PHOSPHORUS

Scalable Design of Resilient Optical Grids

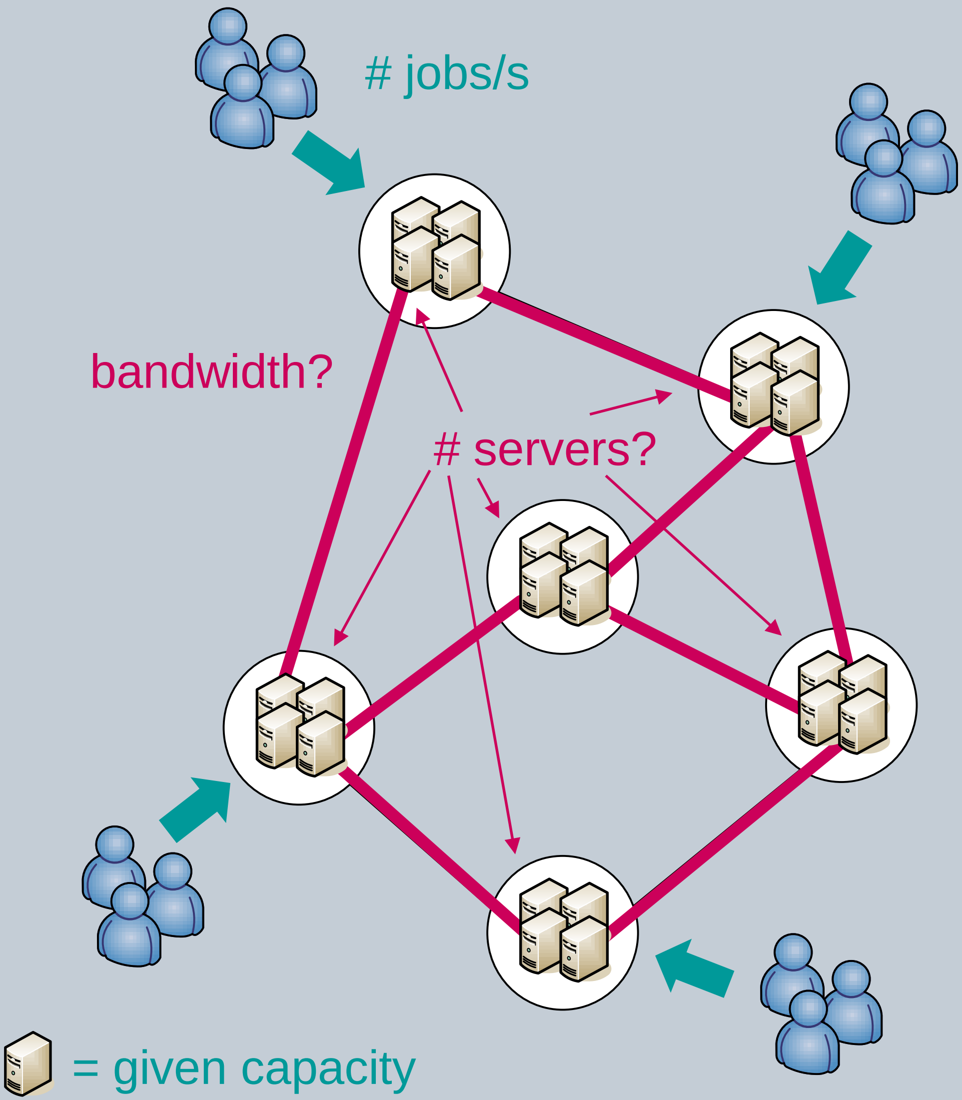
Marc De Leenheer
Ghent University - IBBT

On-Demand Network Services for the Scientific Community
Terena Networking Conference 2009, June 7th, Malaga, Spain

