Chapter 1

The Grid and Grid Network Services

Joe Mambretti

1.1 INTRODUCTION

This chapter presents a brief overview of Grid concepts, architecture, and technology, especially as they relate to Grid networks. Because many excellent explications of Grids have been published previously, including seminal descriptions and definitions [1–3], only a few basic concepts are discussed here.

This chapter also introduces the theme of extending Grid architectural principles to incorporate a wider consideration of communication services and network technologies. This chapter provides a context for subsequent discussions of architecture and methods that allow network resources to be implemented as components that are completely integrated with other resources within Grid environments, rather than as separate, external resources. This integration requires network services to be characterized by a set of attributes that characterize other Grid resources. These attributes are introduced in this chapter, and they are further described in Chapter 3.

1.2 NETWORK RESOURCES AS FIRST-CLASS GRID ENTITIES

Networks have always been essential to Grid environments and, consequently, network services and technologies have been a particularly important topic of discussion within the Grid community. However, network resources have not always been considered what has been termed "full participants" within Grid environments, or "first-class entities": They have not been generally provisioned in accordance with

Grid Networks: Enabling Grids with Advanced Communication Technology Franco Travostino, Joe Mambretti, Gigi Karmous-Edwards

many key Grid architectural principles; they have not been closely integrated with other components Grid environments; they have rarely been implemented as directly addressable, reconfigurable resources.

In part, this circumstance results from the challenge of implementing core network nodes that can be fully administered by Grid processes. This situation is a direct consequence of traditional network architecture and practice, which places an emphasis on stacking services and functions within hierarchical layers, governed by centralized control and management processes. Grid environments are created not as rigid functional hierarchies, but more as collections of modular capabilities that can be assembled and reassembled based on dynamically changing requirements.

Recently, new network architecture and complementary emerging techniques and technologies have provided means by which these challenges can be addressed. These new approaches enable implementations of Grid network resources to be shaped by the same principles as other Grid resources. They also allow dynamic interactions among Grid network services and other components within Grid environments.

1.2.1 WHAT IS A GRID?

Grid networking services are best presented within the context of the Grid and its architectural principles. The Grid is a flexible, distributed, information technology environment that enables multiple services to be created with a significant degree of independence from the specific attributes of underlying support infrastructure. Advanced architectural infrastructure design increasingly revolves around the creation and delivery of multiple ubiquitous digital services. A major goal of information technology designers is to provide an environment within which it is possible to present any form of information on any device at any location. The Grid is an infrastructure that highly complements the era of ubiquitous digital information and services.

Traditionally, infrastructure has been developed and implemented to provide a carefully designed, and relatively limited, set of services, usually centrally managed. Grid architecture has been designed specifically to enable it to be used to create many different types of services. A design objective is the creation of infrastructure that can provide sufficient levels of abstraction to support an almost unlimited number of specialized services without the restrictions of dependencies inherent in delivery mechanisms, local sites, or particular devices. These environments are designed to support services not as discrete infrastructure components, but as modular resources that can be integrated into specialized blends of capabilities to create multiple additional, highly customizable services. The Grid also allows such services to be designed and implemented by diverse, distributed communities, independently of centralized processes. Grid architecture represents an innovation that is advancing efforts to achieve these goals.

As with multiple other technology designs, including those that led to the Internet, the Grid evolved from a need to address the information technology requirements of science and engineering. The Grid is a major new type of infrastructure that builds upon, and extends the power of, innovations that originally arose from addressing the requirements of large-scale, resource-intensive science and engineering applications.

The original architectural concepts developed to address these types of requirements led to the initial formulations of prototype Grid architecture. Early Grid infrastructure was developed to support data and compute intensive science projects. For example, the high-energy physics community was an early adopter of Grid technology. This community must acquire extremely high volumes of data from specialized instruments at key locations in different countries. They must gather, distribute, and analyze those large volumes of data as a collaborative initiative with thousands of colleagues around the world [4].

