Transport and Management of High Volumes of Data through Bounded LAN and WAN Infrastructure at SLAC

S. Luitz[1]*, D. Millsom*[2]*, D. Salomoni*[3]*, J.Y. Kim*[4]*, A. Zele*[5]

Abstract

This talk will address how the Stanford Linear Accelerator Center (SLAC), with its diverse and unique requirements for networking at very high speeds, is managing challenges to provide physicists with state-of-the-art network, in the context of a secure environment.

keywords Networking, LAN

1. Introduction

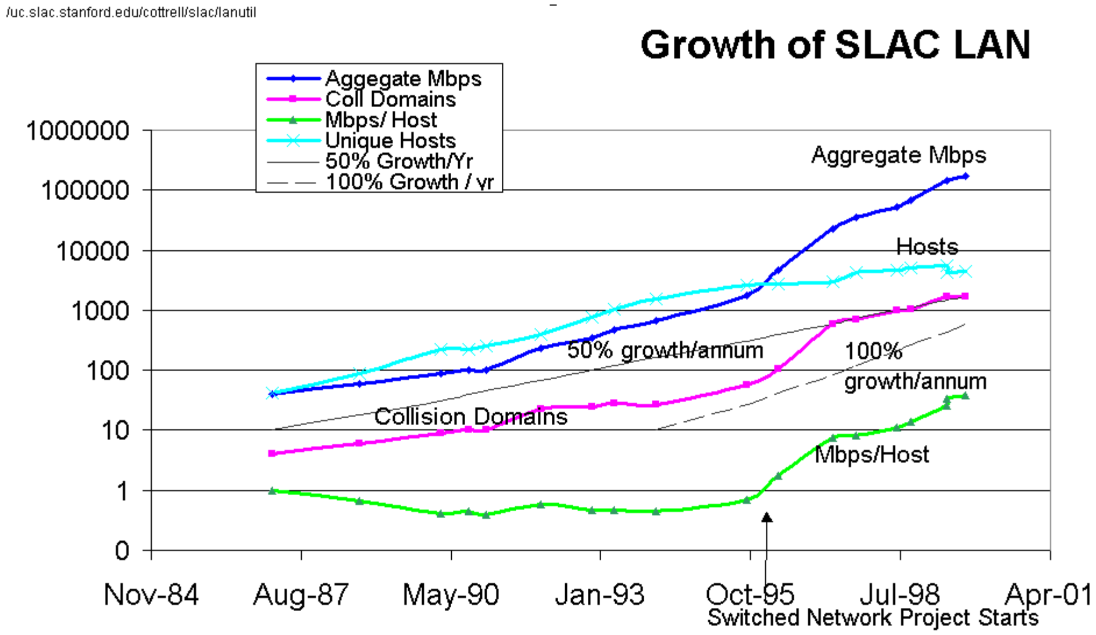
The computing environment at SLAC includes real-time, computational, business systems, high volume physics data and general purpose computing. The laboratory has a full-time staff of 1300 augmented by 700 on-site collaborators. SLAC collaborates with about 200 institutions worldwide. These factors demand a high-performance, highly available network, which provides competent security while remaining flexible and accessible for international collaboration. In addition, network technology is advancing rapidly and must be folded in to the existing framework on an ongoing basis with minimal disruption. The SLAC LAN has been designed to address these (often conflicting) requirements. This paper will describe briefly the architecture chosen to do this and discuss some of the issues encountered.

2. Major Issues

2.1. Bandwidth Growth

Increasing bandwidth demand is the leading factor influencing network design. As traffic increases, the designer is confronted with a number of strategies for providing traffic management. These are: (1) divide and conquer. I.e. segment traffic at layer two and layer three levels, (2) use locality to partition local traffic from non-local, (3) establish a bandwidth hierarchy, providing higher bandwidth to busiest devices, e.g. servers, (4) deploy higher bandwidth technology, e.g. 10 -> 100, 100->1000 Mbps Ethernet including Fast EtherChannel and Gigabit EtherChannel.

