Grid Computing as an Integrating Force in Virtual Enterprises

by

Hongfei Tian
M.S., Software Technology (2002)
Stuttgart University of Applied Sciences

Submitted to Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Civil and Environmental Engineering
at the
Massachusetts Institute of Technology
June 2003

Signature of Author ....
Department of Civil and Environmental Engineering
May 15, 2002

Certified by ....
Amar Gupta
Sloan School of Management

Certified by ....
John R. Williams
Associate Professor of Civil and Environmental Engineering

Accepted by ....
Oral Buyukozturk
Chairman, Department Committee on Graduate Studies

Grid Computing as an Integrating Force in Virtual Enterprises

by

Hongfei Tian

Submitted to Department of Civil and Environmental Engineering on May 15, 2003 in partial fulfillment of the requirements for the Degree of Master of Engineering in Civil and Environmental Engineering

Abstract

Constrained economics, globalization, competition, alliance, outsourcing, and network economy, all force enterprise IT applications to be reviewed from a new perspective. IT managers are frequently asked if the enterprises already have enough IT capacity; if IT departments should be outsourced; how long will it take to connect our ERP with our partners' ERPs; if the real-time business process information across all partners are visible for decision makers, and other question of similar nature.

This thesis investigates grid computing as an integrating platform for inter-enterprise IT applications. Grid computing provides a flexible infrastructure to share Internet-connected resources, including computers, data storages, software, sensors, instruments, and wireless equipment. It implements the model of on-demand resource sharing across administrative domains. Although the resource virtual market mechanisms are still under development, some pioneering enterprises have moved to implement intra-enterprise grid applications and have seen the benefits from such applications. After reviewing the development of enterprise computing, this thesis identifies the requirements of new virtual enterprises computing. Grid fundamental technologies are described based on Globus. Then, four kinds of grid applications are identified and analyzed in the virtual enterprise environment. Business practice case studies are presented after the respective individual application. This thesis also identifies that the key of virtual enterprises is dynamic relationship management. Contract Net, a high level protocol, is adapted to manage this relationship. The concluding section identifies key issues for further research.

Thesis Supervisor: Dr. Amar Gupta, Sloan School of Management
Thesis Reader: John R. Williams, Associate Professor of Civil and Environmental Engineering

Table of Contents

Table of Figures

Acknowledgement

This thesis research constitutes one part of "Virtual Market and Wireless Grids", a collaborative research project supported by National Science Foundation grant #0227879.

This thesis is supervised by Dr. Amar Gupta, Co-Director of Productivity from Information Technology (PROFIT) Initiative at Sloan School of Management at MIT.

Thanks to my team members: Sze Hwei Ong, Madhav Srimadh, Ozlem Uzuner.

Thanks to Professor Lee McKnight, Professor Joon Park, and Professor Junseok Hwang at School of Information Technology at Syracuse University.

Thanks to my classmates at Harvard Business School: Justin Gatlin and Sivakameswari Viswanathan.

1 Does History Replay?

> "If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility.... The computer utility could become the basis of a new and important industry."
>
> ---John McCarthy, speaking at the MIT Centennial in 1961

Forty years ago, researchers at MIT were working hard to verify their theory of MAC (Machine-Aided Cognition or Multiple-Access Computer), which was very different from the concept of batch processing at that time. From a larger perspective, MAC tried to move the area of computing application from information processing to information communities, in which people would be able to access the computing power at the same time and would be able to share knowledge. It was depicted in a paper from Scientific American that appeared in 1966:

> "The system makes it possible for the users to carry on a discourse with one another through the machine, drawing on its large stores of knowledge and its computing speed as they do so. The time-sharing computer system can unite a group of investigators in a cooperative search for a solution to a common problem, or it can serve as a community pool of

knowledge and skill on which anyone can draw according to his needs. Projecting the concept on a large scale, one can conceive of such a facility as an extraordinarily powerful library serving an entire community-in short, an intellectual public utility." [1]

The MAC goals are surprisingly similar to those of today's grid computing. However, their research objects are very different.

In the 60s, batch processing was the only working model of computers. Users had to wait in line to submit jobs, to enter inputs, to get the data processed, and to receive results one by one. Knowledge sharing was through sharing the punch cards, which were encoded with programs. MAC tried to divide the computing time into pieces and to enable many users access the computing power at the same time. Later, MAC also developed mechanisms to enable storage sharing. Eventually, users could share information and knowledge in a centralized utility.

During the period from the sixties to the present, computers have been greatly empowered and miniaturized so that computing power has been dispersed everywhere. Varied hardware, instruments, and software are connected by the pervasive Internet and managed by automotive individuals without centralized control. It becomes very hard to share and coordinate these geographically distributed resources, administered by autonomic entities, to work on a common problem.

Starting from the centralized resources, MAC was aiming to distribute computing power everywhere; in contrast, grid computing is trying to collect distributed computing resources. Both of them share the same goal of knowledge sharing.

This thesis elaborates on how grid computing could serve as a strong support platform for virtual enterprise computing, which demands decentralized, dynamic, flexible, short-term, and across domain resource sharing. In the second chapter, the research objects and the scope of grid computing are introduced. The key characters of Grid computing' key characters are identified. The third chapter reflects the development path of distributed computing and its impacts on enterprises. In the new networked economy, virtual enterprises challenge information technologies to support flexible, dynamic, and cross domain resource sharing relationships. The proposed grid generalized technology is based on Globus. The fourth chapter is focused on grid application. Generally in terms of functionalities, business scenarios of data grid, computing grid, application grid, and collaborate grid, are depicted; a business case analysis follows each application. The chapter on Contract Net and virtual market introduce how grid technology can be used to help enterprises search decentralized resources, negotiate service level agreement, and collaborate in a virtual market. The chapter on wireless grid computing focuses on identifying the business requirements of wireless grid computing. The concluding chapter predicts the future development path of grid application and also highlights several open questions.

2 Introduction to Grid Computing

Internet has connected computers and, in essence, people. However, only email and the WWW systems have been widely used. Many other potential applications, like resource trading and virtual market, are still under development. The Internet connects dispersed resources but does not support coordination of these resources to resolve a common problem. New protocols such as resource discovery, negotiation or coordination protocols are necessary for trading resource on the Internet. As more and more resources are converted from atoms to bits and are connected to the Internet, a framework is required to approach all these resources on the net, no matter what protocols each of them uses and where they are located, and to collect them together in order to help people collaborate and work on a common and complicated problem. Some applications have appeared. For example, in 1994, 600 volunteers around the world collaborated through Internet to unscramble the RSA encrypted message successfully in eight months rather than 40 quadrillion years that the RSA inventor had claimed; in 2000, volunteers from around the world contributed idle computing cycles to look for communication with aliens in the SETI@Home [2] project.

As the WWW was born in the CERN, grid computing was born and tested in the academic world in the 1980s, when researchers looked for a way to share expensive computing resources and experiment equipment. This concept evolved from the metacomputer [3]. The underline technologies are based on parallel computing and cluster computing. Ian Foster et al., define [4] grid computing and its research objects as

"coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations."

It is worthwhile to note that the grid concept evolved and is still developing. The grid application started from computing power tapping and until now, most commercial applications and most people's knowledge are still limited to SETI@Home [2]. Actually, as computing power becomes cheaper and cheaper, the potential of the grid serving as a resource integration platform becomes more important than simply acting as a cheaper alternative for super computers. The resource can include data, software, computing cycles, sensors, and other instruments. The concept of grid as a resource sharing and collaboration tool has been used in academia. However, persons start noticing its potential commercial value due to the new business requirements required by such as virtual enterprises, outsourcing, and IT consolidation. The grid concept evolves from computing power sharing, resource sharing to being used to build service-oriented and components based integrated virtual organizations. This evolution path of virtualization continually abstracts end users from heterogeneous low-level technical implementations.

The Computing grid has many of the same characteristics as the electric grid, such as [5]:

- **Complexity**: characterized by complicate connections across multi administrative domain.
- **Standard**: enabled through research, business drivers, and the government

standard enforcement

- **Distribution**: characterized by several distribution levels with different complexity and different capacity.

As Jim Gray pointed out [6], "what makes grid successful is ubiquitous". Therefore, grid computing also needs to be pervasive, dependable, consistent, and inexpensive to access [5]. Grid computing is going to play a key role in the e-business as a resource sharing and collaboration platform.

3 Virtual Enterprise Computing

3.1 History of High Performance Computing

In terms of infrastructure, there are several levels of high performance computing. They are uni-processor, Massively Parallel Processor (MPP), cluster computing, and high performance distributed computing. Grid computing is closely related to the high performance distributed computing.

High performance uni-processor systems are normally very expensive. As an alternative, the Massively Parallel Processor (MPP) systems utilizing multiple commercial CPUs in one chassis are more cost efficient. As fast speed local area network (LAN) technology has evolved and workstation power has increased, cluster computing has become more attractive from the cost benefit perspective. Cluster computing is built on unit processors and commodity operation systems.

As early as in 1992, L. Smarr et al., described the concept of metacomputing [3], which was built on several geographically distributed super computers connected by fast speed networks. Metacomputing has been applied on large computing simulation, remote instrument/sensor control, and large datasets management. The authors also elaborated the analogy between the metacomputer and PC and suggested the concept of utility computing. This is the original concept of grid computing. Grid computing evolves directly from metacomputing, but with more emphasis on variety resource sharing.

3.2 Development of Enterprise Computing

As the computing equipment has moved from back offices to people's desktop, commercial IT applications have become strategically important and have begun to shape the development of information technology.

In the sixties and seventies, centralized mainframes dominated the enterprise computing platform, as exemplified by a very powerful machine connected with many dumb terminals. The enterprise computing of that time was regarded chiefly as back office processing power to enhance enterprises' operations. It was rarely considered a significant part of enterprises' business strategies.

Advances in technology reduced the computing prices so that powerful workstations and PCs could be massively deployed in the eighties. Individuals were able to promote productivity by leveraging on desktop computing power. As these kinds of software rarely helped collaboration, they did not change the work pattern to any significant degree. The advent of client/server infrastructure and TCP/IP changed people's work behavior permanently. For example, global teamwork began to be valued and project could be conducted 24 hours per day; business reengineering became a buzzword due to the huge IT impact on enterprises. However, on the other side, this trend also led into the advent of disparate information systems, directly or indirectly. First, as the technology became cheaper, the power to make investment decisions was lowered to the department level in companies and this resulted in fragmented investment and heterogeneous systems. Second, constrained by the scarcity of IT engineers, the fast developing business requirements, which were increasingly dependent on IT, could not be met on time and thus the users had to either develop the systems themselves without help from IT department (called end user development) or bring in outside IT engineers (called outsourcing development). Both approaches were potentially beyond the scope of standard IT control. Many disconnected information systems with varied standards were developed inside individual departments. Third, partially attributable to the lack of a common standard in distributed systems is the continuously evolving and yet incompatible information technologies. The development of client/server infrastructure also raised the debate regarding fat client/thin server or thin client/fat server, which correspond to the debate regarding centralization and decentralization organizations (discussed in detail later).