These requirements, as well as multiyear development implementation cycles, lead to architectural designs that eliminate dependencies on particular hardware and software designs and configurations. Scientists designing and developing major applications cannot become dependent either on static infrastructure, given the rate of technology change, or on infrastructure that is subject to continual changes in basic architecture. They require a degree of separation between their applications and specific, highly defined hardware and configurations.

Many computationally intensive science applications are supported by parallel processing. This degree of separation between applications and infrastructure enables the computational scientists to develop methods for distributed parallel processing on different types of computational systems across multiple distributed domains. Such distribution techniques allows applications to take advantage of large numbers of diverse distributed processors.

Although the Grid was initially developed to support large-scale science projects, its usefulness became quickly apparent to many other application communities. The potential of its architecture for abstracting capabilities from underlying infrastructure provide a means to resolve many issues related to information technology services.

1.3 THE GENERAL ATTRIBUTES OF GRIDS

Many characteristics of Grids are common to other information technology environments, and other characteristics differ by only small degrees. However, some attributes are generally implemented only within Grid environments. Grid environments are usually not designed as rigid hierarchies of layered functions. Instead, they resemble collections of modular capabilities that can be assembled and reassembled based on dynamically changing requirements. This approach does not preclude creating hierarchical stacks of functions; however, its basic design is oriented horizontally across resources rather than vertically. Usually, basic design considerations are the prerogative of the infrastructure designer. Using Grids, these choices can be determined by the application and service designers.

Decisions about the placement of capabilities within specific functional areas are particularly important when creating an architectural model. Recently, architectural designs have tended to allow an increasing number of capabilities within any functional area. Also, determining the layers at which to place functions has been a challenge, and such placement has been a subject of considerable debate. For example, it has been strongly argued that the optimal approach is placing functionality at the edge rather than at the core of systems. Grid environments do not necessarily predetermine or presume a "right answer" with regard to placement of capabilities within functional areas or functional areas within predefined layers. They provide options and allow the communities using the environment to make these determinations.

General Grid characteristics include the following attributes. Each of these attributes can be formally expressed within an architectural framework. Within Grid environments, to a significant degree, these determinations can be considered more art than craft. Ultimately, it is the application or service designer who can determine the relationship among these functions.

(a) *Abstraction/virtualization*. Grids have exceptional potential for abstracting limitless customizable functions from underlying information technology infrastructure and related resources. The level of abstraction within a Grid environment enables support for many categories of innovative applications that cannot be created with traditional infrastructure, because it provides unique methods for reducing specific local dependencies and for resource sharing and integration.

(b) *Resource sharing*. One consequence of this support for high levels of abstraction is that Grid environments are highly complementary to services based on resource sharing.

(c) *Flexibility/programmability*. Another particularly important characteristic of the Grid is that it is a "programmable" environment, in the sense of macro-programming and resource steering. This programmability is a major advantage of Grid architecture – providing flexibility not inherent in other infrastructure, especially capabilities made possible by workflow management and resource reconfigurability. Grids can enable scheduled processes and/or continual, dynamic changing of resource allocations and configurations, in real time. Grids can be used to support environments that require sophisticated orchestration of workflow processes. Much of this flexibility is made possible by specialized software "toolkits," middleware that manages requests and resources within workflow frameworks.

(d) *Determinism*. Grid processes enable applications to directly ensure, through autonomous processes, that they are matched with appropriate service levels and required resources, for example through explicit signaling for specialized services and data treatments.

(e) *Decentralized management and control*. Another key feature underlying Grid flexibility is that its architecture supports the decentralization of management and control over resources, enabling multiple capabilities to be evoked independently of processes that require intercession by centralized processes.

(f) *Dynamic integration*. Grids also allow for the dynamic creation of integrated collections of resources that can be used to support special higher level environments, including such constructs as virtual organizations.

(g) *Resource sharing*. Grid abstraction capabilities allow for large-scale resource sharing among multiple, highly distributed sites.

(h) *Scalability*. Grid environments are particularly scalable – they can be implemented locally or distributed across large geographic regions, enabling the reach of specialized capabilities to extend to remote sites across the world.

