Introduction to Grid Computing

Dr Aidan Murphy

School of Computer Science and Informatics
University College Dublin
Ireland

Recall

Main differences between symmetric cryptographic and public-key cryptographic algorithms:

- Symmetric encryption uses a single key that needs to be shared among the people who need to receive the message while asymmetrical encryption uses a pair of public key and a private key to encrypt and decrypt messages when communicating.
- Symmetric encryption is an old technique while asymmetric encryption is relatively new.
- Asymmetric encryption was introduced to complement the inherent problem of the need to share the key in symmetrical encryption model, eliminating the need to share the key by using a pair of public-private keys.
- Asymmetric encryption takes relatively more time than the symmetric encryption.

What is the “Grid”?

- Based in the idea of computing resources being consumed like electricity is from the power grid.
- Resources supported by heterogeneous computer hardware, operating systems, programming languages and applications.
- Management required to co-ordinate the use of resources.
  - Are clients getting what they need?
  - Can services afford to allow access?
- Security also an important consideration.

What is the “Grid”?

- Middleware designed to enable the sharing of resources on a very large scale: files, computers, software, data and sensors.
- Typically shared by groups of users in different organisations who are collaborating to solve a problem:
  - via sharing of data
  - via sharing of computing power
- Distributed system that is loosely coupled, heterogeneous, and geographically dispersed.

Example: World Wide Telescope

- The World Wide Telescope is a project that demonstrates the type of data-intensive application that the Grid is most useful for.
- Project owned by Microsoft Research to deploy data resources that are shared by the astronomy community.
  - Archives of observations covering a particular time period, part of the electromagnetic spectrum and part of the sky.
  - Made by different instruments distributed worldwide.

Example: World Wide Telescope

- Each team's data is gathered locally (order of terabytes and growing exponentially).
- As data is gathered, it is analysed and stored as derived data for use by others: common data labelling/format required.
  - Metadata to describe when it was collected, where it was collected and the instrument used.
  - Derived data specifies parameters used when filtering from the raw data.

Example: World Wide Telescope

- Calculation of derived data requires heavy computational support.
- Frequently has to be recalculated as techniques improve.
- Considerable expense to the data owner.
- Aim of the project is to unify the world's astronomy archives into a giant database containing astronomy literature, images, raw data, derived datasets and simulation data.

Requirements of a Grid

1) Remote Access to Resources (information required from the archives).
2) Processing where it is stored and managed: Typical query involves a small quantity of data from one or more massive archives.
3) Dynamic Service Creation: Resource manager of an archive should create service instances dynamically to deal with particular sections of data that are required.

Requirements of a Grid

4) Metadata: to describe
   - characteristics of the data in the archive (e.g. area of the sky, date, time, instrument used).
   - characteristics of the service (e.g. cost to use, location, publisher, load, space available).

5) Directory Services: to find services based on the above metadata.

6) Management Software: to manage queries, data transfers, reservation of resources. Should take into account that resources may need to be rationed.