[1] Stanford Linear Accelerator Center, luitz@slac.stanford.edu
[2] Stanford Linear Accelerator Center, millsom@slac.stanford.edu
[3] Stanford Linear Accelerator Center, salomoni@slac.stanford.edu
[4] Stanford Linear Accelerator Center, jyk@slac.stanford.edu
[5] Stanford Linear Accelerator Center, tonyz@slac.stanford.edu



The design attempts to use all four strategies. The network is partitioned into multiple subnets using VLANS. Computers are allocated to VLANs based on where their traffic pattern is expected to be localized. Thus, although VLANS can span large physical distances, traffic is expected to be localized to the VLAN and is routed only when destined for other subnets. Since VLANS are implemented using Ethernet switches, traffic is further constrained to those ports and trunks required to transport it within the VLAN.

Secondly, a bandwidth hierarchy places bandwidth only where needed. In the star topology, this tends to be towards the center of the star. Devices with higher bandwidth requirements and also those needed to be accessed by multiple subnets are located towards the center.

This kind of traffic management makes it possible to budget and allocate resources appropriately.

2.2. Reliability/Availability

From our experience we have found that maximizing the reliability and availability of the network can be achieved by addressing the following issues: redundancy, single vs multiple vendors, minimizing equipment types and clear delineation between real-time and general purpose networks.

2.2.1. Redundancy.

We have provided redundant pathways between major network components including the interior routers and between the central and peripheral switches. In addition, all critical switches and routers are fitted with dual power supplies and redundant supervisor modules. All of the central (computer center) network equipment is supplied by a UPS giving it about a four hour window if the power fails.

2.2.2. Single Vendor

Our experience with multiple vendor equipment has shown that the problems with interoperability, maintenance, software/hardware maintenance and management are expensive and time-consuming. Although multi-vendor sites might be practical when generous budgets allow sufficient personnel and other resources, we have taken the more expedient path by selecting a single vendor for major network components. We have done this in light of our own attempts to manage multiple vendors and reports that each additional vendor costs about an additional FTE[6]. In our case, we chose Cisco Systems, Inc.

2.2.3. Minimizing Equipment Types

We have endeavored to select equipment from as small a set of equipment types as possible and to use equipment, which has interchangeable components. This has simplified parts management and facilitated a relatively inexpensive spares pool. This also minimizes the amount of training needed for technical support.

2.2.4. Clear delineation between real-time/on-line networks and general-purpose networks.

Although this is a fairly obvious point because real-time and general purpose networks have different requirements for availability, it is worthwhile noting that in an integrated network such as SLAC's, it is very easy for interdependencies to develop between real-time components and non real-time components. For example, a system providing a real-time function might inadvertently depend on a file system on another part of the network thus making it critical to the real-time function. This has been addressed on two levels. On the system level, we strive to promote and maintain independence of the real-time systems from other systems on-site. At the network level, we have constructed bridge group topologies and deployed subnets in such as way as to prevent network events on the general purpose network from affecting the real-time network and vice versa.

2.3. Matching technology against performance requirements

It is typical for high-energy physics to demand the highest bandwidths, which can be obtained at any given time. As a result the latest technology is deployed as soon as it becomes available and often before it has been well tested in the field. This creates a particular challenge to maintain network stability while installing the latest (and often, not stable) hardware and software.

[6] Full Time Equivalent person

In our experience there is no obvious solution to this except to deploy new technology in well-chosen locations where it will have minimal impact on overall network availability and performance and to test equipment in isolation where possible.

