QoS-driven Web Services Selection in Autonomic Grid Environments

Danilo Ardagna
Gabriele Giunta
Nunzio Ingraffia
Raffaela Mirandola
Barbara Pernici
Introduction

In SOA, complex applications can be composed as business processes invoking a variety of available WS.

Develop applications by specifying component WS only through their required functional characteristics and non-functional constraints.

Select WS during process execution from a registry of available services, guaranteeing QoS constraints.
Satisfying QoS constraints

- QoS requirements are difficult to satisfy especially for the high variability of workloads and need autonomic computing self-managing techniques
- Grid provides basic mechanisms to manage a service center infrastructure, simplifying the re-configuration of the physical resources
- We propose a reference framework to support the execution of e-business applications in autonomic grid environments
- The problem of selection of WS will be analyzed in depth, assuming variable QoS profiles for WS and the long term execution of the composed services
Outline

- An e-business case study
- Preliminary definitions
- Reference Framework
- Optimization Problem Formulation
- Experimental Analyses
- Conclusions and Future work
Case study

- VO composed by a set of small and medium enterprises
- ERP software execution (*AgileERP*)
- The cost allocation process

€ 1,500

\[
\begin{align*}
C_n &= \text{Transfer Cost Center} \\
C_{n.1}, C_{n.2}, C_{n.3} &= \text{Receiver Cost Center}
\end{align*}
\]
Case study

- VO composed by a set of small and medium enterprises
- ERP software execution (AgileERP)
- The cost allocation process

€ 1,500
Case study

- VO composed by a set of small and medium enterprises
- ERP software execution (AgileERP)
- The cost allocation process

€ 1,500
The cost allocation process
The allocate costs sub-process

For $i=1..|CC|$

FALSE

TRUE

CostAccounting. CalculateCharge Balance(CC[i])

FALSE

TRUE

AR=0

CostAccounting. CalculateTotalCost Transferred (AR)

CostAccounting. AllocateCosts

CostAccounting. NavigateCost Hierarchy

End of CH

FALSE

TRUE
Preliminary Definitions

- A Web service is modeled as a software component which implements a set of operations.
- Service Registries stores also for each operation $o$ of a Web service $j$, the values of QoS $q_{j,o}^n(t)$ and the number of instances $N_{j,o}(t)$ that can be executed during a specific time instant.
- A composite service is specified as a high-level business process in BPEL language in which the composed Web service is specified at an abstract level.
- Abstract Web service (task $t_i$) and concrete Web services ($ws_j$).
Preliminary Definitions

BPEL annotations:
- global and local constraints on quality dimensions
- cycles maximum number of iterations and expected frequency of execution of conditional branches
- user preferences, \( \{\omega_1, \omega_2, \ldots, \omega_N\}, \sum \omega_n = 1 \)
- Web service dependency constraints

Constraints and BPEL annotations specified by WS-Policy:
- Global Constraints:
  - \( \text{ExeTime} \leq 2 \text{ days} \)
  - \( \text{ExeTime}(\text{BalanceVerification}) + \text{ExeTime}(\text{NotificationError}) \leq 20 \text{ sec} \)
- Web service dependency:
  - \( \text{BalanceVerification, NotificationError} \)
QoS Profiles

\[ q^n_{j,o}(t) \]

\[ u=13 \]

\[ u=28 \]

\[ \Delta \]

\[ T \]

\[ 2T \]

\[ t \]
Grid Reference Framework

Virtual Organization

Local Grid

Local Resource Allocator

Broker

Service Registry

Grid Physical resources
Web Services

Virtual Organization

Local Grid

Local Resource Allocator

Broker

Service Registry

Grid Physical resources
Web Services

WAN

Virtual Organization

Local Grid

Local Resource Allocator

Broker

Service Registry

Grid Physical resources
Web Services

Virtual Organization

Local Grid

Local Resource Allocator

Broker

Service Registry

Grid Physical resources
Web Services
Resource Management in Grid Environments

Resource management introduces two different optimization problems which correspond to the VOs (providers) and users perspectives:

- each VO would like to maximize the SLA revenues and the use of physical resources
- the end user is interested in the maximization of the QoS of the composed service execution
Local Resource Allocation

- Each VO needs to allocate local grid physical resources to different Web service operation invocations $ws_{j,o}$ in order to maximize the revenues from SLA, while minimizing resource management costs.
- Autonomic techniques allow the dynamic allocation of physical resources among different Web services on the basis of short-term demand estimates.
- The local resource allocation is performed periodically with period $\Delta'$ (e.g., 10-30 minutes).
Local Resource Allocation

