Network monitoring for EGEE and beyond

Etienne Dublé - CNRS/UREC EGEE SA2
Xavier Jeannin - CNRS/UREC EGEE SA2

- Third TNLC - July 3, 2009 -

www.eu-egee.org

Content

- **Introduction**
- **Approach and status of the analysis**
  - Actors and their requirements
  - Services to be provided
  - Technical aspects to be considered
  - Available architectures to build on
- **Foreseen points of discussion for this TNLC**

Introduction

Main goals and reasons for having such a monitoring system

Introduction

- No network monitoring solution at the **EGEE grid level**:
  - PerfSONAR Lite TSS is troubleshooting-oriented
  - Need (also) a system for continuous network monitoring, for EGEE and the future (EGI / NGIs)
- SA2 was initially not in charge of this design work
- However, since it is an important problem, an SA2 team was set up and started working on this subject with the support of DANTE and NRENs of the TNLC:
  - Etienne Dublé (CNRS)
  - Xavier Jeannin (CNRS)
  - Mario Reale (GARR)
  - Alfredo Pagano (GARR)

Introduction

- Main problems:
  - ✓ Should be an end-to-end cross-domain solution
  - ✓ High number of sites (around 300)
    - Impact on deployment
    - Impact on performance
  - ✓ Very high number of possible site-to-site paths to monitor
- This on-going analysis is written in the document https://edms.cern.ch/document/1001777

Approach and status of the analysis

1) Actors and their requirements

Actors and their requirements

- Classify applications and middleware module according to needs in term of network metrics (QoS RFC 4594):
  - EGEE gLite-based applications and gLite itself have low requirements
  - Another use case is DORII (Remote Instrumentation Infrastructure) because it is partially based on EGEE
    - ➢ DORII has higher requirements
  - Other use cases / projects maybe in the future
- Middleware functional needs
  - ✓ Provide monitoring data to the File Transfer Service
  - ✓ Other needs specified in earlier documents (ex DJRA4.7)

Actors and their requirements

- ROC / Grid Operations and Sites (to be precised)
- ENOC
  - Same needs as ROC and Sites but at project level
  - The system should also check the SLAs
- Grid Users
  - Want to know why a job failed / was too slow

Approach and status of the analysis

2) Services to be provided

Services to be provided

- User interface
  - Geographical maps / Other views…
  - adapted views to a given site / a given VO / a given ROC or NGI / considering only the SLAs
  - Instant view
  - Historical data, need of state of reference to compare current state with
- Alerts by email
- Web services to allow other systems to retrieve monitoring data

Approach and status of the analysis

3) Technical aspects to be considered

Technical aspects to consider

- Set up of the system should be an incremental process (in order to provide the first services ASAP)
- The system will live in a constantly changing environment
- The system should be acceptable for grid sites (costs, security)
- Aspects related to metrics acquiring
  - Client/server based probes have to be synchronized (NTP)
  - Probes must be accurate enough (JRA4-TEC-475908)
  - Measurements should be done in both ways (but one way would be acceptable in a first step)
  - Probes behavior should not be affected by heterogeneous hardware
  - If metrics are aggregated, proof of correctness should be provided beforehand

Technical aspects to consider

- The system should provide the following metrics, for one end-to-end path:
  - RTT and, if possible, OWD
  - Jitter
  - Packet Loss
  - Bandwidth (TCP is required):
    - One of (or, better, both):
      - ✓ Achievable bandwidth (tools like iperf)
      - ✓ Available bandwidth (tools like pathchirp)
    - And optionally the end-to-end path Capacity
  - The MTU
  - Maybe topology changes
- For short-term results, focus on metrics easy to set up first

Technical aspects to consider

- The combination of possible site-to-site paths is around 300*299/2 = 44850 https://twiki.cern.ch/twiki/bin/view/EGEE/RoutesMAP

Technical aspects to consider

- We have to choose the paths we will monitor
- List of ideas to determine them:
  - Monitor paths where a SLA was established
  - Discard path A-B if sites A and B do not host a common VO
  - Reflect the network architecture of the mains projects (LHC experiments, etc…)
  - Consider with a lower priority the paths with many hops (imprecise info when looking for a problem, and paths with much traffic are usually short)
  - Consider the geographical distance between sites