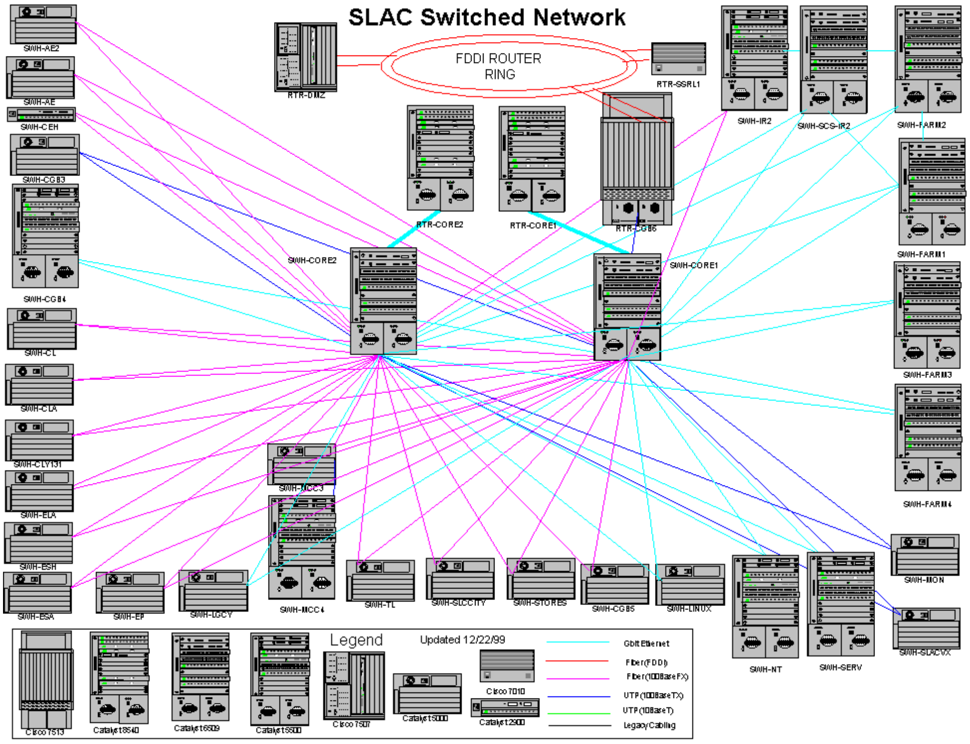
2.4. Open Network for Collaboration vs. Security Requirements

SLAC is an inherently open environment, which encourages collaboration within the lab as well as with other institutes. The lab collaborates with over 200 institutions worldwide. This creates a constant inflow of software and hardware, a requirement for extensive access from offsite and extensive access to external sites from SLAC. As a result, the computing environment is impossible to control and manage in detail and that conventional firewall partitioning of the network from the external network is not possible.

At the same time, there are essential security requirements, which must be met. These include (1) protection of real-time control systems and equipment under computer control, (2) business services, particularly including sensitive personal and other financial information, (3) protection of the network infrastructure including routers, switches, various servers, e.g. name and mail servers, which if damaged or compromised could disrupt service, (4) DOE security standards, (5) protection of individual workstations from sources of obvious attacks such as denial of service, spoofing, etc.

To achieve this, a number of steps have been taken. These include, continual monitoring of the network for intrusion, regular auditing of individual devices for known weaknesses (requiring patches, for example), the disabling of external routing (i.e. the establishment of an Internet Free Zone, (IFZ)) for all infrastructure addresses (e.g. router interface addresses, sensitive servers) which need not be externally visible, the establishment of internal firewalls and islands for very sensitive subnets and the extensive use of ACLs (Access Control Lists) to protect against known methods of attack and to disable all but a well defined set of protocols.

3. Network Architecture



From the preceding diagram it can be seen that the basic network structure is a physical star radiating from the computer center. In addition, a number of sub-stars are deployed for specialized functions. These include the compute farm, the main control center, IR2[7] and business services. At the center of each star are one or more level-two switches providing VLAN[8] switching and trunking to the end-points. Also at the center, routing is provided to move traffic between VLANs, and in the case of peripheral stars, between the star and the core routers.