On the other side, in the business world business process reengineering forced decision-makers to think of IT applications at the strategic level rather than for limited use within individual departments. This in turn highlighted the needs for information and knowledge sharing capabilities across department borders even for a highly distributed enterprise. Destroying department information islands and building an integrated ERP system have become the choices of most enterprises.

However, ERP only makes the size of the information islands larger within the networked economy. When mergers, acquisitions, divestitures, or new partnerships occur, the task of correspondent ERP system separation or integration becomes a nightmare for CIOs. Some enterprises had to look at the degree of compatibility of their ERP systems are compatible as one of the partner selection criteria, when evaluating potential acquisitions. The Internet connects many kinds of distributed resources but does not provide a framework to coordinate these resources to support the set of so-called Internet Applications. The required framework needs to help resource abstraction, resource discovery, resource inter-operation at the semantic level, and resource access control. This framework is necessary for connecting the separated information systems at the semantic operation level. VPN has been used to connect enterprises, but this kind of infrastructure is basically isolated and static as well as difficult to change even though it is very secure. In the networked economy, enterprises compete by groups or value chains instead of by individuals. Successful business will depend on successful collaboration with partners and customer through information and knowledge sharing. The relationships among the business entities have become increasingly dynamic. Your customers can be your suppliers at the same time. Research I supply chain management reminds people that integration forces in business are extremely important. It is highly desirable to build a dynamic and flexible resource sharing relationship among enterprises.

To fulfill these requirements goes beyond the traditional client-server and Internet technologies. Researchers have proposed a new infrastructure that can support this new emerging paradigm.

3.3 Virtual Enterprise Computing

3.3.1 Why Virtual?

As the competition becomes increasingly fierce, consolidation and alliance become increasingly attractive from a competitive viewpoint. In order to achieve the leading position in the market segment, enterprises focus on their core competences and attempt to bring either the most unique or the greatest value to the customers. In areas where enterprises are second level or tertiary level players, they are trying to outsource and build partnerships with the leading players in that area. In a study by PriceWaterhouseCoopers, it was found that "outsourcing has moved markedly from performing a single function more efficiently to reconfiguring or re-bundling whole processes in new ways to generate greater shareholder value across the enterprise" [7]. On the other hand, outsourcing service providers are able to capitalize on the economies of scale and scope. Therefore, modern business competition is not among several enterprises, but among virtual enterprises. The competence of one virtual enterprise is dependent on how well these partners collaborate, how they define the coordination policies, and how they build a flexible and robust coordination mechanism to support their policies.

Two coordination frameworks, e-market and intercompany operating tie [8] could be used to build a virtual enterprise. While sharing the same goals, these two choices are also contradictory in many aspects as well. The framework of E-market emphasizes short-term and flexible relationship while the framework of intercompany operating tie aims to build a long-term and highly integrated relationship. Intercompany operating tie scenario exists more often among large companies while the e-market framework often involves many small and medium companies. Choosing which framework to build virtual enterprises depends on the channel cost. Intercompany operating tie reduces channel cost through highly integrated coordination, but it is not scalable to many partners. The e-market framework reduces channel cost through easily searching potential partners and attains the cheapest solution through free marketing system. Therefore, a flexible and dynamic information technology infrastructure is more important for the e-market framework than for the intercompany operating tie framework.

Outsourcing normally consists of three stages [9]:

- Sourcing modules: advances in the industrial age were based on product modularity that enabled the outsourcing.
- Sourcing processes: business process outsourcing will continue to increase as more specialist firms emerge in the reconfigured business network. Process outsourcing is "The delegation of one or more business processes to an external provider who then owns, manages, and administers the selected processes based on measurable metrics. [10]". As processes become more standardized and as the

sourcing market matures with more stable participants, many corporations will recognize the criticality of business process outsourcing.

- Resource coalitions: focuses on the establishment of a resource network, in which the firm is a part of a dynamic network of complementary capabilities. "A corporation becomes, not a conventional portfolio of products or businesses, but a portfolio of capabilities and relationships... Consequently, the strategic leadership challenge is to orchestrate an organization's position in a dynamic, fast-changing resource network." [9]

3.3.2 What is Virtual Enterprise?

According to Venkatraman and Henderson [9], virtualness can be reflected in three aspects: virtual encounter (B2C); virtual sourcing ("virtually integrated in a business network, in sharp contrast to the vertically integrated model of the industrial economy, which manages a dynamic portfolio of relationships to assemble and coordinate the required assets for delivering value to customers" [9]); and virtual expertise ("leveraging diverse sources of expertise within and across organizational boundaries" [9]). A typical example of virtual enterprise is provided at the end of this chapter.

There are many definition of Virtual Enterprise:

- NIIIP (National Industrial Information Infrastructure Protocols): a Virtual Enterprise is "a temporary consortium or alliance of companies formed to share costs and skills and exploit fast-changing market opportunities" [11].

- Byrne defines, cited in Walton and Whicker [12]: "a Virtual Corporation is a temporary network of independent companies – suppliers, customers, even rivals – linked by information technology (IT) to share skills, costs and access to one another's markets. It will have neither central office nor organization chart. It will have no hierarchy, no vertical integration".
- Walton and Whicker [12]: "Virtual Enterprise consists of a series of co-operating 'nodes' of core competences which form into a supply chain in order to address a specific opportunity in the market place".

Although the above definitions have merits, none of them covers all the three aspects of virtualness proposed by Venkatraman and Henderson. The author believes:

The Virtual Enterprise is a network of several companies, which contribute their core competences and share resources such as information, knowledge, and even market access in order to exploit fast-changing market opportunities. The relationship can be long or short term.

Camarinha-Matos believes that "the concept of extended enterprise, the closest 'rival' term, is better applied to an organization in which a dominant enterprise 'extends' its boundaries to all or some of its suppliers, whilst the Virtual Enterprise can be seen as a more general concept including other types of organizations, namely a more democratic structure in which the cooperation is peer to peer." [13] The author disagrees with the notion of excluding "extended enterprises" from "virtual enterprises" because these two concepts share the same purpose and requirements for IT infrastructure.

Virtual enterprises can be described from the following aspects [13]:

- **Duration** of relationship can be based on either single business opportunity or long-term alliances.
- **Topology/geometry**: in a virtual enterprise, "companies can dynamically join or leave the alliance according to the phases of the business process or other market factors." One company can also participate simultaneously in several business networks or be committed to a single alliance exclusively.
- The **Coordination** mechanism in a virtual enterprise can be star-like structure, in which one powerful users dominate others or democratic alliance.
- Individual companies' **Purpose** of joining a virtual enterprise can be for extending the business boundaries and keeping control over vital suppliers for instance in terms of quality control; for complementing its core competencies in order to be able to share market opportunities; for being involved in a consistent supply chain; for bidding for a single opportunity in the market; for increasing the geographical presence or for improving the quality and responsiveness to the market opportunities.

The creation of a virtual enterprise requires dramatic changes in organization, in areas such as teamwork and delegation of decision making to the lower levels in the hierarchy. The most important aspect for it to attain its desired objective is the realization that the evolving virtual enterprise requires a sophisticated IT infrastructure support.

3.3.3 Centralized or Decentralized?

The proximity to the market and as well as to the customer, globalization, and diversification: all of these push the enterprises to decentralize. On the other end, sharing market and sharing resources require high integration. Corresponding to the business organizational structures, IT department in the enterprise also have to be decentralized or centralized. Mainframe machines figured prominently in the development of centralized information systems while client/server technologies have played a key role for decentralization.

For example, Cisco is regarded as a highly centralized company built on the Internet with all manufacturing activities outsourced.

> "Morgridge (John Morgridge, then CEO of Cisco) believed that many Silicon Valley firms decentralized too quickly and did not appreciate the proven ability of the functional organization structure to scale without sacrificing control during high levels of growth. Accordingly, Morgridge maintained a centralized

> functional organization that is still in place today. While Product Marketing and R&D are now decentralized into three "Lines of Business" (Enterprise, Small/Medium Business, and Service Provider), the manufacturing, customer support, finance, human resource, Information Technology (IT) and sales organizations remain centralized.... There is a belief within Cisco that consistency of strategy, goals, organization, and management provides a huge stable benefit to a fast growing, fast moving company." [14]

However, centralization is not a common rule used everywhere. There are several underline reasons that make centralization suitable for Cisco:

- Cisco product lines are relatively narrow.
- Cisco products are quite standard and are used in a high technology environment. Therefore, customization is not important.
- One of Cisco strategies is selecting the right strategic partners, which implies that aims to build a lean organization; keep the core competences tightly and integrated; and leverage outsourcers to develop fast as well as cater to the fast developing market.

Beginning in the year 2000, Cisco's Chinese market sharing is going down quickly due to the competition from Huawei, a Chinese telecom equipment company. The reasons that might explain this trend are as follows:

- The product design is not suitable for the Chinese telecom infrastructure. When there was no alternative in the Chinese market except Cisco China, it was not a problem. However, as Huawei entered this market with cheaper and relatively good quality products that were more appropriate for customer's requirements, Cisco China faced a serious challenge.
- Cisco China only has sale department and few technical support engineers. To provide technical support, it depends on the partnership with Legend, which is the largest Chinese system integration company. Because CISCO engineers do not work with customers directly, it is not a surprise that Cisco product cannot meet Chinese telecom market requirements.

It seems a dilemma for an enterprise to change its IT organizations between centralized and distributed. Each of them has pros and cons depending on the business environment. One of the most important issues is that distributed information systems often result in the separated systems which are very difficult to be integrated and requires high maintenance. Decentralized information systems can support the core competences of enterprises with distributed organization structure. However, even in the decentralized companies, centralized and integrated information access is also highly important.

VeriFone [15], the number one producer of transaction automation systems in the world, focuses its strategy on technical feasibility, distribution capabilities, and geographic location, because its products need to be certified by local regulators. Thus, VeriFone develops a decentralized network of locations in order to deliver high quality and on time products through local distributed channels. Leveraging on this culture, they even value individualism as a source of creativity and build an entrepreneur culture inside the company. Moreover, the CIO, Will Pape, starts his work everyday by accessing all information either from inside the company of all locations or from outside company. He distributes his analysis results to personnel of all locations. "VeriFone's ability to draw data from a single centralized database of globally aggregated information, accessible from anywhere in the world had tremendous advantages. Internally, it guaranteed consistency between operational and planning levels. Externally, it allowed the company to be extremely resilient to external emergent response." In summary, although VeriFone is a highly decentralized organization, the centralized culture and information sharing technologies is the key to keep its organization integrated.

Grid computing as a solution to integrate dispersed resource is an idea computing platform to balance the pros and cons of centralized resource organizations and decentralized resource organizations. It will reduce the negative effects of both organization forms and enable enterprises to employ the best appropriate organizational forms. Grid computing is able to mange distributed resources and still provides a single point of accessing information. The information in the grid can be distributed but still can be consistent. The resources in the grid can be dispersed globally but still can be shared and coordinated together.

