Grid/Cloud Computing over Optical Networks

- Opportunities & Research Issues

Chunming Qiao

Lab for Advanced Network Design, Evaluation and Research (LANDER)

University at Buffalo (SUNY)

OFCNFOEC 2009

THE FUTURE OF OPTICAL COMMUNICATIONS IS HERE

Outline

- Optical Grid Computing for Petascale Science
- Federated Computing and Networking as Next Generation Cloud Computing

Petascale Science

- Sharing of large amounts of data (in PB range) generated by big experiment instruments and observatories
- Supporting thousands of collaborators worldwide
- Distributed data processing
- Distributed simulation, visualization, and computational steering
- Distributed data management

Petascale Science

q1

Science Areas / Facilities	End2End Reliability	Connectivity	Today	5 years	Network Services
Advanced Light Source	-	• DOE sites • US Universities • Industry	1 TB/day 300 Mbps	5 TB/day 1.5 Gbps	• Guaranteed bandwidth • PKI / Grid
Bioinformatics	-	• DOE sites • US Universities	625 Mbps	250 Gbps	• Guaranteed bandwidth • High-speed multicast
Chemistry / Combustion	-	• DOE sites • US Universities • Industry	-	10s of Gigabits per second	• Guaranteed bandwidth • PKI / Grid
Climate Science	-	• DOE sites • US Universities • International	-	5 PB per year 5 Gbps	• Guaranteed bandwidth • PKI / Grid
High Energy Physics (LHC)	99.95+% (Less than 4 hrs/year)	• US Tier1 (DOE) • US Universities • International	10 Gbps	100 Gbps (30-40 Gbps per US Tier1)	• Guaranteed bandwidth • Traffic isolation • PKI / Grid

Current, Near- and Long-term Requirements

Science Areas	Today End2End Throughput	5 years End2End	5-10 Years End2End	Remarks
High Energy Nuclear Physics	10 Gb/s	100 Gb/s	1000 Gb/s	high bulk throughput and sporadic
Climate (Data & Computation)	0.5 Gb/s	160-200 Gb/s	N x 1000 Gb/s	high bulk throughput
Genomics (Data & Computation)	0.091 Gb/s (1 TB/day)	100s of users	1000 Gb/s + QoS for control	high throughput and steering
SNS NanoScience	Not yet started	1 Gb/s	1000 Gb/s + QoS for control	remote control and time critical throughput
Fusion Energy	0.066 Gb/s (500 MB/s burst)	0.198 Gb/s (500MB/20 sec.)	N x 1000 Gb/s	time critical throughput
Astrophysics	0.013 Gb/s (1 TB/week)	N*N multicast	1000 Gb/s	computational steering and collaborations

A Composable Data Transfer Framework

- Dynamic reconfiguration capabilities
  - to support different objectives such as burst, scheduled, and streaming delivery
- Automatic detection of scenarios and use of appropriate/ available
  - transport media (e.g., (circuit-based WDM, VLANs, SONET, etc), and
  - protocols, such as (TCP-variants, UDP-variants, InfiniBand, SCSI, etc.)
- Capability of one-to-many, and many-to-many data transfers,
  - via Application Level Multicast or peer-to-peer approach

Federated Computing and Networking (FCN)

- A FCN system consists of computing facilities (e.g., clusters, data centers) interconnected with wide-area WDM networks
- A FCN ***service provider*** uses its own computing and WDM networking resources (or resources that belong to a third party for which it is a broker)
- FCN: the next generation of Cloud Computing
  - Interact directly with the WDM networks
  - Integrate a larger scale of computing and networking resources
  - Provide stronger Service Level Agreements (SLAs) including high availability and robustness than e.g., Amazon's EC2