GRID NETWORK DESIGN

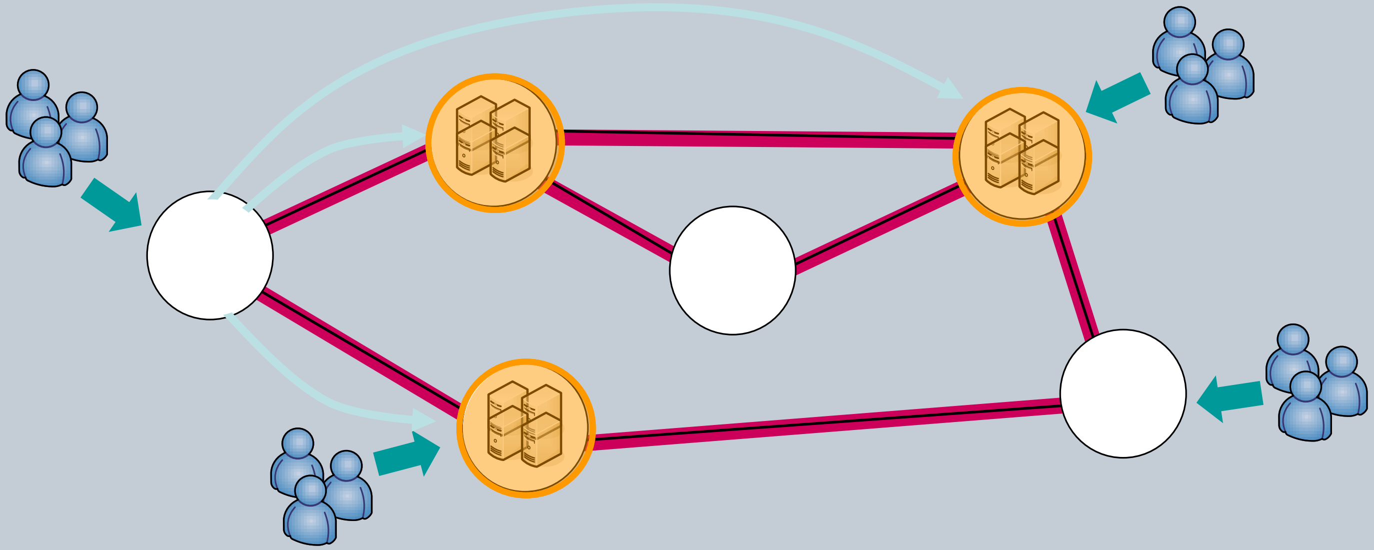
Grid dimensioning - Problem Statement

- Given:
  - Network topology
  - Job arrival process
  - Job processing capacity
  - Target loss rate
- Find
  - Locations of servers,
  - Amount of servers,
  - Amount of link bandwidth
- While
  - Meeting max. loss
  - Minimizing network capacity

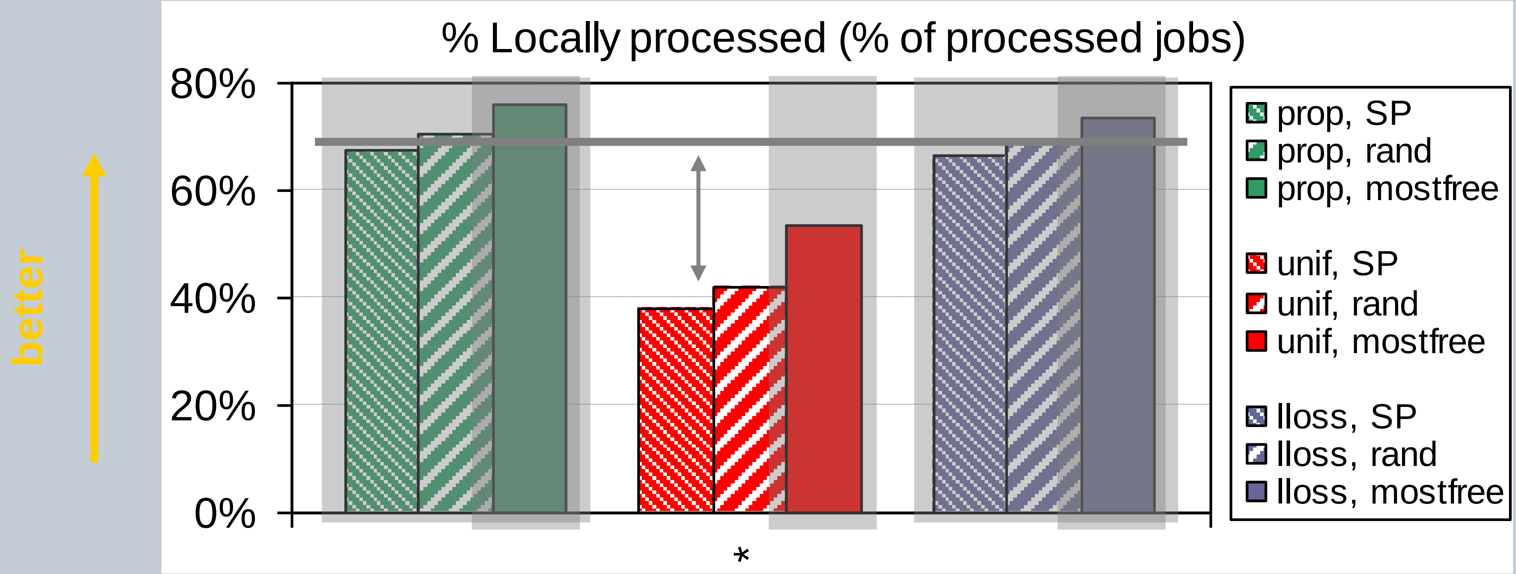


Solution

- Phased approach
  - 1 Determine K server locations (approx., ILP)
  - 2 Determine server capacity (analytical, ErlangB)
  - 3 Determine inter-site bandwidths (simulation)
  - 4 Dimension link bandwidths (=number of wavelengths)