3.3.4 On Demand Integration of Service

Researchers argue for a critical focus on core competencies and an organizational design that best leverages these competences. However, these innovative organization structure changes cannot happen without strong support of computing infrastructure. It seems that the trend towards integration in enterprise computing is contradictory to the trend toward distribution in the business world. Actually, advances in IT can help to resolve the conflict between centralization and decentralization. Several kinds of IT collaboration tools are able to help a team whose members are dispersed globally and work on a common problem; the grid, as discussed later, can help to coordinate geographically distributed resources through virtualizing these resources.

A virtual enterprise IT infrastructure usually should support exchange of information, order status monitoring, quality-related information exchange, extended production planning and control, distributed resource planning, negotiation system to facilitate partner search, contractual processes during the formation of a virtual enterprise, management of a contracts database, and electronic catalogue [13]. To provide the speed and flexibility required by virtual enterprises, information systems need to adapt to a common infrastructure, which will enable them to share resources securely. This common infrastructure should be scalable, evolvable, and adaptive [16].

"When designing an IT platform to support industrial virtual enterprises, certain issues related to information management requirements become especially challenging, such as the physical distribution of data, the enterprise autonomy and privacy enforcement, access rights to shared information, and data visibility levels [17]." To share data resources in virtual enterprise, one requires concurrent access, coherence, and security. The datasets might be very large. Data access should occur without delay; data location should be transparent; secure should be guaranteed. It should not be necessary to modify the application software.

Camarinha-Matos, et al., [18] describe the structure of virtual enterprise execution system (see Figure 1 Virtual Enterprises Computing Components [18]) and identify several demands of virtual enterprise for supporting communication systems. A standard interface is required to enable business objects integration across applications; policy control of integration should be autonomous and flexible across enterprises borders; semantic representation of data should be enabled.

Considering the dispersed nature of a virtual enterprise, the structure of a minimal virtual enterprise information system should comprise of: a process execution and coordination system; a safe communications infrastructure; and a distributed information management system [18].

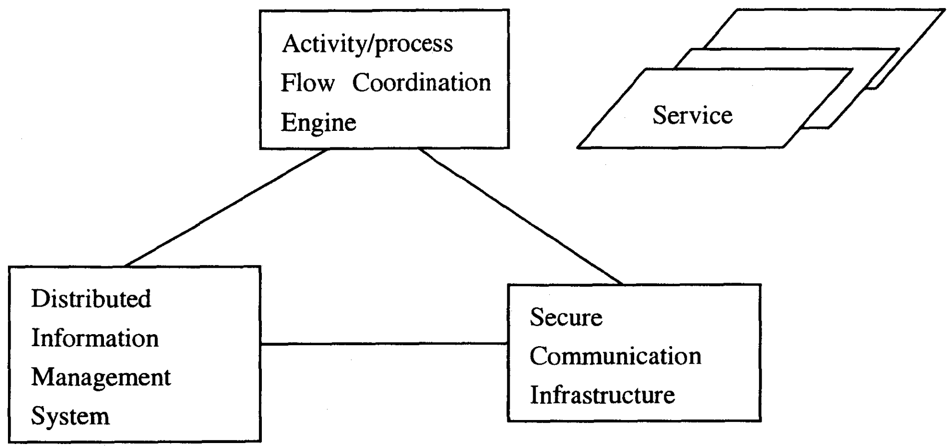

Figure 1 Virtual Enterprises Computing Components [18]

Contracts or cooperation agreements are the main vehicle that regulates the general execution of business processes between enterprises. Such agreements can be used to coordinate the interdependencies among partners in a Virtual Enterprise. "A contract shall establish the conditions for the interdependence/cooperation among partners and, therefore, can be the basis for a coordination policy" [18]. In the chapter entitled "The Contract Net and Virtual Market", a new automatic approach will be investigated to embed this process into the information infrastructure of virtual enterprise. Because contract agreements are normally varied and flexible in term of individual partners, the applicability of this automatic approach will be very limited.

For a long time, research groups from two different academic backgrounds have independently conducted research on resource sharing across administrative domains.

The group from computing research community started research from the viewpoint of sharing computing resources to the notion of sharing all kinds of resources connected by the Internet for the purpose of supporting business requirements. The other group, who coined the term "Virtual Enterprise", comes from enterprise information systems research and started from an analysis of business requirements and the search for an appropriate IT infrastructure. Their research was focused on distributed programming technologies, mostly CORBA. Obviously, these two research groups share similar goals and complement the potential benefits. Grid computing lacks most of high-level functionalities design such as coordination resource sharing while the research on virtual enterprise certainly needs the powerful grid resource-sharing infrastructure.

3.4 Grid Computing-infrastructure for Virtual Enterprise Computing

From the above discussion, the main problem in virtual enterprise information systems is distributed resources sharing relationship management. Grid computing is a framework designed to collect geographically distributed resources to work on one common problem. There are many approaches to implement grid framework. Among them, Platform [19], Avaki [20], DataSynapse [21], and Globus [22] have been widely used. Globus, because of its open source and widely vendor support, has become the de facto standard. Therefore, the following discussion will be based on Globus. However, as it is still at early stage, many functions are still missing in Globus. Globus consists of a suit of protocols supported by API and SDK [23] (see Figure 2

Globus Layer Architecture [22]).

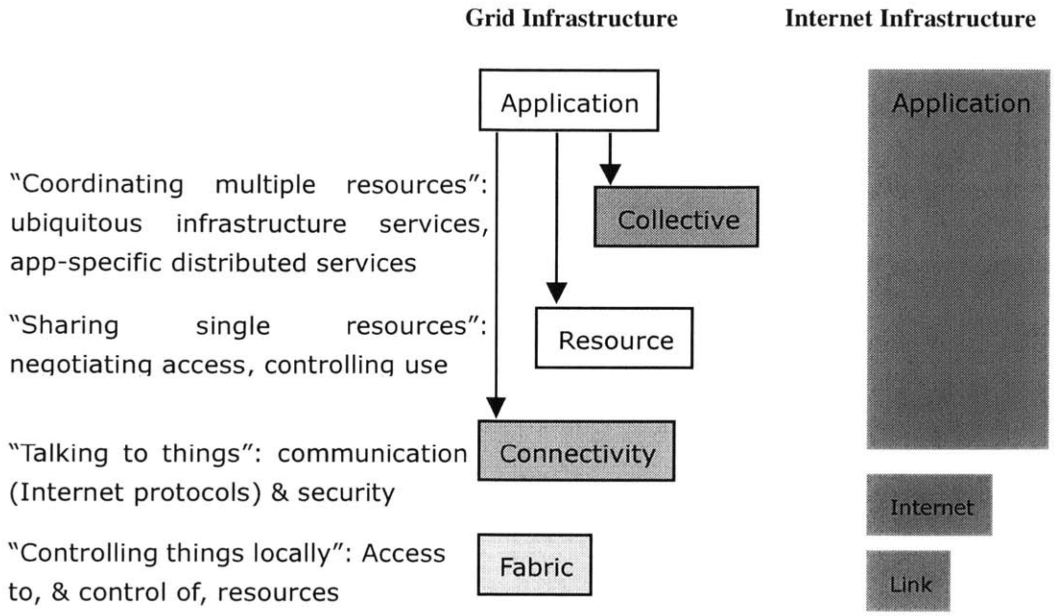

Figure 2 Globus Layer Architecture [22]

Globus adopted the successful Internet protocols design and comprises of five layers. Each layer provides API and SDK to help the application development process. Because every layer is separate from the layer that are above it and below it respectively, it is able to implement diversified protocols in each layer without losing interoperationability (see Figure 3 Globus Layer Components [22**Error! Bookmark not defined.**])

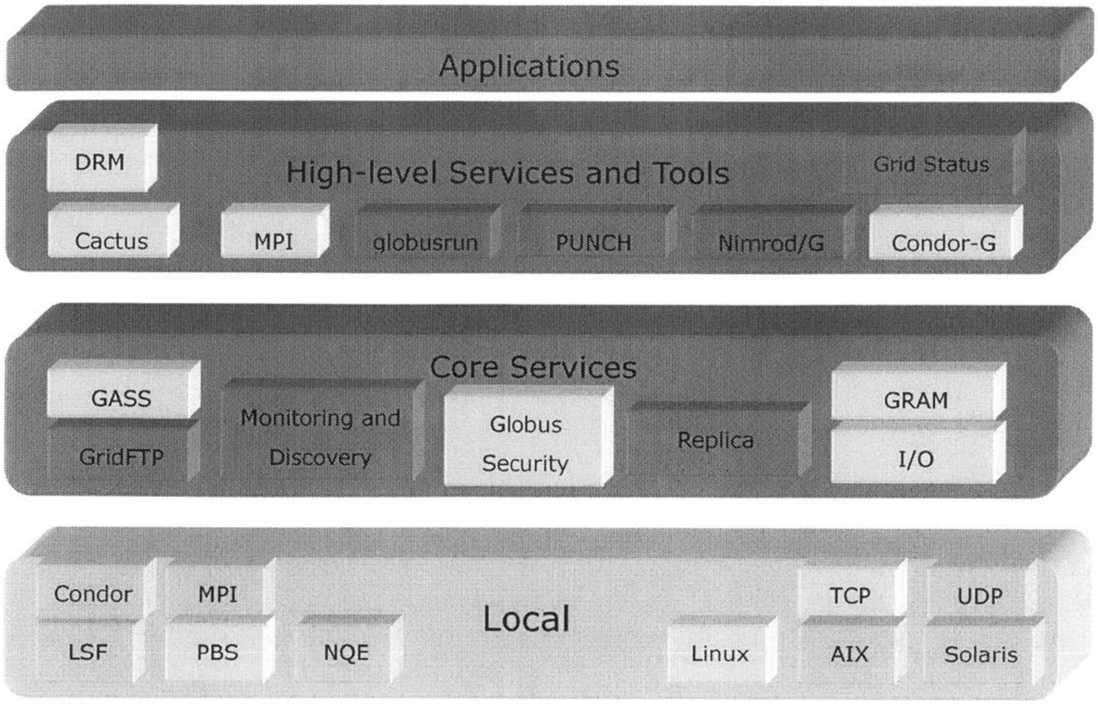

Figure 3 Globus Layer Components [22**Error! Bookmark not defined.**]

3.4.1 Globus Architecture Layers

The Fabric Layer deals with local resource and is resource-specific. The resource can be computing cycles, storage, network, code repositories, or catalogs like databases. In terms of resources, alternative resource operations can be provided. However, the minimum required operations are ones related to enquiry and resource management.

The Connectivity Layer supports core communication and authentication. Because virtual organization system works involves dynamic relationships and the need to establish trust, security is extremely important. The required functions are single sign on, delegation, integration with various local security solutions, and user-based trust relationships. Globus related protocol is grid Security Infrastructure protocol, based on public-key.

Based on the connectivity and the security service of the Connectivity Layer, the Resource Layer defines a suite of protocols on service negotiation, initiation, monitoring, control, accounting, and payment. However, this layer still concerns only the local resource. It deals with two classes of information: information protocols and management protocols. Because management protocols handle the negotiation and sharing relationship initiation, they are embedded with access policies. In Globus Toolkit, Grid Resource Information Protocol (GRIP), Grid Resource Registration Protocol, Grid Resource Access and Management, and GridFTP are defined in this layer.

