Multi Access Inter Domain architecture for QoS provisioning in MPLS/DiffServ networks

Gino Carrozzo, Nicola Ciulli

META Centre
Consorzio Pisa Ricerche
C.so Italia 116, 56125 Pisa, ITALY
Telephone: (+39) 050 915 811
Fax: (+39) 050 915 823
Email: {g.carrozzo, n.ciulli}@cpr.it

Stefano Giordano

Dept. of Information Engineering
University of Pisa
via Caruso, 56122 Pisa, ITALY
Telephone: (+39) 050 2217 511
Fax: (+39) 050 2217 522
Email: s.giordano@iet.unipi.it

***Abstract*— In an advisable network scenario, the peripheral non-DiffServ/non-MPLS access networks (e.g. IntServ networks with the RSVP, H.323, SIP, MPEG-4, etc.) are connected to a MPLS/DiffServ core network. In such a network, border routers manage the interoperability among different domains, providing the correct handling of QoS parameters contained in the different signaling protocol messages. Moreover, a centralized Bandwidth Broker manages network resources, in order to provide a fine-grain/tailored traffic engineering and to handle policy/admission procedures for IP flows accessing the backbone. This paper aims at describing the design issues for network elements operating in a Multiple Access Inter Domain (MAID) architecture based on the MPLS/DiffServ technology.**

I. Introduction

Quality of Service (QoS) over IP networks has a long history of standards and tools, both at the Data Plane level (e.g. Traffic Control algorithms) and at the Control Plane level (e.g. signaling and policy protocols).

From a user perspective, the basic QoS requirements are the dynamism (e.g. the service should last as long as the user needs) and the tailoring (e.g. the network resources allocated for the service should fulfill exactly the end-user requirements) of the end-to-end IP traffic. Though some tools for QoS are available in commercial IP routers, their compliancy to these requirements is still far from being a market reality. Indeed, the main obstacles for such a deployment reside in:

- the different technologies of the backbone networks (e.g. DiffServ, MPLS, IPoATM, etc.), which make hard to guarantee end-to-end QoS, above all when the service has to be deployed across different administrative domains;
- the number of protocols used in the access networks for QoS (e.g. RSVP, H.323, SIP, MPEG-4, etc.), which implies a per-service/per-protocol User-Network-Interface (UNI).

From a Service Provider perspective, other requirements drive the evolution of the services offered by the backbone, such as:

- network scalability, which implies a distributed Control Plane, best fitted if based on MPLS;
- traffic engineering both at the flow and at the resource level, best fitted if based on a centralized domain architecture and on a DiffServ data plane;
- service survivability in case of faults or dynamic network topology changes, easily guaranteed by MPLS recovery strategies;
- interoperation of adjacent domains with the same or different technologies and the interoperation of equipments from different vendors.

In this paper we describe the architecture of a Multi-Access Inter-Domain (MAID) backbone network based on the MPLS/DiffServ technology. The basic elements of such an architecture are the border routers (MA-BR), aimed at handling QoS requests and IP flows from the access networks, and the Bandwidth Broker (BB), aimed at managing network resources and policies and at handling inter-domain communications. In Sec. II the general requirements for the MAID architecture are proposed, while in Sec. III the design issues for the implementation of the network elements are drawn. Some conclusions and some hints for future work are sketched in Sec. IV.

II. Requirements for the MAID architecture

In the MAID network scenario (ref. Figure 1), IP flows with QoS requirements are managed at first by the MA-BR, which provides the inter-working between the access network and the backbone. This network element merges the functionalities of an edge router (ER) and a border router (BR), providing the correct handling of QoS parameters contained in the different signaling protocol messages. Moreover, a centralized network manager, i.e. the BB, manages the backbone resources, in order to provide a fine-grain/tailored traffic engineering and to handle policy/admission procedures for IP flows accessing the backbone. The MAID data plane provides the mapping and forwarding of the access network flows into the proper Diff-Serv PHBs/MPLS Label Switched Paths (LSP) and vice versa. On the other hand, the MAID control plane is responsible for Admission Control (AC) and policy decisions (taken on a per-flow or per-PHB basis) and for the service level agreement (SLA) maintenance. In this reference network, MPLS LSPs with QoS-DiffServ guarantees are established both inside a unique MPLS/DiffServ domain and across multiple domains.