Technical aspects to consider

- Consider the computing resources of the end-sites (i.e. consider with a higher priority the sites with more computing resources)
- Consider the real site-to-site traffic (application logs)
- Possibly subdivide the map in domains and then consider only one of its sites for the inter-domain map
- Consider with a higher priority the sites which have frequent network problems (get this info from GGUS)
- Considering all site-to-site routes, have all the network segments monitored (i.e. at least one of the path we monitor is going through a given segment)

➢ We must find a good combination of some, or all of these criteria.

Technical aspects to consider

- Other technical aspects:
  - Determine the acceptable frequency of measurements
  - Specify
    - How client-server probes will be synchronized
    - How the system will avoid measurements affecting each other (for example two bandwidth measurements going at the same time through the same network sub-path)
  - Determine the archiving process
  - Consider the security aspects

Approach and status of the analysis

4) Available architectures to build on

Available architectures

- Designing a whole network monitoring system “from scratch” is **not** an option
- Therefore we consider adapting / adopting one of the following existing architecture:
  1. PerfSONARLite TSS
     - Extend this solution for continuous monitoring (since for now it is troubleshooting-oriented)
  2. PerfSONAR MDM
     - Collaborate with GEANT / DANTE in order to adapt their solution
  3. The EGEE grid itself
     - Probes would be implemented as grid jobs, resulting in a zero-deployment system

Available architectures

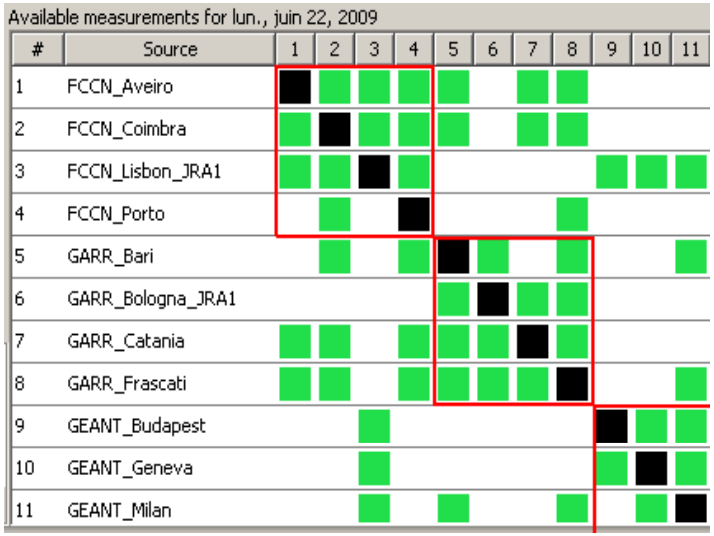
- We need more time to choose one of these architectures
- However we (as SA2) have a close look at the evolution of each of them since:
  - **PerfSONARLite TSS** is being deployed (DFN SA2 Team)
  - **PerfSONAR MDM** is being deployed / tested in Spain (RedIRIS SA2 Team)
  - We are running prototyping tests of a **grid-based architecture** (CNRS SA2 Team)
    For example the network data on this map was retrieved using grid jobs

Foreseen points of discussion

Topics for which your experience would be the most valuable for us

Points of discussion

- perfSONAR UI shows many missing values when probes have to cross several domains.
  For example on this screen dump the measurements are represented by green squares, and left blank when missing.
  We rounded by a red square the probes in the same domain.
  It appears obvious that most of the missing measurements are cross-domain ones (i.e. outside the red squares).
  This seems to show that probes (which are usually UDP or ICMP -based) are often filtered at domain boundaries.
  Do you have any experience you would like to share about efficient **cross-domain network monitoring tools & methods**?



Points of discussion

- Are **aggregated metrics** usable for an operational project or are they still prospective? If they are usable, in which conditions?
- Do you know any side effect when running probes on **heterogeneous hardware**?

Other questions

Thank You.

https://edms.cern.ch/document/1001777