The Collective Layer is used to coordinate multiple distributed resources and to capture interactions across collections of resources. It is not related with any individual resource. Due to the variety of grid applications, there can be many protocols defined in this layer, for example, directory service, collaboration frameworks, software discovery, and etc. The Collective Layer can be general or domain specific. In Globus Toolkit, many protocols are defined.

The Application Layer consists of grid applications, which can be developed on services defined at any layer.

3.4.2 Grid Services

The Grid service was defined to provide standard interface in order to virtualize the grid resource and to facilitate grid interoperationability.

Grid service is a Web service that provides a set of well-defined interface and that follows specific conventions. The interfaces address discovery, dynamic service creation, lifetime management, notification, and manageability; the conventions address naming and upgradeability [24].

By combining web service with Globus, a set of basic grid resource operation interfaces was defined to compose the Open Grid services Architecture (OGSA). These interfaces include Gridservice, NotificationSource, NotificationSink, Registry, Factory, and HandleMap. The Grid service extends Web service through defining transient service instances and notification mechanism, which are significant in the grid environment.

3.4.3 Grid Topology

Based on the layer architecture and the grid service, computing resources are virtualized and can be easily coordinated to build grid application through all the levels. In the following, different grid topologies are discussed [25].

IntraGrid is characterized by deployment limited in one organization and more important, in one security domain. The business drivers behind this deployment are higher utilization of computing resource and consolidation of computing resources.

ExtraGrid connects two or more IntraGrid and, therefore, has several security domains. Every grid is autonomic and has its own access policies. Security management and resource management become very difficult and more important in such an environment. Virtual Enterprises often employ ExtraGrid to integrate their systems and to share resources.

E-utilities aim to provide computing resource in the same way as electricity. Enterprises can purchase computing resource on-demand instead of maintaining their own IT infrastructures. This is driven by the outsourcing trend, xSP business model, and more standardized software components.

3.4.4 Grid Software Components

A grid can optionally consist of the following software components that reside within different layers in the Globus Grid Architecture.

Grid Node Software manages the grid node's local resource, such as authentication, resource registration, resource usage monitoring and reporting, policy management, local job scheduling, tasks execution, and results reporting.

Management Component is responsible for tracking the available grid resources and their capacities, measuring grid's usage pattern and monitoring accounting log.

In the InterGrid environment, **distributed Grid management software** may be needed; further one may need to compose hierarchical or networked management software.

4 Grid Application

The grid infrastructure is adaptable to business requirements allowing for the addition of components to the grid based on the type of application desired (see Figure 4 Globus Data Grid Architecture). Grid applications can be generally categorized into four groups - computing grid, data grid, application grid, and collaboration grid. As grid technologies develop, these categories will grow as well. It is worthwhile to note that these definitions are subjective. In large applications, it is very normal to combine these grids together. In these cases, the grid infrastructure will be rather complicated. For example, NASA's Information Power Grid aims to improve the Aeronautics and Space Transportation design process by supporting large scale product simulation, integrating large sets of data from different sites, and enabling distributed collaborative design.

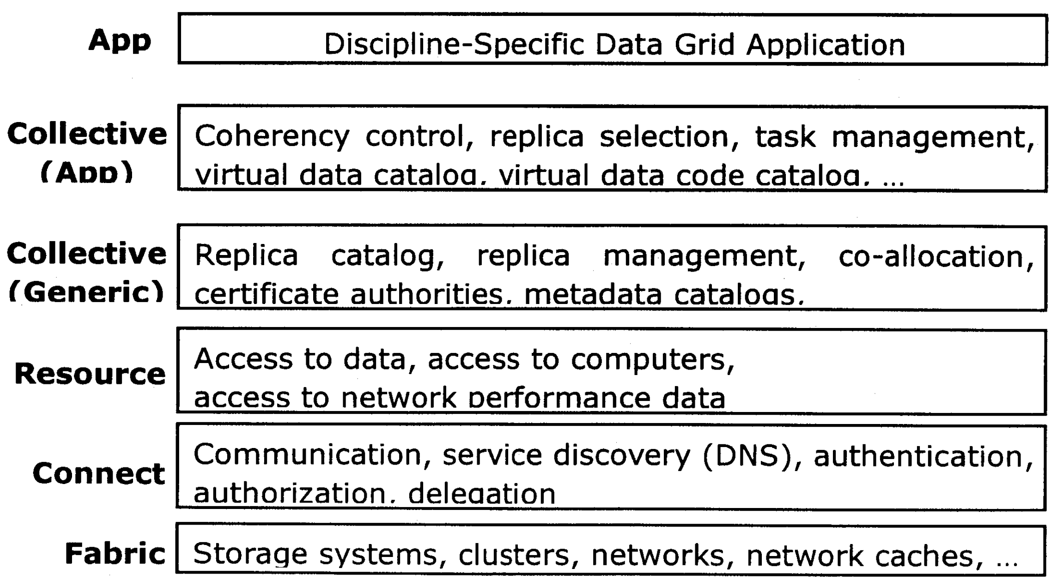

Figure 4 Globus Data Grid Architecture

4.1 Data Grid

Sharing of information sharing is extremely important for a virtual enterprise. The flow of information among the networks of virtual enterprises enables enterprises to respond to external events at the right time with the right actions. The shared information could include market data, production schedules, long/short term demand forecasts, and engineering and design information, and logistics information. These pieces of information are required to be correct, updated, and semantically integrated. However, due to the decentralized organization structure and development history of enterprise computing, these pieces of information are often isolated and managed by different entities in different data formats. In the dynamic virtual enterprise or virtual market, simply finding the right information is very difficult. Although Internet provides ubiquitous access connectivity, often it falls short as a mechanism of effective enterprise data sharing because of the following challenges:

- Search - How to find the right dataset in the a dynamic environment
- Transfer - How to transfer large datasets in the wide area network
- Interoperability - How to bridge the semantic difference among datasets
- Visibility - How to define the data visibility
- Security - How to protect the data
- Usability - How to define a user friendly query interface

In the past, many enterprises dealt with these challenges through Virtual Private Network (VPN) connection and File Transfer Protocol (FTP). But they are inefficient and unpractical for business applications, especially in the virtual enterprise environment where a short-term co-operation implies simple and fast setup of computer infrastructure for data sharing. Companies need to be able to setup and adapt their computer infrastructure quickly to be able to share data with other members of a given virtual enterprise.

Data grid has overlapping goals with heterogeneous distributed database systems, which deal with different kinds of database management system distributed in a heterogeneous environment like different hardware, operation system, network connection, data models, and even DBMS vendors [26]. Both of them aim to resolve the distributed data management tasks including integrated data catalog, data discovery, distributed query, distributed transaction, and semantic integration. However, most distributed database management systems are focused on research under the database circumstances while data grid targets are more complicated as they involve dynamic environments across multiple administrative domains. In addition, distributed databases implement access data location transparently while databases locations are still relatively fixed. In the case of data grid, based on storage sharing, data can be managed by grid and moved based on access efficiency. Simply stated, data in data grid could have no certainly assigned locations. Third, most distributed database systems require a central information server, which is not possible in the dynamic data grid environment. Therefore, the tasks of notification and event mechanism are very important in the case of data grid.

In the academic world, there is a desire to share the expensive experimental data in order to coordinate research. In the business world, data sharing is needed with even higher requirement. Except the above listed requirements, the business data must be in real-time.

For an e-market, it is very important to maintain an updated and consistent product catalog. If we require the suppliers to submit their product information and inventory and update the database, the information is duplicated and at risk of being outdated.

To solve this problem, the key is to eliminate duplication and to be able to authorize the users to access each supplier's database directly. Obviously, this does not work today partially due to the large number of suppliers' databases and the concerns related to security. In addition, because of the poor usability, nobody will use it. Data grid enables users to elicit relevant information from the distributed databases and yet provides integrated information.

Data grids provide an infrastructure to support data storage, data discovery, data handling, data publication, and data manipulation. Enterprise data usually possess the following characters: scale, dynamism, autonomy, and distribution of data sources. Data grid aims to make these transparent to grid applications through a layer of virtualization services. These services support federated access to distributed data, dynamic discovery, of data sources based on content, dynamic migration of data for workload balancing, and collaboration [27].

Data grid consists of data from different source in different forms existing in different physical locations. For instance, the data can be stored in different file systems like FAT32 or NFS, in different databases like DB2, Oracle or even in other Object Oriented databases. The power of data grids is that all of these differences can be hidden from the users, who are able to use the data as they use the data on their local hard disk. Data grids offer a number of advantages over the traditional data access methods such as:

- Transparent sharing of data across organizations
- Transparent sharing of data across data file formats
- Elimination of the inconsistency problem of data because these data are actually not moved.
- Higher level of security through the grid security infrastructure (delegate mechanism), the system security is warranted.

Extending the bioinformatics application scenario described in [22], a scenario of online CD store – "CDGrid Inc." is described below.

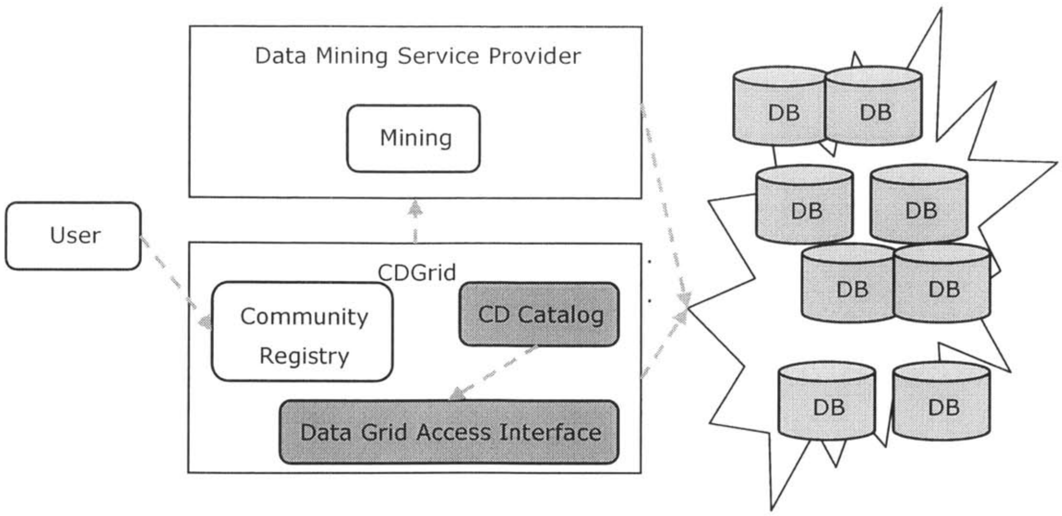

Figure 5 Example: CDGrid Inc.

