Web Services

•XML-based technology for distributed computing

• Web service = a software system designed to support interoperable Machine to Machine interaction over a network

•Web Services Description Language (WSDL), an XML document containing:

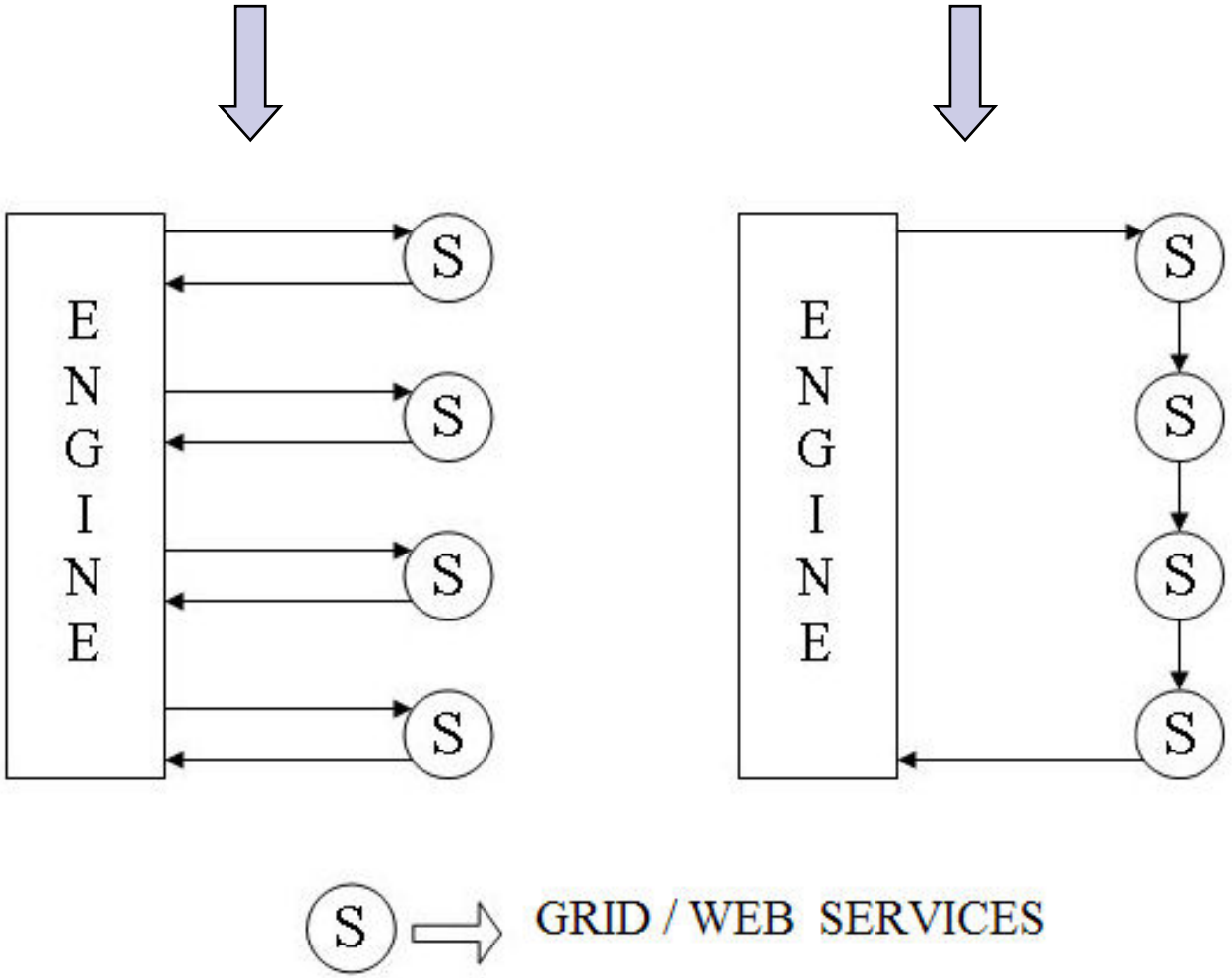
- Message types of the service & Message types to be returned & exceptions
- linked method as “port types”

• Using WSDL, it is possible to know how to access and use a service

Grid Services

- Grid Services extends Web Services to handle stateful and transient services.
- OGSA defines an open architecture for distirbuted services on the Grid.
- WSRF (Web Service Resource Framework) defines an infrastructure specifying the Grid service (WS-resources) defined by OGSA.
- Globus Toolkit 4 implements WSRF and its services and permits to implements distributed applications on computational Grids.