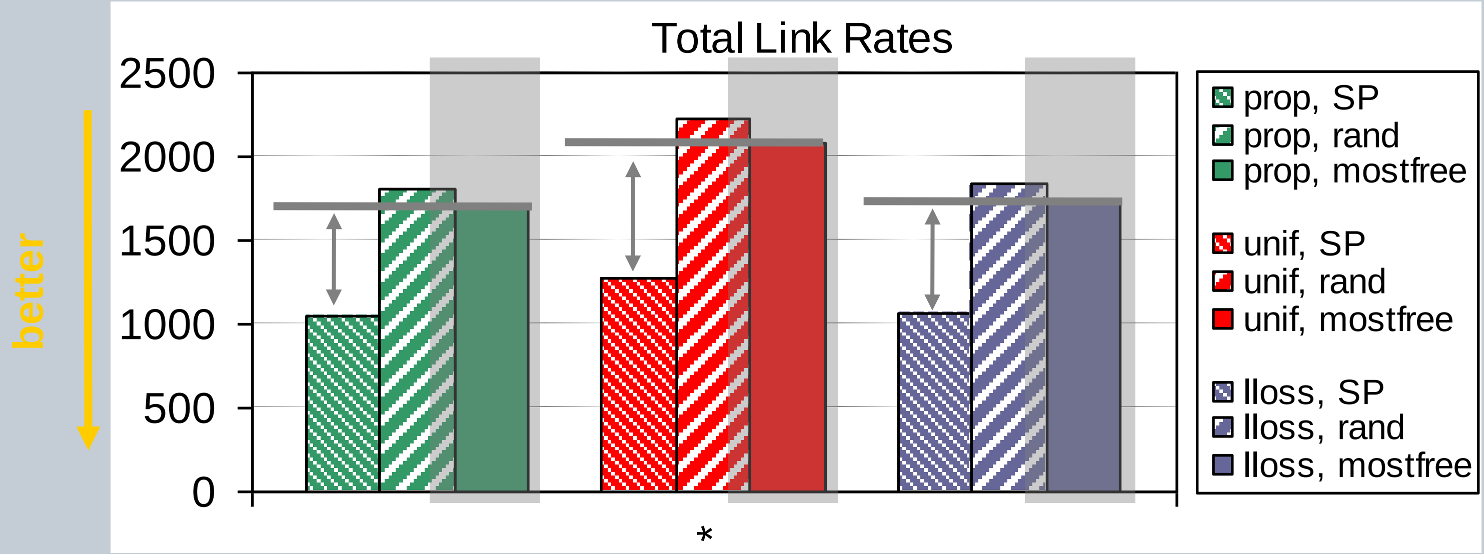
Results: 'Local' processing rate



- Server distribution:
  - ***unif***: uniformly distributed
  - ***prop***: ~ local arrival rate
  - ***lloss***: ~ same (local) loss rate
- Scheduling: local first, if busy then…
  - ***SP***: shortest path
  - ***rand***: randomly pick a free site
  - ***mostfree***: site with most free servers

- Conclusions:
  - ***mostfree*** achieves highest local processing
  - Intelligent server placement (prop, lloss) achieves higher local processing

Results: Link bandwidths



- Link bandwidths:
  - Non-uniform server distribution (prop, llos) leads to significant bandwidth reduction
  - Intelligent scheduling (***mostfree***) comes at a link bandwidth price

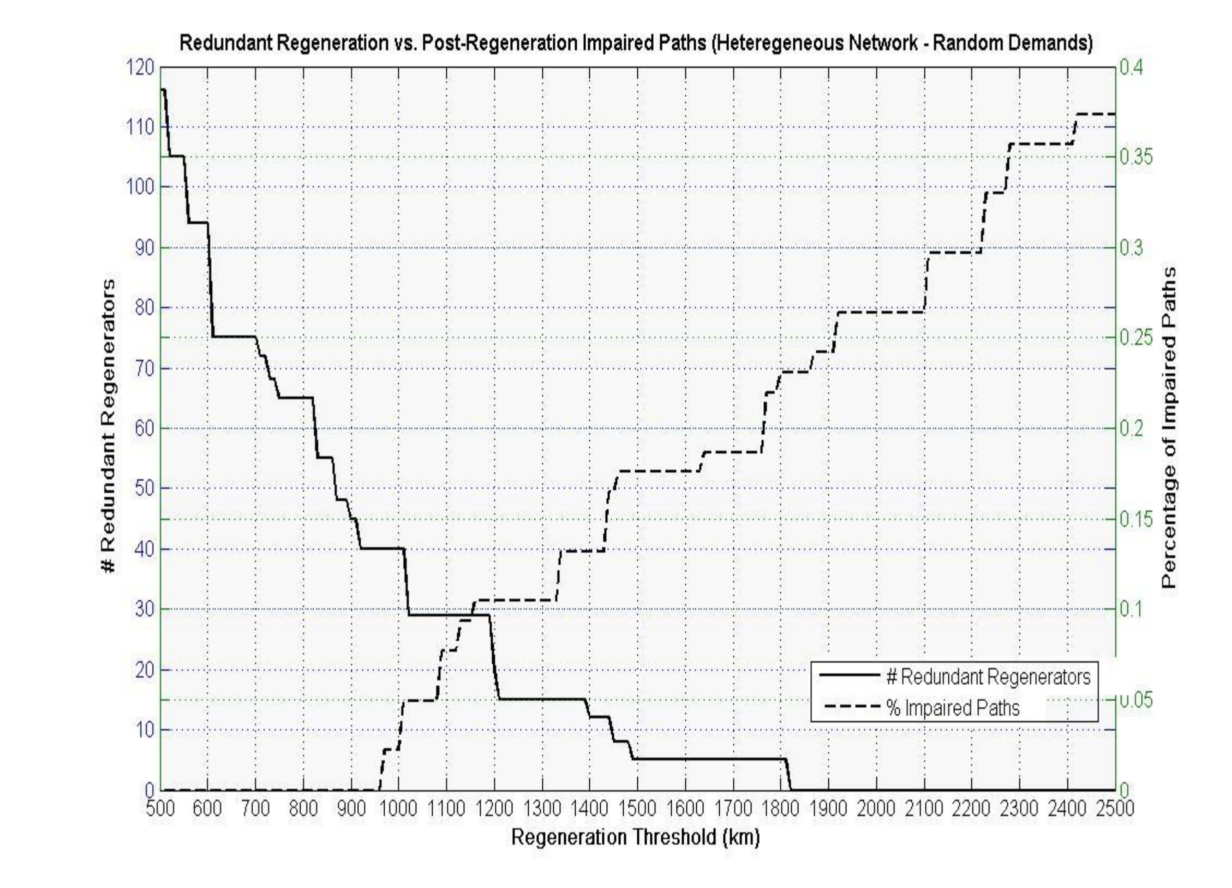
Conclusions wrt dimensioning

- Proposal of dimensioning approach
  - Sequential approach: scalability
  - Combination of analytics and simulation
  - Correlation between dimensioning and scheduling

