Use of Flow Data *for Traffic Analysis and Network performance* characterization

Andrey Bobyshev, Phil Demar, Vyto Grigaliunas, Maxim Grigoriev
Fermilab, PO Box 500, Batavia, IL 60510-0500, USA

E-mail: bobyshev@fnal.gov

Abstract. At Fermilab, we have a long history to use flow data gathered from site routers for various analysis, including network performance characterization, investigating computer security incidents and their prevention, network traffic statistics, and others. Currently, the flow analysis at Fermilab is built as a distributed system to collect data from multiple routers, both at the edge of the site network as well as from local routers and multilayer switches. Daily volume is about 5GBytes of raw data. Despite a high volume of collected information, some analysis is conducted in near real-time to satisfy demands of users communities for quick results. In this paper, we present Fermilab's Netflow Collection and Analysis system, as well as various analysis and tools based on netflow data. As an example of such analysis, we describe traffic characterization and network performance estimation for USCMS Tier1 centre, the tools for checking of traffic consistency for End-To-End circuits and Policy Based Routing and finally, profiling of host's traffic to keep track of their typical behaviour to prevent accidental blocking by site IDS system.

1.Introduction

At Fermilab, analysis of flow data gathered from multiple routers is the source of valuable information for various tasks. We distinguish three major areas where that information is used. These areas are computer security, performance analysis for data movement applications and verification of traffic consistence across site network infrastructure. We have considerable experience to use flow data in two first areas. The last one, verification of traffic consistency is relatively new to us. It is motivated by recent deployment End-To-End circuits for LHC/CMS experiment. These circuits have predictable performance characteristics and dynamic capabilities to use alternative WAN path available for the USCMS Tier1 centre. In this paper, we present analysis tools deployed at Fermilab in all three areas and give their brief overview.

2. Netflow collection system

Over last few years, Fermilab's Flow Collection System has undergone through significant changes due to increasing a number of network devices exporting data, as well as overall volume of information and duration of time to support online access to the historical data. The currently deployed system is depicted in Figure 1.

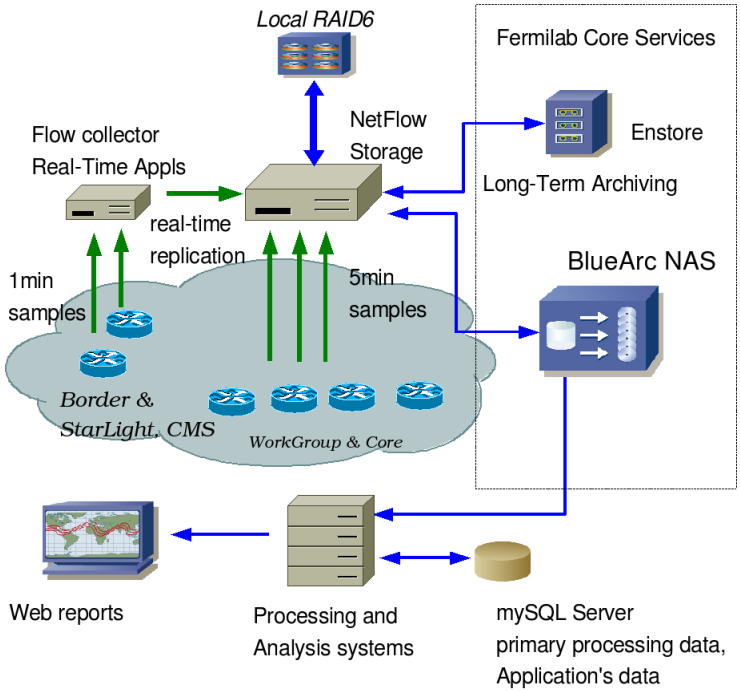

Figure 1: Distributed flow collection and analysis system

For collecting and replicating flows, as well as for preliminary analysis, we use the flow-tools software package[1]. The primary collector is a Dual Xeon 2.4GHz server running Fermi Scientific Linux with SCSI RAID6 and has about 6TB of storage space. For most exporters, information is stored in data sets gathered in five minutes intervals. However, there are a few applications running in near real-time that use flow data gathered in one minute intervals. Usually, such applications run on dedicated machines. That is why, for a few routers, data is exported to these systems directly, and then replicated to primary storage to keep a copy of all data there. In the last year, we started to utilize a BlueArc Network Attached Storage service, which is centrally supported. This service is proven to be a reliable and scalable solution that allows to increase storage space incrementally as needed.

