Designing 21st Century Communications: Architecture, Services, Technology, and Facilities

Joe Mambretti

International Center for Advanced Internet Research, Northwestern University (iCAIR.org), Metropolitan Research and Education Network (www.mren.org)
j-mambretti@northwestern.edu

Abstract. Increasing demand for new applications and services, continuous technology innovation, and rapidly changing economics are motivating the creation of a fundamentally new architecture for 21st century digital communications. Traditional architectural models for communications have been oriented toward meeting exacting requirements of a finite set of well-defined services, essentially, a fixed set of modalities, with well known and well defined parameters. Consequently, this infrastructure has become a restrictive barrier to the deployment of new and enhanced services and capabilities. Meeting the many requirement challenges of continual change requires replacing traditional rigid designs with those that are significantly more flexible and customizable. Often, advances in networking are measured only by increased capacity, and certainly substantially more capacity is required. Fortunately, advanced optical technologies have been created to support 100 Gbps and higher capabilities. However, high capacity alone does not guarantee high performance, and high performance capability does not guarantee required flexibility and determinism. Today, new types of digital communications infrastructure are being designed, prototyped, and provisioned in early implementations. These new design provide for a foundation infrastructure consisting of discoverable, reconfigurable resources that can be dynamically integrated and used. This infrastructure can be considered a programmable platform that can support many more services than traditional deployments, including highly differentiated and deterministic services.

Introduction

Today, there is an unprecedented increase in demand for many types of new applications and services along with accelerating increases in requests for enhancements and extensions of existing applications and services. At the same time, continuous technology innovation is providing many new capabilities, functions, tools and methods, which provide exciting new opportunities to create powerful advanced applications and services. Another macro trend is the changing economics of communications, which is substantially lowering costs at all levels, especially component costs. These decreasing costs allow for more capabilities to be implemented among many additional communities and geographic areas. Collectively, these macro trends are motivating a substantially new architecture for 21st century digital communications.

T. Doulamis et al.: (Eds.): GridNets 2009, LNICST 25, pp. 1–6, 2010.

The traditional architectural development model for communications has been motivated toward meeting the exacting requirements of a finite set of one or a few well-defined services, essentially, a fixed set of modalities, with well known and well defined parameters. This approach leads to the creation of infrastructure that is highly restrictive. Basically, such infrastructure is created to support a few functions extremely well rather than a wide range of capabilities. Furthermore, this infrastructure has been designed and implemented for extremely long lifecycles, with an expectation that major changes would not be implemented for many years. Changes would be undertaken as significant, costly upgrades at a fundamental level. This traditional approach constitutes a substantial barrier to deploying of new and enhanced applications and services.

Meeting the many requirement challenges of continual change requires replacing traditional designs with those that are significantly more flexible and innovative. These new designs provide for a foundation infrastructure consisting of discoverable, reconfigurable resources that can be dynamically integrated, used, and then returned to a resource repository.

These designs provide for communications infrastructure that basically constitutes a large scale distributed programmable platform that can be used to design, implement and operate many types of services, far more than the few based on traditional communication infrastructure deployments, including highly customizable, differentiated and deterministic services. These new designs enable the design, provisioning, and customization of an almost unlimited number of services. These designs provide for not a single centrally controlled network but instead a large scale, distributed facility -- a highly decentralized environment, within which it is possible to create many different networks, each with distinctive characteristics, and each capable of numerous individualized services. These concepts are natural extensions of basic Grid models, i.e., distributed programmable computing environments. Today, the majority of Grids use undifferentiated networks. However, Grid networks extend the Grid concept to network services and infrastructure, creating capabilities for directly addressable programmable networks [1].

Requirement Challenges

Clearly, there are many requirement challenges related to next generation network services. For example, even though there has been much progress in convergence, various major communication service modalities such as voice, video, wireless, and data are still deployed on separate infrastructure, supported by different protocols and technologies. The current, Internet primarily provides a single best effort undifferentiated service, which provides only minimal support for digital media. Traditional Internet architecture and protocols will be challenged as it scales from a service for just over one billion individuals to one that can support over three billion individuals. For example,, Internet security problems are well known as are problems with supporting individual large scale data streams. [2]

These are a few requirements for current general data networking. To preview future networking, it is useful to observe the challenges encountered by large scale data intensive and compute intensive science applications. [3] Large scale, data intensive science applications tend to encounter technology barriers years before they are encountered by other communities. Therefore, the issues that they encounter and the responses to them provide a useful looking glass into the future. Today, one major issue for such applications is bandwidth capacity. Many applications require the gathering, analysis and transport of petabytes of data. Another requirement is a need for dynamic provisioning at all network layers. Another cross cutting requirement at all levels is programmability. Capabilities are required for enabling determinism is be controlled by edge devices.

