Managed by

European Research Consortium for Informatics and Mathematics

ERCIM

Network Monitoring - A scalable approach

Report about the activity of the Network Monitoring project of WP 5.1

Augusto Ciuffoletti

INFN/CNAF Bologna

augusto@di.unipi.it

http://www.coregrid.net

Confluence of autonomous activities

The activity within the project is divided into three main branches.

- Management of the Network Monitoring Layout
  - Specific subtask, resp. CNAF
- Network Monitoring Tools
  - Specific subtask, resp. FORTH
- Data publication
  - Non specific subtask, common proposals

The convergence to a consistent solution is enforced by person to person discussions, writing or revising documents, informal contacts.

A meeting point: the session description

- The whole Network Monitoring activity is split intoSessions.
- Session descriptors are produced by workflow managers as a request for monitoring activity.
- Session descriptors are delivered to specific agents.
- Such agents control the operation of the NetworkMonitoring tools.

A meeting point: the session description

- The Session Descriptor describes the interface between
  - the abstract view of GRID resources reflected by the Network Monitoring layout, and
  - the detailed instructions required by monitoring tools.
- We are currently defining the XSDs for session description, focussing on simplicity (we want a prototype) and expandibility (serving as a basis for real scale applications)

A simple example

```
<session id=1234@this.netmonelem.ip>
    <requestby>broker@nmdomain</requestby>
    <expires>future:time</expires>f
    <toolname>gridmon</toolname>
    <parameters>
      <metric>
         <name>BandwidthUsage</name>
         <scope>
            <source><nmdomain> cnaf.infn </nmdomain></source>
            <destination><nmdomain> forth.ics </nmdomain></destination>
         </scope>
         <protocol_list>
           <protocol> GridFTP </protocol>
         </protocol_list>
      </metric>
    </parameters>
  </session>
```

Behind the Schemas

The capabilities of a Network Monitoring Tool:

- passive monitoring techniques
- dynamically configurable
- supports most relevant metrics

A Network Monitoring layout abstraction:

- represents a domain structure (2-levels hierarchy)
- hosting agents specialized in Network Monitoring
- be dynamically configurable

MAPI - A network monitoring architecture

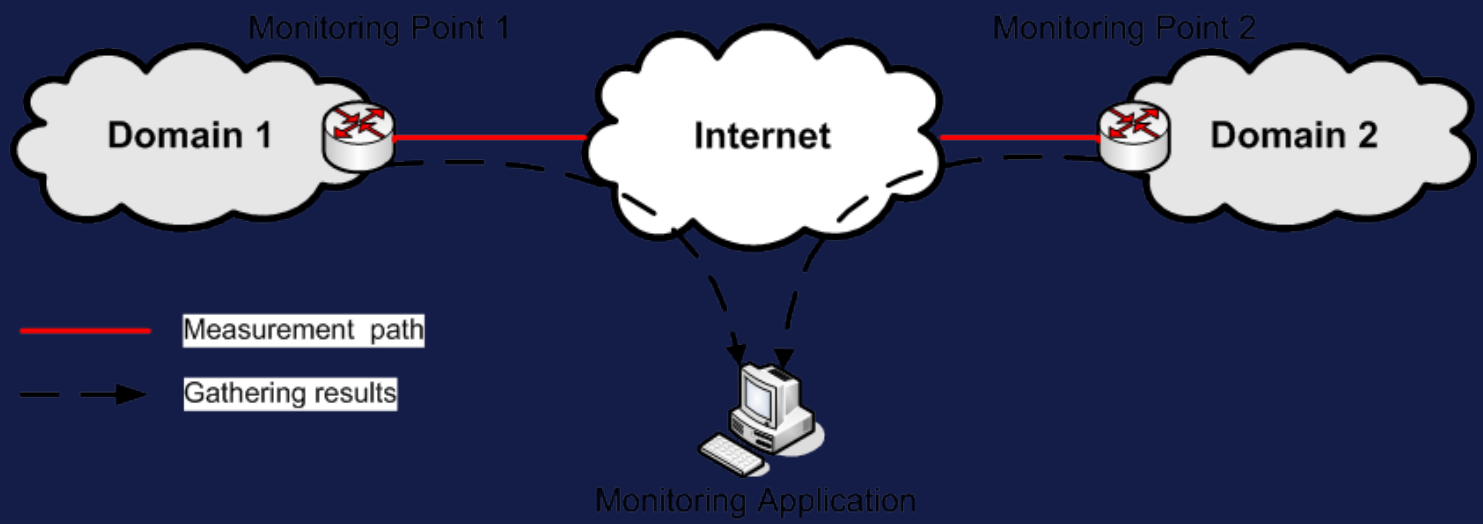
- Based on passive monitoring techniques to measure several network characteristics (scalability)
- It is dynamically configurable to add/remove new measurements (scalability)
- It takes care of security issues (anonymization, aggregation, authentication)
- It can be controlled from remote sites (authentication)

MAPI - Packet loss measurement

Problem: cannot synchronize measurement start & stop.

