

# Extensible Service Provision in Grid Networks: A Case For Resource Visibility and Inter-Domain Exchange

Slobodanka Tomic and Admela Jukan\*

Vienna University of Technology, Institute of Communication Networks, Favoritenstr. 9/388  
A-1040 Vienna, Austria, mailto: {slobodanka.tomic, admela.jukan}@tuwien.ac.at

\*Admela Jukan is now with Georgia Institute of Technology, in Atlanta, GA.

**Abstract—** In this paper, we discuss opportunities and challenges in the network control plane design, critical to the provision of *network-assisted extensible Grid services*. We define a network-assisted extensible Grid service as an atomic infrastructure and operational entity created as control and signaling layer artifact. As such, it provides for dynamic configuration and interconnection of multiple types of distributed resources for high-performance applications (e.g. computing, data storage and visualization) over globally distributed network systems. To this aim, we introduce two concepts: network resource visibility and inter-domain exchange. The network resource visibility is a piece of resource allocation information used to dynamically manage globally distributed resources. The concept of inter-domain exchange is based on what we define as the "Grid" Exchange Point (GXP), which is the equivalent of the Internet Exchange Point (IXP) in the data transport layer (e.g. GMPLS-enabled). With GXPs, service provision can be extended to Grid networks consisting of arbitrarily and dynamically interconnected network domains. We discuss issues critical to the services' creation, provision and operation, and, in particular, for network scenarios where different applications share the same, "virtualized", physical network infrastructure. We show that the concepts presented here are viable in realistic network scenarios and quantify the benefits in service provision performance.

## I. INTRODUCTION

A growing number of high-performance scientific and industrial applications, ranging from tera-scale data mining to complex weather forecast and financial modeling, are increasingly taking advantage of the infrastructure of large geographically-distributed computing, network and data management resources, commonly referred to as "Grid". By capitalizing on a variety of complementary, and often separate, research activities - ranging from the design of user-friendly Grid portals, to the support of numerous application-specific layers, to the implementation of the Grid middleware services, and seamless interactions with intelligent network services – Grid networking will soon be able to provide universal access to what is often referred to as "world-wide-computing/collaboration/visualization".

In Grid-based science toolkits, that typically include architecture of the graphical user interface and middleware, the traditional network functions, such as forwarding, routing and switching, are designed to be sensitive and responsive to the application needs. This is implemented by means of control

and signaling between the application-specific layer and network layer. Many of the proposed toolkits also include advanced features such as real-time resource monitoring, support of QOS guarantees and reliable high-speed data transfer [3]. Typically, these advanced features are used at the time of application invocation, and networks provide for dedicated resources, which remain static for the duration of the services. In this context, various control and signaling protocols have been proposed, differing in the amount of modifications needed to incorporate the requirements for high-performance applications and the level of network infrastructure awareness [2]. Several Grid networking frameworks, such as Quanta, Terascope, ODIN, [1], recommend the network control to be adaptive to the applications needs and, at the same time, take advantage of new networking technologies. Specifically, these frameworks use intelligent optical network services enabled thru the emerging architectures being developed in the standard bodies, i.e. IETF Generalized MPLS, GMPLS [2, 3]. The integration of application specific software technologies with the network control, where a new set of control capabilities will support application-driven adaptive networking, has just begun, and a lot of research issues remain open such as the methods of interworking of distributed Grid resources with the new networking technologies as they are introduced.

In this paper, we discuss opportunities and challenges in the design of such application-adaptive network control plane architectures, and, in particular, in the provision of *network-assisted extensible Grid services*. By enabling dynamic configuration and interconnection of any type of granularity, capacity and connectivity, as dictated by applications, and assisted thru the network control plane, extensible Grid services broaden the network control plane requirements, similarly to what has been defined for generalized virtual private network (GVPN) services in [6,7]. Fig. 1 depicts a simplified, vertical layered structure utilized in this paper to describe the network-assisted extensible Grid services. As used here, the latter are an atomic infrastructure and operational entity created as *control and signaling layer* artifacts (denoted as gS in Fig. 1). Grid *applications specific layers* (ATLAS, ALICE, GENIUS [1]) implement their high-level middleware to invoke extensible services (gS) that allocate network resources according to application logic and architectural requirements. One gS can, through the control layer, extend to different domains, change in time, and be used by more than one application. This exten-

sibility feature is fundamentally important to Grid applications, which typically change over time, e.g., from no network utilization to the enormous network utilizations. Moreover, one extensible grid service represents a hierarchy of services, with various levels of granularity. For example, in Fig. 1, gS-2 can be created as a mesh to cover all existing Grid applications. For applications that require higher granularity, a "sub-Grid" service can be created. At the *network layer*, similar to what has been proposed in [8], we "virtualize" the physical network infrastructure for service provision. While model allows for various applications to properly exploit the high bandwidth and run their specific transport protocols (e.g. RUDP [3], SABUL [9], etc.), it also allow for commercial Internet traffic to co-exist within the same physical network infrastructure.

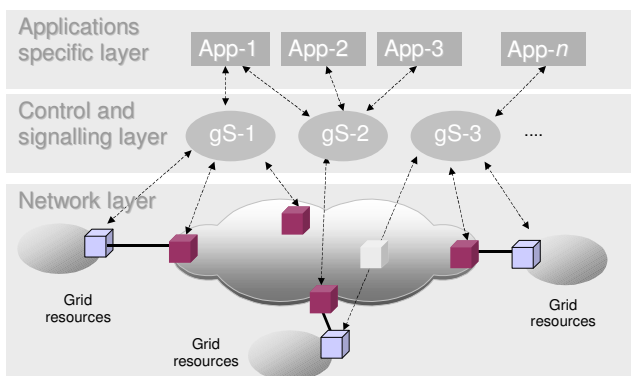


Figure 1. Vertical layered structure for the network assisted extensible Grid services.

