Demand for more computing power

- There are three ways to improve performance:
  - Work smarter
  - Work harder
  - Get help

- In computing:
  - Using optimized algorithms and techniques
  - Using faster hardware
  - Using multiple computers

Cluster Computing

- A cluster is a type of parallel and distributed system, which consists of a collection of inter-connected stand-alone computers working together as a single integrated computing resource.

- Basic element is the node, a single or multiprocessor system with memory, I/O and OS
- Generally two or more nodes connected together
- In a single rack, or physically separated and connected via a LAN
- Appears as a single system to users and applications
- Specialized access, management and programming

Utility Computing History

1990

2010

Grid Computing

Solving large scale problems with parallel computing

Utility Computing

Offering computing resources as a metered service

Software as a Service

Network-based subscriptions to applications

Cloud Computing

Anytime anywhere access to resources delivered dynamically as a service

Grid Computing

- Problem:

  Scientific instruments and experiments provide huge amount of data

- Goal:

  Researchers perform their activities regardless geographical location, interact with colleagues, share and access data

- Solution:

  Networked data processing centers and ”middleware” software as the “glue” of resources.

Once upon a time...

Microcomputer

Minicomputer

Mainframe

Cluster

...up to the Grid

Why not just distributed?

- Distributed applications already exist!
  - But they tend to be specialised system
  - Single purpose
  - Single User Group

- Grids go further!
  - Different kinds of resources
  - Different kinds of interactions
  - Dynamic nature
  - Multiple institutions

Key Concept

ability to negotiate resource-sharing arrangements among a set of participating parties (providers and consumers) and then to use the resulting resource pool for some purpose

Grids in action

- High Energy Physics
  - European Data Grid
  - LHC Computing Grid
- Earth Observation
  - ESA EO Grid
  - Global Earth Observation Grid
- Bioinformatics
  - Genome Grid
- Mathematics
  - Zetagrid
- Geology
  - Earthquake Engineering Simulation
- Astronomy
  - SETI@home

Elements of Grid Computing

- Resource sharing
  - Computers, data, storage, sensors, networks, …
  - Sharing always conditional: issues of trust, policy, negotiation, payment, …
- Coordinated problem solving
  - Beyond client-server: distributed data analysis, computation, collaboration, …
- Dynamic, multi-institutional virtual organizations
  - Community overlays on classic org structures
  - Large or small, static or dynamic