Basic idea: count number of retransmissions during a connection.

Features: non-intrusive, per application, third party monitoring.

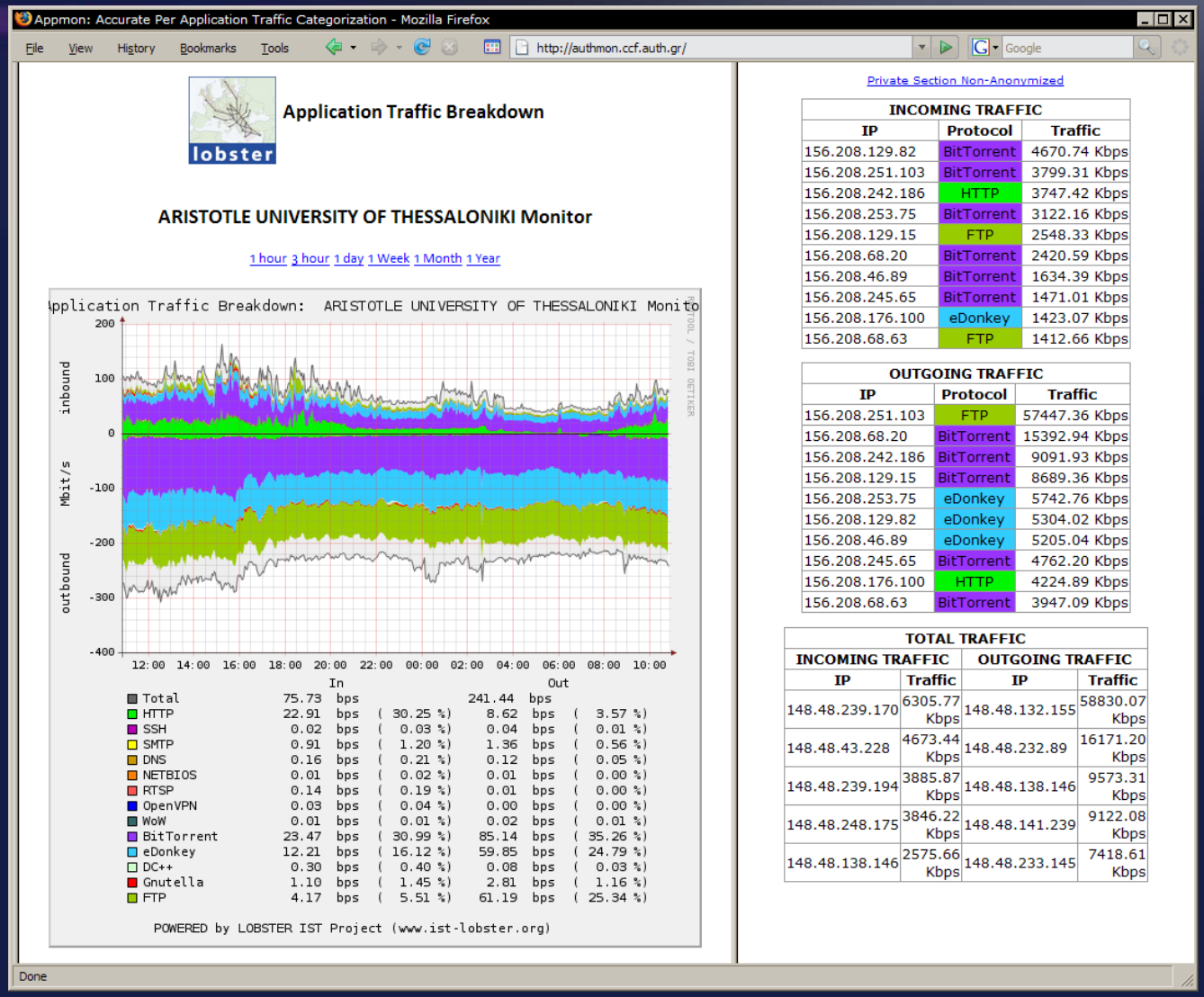


APPMON - Accurate Traffic Characterization

Based on MAPI

- Categorizes and visualizes per-application traffic
- Gives an answer to common questions
  - What applications are running in my network?
  - Which of them consumes most of the bandwidth?
  - Which hosts are consuming most of my organization's network bandwidth and what applications are they running?