The novel contribution of this paper is twofold. First, we introduce the concept of *network resource visibility*, i.e., a critical piece of resource allocation information used in the control and signaling layer to adjust the use of network resources to the applications needs. Second, we consider service provisioning in a novel scenario consisting of *arbitrarily and dynamically* interconnected network domains. By adopting proven Internet concepts into GMPLS-enabled networks, we address the "missing link" in the Grid networking "big picture", i.e., the concept of inter-domain exchange. The latter is based on what we define as the "Grid" Exchange Point (GXP<sup>1</sup>), which is the equivalent of the Internet Exchange Point (IXP), moved down to the lowest possible, data transport layer (e.g. GMPLS-enabled optical layer). Based on [10], we briefly present the control plane-enabled interface called Multi-Provider Edge, MPE, and discuss the implementation of the routing service and mediation of inter-domain resource utilization.

Despite our strong reference to the concepts developed under the GMPLS framework, this paper does not advocate GMPLS control plane as the best alternative for integration with Grid network application middle-ware services or toolkits. In fact, the concept is general enough to be applied for the IP layer networks, if the future Internet architecture will allow for the

proposed network control (see [8]). Although the GMPLS framework can provide some intuitions about the feasibility of our concepts, given its maturity and its intrinsic consideration of heterogeneous network technologies, our study suggests that appropriate design choices and service performance benefits could only be gained thru the experience of building the network control architecture with a Grid application prototype.

The rest of the paper is organized as follows. Section II describes the network architecture reference model and the concept of resource visibility. Here, we present a new type of Grid service-enabling interfaces (MPE), and present the concept of inter-domain exchange. Section III illustrates four possible service provision scenarios that show how the GMPLS-enabled network and Grid service can signal for extensible resource management. Section IV presents several illustrative examples for performance evaluation related to the efficiency of extensible service provision. We conclude the paper and discuss open issues in Section V.

## II. ARCHITECTURE

Our reference architecture is shown in Figure 2. For illustration, we assume that two network assisted Grid services (or simply "services") are provisioned: gS-1 (green) and gS-2 (orange). We represent gS-1 as a service that connects three geographically distributed computing, storage, and visualization and/or application invocation sites. Service gS-2, has four client sites (e.g. collaborative environments). This architecture is building upon what has been defined in [6] for GVPN<sup>2</sup>. Also here, the multi-domain approach is used to capture a variety of architectural features. Similar to the concept of LSP regions [11], domains can represent layers in which only interfaces of a specific switching or computational capability are included (e.g. IP routers, Internet switches, optical nodes). For example, in Figure 2, the domains D1 and D2 can be packet switched, TDM-granular LSP regions, while D3, D4 can be a Lambda-granular LSP region. The multi-domain model is not only appropriate to account for the multi-domain network environment, but it can also be used to represent a variety of Grid resources. Heterogeneous resources (computer clusters, storage, visualization sites), can also be modeled and represented as "domains". For example, gS-1 service internally may be represented with a computational, storage, and a visualization domain.

Akin to the VPN taxonomy, we will use notions of Client<sup>3</sup> and Network Edge (CE, PE), to refer to the high performance

<sup>2</sup> Since VPN services are typically understood for specific technologies and protocols, e.g. IPsec, a more general service description may be more appropriate, similar to the notion of *virtual meshes* [8]. The virtual meshes (virtual networks or supranets) are used as the core abstractions of resource management in value-added service networks, where the providers can efficiently manage and control the resources in a "customized management" way.

<sup>3</sup> The taxonomy related to our reference models deserves a separate study and is outside the scope of this paper. Here we assume that a "client edge" (or user, computer, cluster, or service member) is any party that participates in a Grid service, but does not give its resources for usage to other existing Grid services. If resources have to be shared with the other clients, these clients either become service members, or a client "virtually" separates its resources

<sup>1</sup> The GXP is analogous to the GMPLS-XP we introduced in [10].

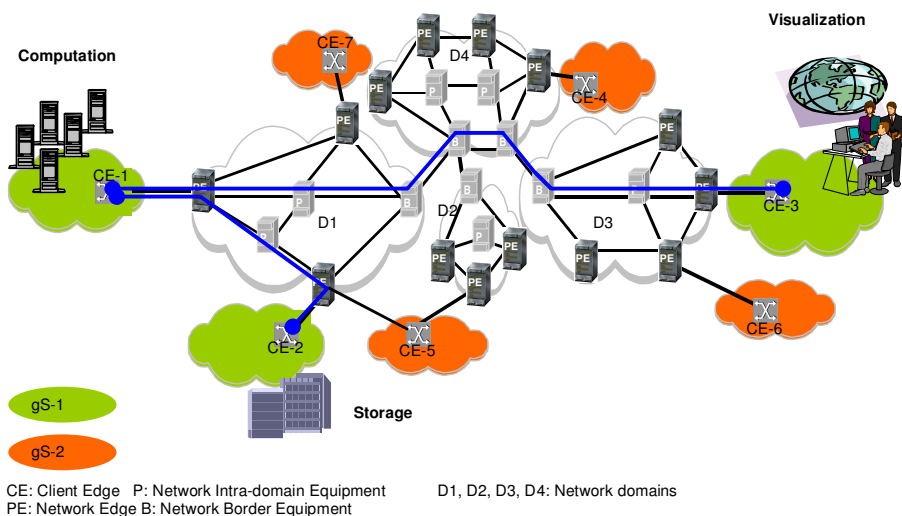


Figure 2. The reference architecture.

equipment on the client side (CE) and to networking resources (PE). Within one administrative domain we will refer to the network intra-domain network resources (P) and domains network border elements (B). The intra-domain and border resources (P, B) have the common feature that they do not interconnect to any client resources. The network border resources (B) play an important role in inter-domain provision and service membership discovery<sup>4</sup>. All edge nodes (“service members”) and network resources involved in service provision create the *Grid service domain*.

The *extensible Grid services* model presented here adopts the concepts from GVPN service definition from [6] and adds to it the properties of global, multi-domain resource reachability, as well as extensibility of the allocated resources during the service lifetime. As in the case of GVPN services, service operation is characterized by two phases: membership discovery and connectivity. In the membership discovery phase, the mapping between CE and PE nodes is made known globally within the service domain. At the same time, the existence of all service members is made known service-internally, i.e. among the service members. In the connectivity phase, each service member establishes service links to other members. Here, we enhance the connectivity phase by the actions to allow *service extensibility*. These actions assist the timely adjustment of resource granularity, connectivity, and service reach, by taking into consideration the applications’ require-

ments and efficient utilization of the network resources. To customize the service granularity, either the service-internal granularity or the granularity supported at the CE-PE interface, or both, can be considered. For example, in Figure 2 we can assume that the interface between CE-1 and PE-1 is a WDM multiplex, over which multiple clients can establish lambda LSPs in order to access computational resources. Over the same lambda LSPs, two TDM-granular LSPs, one to the application site and one to the storage site can be established if service gS-1 internal data exchange dynamics does not require an entire lambda LSP.