(i) *High performance*. Grids can provide for extremely high-performance services by aggregating multiple resources, e.g., multiple distributed parallel processors and parallel communication channels.

(j) *Security*. Grids can be highly secure, especially when segmentation techniques are used to isolate partitioned areas of the environment.

(k) *Pervasiveness*. Grids can be extremely pervasive and can extend to many types of edge environments and devices.

(l) *Customization*. Grids can be customized to address highly specialized requirements, conditions, and resources.

Grid environments can provide these capabilities if the design of their infrastructure is developed within the context of a Grid architectural framework (described in Chapter 3). Increasingly, new methods are being developed that allow for the integration of additional resources into Grid environments while preserving, or extending, these capabilities. For such resources to "fully participate" within a Grid environment, they must be able to support these attributes.

Grids are defined by various sets of basic characteristics, including those that are common to all information technology systems, those that are common to distributed systems, and those that define Grid environments. The general characteristics of a Grid environment described here are those that define basic Grid environments. These characteristics are made possible by the way that resource components are implemented and used within a Grid environment. These individual resource components contribute to the aggregate set of capabilities provided by the Grid. A Grid environment comprised multiple types of resources that can be gathered, integrated, and directly managed as services that can perform defined tasks.

1.3.1 THE GRID AND DESIGN ABSTRACTION

Two key attributes of Grids described in the previous section are those related to abstraction and pervasive programmability. The principle of abstraction has always been fundamental to information technology design. Many important new phases of technology development have been initiated by an innovation based on providing enhanced levels of abstraction. As another phase in this evolution, the Grid builds on that tradition. For example, this abstraction capability makes the Grid particularly useful for creating common environments for distributed collaborative communities.

Grids are used to support virtual organizations. An important benefit of the Grid is its capability for supporting not only individual applications and services but also complete large-scale distributed environments for collaborative communities, thereby "enabling scalable virtual organizations" [2]. Grid developers have always stressed the need to create an environment that can support "coordinated resource sharing and problem solving in a dynamic, multi-institutional virtual organization" [2].

This defining premise has been one of the motivations behind the migration of the Grid from science and engineering to more industrial implementations as well as to other more general domains. Grids can be used to create specialized environments for individuals, large groups, organizations, and global communities. Grids are even being used to support groups of individuals world-wide who are collaborating as if they were all within the same local space – sharing customized global virtual environments.

Grid services abstractions are expressed through standard services definition, middleware, protocols, application programming interfaces, software tools, and reconfigurable infrastructure. These abstraction capabilities are made possible primarily by a set of sophisticated Grid middleware, toolkit suites, which reside between services and infrastructure – separating upper level end-delivered service functionality from lower level resources such as system software, data, and hardware within specific configurations.

Grid application requirements preclude traditional workflow and resource usage, such as those that utilize components as discrete production units. Traditional information technology components have been used as separate components, e.g., computer processors, storage, instruments, and networks. Although these components are connected, they are not integrated.

Grid developers have designed methods, based on services abstractions, for creating environments within which it is possible to discover, gather, and integrate multiple information technology components and other resources from almost any location. Grid architecture provides for an extremely open and extensible framework that makes it possible to create distributed environments using these methods of collecting and closely integrating distributed heterogeneous resources.

1.3.2 THE GRID AS AN ENABLER OF PERVASIVE, PROGRAMMABLE UTILITY SERVICES

The term "Grid" was selected to describe this environment as an analogy to the electric power grid, that is, a large-scale, pervasive, readily accessible resource that empowers multiple different devices, systems, and environments at distributed sites. However, this metaphoric description of the Grid as a set of ubiquitous utility services may overshadow its versatility – its potential for flexibility and reconfigurability. General utility infrastructure is usually designed to deliver a single service, or a narrow range of services. Those services are to be used in the form in which they are delivered. The power Grid is based on a relatively fixed infrastructure foundation that provides a fairly limited set of services, and its underlying topology certainly cannot be dynamically reconfigured by external communities.