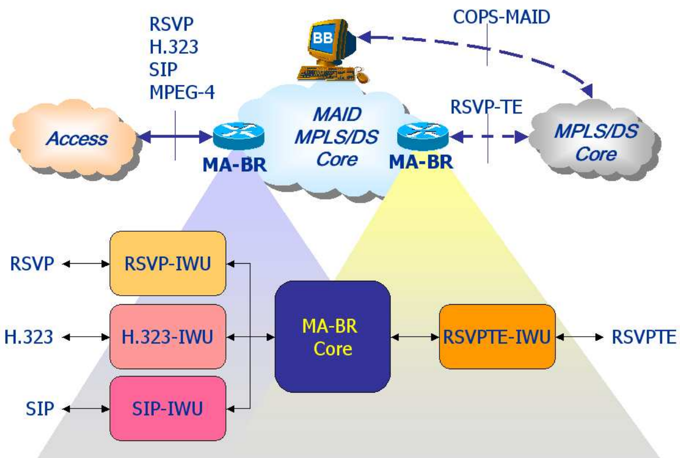

Fig. 1. Multiple Access MPLS/DiffServ network architecture

The signaling protocol for setting up and tearing down LSPs inside a domain is RSVP-TE [1][2][3], while the signalling protocol for the communication between an access network and the MA-BR is application-dependent (e.g. RSVP for IntServ, H.323 or SIP for VoIP, etc.). The MA-BR may either delegate all AC and policy decisions to the BB, working in a totally outsourcing scenario, or it may have limited local AC and policy functionalities, granted by a provisioning scenario. In any case, BB has pre-emption rights on each MA-BR decision, as it provides a centralized AC and policy which is supposed to be optimal with respect to that provided by MA-BR. It is possible that BB configures directly core and border routers according to its criteria (e.g. via SNMP protocol). However, it is desirable to tune an optimum mix of static and dynamical resource allocation (e.g. via MPLS signaling protocols, ref. Figure 2), in order to share architectural complexity between BB module and border NEs and to make the entire system scalable.

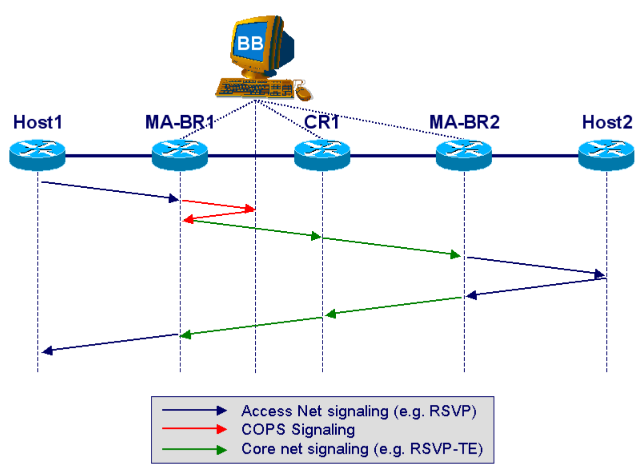

Fig. 2. MAID Intra-domain signaling

A. Communications between MA-BR and BB

In the MAID architecture, the protocol used for the communication between the MA-BR and the BB is the Common Open Policy Service protocol (COPS-[4]). Relying on the client-server model, COPS architecture is based on two fundamental elements: a policy server, called Policy Decision Point (PDP), also addressed as COPS server, and one or more policy clients, called Policy Enforcement Points (PEPs), addressed as COPS clients. At least one policy server must exist in each administrative domain, in order to implement a complete COPS communication with one or more PEPs. A single PEP is able to support multiple client-types. If a client-type is not supported by the PDP, the PDP itself can redirect the PEP to an alternative PDP address and port for a given client-type via COPS. Different applications using different protocols may be viewed as different client types. This is the case of the RSVP protocol, for which the client-type 1 has been assigned [5]. However, the definition of a new client type for each possible protocol of the access network would result in a hard limit to the scalability of the system, because of the duplication of the states installed both in the PDP and in the PEP. A possible solution to this issue might be the use of many PDPs, each supporting one or few client-types; but, all of these COPS servers have either to exchange management information to perform a coherent resource allocation, or they must query for a higher level ”omniscient” BB.

