On-demand VPN Support for Grid Applications

S. Andreozzi, T. Ferrari, E. Ronchieri *
Istituto Nazionale di Fisica Nucleare, CNAF, Italy

Abstract

Quality of Service delivered to Grids by packet-switched networks, is one of the most important factors affecting performance and efficiency of both Grid applications and middleware. The use Virtual Private Network services can improve the overall performance of Grids in many respects. In this paper, we show how the security, privacy and Quality of Service offered by scalable on-demand VPN services can be applied in large-scale Grid scenarios. We propose a novel network resource abstraction for resource discovery of on-demand Virtual Private Networks. It is implemented in a Grid Information Service prototype which was successfully tested both on dedicated infrastructures and production networks. [1]

INTRODUCTION

Virtual Private Network (VPN) is a generic term which is used to refer to the capability of both private and public networks to support a communication infrastructure connecting geographically dispersed sites, where users can communicate among them as if they were in a private network [1]. In VPNs the different groups of users are completely transparent to each other. VPNs can be implemented by different enabling technologies at various levels of the Open System Interconnection reference model (at Layer-3, Layer-2 and Layer-1) and can support a variety of traffic models and services that can increase the efficiency and performance of Grid infrastructures, as explained in the following section.

Layer-3 VPNs *interconnect sets of hosts and routers based on Layer-3 addresses* [2]. This requires the VPN tunnels used for traffic forwarding, to be terminated by PE equipment and the traffic forwarding to be only based on information conveyed in fields from Layer-3 headers. Layer-3 VPN routers support a complete and distinct *logical router* for each VPN. Unlike the former case, Layer-2 VPN services are used to emulate the functionality of a Local Area Network (LAN) in large-scale scenarios. Layer-2 VPNs offer the possibility to extend the traditionally limited LAN reach to the wide area; in this way geographically remote nodes can be virtually connected as if they were part of the same LAN. Traffic forwarding in Layer-2 VPNs relies on Layer-2 frame information. Layer-2 VPNs can connect multiple LAN segments to the same VPN or just two CE devices (either hosts and routers, or Layer-2 switches) [2]. Finally, Layer-1 VPNs connect a number of CE network elements with Layer-1 point-to-point links based on either optical or Time Division Multiplexing technologies. Multiple independent Layer-1 VPNs can coexist on the same physical infrastructure.

In what follows, we show how on-demand VPNs can provide useful services to Grids. We initially illustrate the benefits of supporting different VPN services in various Grid scenarios and we explain why dynamic provisioning of VPN services is important in Grids. In the last part of the paper, we focus on the problem of network resource representation in Grid Information Services (GIS), and we show how this can assist resource brokers during service discovery and matchmaking.

VIRTUAL PRIVATE NETWORKS AND GRIDS

VPNs are applicable to Grids in several different scenarios. VPNs can offer security and privacy to both Grid applications and data management services for large-scale Grid file transfers which rely on storage access protocols not providing security and privacy. In fact, VPNs can support confidentiality and integrity by means of data isolation, i.e. by separating in intermediate forwarding devices the forwarding control plane, the signalling and the routing information of each VPN.

Layer-3 VPNs can provide different traffic forwarding behaviors across the shared public infrastructure. This is achieved by associating priority levels to the packets that are exchanged on specific tunnel instances. In the shared section of the forwarding paths, guarantees are enforced by differentiation mechanisms such as classification, policing, shaping and scheduling. For instance, in MPLS-based VPNs, classification is based on three-bit code points carried in the experimental field of the MPLS Shim Header [3]. Differentiated forwarding behaviors are today viable and the Differentiated Services (DiffServ) architecture [4] is currently supported by several European Research and Education Networks. Less than Best Effort (LBE) and IP Premium [5, 6] are examples of IP-based services offered by the European transport network GEANT. In case of congestion, the former service supports high-end applications by allocating the entire amount of bandwidth to packet-loss sensitive traffic. Conversely, IP Premium constantly minimizes both packet-loss and one-way delay while also supporting guaranteed bandwidth.

* Sergio.Andreozzi/Tiziana.Ferrari/Elisabetta.Ronchieri@cnaf.infn.it

[1] This work is supported by the EGEE project, sponsored by the European Commission under grant IST-2003-508833, and by the INFN project INFN-GRID. It was also funded by the European Commission grant IST-2001-32459 (DataTAG project).