For data processing, we use two Dual Xeon Quad Core 2GHz and three Dual Xeon Dual Core 2.4 systems. All these servers have NFSv3 access to centralized BlueArc storage. Table 1 summaries average daily and monthly volumes of flow data gathered at different layers of the network and actual

traffic passed routers. These measurements are taken by approximating actual volume of allocated storage space for a few months. That is why monthly numbers do not necessarily match multiplied daily readings. Other devices not shown in the table produce approximately 3GB of flow data daily.

Table 1: Volumes of flow data and actual traffic

	Netflow Data		Actual Traffic	
	Daily	Monthly	Daily	Monthly
Border	600MB	17GB	15TB(43TB)	300TB (600TB)
StarLight	200MB	5GB	80TB(120TB)	1.5PB(2.4PB)
CMS	1.2GB	30GB	80TB(120TB)	1.5PB(2.4PB)
CORE	3GB	40GB	N/A	N/A

3. Preliminary tagging of flow data and breakdown of traffic

Many of our applications utilize traffic between different entities, calculated in terms octets, flows and packets. An entity is identified by the name and can represent a single host, IP subnets, a list of IP blocks or the whole site. The raw flow data selected for analysis is run through an initial tagging process to assign names based on the static definitions or generated dynamically, and calculate corresponding traffic. The results of tagging are stored in an mySQL database in tables suitable for further processing or representation. A simplified tree of this multi-layers process is shown in figure 2.

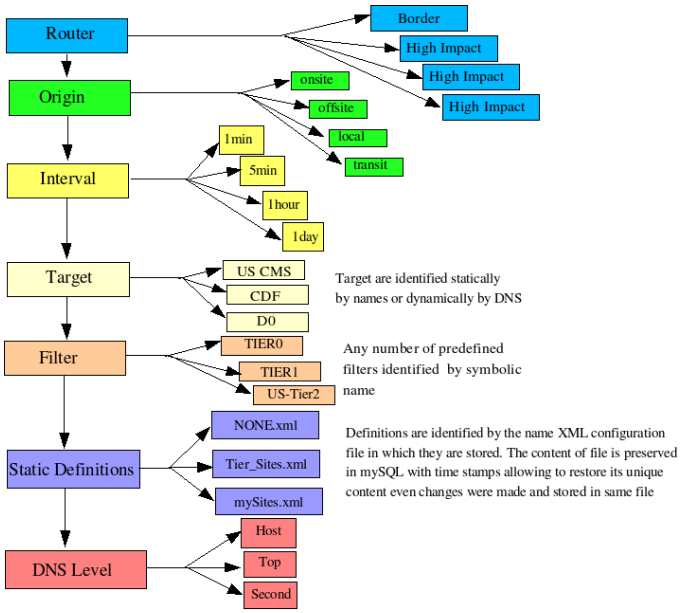

Figure 2: Traffic breakdown tree

A start layer for the tagging process is network layer. At Fermilab, we distinguish several major parts of the network: border, high-impact infrastructure, site core and work groups. The network layer is identified by router, or group of routers, exporting flow data. The next layer is the origin of traffic. Currently we define the following:

- *onsite* - traffic initiated from IP blocks assigned for Fermilab network to any other IP block
- *offsite* - opposite to onsite, traffic initiated from non-Fermilab IP addresses to Fermilab's one.
- *local* - traffic between Fermilab's IP addresses
- *transit* - traffic between non-Fermlab blocks. This is typically RFC1918 private, non-routed IP blocks leaking from work group networks. But also could be some IP reflectors running via VPN connections.

