

Deploying perfSONAR-based End-2-End monitoring for production US CMS networking

Maxim Grigoriev, Andrey Bobyshev, Matt Crawford, Phil DeMar, Vyto Grigaliunas, Don Petravick

Fermilab, PO BOX 500, Batavia, IL 60510-0500, USA

maxim@fnal.gov

Abstract. Fermilab is the US Tier-1 Center for CMS data storage and analysis. End-2-End (E2E) circuits are utilized to support high impact data movement into and out of the Tier-1 Center. E2E circuits have been implemented to facilitate the movement of raw experiment data from the Tier-0 Center at CERN, as well as processed data to a number of the US Tier-2 sites. Troubleshooting and monitoring of those circuits presents a significant challenge, since the circuits typically cross multiple research & education networks, each with its own management domain and customized monitoring capabilities. The perfSONAR Monitoring Project was established to facilitate development and deployment of a common monitoring infrastructure across multiple network management domains. Fermilab has deployed perfSONAR across its E2E circuit infrastructure and enhanced the product with several tools that ease the monitoring and management of those circuits. This paper will present the current state of perfSONAR monitoring at Fermilab and detail our experiences using perfSONAR to manage our current E2E circuit infrastructure. We will describe how production network circuits are monitored by perfSONAR E2E Monitoring Points (MPs), and the benefits it has brought to production US CMS networking support.

1. Introduction

In this presentation, we will describe current state of Wide Area Network (WAN) Monitoring at Fermilab for the upcoming era of the LHC CMS [1] data intensive applications. In Section 2, we will present current and future requirements for Fermilab's WAN monitoring system and will explain our choice of perfSONAR [2] middleware as media layer for network monitoring data. Our experience with deployment of the End-to-End (E2E) network monitoring system services will be highlighted as well. The remainder of the paper is organized as follows: Section 3 briefly describes perfSONAR and perfSONAR-PS middleware APIs; in Section 4, we give a case study to show how these APIs were utilized and enhanced at Fermilab; and finally in Section 5, we unveil our plans on future WAN monitoring developments.

2. Motivations

With upcoming Large Hadron Collider (LHC) experiments at CERN, the task of monitoring WAN paths becomes significantly more complex than with previous collaborations. At Fermilab's Tier-1 Center, we expect to receive approximately one petabyte per month of raw CMS experiment data from

Tier-0 center at CERN, and send out several petabytes of reprocessed data from the Tier-1's data storage facilities for analysis at CMS Tier-2 sites. There are two major WAN infrastructures to monitor. One is LHC Optical Private Network (LHCOPN) [3], which connects Fermilab to the Tier-0 Center (CERN). The other is a heterogeneous set of network links between Fermilab and Tier-2 sites, with multi-gigabit production links and alternative high bandwidth network circuits designed to meet the data movement requirements of CMS.

2.1. LHCOPN

The LHCOPN consists of private optical paths between CERN and Tier-1 sites across Europe, Asia and North America. This is a mission critical network, designed for reliable and resilient 24x7 operations. It is required to have near real-time monitoring capabilities with centralized model of trouble ticket reporting to LHCOPN operational control center. The LHCOPN E2E network links are distributed over countries, continents and administrative domains. Currently, there are two 10Gbps circuits between the Tier-0 facility at CERN and the Tier-1 Center at Fermilab and there are plans to expand to four 10Gbps circuits in the near future (Figure 1).

