

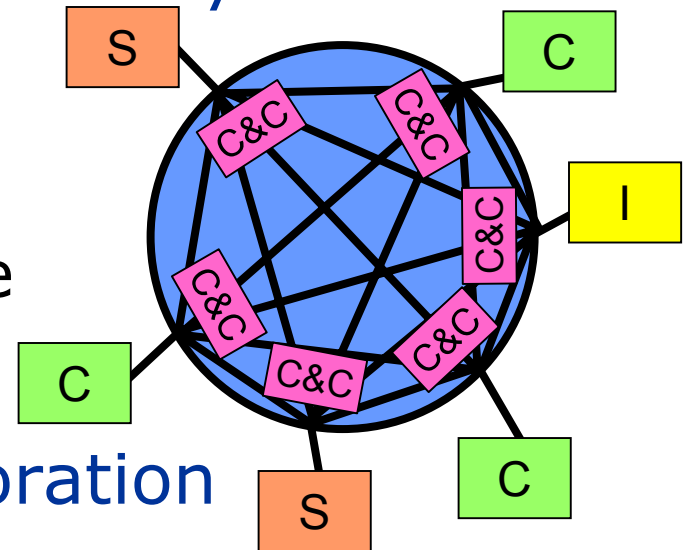
Breakout Group 4

- Charge: “Turn application requirements into middleware capabilities/research needs”

Ian Foster, Dennis Gannon, and numerous participants ...

What Do DOE Scientists Need from Networks?

- Fabric: fast, functional connectivity
 - ◆ High-performance links
 - ◆ Some QoS support
 - ◆ Network cache and compute
 - ◆ Instrumentation
- Services: support for collaboration
 - ◆ Overcome barriers of distribution when accessing information, resources, people
- *Majority of application requirements are concerned with services not fabric*



The Role of Middleware: Enabling New Science

- Translate potential of fast, functional networks into actual scientific progress by enabling easier, faster
 - ◆ Access to, and integration of, remote information, computers, software, and/or experimental devices
 - ◆ Inter-personal communication

What is Middleware?

- “Middleware” is a very broad term!
 - ◆ “Centralized/P2P” services: e.g., directories
 - ◆ “Community” services: e.g., membership
 - ◆ “Resource/site” services: remote access
 - ◆ “Application” libraries, e.g., data transfer
- We assume that focus here on middleware that needs to sit “in the network”: e.g., at centralized locations or at resource sites
 - ◆ Less focus on code that gets linked with apps, although this is also very important

The Middleware Version of the End-to-End Problem

- Middleware capabilities (authentication, discovery, access, transport, ...) need to be delivered end-to-end
- Thus numerous site integration issues, e.g.
 - ◆ Security infrastructures: federated identity
 - ◆ Policy: discovery, pre-emption, ...
 - ◆ Computer schedulers: on-demand access
 - ◆ Parallel file systems for end-to-end transport
- Underestimated in most planning of middleware R&D and deployment

Breakout Group Approach

- Review, expand, and reconcile application requirements
 - ◆ Use material provided by groups 1-3 as a starting point
- For each, discuss from the perspective of
 - ◆ R&D requirements
 - ◆ Deployment and support requirements
- Attempt to prioritize by identifying most important areas

A Note Concerning Distributed Computing

- More to life than Data & Collaboration Grids
 - ◆ Remote access to code (e.g., fusion)
 - ◆ On-demand access to computers (e.g., for enhancement of instrumentation data)
 - ◆ Distributed multidisciplinary simulations
- Most requirements seem to be covered, anyway

What are Key Middleware Capabilities⁸ That Will Enable New Science?

- Secure control over who does what
 - ◆ Prerequisite for collaborative science
- Information integration and access
 - ◆ Enable discovery-driven science
- Coscheduling and QoS
 - ◆ Enable new modalities with real-time demands
- Effective network caching and computing
 - ◆ Enable new modalities based on large data
- Services to support collaborative work
 - ◆ Enable community activities
- Monitoring and problem diagnosis
 - ◆ Enable use of distributed resources

Secure Control over Who Does What

- Why?
 - ◆ Enable access to, and sharing of, remote resources: compute, data, code, devices
- What?
 - ◆ Discovery, negotiation, mapping of identity, authentication, authorization, accounting, auditing, community policies
 - ◆ Firewalls, NATs
 - ◆ Note also policy issues
- How?
 - ◆ Significant R&D and deployment issues

Information Integration and Access

- Why?

- ◆ Enable sharing of resources and information (data, accounting, events, provenance, ..., ...)
- ◆ Enable discovery-driven science

- What?

- ◆ Information integration => Discovery
 - Global name space; registries, information services, metadata, P2P, ...
- ◆ Information access => Reliable transport
 - Reliable transport as a service

- How?

- ◆ Significant R&D and deployment issues

Coscheduling and QoS

- Why?
 - ◆ Enable real-time running of experiments, interactive analysis, efficient distributed processing
- What?
 - ◆ QoS as an end-to-end issue—not primarily a network issue!
 - ◆ Brokers, reservation, allocation, monitoring, adaptation, ...
- How?
 - ◆ Significant R&D and deployment issues

Network Caching and Computing

- Why?
 - ◆ Enable new collaboration modalities, e.g., efficient collaborative science on very large datasets
- What?
 - ◆ Organization, policy, application services for use of such devices
- How?
 - ◆ Significant R&D and deployment issues

Services to Support Collaboration

- Why?
 - ◆ Enable basic collaboration modalities—
discovery, publication, authoring, ...
- What?
 - ◆ Per-site and per-resource: secure remote
access to resources, code, data, instruments
 - ◆ Centralized/P2P: majordomo, brokers, CA,
archival storage, replication, ..., ...
- How?
 - ◆ Significant deployment; some R&D

End-to-End Monitoring and Diagnosis

- Why?
 - ◆ Enable network-wide resource utilization
- What?
 - ◆ End-to-end, top-to-bottom monitoring and diagnosis so that sources of problems in distributed computations can be identified
 - ◆ Automated response to identified problems
- How?
 - ◆ Significant R&D; eventually deployment