<sup>4</sup> In this paper, we do not consider the issue of ownership either [20]. Client and the infrastructure provider (network) can be owned by the same party (e.g. a research network). However, A party (A) can take only one role at a time, i.e. either “client” or “provider”, with respect to another party (B).

To achieve the extensible service features, we next present two new concepts. First, we introduce the concept of *network resource visibility*, i.e. a critical piece of resource allocation information, that network control plane makes available to the Grid application. Second, we present a novel scenario for service provisioning in the network of *arbitrarily* and *dynamically* interconnected network domains, i.e. the “Grid Exchange Point”, GXP.

#### A. Resource Visibility

The notion of *resource visibility* as introduced here is used to describe the globally distributed resources in the service domain. We consider two basic parts of the resource visibility information. The first is supported service-internally, and refers to the resource visibility information that is accessible to all service members. For example, gS-1 in Figure 2 can make the distributed cluster of computers fully or partially visible with respect to the storage and application invocation sites. The second part of the resource visibility information is service-external and can be negotiated and agreed upon among service members and involved network domain. For example, gS-1 can use the TDM-granular LSPs to connect to the application site, but an entire lambda LSP to connect to the storage site.

Let us also consider the example of the service gS-2 (Figure 2). Assume that gS-2 can require lambda LSPs in the network domains with service-internally supported granularity set {TDM, LSC}. Although the service is subscribing to the whole wavelength, if domain D2 makes its TDM-granular resources temporarily visible on that wavelengths, the clients participating in the service gS-2 could internally decide whether to establish a new lambda-LSP or use available TDM-granular resources more transmission efficiently. This feature is particularly interesting for Grid applications. In some cases, e.g. bulk file transfer, the application could gradually “slide” from one computational site highly utilized to the less utilized one. For example, application can gradually migrate from the site CE-4, has increased 70% to the computational site CE-7 (utilization 35%), without service

interruptions. In other cases, e.g. remote visualizations, such adjustments are less desirable during the service life-time and the lambda circuit has to be maintained unchanged for the whole duration of the service life time [12].

Resource visibility is not equivalent to, but incorporates the concepts of control plane interworking models, e.g. peer or overlay. In other words, different levels of integration, ranging from overlay to peer, can provide reduced or full resource visibility. (We will address this relationship more in Section III where the *visibility graphs* are introduced.) For example, service gS-1 can have both TDM and LSC resource visibility information at its disposal. Given that Grid networks mostly operate according to what is closest to the peer interworking model, if the negotiated switching capability is lambda (LSC), the applications can establish a path based on the visibility of the LSC-capable resources. However, network can make only a limited amount of all LSC-capable resources “visible” (e.g. only 30% of all LSC capable resources are visible).

Service extensibility and resource configurability depends on the availability of resource visibility information. This availability is not only dependent on the applications’ awareness of the network resources, but it also depends on the characteristics of the access element, i.e. interface, between the client edge (CE) and network (domain) edge (PE). The GVPN concept [6], supported by BGP for membership discovery and GMPLS for signaling, enables flexible set-up/termination of connections between service-enabling ports without involving configuration changes in network’s domain. We believe that a more flexible solution would be a specific configuration is “negotiated” between application and network. For example, in Figure 2, gS-1 could have a choice to use up to W channels in the WDM multiplex on the corresponding links between the computing and storage site. How a specific configuration can be “negotiated” between the application and network without manual intervention is open to research.

### B. Inter-Domain Exchange

A further important enabler of service flexibility involves the domains’ interconnections, i.e. domain-domain (B-B) interfaces, and CE-PE multi-homing interfaces (e.g. CE-5 in Figure 2). On these interconnections, higher flexibility than that available in traditional, statically provisioned back-to-back links can be achieved, resulting in multi-domain architectural flexibility. To this aim, the concept of multipoint-to-multipoint interfaces in Grid networks will play an increasingly important role in the future, and will be critical to extend the reach of the Grid services. For example, in the reference scenario shown in Figure 1, between domains D1 and D2 as well as Grid sites CE-2 and CE-5, a new interface can be inserted, such that each connected CE can select the most suitable network domain when invoking a service. On the other hand, a remote client can also chose a network domain to improve reachability with respect to a remote Grid resource site. This new interface, hereafter referred to as Multi-Point Edge (MPE), can enable new features by introducing proven Internet concepts into the world of transport networks, i.e. the con-

cept and the application analog to those of the Internet Exchange Point, i.e. “Grid” Exchange Point. For example, with switching function incorporated in MPE, applications have an intelligent choice of domains (and, hence resources reachable thru these domains).

In [13], we proposed the architecture for MPE in the multi-provider scenario and compared it to the traditional architecture with static back-to-back links. We evaluated advantages of the provider-neutral maintenance of manifold Service Level Agreements (SLA), service monitoring capabilities and consistency of network views. For example, the local policies and single client/server (or, customer/provider), SLAs can be imported into the MPE, and translated into “higher level SLAs”. SLAs migrated to MPE can be modified without invoking a huge amount of routing information updates, particularly if the modification does not affect other coexisting SLAs.

In [10], we introduced the concept of GMPLS-XP. Figure 3 illustrates how functional flexibility of the GMPLS-XP can be exploited even further by integrating it with the MPE routing service. Via the MPE, the GMPLS-XP imports the routing information and policies from the connecting domains. As a result, it may collect the complete routing and policy information and, therefore, act as a routing proxy for the connected domains. For example, the policy can specify the metric for which the least-cost is to be achieved, the preference list of domains, etc.