Definitions

We define a **Grid** as a system that

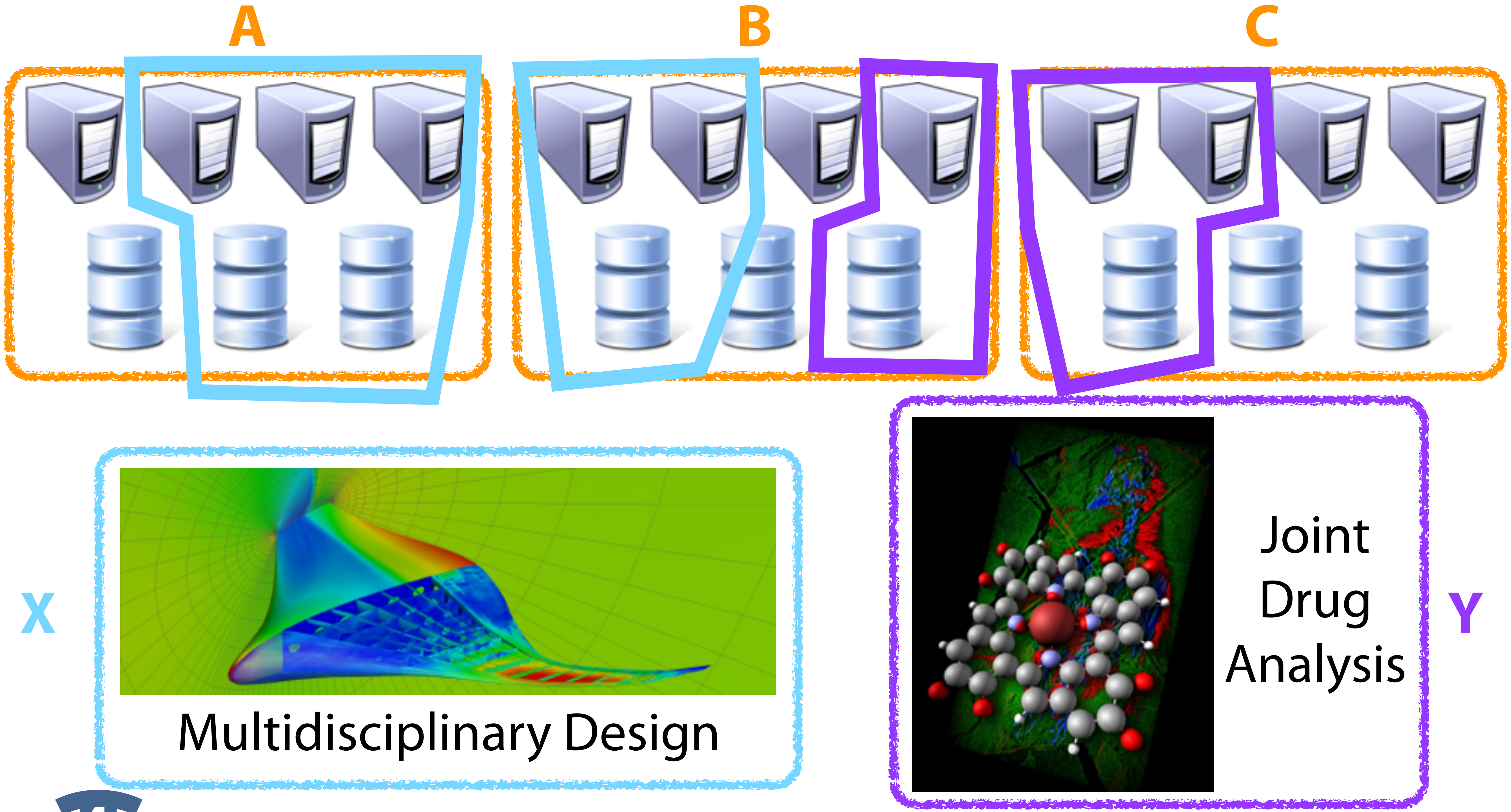
- **coordinates distributed resources**
  - integrating and coordinating resources and users that live within different control domains
  - addressing the issues of security, policy, payment, membership, and so forth that arise in these settings
  - Otherwise, we are dealing with a local management system.
- **using standard, open, general-purpose protocols and interfaces**
  - built from multipurpose protocols and interfaces that address such fundamental issues as authentication, authorization, resource discovery, and resource access.
  - Otherwise, we are dealing with an application-specific system.
- **to deliver nontrivial qualities of service**
  - resources to be used in a coordinated fashion to deliver various qualities of service (e.g., response time, throughput, availability, and security)
  - coallocation of multiple resource types to meet complex user demands, so that the utility of the combined system is significantly greater than that of the sum of its parts

Virtual Organizations

- Components
  - set of individual/institutions
  - set of resources
  - set of sharing rules
- Dynamic set of individuals and/or institutions defined by a shared goal and a set of sharing rules
- May vary in size, scope, duration and structure
  - Example: class students for cooperative lecture writing
  - Example: industrial consortium building a new aircraft
- The sharing is highly controlled, with resource providers and consumers defining clearly and carefully just what is shared

Example of VOs

- Three physical organizations (A, B, C)
- Two virtual organizations (X, Y)