In contrast, the information technology Grid can be used to create an almost unlimited number of differentiated services, even within the same infrastructure. The Grid is an infrastructure that provides a range of capabilities or functions, from which it is possible for multiple distributed communities, or individuals, to create their own services.

The Grid is "programmable," in the sense of high-level macro-programming or "resource steering" – providing capabilities for dynamically changing underlying infrastructure. This potential for dynamic change is a primary benefit of Grid environments, because it provides an almost endless potential for creating new communication services as well as for expanding and enhancing existing services. Grid services are self-referential in that they include all information required to find, gather, use, and discard resources to accomplish goals across distributed infrastructure. Grid services are also highly modularized so that they can be advertised to other Grid services and related processes and combined in ad hoc ways to accomplish various tasks. This flexibility is being extended to all Grid resources, including Grid networks.

1.4 TYPES OF GRIDS

Grid architecture continues to evolve as the overall design concepts continue to improve and as it is employed for additional tasks. Grids are often associated with high-performance applications because of the community in which they were originally developed. However, because Grid architecture is highly flexible, Grids have also been adopted for use by many other, less computationally intensive, application areas. Today, many types of Grids exist, and new Grids are continually being designed to address new information technology challenges.

Grids can be classified in various ways, for example by qualities of physical configuration, topology, and locality. Grids within an enterprise are called intra-grids, interlinked Grids within multiple organizations are called inter-grids and Grids external to an organization are called extra-grids. Grids can have a small or large special distribution, i.e., distributed locally, nationally or world-wide. Grids can also been classified by their primary resources and function, for example computational Grids provide for high-performance or specialized distributed computing. Grids can provide modest-scale computational power by integrating computing resources across an enterprise campus or large-scale computation by integrating computers across a nation such as the TeraGrid in the USA [5].

Data Grids, which support the use of large-scale distributed collections of information, were originally developed for the distributed management of large scientific datasets. Many data Grids support the secure discovery, utilization, replication, and transport of large collections of data across multiple domains. For most data Grids, the primary design consideration is not access to processing power but optimized management of intensive data flows. Data Grids must manage and utilize data collections as a common resource even though those collections exist within multiple domains, including those at remote locations [4,6].

Grids continue to integrate new components and innovative methods, to meet the needs of existing and new applications. Application Grids are devoted to supporting various types of applications. Examples include those which support visualization, digital media, imaging, and collaborative communication (such as the Access Grid, a specialized communications environment), storage grids (which support massive data repositories), services grids (which are devoted to general or specialized services), sensor grids, Radio Frequency Identification Systems (RFID) Grids, and security grids. Grids can even exist on a very small scale, for example, across collections of tiny devices, such as electronic motes.

At the same time, new types of world-wide network facilities and infrastructure are being created and implemented to support global high-performance services. For example, “Global Lambda Grids,” which are based on high-performance optical networks, are supporting major science projects around the world [7]. One research project is exploring new tools for scientific research based on large-scale distributed infrastructure that uses advanced, high-performance optical technologies as a central resource [8].

1.4.1 GRIDS AND GRID NETWORKS

Extending general Grid attributes to communication services and network resources has been an evolutionary process. A key goal has been to ensure that these services and resources can be closely integrated with multiple other co-existent Grid services and resources. This close integration is one of the capabilities that enable networks to become "full participants" within Grid environments, as opposed to being used as generic, accessible external resources.

Almost all Grids are implemented as distributed infrastructure. Therefore, from the earliest days of their design and development, Grids have always utilized communications services, especially those based on TCP/IP (transmission control protocol/Internet protocol). Grids could not have been developed without the Internet, a widely deployed, inexpensive data communications network, based on packet routing. As discussed elsewhere in this book, the Internet and Grids share a number of basic architectural concepts.

Many fundamental Grid concepts incorporated new approaches to networking created for specialized projects, such as the innovative I-WAY project (Information Wide Area Year), which was based on an experimental broadband network implemented for Supercomputing 95 [9]. The I-WAY project demonstrated for the first time that a national network fabric could be integrated to support large-scale distributed computing. The software created for that project became the basis for the most widely implemented Grid software used today [10].