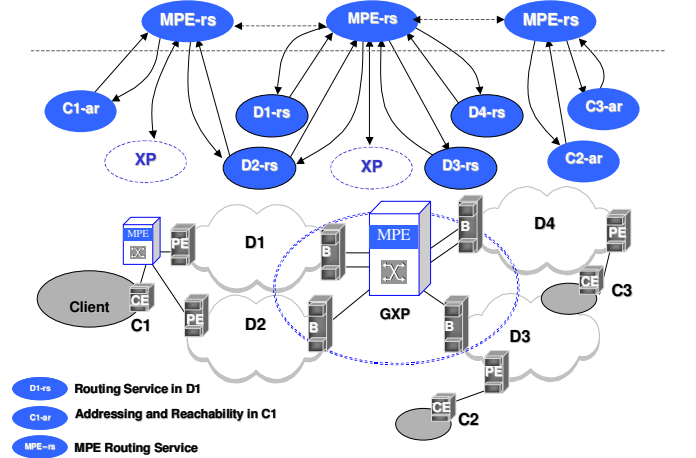


Figure 3. GPX with the embedded MPE.

### C. Exchange Points Revisited

Surprisingly, relatively little research has been reported so far related to the class of problems pertaining to multi-domain issues. Two major flavors of multi-granular provider-provisioned Virtual Private Network (VPN) services have been identified and proposed for further standardization within the IETF: (i) the Generalized MPLS/BGP VPN (GVPN) defined in [6] reuses proven concept of MPLS/BGP that utilizes BGP for distribution of the VPN information and MPLS tunneling; (ii) the Virtual Optical Cross-Connect Service (VOXC) defined in [7] reuses the concept of the “virtual

router"-based VPN service. Defined as a generic "network-in-network" service, GVPN/VOXC is applicable for different provider-customer relationships.

The necessity for flexible inter-connections clearly exists in other networks, both wireline and wireless. In the conventional wireline IP networking, at the Internet Exchange (IX) or Internet Business Exchange (IBX) [14, 15], the Autonomous Systems (AS) are statically interconnected either directly or through a layer 2 switch. In the architecture of all-IP wireless (or mobile) networks, the concept of the exchange has its representation in GRX Exchange Points where the providers of the GPRS Roaming Exchange service (GRX) [16] are interconnected according to GRX peering agreements. GRX plays the crucial role for users' roaming, enabling not only the global connectivity but also new mobile VPN services.

In the optical network domain, by introducing the "distributed exchange" concept based on the optical BGP [17], the CA\*net4 research network has proposed an important approach for the global optical fiber network. In the CA\*net4, the institutions interested in direct interconnecting acquire dark fibers from different network carriers and connect to the optical cross-connects of the optical core of the CA\*net4 network. The optical BGP distributes the reachability over the optical core. When one institution wants to establish a direct peering session for high-bit-rate applications to another reachable institution an IP BGP session is initiated and the lightpath is established between these two institutions. The Lightpath Route Arbiter is the component responsible for the lightpath establishment. When a direct lightpath is established, the interconnected institutions can establish BGP peering sessions between them. In this way, the optical core acts as a re-configurable distributed exchange point.

In the MPLS community, the application of exchange points is gaining attention. In [18] the exchange architecture based on MPLS technology called MPLS-IX was proposed. MPLS-IX is data-link independent and can unify two IX architectures prevailing today, being (i) Local Area Networks (LANs), such as FDDI, Ethernet or Gigabit Ethernet, and (ii) Permanent Virtual Circuits (PVC) ATM. The MPLS mechanisms are used to configure virtual back-to-back links, or back-to-back LSPs between interconnecting domains. Over those virtual interconnections, traditional bi-lateral peering models can be deployed.

In Grid networks, the exchange points and the mechanisms for inter-domain inter-working are crucial for operating and expanding the global network, and for provisioning of services with global reach. The GXP exchange architecture will become even more important, if the process of interconnection set-up is integrated in the on-line network operation supported by the mechanisms of the control plane. With GXP applications, service can globally extend reach, remotely discover computational and network resources. In addition, it will facilitate grid networks interoperability and its co-existence with the commercial applications.

### III. STRATEGIES FOR EXTENSIBLE SERVICE PROVISION

We will now present a few service provision strategies that illustrate how the concepts of resource visibility can be used. To this aim, we propose that the so-called *visibility graphs* are created in the service membership discovery phase, and are then used to determine LSPs, which as a result of service-internal traffic dynamics have to be set-up or released. The creation of the visibility graphs as described here follows the same principle regardless of whether the network is deploying the concept of multi-point edges. However, with MPEs in place, the full potential in deploying the resource visibility information can be achieved.

Examples that follow are based on the reference architecture in Figure 2, where we assume that supported internal granularities are {TDM, LSC} for both gS-1 and gS-2. In other words, every service can support applications that are using optical paths with either TDM or LSC granularity (or both). As illustrated in Table I, for each scenario we consider (i) service-internal granularities, e.g. as dictated by the applications; (ii) granularities supported by the network, e.g. LSC, TDM; (iii) interworking models, e.g. peer or overlay<sup>5</sup>; (iv) routing approaches, e.g. "discrete" or "combined". When the network switching capability is LSC, we refer to these cases as *reduced visibility* scenarios (in Table I, in scenarios a, b, and c). In scenario (d), the network is making visible all {TDM, LSC} resources, and we refer to this case as *extended visibility*. The TDM and LSC visibility graphs corresponding to each scenario are given in Figure 4.

In scenarios (a) and (b), either the TDM-capable resources or LSC-capable resources are visible. In scenario (a), network-internal resources are not visible due to the overlay model ("clouds"). In scenario (b), the LSC-graph includes all advertised network LSC-capable resources (peer).

Figure 4c shows the scenario with reduced visibility. The LSC-graph created in this scenario is the same as in scenario (b). However, the TDM-graph is considerably different. The TDM graph shows the same topology as the LSC graph, however the links are now specially marked (dotted in Figure 4c). The dotted edges represent LSC-only capable links marked as *transit*. Transit is used to refer to those LSC-links that does not start and end at the TDM capable interface. In other words, only a *concatenation* of dotted links that starts and ends at the TDM capable interface is considered as a TDM-granular resource that can be used for a TDM-LSP set-up.