- Specific dimensioning studies
  - Computational resources
  - Data consolidation (computational & storage resources)
  - Impairment-aware network design
  - Studies related to Optical Burst Switching

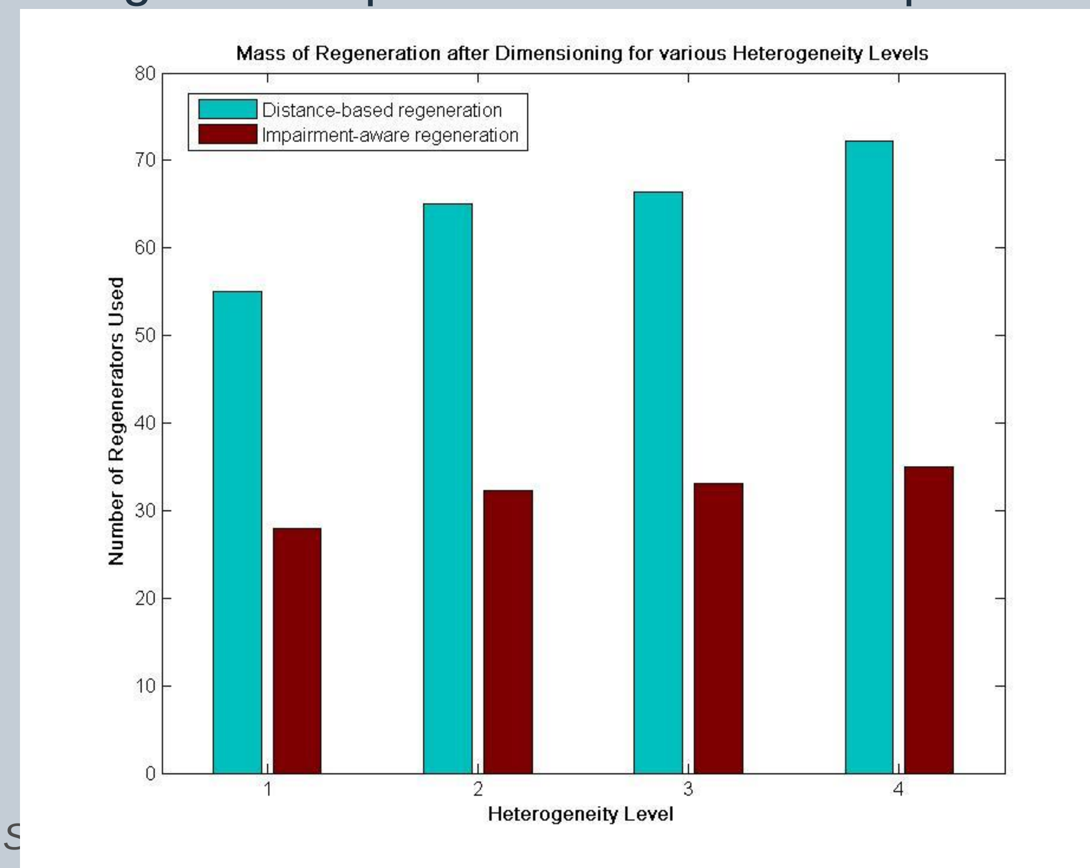
Impairment-Aware Design: Context

- Impairment-aware (IA) design of Grid optical networks
  - Link selection
  - Dimensioning of: fibres per link, wavelengths per fibre, switch sizes
  - In addition: place regenerators at design time to rectify signal over impaired connections



Impairment-Aware Design: Results

- Network dimensioning: optimal solution using integer programming
- Regenerator placement:
  - Based on analytical calculation of BER across candidate lightpaths
  - Integrated into the integer program
- Comparison with regenerator placement based on a predefined optical reach value



RESILIENT GRID NETWORKS

Resilient Grid Networks

- **Goal**: Protection and restoration techniques for failures **in network, resources, or both**.
- Network Resilience
  - Path Provisioning under Multiple Failures
  - Resilient Grid network design
  - Resilient physical-constraints-aware routing
  - Differentiated Resilience with Dynamic Traffic Grooming for WDM Mesh Networks
  - Differentiated Resilience for Anycast Flows
- Resource Resilience
  - Job Relocation
  - Joint Resilience
- Some sample results

