FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

A Novel Approach to Workflow Management in Grid Environments

Frank Berretz*, Sascha Skorupa*,
Volker Sander*, Adam Belloum**

15/04/2010

* FH Aachen - University of Applied Science, GER

** University of Amsterdam, NL

Outline

Taxonomy of Grid Workflow Systems

> Common Architecture

Push-based Job Distribution

> Requirements

> UNICORE Workflow System

> Consequences

Pull-based Approach

> Benefits & Challenges

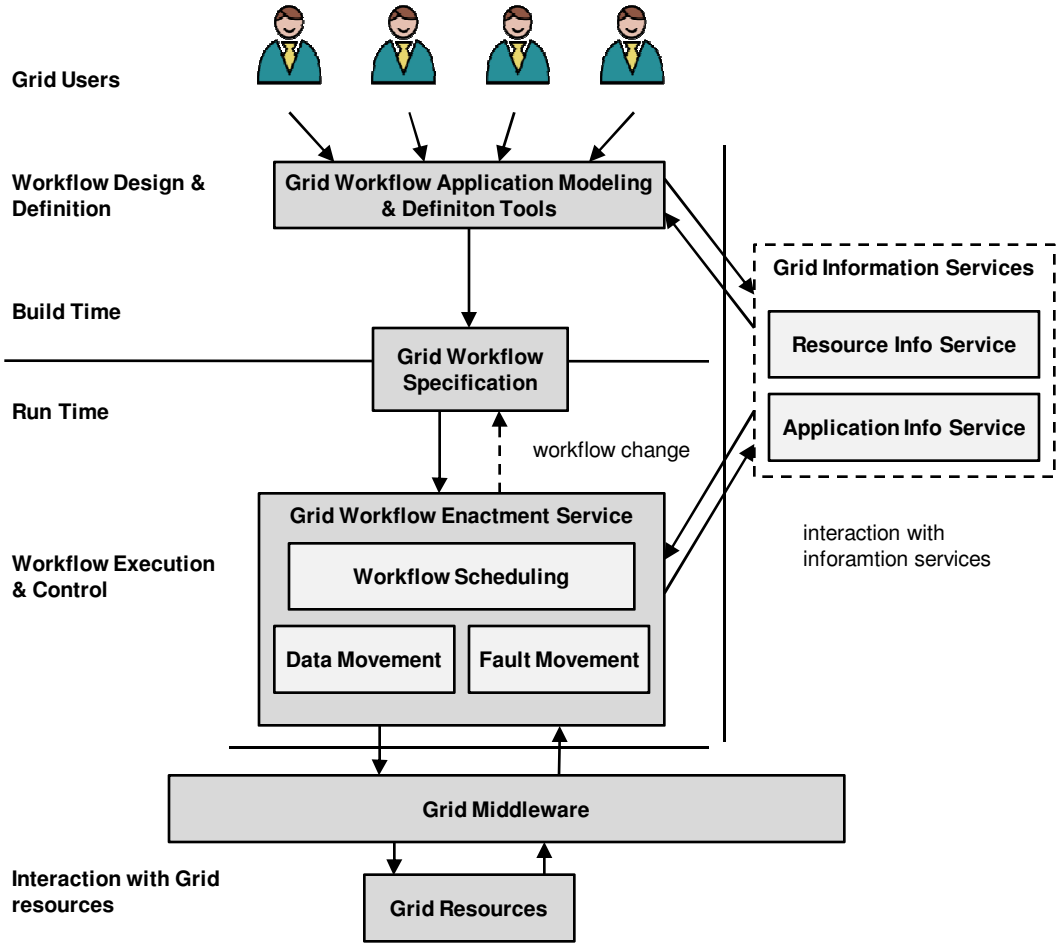
> General Architecture

Prototype Implementation

> UNICORE Grid Middleware

> jBPM Workflow Engine

Taxonomy of Grid Workflow Systems Common Architecture



Source: Jia Yu and Rajkumar Buyya

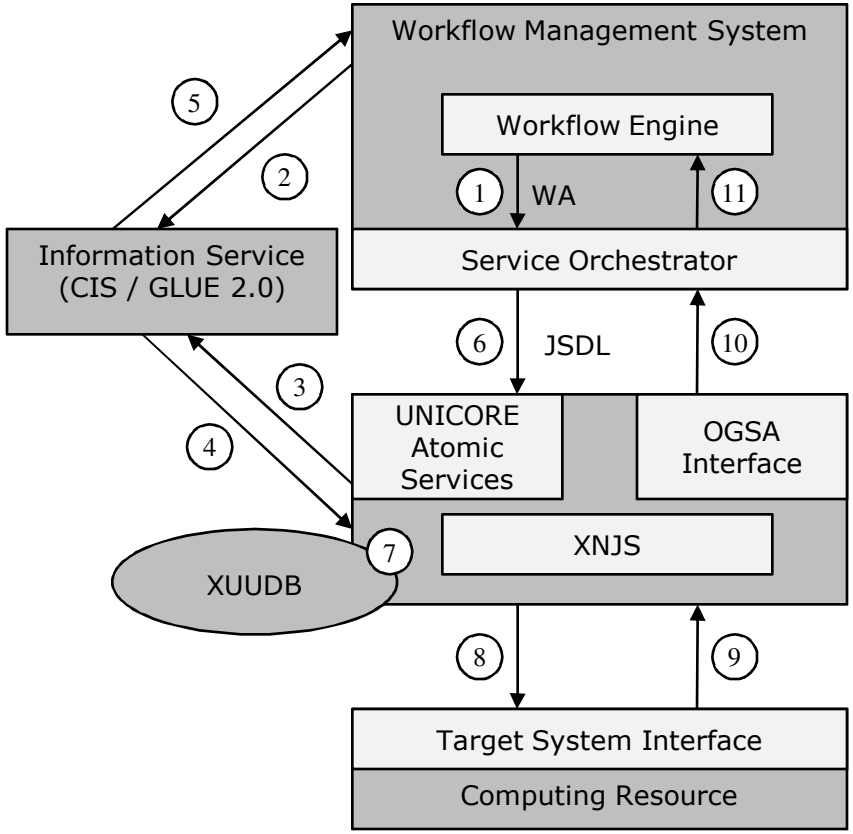
Push-based Job Distribution Requirements

Push-based job distribution requires

- Efficient resource discovery and selection processes
- Detailed knowledge of available resources
- Well-defined interfaces of resources
- Up-to-date and confidential information systems
- Adapt VO schedulers and local schedulers
- Proper Access Control Lists from resource providers

Push-based Job Distribution UNICORE Workflow System

1. Splitting workflow into sequence of WAs and send them to a orchestrator.