APPMON – A prototype application



- Graph showing the categorized per-application traffic
- Report of the *"top 10"* bandwidth consuming IP addresses
- Custom selection of the monitored subnet/hosts
- Deployed for administrative monitoring

Network Monitoring Layout management - Generalities

- Partitions the GRID into domains, based on link level topology (scalability)
- Introduces agents that are specialized in the management of Network Monitoring (security)
- Enforces coordination between such agents (scalability?) to:
  - share knowledge about domain membership (security)
  - share knowledge about network monitoring capabilities (GGF?)
  - establish the membership of NM agents (security)

Domain partitioning

- Grid resources that share similar connectivity characteristics with the rest of the network are represented as a unique domain (not DNS)
- The Network Monitoring layout consists in the full (DxD) mesh (scalability?)
- We introduce identifiers for Domains, and represent Network Elements as Domain pairs (scalability)
- In principle, monitoring activity is on demand (scalability)

Network Monitoring Elements (NMEs)

- NMEs hold the knowledge of the Network Monitoring Layout
  - mapping resources to domains
  - mapping other Network Monitoring Agents to domains
- NMEs serve as proxies to access local Network Monitoring capabilities:
  - they accept Network Monitoring requests from applications/users (security)
  - they offer a Network Monitoring Layout directory service to local application/users (scalability)

Coordination among Network Monitoring Elements

- A relevant issue for scalability: we need to avoid single points of failure, and security leaks
- We observe that the membership is quite stable (although not static): high update latency is not an issue
- We address a solution with a O(1) cost (consumed resources), and a O(N) latency
- We privilege scalability for latency: other options are possible

Note: this issue already indicated in a GGF document about network monitoring requests

A token passing based coordination protocol

- A single token circulates in the system
- At each token exchange the NM Layout database of the destination is synchronized with that of the source
- Token destination is selected randomly among all NM Agents
- Token loss is handled by timeout and removal of spurious tokens
- Join operations introduce a spurious token
- The underlying protocol ensures security using certificates contained in the Layout database

Secure Token Passing protocol

Type of packets:

- Move pkt: submits token exchange.
- Acknowledge pkt: accepts packet exchange
- Commit pkt: confirms token exchange
- EarlyStop pkt: stops resending Commit pkts

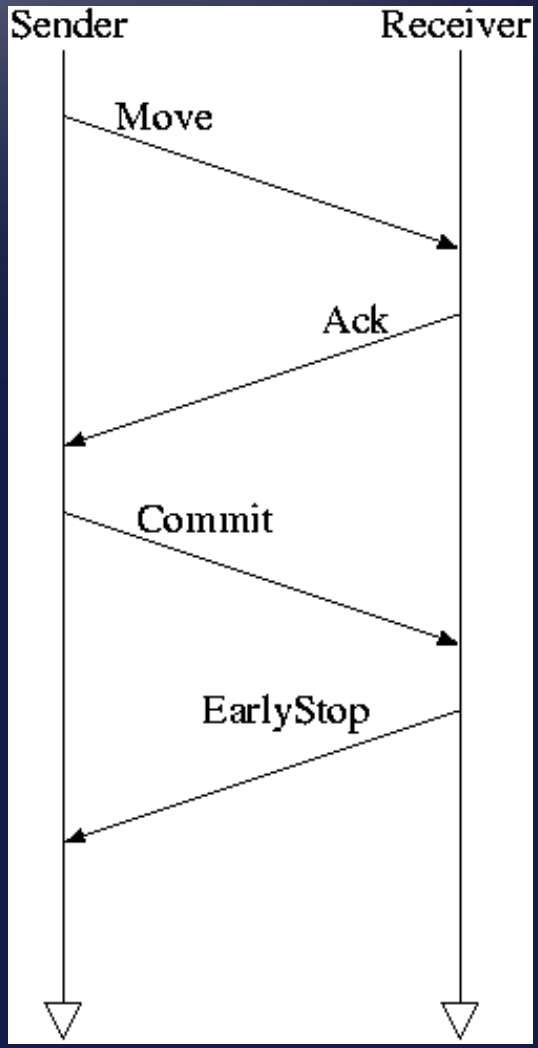
All types except EarlyStop can be resent

Non conformant packets are silently dropped

All packets are signed

Token passes if and only if

- source predicate: ack received
- receiver predicate: commit received



Results: simulations and experiments

- Group Membership protocol: simulation of 300 agents for a real-time application in an Internet Scale network (DAPSYS 2006).
- A multi-token protocol (lower latency) in preparation.
- A resilient token passing protocol passed a long run trial in the Internet(18 days operation before token loss, one token passing operation every 10 seconds, p(loss)<10**5).

Conclusions

- Activity follows two distinct branches, carried out by the two participants (FORTH and CNAF).
- A limited number of CoreGRID papers give evidence of the coordination among the two branches.
- Other non-CoreGRID papers assess the autonomous validity of the work of the teams.
- One report/paper in preparation about common topics (CoreGRID label).


Information Society
Technologies


CoreGRID