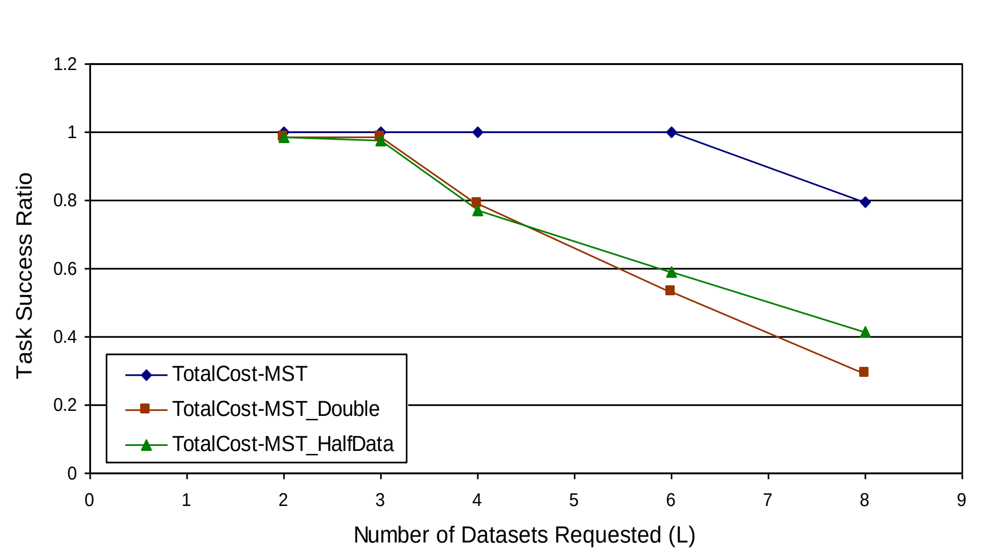
Data Consolidation and Resiliency

- Data Consolidation
  - Combine data from multiple sites at a processing site
- Combination of Data Consolidation schemes with resiliency techniques:
  - Double Site: select two Data Consolidation sites, the first and the second “best”, according to the corresponding DC scheme used and transfer the task’s data to both sites.
  - Half Data: again select in the same way two DC sites, however in the second-“best” site we transfer only half of the data needed by the task.
- We proposed the TotalCost_MST Data Consolidation scheme:
  - Selects the data replicas and the data consolidation site similarly to the TotalCost algorithm
  - Routing uses a Minimum Spanning Tree (MST) instead of Shortest Path Tree (SPT).

Data Consolidation and Resiliency

- A task fails when no resource is found with sufficient free storage space where the task's datasets can consolidate.

Task success ratio

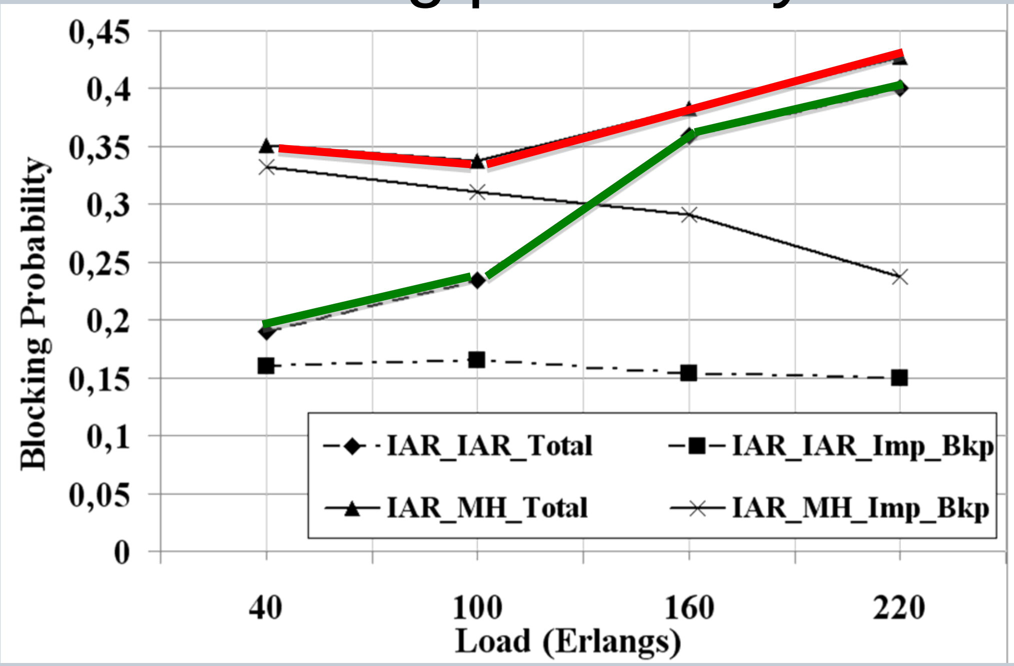


- The resiliency techniques applied increase the load in the network and as a result the task delay. This results in longer reservation times of the storage resources and to more task failures.
- TotalCost_MST algorithm: the resiliency methods use network resources more efficiently, leading to larger task success ratios than when other DC schemes are used.

Impairment-aware Resilient Routing

- Physical impairments considered as a routing criterion in routing both working and protection paths
- Tested in the Shared Backup Protection Path (SBPP) scheme
- Evaluation against the approach that maximizes resource sharing among backup paths