It is this concatenation that we can use to *extend* the service reach if, for example, a path cannot be found within the TDM graph. In fact, in the *extended visibility* scenario (d), dotted edges are also included in the TDM-graph. Here, in addition to the LSC-only capable resources, network advertises also

---

<sup>5</sup> Grid networks today mostly operate according to the "peer model". However, without advocating the "carrier cloud" model from the commercial world, we believe it is important to take the overlay model into consideration given that Grid services can expand over commercial and/or carrier-owned networks.

the TDM-switching. Thus the TDM-graph includes the transit links as in scenario (c), but also the TDM Traffic Engineering (TE) links (an IETF standard terminology, see [11]), being established in the previous cycles of path provision actions or

being advertised by the network. We will explain next in more detail how the visibility graphs are used for services link set-up/release and we will also explain how “discrete” and “combined” routing can be performed.

TABLE I. FOUR REPRESENTATIVE SERVICE PROVISION SCENARIOS.

Scenario	Service	Service Internal Granularity	Network Resources Granularity	Interconnection-model	Routing	Visibility
Fig 3 (a)	gS-1	TDM, LSC	LSC	Overlay	Discrete	Reduced
Fig 3 (b)	gS-1	TDM, LSC	LSC	Peer	Discrete	Reduced
Fig 3 (c)	gS-1	TDM, LSC	LSC	Peer	Combined	Reduced
Fig 3 (d)	gS-2	TDM, LSC	TDM, LSC	Peer	Combined	Extended

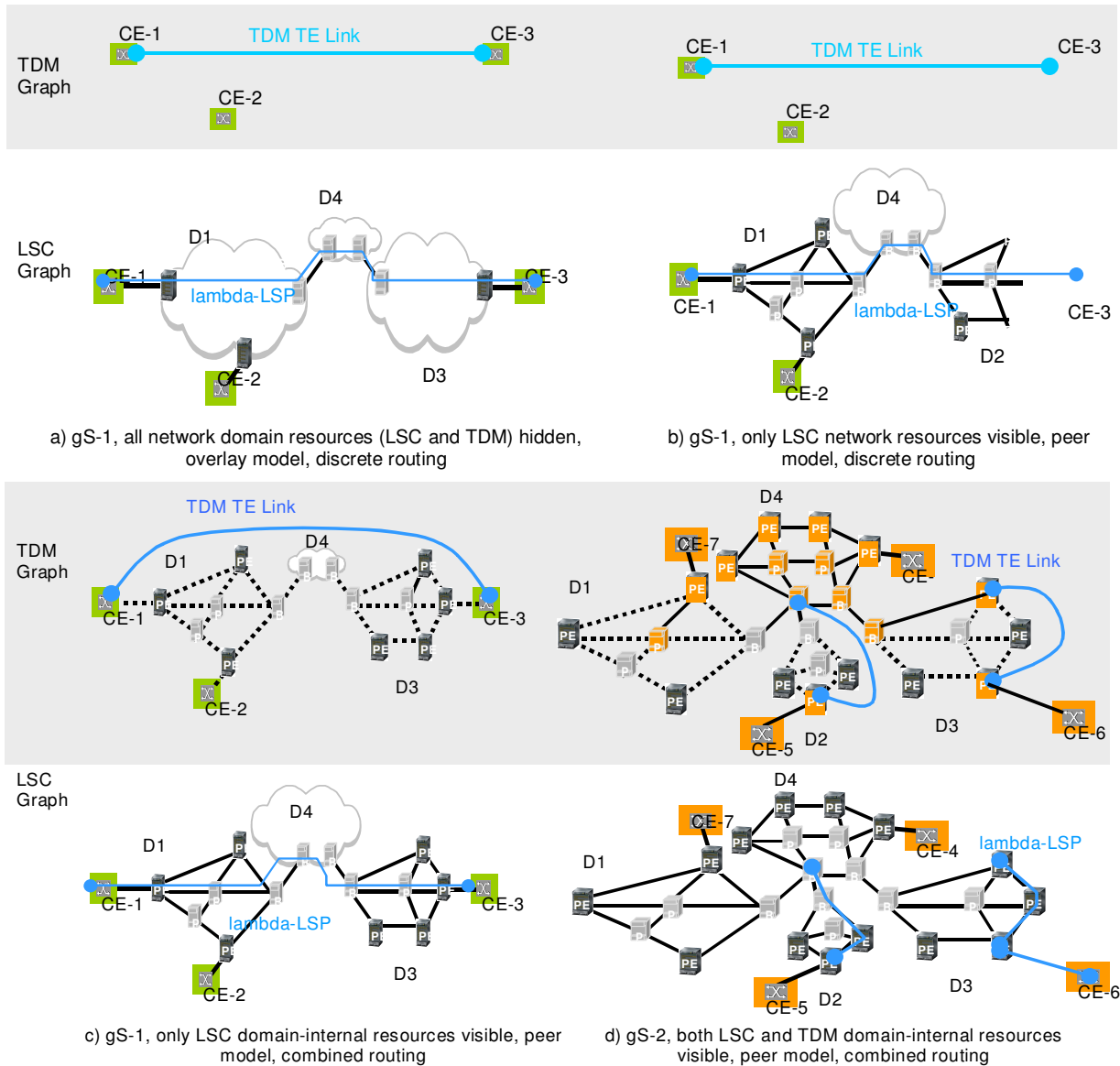


Figure 4. Visibility graphs used to for extensible service provision as defined in Table 1

### A. Connection set-up

For the reduced visibility scenarios (a) and (b), the connectivity phase is characterized by discrete (layer-by-layer or graph-by-graph) routing, where TDM and LSC visibility graphs are considered separately (“discretely”). First, the TDM-graph is used for selecting the path of the requested TDM-LSP between two service members (e.g. storage and computing entities in gS-1). The connection set-up can follow a specific strategy, such as max-granularity (TDM). If a path is found, the requested TDM-LSP can be set-up. If such path does not exist in the TDM-graph, or, if during the service life time the application request for more bandwidth or additional connectivity, the service can be *extended*. In that case, the LSP-source (CE-1) starts a traffic engineering action by deciding where to extend the lambda-LSP. This decision can be made according to the path availability in the LSC-graph. If a path can be found in the LSC-graph, the corresponding lambda-LSP can then be signaled. Upon the successful set-up of the new lambda-LSP a new *TE-link* will be inserted in the TDM-graph between the source and the destination.

