Multifaceted Resource Management for Dealing with Heterogeneous Workloads in Virtualized Data Centers

Íñigo Goiri, J. Oriol Fitó, Ferran Julià, **Ramón Nou**, Josep Ll. Berral, Jordi Guitart and Jordi Torres

Barcelona Supercomputing Center (BSC) and UPC Barcelona Tech.

- Provider benefit depends on
  - Reducing energy consumption
  - Avoid violating SLAs
- Virtualization is used to
  - Consolidate different tasks in the same physical host
  - Save energy and reduce management complexity
- Execute heterogeneous tasks on top
  - HPC tasks based on deadline
  - Web-based applications based on response times and uptime

- Virtualization adds overheads
  - Creation time
  - Migration
  - Disk management
- Aggressive consolidation for saving energy
  - May incur in performance loss, i.e. less benefit
- SLA violation $\rightarrow$ provider pays penalty
  - Delay finish time of HPC tasks
  - High response time of Web applications

- We propose a new scheduling policy
  - Increases provider benefit
  - Reduces energy consumption
  - Manages virtualization overheads
  - Reduces SLA violations
  - Allows outsourcing of resources
  - Supports fault tolerance
  - Runs HPC and web-based applications

Economic-based scheduling

Scheduling algorithm

- Decides where to run a VM dynamically
  - Evaluates every VM allocation in every host (physical server)
- Calculates benefit for each allocation
  - Aggregation of revenues and costs
  - Higher benefit is better
- Finds VM scheduling with higher benefit

Economic-based scheduling

Calculate benefit

- Benefit of a tentative allocation of virtual machine *VM* in host *H*
- Aggregation of revenues and costs
  - Execution revenue
  - Virtualization overhead
  - Energy cost
- $Benefit = Revenue - \sum Costs$

Economic-based scheduling

Cost calculation: Hardware, software, and resource requirements

- If the host cannot fulfill the VM requirements:
  - Lacks required hardware: number of CPUs, disk...
  - Lacks required software
  - Lacks required hypervisor
  - ∞ cost
- If the host does not have enough free resources:
  - Not enough CPU, memory...
  - ∞ cost

Economic-based scheduling

Cost calculation: Virtualization overhead

- Overhead introduced by virtualization management
  - Time to create the VM
  - Time to migrate the VM
- Estimate SLA penalty
  - Extra time added to HPC tasks
  - Loss of performance

Economic-based scheduling

Cost calculation: Energy consumption

- Estimate energy consumption
  - Get host load
  - Get hosts energy consumption
  - Assess consumption per each VM
  - Transform it to cost using electricity price

Economic-based scheduling

Other features

- Virtualized datacenters provides new capabilities
  - Outsourcing
  - Checkpointing
- Our scheduling policy supports them

Economic-based scheduling

Outsourcing

- Run VMs (resources) in external IaaS providers
- Model an external provider as a big local host
  - Revenue of executing the task
  - Cost to outsource
  - Overhead of starting VMs in an external provider
- Add a new virtual host per each external provider

Economic-based scheduling

Fault tolerance

- Uncertain hosts crashes
  - Estimate hosts reliability
- Resume failed VMs
- Evaluate VM tolerance to failure
  - Web applications: state-less
  - HPC tasks: need checkpoints

Solving scheduling

- Calculate benefit of the current VMs allocation
- Optimize system in order to get higher benefit for VM allocations
  - Hill climbing algorithm
- Possible changes to be applied to the system:
  - Create VM
  - Execute (outsource) VMs in external providers
  - Migrate VM among nodes
  - Cancel VMs execution
  - Keep VMs that cannot be executed in a queue
  - Apply turn on/off policy

Energy savings

Turn on/off policy

- Turn on/off approach: save energy
  - Use two thresholds
  - Turn off idle servers
    - Turn off servers as soon as they are not used
  - Turn on new machines if they are required
    - Wait until machines are needed
- Use of consolidation
  - More energy savings
  - Lower performance