Traffic is analysed for different time intervals, typically 1min, 5min, 15min, 1hour, or 1 day. A start date and time can be also specified to generate historical results. Tagging of raw data can be done at any time, but for most applications, it runs when new flow sets are created and become available on BlueArc NAS server. New intervals can be added as needed. The next level is the target, which represents defined list of IP blocks assigned to particular users group or community. The USCMS Tier 1 LAN is an example of a user group address block. A target's traffic is going to be inspected based on set of predefined filters. For example, analysing appropriate traffic to specific destinations. For each target there may be several filters defined. Filters have unique names and are stored in a database with time stamp allowing to track changes. While inspecting traffic, we may use some specific names for destinations or sources. For example, we don't need to known all details of traffic to each node of the destination cluster. Instead, we need to know traffic to whole cluster, and use a predefined name for the address block. The next level is the static definition of rules specified in separated XML files. This level is identified by name of XML configuration file. The content of configuration files may be changed over time, due different reasons, i.e. readdressing of nodes, moving them to different subnets and so on. To track these changes, the content of configuration files is stored in database with time stamp. Anything that is not statically defined will be resolved based on DNS level, i.e. top level, second, third and so on, as needed.

4. Tools for Flow analysis

4.1. Computer security applications

When investigating the computer security incidents, such as aggressive scanning, Denial of Service attacks, spreading of worms and computer viruses, raw flow data provides information about source, destination IP addresses, protocols, ports, as well as traffic information in byte, flow and packet count information. In addition to CLI tools supplied with the flow-tools package, we have developed a Web interface that simplifies the process of selecting traffic for analysis based on multiple matching criterion as demonstrated in figure 3.

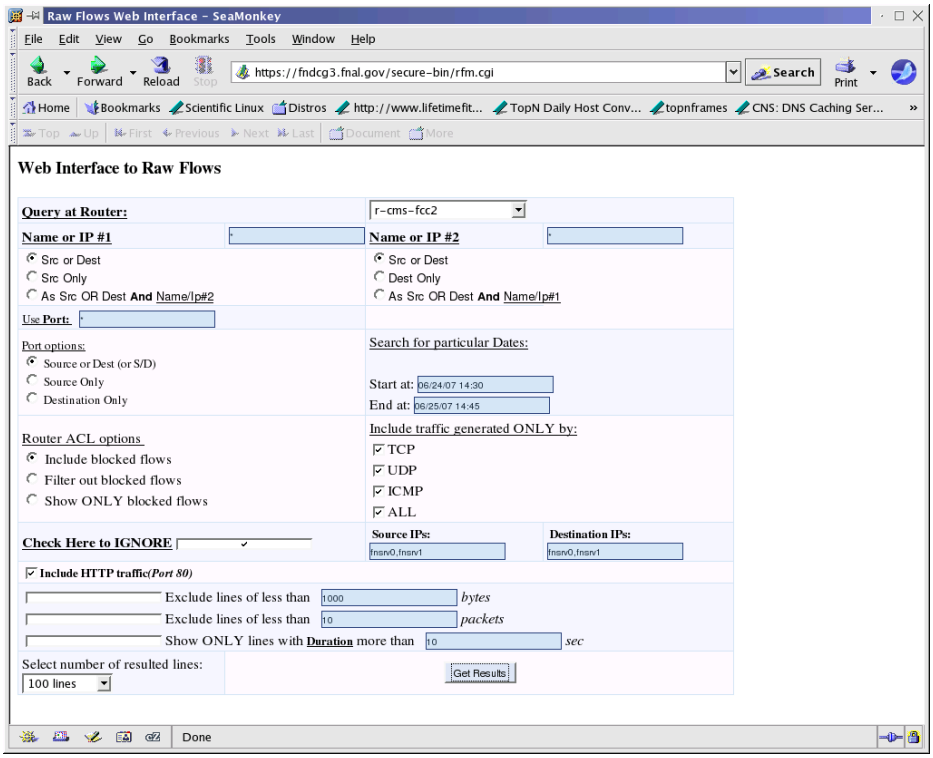

Figure 3: Raw Flow GUI

Usually, aggressive network scanning and probing of single hosts precedes spreading of worms or viruses. The information provided in the topN reports that we generate helps to recognize and prevent such attempts. The Top scanners reports are distributed vie e-mail on the regular basis, as well as being available on the web. The figure 4 illustrates an example of output from that tool. Daily, hourly and shorter intervals are available. Also, it is possible to see historical reports by specifying a start date and time.