2. Filter appropriate resources by requesting an information service.
3. Information Service requests all available Grid sites.
4. Evaluate requests from multiple VOs.
5. Response concrete resource endpoint.
6. Service Orch. forwards JSDL to known interface.
7. Site performs authorization by mapping CA to a local account.
8. XNJS sends job through the TSI to a physical **computing** resource.
9. – 11. Callback chain



Push-based Job Distribution Consequences

Scalability: Information systems and schedulers may become bottlenecks with respect to the amount of…

> Users and resources

> Parallel branches inside a workflow (parameter studies)

Cross Grid Scheduling:

> Side effects caused by resources in multiple Grids

> Limitation of resource candidates to resources of a particular Grid (Open community approach?)

Heterogeneity: Grid Workflow Systems typically deal with computational resources

> Cumbersome integration of special resources like human interaction (lack of integrating emerging standards)

> Complex decision processes by humans might influence further workflow steps (e.g. qualitative assessment criteria)

Alternative Idea
Pull-based Job Distribution Strategy

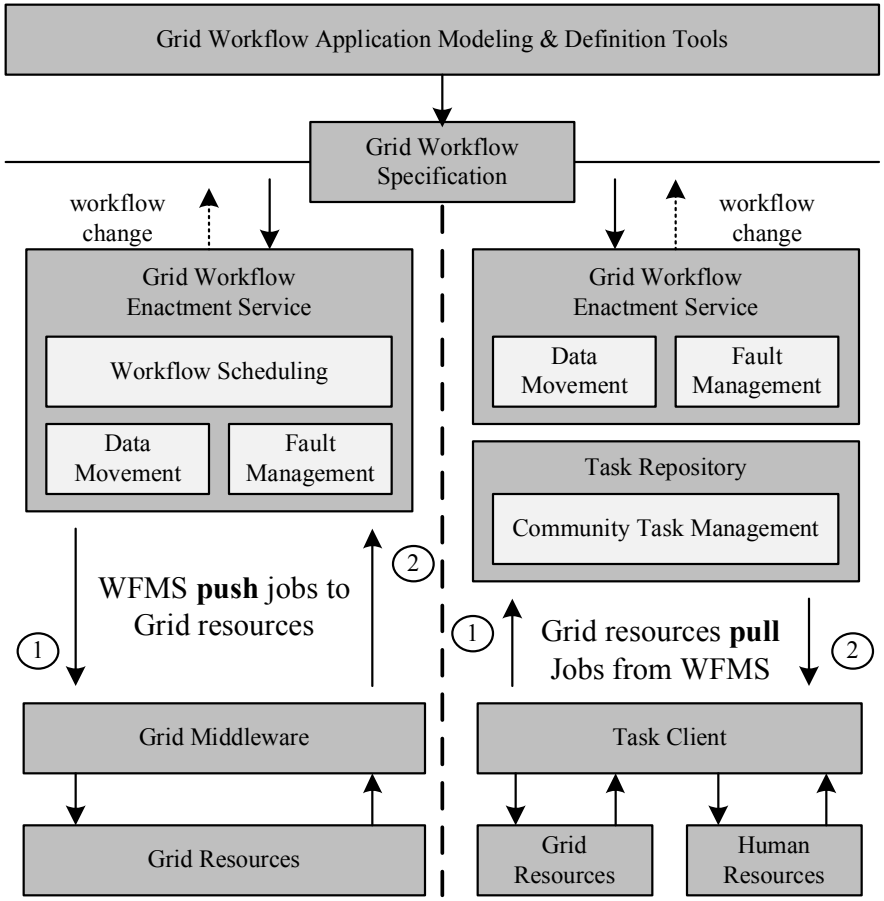
Workflow system sends task to an intermediary repository