The physical network consists of single-mode and multi-mode fiber running Ethernet at 100 and 1000 Mbps. The inter-switch traffic uses Cisco ISL protocol to provide VLAN trunking. ISL was chosen because it was the only trunking protocol available when the network was being developed. Consideration has been given to converting to 803.1Q but due the various problems with that standard it has not been deployed.

VLAN trunking is also deployed between the core switches and routers. This greatly reduces the number of physical interfaces required on each router because each new subnet does not require a new router interface.

The switch to router connection also uses Giga EtherChannel providing a bandwidth of up to four gigabits/sec per channel. Additional bandwidth can be obtained by distributing the subnet/VLAN pool over multiple GigaEtherChannels.

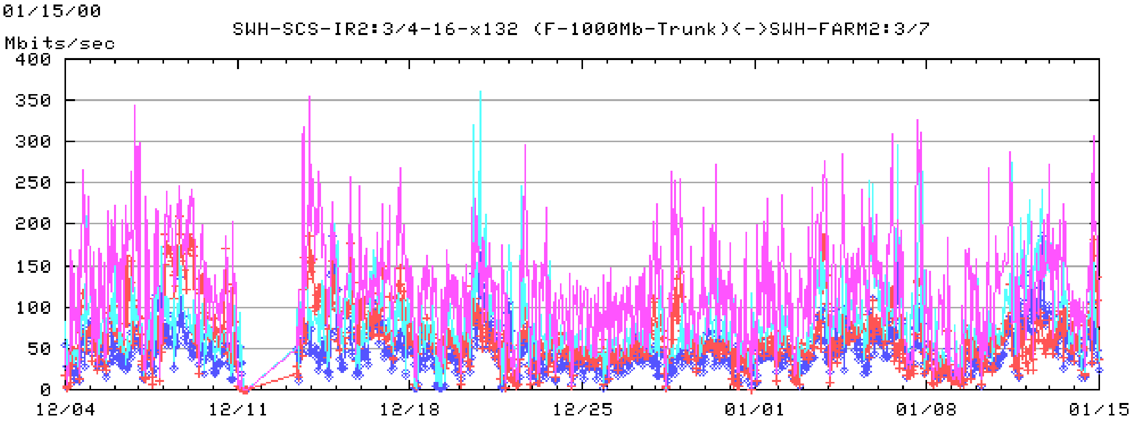
4. Performance

The network is well instrumented allowing throughput and utilization to be monitored down to the switch port level. A set of graphs is produced daily summarizing the performance over the last 24 hours and summarizing the performance of the proceeding week and preceding six weeks and six months. From these graphs we are able to ascertain that there are no components which are approaching saturation. This includes port utilization, backplane utilization and CPU utilization. All known limitations to network performance are related to end-node limitations such as bus bandwidths and NIC card implementations. For example, we know of no gigabit interfaces, which are capable of attaining throughputs of 750 Mbps.

The following graph shows utilization of one of the busiest inter-switch trunks on the network.

[7] Interaction Region 2 (BaBar Detector)

[8] Virtual LAN



5. Discussion

We believe that we have designed a network, which meets the requirements of SLAC's collaborative environment by providing a high-bandwidth, robust, secure and highly available infrastructure. This has been achieved by deploying virtual LANs and centralized routing. The current design was commenced four years ago and continues to be expanded using the same design goals and principles. Although new technology as arrived (Gigabit Ethernet, Fast EtherChannel and GigaEtherCHannel), these have been successfully incorporated into the network and successfully matched the increasing bandwidth demand.

On reflection, we have found that the main problems have been (1) management and configuration of layer-2 switching and spanning trees, (2) attempting to mix standards, particularly ISL and 803.1Q even in a non-overlapping configuration, (3) monitoring traffic, given inadequate monitoring technology and the difficulties of effective data gathering in a distributed switching environment, (4) diagnosis of equipment failures in the switching environment. However, the reliability of the network we have constructed is such that once having established the network, the above mentioned problems are not in-surmountable.