The blue links in Figure 4 refer to the newly established TE links (extended). After the new TE link is inserted, routing in the TDM graph becomes trivial as it involves only one-hop. Here, the service link that has triggered the TE action uses either the whole TE link capacity or a subset of it. In the latter case, the remaining capacity remains available for the future requests. Finally, the process of TE link set-up supports the connectivity establishment and modification during the service lifetime. For example, if a file-transfer application requires an extension of the capacity or increase in the transmission rate, this action can be performed in-situ by activating signaling protocols supporting the service (reach, capacity) extension.

Slightly different is the case of reduced visibility when *combined routing* is applied (Figure 4c). In contrast to discrete routing, the combined routing considers both TDM and LSC graphs jointly. With combined routing, application “always finds” the explicit path. “Always found” implies that a path in the TDM graph is searched with the awareness of available capacity, but without knowing the granularity of the LSP that has to be invoked to set-up that path. Consequently, the resulting path can consist of multiple segments containing concatenated of transit (dotted) links. Similarly to scenario (b), these path segments (lambda-LSPs) require to be set-up first. In this case, the cost of the TE links is also crucial. The TE links already established should be assigned a cost lower than that of the corresponding lambda-LSPs. In all scenarios with reduced visibility, a new TE link is added between the request source and the destination, independently of the service-internal traffic dynamics. If the capacity of the resulting TE link is not fully utilized, reserving entire lambda-LSP can prove cost inefficient for a service.

Scenario (d) illustrates the extended visibility with combined routing. Due to extended visibility, in addition to transit (dotted) and TDM TE links created service-internally, the TDM

graph contains also the TDM-capable resources. The request for a TDM-LSP is again “always found” in the TDM-graph. The shortest path found in the TDM graph can now include segments that are (i) concatenation of available transit links (dotted), (ii) TDM capable resources, and (iii) TDM TE links. In Figure 4d, TDM TE links with capacity higher than required (i.e. larger than  $b$ ) are given as solid black and blue, if established over one LSC-capable transit link and over their concatenation, respectively. For a lambda-LSP that supports a concatenation of transit links, the new capacity should be preferably bundled to a TE link if such already exists. Otherwise, a new TE link in the TDM visibility graph will appear. Lambda-LSP set-up actions also affect the LSC graphs such that the capacity of all affected LSC-links have to be updated.

It is worth noting that the considerations presented above strictly relate to the scenarios defined in Table I. For example, we assumed that each service link request corresponds to a TDM-granular LSP. The requests for lambda LSP are also possible but those require LSC visibility graphs only and, thus, do not illustrate more granular (or, extensible) service requirements. Other scenarios, such as extended visibility with discrete routing, are also possible but are not illustrated here. How the visibility graphs are created, updated and used in different scenarios is a strategic design choice that depends on specific cases considered.

Finally, we could not discuss the potential burden in implementing the above concepts in the existing protocols. Our expectation is that the services and applications will be able to automatically run signaling protocols capable of adjusting the GMPLS-controlled lambda paths with respect to capacity or reachability needs of a particular application, while at the same time considering the network resource efficiency. We also believe that above architecture can be equally feasible for scheduled and on-demand invocation of grid services.

### B. Connection release

Extensibility does not mean that the service is always “expanding”. In some cases, applications may need to reduce the resource utilizations for variety of reasons, e.g. scheduling of a high priority application. We define three strategies for connections release referred to as: “*never release*”, “*release when idle*” and “*scheduled release*”.

With “*never release*”, the service connectivity and reach is never reduced, and the links are never released for the whole duration of service, even if applications are actually not using them. The “*never release*” strategy is safe for critical applications, but is resource efficient only if the idle resources are made visible for the future traffic.

With “*release when idle*”, service links are released as soon as they are not being utilized. This strategy can help to better share the wavelengths on the WDM links.

The third strategy, “*scheduled release*” works as the following: the resource scheduler can mark unused links as “not to be used for further requests”. In this case, the unused TE link would be avoided in future requests, and would eventually,

after a scheduled resource availability checking, labeled as unused, and released. This strategy can support graceful pre-emption.

#### IV. PERFORMANCE

We will now illustrate the network performance impacts when service extensibility strategies are used. Our test network comprises three interconnected GMPLS-enabled network domains (Figure 5). Each of the domains' nodes is connected to a Grid resources or users' site, represented by one CE node. In all domains, each node and each client access implement LSC and TDM switching capability. We furthermore assume that the capacity of the established TDM TE link (i.e. the capacity of one established lambda-LSP) is four, e.g. one lambda-LSP results in  $4 \times \text{STM-48}$  TDM containers (Figure 4b). All intra-domain links are single fiber WDM systems with 16 wavelengths each, whereas the B-B links have 64 wavelengths.

Figure 5 shows the traditional network topology, where network domains are interconnected at several nodes through static interconnections (B-B links). When used, the MPE topology is created when multi-point edge (MPE) instances are activated (here: 6 randomly placed, dashed circles). In contrast to back-to-back static connectivity, MPE can arbitrary interconnect all collocated parties (any-to-any). As for the routing, we assume that routing can take into account link busy states. The shortest path from source to destination in different domains is a concatenation of shortest paths in each single domain that it traverses.

We first study the efficiency of WDM link utilization for service strategies under the same load conditions. The network topology without MPEs is assumed. Since the definition of various Grid service is not relevant for these results, we make an assumption that all clients belong to one single Grid ser-

vice. The requests for service links are on-demand, i.e. arrive and last according to negative exponential distribution. The requests are not symmetric, but each client is equally probable to generate requests to any other client and we do not distinguish between type of Grid resources, i.e. whether computational, storage, etc. With respect to the network domains, we distinguish between requests that occur *intra-domain* and *inter-domain*, the latter ones making 10% of the total requests. We assume that each intra-domain service link allocates the resources of its corresponding domain only. If these resources are exhausted, the request is blocked.