However, until recently, the mechanisms that allow networks to be fully integrated into Grid environments did not exist, in part because Grid architectural concepts differ from those that have governed the design of traditional networks. Before the Internet, traditional networks were designed specifically to support a narrow range of precisely defined communication services. These services were implemented on fairly rigid infrastructure, with minimal capabilities for ad hoc reconfiguration. Such traditional networks were designed with the assumptions that target service requirements are known, and that the supporting infrastructure would remain relatively unchanged for many years. Traditional networks were provisioned so that they could be used as resources external to other processes, with minimal capabilities for dynamic configurations or ad hoc resource requests. They have been centrally managed and controlled resources.

The Internet design has been a major benefit to Grid deployments. Unlike legacy telecommunications infrastructure, which has had a complex core and minimal functionality at the edge, the Internet places a premium on functionality at the edge supported by a fairly simple core. This end-to-end design principle, described in Chapter 10, enables innovation services to be created and implemented at the edge of the network, provides for high-performance network backbones, and allows for significant service scalability.

Because the Internet generally has been provisioned as an overlay on legacy communications infrastructure, its potential to support Grid communications services has not yet been completely realized. To enable networks to be utilized with the same flexibility as other Grid resources, Grid networks should incorporate the design goals that shape the larger Grid environment within which they are integrated. Currently, various initiatives are creating frameworks that allow for Grid network resources to accomplish this goal. These initiatives are also beginning to create capabilities that provide for interactivity among multiple high-level Grid services, processes, and network resources. These methods can be used to integrate network resources much more closely with other resource components of Grid environments.

1.4.2 ATTRIBUTES OF GRID NETWORKS

The architecture and methods that are being created for enabling network resources to be more closely integrated into Grid environments are directed at enabling those resources to have the same characteristics as the general Grid environment. The key attributes of Grid network features comprise basic themes for this book, such as capabilities for abstraction, programmability, services oriented architecture, and related topics.

1.4.2.1 Abstraction

One of the most important features of a Grid is its potential for abstracting capabilities from underlying resources and enabling those capabilities to be integrated to support customized services. The Grid architectural model presupposes an environment in which available modular resources can be detected, gathered, and utilized without restrictions imposed by specific low-level infrastructure implementations. This architecture does not specify the complete details of all possible resources, but instead describes the requirements of classes of Grid components. For example, one class of components comprises a few basic abstractions and key protocols that are closest to applications. Another set consists of capabilities for discovering, scheduling, gathering, interlinking, coordinating, and monitoring resources, which can be physical or logical. Another set comprises the actual resources, sometimes termed the Grid "fabric."

The virtualization of resources is as powerful a tool for creating advanced data network services. A major advantage to the virtualization of Grid network functionality through abstraction techniques is increased flexibility in service creation, provisioning, and differentiation. It allows specific application requirements to be more directly matched with network resources. Virtualization also enables networks with very different characteristics to be implemented within a common infrastructure and enables network processes and resources to be integrated directly with other types of Grid resources. For example, low-level functionality within the core of a network can be extended directed into individual applications, allowing applications to signal directly for required network resources.

Using high-level abstractions for network services and integrating network capabilities through Grid middleware provides a flexibility that it is not possible to achieve with traditional data networks. Traditional data networks support only a limited range of services, because they are based on rigid infrastructure and topologies, with restricted abstraction capabilities. General network design and provisioning is primarily oriented toward provisioning highly defined services on specific physical infrastructure, making enhancements and changes difficult, complex, and costly.

1.4.2.2 Resource sharing and site autonomy

The Global Grid Forum (GGF), described in Chapter 4, is engaged in specifying the open Grid services architecture and leveraging the Web Services framework, one component of which is the Web Service Resource Framework (WSRF), also described in Chapter 4. The Grid development communities are engaged in implementing Grid infrastructure software with Web Services components. These components provide access to sets of building blocks that can be combined easily into different service combinations within classes, based on multiple parameters. They can be used to customize services and also to enable shared resources within autonomous environments.

