Secure multi-level parallel execution of scientific workflows on HPC Grid resources by combining Taverna and UNICORE services

Sonja Holl*, Olav Zimmermann*, Bastian Demuth*, Bernd Schuller*, Martin Hofmann-Apitiu+

*Jülich Supercomputing Centre (JSC)  
Institute for Advanced Simulation (IAS)  
Forschungszentrum Juelich, Germany

*Fraunhofer Institute for Algorithms and Scientific Computing SCAI  
Schloss Birlinghoven, Sankt Augustin, Germany
Outline

- Motivation / Use case
- Architecture
- Security issues
- Results
- Conclusion
Taverna Life Science Workflows on the Grid

- Compute intensive
- Often contains independent tasks
- Benefit from High Performance Computing (HPC) resources
- Access to external data repositories or databases
- High security requirements (licence agreement, resource access)
- Explicit and reliable resource access

Combination of Taverna and UNICORE

May 30th 2012
Sonja Holl
HPC Support for Taverna

- Execution of individual jobs on the Grid
- Set of applications can be easily extended
- X.509 security mechanism
- Benefit: easy creation, execution and sharing of compute-intensive workflows in Taverna
Use Case

- **Optimization of scientific workflows**
  - Stochastic optimization to search for optimal parameters
  - Genetic Algorithms (GA) or Particle Swarm Optimization (PSO)
  - Enable various parameter constraints (functions, mathematical dependencies, fixed parameter sets,..)

- Scenario with many independent workflow instances
Use Case

Parameter Constraints

Start

Initialize New Parameter

Create Workflow Population

Run Workflows

Evaluate Generation Results

End

Optimal Parameter Set

May 30th, 2012

Sonja Holl
Take more advantage of HPC-resources

- Typical optimization design using GA:
  - 50 individuals, 10 generations
  - 50 workflow executions in parallel 10 times

- Workload and memory requirement too high for execution in the Taverna Client

- *Enable execution of individual workflow instances on the Grid*
Taverna Workflow Grid Plugin

- Extend Taverna Workbench by a new workflow submission mechanism
- Development of a new activity for workflows
  - Identify and consume inputs
  - Submit the workflow
- Reuse parts of the submission mechanism from the previously developed UNICORE-Taverna Plugin\(^1\)

\(^1\) S. Holl, O. Zimmermann, M. Hofmann-Apitius
A UNICORE Plugin for HPC-enabled Scientific Workflows in Taverna 2.2
Proceedings of the 2011 IEEE Seventh World Congress on Services (SERVICES 2011), July 2011
The architecture – three layer parallelism

**Client**
- **Taverna**

**Grid**
- **UNICORE Service Orchestrator (SO)**
- **Taverna**
- **UNICORE SO**

**HPC resource**
- **UNICORE SO**
- **UNICORE SO**
- **UNICORE SO**

- Parallel execution of workflows
- Distribution on different Grid resources
- Parallel submission of independent WF steps
- Distribution on different Grid resources
- MPI or thread-parallel application
New security issues

- Taverna uses Credential Manager to access X.509 certificate
  - basic security feature for job submission
- Requires extended security mechanism
  - Trust Delegation
New security issues

Taverna requires a user certificate for submission

Client

Grid

HPC resource

Sonja Holl

May 30th, 2012
Security propagation mechanism

Client

Grid

HPC resource

Publish server certificate DN via IDB

May, 30th 2012
Sonja Holl
Security propagation mechanism

Client

Grid

HPC resource

Use SAML assertion for trust delegation to submit job
Results

- CPU load on the client machine

![Graphs showing CPU load for different execution scenarios]
Results

- Performance issues during Grid execution
  - Scaling is independent of workflow size and parallel executions (GA parameterization)
  - Usage of UNICORE parameter sweep extension would reduces the CPU load to one job
Conclusion

- Execution of Taverna workflows on the Grid
- New security propagation mechanism via trust delegation

Outlook

- Speed up the execution of Taverna Server
- Decrease Taverna client workload and submission overhead by adaption of UNICORE parameter sweep extension
- Further scaling tests
Questions?
Thank you!