Evaluation

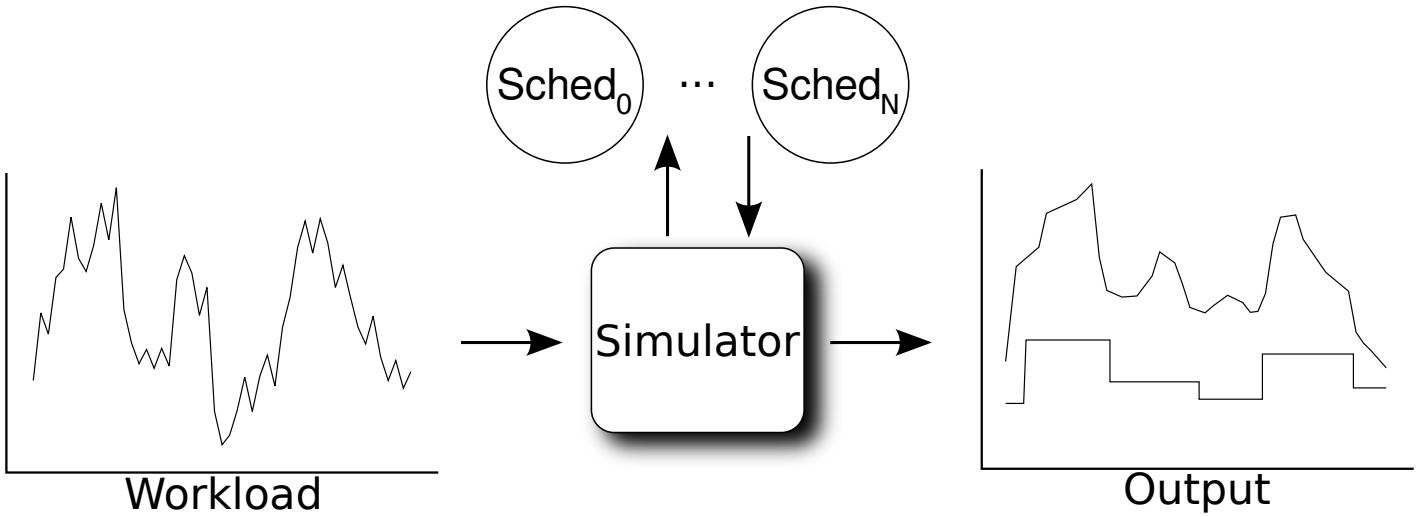
Environment

- Simulated environment
- Heterogeneous real workload (one week)
  - Grid 5000:
    - $\sim$2000 tasks with an average of $\sim$5000 seconds per task
  - Anonymous European ISP
    - Aggregation of several web-based services
- 100 virtualized hosts