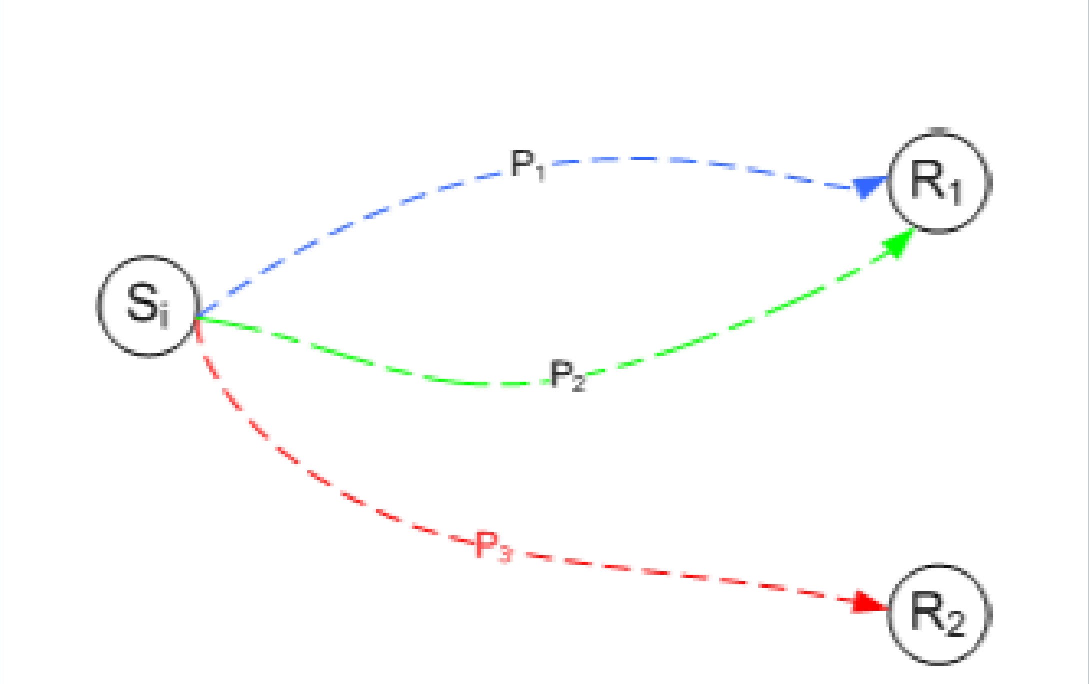
Blocking probability



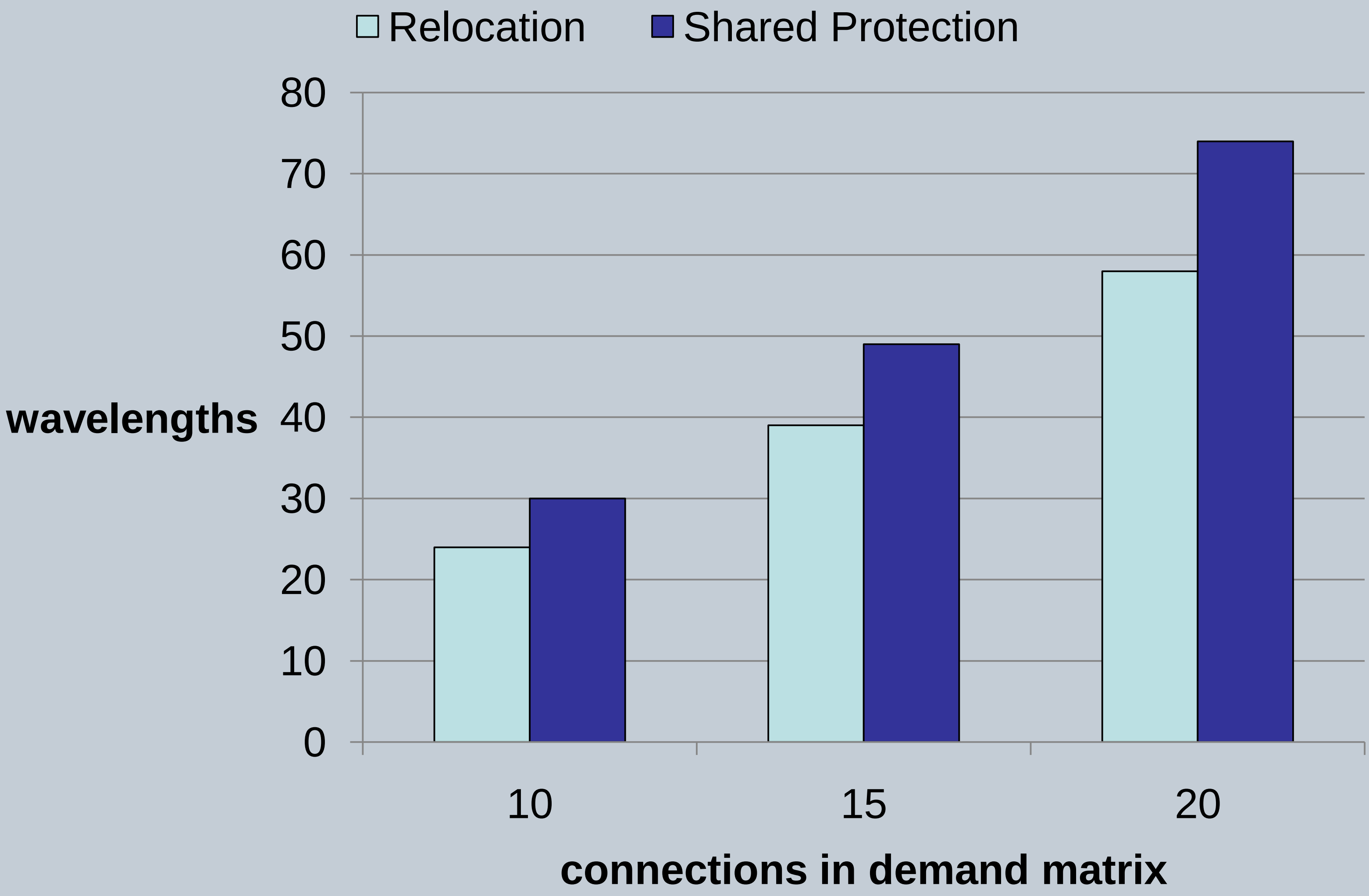
- IAR for primary paths
- IAR or minimum hop for backup paths
- IA-routing at both primary and protection paths lowers total blocking
- Benefit diminishes for higher loads, still remained more efficient

Exploiting relocation to improve network utilization

- Given
  - Network topology, job arrival rates
- Find
  - Primary path p_1 and secondary path p_2 to primary resource r_1
  - Secondary path p_3 to secondary resource r_3
- Trade-off
  - Dedicated vs shared
  - Network vs resource cost
- ILP formulation