DiffServ services can be integrated in Grids to achieve high-performance in data transmissions among critical computing and storage nodes. This gives the possibility to establish a relationship of virtual *closeness* [7] between Grid services that can drive the behavior of brokers during the resource discovery phase. The resource broker scheduling policies typically dispatch jobs to the Computing Elements (CEs) that can retrieve input files from local Storage Elements (SEs). Unfortunately, the effect of this scheduling policy is the concentration of the workload on the Grid sites that host large data storage facilities. With QoS-capable VPNs directly linking small Grid domains to the larger Grid data sources, a number of remote SEs can be classified as *close*. This results in an increasing use of small computing sites and, consequently, in a more efficient workload distribution. In this way, not only computing resources can be used more efficiently, but also the deployment of network capacity can be optimized, as in this scenario extensive data replication across Grids is not required.

Not only Layer-3, but also Layer-2 VPNs can be successfully applied to Grids, as they can help to bypass firewalls in order to prevent the performance penalties that typically negatively affect data-intensive high-end applications. Layer-2 VPNs can also be used to dynamically cluster geographically dispersed resources belonging to the same Grid Virtual Organization.

The replication of large data files in wide area Grids can be assisted by the dynamic establishment of optical Layer-1 VPNs offering high-speed links between few well-defined Grid nodes. The use of a dedicated high-speed communication channel for single-flow data transmissions, avoids contention between multiple concurrent streams and, consequently, improves the efficiency of reliable transport protocols that are sensitive to loss rate and loss pattern, especially in high-speed long-haul networks [8]. In particular, in case of optical Layer-1 VPNs, *Wavelength Division Multiplexing and tunable technologies in combination with optical switching can provide dynamic control and allocation of bandwidth at the fiber, wavelength band, wavelength or sub-wavelength granularity in optical circuit, burst, or optical packet systems* [9]. Layer-1 VPNs require Grid customers to be responsible for the management and control of the Layer-1 network infrastructure. In addition, on-demand optical Layer-1 VPNs can offer relatively low latency and a large amount of low cost bandwidth, which however needs to be provisioned and scheduled on-demand. The availability of standard protocols for routing, establishment of end-to-end paths and configuration/control of optical cross-connects, is essential.

DYNAMIC ALLOCATION AND ADVANCE RESERVATION

Scalability is an inherent property of Grid systems and it is one of the main factors that drive their design. Consequently, in order to effectively use VPNs in large-scale Grids (e.g., to be capable to address an increasing number of users ubiquitously), stable and scalable VPN services are necessary [10]. Dynamic provisioning is needed in order to reduce management costs together with the number of Grid VPNs that the public network have to support concurrently.

Dynamic provisioning relies on the availability of a suite of protocols which perform discovery of available services, agreement negotiation and agreement establishment between initiators (the Grid user or proxy) and providers (e.g. Grid resource brokers). Of course, a control plane – capable of establishing, managing and tearing down services – is necessary for the actual provisioning of the service [11].

Advance reservation is a mechanism that allows a user to request exclusive access, for a specific time interval in the future, to a set of services that satisfy specific requirements. The user who issues an advance reservation request, has to be authenticated and authorized on the basis policy rules. The actual resource allocation is performed by a resource manager that hides the complexity of the resource-specific allocation tasks. During the reservation life cycle, the Information Service offers vital information in a number of reservation phases. In particular, during the resource discovery, it provides the list of resource instances that satisfy the requirements, and for each instance, information about its properties and their corresponding authentication/authorization manager.

In the following, we propose an approach to scalable on-demand VPN services for Grids. In particular, we detail how service allocation requests are expressed and we illustrate our Grid network resource abstraction is used for service discovery. Detailed information on the overall architecture and heterogeneous resource management techniques is provided in [12].

PATH ELEMENT

On-demand allocation of VPN services requires the initial discovery of resource availability. Grid middleware supports this by relying on information models responsible for capturing structures and relationships of the involved entities. Abstractions are necessary for information publishing in the GIS.

Several are the entities to be represented for VPN service discovery: the services and general capabilities offered on network paths, the agreement providers – contacted during request submission, and the *manager services* (also known as service providers) – responsible for enforcing the agreed service guarantees. In what follows we concentrate on the issue of resource representation.

To cope with the heterogeneity of the network infrastructure when making advance reservations, we propose a new technology-independent network resource abstraction: the *Path*. The Path definition is based on the notion of *Path Element* (PE). The PE provides unidirectional connectivity between two network nodes, where the network node can represent a single device (e.g., an end-system, a router or a switch) or a network domain (e.g., an Autonomous System,

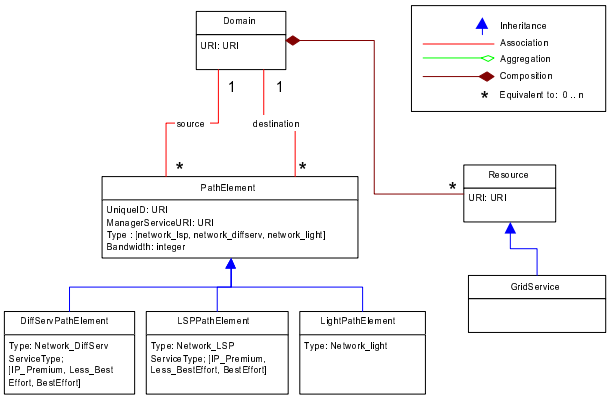

Figure 1: PE Information Model

an IP network or LAN).