- CDGrid wishes to create an integrated database to keep real-time product information based on all its suppliers' inventory data. The database should be publishable as the electronic catalog on the website and support query by CD titles or singers.
- CDGrid further intends to provide the customized CD catalog service for the users. Users may expect the CD to meet certain criteria, such as the list to be frequently updated. CDGrid can then create a database on its own server or, if the user allows, create a database on the user's computer, in order to provide faster access speed to the users.
- When users query the electronic catalog, CDGrid will initiate a distributed querying on the data grid.
- The data grid's constituent databases can dynamically join or leave the data grid configuration at any time. When CDGrid cannot find the CD, CDGrid may connect one data mining service providers in the grid computing virtual market

that can be dynamically located.

- This data mining service provider will use computing grids as well as data grids to search all its own databases, as well as its partners' CD producers databases.
- The query results will be provided to the user or cached into the database created for this purpose. The data movement is also facilitated by the data grid technology.
- When a change occurs to the suppliers' inventory change or the prices at any time, their databases will send a notification to the cached database and update it. This is enabled by the notification function of Grid Services.

Cast study: [28]

Pfizer is the world's largest pharmaceutical company; its products cover human health, animal health, consumer healthcare products, confectionary, and shaving products. Pfizer is decentralized organization composed of business groups around the world. Corresponding to the organization, IT resources are located in individual business department. The IT Leadership Team, which consists of vice presidents from business departments, coordinates the IT infrastructure standards and enterprise-wide applications.

Innovation is extremely important for Pfizer. Drug discovery and development are very complicated processes. Huge amount of data are needed and created during these processes. The processes include target identification, prioritization/validation, lead identification, optimization, preclinical technology, chemical manufacturing controls, pharmacology/toxicology, Phase I clinical study, and Phase II clinical studies. Apart from the data related to these steps, other information is gathered from multiple sources including published papers. Hauck said, innovation "will happen from query language joins of data across some of our 3,000 applications." As the notion of "customized drugs" and drugs' requirement adapted for local customers, the size of the data involved will grow dramatically. Managing and analyzing these distributed data sets in a consistent manner is a formidable task.

The use of data grids can help to mitigate this problem. Data grids can enable drug researchers to gain access to comprehensive patients records that meet designated characteristics, such as age, race, geographical areas et; such data could come from large disparate databases located in hospitals at many geographic places. Drug companies could also share experiment data with their partners across administrative domains. (see Figure 6 UK e-science X-informatics Grid [])

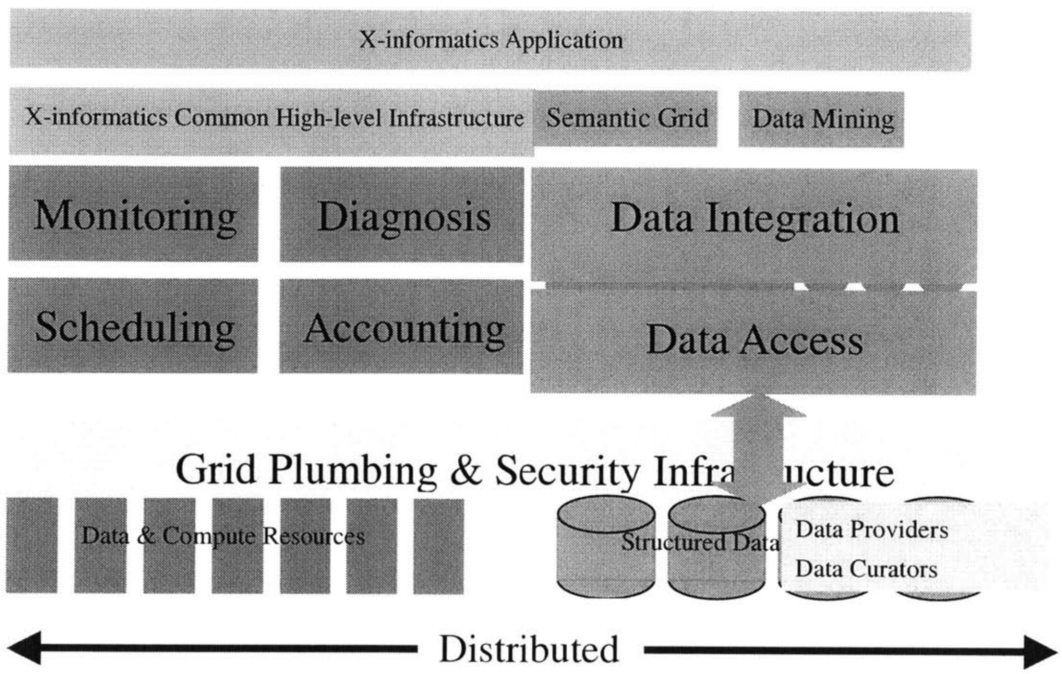

Figure 6 UK e-science X-informatics Grid [29]

4.2 Computing Grid

In the last several decades, as IT becomes an important business driver for the companies, we have seen numerous investments in IT, from IBM main frames, DECs to PCs. Companies invest half their budget into IT without conducting financial analysis. The failure of e-commerce and numerous failure ERP projects disappointed people so that researchers start questioning the IT investment and looking for suitable financial measurements for IT projects. Most enterprises began cutting IT investment and consolidating existing IT resources.

As discussed before, decentralization, fast IT development, rapid market opportunities all led into the fragmentation investment in IT. It is very common to see that every server machine runs only one application. In some cases, this makes sense due to security concerns. However, most of time, people just didn't spend time on monitoring their applications' running patterns and their computing servers' usage. Actually, computing resources could be integrated and shared among applications assuming that different applications have different running patterns. For example, web servers are very busy in the day and light-loaded in the evening. But transaction validation and financial model simulation for reporting are normally conducted in the evening. For global companies, data centers in Asia work at the different time from the data centers in the America. These applications can definitely share computing resources without losing any efficiency. Because of the tight economic condition, some companies start consolidating data centers to cut cost through installing system health monitoring software to understand application running patterns.

Computing grid provides the ability of sharing computing power across heterogeneous operating platforms, locations, and administrative domains. The computing grid is a strong platform to consolidate computing resources and cut IT investment and operational cost. Computing grid is able to monitor the server computing resource usage, discover software's usage pattern, reserve computing resources for certain applications in terms of priority policies, forecast the application running behavior, and allocate appropriate computing resources to different application in terms of policies as well as software requirements. Once each server is grid enabled, they are treated as a super server and servers can share computing power. Several applications can run on one machine as well as one application can run several servers when high performance is required. In the virtual enterprise environment, the cooperated member can even share their computing server resources across administrative domains. If employing accounting mechanism, the members can even trade their excessive computing resources.

Different workload management software has different implementation. For example, Condor, developed at University of Wisconsin in Madison, is composed of three basic components: central manager, execution hosts, and submission hosts. The central manager collects the running status of all computing server in the clusters and match computing resource request with a server that can meet this requirement. Execution machines are servers executing the assigned jobs. Submission hosts are servers that request computing resources. Note that this server division is based on functions. In fact, every server in the cluster can be configured execution machine and submission machine at the same time. There is always one central manager in one cluster although it can be co-located with submission machines and execution machines.

Case: Postgirot Bank [30]

Postgirot Bank was founded in 1925 to implement cashless payments through the giro system. It serves all Swedish business activities and there were 155,000 accounts using on-line services by may 2000. Postgirot Bank is a leader in IT application. Its IT department is decentralized. Most Postgirot Bank's computing servers were distributed in departments and were added as new application needed. There are three reasons leading Postgirot Bank into decentralized IT governance. First, Postgirot Bank business grows very fast that business can't wait IT planning. Second, there are IT skilled professionals inside every Postgirot Bank department. Third, Postgirot Bank uses decentralized purchase policy. In 1999, the owner of Postgirot Bank, the Swedish government, decided to sell Postgirot Bank. In order to favor the future buyers, Postgirot Bank decided to consolidate all computing servers into one data center to cut IT cost. By employing Provment AB's system health monitoring software package, which is used to monitor the utilizing pattern of server and provide productivity reports understandable for both IT and non-IT people, Postgirot Bank found that their computing servers were extremely underutilized. Many servers were most likely not needed.

Running several applications on one computing server could significantly cut IT cost. However, leveraging on Computing grid, more cost could be cut further through sharing computing power among computing servers given different servers have different usage patterns, for example peak time. Grid resource management system can even prioritize the computing tasks by enforcing enterprise policies in order to meet the overall enterprise business requirement with minimum IT cost. For example, transaction application might have higher priority than normal office application. Therefore, the transaction server always has the higher priority to receive the computing power. In summary, computing grid gives data centers more latitude and stronger tools to make use of computing resource more efficiently.

4.3 Application Grid

In the grid, software applications can also be shared. For example, when one expensive software can only be installed on one machine in the grid, all data have to be sent to those machines to be processed. Grid resource management systems and schedulers can be used to manage the access policies or priorities [25]. Another sense of application grid is that there are many application components deployed in the grid. Users can use a workflow language to link them into an application. This kind of Application grid is based on web service and each node in the grid provides a grid service. Grid services extend web services in the following aspects:

- Web services provide a mechanism of services discovery and invocation. However, web services only manage persistent states of services. This means, once the web service is programmed into the users' application, it is fixed and hard to change. This is not suitable for the grid's requirement, dynamic joining and dynamic leave. Grid needs transient information. Therefore, GSDL extends WSDL and provides information of lifecycle management, service creating, notification, etc.
- Grid services provide notification mechanism. Grid services implement both push and pull information models.

In the application grid, Users submit their jobs as a workflow description language, which explains the required software components with specifications and execution orders. The execution of software components might be sequential or parallel (see an example application grid: Figure 7 Fraunhofer Resource Grid []). Look at the following examples.