The novel and original solution we propose through the MAID architecture is to define a unified COPS semantic [6], in order to integrate all the QoS information carried out by the different protocols of the access domain. This semantic will translate the different QoS information into a common format besides of the specific protocol (ref. TABLE I, TABLE II and [6] for details). This solution transfers the complexity of the system on the border routers, where appropriate Inter Working Units (IWUs) are used to map protocol specific messages into generalized client messages (ref. Figure 1-left side). Thus, a unified COPS client-type transmits all the information to a unique PDP, and there is no need to develop different COPS server (PDP) for each supported COPS client or to refer to a higher-level device when performing resource allocation, because the unique COPS server (PDP) could be located inside the BB itself.

B. Inter-Domain Communications

The research community is dealing with a number of open issues regarding inter-domain communications (e.g. the optimal TE routing, the NNI signalling protocol extensions, etc.). In this context, the MAID architecture arises as an effective solution, because of the centralized action of the BB and of the modularity of the MA-BR.

The details for an inter-domain operation are out of the scope of this work. However, two possible strategies for the inter-domain connection setup are possible and are sketched here to prove the architectural flexibility:

- Inter-BB communication via COPS-MAID interface (ref. Figure 3), which has network granularity;

TABLE I
COPS-MAID EXTENSION FOR THE REQUEST PHASE

Message Type	Direction	Contents	
Request	PEP → PDP	Traffic originator	- Source host - Ingress MA-BR I/F + Label
		Traffic Terminator	- Destination host - Egress MA-BR I/F
		Traffic description	- resource class/color - setup/ holding lsp priority, - multiple {lsp diffserv type (e-lsp, l-lsp), traffic characterization (RP, LBAP, 3D-LBAP, etc.)}
		QoS description	- Bandwidth - Delay - Jitter - Loss probability
		LSP recovery behaviour	- Recovery type (path prot., path rest., …) - Diversity type (node, link, SRLG)
		Temporal infos	- start time - end time

TABLE II
COPS-MAID EXTENSION FOR THE DECISION PHASE

Message Type	Direction	Contents	
Decision	PDP → PEP	… Request Message Objects …	
		Label type	- DiffServ, - ATM, - MPLS
		Label	- DSCP, - LSPId
		Explicit Route Object	- primary ERO - backup ERO

- Inter BR communications via strict Network-to-Network Interface (NNI, ref. Figure 4), which has node granularity.

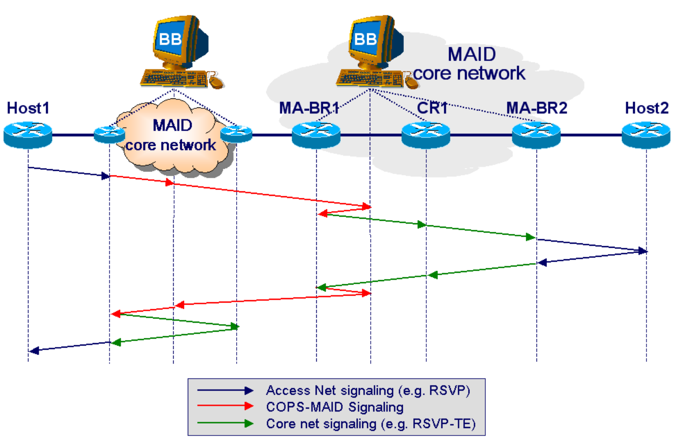

Fig. 3. MAID Inter-domain signalling: inter BB case

In the first case, the messages exchanged via COPS-MAID between adjacent BBs allow to:

- send/receive a request of service for/from another domain;
- send/receive a response about a service request.

Due to the client-server nature of COPS, both the client and the server sides have to be implemented on each BB.