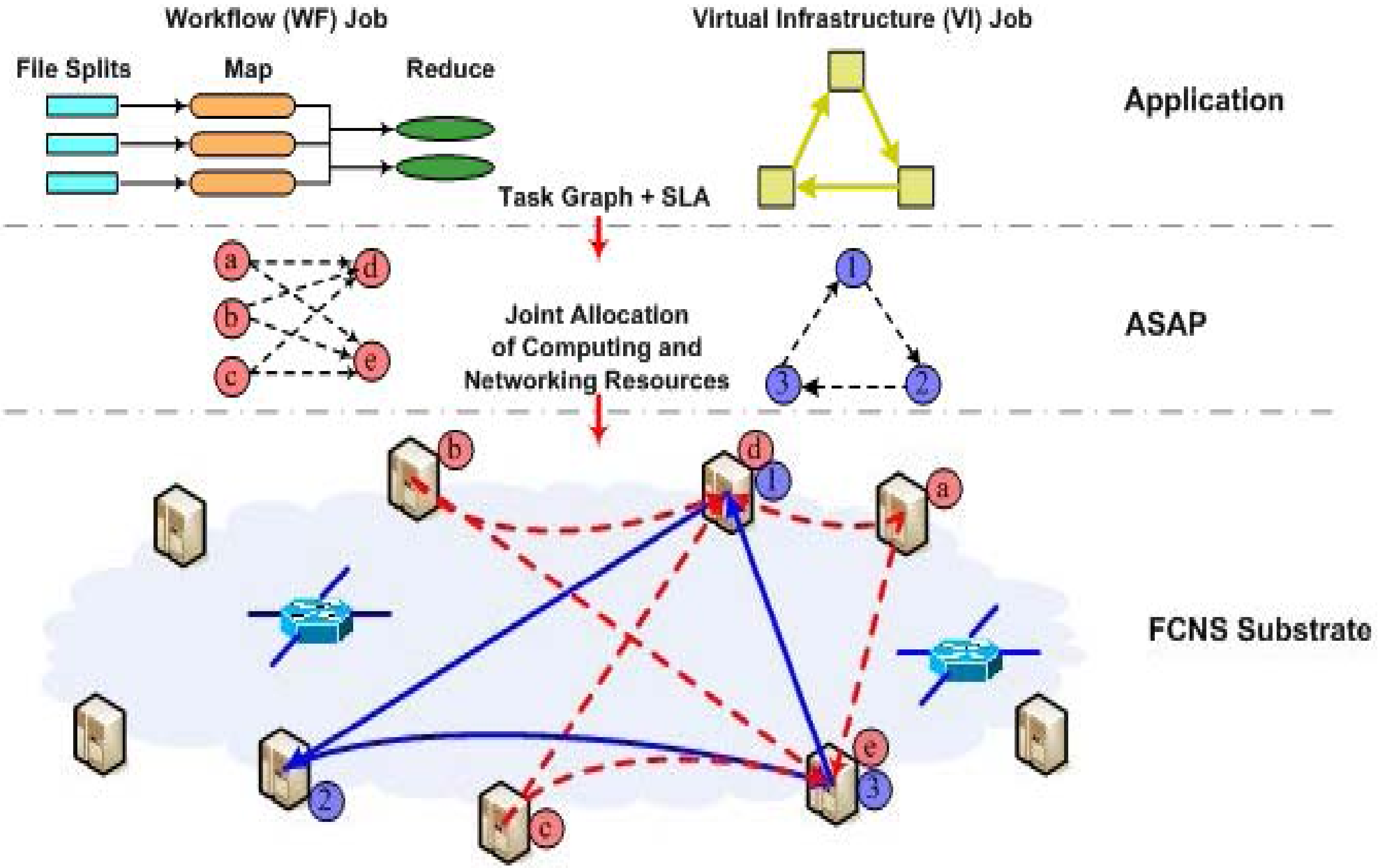
VI Job & WF Job

- Two general types of distributed jobs / apps
- **<u>Virtual Infrastructure (VI)</u>**
  - specifies a set of computing resources (e.g., processing clusters), and their connectivity (in terms of topology, bandwidth, and delay) for a specific period of time
  - Typically represented using a general directed graph
- **<u>Workflow (WF)</u>**
  - involves large data sets to be distributed among many sites
  - Represented using a directed acyclic graph, or DAG, where directed edges imply precedence among the tasks

Support VI/WF Jobs in FCNS: the ASAP Platform

- **<u>Provision Application-Specific, Agile, and Private (ASAP) platform</u>**
  - Given: a VI or WF job request,
  - Determine: the mapping of the tasks to computing facilities, and the routes as well as wavelengths to be used for connections over the WDM networks,
  - Objective: to satisfy the job's requirements with some optimization goals

Illustration of FCNS



Example Research Issues

- Advanced Network Provisioning Technologies
  - enable dynamic, multi-layer, end-to-end, circuit-based services across federated networks
  - Extensions of existing control plane technologies such as (GMPLS, MPLS, etc.) to accommodate scheduling, and reservation
  - unified control plane technologies, path computations, and traffic engineering for multi-layer and multi-domain networks offering hybrid best-effort IP, burst and switched circuit services

Example Research Issue II

- **<u>Resource co-scheduling to improve data transfer or data analysis job performance</u>**:
  - Offline/online provisioning of data transfer request(s)
    - Optimal co-scheduling of computing resources (e.g., storage/caching) and network resources
  - Offline/online provisioning of data analysis job(s)
    - Decide the execution host(s) for the job(s), and establish network paths to stage missing input files locally

Example Research Issue III

- Fault Diagnosis and Tolerance
  - Dynamic performance monitoring over heterogeneous multi-domain networks
  - Fault location and diagnosis
  - Protection/Restoration approaches to survive various failure scenarios
  - Proactive replication to increase the availability of data
  - Network coding to reduce storage and bandwidth requirements

Research Issue IV

- SLA-driven, cost-effective algorithms for provisioning ASAP platforms,
  - addressing the optimal joint task assignment & scheduling and lightpath establishment (as well as traffic grooming) problems,
  - subject to heterogeneous computing resources and limited optical networking resources
- Robust and resilient approaches to survivable ASAP platforms
  - considering tradeoffs involving SLA guarantee and resource usage, under various failure scenarios

Previous Results

- **"Performance Comparison of Optical Circuit and Burst Switching for Distributed Computing Applications" -** OFC 2008

- **"Survivable Optical Grids" -** OFC 2008

- **"Task Scheduling and Lightpath Establishment in Optical Grids" -** INFOCOM 2008 Mini-Conference

Recent Works

- Maximizing the Revenues for Distributed Computing Applications over WDM Networks - OFC 2009, OMG2
- **"Survivable Logical Topology Design for Distributed Computing in WDM Networks" -** OFC 2009 OMO3
- **"Robust Application Specific and Agile Private (ASAP) Networks Withstanding Multi-layer Failures" -** OFC 2009 OWY 1 (Wed)
- **"Online Job Provisioning for Large Scale Science Experiments over an Optical Grid Infrastructure" -** HSN 2009 in conjunction with INFOCOM 2009
- **"Application-Specific, Agile and Private (ASAP) Platforms for Federated Computing Services over WDM Networks -** in INFOCOM 2009 Mini-Conference

Thank you!