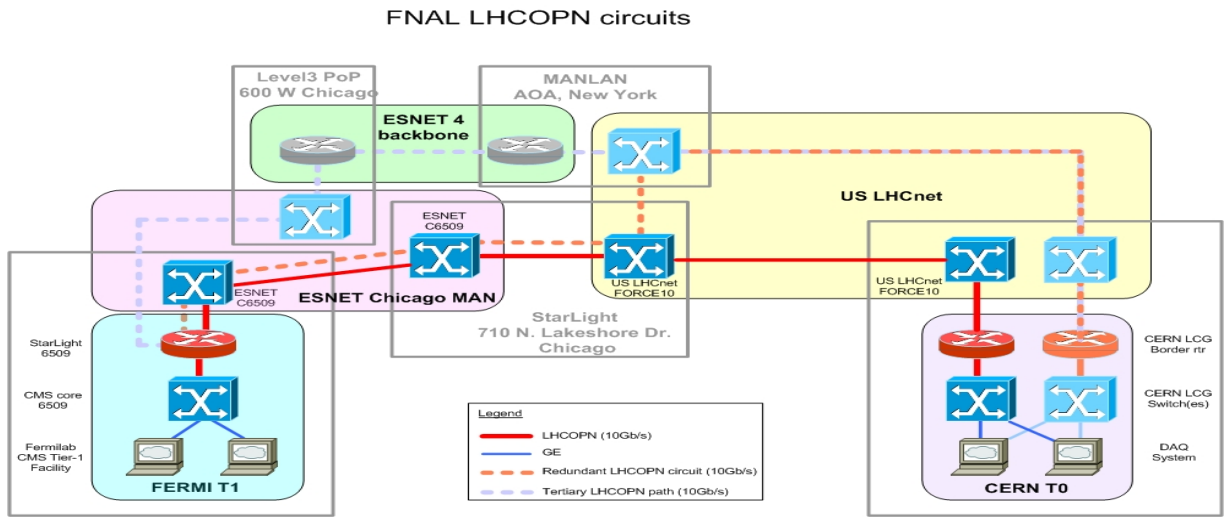


Figure 1: Fermilab's LHCOPN circuits

Supported levels of network monitoring capabilities and provider service level agreements differ widely across the research networks and institutions that make up the LHCOPN. The major task therefore is to create a unified monitoring data exchange layer and formalize the interactions between the various administrative entities. The WAN group at Fermilab has attempted to answer these requirements by adapting perfSONAR middleware to our network monitoring system and by undertaking the cohesive efforts with international LHCOPN network monitoring groups.

2.2. US CMS requirements

There are seven US-CMS Tier-2 [4] sites, as well as two Tier-2 sites in South America that are closely allied (Figure 2).