Meeting Bandwidth Capacity Requirements

Providing for capacity requirements is a particularly interesting issue. Today, there are major on-going increases in capacity demands. Often, advances in networking are measured only by increased capacity. Fortunately, advanced optical technologies have been created to support extremely high volume capacities, including 100 Gbps and higher. [4, 5] These capabilities are beginning to be deployed today. However, high capacity alone does not guarantee high performance, and high performance capability does not guarantee required flexibility and determinism. To fully optimally utilize such capacity, it will be necessary to address a range of issues related to architectural designs, provisioning methods, signally, core components, and others.

New Architectural Designs for Communications

Despite the difficulty and complexity of these challenges, new communications architectural designs methods are being created for meeting them. [1] These designs represent a fundamentally transformation of the traditional models. In the past, many of the most challenging problems in information technology have been addressed by creating architecture and technology that provides for a higher level of abstraction than those used by the current generation. This basic and proven approach is being used to design new architecture for communication services and infrastructure.

Some of these techniques are leveraging developments in other areas of information technology. For example, Services Oriented Architecture (SOA) is being widely used to enhance general services provisioning on distributed and data center infrastructure. SOA enables resources to be highly customized and fine tuned using granulated component identification, selection, integration, and partitioning. Today, a number of networking research organizations are creating customizable SOA based models for communication services, especially to match precisely requirements at different sites with customized services.

Another area of leveraging is the trend toward implementing virtualized services, which mitigate or eliminate restrictions of local implementations and configurations, especially physical hardware. Virtualization transforms infrastructure from rigid physical resources to sets of individual software objects, enabling the creating of distributed programmable platforms. Virtualization can significantly enhance communication services by decoupling services from specific implementations and configurations of underlying hardware and software. Virtualization provides high levels of flexibility in resource discovery, selection, integration, application, reconfiguration, customization, and response to changing conditions.

At the same time that these architectural models are being developed, technology innovations continue to advance at an accelerating rate. New protocols are being created, as are new types of signaling, switching fabrics, routing functions, appliances, optical transport components, and many other innovations. The large range of new innovations gives rise to a reconsideration of the traditional network architectural model.

Deep and Wide 4D Networking

The traditional network architectural model is based on a hierarchical set of defined layers, which for many years has assisted in formulating contextual framework descriptions of the relative placement and relationship of services and functions. Over the last few years, this basic model has been challenged as recent architectural concepts have proposed new types of mid-layer services, processes that enable non-hierarchical transitions among layers, increasingly sophisticated and complex services at each layer, and many more types of services at each layer. If the traditional OSI seven layer model could be considered a horizontal axis, the additional services at each layer could be considered a vertical axis. For example, today many variant transport protocol stacks are being created. It is possible to consider them not as mutually exclusive but as potentially selectable options, resources within a highly distributed programmable communications environment. Individual services could dynamically select specific transport protocols from among multiple available options. Variations among those services could be considered a third (Y) axis, resulting in a 3D model of functionality. The fourth dimension, time, is important to note because of the increasing implementation of dynamic provisioning capabilities at all layers, including dynamic lighpath provisioning. For example, over the last few years, capabilities for dynamic lightpath provisioning have been significantly advanced. [6,7,8] Traditionally, services at different layers and using different protocols have been fairly compartmentalized. However, using new types of abstraction techniques, it is possible to envision high level services being created by combining multiple layers of services and protocols, and even creating those services dynamically and continually adjusting them as requirements and conditions change.

Dynamic Services Provisioning

The benefits of this new model can be demonstrated by noting the advantages of dynamic services provisioning. An especially challenging communication services issue has been providing support for large scale high volume individual data flows while maintaining fairness to other much smaller sized flows. Because traditional data services and infrastructure are designed to support large numbers of individual data flows, managing single high volume flow is problematic. However, using new architectural models that can effectively dynamically reprogram core resources, it is possible to provide services for such data intensive streams, including at the petascale and terascale level.

Edge Signaling

Another implication of this model is that it will enable much more functionality and capabilities at edge sites. This model allows for transition from highly centralized hierarchical management and control systems to highly distributed management and control. Therefore, instead requiring edge sites processes to accept generalized communication services as only designed and delivered, the new approach will provide them with options for customizing those services to meet precise local requirements. In fact, this architecture will allow edge site to create their own customized services using core resources as fundamental building blocks.