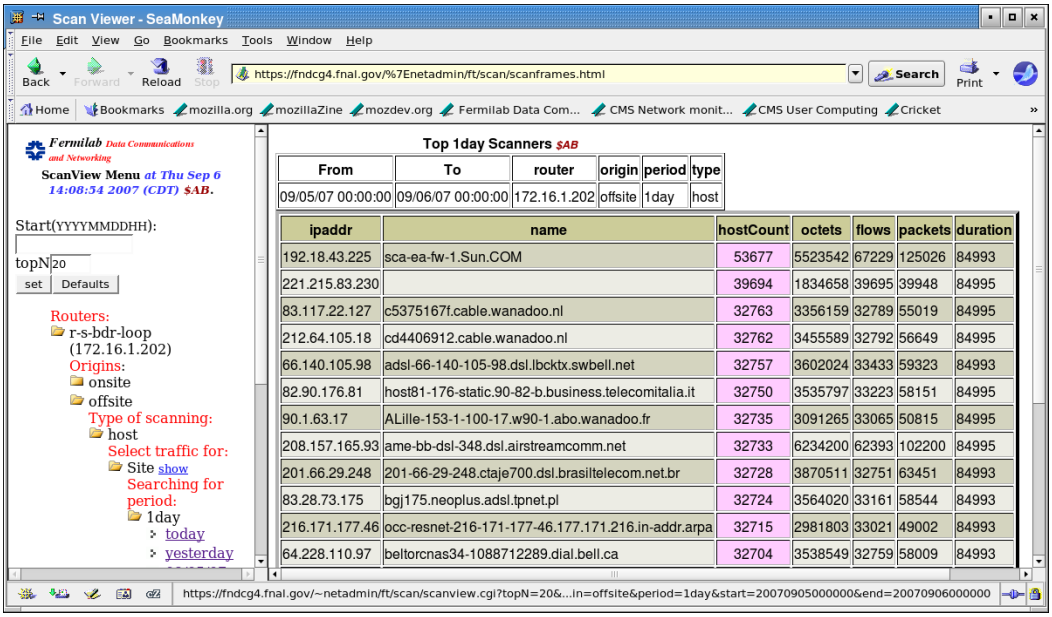

Figure 4: Example of TopScan output

In addition to the reports of top scanning activity, we deployed an automatic tool that searches scanners, both offsite and onsite, and block/unblock them automatically. This tool, called Autoblocker, has been running at Fermilab for about five years. It has been presented at CHEP2004[3]. In this paper, we would like to introduce a few new features that have been added. In first, a new detector to search a slow scanning has been developed. It runs as a standalone process and sends alerts to AutoBlocker via SOAP protocol. Second, AutoBloker has now the capability to recognize applications based on their typical behaviour in terms of its metrics. This feature has significantly reduced the rate of false positive blocks for the Grid applications. The current exception system utilizing this features is depicted in figure 5.

