Automatic Network Services Aligned with Grid Application Requirements in CARRIOCAS Project

(Invited Paper)

D. Verchere[1], O. Audouin[1], B. Berde[1], A. Chiosi[1], R. Douville[1], H. Pouyllau[1], P. Primet[2], M. Pasin[2], S. Soudan[2], T. Marcot[3], V. Piperaud[4], R. Theillaud[4], D. Hong[5], D. Barth[6], C. Cadéré[6], V. Reinhart[7], and J. Tomasik[7]

[1] Alcatel-Lucent Bell Labs Nozay, 91650 France
Dominique.Verchere@alcatel-lucent.com
[2] INRIA
Pascale.Primet@ens-lyon.fr
[3] Orange Labs
Thierry.Marcot@orange-ftgroup.com
[4] Marben Products
Veronique.Piperaud@marben-products.com
[5] N2Nsoft
Dohy.Hong@n2nsoft.com
[6] PriSM
Dominique.Barth@prism.uvsq.fr
[7] Supélec
Joanna.Tomasik@supelec.fr

Abstract. Automatic Service framework named Scheduling, Reconfiguration and Virtualization (SRV) is developed in CARRIOCAS project to enhance existing Telecom network infrastructures for supporting grid applications sharing IT resources interconnected with ultra-high performance optical networks. From the requirements of Grid applications a classification is proposed to specify the network services and their attributes. In large-scale collaborative environments, SRV solution is described to enable automatic network service operations according to high-performance computing service access. The resources hosted at datacenters are virtualized to be attached to transport network infrastructure offering uniform interfaces towards external customers. New level of service bindings is defined with network services during executions of Grid applications' workflows. On-demand intensive computing and visualization services scenario is described in Telecom environment.

Keywords: network services, infrastructure control and management, virtualization, collaborative applications.

1 Introduction

New management and control functions are required to adapt existing Telecom network infrastructures to deliver commercial IT services for company customers.

P. Primet et al. (Eds.): GridNets 2008, LNICST 2, pp. 196–205, 2009.

Large scale distributed industrial applications –termed as Grid applications – require use of ultra-high performance network infrastructure and multiple types of Grid resources such as computational, storage and visualization. Collaborative engineers need further to interact with massive amounts of data to analyze the simulation results over high resolution visualization screen fed by storage servers. Network operators are looking forward to delivering IT computing services for their company customers with several advantages. First is the practice of Virtual Organization (VO) with the related infrastructure access: IT, Labs or Equipments. The enterprises come together to share their core competencies. An efficient networking solution is key for implementing a productive VO. Secondly collaborations enable to better execute the projects attached to market opportunities. The third motivation is to split the fixed costs of maintenance and the management by outsourcing the infrastructure. But an «Utility computing» model still require standard interfaces [7] at the network service management [4]. The advantage for Telecom operators is the central role in interconnecting datacenters and company customers with the possibility of managing more dynamically the network services to Grid Service Providers (GSP): Fig.1. By enabling the network service management to intercept the grid application workflows, the operator can manage better the resources required to fill end-to-end QoS requirements. These workflows require explicit reservation of network resources with other types of Grid resources (e.g. computational, storage), then intercepting information of the workflows will enable operators to improve the utilization of their infrastructure.

The enterprises will express to GSP their QoS requirements including the maximum cost they are willing to sustain and a time window within workflows has to be completed. On the other side, the infrastructure providers will publish to the GSP their services according to their policies and negotiations to maximize utilization rates together with Telecom and datacenter operators. The binding of customer applications' workflows over connectivity services triggering network service reconfigurations is a capability that Telecom operators are now integrating. This binding requires the vertical service interactions allowing the negotiation of connectivity services between a network operator and GSPs. The companies sharing their datacenters have

