Progetto SCoPE

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

CHEP 2009

21 - 27 March 2009, Prague, Czech Republic

A. Ciuffoletti [c], L. Merola [a], F. Palmieri [a], S. Pardi [b], G. Russo [a]

[a] University of Napoli Federico II – Napoli, Italy

[b]INFN-Napoli Section - Italy

[c] University of Pisa – Largo B. Pontecorvo – Pisa, Italy

THE SCoPE PROJECT

The S.Co.P.E. Project [1] aims to the implementation of an open and general purpose Grid infrastructure linking together the departments of University Federico II distributed in Naples on a metropolitan scale.

Scientific and technological research, implementation of innovative concepts in strategic research fields.

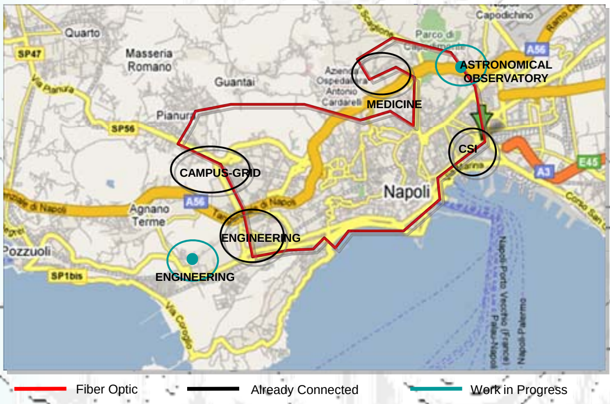
NETWORK- AWARE GRID

Network performances can affect dramatically job computation time, especially when processing remote bulk dataset and during data replication activities.

The Grid infrastructure of main Grid deployments works below the best effort threshold: the network is considered *pure facility* and middleware components act without taking into account network parameters.

An active network-aware approach needs a strict integration between many partners:
- the Grid resource management logic
- the requesting application
- the network entities that offer the connectivity services

The goal is to take job scheduling and resource allocation decisions based on network performance.



THE NETWORK-AWARE SERVICES

The measurement provided by GlueDomains are used by network aware services to estimate *closeness* between nodes.

The *closeness* criteria is uses a simple model that estimates a network cost using network measurements. According with the DIANA approach [3]:

$$netCost = \frac{Losses}{MaxBand}$$

where

$$Losses = w_1 RTT + w_2 Loss + w_3 |Jitter|$$

RTT, MaxBand, Loss and Jitter are rispectively the values of Round Trip Time, MaxBand estimates, packet loss rate and Jitter during the period of observation.

A replica optimization service has been implemented using the above network cost estimationas.

A light-weight Web service interface is used to select closest replica of a dataset used by jobs in a Computing or Storage Element, addressed using their Site URL.

The web service interface uses the nusoap library, that provides a set of PHP classes to create and consume web services based on SOAP, WSDL and HTTP.

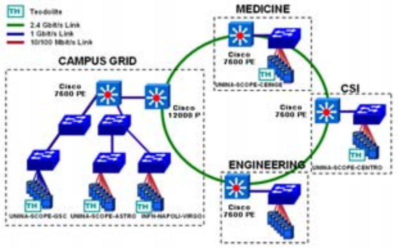
```
<message name="netCostRequest">
   <part name="lfn" type="xsd:string"/>
   <part name="dest" type="xsd:string"/>
</message>
<message name="netCostResponse">
   <part name="neCost" type="xsd:string"/>
</message>
<portType name="Network Resource ManagerPortType">
<operation name="netCost">
   <input message="tns:netCostRequest"/>
   <output message="tns:netCostResponse"/>
</operation>
```

THE GLUE DOMAIN DEPLOYMENT

The middleware layer supporting the domain-based INFN Grid network monitoring activity is powered by GlueDomains [2].

The GlueDomains service is managed by a dedicated central Server node, located in the main Campus Grid site.

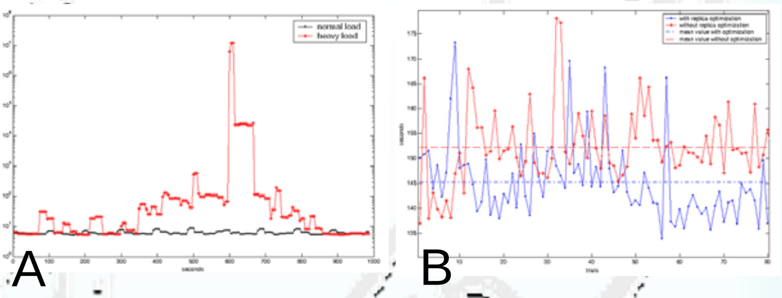
The Server periodically checks and configures a number of Theodolites, deployed across five sites in the SCoPE infrastructure. The activity of Theodolites consist of running autonomuosly a number of active network performance probes, as shown in figure.



Theodolites activity is performed by the Storage Elements and the monitoring topology is a full mesh.

Round-trip Time, Packet Loss Rate and One-Way Jitter performance measurements are periodically measured for each domain-to-domain path.

The data produced by Theodolites are periodically published through the GridICE web interface which is also used to check theodolite operation.



The figure on the top, show the results of some tests performed with the replica optizimation service implemented.

Regarding the network cost trend we have made 12 hours of observation of the cost betwen two sites, the results show an average netcost value of 6.3 with 0.8 standard deviation, during the normal network activity. On the other side, during the different phases of the heavy network workload we measured a significant growth of the netcost value with its maximum in the order of 107.

In figure A we showed the netcost behaviour varying in time under normal and heavy load, calculated between the UNINA-SCOPE-ASTRO and INFN-NAPOLI-VIRGO sites' theodolites.

The second experiment consists in downloading in the storage element of the INFN-NAPOLI-VIRGO site, a set of 100 1.2 GB sized files that are replicated in the UNINA-SCOPE-ASTRO, UNINA-SCOPE-GSC, UNINA-SCOPE-CEINGE and UNINA-SCOPE-CENTRO sites. The files are registered in the SCoPE logical file catalogue. The tests have been split in two phases:

– Replication of 100 files on the INFN-NAPOLI-VIRGO site by using the lcg-utils tools and the logical file name.

– Replication of 100 files on the INFN-NAPOLI-VIRGO site aided by the network-aware replica optimization service.

We measured an average transfer time on 100 files of 154 seconds for each file by using the lcg-utils versus 145 seconds by obtained by using the replica optimization services. In fig. B we show the replica optimization services effects in reducing the data transfer time.