The representation of the PE in the GIS needs to capture both general and technology-dependent properties, as illustrated in the Unified Modeling Language class diagram in Figure 1. The list of general attributes includes: the PE unique identifier, the manager service end-point, the PE type – it identifies the services supported such as MPLS, DiffServ, Light Path, etc. – and metrics such as the maximum available bandwidth.

The PE class is a generalization from which technology-dependent classes are derived. Figure 1 provides some examples. The instances of the Layer-3 `DiffServPathElement` class are characterized by the attribute `ServiceType` (e.g., IP Premium, LBE, Best-effort, etc.). Similarly, the Layer-2/3 `LSPPathElement` class has instances which differ about their service type. The Layer-1 `LighPathElement` class is used to represent Ligh Paths.

We define an additional novel abstraction called *Domain* to represent membership and topology information. It describes a collection of networked resources. PEs can connect either to end-systems or domains. Each domain can host several PE termination points, one for each PE associated with it. To represent this abstraction, we introduce the `Domain` class and the network `Resource` member class.

The Domain and the Resource class are both uniquely identified by an `URI`. The Domain class has two PE associations, one referring to the PE as a `source` and the other referring to it as a `destination` (see Figure 1). In presence of one or more PE source termination points, resources in the source domain are entitled to use the PE VPN services only for outgoing traffic, as PEs are unidirectional. Similarly, in case of one or more PE destination termination points, the use of PE VPN services is restricted to incoming traffic.

The proposed network resource abstraction is specialized in the `GridService` class, which is associated with an inherited unique identifier. This class is a potential merging point with the GLUE Schema [13], as the Grid service can be subsequently specialized in the CE and SE concepts.

A prototype of the classes herein proposed was implemented and integrated with a GIS experimental server supporting the Globus Monitoring and Discovery Service (MDS), version 2 [14]. This required the use of the Lightweight Directory Access Protocol (LDAP) data model and was implemented in compliance to the rules that map the GLUE Schema into the LDAP data model (see [15]). In the following, two LDAP queries are presented. The first one shows the LDAP query needed to retrieve SEs and their domain membership. It requires the extension of the SE GLUE schema by the addition of the domain concept and of the unique ID. The second one allows the retrieval of the list of DiffServ PEs whose destination domain is equal to a specific string (e.g. `/infn/cnaf`).

```
ldapsearch -h hostname -p 2135 -x -b
           "mds-vo-name=local,o=grid"
           '(objectclass=GlueSE)'
            GlueSEUniqueID GlueSEDomainURI"

ldapsearch -h hostname -p 2135 -x -b
           "mds-vo-name=local,o=grid"
           '(&(objectclass=GluePE)(GluePEType=lsp_diffserv)
            (GluePEDestinationDomain=/infn/cnaf))'
```

In our prototype Grid services need to be aware of their domain membership with regards to the PEs. The class definition approach presented here is one of the possible choices to represent resource membership. Alternative solutions, where the relationship is represented independently of the service itself, are possible.

PATH DISCOVERY

In our proposal, VPN service requests are submitted to a Grid resource broker responsible for performing resource discovery. The request includes the service description terms, which are subsequently extracted by the broker and are provided in input to the Matchmaker, according to the DataGrid Workload Manager architecture described in [16].

Service attributes can be grouped into four major categories. Of course, the list is non-exhaustive and more attributes can be added for a fine granularity resource match.

1. **Type**: attributes in this group provide information about the *Type* of the reservation (which can be atomic or compound). Other type attributes in this category distinguish service requests from job submission requests. The *ServiceType* specifies the type of the service depending on the service name. For example, in case of service name equal to "Network_LSP", the types can be: "IP_Premium", "BestEffort", "Less_BestEffort", "Invalid", "Other", etc.

2. **Time**: they provide the time information of the reservation. Some attributes of this group are: *AllocStartNow*, *AllocEndInfinite*, *AllocStartTime*, *AllocEndTime* and *AlloDurationTime*.

3. **Application**: attributes in this category specify the information that is specific to a service request instance,

such as service run-time properties. Examples of run-time properties are the TCP/UDP port numbers, IP addresses, and protocol types.

4. **Service Properties**: they include the specific metrics that quantitatively describe the service agreement specification. Each service has its own domain-specific parameters.