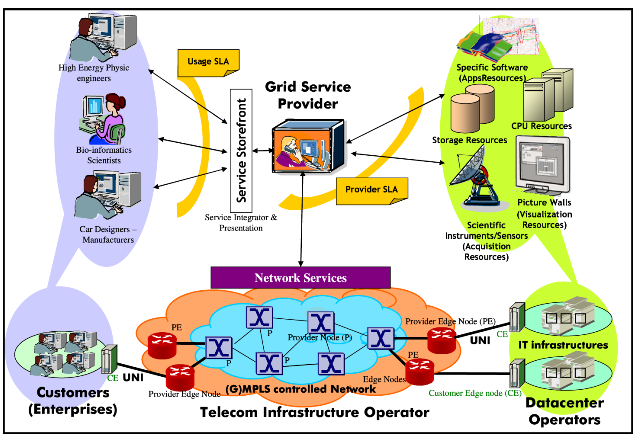

Fig. 1. Actor interactions positioning the central role of Telecom network infrastructure

incongruous infrastructure needs. As described on Fig.1 SLA usages and SLA providers are bound by the GSP. The parameters include the amount of IT + network resources required, the type required, the class of end devices used to deliver the services, the duration, connectivity service, the service accounting.

This paper is organized as follows: Section 2 gives an outline of the grid application classifications and their requirements for defining the network service management layer. Contributions on Scheduling, Reconfiguration and Virtualization (SRV) function with the network architecture to support the delivery of network services for GSPs are presented in section 3 and the SRV functional architecture in section 4. The section 5 presents the management and control function extensions for the provisioning of scheduled connections. Section 6 presents SRV oriented Transport Network scenario considered in CARRIOCAS network pilot. Conclusion is drawn in Section 7 with statements of potential standardization of a Network Service Interface [10] for Telecom infrastructure management and network services.

2 Grid Applications Requirements for the Network Services

It is complex to define a generic classification of all the distributed applications. The focus is on large-scale grid applications requiring ultra-high bit rate network services at the order of the transmission capacity of the network infrastructures. Seven types of Grid applications are listed depending on their connectivity characteristics, capacity requirements, localization constraints, performance constraints and scenarios between networks and grid services offered to end-users [1]. Classification of distributed applications in Table 1 was specified from developers/users contributing in CARRIOCAS, and through specific workshops and interviews [7]. These applications combine the following requirements: (i) Real-time remote interactions with constant bandwidth requirements. (ii) Many large data file transfers between distant sites whose localization is known in advance (distributed storage and access to data centers) and with sporadic bandwidth requirement. (iii) Many data-file streaming between anonymous sites (e.g. multimedia production) requiring statistical guarantees, (iv) Data and file transfers between locations carried on a best-effort network services.

As the connectivity service needs are very different, the network infrastructure must be able to support the class of services with ultra-high performance parameters, including bandwidth and edge-to-edge latency. The connection configuration must be dynamic to be adapted to different grid application usages. For applications which user sites and datacenter sites must be tightly coupled and reconfigured, performance of network infrastructure should make possible to abstract the location constraints of IT resources over a wide area network, this is one of the challenges to be demonstrated. The actors have their objective to maximize their utility functions [7]. For customer, the objective means to execute the submitted jobs at low costs within a required time window. For GSP it means to deliver connectivity services and the other IT services at the highest quality including performance and security and lowest cost. Each operator wants to maximize infrastructure utilization (CAPEX) with the minimal operation efforts (OPEX) to finally obtain best possible return on investment (ROI). These connectivity services require a pool of resources to be explicitly reserved and the resource pools are computed by the network resource planning

Table 1. Classification of Distributed Applications

Type	Characteristics	Capacity Requirement	Localization constraint	Performance constraint	Applications Examples
Type A Parallel Distri-buted computing Tight synchronized computation	Scientific applications involving tight coupled tasks	High	Low	Computational, connectivity latency	Grand scientific problems requiring huge memory and computing power
Type B High throughput low coupling computation	Great number of independent and long computing jobs	high	low	Computation	Combinatory optimization, exhaustive research (crypto, genomics), stochastic simula-tions (e.g. finance)
Type C Computing on demand	Access to a shared remote computing server during a given period	Medium	low	Computational, high bit rate, connectivity latency	Access to corporate computing infra., load balancing between servers, virtual prototyping
Type D Pipeline computing	Applications intensively exchanging streamed or real time data	Medium	Average	Computation, bit rate, latency, storage	High performance visualization, real time signal processing
Type E Massive data processing	Treatment of distributed data with intensives communication between computer and data sources	Low	High	High bit rates and connectivity latency	Distributed storage farms, distributed data acquisitions, Environmental data treatment, bioin-formatic, finances, high energy physics
Type F Largely distributed computing	Research, mo-dification on dis-tributed data-bases with low computing and data volumes requirements	Medium	High	Connectivity latency	Fine-grained computing applica-tions, network con-gestion can cause frequent disruptions between clients.
Type G Collaborative computing	Remote users interacting on a shared virtual visual space	Medium	High	Bit rate and latency (for interactivity)	Collaborative visualization for scientific analysis or virtual prototyping