Resources act autonomously and adapt to the repository

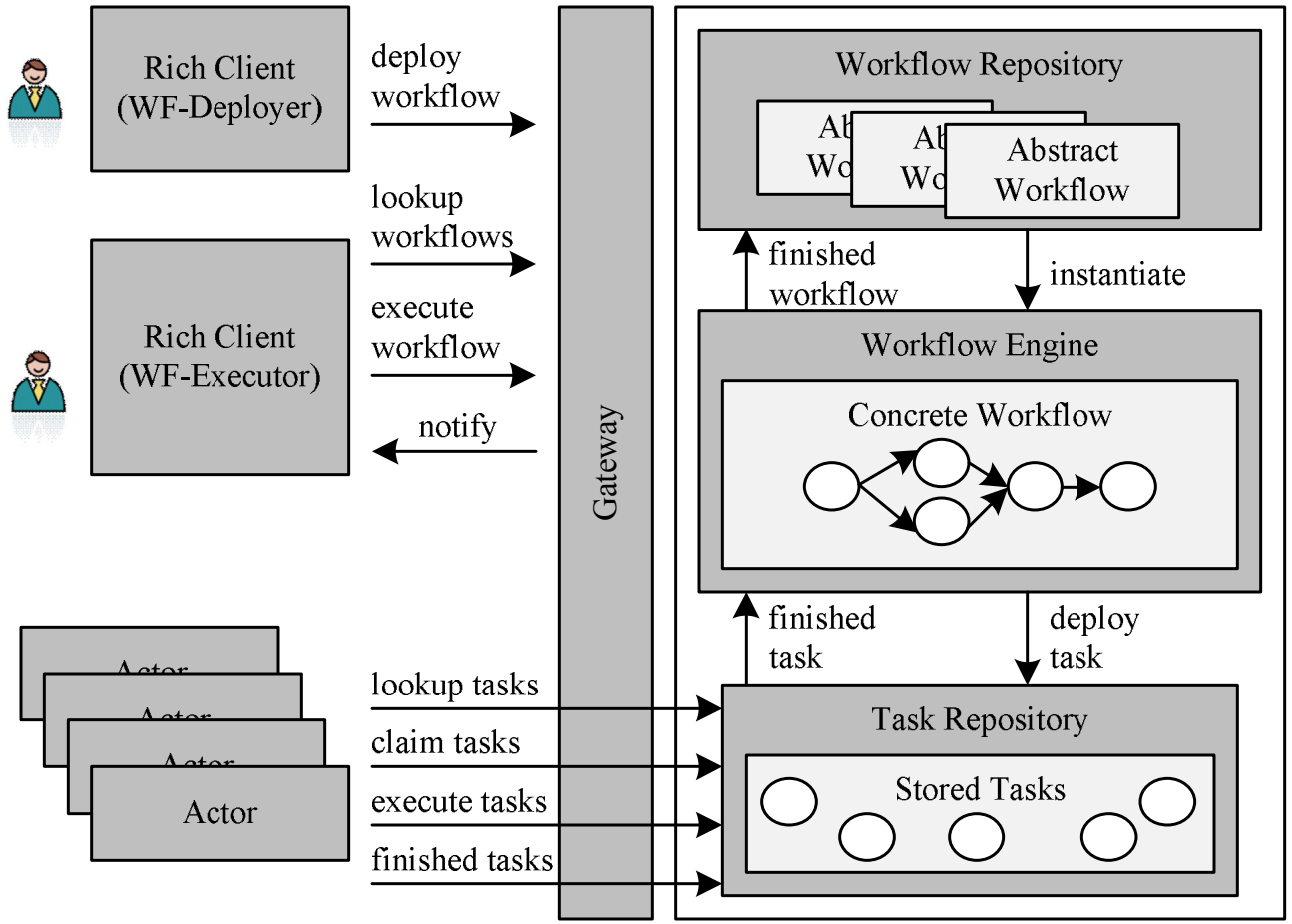
Any kind of resource can actively request the repository