Within the Workload Manager Service, the Matchmaker relies on matchmaking and ranking algorithms in order to determine the list of suitable CEs for a given job. The best CE in the list will be the one in charge for running and completing the job. For example, the matchmaking algorithm in [16] adopts different discovery strategies in three scenarios: direct job submission and job submission with or without data-access requirements. These matchmaking rules have been changed in order to cope with the introduction of the PE. In particular, new matchmaking and ranking rules have been defined for the selection of CE/SE pairs, which consider the availability of Layer-3 VPN services on PEs connecting remote CEs and SEs when making choices.

EXPERIMENTAL RESULTS

MPLS-based Layer-2 VPNs with DiffServ QoS have been tested both on a dedicated network infrastructure connecting Geneva and Chicago and across production networks (GARR, the national research and education network in Italy), and GEANT connecting INFN CNAF (Italy) and CERN (Switzerland). We were able to demonstrate that the VPN services for Grids proposed here are viable. In particular, end-systems at INFN and CERN could be dynamically configured to be part of one or multiple Layer-2 VPNs at the same time, and different VLANs were set up dynamically with different packet forwarding behaviors. On-demand MPLS connectivity relied on the presence of two static paths: an IP Premium path and an LBE path, whose attributes were provided by the experimental GIS described in this paper. Traffic exchanged between hosts on the IP Premium VLANs was successfully protected from congestion caused by competing traffic on the LBE VLANs.

CONCLUSION

In this paper, we have illustrated a number of different application scenarios of VPN services for Grid applications and middleware. Extensions to existing Grid services are needed to implement on-demand VPN services in Grids. We have proposed various entity abstractions which need to be supported by the GIS in order to assist resource brokers during the resource discovery phase: the Path Element, the Domain, the Resource and the Grid Service. We have outlined the characteristics of our prototype based on the proposed resource abstraction hierarchy. This prototype was successfully tested for making advance reservations of different VPN services in a large scale network infrastructure.

REFERENCES

[1] Gleeson, B. et al.; *A Framework for IP Based Virtual Private Networks*; RFC 2794, Feb 2000.

[2] Andersson, L.; Madsen, T.; *Provider Provisioned VPN Terminology*; IETF draft: draft-ietf-l3vpn-ppvpn-terminology, Sept 2004, work in progress.

[3] Le Faucheur, F. et al.; *Multi-Protocol label Switching (MPLS) Support of Differentiated Services*; RFC 3270, May 2002.

[4] Blake, S. et al.; *An Architecture for Differentiated Service*; RFC 2475, Dec 1998.

[5] Ferrari, T. et al.; *Experiments with Less than Best Effort (LBE) Quality of Service*; Deliverable D9.9, IST project IST-2000-26417 (GEANT), Aug 2002.

[6] Campanella, M. et al.; *Specification and Implementation Plan for a Premium IP Service*; Deliverable D9.1, IST project IST-2000-26417 (GEANT), Apr 2001.

[7] *Expressing Relationship between Computing and Storage Services*, GLUE Schema (`http://www.cnaf.infn.it/~sergio/datatag/glue/v11/CESE/index.htm`).

[8] *High Performance Transport*; Deliverable 2.2, DataTAG, IST project IST-2001-32459, Mar 2004 (`http://edms.cern.ch/document/431718`).

[9] Simeonidou, D. et al.; *Optical Network Infrastructure for Grid*; draft-ggf-ghpn-opticalnets, Grid High Performance Networking Research Group, Global Grid Forum, Sept 2004.

[10] Nagarajan, A. (Ed.); *Generic Requirements for Provider Provisioned Virtual Private Networks (PPVPN)*; RFC 3809, Jun 2004.

[11] Andrieux, A. et al.; Web Services Agreement Specification (WS-Agreement); GGF Grid Resource Allocation Agreement Protocol draft, work in progress (`https://forge.gridforum.org/projects/graap-wg`).

[12] *Demonstration of Advance Reservation and Services*; Deliverable 2.5, DataTAG, IST project IST-2001-32459, Mar 2004 (`https://edms.cern.ch/document/431913`).

[13] *GLUE Schema - Resources* (`http://www.cnaf.infn.it/~sergio/glue`).

[14] Czajkowski, K. et al.; *Grid Information Services for Distributed Resource Sharing*; in Proc. of the 10^{th} IEEE Int. Symposium on High-Performance Distributed Computing, IEEE Press, Aug 2001.

[15] Andreozzi, S.; *GLUE Schema Implementation for the LDAP Data Model*; to appear as Tech. Report of the Istituto Nazionale di Fisica Nucleare. Sep 2004 (`http://www.cnaf.infn.it/~sergio/publications/Glue4LDAP.pdf`).

[16] *Definition of Architecture, Technical Plan and Evaluation Criteria for Scheduling, Resource Management, Security and Job Description*; DataGrid Deliverable DataGrid-01-D1.2, Dic 2001 (`https://edms.cern.ch/document/332413`).