A

B

C

X

Multidisciplinary Design

Joint Drug Analysis

Y

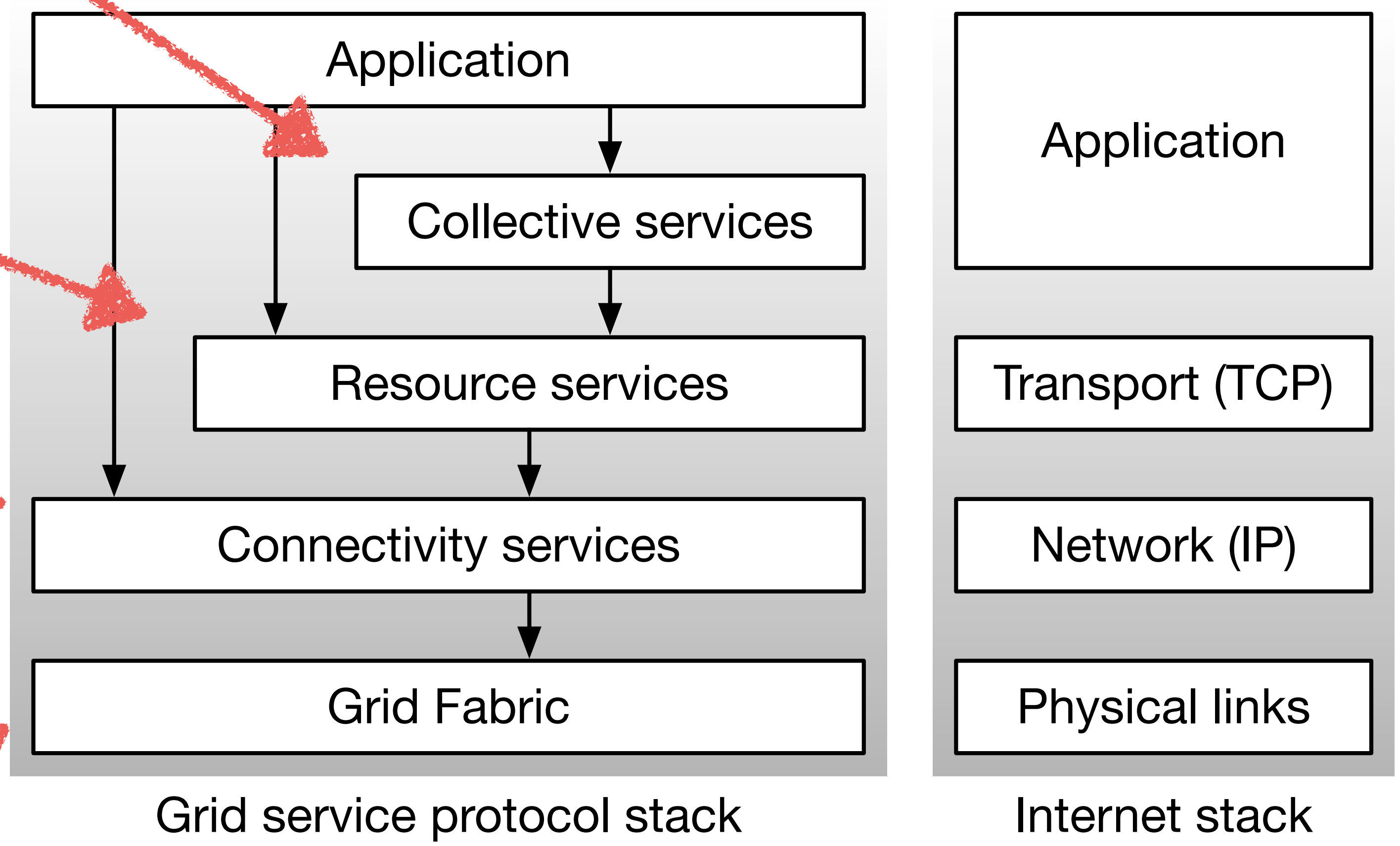
Grid Architecture

“Coordinating multiple resources”:
Ubiquitous infrastructure services,
application-specific distributed services