Resources apply for defined roles to receive tasks according to their capabilities

Resources have to authenticate against the task repository



Architectural Concept for the Pull Model



Benefits & Challenges of Pull-based Approach

Benefits

- Scheduler and brokering components are now optional
- Simplified integration of special resources like humans, telescopes or medical devices
- Reduced administrative VO management overhead at resource sites
- Support actors across organizational boundaries (community approaches)

Challenges

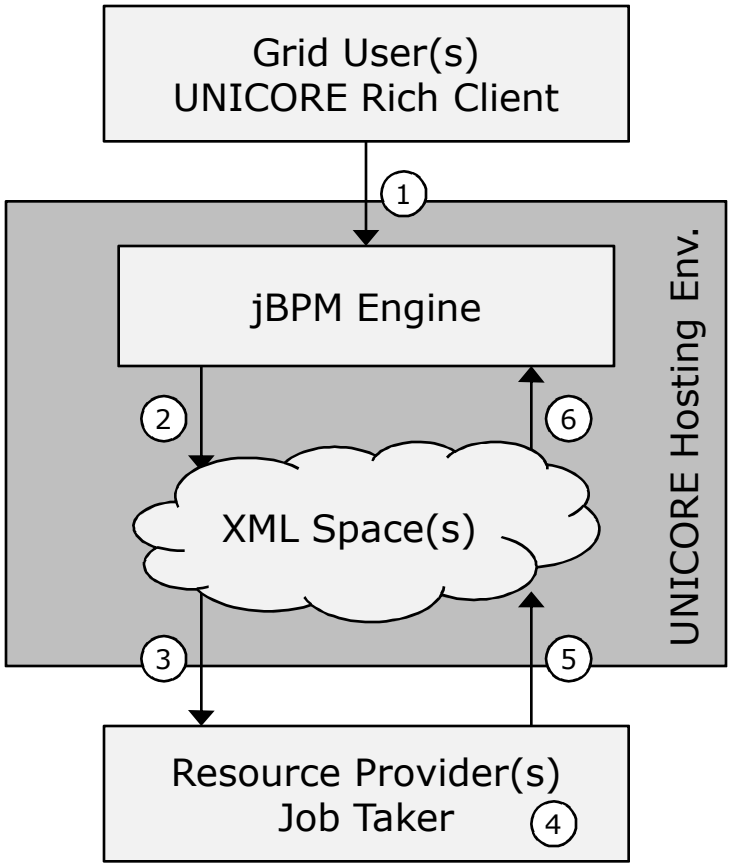
- Bottleneck problem should not be shifted to the task repository
- Submitted jobs run the risk of starvation (SLAs!)
- Appropriate security and provenance frameworks needed

Integrate Pull Approach to an existing Grid Middleware

Extending UNICORE Grid middleware

Use existing XML Tuple Space as repository

Integrate jBPM Workflow Engine as client for the space



1. User starts workflow
2. Engine writes job to space
3. Resource takes job from space
4. Resource executes job locally
5. Resource finishes job
6. Engine receives notification by the space and resumes workflow

Summary and Outlook

Conclusions:

> Pull-based job distribution strategies are currently missing in Grid systems

> But it could be used as an alternative model for certain application scenarios

(heterogeneous resources, high-throughput computing, …)

Summary and Outlook

Further steps:

- XML space should be replaced by a scalable task repository
- Among the simple Job-Takers more complex client systems should be implemented to integrate special resources (humans) into Grids
- Hybrid push/pull distribution strategies as an option
- Performance and scalability analysis

This work should result in a refined architecture to address the challenges of pull-based approaches

FH Aachen University of Applied Science
Campus Jülich

Ginsterweg 1
52428 Jülich
T +49. 241. 6009 53277
berretz@fh-aachen.de, skorupa@fh-aachen.de
www.fh-aachen.de