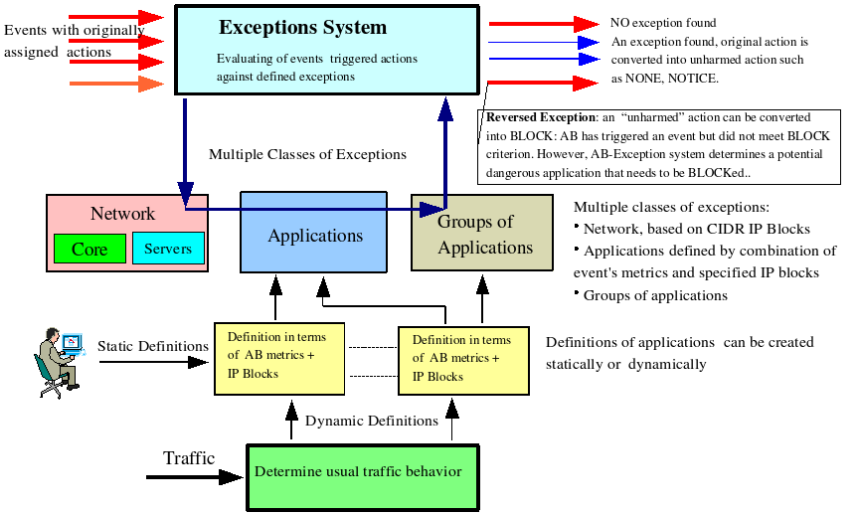

Figure 5: AutoBlocker Exceptions System

4.2. Performance analysis

Knowing transfer rates for data movement applications is very important. Often, neither SNMP monitoring of router/switch interfaces, nor applications themselves can provide that information. For this purpose, we use passive monitoring based on netflow information. Several tools have been deployed at Fermilab to address this need.

First, on the regular basis, we generate reports with topN senders, receivers and conversations for specified interval of time between the entities that we need to monitor. The reports are produced using the results of traffic tagging described earlier. The output of the WEB based interface is very similar to Top Scanning reports.

A new tool that was deployed recently is the US CMS Network weather map. This is an interactive tool showing directions and rates from/to US CMS Tier1 based on static and dynamic definitions. The examples a few of its pages are presented in the figures 6,7,8. The top page shows the most recent rates to/from US CMS Tier2 sites and two Tier2 sites at South America. There is a link associated with each Tier2 icon. Links between icons have a pop up historical graphs and topN tables. The CMS sites that need to be monitored are defined by a list of IP blocks in an XML configuration file.