functions i.e. Network Planning Tool (NPT). Today network services are provisioned in the infrastructure and are not automatically related to how the business rules operate during the publication, the negotiation, and the service notification to the customers. Second, network management system (NMS) must cope with the scheduled service deliveries enabling connection provisioning to be more autonomic. The extensions of NMS are to ensure that infrastructure delivers the on-demand or scheduled network services asked by external entities such as the GSP. The network services must be adjusted in response to the changing customer environments such as new organizations joining and then connecting on network + datacenter infrastructures (e.g. car designer is going to be connected to Complex Fluid Dynamic simulation application the next month during two weeks) or a connected organization that is going to stop to use the IT services during the next week. The CARRIOCAS network has to provide dynamic and automatic (re)configuration of the network devices according to the roles and the business rules of different actors.

Provider network services (respectively connectivity services) are described from the network infrastructure operator (respectively the customer) point of view which are distinct from the solutions implemented through the use of customer edge (CE) node based solutions [4]. Virtual Private Network services (VPN) provides secure and dedicated data communications over Telecom networks, through the use of standard tunneling, encryption and authentication functions. These services are contrasted from leased lines configured manually and allocated to one company. VPN is a network service dedication achieved through logical configuration rather than dedicated physical equipments with the use of the virtual technologies. To reconfigure automatically the provisioning of VPNs, the SRV should accommodate addition, deletion, moves and/or changes of access among datacenter sites and company members with no manual intervention. In CARRIOCAS network pilot, three classes of connectivity services named Ethernet Virtual Circuit (EVC) are proposed to customers: point-to-point EVC, multipoint-to-multipoint EVC and rooted multipoint EVC. The description of each EVC includes the User-to-Network Interface (UNI) corresponding to the reference point between the provider edge (PE) node and CE node. At a given UNI more than one EVC can be provisioned or signaled according the multiplexing capabilities. CARRIOCAS network envisions the reconfiguration functions for wavelength connections thanks to the reconfigurable Add-Drop Multiplexers (R-OADMs) node and the network control functions. GMPLS controllers enable to establish the wavelength switched connections on-demand by communicating connectivity service end-points to PE node embedding a dynamic protocol message exchange in the form of RSVP-TE messages.

3 SRV Oriented Transport Network

To establish automatically network services, the connectivity requirements of the grid application workflows gathered by the GSP are intercepted automatically by the network service management functions. CARRIOCAS transport network architecture is extended to position the functional components of the network service management layer between the GSP and the network resource management and control functions.

The network service management functions are organized in the Scheduling, Reconfiguration and Virtualization (SRV) service components separately from network resource management and control functions. The SRV part of the network infrastructure operations is central and integrates different service interfaces according to the entities it interacts with. The SRV north bound interface dependent of the technology implemented at the GSP interfaces enable the SLA provider to be described. Towards the network infrastructure, the SRV south bound interface is based on two options: SRV requests for switched connections through the provider UNI (UNI-N) if it is a GMPLS controlled network [2] or SRV requests for permanent connections or soft-permanent connections through the Network Management Interface based on MTOSI from Tele-Management Forum [3]. Connections are dynamically established (provisioned through NMS or signaled through UNI) with guaranteed QoS defined in SLA provider. In the first version, the selection of computing and storage resources is performed by the grid application middleware hosted at end-users or GSP, but externally of the network infrastructure management. The SRV East-West interface supports edge-to-edge connectivity service covering multiple routing domains. The SRV interfacing the GSP manages the network service at the ingress network domain and the interdomain connectivity services through peer operation information exchanges with the SRVs involved in the chain [10].