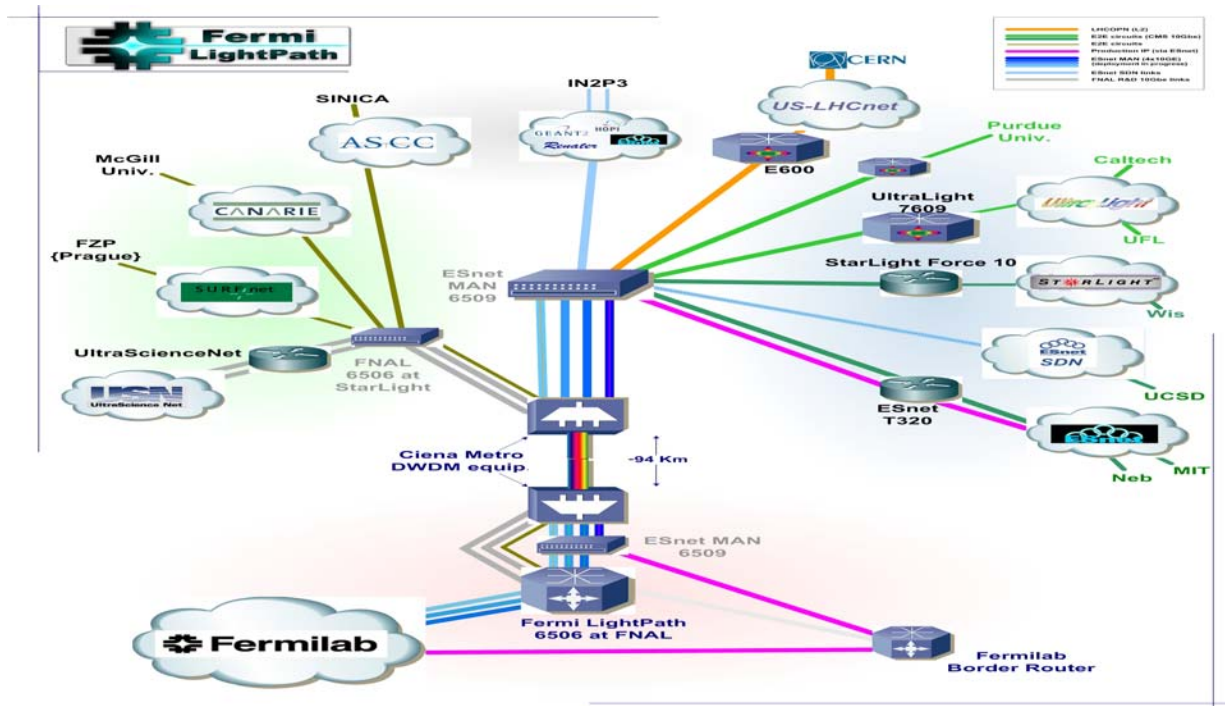


Figure 2 Fermilab WAN connectivity

The network monitoring framework for US CMS Tier-1 site must be integrated into the general network monitoring system at Fermilab. However, connectivity to Tier-2 sites is commonly met with dedicated E2E circuits in order to provide a preferable, high bandwidth path for CMS data traffic. E2E circuits are based on various network technologies, such as MPLS tunnels or Layer2 VLANs. Monitoring individual components of circuits that cross multiple administrative domains brings additional complexity for the whole system as well. Our group identified the list of network monitoring requirements for US CMS:

- near real-time availability of the results with alarm latency within 5 minutes
- clear indication and localization of failed circuit's fragments
- real-time status of complete circuits, with historical reports
- option to show physical components of circuits
- interface to on-site monitoring and alarm/notification systems
- expendable & easy configurable for new circuits
- up-to date contact information for network admins of the failed segments

These requirements are quite challenging and demand a unified metadata layer, provided by perfSONAR.

3. perfSONAR middleware

3.1. Overview

perfSONAR [6] is a network monitoring middleware developed by consortium of network developers from the Geant2 [7], Internet2 [8], ESnet [9] communities, with code contributed from

several research institutions in Europe and US. The project's website is the best source of information on API and development curve.

Conceptually, deployment of a perfSONAR framework would make network performance information available from any participating network domain to any authorized network monitoring system in the world. perfSONAR is a collection of web services, where each service might be used independently or in collaboration with others. The data collection service is named the Measurement Point (MP), and it may publish results of network measurement data to a storage service called the Measurement Archive (MA). All services are discoverable through an independent Lookup Service (LS). The development of the perfSONAR API is still an effort-in-progress, and some of the services currently available are either pre-production quality or still in the beta stage of development (like the Authentication and Authorization Service). The perfSONAR messaging flow is based on XML schemas developed by Global Grid Forum (GGF) Network Measurement Working Group NMWG [10], and is implemented as interoperable SOAP messages. This makes integration of services into the existed frameworks a straightforward task, and enables development of clients with any programming language that supports a SOAP API .

3.2. Utilized Services

At Fermilab, we have deployed several perfSONAR services. The first one was Round Robin Database (RRD) MA which provides us with link utilization on network physical interfaces involved in high impact CMS data movement. In order to satisfy our LHCOPN requirements, we have deployed an E2E circuit status MP and MA. We have also developed and deployed PingER MA. Eventually, we hope to see PingER MPs deployed at every remote US CMS Tier-2 site, allowing us to provide close to real-time E2E active monitoring for those circuits as well.

3.3. perfSONAR-PS API

perfSONAR-PS [11] is a collaborative effort by Internet2, SLAC, Fermilab, IU, University of Delaware, ESnet, and Georgia Tech to make perfSONAR middleware more “user friendly” and oriented towards network administrators. The project started in March, 2007 and has resulted in several core perfSONAR services implemented in Perl. Perl was chosen as the programming language most comfortable to work with by network administrators. Perl APIs make the task of integration with currently deployed site Network Monitoring Systems (NMS) more straightforward, since Perl is by definition the most comprehensive “glue” language. PingER MA and MP are being developed as part of perfSONAR-PS framework.