For different combinations of visibility, link-release (LR) strategy and routing (discrete, combined), we define eight scenarios in which the distribution of WDM link occupancy is evaluated (Figure 6). We use the link occupancy to measure the load distribution. The higher the link occupancy for a certain load, the lower is the efficiency of link utilization and consequently the probability that a new connection request can be accommodated. In all scenarios, we assume the peer interworking model. The Y-axes show the percentage of WDM links with the link occupancy of 0%, 1-20%, 21-60%, 61-80%, 81-99%, and 100%. As it can be observed from Figure 5, all strategies perform similarly relative to each other within a domain. For example, RED-LBR-NR shows in all three domains higher number of links with link occupancy over 60% than the RED-C-REL strategy. High link occupancy can be an important operational objective for multi-administrative environments. Most significantly, the "release" strategy introduces considerable improvement for the options in visibility and uses the available resources in the most efficient way.

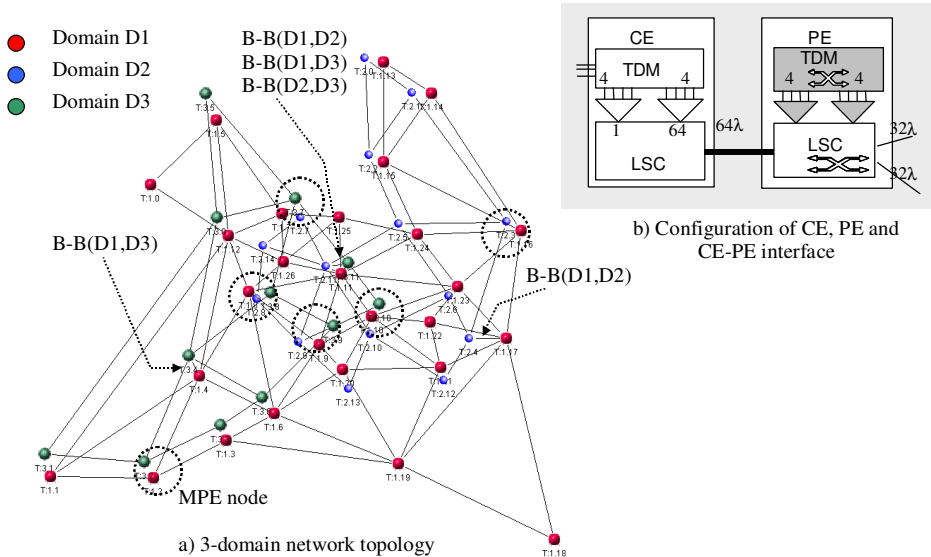


Figure 5. (a) Example topology with 6 MPEs (dashed circles), and (b) supported switching capabilities at LSRs.

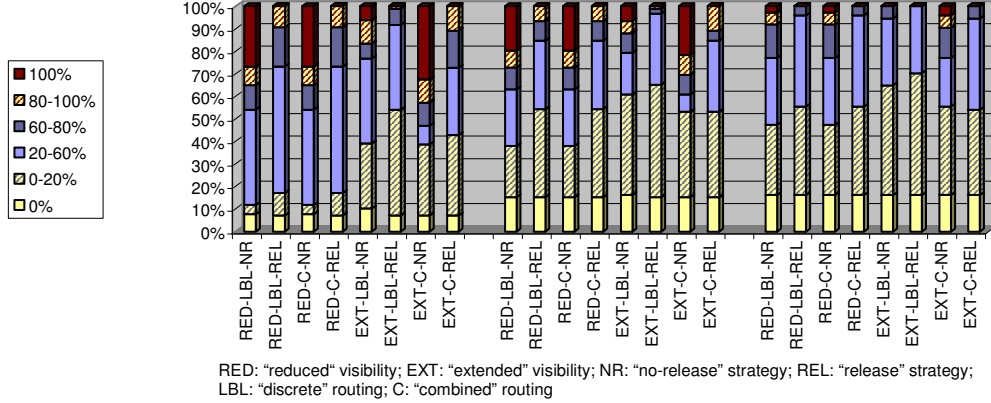


Figure 6. Distribution of link occupancy per network domain.

TABLE II. SUCCESS RATIO FOR CONNCTION EXTENSIBILITY REQUESTS FOR gS-1 AND gS-2 SERVICES.

Scenario	Visibility	Link Release Strategy	Routing	Success Ratio (%)					
				gS-1			gS-2		
				L	M	H	L	M	H
S1	RED	NR	Disc	99.7	86.7	48.0	100	99.5	77.0
S2	EXT	REL	Comb	99.9	94.2	62.0	100	99.7	83.8
M1	RED	NR	Disc	100	99.9	90.6	100	100	99.45
M2	EXT	REL	Comb	100	100	96.7	100	100	99.85

We next illustrate the performance impact when MPE on the example two different gS services, gS-1 and gS-2. In the first test case, gS-1 is created as a D2-internal service, with 5 nodes and full-mesh connectivity. In the second test case, gS-2 is also configured with 5 nodes and full-mesh connectivity, but now over all domains D1, D2 and D3. In both cases, also the inter-domain gS from the previous example is active and is now used to generate the so-called *background traffic* (e.g. commercial Internet traffic). For one specific background traffic load (Erl=1) and three typical service loads (L=10 Erl, M=20 Erl, H=40 Erl, corresponding to network operation under low, medium and high traffic load), we present the performance results for four provision scenarios described in Table II. Scenarios S1 and S2 are performed in the B-B topology only. The scenarios M1 and M2 assume the presence of MPEs.

From Table II, it can be seen that the service provision success (accommodated connection requests) is significantly better for all scenarios with extended visibility. The MPE-based services perform considerably better in general. The

gS-1 (intra-domain service in D2) is unsuccessful in the B-B topology due to the exhaustion of the domain-internal resources. In the case of gS-2 (inter-domain service spanning all domains), this is not the case and improves with usage of MPE. The benefits of MPE are higher for gS-2 service with respect to gS-1.