By integrating the concepts of virtualized infrastructures (datacenters and transport networks), the virtualized services can be delivered by the second version of SRV. A virtualized service is a composition of computing, storage and connection as a managed composite service. Further specific software applications can be virtualized and integrated as an element of the composite service too. The virtualized services require explicit reservation of different types of resources offering the SRV the

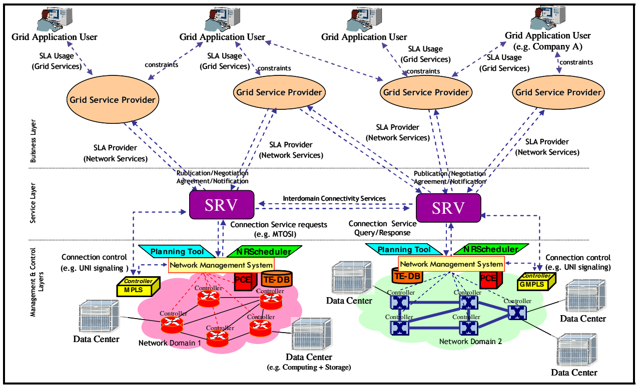

Fig. 2. SRV oriented Transport Network for Grid Service Provider

possibility to co-allocate them. After the resource reservation phase is complete and before the allocation phase starts, cross-optimization can be executed through different types of resources. During the selection of one computational server at a datacenter location to execute a workflow, the virtualized service routing functions will take into account heterogeneous criteria including the computational capability of the server (e.g. Operating Systems, cycle speeds, RAM), the available network capability (interfaces, bandwidth, latency) on the selected connection path. Such an extension requires an additional interface between the SRV and the Grid Resource Management System (GMS) at the datacenter infrastructures (not represented on Fig. 2) to exchange management/control information between grid application and network services, which capability is not supported in standard infrastructures [2]. However several technical challenges still need to be overcome such as the amount of information to be managed, the security and confidentiality constraints between the network operators and the other business actors, the reconfiguration time delay of different resource types of the virtualized infrastructures.

4 SRV Functions and Scheduled Network Management

The SRV internal architecture is composed by the functions to deliver connectivity services fulfilling the SLA requests towards a GSP. Three internal functional layers identified are: the GSP interface, Network Management System interface and the mediation layer as represented in Fig. 3.

The SRV functional components interacting with the GSP are the service publication, service negotiation and service notification. These components allow publishing, negotiating and notifying respectively the connectivity services from the SRV towards external GSP customers. This interface is based on the VXDL description language [6] of the services (connectivity services or virtualized services) compliant with the specifications of the Open Grid Service Architecture (OGSA). Based on XML, VXDL has the advantages to enable the composition of multiple service elements by the GSP (e.g. a computing service combined with a connectivity

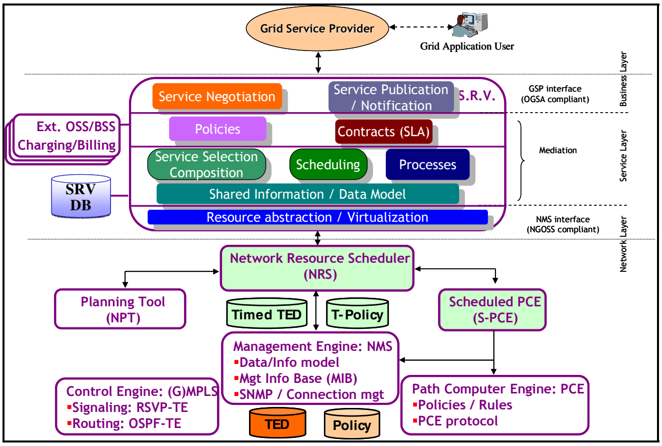

Fig. 3. The SRV functional model

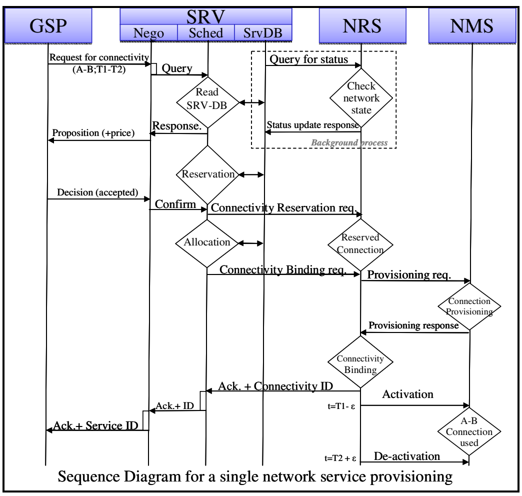

Fig. 4. Connectivity service request and network service provisioning sequence diagram