4. Monitoring at Fermilab

4.1. E2E monitoring system

The major release of perfSONAR API is still Java-based. It uses a Tomcat webserver as a servlets container and as a provider for an eXist XML database service, which is used for metadata storage. In order to publish monitoring data from our current Cricket [12] based SNMP monitoring system, we developed *cricket2perfonar* tool. It allowed us to convert our Cricket configuration files into the perfSONAR XML service descriptions.

4.2. Network circuits visualization

We have deployed and adapted the web-based E2E Monitoring system developed by DFN [13]. The normal view of the monitoring system for Fermilab's primary LHCOPN E2E circuit is shown in Figure 3. It is network accessible at [14]. Above the link view, there are records of the latest operational problem identified on the circuit. This feature, as well as a historical list of all previous incidents, were added to give WAN administrators at the Fermilab Tier-1 site more informative description of LHCOPN circuit problems.

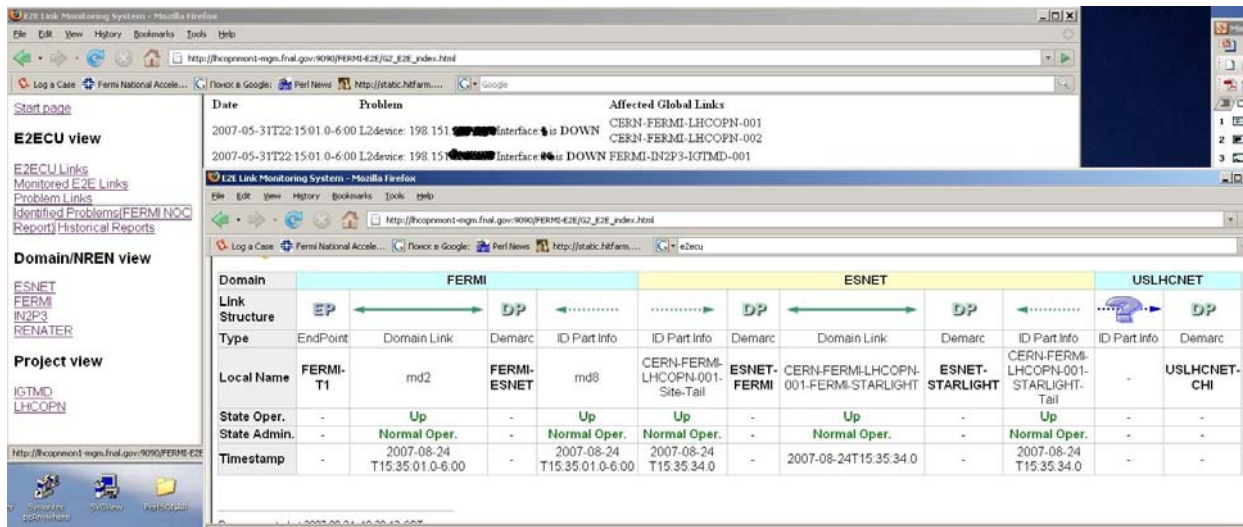


Figure 3 E2E Monitoring system

4.3. Monitoring system configuration management

At Fermilab, we developed a configuration management utility for the E2E circuit MP and MA, and created the FERMI E2E Configuration Manager web-interface (Figure 4). It utilizes Web 2.0 approach, and provides a fully interactive webpage for local administrators, who are now able to change any part of monitoring system without coding XML configuration files.

