approval are not clear and are thus not working well. This should be remedied at highest priority. In preparing/negotiating the support agreements we recommend to consider that NOνA needs will be shifting from construction/commissioning to operations/data taking during the next years.

Computer security of the far detector network should be reviewed. Procedures for rapid network isolation, operations during network isolation, access requirements to VLANs during commissioning and during data taking need to be developed. We recommend NOνA to write a minor application security plan and to submit it for review to CS.

Procedures need to be worked out for several tasks: monitoring, fault detection, problem diagnosis, emergency repairs, power outages (including partial) and power up sequence. Those can only be prepared after roles in the support and operation of the network are agreed upon.

Working out a complete set of requirements, the roles of people and groups, the responsibilities of each, a computer security plan and the various procedures will require significant technical effort and we expect this to lead to design updates/revisions and changes and/or additions to the network setup (and potentially even other infrastructure at Ash River). The scope and depth of these modifications will be driven by how special the far detector networking requirements are and will become clearer as requirements are better understood.

The committee offers to meet periodically with NOνA, after the collaboration has identified people and effort, to provide consulting and routing for the new people/groups and to review network design and implementation changes and other actions this review triggered.

References

[1] A.J. Norman, *Networking Review Materials*, NOVA-doc-8423 (2013), http://nova-docdb.fnal.gov:8080/cgi-bin/ShowDocument?docid=8423.

[2] J.W. Cooper, *Technical Design Report for CD-2/3a*, NOVA-doc-2678 (2007), http://nova-docdb.fnal.gov:8080/cgi-bin/ShowDocument?docid=2678.

[3] A.J. Norman, *Nova Bandwidth Requirements*, NOVA-doc-7109 (2012), http://nova-docdb.fnal.gov:8080/cgi-bin/ShowDocument?docid=7109.

[4] C. Group, M. Messier and A.J. Norman, *NOvA scientific computing needs*, NOVA-doc-8443 (2013), http://nova-docdb.fnal.gov:8080/cgi-bin/ShowDocument?docid=8443. (The estimated data volume of 7.4 TB/year $*$ 100 in the presentation is for tape storage of two copies.)

[5] A.J. Norman, *Network and DAQ Hardware*, NOVA-doc-5525 (2011), http://nova-docdb.fnal.gov:8080/cgi-bin/ShowDocument?docid=5525.

[6] A.J. Norman, *NDOS Network Configuration Data Sheets*, NOVA-doc-5745 (2011), http://nova-docdb.fnal.gov:8080/cgi-bin/ShowDocument?docid=5745.

[7] R. Kennedy *et al.*, *FNAL Foundation Service Level Agreement*, CS-doc-4042 (2013), http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=4042.

[8] L. Lueking, *NOvA Construction Project Computing MOU*, CS-doc-4768 (2012), http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=4768.

[9] M. Kaiser, D. Petravick and D. Whitten, *Fermilab Change Management Process and Procedures*, CS-doc-3530 (2010), http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=3530.

Charge for Review of the NOvA Network

Proposed Review Committee:
Kurt Biery
Andrey Bobyshev
Stuart Fuess
Stephan Lammel
Ruth Pordes
Tammy Whited

Dear Committee,

The NOvA deputy project manager, Rick Tesarek, has requested a review of the networking for the NOvA far detector. His request included review of a number of different components ranging from the technical design of the networking to the procedural aspects of the operational support for the networks. His exact, original request is included at the end of this charge.

I would like this committee to conduct a full review of the networking and procedures being used by the NOvA project and experiment. The focus of the this review should be to ensure that the designs and implementations of the networking for NOvA will allow the experiment to be successful in meeting both its construction project goals, as well as its long term scientific goals.

The review should address the long term support and operations of the NOvA networking infrastructure. This should include changes that the project and experiment will need to implement in moving from the current "development/debugging" environment to the "production" environment that will be required for detector operations.

In particular you should address the following topics:

1. Are the designs for each of the NOvA networks adequate to address the technical and operational requirements of NOvA experiment?
2. Do the network hardware and configurations that have been deployed to Ash River satisfy the designs of each of the NOvA networks?
3. Is the documentation that has been compiled by NOvA sufficient for support and maintenance of the network as the experiment moves into the build/operations phase of the far detector?
4. Are there appropriate procedures in place for affecting changes to the network configurations at Ash River?

5. Are there appropriate procedures in place for incident reporting and for tracking incident resolution?

The final result of your review will be a document answering the above questions and providing recommendations detailing changes that should be made to the NOvA networking and to the procedures that are used by the project and experiment to operate, maintain and change the networking systems and configurations. I will circulate this document to the NOvA project's management for their consideration.

In arranging this review there will be a substantial need for supporting documents on different subjects from members of NOvA. I can assist you contacting the appropriate people who will provide you with or direct you to the documentation.

Due to upcoming conflicts that some members of the committee will have with conference travel the week of Nov 12th, I would propose that the committee formally meet to conduct the review either the week of November 19th or the 26th, but no later than the 1st of December.

If you have questions or concerns please let me know.

Thank you very much for assisting NOvA with this review.

Andrew Norman
SCD NOvA Liaison

Original Request from Rick Tesarek:

```
Hi Andrew,
        As noted before the collaboration meeting, I would like you to call a review
of the NOvA far detector network.  The scope of this review should include
aspects of the following:
  o Is the network design adequate for our needs?
  o Does the existing network match reality at Ash River?
    (how do we know).
  o Is the network documented, if so, where?
  o What is the mechanism for making changes to the network?
  o Is the knowledge base redundant (more than one person capable of
    fixing/modifying the system) for NOvA operations.

In my mind, I would like to see a complete, line-by-line review of the connection
map (NOvA-6642)and that that map reflect how things are put together physically
down to which object is plugged into which port (with correct location names).
Do we have confirmation what is plugged in where?  Similarly with IP addresses
and MAC addresses. Finally do we have a procedure to make changes to the network
(approvals, implementation and documentation of the final result).


I'm sure the language needs to be worked out for an official change and there
may be additional items that you may wish to cover in such a review and we
should add these to the charge.

                                                  Rick
```