service). The network services are separated from the description of the connectivity services and are associated to the NMS to provision the resources. The interactions between GSP and SRV are based on Customer - Producer pattern including the SRV database. Three types of interactions between SRV and GSPs are supported: publish/subscribe query/response, and notification. The SRV acts upon a request initiated by GSP, the query/response interaction pattern is implemented (see Fig. 4). Other types of negotiation sequences are also considered as evaluated in [5].

The service publication function is attached to the creation of a SRV-DB that is the repository of the information related to the states of the connections, from which the SRV chooses a class of connectivity service according to the GSP requirements. The high-level description of the service offerings exposed by the SRV belongs to the class of *CustomerFacingServiceSpec* of the connectivity service specifications that are the service functions that GSP can request and purchase to the network operator [3]. The parameters of the connectivity services provided by the GSP in the SLA provider are derived from the grid application workflow.

The connectivity service negotiation enables the GSP to negotiate for the *QoSService* and other attributes such as confidentiality. The sequence associated to a SLA for connectivity services is complicated by the need to coordinate the (re)configuration of multiple elements simultaneously in the network infrastructure. At the NMS interface layer, the network infrastructure is virtualized according to the *Service* class of the shared information data (SID) model [3]. The abstraction depends of the connectivity services published to the GSPs. The network services are generalized according to Lx-VPN, with x=1, 2 or 3. The basic elements described belong to the class of *ResourceFacingServiceSpec* of the network services. At the mediation layer of the SRV, the selection of network service elements and their composition enable to bind the GSP based connectivity service descriptions according to the rules defined by the network operator on provisioning (Policies). The Policy component insures the continuity from the GSP connectivity service requirements to the connection requests towards the NMS interface. The rules select the appropriate process to provision the connections via the NMS or to signal the establishment of the connections via the ingress node controller. The bulk data transfer scheduling (BDTS) service enables scheduling of network service end-points by providing time guarantees that a specified amount of data streaming is transferred within a strict time window [8]. The service Scheduling (Sched) component interfaces with network resource scheduler (NRS). The NRS is an extension of the NMS to integrate the time constraints expressed by the GSP according to the execution of the grid application workflows. NRS can signal scheduled connections e.g. Label Switched Paths for RSVP-TE signaled connections. The Scheduled-PCE (S-PCE) is an extension from the Path Computing Element architecture [9] and its PCE Protocol (PCEP) to enable the selection of the available network resources according to the time window constraints. Each connection provisioning request from the NRS is characterized by a source node, a destination node, a QoS (bandwidth and availability), a starting time noted START, and a ending time noted END. The times START and END define a time window managed by NRS for resource allocation. The provisioning sequence for one or several connections is divided into four phases: (i) resource request phase (e.g. between NRS and network elements), (ii) reservation phase, (iii) intermediate phase during which modification of the reservation is possible, and (iv) resource allocation phase during which the network service is provisioned as recommended in [2].

5 Use-Case and CARRIOCAS Test-Bed

A lot of R&D projects require massive high-performance computing (HPC) powers generating large amounts of data which push enterprises to access storage. The data storage demands are increasing because of the meshing resolutions (3D, time, temperature) of their applied mathematics methods. It is often required to process the raw data, to visualize the results on visualization means and to interpret them immediately and to restart a new cycle of application workflows. These types of scenarios named VISUPIPE are developed for industrial companies connected to CARRIOCAS network [6].