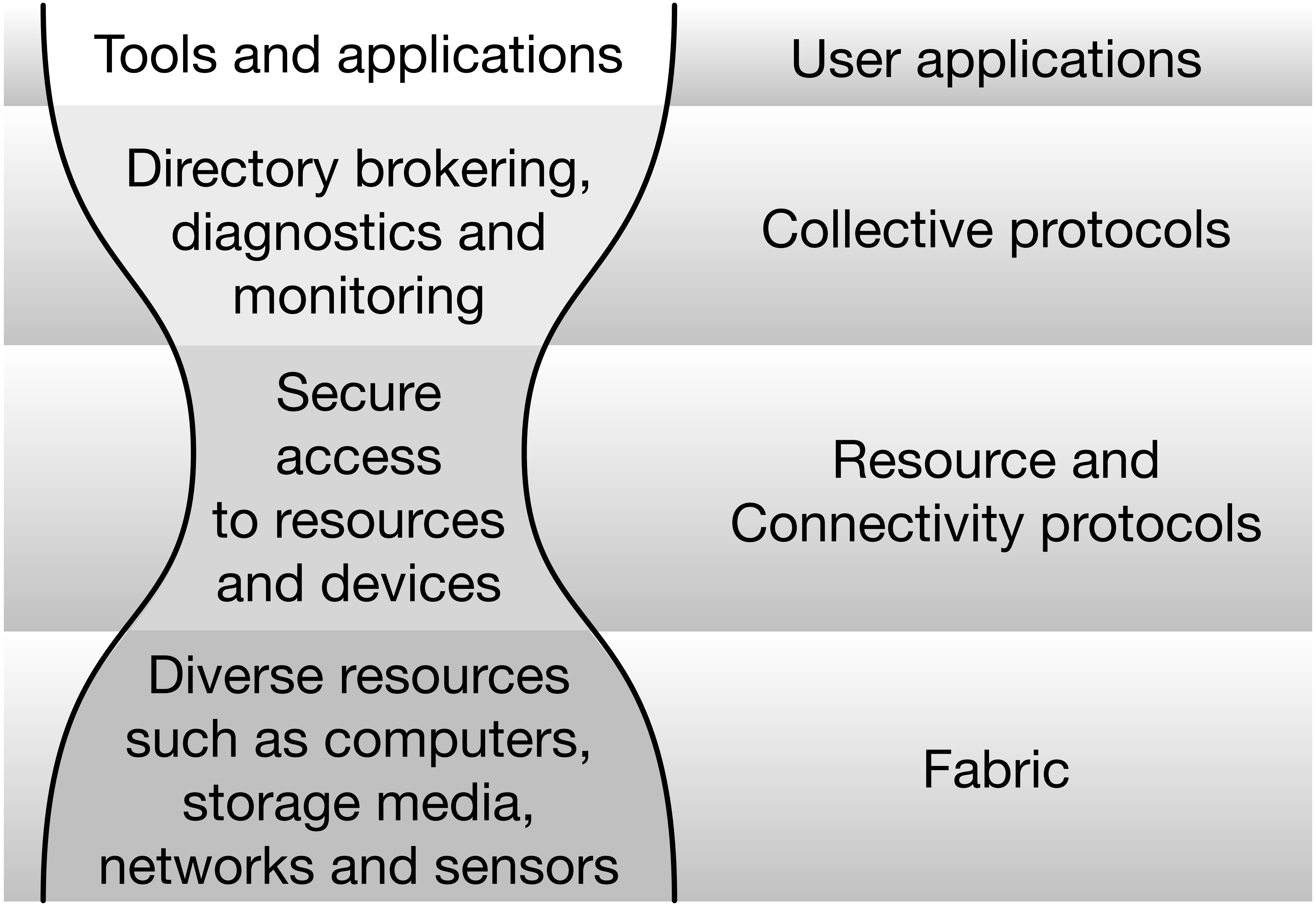
“Sharing single resources”:
Negotiating access, controlling use

“Talking to things”:
Communication (Internet protocols)
& security

“Controlling things locally”:
Access to & control of resources



The Hourglass Model



Tools and applications	User applications
Directory brokering, diagnostics and monitoring	Collective protocols
Secure access to resources and devices	Resource and Connectivity protocols
Diverse resources such as computers, storage media, networks and sensors	Fabric

Fabric Layer

- Just what you would expect: the diverse mix of resources that may be shared
  - Individual computers, Condor pools, file systems, archives, metadata catalogs, networks, sensors, etc.
- Few constraints on low-level technology: connectivity and resource level protocols form the “neck in the hourglass”
- Defined by interfaces not physical characteristics

Connectivity Layer

- Communication
  - Internet protocols: IP, DNS, routing, etc.
- Security: Grid Security Infrastructure (GSI)
  - Uniform authentication, authorization, and message protection mechanisms in multi-institutional setting
  - Single sign-on, delegation, identity mapping
  - Public key technology, SSL, X.509, GSS-API
  - Supporting infrastructure: Certificate Authorities, certificate & key management, …

Resource Layer

- Grid Resource Allocation Management (GRAM)
  - Remote allocation, reservation, monitoring, control of compute resources
- GridFTP protocol (FTP extensions)
  - High-performance data access & transport
- Grid Resource Information Service (GRIS)
  - Access to structure & state information
- Others emerging: Catalog access, code repository access, accounting, etc.
- All built on connectivity layer: GSI & IP

Collective Layer

- Index servers a.k.a. meta-directory services
  - Custom views on dynamic resource collections assembled by a community
- Resource brokers
  - Resource discovery and allocation
- Replica catalogs
- Replication services
- Co-reservation and co-allocation services
- Workflow management services
- etc…