

Dynamically Forecasting Network Performance of Bulk Data Transfer Applications using Passive Network Measurements

Tuesday, February 14, 2006 5:20 PM (20 minutes)

High Energy and Nuclear Physics (HENP) experiments generate unprecedented volumes of data which need to be transferred, analyzed and stored. This in turn requires the ability to sustain, over long periods, the transfer of large amounts of data between collaborating sites, with relatively high throughput. Groups such as the Particle Physics Data Grid (PPDG) and Globus are developing and deploying tools to meet these needs. An additional challenge is to predict the network performance (TCP/IP end-to-end throughput and latency) of the bulk data transfer applications (bbftp, ftp, scp, GridFTP etc) without injecting additional test traffic on to the network. These types of forecasts are needed for: making scheduling decisions, data replication, replica selection and to provide quality of service guarantee in the Grid environment. In this paper, we demonstrate with the help of comparisons that active and passive (NetFlow) measurements are highly correlated. Furthermore, we also propose a technique for application performance prediction using passive network monitoring data without requiring invasive network probes. Our analysis is based on passive monitoring data measured at the site border of a major HENP data source (SLAC). We performed active measurements using iperf and passive (NetFlow) measurements on the same data flows for comparison. We also take into account aggregated throughput for applications using multiple parallel streams. Our results show that active and passive throughput calculations are well-correlated. Our proposed approach to predict the performance of bulk-data transfer applications offers accurate and timely results, while eliminating additional invasive network measurements.

Summary

In this paper we will explain in detail two common approaches for network monitoring i.e. passive and active monitoring and we will also discuss how to get best of both the worlds. We will then describe different techniques which we used to calculate throughput from passive data, flow sorting and multiple parallel flow aggregation while performing comparison between active and passive network measurements. After this we will compare the results of passive and active measurements and study in detail the cases and reasons for poor agreement. In the next section we describe and discuss the challenges of our proposed approach to predict the network performance of the bulk data transfer application using passive monitoring data. A comparison of our results with other active forecasting techniques applied to the data from site border of a major HENP data source (SLAC) will also be discussed. Furthermore, we will investigate the reasons for multiple modes (like: diurnal effect, network performance changes etc) in our passively calculated throughput data. Our paper ends with an evaluation of the results, and a description of our future work.

Primary authors: Ms LOGG, Connie (Stanford Linear Accelerator Center (SLAC)); Mr NAZIR, Fawad (Stanford Linear Accelerator Center (SLAC)); Mr MEI, I-Heng (Stanford Linear Accelerator Center (SLAC)); Dr COTTRELL, Les (Stanford Linear Accelerator Center (SLAC)); Mr CHHAPARIA, Mahesh (Stanford Linear Accelerator Center (SLAC))

Presenter: Dr COTTRELL, Les (Stanford Linear Accelerator Center (SLAC))

Session Classification: Computing Facilities and Networking

Track Classification: Computing Facilities and Networking