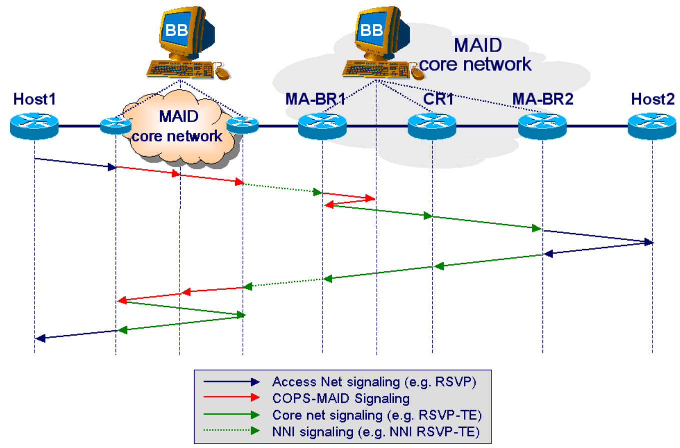

Fig. 4. MAID Inter-domain signalling: strict-NNI case

III. NETWORK ELEMENTS INTERIOR DESIGN

The MAID architecture is going to be tested on a PC-based test plant in the context of the TANGO project [7] (some preliminary results are available in [8]). The modules implementation is targeted for Linux-based platforms with particular emphasis on:

- system overall modularity, provided by socket interconnection between the different modules, achieving also the possibility to distribute the MA-BR (or the BB) modules on different machines and to scale the system easily;
- standard compliancy of the code for the MPLS suite, the DiffServ, the COPS, the signalling protocols for the access networks (e.g. SIP, H.323, RSVP, MPEG-4);
- interoperability with commercial routers;
- possibility to add only the MAID control plane to commercial routers (e.g. on a companion equipment sniffing the standard signalling flow), providing new functionalities to a commercial MPLS/DiffServ core network;
- use of open source code and tools (e.g. the Linux Traffic Control tools, the MPLS patch to Linux kernel, the ISI RSVP/RSVP-TE, the OpenH323 suite, the Zebra routing suite, the MySQL suite).

In this section, the interior design of the two basic MAID elements is sketched:

- the MA-BR, which exposes interfaces towards the access network, the adjacent administrative domains, the BB (mainly via COPS-MAID), the core routers (CR);
- the BB, which exposes interfaces towards the MA-BRs, the CRs and the adjacent BBs.

A. Multi Access Border Router

In the MAID architecture, the MA-BR is designed in order to manage:

- signalling messages from the access network(s);
- IP control plane towards the core network (e.g. protocols RSVP-TE and OSPF-TE);
- IP management plane (e.g. protocol SNMP);

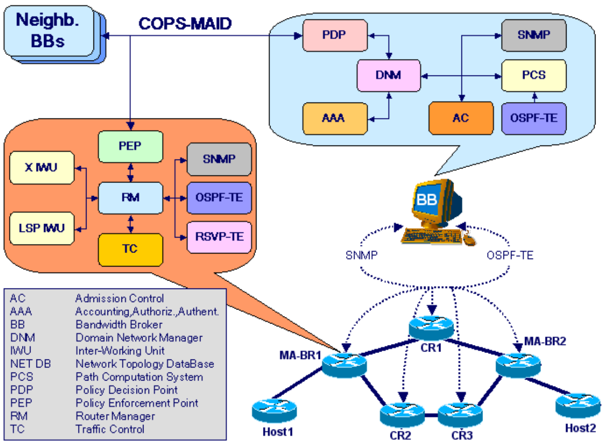

Fig. 5. MAID network element modular decomposition

- local admission and policy of traffic flows, in the case of a resource provisioning scenario;
- the request for traffic admission towards a BB, in the case of an outsourcing scenario.

The MA-BR has a number of functions in common with a CR and a number of functions dedicated to the translation between the access protocols and COPS-MAID. The two functional areas are distinguished by the two differently shaded areas in Figure 6.

The upper part is configured either via a local command-line user interface or via COPS by the BB. The lower part receives configuration commands in three ways:

- from the BB, by means of the SNMP protocol;
- from the (command line) user interface.
- from the upper part, through the interface between Signalling Translation Module (STM) and the MA-BR Resource Manager (MA-BR-RM);