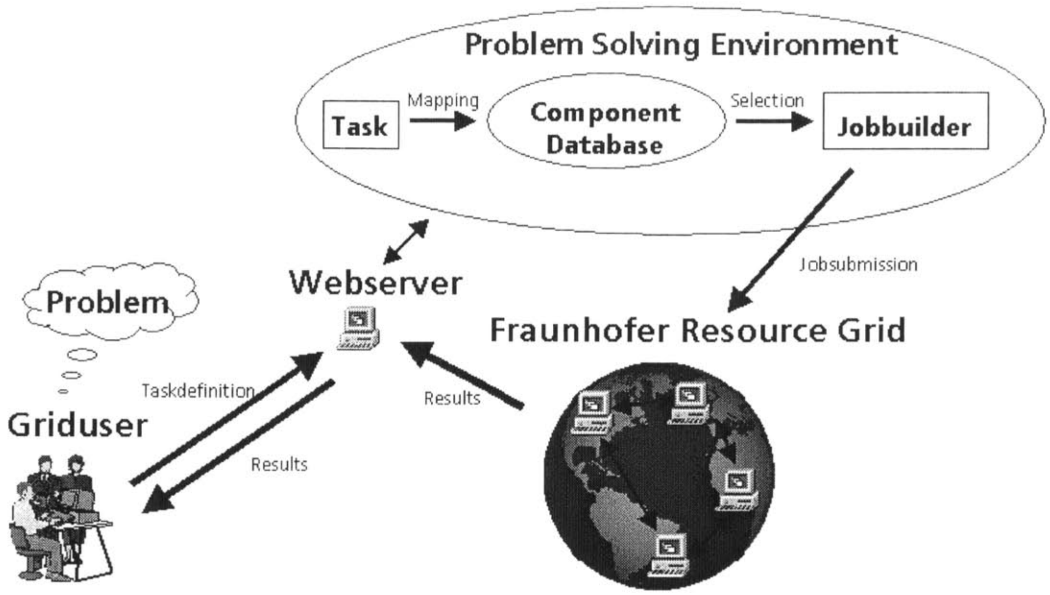

Figure 7 Fraunhofer Resource Grid [31] [32]

In [33], an application grid scenario is described. This application grid provides the users with a virtual fly-through of Hawaii on a desktop computer. The required data sources include satellite terrain data, a data file containing movie camera positions (captured by user's mouse movement on the desktop computer), and a MPEG "skeleton" parameter file. The required applications include a frame generation program that will render 3D images for each camera position into image files and MPEG encoder program. The frame generation program is a parallel application running on two Sun Solaris E420s. The MPEG encoder runs on a four processor SGI. The client side runs on a PC. The client submits a workflow file, which is parsed by the workflow parser running on the broker server. This server is responsible for searching suitable software components, a frame generator and a MPEG encoder in this case, through looking at the grid Information service server. Once the resource broker finds a suitable target resource, it returns the hardware attributes and the software attributes of the "frame generator" application. The same process happens with the MPEG encoder. According to the workflow instruction, the broker server moves the user input camera position parameters to the Frame Generator application. After the frame generation, the broker server transfers the movie frame images to the server that runs the MPEG encoder. Then the broker server fetches the resulting MPEG movie to the client's desktop.

There are some alternative scenarios. For example, the broker server could divide one frame generation job into several small jobs and assign them to many frame generators in order to achieve the highest cost/efficiency and speed. In the grid, there might be many frame generators. Although their underline implementation might be different, as long as their processing capabilities meet the computational requirements, they can be used to generate frames in a parallel fashion. Cybenko, et al., [34] describe a negotiation protocol that is suitable for parallel operations, and considers the issues of the quality of the result.

The second alternative scenario occurs when the author does not elaborate on how the Frame Generator can find the terra data files. These huge terra datasets could be located concurrently on several geographically distributed machines. In one case, the data could be moved to the machine where the frame generator software will be executed. However, when the dataset are too large and the bandwidth of the network is limited, the grid system might choose to move the frame generator software (mobile code) to the machines that host the data. Only the processed results need to be transferred back to the network.

In the computing grid, SETI@Home [2] for example, the working machines receive software from the submission machines. However, in the application grid, the grid node provides not only the computing power but also its own software application resources. This is a major difference between a computing grid and an application grid.

In the virtual enterprise environment, it is always required to coordinate partners' applications and integrate them to work on a common business issues. Sometimes this is required in order to facilitate information flow among virtual enterprise partners. Sometimes software components sharing is needed to leverage the best practice from partners. For example, several enterprises might cooperate to work on the design of a new automobile model. Each company has the best practice in designing certain component as well as correspondent simulation software. During the phase of integrated simulation, those pieces of simulation software have to be integrated and coordinated to test the whole automobile model. Application grid can meet such business requirements without disclosing the component design data and simulation software because the software are executed at their own servers. Furthermore, enterprises can even trade software components in the virtual market.

Application grid is enabled by modularization of software, which represents a very important advance in the software industry. Christensen describes that the product development from interdependency to modularity represents the advance and maturity of this product [35]. Modularity breaks the integrated products. Producers are able to compete on the performance of individual components. For example, in the mature automobile industry, "GM, Ford, and Chrysler operate more like assemblers and integrators of subsystems than traditional vertically integrated manufacturers. Many people are worried that outsourcing software components or using grid service provides by other companies in the application grid will compromise their core competence and increase dependence on outside forces. However, Baldwin and Clark argue that in a new age of modularity, the value-adding role of a corporation is less in the manufacture of a critical component than in the creation of a product or service architecture [36]. Besides, the flexibility of switching software components actually reduces the dependence on certain software providers. Therefore, it becomes important that the information system designer assembles multiple software components in the application grid to create a solution in a short period to meet the requirements of virtual enterprise computing. [9].

The virtual enterprise information systems built on application grid have three advantages: first, using modularized software components improves the flexibility of leveraging the best outside knowledge; second, competition becomes possible among the software component producers and help to promote software quality and reduce price; third, on-demand plug in/out software components becomes possible. Therefore, the information systems could respond to the changing business requirements faster.

Web service is a framework and a set of protocols for resource discovery and interoperation of software components. Grid service enhances web service through defining notification and transient mechanism. Application grid leverages the grid service mechanism to implement the virtual enterprise applications, which actually work in the similar manner as the normal software applications. Web service components are similar to the software libraries in the PC programming environment while application grids provide a framework to coordinate them into one Internet scale application.

Case: JP Morgan [37]

With reduced budget for IT and growing computing demand, JP Morgan needed to consolidate the IT infrastructure. Noticing that hardware become cheaper and cheaper and what costs a lot are restoring the application states (which means that application saves its status on the machines on which it runs) and customizations, JP Morgan chose to initiate the project of "Compute Backbone", which aimed to separate IT infrastructure from applications. Leveraging on grid computing, JP Morgan built a standard infrastructure based on "disposable computing" components like cheap CPU. When something goes wrong in the infrastructure components, it could be just replaced quickly instead of being fixed. All the software state information are designed to be stored separately on servers on the net. JP Morgan separates its software by four levels: user interface, business logic, data layer, and processing layers. The tasks related to four levels used to be run on a single machine. In the "Compute Backbone", all but user interface float on the network. Software and computing power are priced by usage pattern in order to cut cost. This infrastructure also implements on-demand computing and the computing power can be scaled with business demand.

4.4 Collaboration Grid

"Business organizations are, in essence, mechanisms for coordination. They exist to guide the flow of work, materials, ideas, and money, and the form they take is strongly affected by the coordination technologies available" [38]. In the virtual enterprise environment, collaboration among partners becomes even more important. From instant messengers to virtual reality computing tools, many software tools have been used to facilitate communication among geographically distributed enterprises. Researchers need to share expensive experimental data and analyze the simulation results together from different places; engineers hope to work on different components of a CAD project concurrently without disclosing their own proprietary technologies. Grid computing is able to discover and aggregate resources on the Internet and to allow the researchers to manipulate or coordinate the geographically dispersed experimental instruments from their desktops.

Although collaboration is not merely a technology problem, it is certain that the technology barriers in information sharing in the past have seriously impeded the development of good collaboration paradigms. In the virtual enterprises, companies of different shapes and sizes are inclined to share their real-time data with their customers and partners in order to leverage the best practice, improve production design, meet the market requirement on time, and pool intellectual capital on a global basis [39].

In the manufacturing industry, an increasing fraction of the equipment can be automatically controlled through software. The performance data can be collected and published on the networks. Distributed information systems have been developed to connect manufacturing equipment and to share producing data among production stages [40]. However, currently these efforts are mostly focused on collaboration and automation inside one enterprise. The grid-enabled virtual enterprises manufacturing systems aim at improving manufacturing performance in a virtual enterprise by integrating the heterogeneous manufacturing systems and bridging the gap between the higher planning levels (ERP, for example) and the lower distributed production floor control levels (e-factory, for example). Leveraging the coordination frame used to collect computing cycles, collaboration grid is able to coordinate distributed robotic machines to work on one production task. Production capacity becomes an important resource in the manufacturing collaboration grid. These systems based on coordination grid could address major business requirements that are important for the virtual enterprises like generation of production plans that are feasible and optimized with respect to the virtual enterprises, production capacity monitoring, production coordination, remote machine control, access policy management, service agreement, quality control and etc.

Built on the grid-enabled manufacturing systems, massive customized design or manufacture grid service can be published on the grid. The individual customer can use a web interface to access the virtual enterprise manufacturing system to design their own product and display how it would look. Powered by teleimmersion enabled by the grid platform the customer can even test how the product would perform. If the users are interested, they are even able to monitor the production process and control the robotic machines through friendly user interfaces. The products which are attracting for massive customized production include clothing, furniture, and cars. The virtual enterprise manufacturing system will schedule the enterprise resources like material supply, logistics, manufacturing machines, and distribution. Supported by Computer Aided Design (CAD) and Computer Aided Manufacturing (CAM) technologies, this system can streamline the production floor production plan and even automatically control machines like fabric cutters although these robotic machines possibly are dispersed globally and belong to autonomous enterprises. This make-to-order system will reduce the inventory cost dramatically [41]. The following case studies how a trade company can coordinate production globally without owning any factory. Collaboration grid is able to automate all these coordination tasks and improve the efficiency further.

Case: Li & Fung [42]

Li & Fung was founded in 1906 and is an export trading company in Hong Kong. Its source networks are based on mainly East Asia, Indian sub-continent, the Mediterranean, and Caribbean basins. Li & Fung serves the US and European markets with high-volume, time-sensitive consumer goods, such as apparel, handicrafts, and etc.

Rather than a simple mediate import-export trading company, Li & Fung organizes the whole supply chain from raw material procurement to sale of finished product. It provides value-added service through the value chain, from product development, raw material sourcing, production planning, factory sourcing, manufacturing control, quality assurance, export documentation, and shipping consolidation, but without owning any one of the nodes in the chain. The production in this business network is very complicated. For example, "a down jacket's filling, might come from China, the outer shell fabric from Korea, the zippers from Japan, the inner lining from Taiwan, and the elastics, label, Velcro, and other trim from Hong Kong. The garment might be dyed in South Asia, stitched in China, then sent back to Hong Kong for quality control and finally packaged for delivery." Sometimes, when large orders are required to be processed in a constraint time, Li & Fung divides the huge task into small order and manufacturing parellelly in the geographically distributed factories.

This system runs very well under the long-term partnerships with its suppliers. To defend itself from the Internet startup and speed response to the market, Li & Fung built intranet and extranet to facilitate information flow among its business partners. This information platform enables the customers to track the order status and change the order just before process. Lifung.com was also developed to meet the small and medium-sized customers, who only have small orders and were usually ignored. "eSo" was developed to help Li & Fung's suppliers to sell their extra stock online.

What grid can make the above process even better? With production floor real-time information, such as production capacity, status of real-time manufacturing, machine availability, Li & Fung will be more efficient on distributed production scheduling and concurrent manufacturing among its globally distributed business network. The fast grid network connection will enable concurrent design.

5 Contract Net and Virtual Market

The fast development of grid computing technologies and web services enhance the interoperationability among information systems. It will be easier and take short time to integrate heterogeneous systems from either inside or outside of the enterprises. For instance, in the on-demand computing world, an Application Service Provider (ASP) might accept a huge task that may be beyond its processing capacity. How does this ASP build an integration system including all the required services like order processing, production planning, logistics, and human resource management in a short response time? Web service integrators like Jamcracker and a few other startups are beginning to address parts of this field. However, their aggregation services are based on static partnerships and fixed network connections, which are not scalable and flexible.