Orchestration and Choreografy



ProCKSI

- Limited scalability of ProCKSI
  - Collect all info to a central point
  - Time consuming methods
- Actual configuration permits to deal with about 400 proteins
- As of July 31, 2007, 41298 protein structure in PDB (cite Azhar)
  - About 850 million of pair wise comparison
  - We need Grid computational power and storage

Distributed ProCKSI Scenario

- A user want to build a **workflow**, composing of basic services, as preprocessing steps, protein comparisons, clustering, visualizing, etc..
- This application must work on a large sets of proteins, he needs a lot of resources (cluster/grid)
- ProCKSI should automatically run the workflow on a **dedicated** cluster or on computational grids
- Additional properties: failure recovery, checkpoints, scheduling using the best resources available

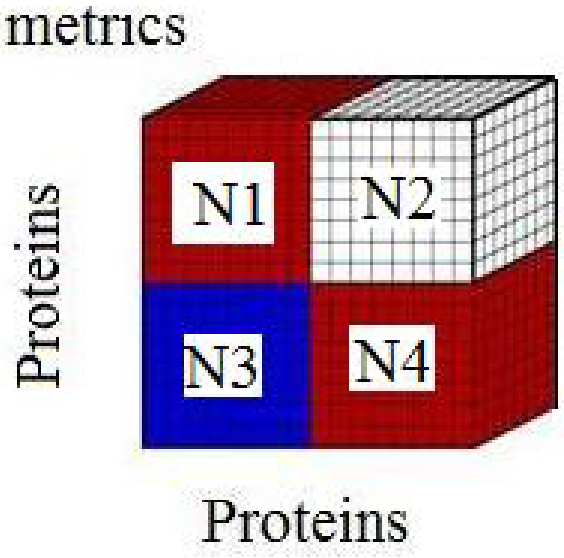
Distributed ProCKSI

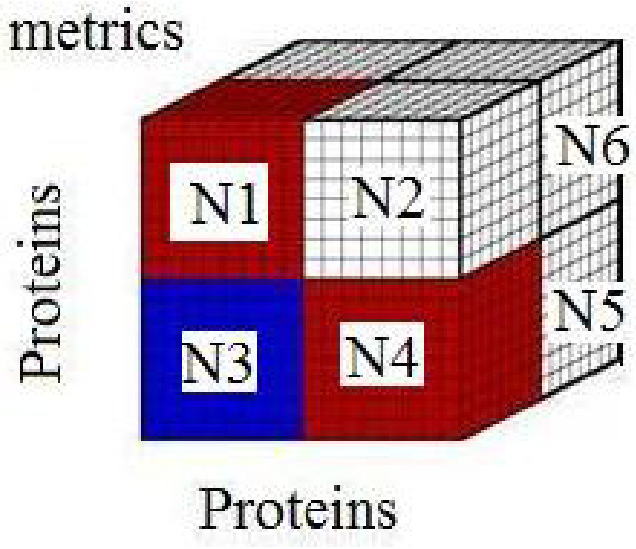
- Web/Grid Service implementation of methods as MaxCMO, USM, FAST, etc.., receiving two proteins (or to receiving a set of proteins or a matrix) and returning a similarity vector (maybe stateful and transient)
  - Most computing expansive (MaxCMO) could be run on clusters
- Web Service for computing Contact Map
- Distribution of data (moving model instead of data)
- Consensus methods more adapt to Grid
  - Data all too large to be collected in a central place (models could be collected)

Decompositon of the domain

- The ProCKSI tridimensional matrix (proteins, proteins, metrics)
- Consider a set of resources (nodes of the clusters or machines on the grid) N1, N2, N3…

- Actual decomposition decompose along the metrics axis





Issues in Distributed ProCKSI

- Grain of the Web service
  - I run a web service computing a single pair wise comparison or a web service computing a portion of matrix
- Scheduling tasks in presence of multiple users
- Asking more grid resources to university and external institutes
- Needs of local repositories or cache to maintain information
- Exploiting potentialities of GT4 for trasferring data (RFT and GridFTP) and (maybe) OGSA-DAI for replicating data

BPEL4WS

Using BPEL4WS for building user defined workflows

Business Process Execution Language for Web Services

- XML-based Language
- Composition of Web Service distributed on the Web
- Supports stateful Web Services and Grid Services
- Drag & drop based tools for generating BPEL code.
- Uses typical construct of programming languages (if, while,…) to define the flow of web services (workflow)