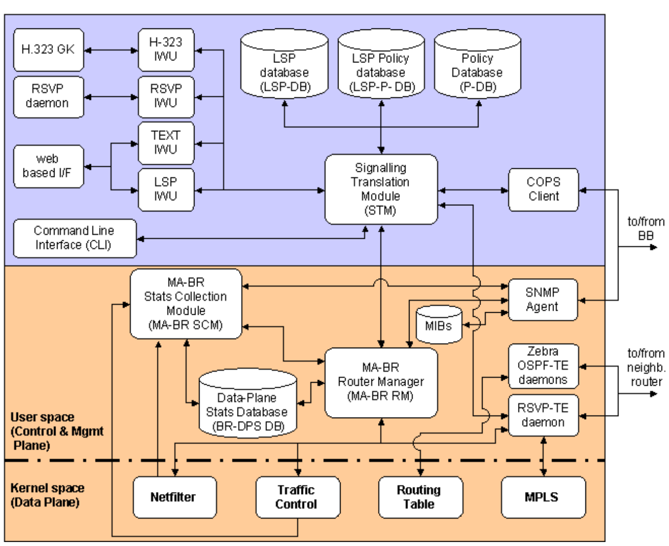

Fig. 6. MA-BR internal architecture

In the lower part, a further distinction is possible between the user space (control and management plane) and the kernel space (data plane).

The rectangular blocks in Figure 6 represent the different processes/threads, while the cylindrical blocks represent the databases.

Focusing on the MA-BR control and management plane, the basic action of the MA-BR is to understand service requests expressed by means of various signalling protocols and to manage them by granting or refusing access to the MPLS/DiffServ domain. This service management is made up of two basic actions:

- Requester authentication;
- Admission Control (AC), in terms of Authorization and resource Admission Control.

Both actions can be performed either locally at the MA-BR or by querying the domain Bandwidth Broker. The MA-BR may operate passively and actively in each signalling transaction associated with a service request. In the passive case, it handles the signalling messages to distinguish the different call phases (e.g. setup, data transfer and teardown) and extracts the basic information for authentication and AC. While, in the active case, MA-BR may alter the normal signalling evolution according to the authorizations and/or availability of MPLS/DiffServ network resources. Both kinds of intervention (passive and active) depend on the adopted signalling protocol and so they should be carried out by the protocol-specific IWU.

The MA-BR data plane is mainly based on the ”traffic control”, which classifies packets, maps incoming flows into the appropriate PHB according to specific criteria defined by the control plane, and forwards data to the core network. In order to provide MPLS functionalities, even the MPLS module is active on the data plane; it manages labeling of incoming packets on the basis of the control plane decisions.

B. Bandwidth Broker

The Bandwidth Broker (BB) plays a key role in the MAID architecture. As shown in Figure 7, BB communicates with MA-BRs through the Common Open Policy Service (COPS) server block, making AC/policy decisions based on information coming from the core network. This information is collected alternatively and complementarily by the SNMP API and by the OSPF-TE module. BB may provide the configuration of its area routers (e.g. CRs, simple MPLS/DiffServ BRs and MA-BRs) by means of SNMP protocol and COPS decision messages. However, in a more desirable scenario, BB should delegate the complete configuration of resources within the domain (e.g DiffServ and MPLS) to RSVP-TE, limiting its action to request validation (e.g. policy/AC), route computation for LSPs and response generation.

The Domain Network Manager (DNM) provides a basic coordination function among all the modules of the BB. Upon receiving a request for policy decision through the COPS Server, DNM triggers an authentication/authorization/accounting procedure in the Policy Engine (PE) and a LSP

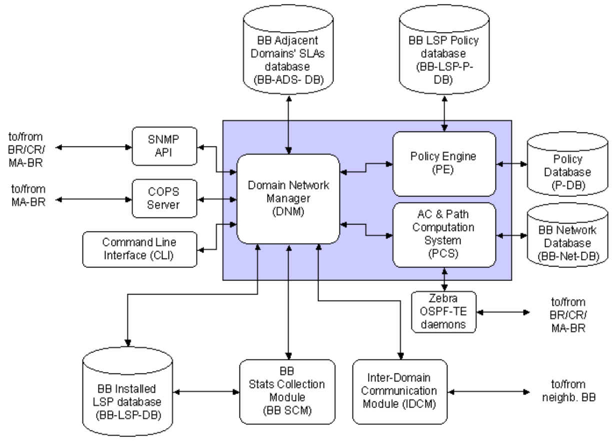

Fig. 7. BB internal architecture

route computation in the Path Computation System module (PCS), together with an Admission Control procedure. On the basis of the response from the mentioned modules, the DNM decides if:

- the received request can be answered by a domain-local AC (e.g. if the destination is inside the domain or if the destination is outside the domain but a proper SLA is already in place with the adjacent domain);
- then AC request must be propagated to an adjacent BB through the Inter-Domain Communication module (IDCM). For this purpose, the DNM accesses the BB Adjacent Domains SLAs DB (BB-ADS-DB).