Within a Grid network services context, these capabilities provide new mechanisms for network services design and provisioning, especially new methods for directly manipulating network resources. This approach allows for the creation of customized services by integrating different services at different network layers, including through inter-layer signaling, to provide precise capabilities required by categories of applications that cannot be deployed, or optimized, within other environments. Using these techniques, novel network services can be based on multiple characteristics, e.g., those based on policy-based access control and other forms of security, priority of traffic flows, quality of service guarantees, resource allocation schemes, traffic shaping, monitoring, pre-fault detection adjustments, and restoration techniques.

1.4.2.3 Flexibility through programmability

An important characteristic of the Grid is that it is a programmable environment. However, until recently, Grid networks have not been programmable. This programmability provides a flexibility that is not characteristic of common infrastructure. As noted, network infrastructure has traditionally been designed to support fairly static services with fixed parameters. As a result, network services are costly to deploy and reconfigure, because major changes are primarily accomplished through time-consuming physical provisioning and engineering.

To date, almost all Grids have been based on communication services provided by statically provisioned, routed networks, and the common accessible data service has been a single, undifferentiated, "best effort" service, with minimal potential for service determinism, flexibility, and customization.

In the last few years, several initiatives have been established to create a Grid network services architecture that enables communication services to be substantially more flexible. Using these new methods, Grid network services can be provisioned as "programmable," allowing continually dynamic changing of service and resource allocations, including dynamic reconfigurations. Similarly, these methods make it is possible to initiate processes that can implement instantiations of Grid network services, for short or long terms, with static attributes or with continually changing attributes.

1.4.2.4 Determinism in network services

Because the Grid is flexible and programmable, it allows applications to be matched with the precise resources required. This ability to request and receive required resources and to define precisely matching service levels is called "determinism." Determinism is especially meaningful to Grid networking. Grids have usually been based on common "best effort" data communication services, not deterministic services. Often, the networks on which Grids are based do not provide consistent levels of service, and there have not been any means by which specific levels of service could be requested or provided.

A primary goal of Grid network research is to create more diverse communication services for Grid environments, including services that are significantly more deterministic and adjustable than those commonly used. New methods are being created that allow individual applications to directly signal for the exact levels of network service required for optimal performance. Network service responsiveness, such as its delivered performance, is determined by the degree to which network elements can be adjusted – managed and controlled – by specialized explicit signaling.

Deterministic networking is important to achieving optimal applications performance. It is also a key enabling technology for many classes of applications that cannot be supported through traditional network quality of service mechanisms. This capability includes mechanisms both for requesting individual network services that have specific sets of attributes and also, when required, for reconfiguring network resources so that those specific services can be obtained. This capability is critical for many classes of applications. For example, Grid technology is used to support many large-scale data-intensive applications requiring high-volume, high-performance data communications. Currently, this type of service is not well supported within common Internet environments: large data flows disrupt other traffic, while often failing to meet their own requirements.

1.4.2.5 Decentralized management and control

An important capability for Grid environments is decentralized control and management of resources, allowing resource provisioning, utilization, and reconfiguration without intercession by centralized management or other authorities. During the last few years, various technologies and techniques have been developed to allow decentralized control over network resources. These methods allow Grid networks to be "programmed," significantly expanding Grid network services capabilities. Today, methods are available that can provide multiple levels of deterministic, differentiated services capabilities not only for layer 3 routing, but also for services at all other communication layers.

Some of these methods are based on specialized signaling, which can be implemented in accordance with several basic models. For example, two basic models can be considered two ends of a spectrum. At one end is a model based on predetermining network services, conditions, and attributes, and providing service qualities in advance, integrated within the core infrastructure. At the other end is a model based on mechanisms that continually monitor network conditions, and adjust network services and resources based on those changes. Between these end points, there are techniques that combined pre-provisioning methods with those based on dynamic monitoring and adjustment. Emerging Grid networking techniques define methods that provide for determinism by allowing applications to have precision control over network resource elements when required.