- Short-term workload predictor
- Grid Monitor
- Workload statistics
- Local Resource Allocator

- Long-term workload predictor
- Historical workload statistics
- Local Resource Allocator Algorithm (Simulation)

\[ N_{j,o}(t) \]

\[ q^n_{j,o}(t) \]

\[ f_{j,o}(t) \]
Maximizing QoS for the End User

- Execution Path ($ep_k$): a set of tasks $\{t_1, t_2, ..., t_i\}$, such that $t_1$ is the initial task, $t_i$ is the final task and no $t_{i1}$, $t_{i2}$ belong to alternative branches
- Execution Plan: a set of ordered triples $\{(t_i, ws_{j,o}, x_i)\}$, indicating that task $t_i$ included in $ep_k$ is executed at time instant $x_i$ by invoking $ws_{j,o}$
- Global Plan: a set of ordered triples $\{(t_i, ws_{j,o}, x_i)\}$, which associates every task $t_i$ to a given WS invocation $ws_{j,o}$ at time instant $x_i$
- $q^n(k)$: n-quality dimension value along execution path $k$
Execution Paths

Execution path $ep_1$

1. CostHierarchy. BalanceVerification(CC[])
2. CostAccounting. CalculateCharge
3. Balance(CC[1])
4. CostAccounting. CalculateTotalCost
5. Transferred (AR)
6. CostAccounting. AllocateCosts
7. CostAccounting. NavigateCost Hierarchy

Execution path $ep_2$

1. CostHierarchy. BalanceVerification(CC[])
2. CostHierarchy. NotificationError

Additional steps:
- Report.GenerationPDF
- Report.GenerationExcel
- CostAccounting. ConsolidateCosts
Problem Formulation

- The WSC problem is multi-objective and a simple additive weighting technique is used to evaluate the overall value of QoS from multiple quality dimensions.

- Decision variables:
  - $y_{i,j,o,u}$ equals 1 if the Web service operation $ws_{j,o}$ executes task $t_i$ during the time interval $u$, 0 otherwise.
  - $w_{i,u}$ is equal to 1 if the task $t_i$ is executed during time interval $u$, 0 otherwise.
  - $x_i$ start time of task $t_i$. 

D. Ardagna, G. Giunta, N. Ingraffia, R. Mirandola, B. Pernici - GADA 2006 3rd November
Maximizing QoS for the End User

\[
P1) \quad \max_{\sum_{i=1}^{N} \sum_{j \in WS} \sum_{o \in OP} \sum_{u=1}^{U} \gamma_{i,j,o,u} \cdot y_{i,j,o,u}}
\]
Maximizing QoS for the End User

\[
x_I + q^1_I \leq E \\
q^n(k) \geq Q^n \quad \forall k, n \\
x_i \in R^+ \quad \forall i \\
q^n_i \in R^+ \quad \forall i, n \\
q^n(k) \in R^+ \quad \forall k, n \\
y_{i,j,o,u}, w_{i,u} \in [0, 1] \quad \forall i, j, o, u
\]

The problem is NP-hard, equivalent to a Multiple choice Multiple Dimension Knapsack Problem
Experimental Results

- Analyses performed on a wide set of randomly generated problem instances
- Problem solution obtained by CPLEX, a state of the art integer linear programming solver (execution time limited to one minute, approximate solution)

Percentage gap between the approximate and global optimum solutions
Conclusions and Future work

This paper presents a framework for the development of e-business applications built on autonomic grid computing infrastructure. Service selection and composition is performed to guarantee that the overall quality perceived by the user is maximized. Implementation of the proposed approach and its validation on a real industrial test bed. Optimization of a set of instances.
Thanks! Any questions?
Local Constraints

\[ \sum_{j \in WS_{i_1}} \sum_{o \in OP_j} \sum_{u=1}^{[E/\Delta]} P_{j,o,u} y_{i_1,j,o,u} \leq \bar{P} \]
Web Service Dependency Constraints

\[
\sum_{o \in OP_j} \sum_{u=1}^{[E/\Delta]} y_{i1,j,o,u} = \sum_{o \in OP_j} \sum_{u=1}^{[E/\Delta]} y_{i2,j,o,u} \quad \forall j \in WS_i_1 \cap WS_i_2;
\]

\[
\sum_{o \in OP_j} \sum_{u=1}^{[E/\Delta]} y_{i1,j,o,u} = 0, \forall j \in WS_i_1 \setminus WS_i_2;
\]

\[
\sum_{o \in OP_j} \sum_{u=1}^{[E/\Delta]} y_{i2,j,o,u} = 0, \forall j \in WS_i_2 \setminus WS_i_1.
\]