Moreover, DNM manages statistical information coming from the network elements through the SNMP API module; DNM delegates statistics interpretation to the Stats collection module.

LSPs for BB response to flow requests are computed in the PCS module. The path between an ingress MA-BR and an egress MA-BR is computed according to the topology information achieved by the routing software running on BB. Path computation algorithm is based on a Constraint-based Shortest Path First (CSPF) implementation, which takes into account TE constraints (bandwidth, delay, jitter, packet loss, resource class/colors). The path calculator returns a Label Switched Path (LSP), if any, in the form of an RSVP-TE Explicit Route object. For survivability purposes, the PCS may process a recovery description for the requesting traffic flow. This description is based on:

- a recovery type (e.g. Unprotected, Path/Link Protection, Path/Link Restoration);
- a requested recovery diversity type the returned LSPs should meet (e.g. Node Diversity, Link Diversity, SRLG Diversity).

The algorithm for such a computation is based on specific topology transformations and its description is out of the scope in this work. In the MAID architecture, the routing protocol (e.g the Zebra suite) is needed for flooding MPLS topology information throughout the core network and, above all, towards the BB. So, only the OSPFv2 with TE extensions module is used. Nevertheless, a proper configuration of BB link cost and TE information is needed in order to avoid BB participating in LSP like a standard MPLS/DiffServ network element. A specific interface has been developed to the Zebra code in order to export the OSPF LSDB to the PCS data structures.

IV. Concluding Remarks and Future Work

In this paper the architecture of a Multi-Access Inter-Domain (MAID) backbone network has been sketched, focusing on the requirements for providing end-to-end QoS to IP flows. The backbone network is assumed to be based on the MPLS/DiffServ technology, as this solution promises to be the most effective for QoS purposes both at the user and at the Service Provider level. The basic elements of the MAID architecture (e.g. the MA-BR and the BB) have been detailed in their modular composition and implemented for Linux-based platforms. The overall architecture is going to be tested on a PC-based test plant. Other design and implementation activities are in progress, above all on the inter-domain issues, as the final purpose of the TANGO project will be the demonstration of a dynamical inter-working with QoS guarantees between different administrative IP domains owned by some of the research units involved in the project.

Acknowledgements

This work has been supported in part by the Italian Ministry of Education, University and Research through the project TANGO (MIUR Protocol No. RBNE01BNL5).

References

[1] D. Awduche et al., *RSVP-TE: Extensions to RSVP for LSP Tunnels*, IETF RFC 3209, December 2001.
[2] F. Le Faucheur (Editor) et al., *Multi-Protocol Label Switching (MPLS) Support of Differentiated Services*, IETF RFC3270, May 2002
[3] D.Black, S. Brim, B. Carpenter, F. Le Faucheur, *Per Hop Behavior Identification Code*, RFC 3140, June 2001
[4] D. Durham, Ed., J. Boyle, R. Cohen, S. Herzog, R. Rajan, A. Sastry, *The COPS (Common Open Policy Service) Protocol*, IETF RFC 2748, January 2000
[5] S. Herzdog, Ed., J. Boyle, R. Cohen, D. Durham, R. Rajan, A. Sastry, *COPS Usage for RSVP*, IETF RFC 2749, January 2000
[6] G. Carrozzo (Editor), *COPS-MAID, COPS Usage for Multi-Access Inter-Domain MPLS-DiffServ Networks*, draft-cpr-rap-cops-maid-00.txt, Internet Draft, Work in progress, November 2003.
[7] *Traffic models and Algorithms for Next Generation IP networks Optimization (TANGO) Project*, http://tango.isti.cnr.it/.
[8] N. Ciulli et al., *An Architecture for Transparent End-to-End QoS-IP Transport Service Provisioning*, Proc. of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA '03), Las Vegas (NV) USA, June 23-26 2003