Flexible Scheduling in QoS-aware Cooperative Service Execution Environments

Luís Nogueira, Luís Miguel Pinho

$\{$luis,lpinho$\}$@dei.isep.ipp.pt.

IPP-Hurray! Group

School of Engineering, Polytechnic Institute of Porto

Porto, Portugal

Highly dynamic scenario

- Consider a network with several heterogeneous nodes
  - Each node may supply a **distinct set** of resources, with **different capacities**

- There will be several services to be executed
  - Tasks arrive/leave **while** others are being executed
  - Each service has **users' specific QoS constraints**
  - All compete for the **finite set** of resources

- Service cannot be satisfied by a single node or handles it inefficiently
  - Nodes must **cooperate** to fulfil resource and performance demands
  - Split service's tasks among the **best subset** of nodes

Requirements (1)

- Supporting QoS requirement's specification **layering**
  - User (e.g. high quality)
  - Applications (e.g. frame rate)
  - System (e.g. period and cost)

- Specify QoS allowing **quality tradeoffs**
  - Ranging from a preferred to a minimum tolerable SLA
  - Maximising user's influence on provided QoS

- It may be **useless to compute optimal** service allocations
  - **Tradeoff** computation **time for quality** of results
  - **Anytime algorithms** for framework's management

Requirements (2)

- **Dynamic** QoS
  - Initial SLA may have to be **downgraded**
  - Current SLA can be **upgraded** on underutilisation
  - Dynamic **adaptation** of promised stability
    - High reconfiguration rate may be undesirable

- Resource usage must be **monitored in run-time**
  - Decisions based on **actual usage** and not on requested assumptions

- Schedule set of tasks with **variable execution times**
  - Using soft guarantees based on average times
  - Giving **priority to users' tasks** in overloads
  - Mix of **isolated** and **non-isolated** servers