Researchers in the grid computing resource management community have been working on distributing standard subtasks based on master-worker model. Relying on research on agents, researchers have developed market-based protocols for service negotiation. However, these protocols are focused on computing resource sharing and the processing tasks are relatively simple. Such approaches cannot meet the virtual enterprise computing requirements, which are characterized by varied processing tasks and constrains of business service contracts.

Coordinating distributed resources without central control, using standard interface protocols, and working on a common problem are three key characters of grid computing. The Contract Net protocol provides a solution to coordinate varied distributed resources based on the following design purposes: processing task should be a local process that does not involve centralized control; Contract Net should be a high level protocol independent from physical architectures; and processing tasks have to work on a common problem.

5.1 Introduction of Contract Net Protocol

Contract net protocol is a high level protocol used to solve complicated problems in a distributed environment [43]. Unlike the low level bit stream connection protocol as TCP/IP, Contract Net Protocol focuses on the semantic information transmission and decides what the nodes should say to each other rather than how to say it [43]. The nodes in Contract Net are functionally assigned as managers, contractors or both at the same time for different contracts. A Manager is responsible for looking for contractors, distributing jobs, managing the status of the jobs, and collecting results. A Contractor is responsible for looking for a job actively, running a job, notifying job status, and sending back results. A Manager-Contract relationship is settled by a process of negotiation, which is started either from a Manager publishing a task to many potential Contractors or from a Contractor searching for a Manager with potential jobs. Under both conditions, both Manager and Contractor might receive a list of opportunities. Based on their polices, Managers and Contractors can choose to agree on the deal, prioritize response to different task publish, or explicitly reject the job assignment. A Contractor can even further divide its assigned contract into small contracts and assign them to its sub contractors. Under such conditions, it is no more a Contractor than a Manager.

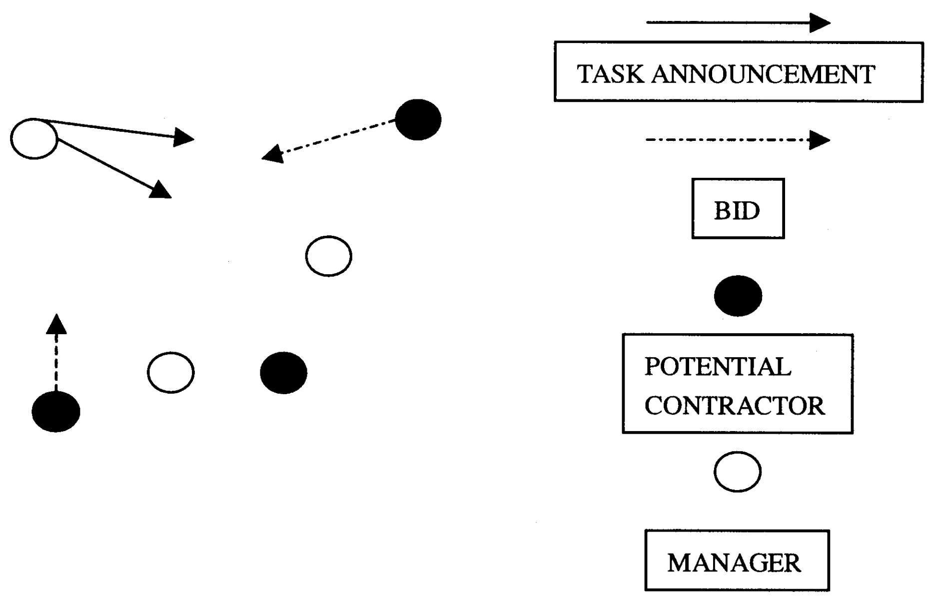

Figure 8 Contract Net

Contract Net Protocol was original designed for sensor networks, which is highly constrained by power sources, processing ability, and communication resources. All of these problems do not exist in the grid computing environment. In the computing grid, a suite of complicate protocols and parsers are required to warrant the efficient and secure network operation.

5.2 Service Discovery

Decentralized systems have the advantages of flexibility, reliability, and scalability. However, because decentralized systems are dynamic and do not have a central server monitoring the whole system status, it is very difficult to find certain resource in those systems.

An intermediate solution is to build a centralized server in the form of directories or e-market. In the PRODNET II [13] design of virtual enterprises system, a Network Coordinator is designed to act as a regulatory component. It is responsible for maintaining the whole network directory information and generating news about the network configuration. This kind of information systems is different from the systems based on Contract Net, which emphasizes on no central control. However, these two approaches are not totally contradictory. The systems based on PRODNET II framework are able to make use of Contract Net Protocol to achieve certain flexibility as well.

The service discovery problem also is a concern in P2P research. Many approaches, such as Distributed Hash Table (DHT), Napster, Domain Name System, and etc., have been designed [44]. Most of them aim to provide a robust resource lookup method in a scalable and decentralized system. These search methods are based on simple attribute information in a loosely controlled environment and are not suitable for the virtual enterprise computing environment, in which resources are highly varied and administrated by autonomous entities. In the virtual enterprises, trust relationships are extremely important in the networks although the e-commerce regulation and security technology will become more mature and the benefits of virtual enterprises will justify the risks.

Based on the above analysis, a partner list is designed into the Contract Net Protocol. Every Manager or Contractor maintain a trust partner list, which might include its sub contractors and partners to compliment their competences. When a Manager wants to contract one job, it first looks up its trust partner list and writes a Request For Information or Request For Proposal notice to them. When a Manager might not have capacities to fulfill a large order, it could quickly query its trusted sub Contractors' real-time capacities through collaboration grid and decide if they can fill this order together. From the other way, the Contractors can also poll its partners, the potential Managers, in its trust partner list and ask for a job. The partner list can be built based on the past cooperation experience or referenced by its trusted partners. A contractor can also choose not to take the job itself but to recommend some of its partners.

The advantages of the above lookup method are, first, it avoids broadcast, which is unpractical for an Internet wide application. Second, it embeds trust relationship inside the negotiation. The disadvantages are: first, it scalable not well; second, a way to prevent the search loop should be designed.

Once the potential partners or required services are founded, the cooperation enters into phase of negotiation.

5.3 Negotiation

After the partner and resource are found, a couple of Manager and Contractor are able to build a direct connection to start the negotiation. Contract Net Protocol, as a high level language, can be semantically very rich. Each message in the Contract Net Protocol has a set of defined information format to describe individual task, whose name has to be standardized in the whole network. However, standardizing task name is not practicable in the Internet. The long-term solution of this problem depends on the development of semantic web and ontology research. The short-term solution is to adopt a certain standard within one industry.

Intelligent agent technologies can be leveraged to automatize the contract process. However, the enterprise tasks are usually very complicated and possess multiple objectives so that this process is very difficult to be automatized. Many market mechanisms such as bidding might be used in the negotiation process. One of computing grid's key characters is sharing resources across borders of autonomous entities. Every entity still controls its own resources and defines its own access policies. This means every node priorities self's interest before the cooperation's and group's interest. Different from traditional researched cooperative agent systems, in the virtual enterprise contract net every agent is assumed to be self-interested. This difference is very important because agents in the virtual enterprise contract net are chasing the maximum of own interest instead of the interest of the whole system. That is, every agent will not victimize own interest. Sometime, one agent might express hostile to another agent.

Recently, the notion of Service Level Agreement has been widely used to define the contract in the IT service industry. In the computing resource sharing level of the computing grid, a protocol was researched to describe the service Level Agreement [45] and is expected to appear in Globus GRAM-2. In the virtual enterprise environment, a business level service Level Agreement Protocol is required but no one has done research in this area yet, partly because the business jobs are too complicated to be expressed in the agent languages. Some part of agreement like decommitment from contract net is especially important for the business applications.

5.4 Virtual Market

Virtual enterprises are more significant for Small and Medium Businesses (SMB) as they are often constrained on resources.

In the virtual market trading computing resources, SMB's IT requirements for xSP are: rent software, limited functionality, low cost, and hosted elsewhere. As software becomes more standardized and modularized, the value in the software industry will shift from products into services. The xSP value chain will consist of networks, platforms, operations, applications, and end services. Grid computing will compliment xSP model by establishing dynamic markets for trading computing and storage resources, hence overcoming the limitations of current static configurations.

Eventually this virtual market in computing grid will also facilitate the information system integration and production resource trading across enterprise borders. This trend will lead the participants of virtual market not merely exchange information but focus on creating collaboration advantages over the competitors [46].

6 Wireless Grid

Although the speed of conversion from atoms to bits is dramatically fast, it still cannot surpass expectations of users. Many applications today are still short of real-time information from the physical world. This requires huge amounts of wireless devices, sensors, and actuators to acquire information. However, communication and resource sharing among these devices are also important in order to achieve a common goal of solving complicate problems. As stated before, grid computing is a platform for collecting and coordinating resources. Therefore, wireless devices in the computing grid platform are also an important research area.

At first glance, it seems that computing grids are unable to be deployed onto wireless devices. This is because grid computing's original application is to tap distributed idle computing cycle. Wireless computing is constrained by power, processing ability, network bandwidth, as well as storage. However, on the other side, when a resource becomes so scarce, it becomes more necessary to share it. Phan et al., [47] strongly argues that wireless grid is a viable solution in the near future.

However, it is worthwhile to recognize that there are different kinds of understanding of wireless grid.

- Wireless device grid access interface [48] is a concept extending the grid console to mobile devices through adapting the user interface to small screen and little keyboards. This mobile access interface has to deal with the connectivity of mobile devices.

- Integrating mobile devices into grids is an approach to actually run computing jobs or store data on the mobile devices. This concept has a much higher challenge than the first understanding. To abstract the mobile device's complicate communication layer and heterogynous platforms from grid service interface, most wireless grid infrastructures use a proxy to distribute jobs and handle connectivity for the wireless devices [47].
- Ad hoc wireless grid can be compared with the normal computing grid. Through some designated protocol, the wireless devices can connect each other and share resources such as wireless bandwidth, storage places, and computing cycles [49]. For example, in field operations like military or disaster recovery, in order to deal with a huge computing task, it is necessary to compose a grid based on hand-held devices; the other possible application is mobile gaming. Wireless grid can enable more powerful game in the mobile devices.

7 Conclusion

7.1 Barriers of Grid Development

While grid computing has the potential to become a revolutionary technology impacting businesses in fundamental ways, the following issues have to be resolved before that becomes a reality.

First, the use of computational grids for solving real-world problems is still limited to research labs and highly specialized scientific applications funded by government agencies. Pushing grids into mainstream computing will first require major advances in **grid programming tools and application development** and the development of standards. Independent Software Vendors will have to come forward and develop applications for the grid.