1.4.2.6 Dynamic integration

Grid architecture was designed to allow an expansive set of resources to be integrated into a single, cohesive environment. This resource integration can be accomplished in advance of use or it can be implemented dynamically. Traditionally, the integration of network resources into environments requiring real-time ad hoc changes has been a challenge because networks have not been designed for dynamic reconfiguration. However, new architecture and techniques are enabling communication services and network resources to be integrated with other Grid resources and continually changed dynamically.

1.4.2.7 Resource sharing

A primary motivation for the design and development of Grid architecture has been to enhance capabilities for resource sharing, for example, utilizing spare computation cycles for multiple projects [11]. Similarly, a major advantage to Grid networks is that they provide options for resource sharing that are difficult if not impossible in traditional data networks. Virtualization of network resources allows for the creation of new types of data networks, based on resource sharing techniques that have not been possible to implement until recently.

1.4.2.8 Scalability

Scalability for information technology has many dimensions. It can refer to expansion among geographic locations, enhanced performance, an increase in the number of services offered and in the communities served, etc. Grid environments are by definition highly distributed and are, therefore, highly scalable geographically. Consequently, Grid networks can extend not only across metro areas, regions, and nations but also world-wide. The scalability of advanced Grid networks across the globe has been demonstrated for the last several years by many international communities, particularly those using international networks.

Currently, the majority of advanced Grid networks are being used to support global science applications on high-performance international research and education networks. This global extension of services related to these projects has been demonstrated not only at the level of infrastructure but also with regard to specialized services and dynamic allocation and reconfiguration capabilities.

1.4.2.9 High performance

Because many Grid applications are extremely resource intensive, one of the primary drivers for Grid design and development has been the need to support applications requiring ultra-high-performance data computation, flow, and storage. Similarly, Grid networks require extremely high-performance capabilities, especially to support data-intensive flows that cannot be sustained by traditional data networks. Many of the current Grid networking research and development initiatives are directed at enhancing high-performance data flows, such as those required by high-energy physics, computational astrophysics, visualization, and bioinformatics.

For Grid networks, high performance is measured by more than support for high-volume data flows. Performance is also measured by capabilities for fine-grained application control over individual data flows. In addition, within Grid networks, performance is also defined by many other measures, including end-to-end application behavior, differentiated services capabilities, programmability, precision control responsiveness, reconfigurability, fault tolerance, stability, reliability, and speed of restoration under fault conditions.

1.4.2.10 Security

Security has always been a high-priority requirement that has been continually addressed by Grid developers [12]. New techniques and technologies are currently being developed to ensure that Grid networks are highly secure. For example, different types of segmentation techniques used for Grid network resources, especially at the physical level, provide capabilities allowing high-security data traffic to be completely isolated from other types of traffic. Also, recently, new techniques using high-performance encryption for Grid networks have been designed to provide enhanced security to levels difficult to obtain on traditional data networks.

1.4.2.11 Pervasiveness

Grid environments are extensible to wide geographic areas, including through distributed edge devices. Similarly, Grid network services are being designed for ubiquitous deployment, including as overlay services on flexible network infrastructure. Multiple research and development projects are focused on extending Grids using new types of edge technologies, such as wireless broadband and edge devices, including consumer products, mobile communication devices, sensors, instruments, and specialized monitors.

1.4.2.12 Customization

Just as Grids can be customized to address specialized requirements, new Grid network architecture and methods provide opportunities for the creation and implementation of multiple customized Grid communication services that can be implemented within a common infrastructure. Grid networks based on capabilities for adaptive services, resource abstraction, flexibility, and programmability can be used to create many more types of communication services than traditional networks. New types of communication services can be created through the integration and combination of other communication services.

For example, such integration can be accomplished by integrating multiple types of services at the same network layer, and others by integration services across layers. New services can be also created by closely integrating Grid network services with other Grid resources.