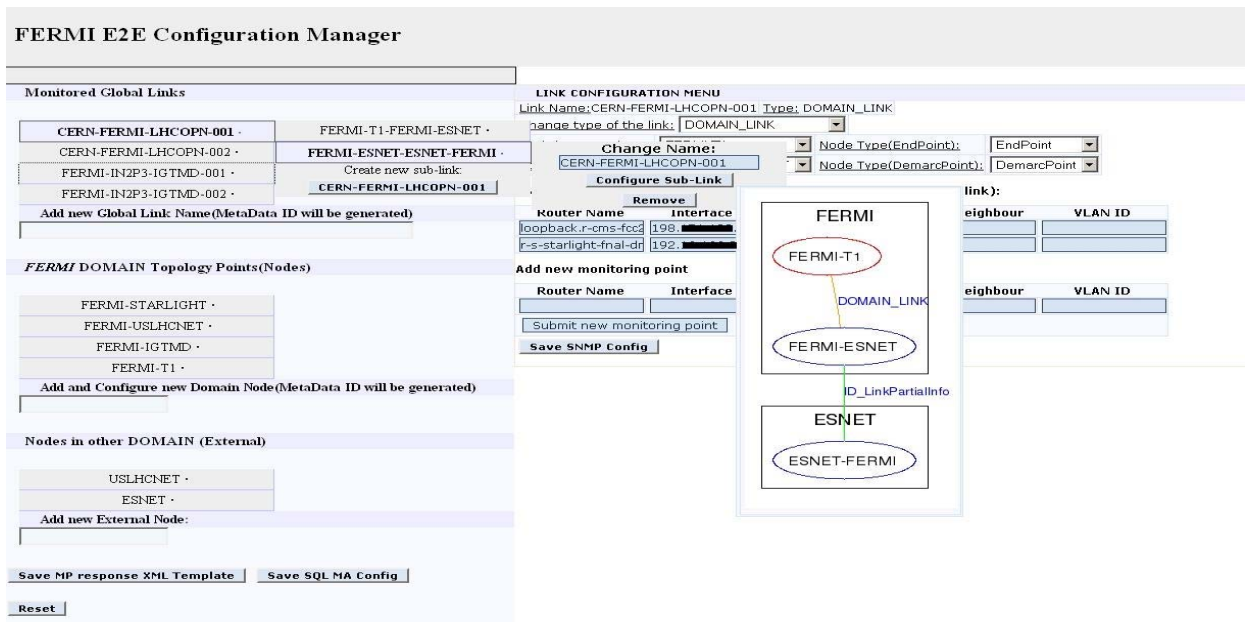


Figure 4 Configuration Management for E2E Monitoring system

4.4. Deployment issues

Several deployment issues were uncovered while trying to deploy perfSONAR services. Most notably, the heavy dependencies were found to exist on the version of Java, Tomcat and eXist DB by perfSONAR API. For network administrators the hardest part was dealing with multiple unfamiliar XML configuration files, a practically unknown Java environment and a cumbersome installation process.

perfSONAR-PS is still a work-in-progress, and it will take time to develop the configuration and installation process into something more manageable. Scalability issues in real life production environment still need to be addressed by its software developers.

5. Future plans

In the time remaining before the startup of LHC, the WAN group at Fermilab will be working closely with the perfSONAR and perfSONAR-PS developers to satisfy all requirements for network monitoring of the LHCOPN and the US CMS network infrastructure.

References

- [1] LHC CMS experiment at CERN, <http://cms.cern.ch>
- [2] A. Hanemann et al., "PerfSONAR: A Service Oriented Architecture for Multi-Domain Network Monitoring", In "Proceedings of the Third [International Conference on Service Oriented Computing](#)", [Springer Verlag, LNCS 3826](#), pp. 241–254, ACM Sigsoft and Sigweb, Amsterdam, The Netherlands, December, 2005
- [3] LHCOPN wiki, <https://twiki.cern.ch/twiki/bin/view/LHCOPN/WebHome>
- [4] US CMS, <http://www.uscms.org>
- [5] MonALISA project, Caltech: <http://monalisa.cacr.caltech.edu>
- [6] perfSONAR project, <http://www.perfsonar.net>
- [7] GEANT2 network, <http://www.geant2.net>

- [8] Internet2, <http://www.internet2.edu>
- [9] ESnet, <http://www.es.net>
- [10] NMWG, <http://www-didc.lbl.gov/NMWG/>
- [11] perfSONAR-PS project, <https://wiki.internet2.edu/confluence/display/PSPS/Home>
- [12] Cricket project, <http://cricket.sourceforge.net/>
- [13] E2Emon MP, http://wiki.perfsonar.net/jra1-wiki/index.php/E2EMon_MP
- [14] Fermilab E2E Monitoring page, http://lhcopnmon1-mgm.fnal.gov:9090/FERMI-E2E/G2_E2E_index.html