Table 1 Barriers for Grid Development

Barriers/Challenges to Grid Development	
Grid Programming Tools	Require general programming tools/problem solving tools for developers to write to grid platforms
Application Development	Require critical applications to run on the grid infrastructure
Security	Require multi security domain management system
Intellectual Property	Require regulation for intellectual property publishing on the grid
SLA/QoS	Require the technology and methodology to ensure the SLA and QoS
Bandwidth	Require fast speed Internet connection
Resiliency	Require a self recovery system.
Grid Economy	Require new business models

Second, **security** in such a dynamic computing environment as grid is quite complicated. Grid enables resource sharing across administrative domains, each of which has its own security policies; further, to ensure scalability, there is no central control server. Although grid researchers have implemented a complete grid security framework through all the infrastructure levels, it needs to be tested in the complicated business environment before the grid can be put into broad use. Industry players predict that grid usage will follow the course of credit card usage on the Internet in terms of security and in time, security will be but a small consideration for users of Internet distributed computing - and those few firms with ultra-high security requirements will pay a premium to have their jobs routed to a fixed set of physically secured nodes.

Related to the issue of security is the issue of **intellectual property.** There is a view that only jobs free of intellectual property considerations should be considered for grids since transmitting patented data over a grid could be considered publishing it, which could prevent future patent claims on that data.

Third, how to define a **service Level and Agreement** (SLA) and how to ensure the **Quality of service** (QoS) are still unresolved issues. Currently there is very limited understanding of making SLA and ensuring QoS in an environment with multiple service providers. From the technical perspective, even monitoring the health condition of the dynamic grid is very hard.

Fourth, **network bandwidth** issues will have to be resolved for the grid to be effective in commercial applications. Submissions need to be "throttled" based on network traffic, and intelligent caching of data and applications is a must. Due to the on-demand nature of the computing grid, adaptiveness of network is very important.

Fifty, **resilience, reliability, and disaster recovery measures** have to be developed for the grid to provide zero downtime service.

Finally, a **grid economy** with service providers, products and services, and pricing models will have to develop. Software licensing models will have to change. Issues of access and authorization need to be addressed.

7.2 Grid Market Entry Strategy

Although grid technology has been experimented within academia for a few years, it is still in the early stages of commercialization. The more important participants like xSPs have not been actively involved. They will be the main market drivers in the future.

Because the grid was born in the academic world and was designed in terms of the research environment requirements, the grid will be firstly used in the research and development departments in the enterprises, especially in the life science industry, which involves data intensive computing and has been invested heavily in IT. Other possible first movers are engineering simulation and design departments in the manufacturing industry and applications in the financial industry. After grid technologies become mature, enterprise-wide grid will be deployed. Grid will be used to federate distributed heterogeneous data and help users collaborate in the decentralized enterprises. Eventually after the barriers mentioned above are eliminated, virtual enterprises supported by ExtraGrids will become the reality.

7.3 Open Questions

Observing the above set of evolving scenarios, many persons will ask the following questions.

On-demand computing will not be possible because many companies like Cisco will retain their information systems as their core competences. If everyone subscribes to standard software components from ASPs, what will be its set of differentiators?

This question can be answered from the following perspectives. First, because computing resources are virtualized in the grid, it will not make different whether applications run inside the company or outside the company. Tangible, physical assets will be outsourced but the abilities of creating, nurturing, and deploying key intellectual and knowledge assets will still be kept [9]. Second, grid only provides an easy way to buy and assemble knowledge-embedded, distributed software components from outside. Enterprises still have enough latitude to make use of their knowledge to choose and assemble these resources into grid applications with right partners through building virtual enterprises. Future enterprise information systems will use more pieces of software components distributed by ASPs, as compared to software components embedded into core competences. As discussed earlier, Li & Fung, GE, and Nike outsourced their manufacturing needs and retained their core competences such as market channels, design, and infrastructures to assemble bought components. These companies maintained, not a conventional portfolio of products or businesses, but a portfolio of capabilities and relationships. The challenge is to orchestrate a virtual enterprise in a dynamic, fast-changing resource network [9]. The real challenge lies not in designing the technological platform to coordinate work but in designing the organization structure and processes in order to share knowledge effectively.

Researchers predict the future of grid computing from different perspectives. Different computing vendors companies also contribute to grid computing development from varied ways. Among of them, IBM and HP have been doing grid computing development from different views.

IBM, as the developer and supporter of Globus, focuses on integrating heterogonous computing resource across administrative domain on the Internet.

The researchers at the HP Lab are mainly focused on data center technologies [50]. This concept is more like the computing power plant or computing utility. The HP data center is able to consolidate heterogonous computing resource and implement on-demand resource allocation even in terms of infrastructure topologies.

7.4 Overall Trends

Although commercial use of grid computing is at the early stage, we can still expect significant changes to occur in the IT industry in the near future.

First, consolidation of computing resources will become increasingly important and will receive more attention.

Second, modularity of software will evolve rapidly supported by web service. The degree of interoperation among software modules will become increasingly important and will be supported by grid computing. The value in the IT industry will shift from developing integrated software to understanding business requirements and assembling software modules.

Third, data grid will serve as a better platform to manage huge amounts of unstructured and distributed data as compared to distributed database management systems.

Fourth, in the networked economy, collaboration among enterprises will become increasingly important. Grid computing, as the platform of virtual enterprise computing, will be the ideal platform to coordinate enterprise resources to work on a common business problem.

Reference

1 Simson L. Garfinkel, edited by Hal Abelson, Architects of the Information Society – Thirty-Five Years of the Laboratory for Computer Science at MIT, The MIT Press, 1999

2 http://setiathome.ssl.berkeley.edu/

3 L. Smarr and C. Catlett. "Metacomputing," Communications of the ACM, June 1992

4 Ian Foster, Carl Kesselman, and Steven Tuecke, The Anatomy of the Grid, International Journal of Supercomputer Applications, 2001

5 Ian Foster and Carl Kesselman, The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann Publishers, Inc. 1999

6 Jim Gray, www. Globusworld.org, 2003

7 Carl Beckett and Mohammad Waqas Khan, Trends in IT Outsourcing – A Vendor Perspective, Thesis of Master of Science Management in MIT Sloan

8 Jonathan L.S. Byrnes and Roy D. Shapiro, Intercompany Operating Ties: Unlocking the Value In Channel Restructuring, Harvard Business School Working paper No. 92-058

9 N. Venkatraman and John C. Henderson, Real Strategies for Virtual Organizing, Sloan Management Review, Fall 1998, Volume 40, Number 1

10 G2 Research

11 NIIIP, 1996, The NIIP Reference Architecture, www.niiip.org

12 Walton, J. and Whicker, L., Virtual enterprise: myth and reality, Journal of Control, 22-25, 1996

13 L.M. Camarinha-matos etc, Towards an architecture for virtual enterprises, Journal of Intelligent Manufacturing (1998) 9, 189-199

14 Richard L. Nolan, Cisco Systems Architecture: ERP and Web-enabled IT, Harvard Business School case: 9-301-099

15 F. Warren McFarlan, "VefiFone: The Transaction Automation Company (A) Abridged", Harvard Business School, case 9-397-116

16 Fadisandakly etc, Distributed shared memory infrastructure for virtual enterprise in building and construction, Journal of Intelligent Manufacturing 12, 199-212, 2002

17 C. Garita etc, The prodnet cooperative information management for industrial

virtual enterprises, Journal of Intelligent Manufacturing 12, 151-170, 2001

18 Luis M. Camarinha-Matos and C. Pantoja-Lima, Cooperation coordination in virtual enterprises, Journal of Intelligent Manufacturing12, 133-150, 2001

19 www.platform.com

20 www.avaki.com

21 http://www.datasynapse.com/home/index.jsp

22 www.globusworld.org

23 Ian Foster, Carl Kesselman, and Steven Tuecke, The Anatomy of the Grid – Enabling Scalable Virtual Organizations, International Journal of Supercomputer Applications, 2001

24 Ian Foster, etc., The Physiology of the Grid-An Open Grid Services Architecture for Distributed Systems Integration

25 IBM Red Book, Introduction to Grid Computing with Globus

26 Gartner Research, Data Grids Begin to Emerge, 24 March 2003

27 Vijayshankar Raman, etc., Data Access and Management Services on Grid, http://www.gridforum.org/Meetings/ggf5/pdf/dais/document2.pdf

28 F. Warren McFarlan and Brian Delacey, "Pfizer's Virtual CIO", Harvard Business School N2-303-090

29 Malcolm Atkinson, www.globusworld.org

30 F. Warren McFarlan and Nancy Bartlett, " Postgirot Bank and Provment AB: Managing the Cost of IT Operations", Harvard Business School case 9-302-061

31 http://www.fhrg.fhg.de/

32 Andreas Hoheisel, and Uwe Der, An XML-based Framework for Lossely Coupled Applications on Grid Enrironments, to be published in Lecture Notes in Computer Science

33 Hugh P. Bivens, Grid Workflow, http://vir.sandia.gov/~hpbiven/ggf/draft-bivens-grid-workflow.pdf

34 George Cybenko, Guofei Jiang, and Daniel Bilar, Machine Learning Applications in Grid Computing

35 Clayton Christensen et. al, " Skate to Where the Money Is", Harvard Business Review, November 2001

36 C. Baldwin and K. Clark, Managing in an Age of Modularity, Harvard Business Review, Volume 75, September-October 1997, pp.84-93

37 "JP Morgan Unveils Project Compute Backbone", Waters Advanced Technology issue October 2002

38 Thomas W. Malone and Robert J. Laubacher, "the Dawn of the E-lance Economy", Harvard Business Review, September-October 1998

39 Cathleen Benko and F. Warren McFarlan, Aligning Projects with Objectives in Unpredictable Time-Connecting The Dots, Harvard Business School Press, 2003

40 Howard M. Bloom, Software and Computer Integrated Manufacturing, http://www.mel.nist.gov/msidlibrary/doc/bloom90.pdf

41 John McCarthy, Man and Computer. Proc. Int. Conf., Bordeaux 1970

42 F. Warren McFarlan and Fred Young, "Li & Fung (A): Internet Issues", Harvard Business School case 9-301-009

43 Reid G. Smith, The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver, IEEE Transactions on Computers Vol C-29, No 12, December 1980

44 Hari Balakrishnan, etc., Looking Up Data in P2P Systems, Communications of the ACM

45 Karl Czajkowski, etc., SNAP: A Protocol for Negotiating Service Level Agreements and Coordinating Resource Management in Distributed Systems, 8th Workshop on Job Scheduling Strategies for Parallel Processing, Edinburgh Scotland, July 2002

46 Rosabeth Moss Kanter, Collaborative Advantage: The Art of Alliances, July-August 1994

47 Thomas Phan, Lioyd Huang, and Chris Dulan, Challenge: Integrating Mobile Wireless Devices Into the Computational Grid, MOBICOM'02, September 2002, ACM

48 Condor Grid Computing from Mobile Handheld Devices, Javier Vales-Alonso etc, Mobile Computing and Communications Review, Volume 6, Number 2

49 Mateen Khan and Suresh Singh, Extensible, Reliable Distributed File System for Palmtop Peers

50 Sven Graupner, Vadim Kotov, Artur Andrzejak, and Holger Trinks, Control Architecture for Service Grids in a Federation of Utility Data Centers, www.hpl.hp.com