Reduction of network dimensions by relocation



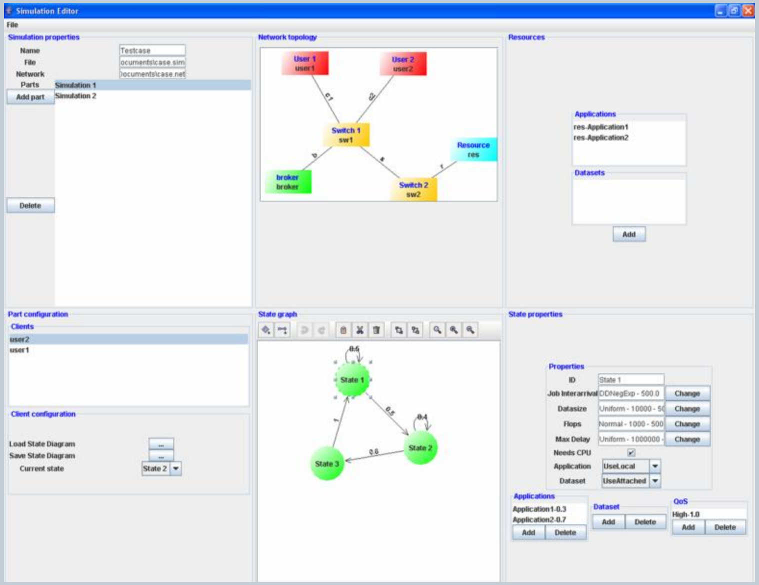
SIMULATION ENVIRONMENTS

Grid Simulation Environment

- Basic framework developed by IBBT
- Extensions implemented by other partners (AIT, CTI, UniBonn, ULeeds)
- Features
  - Java, no dependencies, discrete event
  - Modeling network and Grid resources
  - Dynamic OCS & OBS path set-up and tear-down
  - Flexible job models (based on Markov states)
  - GUI to define network topology and traffic models
- GPL license

- **Topology**
  - Job sources
  - Switches
  - Resource broker
  - CPU/storage resource

- **Job model**
  - Multiple (Markov) states
  - Transition probabilities
  - Given Job IAT and job size distribution in each state



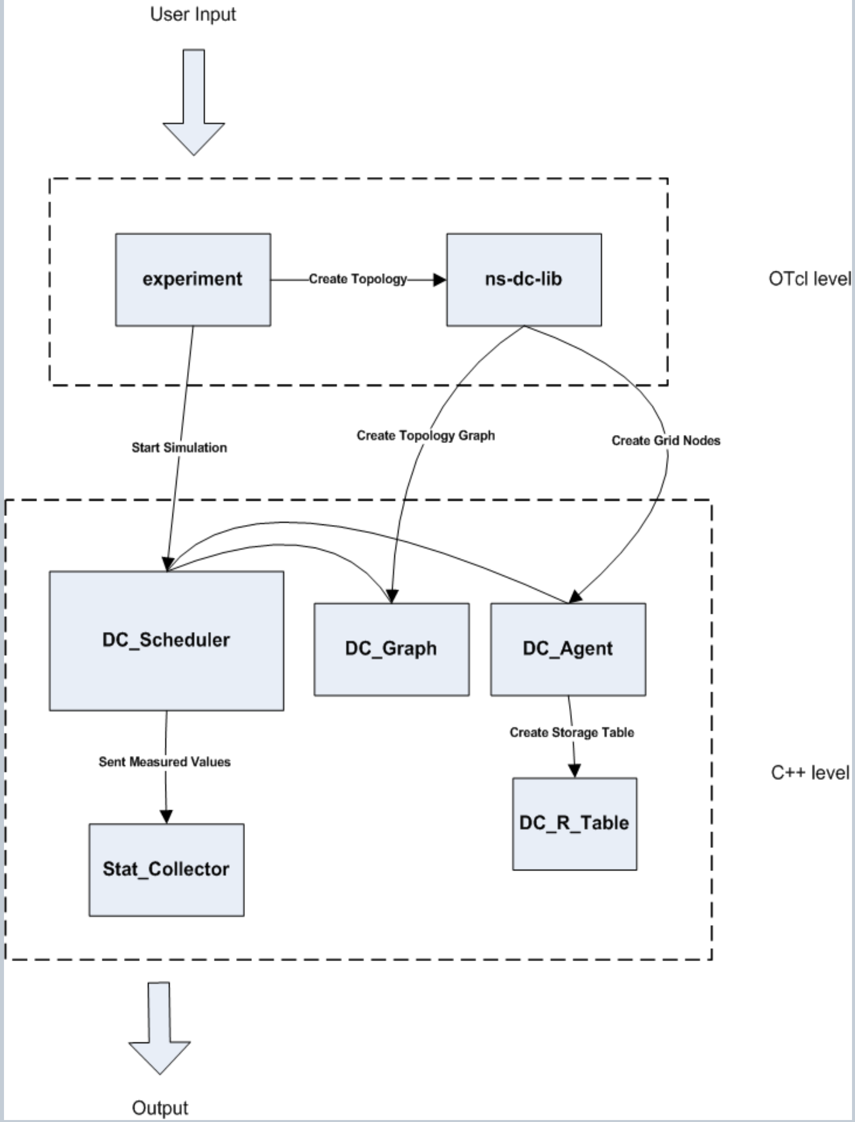
GridNs Module

- **Dimensioning and Fault Tolerance Simulation Studies for Data-Intensive Applications**