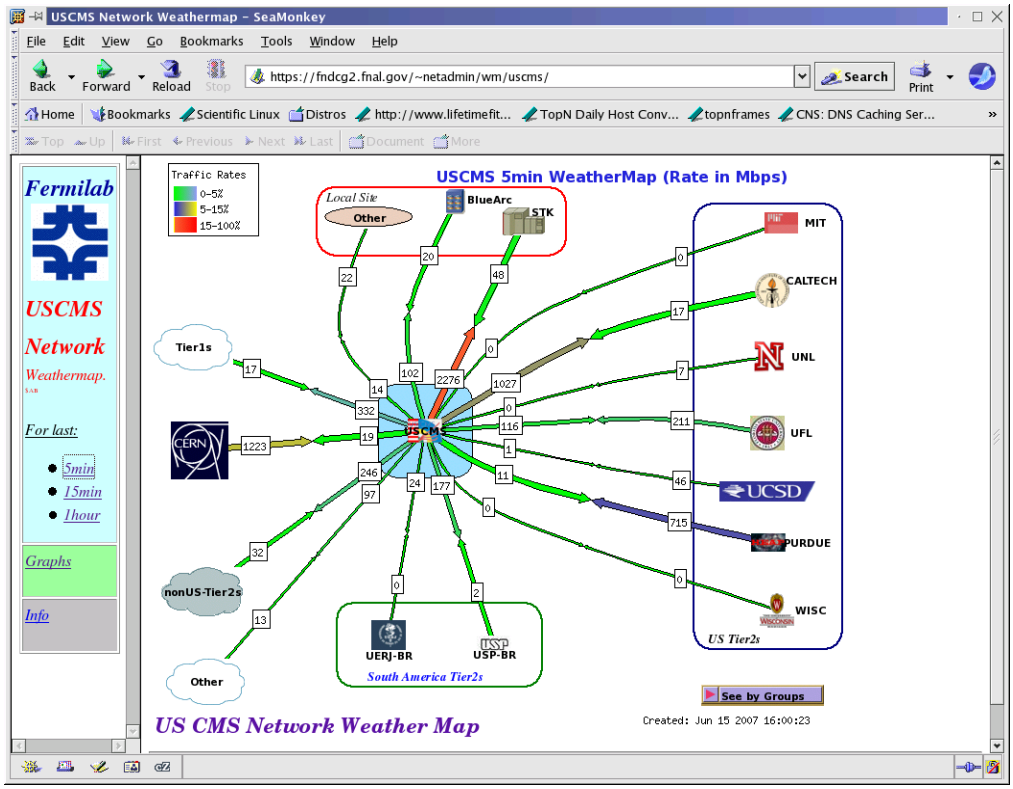

Figure 6: US CMS Tier2 and South America sites

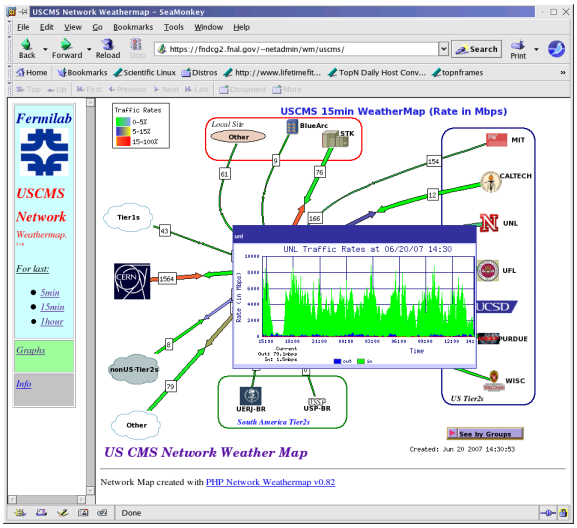

Figure 7: USCMS Weathermap, pop-up graphs

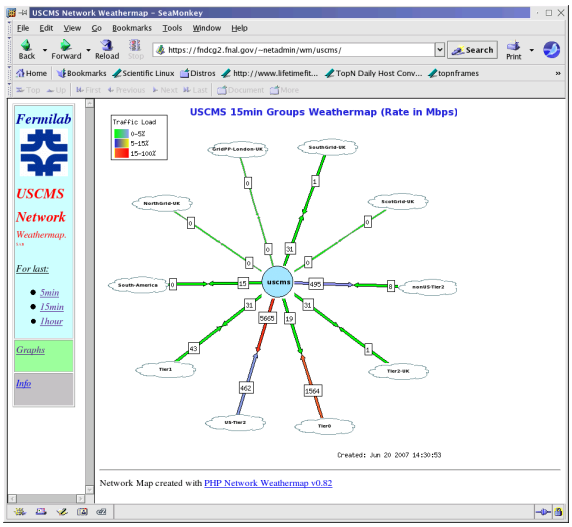

Figure 8: Throughput by groups

4.3 Traffic consistency

In the two last years, Fermilab has created a number End-to-End(E2E) circuits to provide a preferable network path between US CMS Tier1 centre and Tier0, various Tier1 and Tier2 centres [5]. When steering traffic into such circuits, statically or dynamically[4], asymmetry of data flow can occur. Although such behaviour is not fatal and is even sometimes unavoidable, it could affect throughput of data movement applications. Our objective is to detect such conditions automatically and notify if it becomes a long term situation. After preliminary investigation, we determined that flow data collected from routers at the edge of the campus network could be used for that purpose by comparing flow rates at various points of network. Flow rates for a symmetric path should be very close in both directions, inbound and outbound. We developed a new tool, called ***bfp*** that can be used interactively to show asymmetric traffic conditions. Currently, we are working on automated alerts. The figure 9 demonstrates this idea in tests with Lambda Station[4], which was steering production traffic between USCMS Tier1 at Fermilab and Tier2 centre at Caltech. When Lambda Station is operating, the traffic switched into a high-impact path is almost always symmetric in all routers. However, when Lambda Station is not operating, then traffic is going to be asymmetric at the Fermilab border and StarLight's the point of presence. From graphs at the right we can see that traffic at work group router is almost symmetric. However, at the border router we are going to observe only outbound flows (second graph) and at the high-impact infrastructure providing alternative path, only inbound flows are observed. Thus, by comparing flow rates at the different routers along potential traffic path we successfully detect