Testbeds and Early Prototype Implementations

This new architectural model is being investigated using multiple local, national, international advanced network research testbeds, including by Grid network research communities. However, it is already emerging from research organizations and it is beginning to be implemented within metro, national, and international prototype facilities. These facilities are being used to demonstrate a wide spectrum of innovative, high performance large-scale applications, advanced data services, and specialized networks. Examples of early implementations consist of innovative high performance Grid networks, Cloud networks, such as the Open Cloud Testbed (www.ncdm.uic.edu), science research networks, digital media networks, and extremely large scale high performance computing networks. [9,10,11]

Several national governments world-wide have funded large scale implementations of next generation network testbeds and prototypes. In the US, the National Lambda Rail (NLR) is a national scale distributed facility based on lightpaths supported by an optical network using fiber leased by a university consortium (www.nlr.net). The NLR supports multiple large scale testbeds, including many that are developing new types of high performance services. Internationally, the Global Lambda Integrated Facility (GLIF) is a large scale distributed facility based on lightpaths supported by optical networks provisioned among several continents (www.glif.is). The core nodes of the GLIF consist of multiple international exchange points, facilities that interconnect regional, national, and international research and education networks, and a few corporate research networks. One such facility, the StarLight international communications exchange on the Chicago campus of Northwestern University provides support for over 80 10 Gbps lightpaths interconnecting sites in the metro area, across the state, regionally, nationally and world-wide (www.startap.net/starlight). StarLight is being used to design, implement, and operate multiple innovative advanced communication services for a wide range of advanced applications.

Summary

Emerging 21st communication applications and services have much more aggressive requirements than those that are commonly deployed today. Future applications and services will be much more powerful, flexible, customizable, and reliable. To provide these benefits, fundamentally new communication services and infrastructure architectural models must be created and deployed. Today, in response to these challenges, early architectural concepts are being designed, investigated, and implemented in prototype. The early results demonstrated by these prototypes have indicated that the new architectures models will provide a wide range of new benefits.

References

[1] Travostino, F., Mambretti, J., Karmous-Edwards, G. (eds.): Grid Networks: Enabling Grids with Advanced Communication Technology. Wiley, Chichester (2006)

[2] Gu, Y., Grossman, R.: UDT: UDP-based data transfer for high-speed networks. Computer Networks 51(7) (May 2007)

[3] Special Issue, Journal of Future Generation Computer Systems 22(8) (October 2006)

[4] Mambretti, J., Aoyama, T.: Report of the Interagency Optical Networking Testbed Workshop 3, Networking and Information Technology Research and Development's Large Scale Network Coordinating Group (May 2007)

[5] Petravick, D., Bill Wing, N.G.J.M., Yu, N.R.D.: US Department of Energy Office of Science. In: Workshop Report on Advanced Networking for Distributed Petascale Science: R&D Challenges and Opportunities, April 8-9 (2008)

[6] Mambretti, J., Lillethun, D., Weinberger, J.: Optical Dynamic Intelligent Network Services (ODIN): An Experimental Control Plane Architecture for High Performance Distributed Environments Based on Dynamic Lightpath Provisioning. IEEE Communications 44(3), 92–99 (2006)

[7] Lillethun, D., Lange, J., Weinberger, J.: Simple Path Control Protocol Specification, IETF draft-lillethun-spc-protocol-00.txt

[8] DeFanti, T., Brown, M., Leigh, J., Yu, O., He, E., Mambretti, J., Lillethun, D., Weinberger, J.: Optical Switching Middleware For the OptIPuter, Special Issue on Photonic IP Network Technologies for Next-Generation Broadband Access. IEICE Transactions on Communications E86-B(8), 2263–2272 (2003)

[9] DeFanti, T., De Laat, C., Mambretti, J., St Arnaud, B.: 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 (2003)

[10] Smarr, L., Chien, A., DeFanti, T., Leigh, J., Papadopoulos, P.: The OptIPuter. Special Issue: Blueprint for the Future of High Performance Networking Communications of the ACM 46(11), 58–67 (2003)

[11] Mambretti, J.: Progress on TransLight and OptIPuter and Future Trends Towards Lambda Grids. In: Proceedings, Eighth International Symposium on Contemporary Photonics Technology (CPT 2005), Tokyo, January 12-14 (2005)