- Based on Network Simulator 2 (NS-2)
  - NS-2 simulates a large number of network-related parameters and characteristics

- Extensions for Grid:
  - Computational and storage resources
  - Data Consolidation algorithms

- GPL license



A Simulator for Examining Differentiated Resilience for Anycast Flows

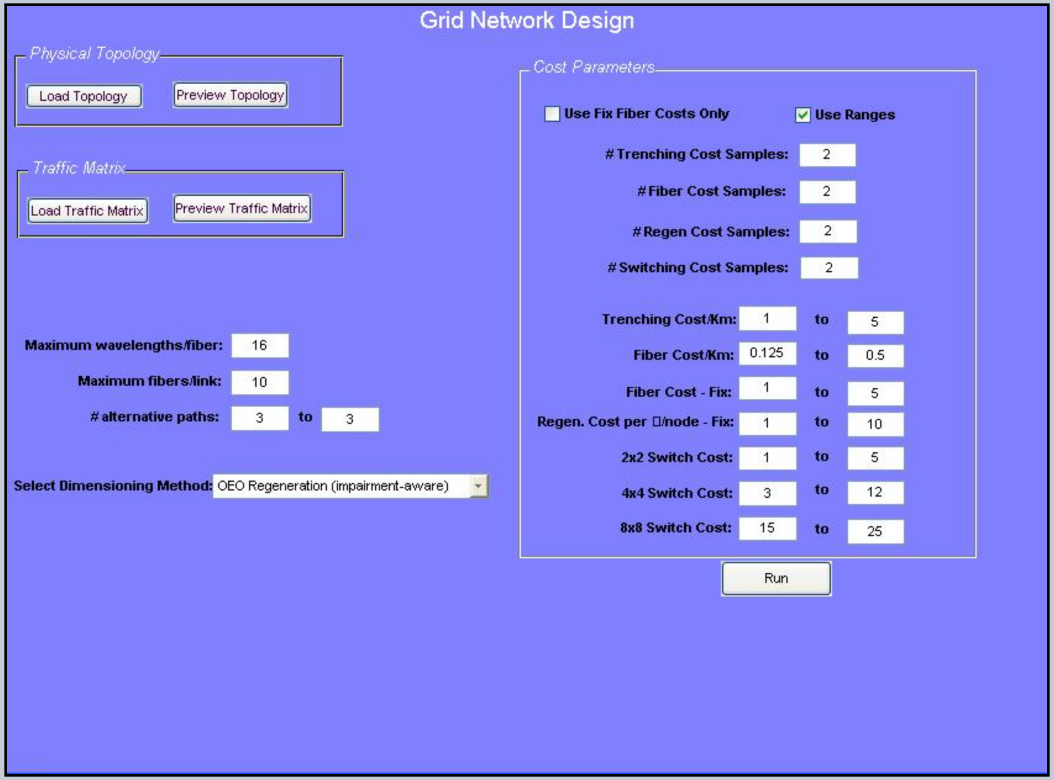
- Anycast request can be served by any suitable replica server
- Modular design

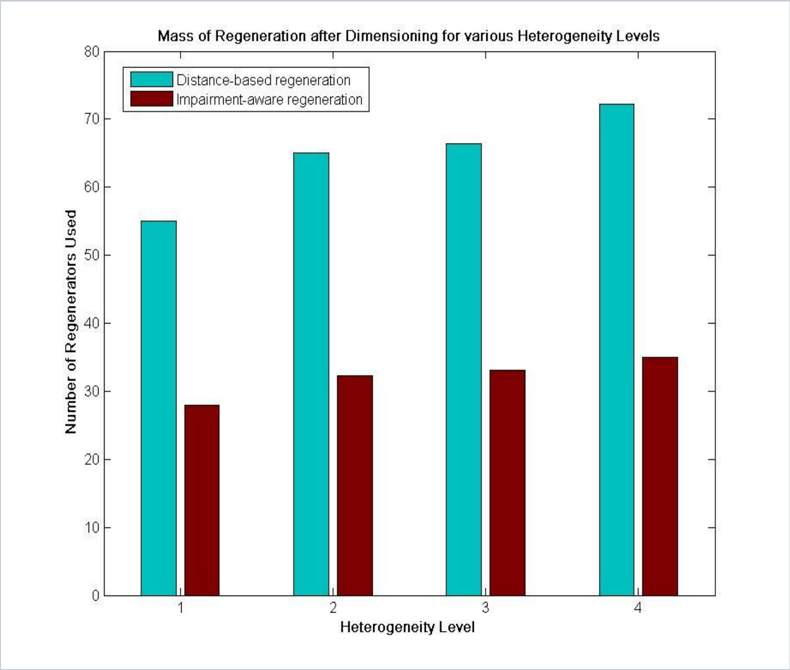
- Routing algorithms:
  - Constraint Shortest Path First (CSPF) algorithm
  - Least Interference Optimization Algorithm (LIOA)

- Server selection algorithms
  - Hop Number Server (HNS)
  - Residual Capacity Server (RCS)
  - Hop Number Widest Server (HNSW)

Node simulation cycle

NeDeTo – Network Design Tool





- Minimum-cost **WDM Network Dimensioning** using Integer Linear Programming
- **Jointly with Regenerator Placement (RP)**
- Three RP approaches implemented:
  - **No Regeneration (benchmark)**
  - **Length-based Regeneration**
  - **Impairment-aware Regeneration**
  - Tool accepts user specified input topology, traffic matrix (input files) and costs
  - Output: various evaluation statistics (e.g. total network cost, #regenerators etc.)

Questions?