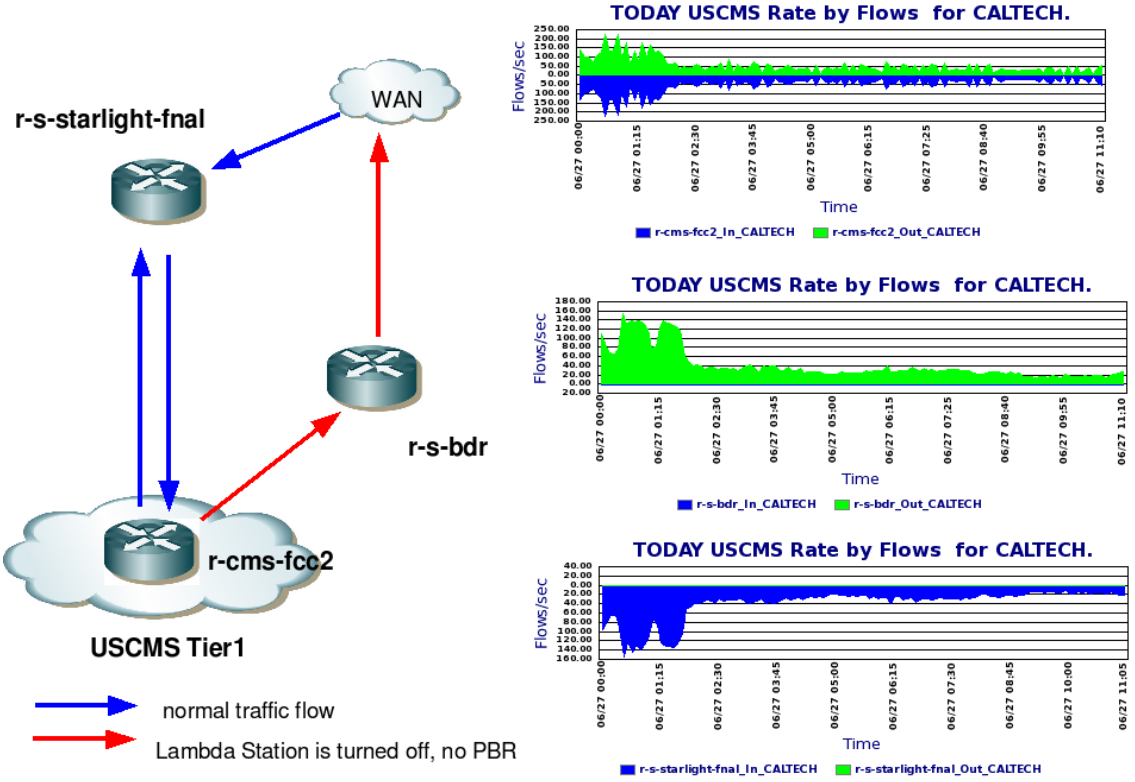

Figure 9: Detection of traffic asymmetry

asymmetry conditions.

5. Summary

Passive analysis based on flow data complements generic monitoring infrastructure at Fermilab. It provides the valuable results in several areas such as computer security, performance analysis and traffic verification. At Fermilab, we are using a few commercial products and public domain software when applicable. However, taking into consideration our specific requirements we still need to develop our own tools. Cooperation with other sites participating in the LHC experiments, or having similar requirements could be productive and beneficial for all participants.

6. References

[1] The OHIO State University Flow-tools, http://www.splintered.net/sw/flow-tools

[2] A.Bobyshev, M.Grigoriev, Methodologies and techniques for analysis of network flow data, Proceedings of CHEP04, CERN, Interlaken, Switzerland, 27th September - 1st October February 2004

[3] A.Bobyshev,D.Lamore,P.Demar, A real-time system for detecting and blocking of network scanning based on analysis of netflow data, Proceedings of CHEP04, CERN, Interlaken, Switzerland, 27th September - 1st October February 2004

[4] A.Bobyshev et al., Lambda Station: Production Applications Exploiting Advanced Networks in Data Intensive High Energy Physics, Proceedings of CHEP06, TIFR, Mumbai, India, 13-17 February 2006.

[5] P.Demar et al., Use of Alternative Path WAN Circuits at Fermilab, to be presented at CHEP07, Victoria BC, Canada, 2-7 September 2007.