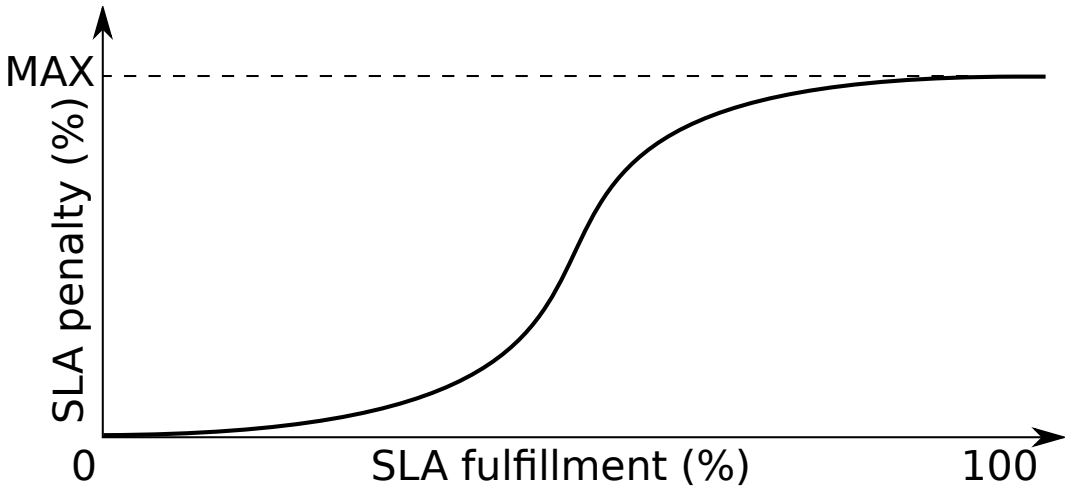
Evaluation

Environment: power simulator

- Simulate nodes with different features
  - Fast and reproducible results
- Scheduling policies:
  - Random
  - Round robin
  - Backfilling
  - Dynamic backfilling
  - Economic-based



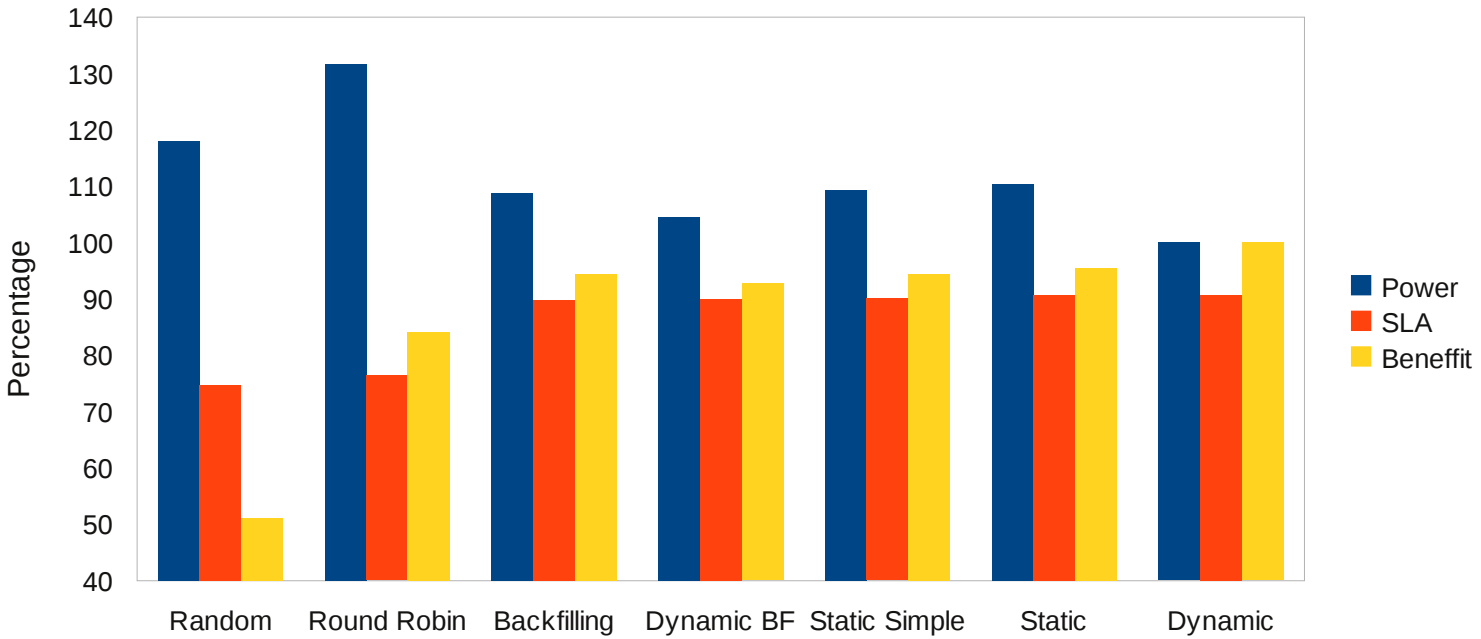
- Metrics:
  - Energy consumption (Wh)
  - Client satisfaction (SLA fulfillment)
  - Benefit (euros)
- SLA penalties:



Evaluation

Local allocation

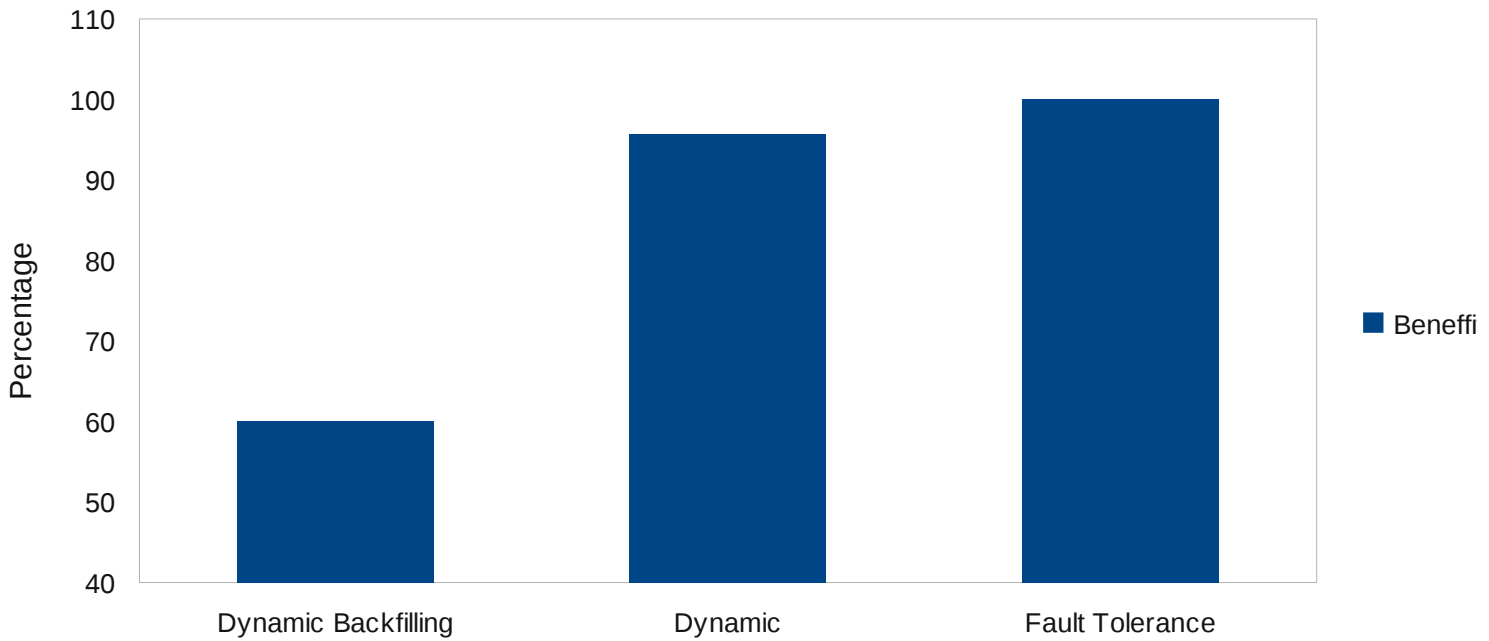
- Metrics
  - Normalized average power to the best policy
  - Client satisfaction
- Scheduling policies
  - Static: no migration
  - Dynamic: use migration to consolidate