1.5 GRID NETWORKS AND EMERGING COMMUNICATION TECHNOLOGIES

This chapter describes basic Grid environment attributes that could be used as general design goals for infrastructure development, implementation, or enhancement. These attributes are being formalized through architecture being created by standards committees (described in Chapter 4). Various initiatives, including those established by standards committees, are extending these attributes to Grid network resources. Currently, Grid research communities are creating a new architectural framework that will enable network resources to be used within Grid environments as easily as any other common resource. These initiatives are extending current Grid architectural principles, inventing new concepts, and creating new protocols and methods.

These research and development efforts are taking advantage of recent network innovations to accomplish these goals, especially innovations related to data network services and those that are allowing the integration of services across all of the traditional network layers. Given the importance of the Internet, many advanced techniques have been developed for more sophisticated routed packet-based services and for Internet transport. These topics are discussed in Chapters 8, 9, and 10. Advanced architecture is also being designed to take advantage of innovations related to other types of transport networking, including techniques for high-performance switching and dynamic path provisioning. These topics are discussed in Chapter 11.

Other important recent research and development activities have focused on the potential for Grid communications based on lightpath services, supported by agile optical networks, especially those based on dynamically provisioned lightpaths. A variety of optical technologies are being tested in advanced Grid environments to demonstrate how their capabilities can be used to complement traditional network services. These topics are discussed in Chapter 12.

In addition, many emerging communications technologies are being investigated for their potential for supporting Grid networking environments. Various research projects are developing powerful, advanced communication technologies based on innovative technologies, including those related to wireless, free space optics, LED technology, and optical switches. These topics are discussed in Chapter 15.

REFERENCES

[1] I. Foster and C. Kesselman (eds) (2004) *The Grid: Blueprint for a Future Computing Infrastructure*, 2nd edn, Morgan Kaufmann Publishers.

[2] I. Foster, C. Kesselman, and S. Tuecke (2001) “The Anatomy of the Grid: Enabling Scalable Virtual Organizations,” *International Journal of Supercomputer Applications*, 15(3), 200–222.

[3] F. Berman, A. Hey, and G. Fox (2003) *Grid Computing: Making The Global Infrastructure a Reality*, John Wiley & Sons, Ltd.

[4] I. Foster and R. Grossman (2003) “Data Integration in a Bandwidth-Rich World,” special issue on “Blueprint for the Future of High Performance Networking,” *Communications of the ACM*, 46(1), 50–57.

[5] www.teragrid.org.

[6] H. Newman, M. Ellisman, and J. Orcutt (2003) "Data-Intensive E-Science Frontier Research," special issue on "Blueprint for the Future of High Performance Networking," *Communications of the ACM*, 46(11), 68–75.

[7] T. DeFanti, C. De Laat, J. Mambretti, and B. St. Arnaud (2003) "TransLight: A Global Scale Lambda grid for E-Science," special issue on "Blueprint for the Future of High Performance Networking," *Communications of the ACM*, 46(11), 34–41.

[8] L. Smarr, A, Chien, T. DeFanti, J. Leigh, and P. Papadopoulos (2003) "The OptIPuter," special issue on "Blueprint for the Future of High Performance Networking," *Communications of the ACM*, 46(11), 58–67.

[9] T. DeFanti, I. Foster, M. Papka, R. Stevens, and T. Kuhfuss (1996) "Overview of the I-WAY: Wide Area Visual Supercomputing," *International Journal of Supercomputer Applications and High Performance Computing*, 10(2/3), 123–130.

[10] I. Foster, J. Geisler, W. Nickless, W. Smith, and S. Tuecke (1997) "Software Infrastructure for the I-WAY High Performance Distributed Computing Project," *Proceedings of 5th Annual IEEE Symposium on High Performance Distributed Computing*, pp. 562–571.

[11] K. Czajkowski, S. Fitzgerald, I. Foster, and C. Kesselman (2001) "Grid Information Services for Resource Sharing." *Proceedings of the 10th IEEE International Symposium on High Performance Distributed Computing (HPDC-10)*, IEEE Press.

[12] I. Foster, C. Kesselman, G. Tsudik, and S. Tuecke (1998) "A Security Architecture for Computational Grids," *Proceedings of the 5th ACM Conference on Grid and Communications Security Conference*, pp. 83–92.