One of the issues that requires more study is the responsiveness of the network to changing requirements of the applications and the allocation strategies that applications can use in order to acquire resources when competing with other services in the network. To this aim, we performed numerous studies and simulations related to the sharing of resources for competing services. As service virtual topology is evolving over the network infrastructure, depending on visibility of the switching capabilities, resources of different granularities can be acquired and allocated in course of the accommodation of virtual connections. Sometimes, it is necessary to build the whole switching and control hierarchy in order to use only one lower-granular resource. Moreover, if the resources that are made available are visible not only to the service that initiated the resource con-

figuration, they can be shared with other services. At the first sight, the concept of resource sharing seems beneficial. However, our performance studies showed that when service connectivity is changing, i.e. as in the case of extensible services, sharing of the resources may have negative impact on the connection set-up. We performed one such experiment on a 15-node network, where four services shared the network resources. In this case, full (extended) visibility strategy was superior to reduced visibility strategy. We compared this case to the case with services that exclusively use the resources, i.e. dedicated resources without sharing. Surprisingly, our results showed that, for extended visibility, all services performed better without sharing. With the reduced visibility, however, blocking of service extension requests was lower when sharing was enabled. When network is operating with higher loads, the exclusive usage of resources is a better strategy. This lead us to the conclusion that although sharing may look as a globally better strategy the application-internal proactive allocation with the dedicated resources that use the prediction of the anticipated application needs may be a better strategy.

#### V. SUMMARY AND FINAL REMARKS

This paper discussed opportunities and challenges in control plane design and operation for the provision of network-assisted *extensible* Grid services, defined as an atomic infrastructure and operational entity created as control and signaling layer artifact. To this aim, we introduced two novel concepts: network resource visibility and inter-domain exchange. The network resource visibility is a piece of resource allocation information used to dynamically manage globally distributed resources. The concept of inter-domain exchange is based on what we define as the "Grid" Exchange Point (GXP), which is the equivalent of the Internet Exchange Point (IXP) in the data transport layer (e.g. GMPLS-enabled). We showed that the concepts presented here are viable in realistic network scenarios and quantify the benefits in service provision performance.

In our rather conceptual framework, we assumed that the Grid applications are truly network aware share the same, "virtualized", physical network infrastructure. We believe that this interdependence between the Grid applications and intelligent networks will remain, at least in the situations where applications have to react to network outages and adapt to new network states. For extensible network services, the ability to dynamically adapt to the application needs and network resource availability has to come hand-in-hand with the ability to coexist with other Grid services over the same physical network infrastructure. While the performance results have clearly demonstrated that significant benefits can be achieved in link utilization and routing performance by choosing the appropriate resource visibility and interworking architecture, a number of related issues are still under investigation. For example, how will the extensible service strategies scale with respect to network switching hierarchy and multiplicity of services and applications? Will the performance improvement achieved through MPE-like interfaces ("Grid Exchange") justify their

introduction between domains? How will Grid Exchange impact *trusted* visibility mediation? Although the answers to these questions may depend on specific future applications and their requirements, the service provision strategies introduced in this paper are critical for enabling distributed scientific applications with increased resource utilization for networks and improved availability of Grid resources for users.

#### VI. REFERENCES

- [1] Future Generation Computer Systems (FGCS), Special Issue IGrid 2002, Editors: T. De Fanti, M. Brown, C. de Laat, Vol. 19, Nr. 6, Aug. 2003.
- [2] J. Mambretti, et al: "The Photonoc TeraStream: Enabling Next Generation Applications through Intelligent Optical Networking", FGCS, Vol. 19, Nr. 6, Aug. 2003, pp. 897-908.
- [3] E. He, et al: "Quanta: a toolkit fro high performance date delivery over photonic networks", FGCS, Vol. 19, Nr. 6, Aug. 2003, pp. 919-933.
- [4] E. Mannie, et. al.: "GMPLS Architecture", draft-ietf-ccamp-gmpls-architecture-01.txt, work in progress.
- [5] K. Kompella, Y. Rekhter: "LSP Hierarchy with Generalized MPLS TE", draft-ietf-mpls-lsp-hierarchy-07.txt, work in progress.
- [6] H. Ould-Brahim, et. al.: "GVPN: Generalized Provider-provisioned Port-based VPNs using BGP and GMPLS", draft-ouldbrahim-ppvnp-gvpn-bggpmpls-01.txt, work in progress.
- [7] H. Ould-Brahim, et. al.: "VPOXC Provider Provisioned Virtual Private Optical Cross-Connect Service", draft-ouldbrahim-ppvnp-gvpn-bggpmpls-01.txt, work in progress.
- [8] K. Birman: "The league of Supernet", IEEE Internet Computing, Sept/Oct 2003, pp. 93-98.
- [9] R. L. Grossman, et al: "Experimental studies using photonic data services at Igrid2002", FGCS, Vol. 19, Nr. 6, Aug. 2003, pp. 945-953.
- [10] S. Tomic, A. Jukan: "GMPLS-based Exchange Points: Architecture and Functionality", HPSR 2003, Torino, Italy, 2003.
- [11] K. Kompella, et al: "Multi-area MPLS Traffic Engineering", work in progress draft-kompella-mpls-multiarea-te-03.txt
- [12] M. Veeragavan: "Cheetah", WORKSHOP on Optical Control Planes for the Grid community", Chicago, April 2004.
- [13] S. Tomic, A. Jukan: "MPFI: The Multi-provider Network Federation Interface for Interconnected Optical Networks", Globecom 2002, November 2002, Taipei, Taiwan.
- [14] Y. Xu, A. Basu, Y. Xue, "A BGP/GMPLS Solution for Inter-Domain Optical Networking", work in progress draft-xu-bgp-gmpls-02.txt.
- [15] Metz, C. "Interconnecting ISP networks" IEEE Internet Computing, Volume: 5 Issue: 2, March-April 2001 Page(s): 74 -801
- [16] K. J. Blyth, et al. "Designing a GPRS roaming exchange service", Second International Conference on 3G Mobile Communication Technologies, 2001., 2001 Page(s): 201 -205
- [17] Bill St. Arnaud, et. al "BGP Optical Switches and Lightpath Route Arbiter" Optical Networks Magazine March/April 2001
- [18] I. Nakagawa, et al "A design of a next generation IX using MPLS technology" Applications and the Internet, 2002. (SAINT 2002). Proceedings. 2002 Symposium on, 2002 Page(s): 238 -245.
- [19] P. Chandra, et. al.: "Darwin: customizable resource management for value-added network services", IEEE Network, Jan/Feb 2001.
- [20] W. Alanqar, A. Jukan: "Extending End-to-End Service Provisioning and Restoration in Carrier Networks: Opportunities, Issues and Challenges", IEEE Communications Magazine, January 2004.