The BDTS service (Sched) allows reserving in advance network services with negotiated QoS and then allocating each service between two or more access end-points [8]. The Grid resources are located at different points. The GSP workflow requests are modeled as a graph. Each vertex characterizes a task and QoS requirements (computational, storage and visualization), each edge represents the connectivity QoS requirements (bandwidth and latency). In the first version when datacenter resources are selected the vertices of the graph are defined, then GSP requests for connectivity services to SRV. The sequence diagram (Fig. 4) presents a scheduled connectivity request served by a connection provisioned with NMS. The sequence diagram of a switched connection differs slightly. The background process is monitoring the states (available or used) of the connections on the network and logs their states in SrvDB. When the connectivity service is bound to the network service delivered by the NRS (i.e. after the connectivity binding stage), the NRS returns a connectivity service ID to SRV that can be used to reference the connectivity service request from GSP. Before the grid application VISUPIPE workflow starts (i.e. T1-ε) the provisioned connection is activated by the NRS. Similarly after the time T2 negotiated by the GSP and accepted by SRV, the NRS deactivates the connection provisioned by the NMS. The connection state is changed to available.

6 Conclusion

SRV oriented network infrastructure provides multiple network services automatically through a generic network service interface according to the connectivity requirements of the grid application workflows. Integrated services rely on the virtualized infrastructures with the connectivity service storefronts and the transport network infrastructures owned and operated by different organizations. Immediate or scheduled network services will have to be supported. The specifications of a standardized network service interface will facilitate the delivery of on-demand connectivity services and their integration with other services delivered by datacenter infrastructures [10]. CARRIOCAS aims to demonstrate a reconfigurable transport network with 40Gb/s optical transmission links providing connectivity services between the enterprise clients to remote datacenter servers to run computing and data intensive applications. In the first part of the project, after having analyzed the requirements of the distributed applications and evaluated the optimal network system architecture, the second phase of the network and application server infrastructure deployments is complete. The specifications of the SRV are complete and the design phase is under progress. Beyond the production network pilot used for the advanced tests and the experimentations, one of the objectives is to define new commercial usages of virtualized infrastructures. Further the CARRIOCAS is a collaborative platform for the design and the analysis of complex numerical systems and is fostering on new business models.

Acknowledgments

The authors thank all the partners of CARRIOCAS project including O. Leclerc (Alcatel-Lucent), L. Tuhal, J.-L. Leroux, M. Chaitou and J. Meuric (France-Telecom), X. Gadefait (Marben Products), J.-M. Fourneau and N. Izri (PRiSM), L. Zitoune and V. Vèque (IEF), M. Gagnaire and R. Aoun (Telecom Paris-Tech), I. Wang and A. Cavali (GET/INT), E. Vieira and J.-P. Gallois (CEA-list), C. Mouton (EDF) and J.-C. Lafoucrière (CEA Ter@tec) provided valuable inputs for this paper. CARRIOCAS project gathers 22 institutions (industrials, SME's, and academia) and is carried out in the French cluster "system@tic Paris-Region".

The project is funded by the French Industry Ministry, the Essonne, Hauts-de-Seine and Paris general councils: http://www.carriocas.org

References

1. Audouin, O., et al.: CARRIOCAS project: An experimental high bit rate optical network tailored for computing and data intensive distributed applications. In: APOC 2007, Wuhan, China (November 2007)
2. Requirements for Automatic Switched Transport Networks (ASTN), ITU-T Recommendations G.807/Y.1302 (July 2001)
3. Multi-Technology Operations Systems Interface (MTOSI) 2.0, TMF Forum (May 2008)
4. Barroncelli, F., Martini, B., et al.: A Service Oriented Network Architecture suitable for Global Grid Computing. Personal Communication (October 2004)
5. Cergol, I., et al.: Performance Evaluation of a SLA Negotiation Control Protocol for Grid Networks. In: Gridnets 2008 ICST Conf., China (2008)
6. Koslovski, G., Primet, P., et al.: VXDL: Virtual Resources and Interconnection Networks Description Language. In: Gridnets 2008 ICST Conf., China (2008)
7. Utility computing enhanced with high bit rate networks: Opportunity and requirements. In: Gridnets 2007 workshop, Lyon (October 2007)
8. Primet, P., et al.: Supporting bulk data transfers of high-end applications with guaranteed completion time. In: ICC 2007, Glasgow (June 2007)
9. Farrel, A., et al.: A Path Computation Element (PCE)-Based Architecture. IETF RFC4665 (August 2006)
10. Network Service Interface, Open Grid Forum-GHPN working group OGF24 Singapore (September 2008), http://www.ogf.org