Multi-Leveled Hierarchical Control to Optimize Workload of a Service-Oriented Platform

Maksim Khegai, Tatiana Kharchenko, Dmitrii Zubok and Alexandr Maiatin
ITMO University, St.Petersburg, Russia
Table of Contents

- Introduction and motivation
- Basic architecture of the system
- Architecture of the system
- The algorithm
- Experiments results
- Conclusion
Introduction and motivation

- A performance control was and will be a very important part of every computing system
- Modern cloud systems utilize complex system architecture and as a result have a lot of parameters for performance control
- Simultaneous use of several optimization techniques is not widely researched
Basic architecture of the system: Levels

• Physical level
• Virtual level
• Applications level
• Jobs queuing level
• Supplementary level
Basic architecture of the system: Controller

- Receive a job
- Decide on which path a job must take
- Send data to application
- Receive data from application
- Send a result to user
Basic architecture of the system: Optimization

- Monitor performance and available resources
- Search for a path
- Store knowledge about performance for future use
Architecture of the system
The algorithm

- Jobs queuing step
- Applications redistributing step
  - Determine application with lowest performance
  - Determine virtual machine with highest performance and maximum free resources
  - Check if there are enough resources
  - Reallocate application
- Migration and creation of virtual machines
- Starting or stopping physical servers
Experiments results:
First step only
Experiments results:
Second step only
Experiments results: 
First and Second steps together
Conclusion

- The performance of the system was improved by using two optimization methods.
- Use of a multi-leveled optimization shows its potential.
- Future work lies in implementing two last steps.
Thank you for your attention