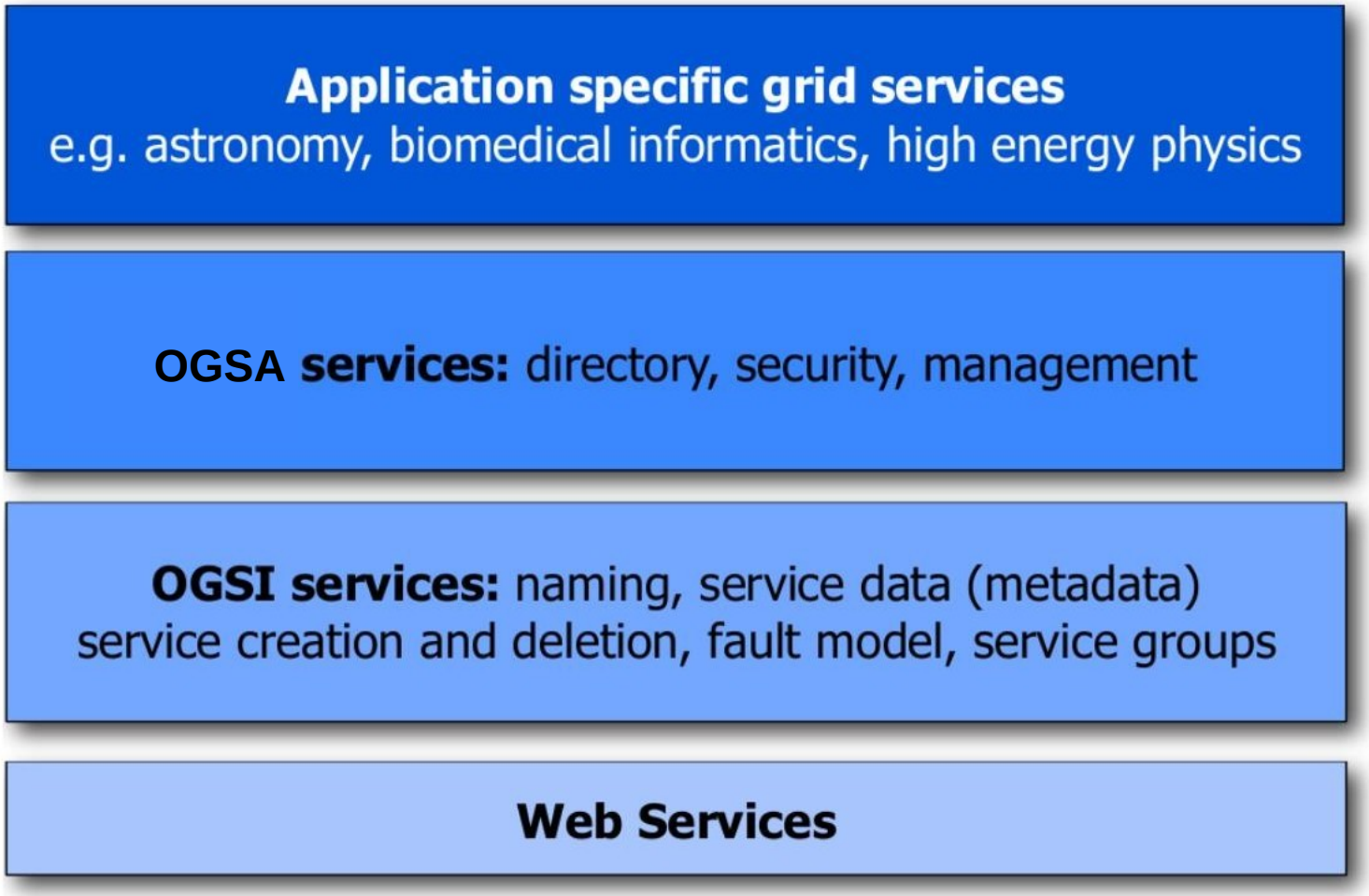
Requirements of a Grid

- Web Services can deal with the first two requirements (remote access/local processing).
- Requires that each application provides a service description that includes a set of methods for accessing the data.
- Grid middleware deals with the rest of the requirements (including security, if relevant).
- Note: World Wide Telescope is data-intensive. Grids also used for computationally-intensive tasks also (e.g. image analysis)

Open Grid Services Architecture

- OGSA is a standard for grid-based applications.
- Framework within which the requirements above may be met.
- Based primarily on web services.
- Resources managed by application-specific grid services.

Open Grid Services Architecture



application specific interfaces

standard grid service interfaces e.g. **GridService** Factory

Application level grid services

1. Web services that implement standard grid-service interfaces in addition to their application-specific interfaces.
   a) interface to a set of data (service data) that contains metadata about the service (e.g. recent results, average values, characteristics of the data or other service characteristics).
   b) the context in which a service runs must provide the ability to create, start and stop service instances. Relies on the naming facilities of the OGSI layer

Application-level grid services

2. Make use of standard grid services: provided as two layers (OGSI and OGSA).

a) OGSA Layer includes:

- Directory service: allows clients to select suitable services based on metadata collected from services instances that have registered with the directory service.
- Management service: monitors services, deal with failures and control service instance lifetimes.
- Security services: provide single logins, delegation as well as authentication and encryption.

Application-level grid services

b) Open Grid Services Infrastructure (OGSI) layer includes:

- Implementation of naming scheme for service instances.
- Definition of standard service data elements that must be implemented by every application-level grid service instance, along with operations to get and set their values.
- Definitions of the interface to the factory to create new instances.
- Fault model for use by all grid services
- Notification services: enable services to publish information about service data and others to subscribe
- Service groups so services can cooperate

OGSI Layer

- Service instances are created dynamically.
- Two-level naming scheme:
  - Long-lived globally unique identifier Grid Service Handle (GSH). Starting a new instance in the same state later can re-use the GSH.
  - Grid Service Reference (GSR): refers to the location of the service (how and where it can be contacted).

OGSI Layer

- Service Data
  - Clients can request a service instance to return information about its current state (e.g. capacity, free space, errors, recent results, load)
  - Requires that each service instance must store this data and offer operations to access it.
  - Standard set of operations provided in the GridService interface: includes operations to access the names of other interfaces supported, names of service data elements, identity of the factory that created it, GSH/GSR, termination time.
  - Every grid service must implement GridService

OGSI Layer

- Fault Model
  - Common approach for reporting of faults.
  - Used by all OGSI-level services.
  - Recommended for application-level grid services.
  - XML scheme that requires at least a timestamp and the originating service to be specified.
  - Optional extras: fault description, fault, error code.

OGSA

- Higher level services built on OGSI services.
- Variety of OGSA services will meet specific requirements of application-level services.
- Some are so widely-applicable that they are relevant to any system.
  - e.g. directory services, management, monitoring, security.
  - Common part of grid toolkits

OGSA

- Management Services
  - Management of any resource that can be shared or exploited.
  - Arranging for services to be used and monitoring their state.
  - Issues include task submission, Quality of Service agreement, advance reservation.
    - May involve Service Level Agreements (SLAs), which make guarantees to clients as to the type of service being provided and allow the resource owner to maintain control over how it's used and how much information is exposed to the client.

Globus Toolkit

- Globus Project began in 1994 to provide software that integrates and standardises the functions required by a family of scientific applications (including directory services, security and resource management).
- First version of the toolkit appeared in 1997.
- OGSA came from second version.
- Major version is version 5 (2010)
- Latest release October 2014 (version 6.0)
- Maintained by Globus Alliance as an open source project (www.globus.org)

Summary

- The Globus Toolkit is the defacto standard for grid computing.
- It is founded on the Web Services Core.
- Four key services are built on top of this:
  - Execution Management,
  - Data Management
  - Security.
  - Common Runtime.
- Most current work on grid computing assumes the use of the Globus Toolkit.

Thank you