Evaluation

Fault tolerance

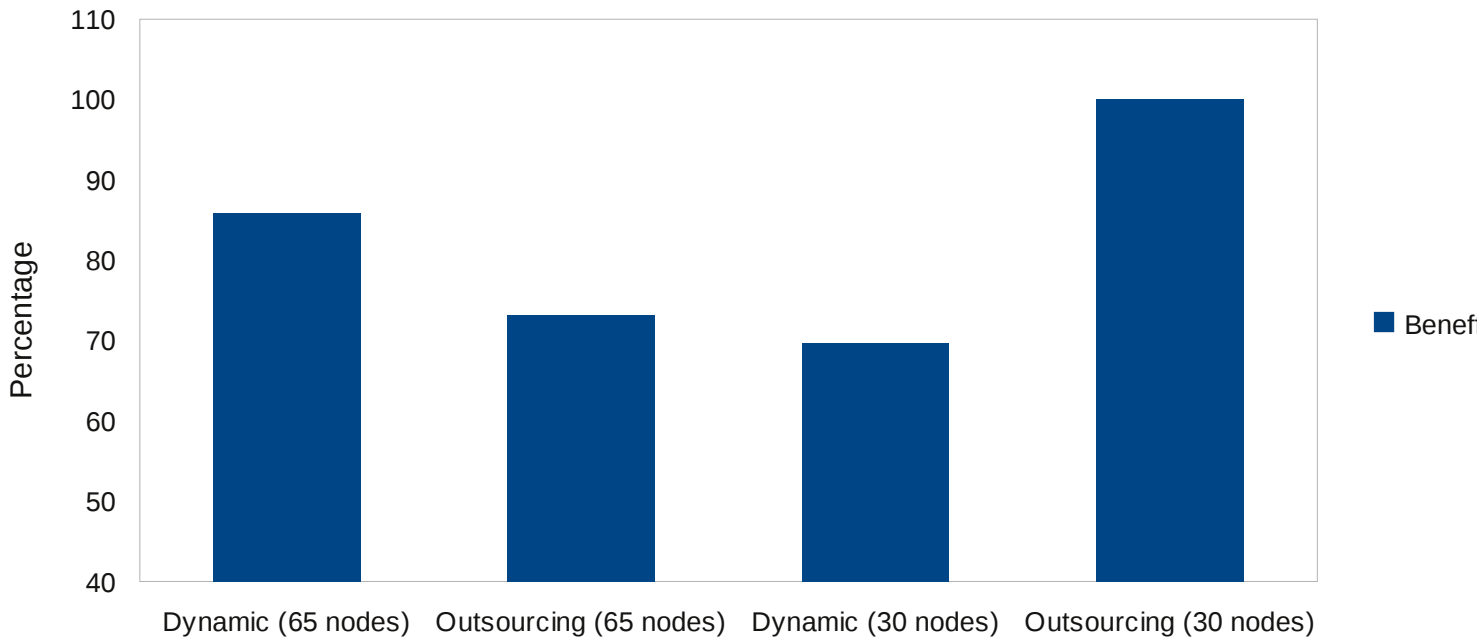
- Scheduling policies in a faulty environment
  - 99.9% node reliability: an average of one or two failures per week
- “Fault tolerance” policy
  - Performs checkpoints of HPC tasks
  - Able to manage with node crashes



Evaluation

Outsourcing

- Outsource resources in order to withstand period of peak loads
  - Using 65 nodes: support peak load and high CAPEX
  - Using 30 nodes: outsource peak load and save CAPEX



Conclusions and future work

- Improve energy efficiency
- Deal with virtualization overheads
  - More SLAs are fulfilled
- Increase provider's benefit
- Future work
  - Dynamic SLA enforcement
  - Powering on/off servers with dynamic turn on/off thresholds
  - Add more heterogeneity support

Multifaceted Resource Management for Dealing with Heterogeneous Workloads in Virtualized Data Centers

Íñigo Goiri, J. Oriol Fitó, Ferran Julià, **Ramón Nou**, Josep Ll. Berral, Jordi Guitart and Jordi Torres

Barcelona Supercomputing Center (BSC) and UPC Barcelona Tech.