SERVICE AUTOMATION AND DYNAMIC PROVISIONING TECHNIQUES IN IP/MPLS ENVIRONMENTS

Christian Jacquenet, Gilles Bourdon and Mohamed Boucadair

All at
France Telecom, France
SERVICE AUTOMATION AND DYNAMIC PROVISIONING TECHNIQUES IN IP/MPLS ENVIRONMENTS
The ‘Wiley Series in Communications Networking & Distributed Systems’ is a series of expert-level, technically detailed books covering cutting-edge research, and brand new developments as well as tutorial-style treatments in networking, middleware and software technologies for communications and distributed systems. The books will provide timely and reliable information about the state-of-the-art to researchers, advanced students and development engineers in the Telecommunications and the Computing sectors.

Other titles in the series:

Wright: Voice over Packet Networks 0-471-49516-6 (February 2001)
Jepsen: Java for Telecommunications 0-471-49826-2 (July 2001)
Sutton: Secure Communications 0-471-49904-8 (December 2001)
Stajano: Security for Ubiquitous Computing 0-470-84493-0 (February 2002)
Martin-Flatin: Web-Based Management of IP Networks and Systems, 0-471-48702-3 (September 2002)
Berman, Fox, Hey: Grid Computing. Making the Global Infrastructure a Reality, 0-470-85319-0 (March 2003)
Welzl: Network Congestion Control: Managing Internet Traffic 0-470-02528-X (July 2005)
Raz, Juhola, Serrat-Fernandez, Galis: Fast and Efficient Context-Aware Services 0-470-01668-X (April 2006)
Heckmann: The Competitive Internet Service Provider 0-470-01293-5 (April 2006)
Dressler: Self-Organization in Sensor and Actor Networks 0-470-02820-3 (November 2007)
Berndt: Towards 4G Technologies: Services with Initiative 978-0-470-01031-0 (March 2008)
SERVICE AUTOMATION AND DYNAMIC PROVISIONING TECHNIQUES IN IP/MPLS ENVIRONMENTS

Christian Jacquenet, Gilles Bourdon and Mohamed Boucadair

All at
France Telecom, France
Contents

Preface xi
Acknowledgements xiii

PART I ARCHITECTURES AND PROTOCOLS FOR SERVICE AUTOMATION 1

1 Introduction 3
  1.1 To Begin With 3
    1.1.1 On IP Networks in General, and Routers in Particular 3
    1.1.2 On the Usefulness of Dynamic Routing Protocols in IP Networks 5
    1.1.3 On the Inability of an IGP to Address Interdomain Communication Needs 7
    1.1.4 On the BGP-4 Protocol 9
    1.1.5 The Rise of MPLS 10
  1.2 Context and Motivation of this Book 13
    1.2.1 Classifying Capabilities 14
    1.2.2 Services and Policies 14
    1.2.3 The Need for Automation 15
  1.3 How this Book is Organized 16
  1.4 What Is and What Should Never Be 16
References 16

2 Basic Concepts 19
  2.1 What is a Policy? 19
  2.2 Deriving Policies into Rules and Configuration Tasks 19
    2.2.1 Instantiation 20
    2.2.2 Device Identification 20
    2.2.3 Translation 21
  2.3 Storing Policies 21
  2.4 Policy and Device Configuration 21
  2.5 Policy-based Management Model 22
    2.5.1 Reaching a Policy Decision 24
    2.5.2 Requirements for a PEP–PDP Communication Protocol 24
References 25
5 The Common Open Policy Service (COPS) Protocol

5.1 A New Scheme for Policy-based Admission Control
5.2 A Client–Server Architecture
5.3 The COPS Protocol
   5.3.1 The COPS Header
   5.3.2 The COPS Message Objects
5.4 COPS Messages
   5.4.1 Client-Open (OPN)
   5.4.2 Client-Accept (CAT)
   5.4.3 Request (REQ)
   5.4.4 Decision (DEC)
   5.4.5 Other COPS Messages
5.5 Summary of COPS Operations
5.6 Use of COPS in Outsourcing Mode
5.7 Use of COPS in Provisioning Mode
   5.7.1 On the Impact of Provisioning Mode on COPS Operations
   5.7.2 On the Impact of Provisioning Mode on PEP–PDP Exchanges
5.8 Security of COPS Messages
5.9 References

6 The NETCONF Protocol

6.1 NETCONF at a Glance
   6.1.1 Introduction
   6.1.2 Motivations for Introducing NETCONF
   6.1.3 NETCONF, an IETF Initiative
   6.1.4 Missions of the IETF NETCONF Working Group
   6.1.5 NETCONF-related Literature
   6.1.6 What is In? What is Out?
6.2 NETCONF Protocol Overview
   6.2.1 Some Words about XML
   6.2.2 NETCONF Terminology
   6.2.3 NETCONF Layer Model
   6.2.4 NETCONF Communication Phases
   6.2.5 NETCONF Data
   6.2.6 NETCONF Capability Exchange
   6.2.7 RPC Layer
   6.2.8 NETCONF Filtering
6.3 NETCONF Protocol Operations
   6.3.1 Retrieve Configuration Data
   6.3.2 Get
   6.3.3 Delete Configuration Data
   6.3.4 Copy Configuration
   6.3.5 Edit Configuration Data
   6.3.6 Close a NETCONF Session
   6.3.7 Kill a Session
6.4 References
6.3.8 Lock NETCONF Sessions
6.3.9 Unlock NETCONF Sessions
6.3.10 Validate Configuration Data
6.3.11 Commit Configuration Changes
6.3.12 Discard Changes of Configuration Data
6.3.13 NETCONF Notification Procedure
6.4 NETCONF Transport Protocol
  6.4.1 NETCONF as Transport-independent Protocol
  6.4.2 Transport Protocol Alternatives
6.5 NETCONF Capabilities
  6.5.1 URL Capability
  6.5.2 XPath Capability
  6.5.3 Writable-Running Capability
  6.5.4 Candidate Configuration Capability
  6.5.5 Confirmed Commit Capability
  6.5.6 Validate Capability
  6.5.7 Distinct Startup Capability
  6.5.8 Rollback on Error Capability
  6.5.9 Notification Capability
6.6 Configuring a Network Device
6.7 NETCONF Content Layer
References

7 Control and Provisioning of Wireless Access Points (CAPWAP)
  7.1 CAPWAP to Address Access Point Provisioning Challenges
  7.2 CAPWAP Concepts and Terminology
  7.3 Objectives: What do we Expect from CAPWAP?
  7.4 CAPWAP Candidate Protocols
  7.5 The CAPWAP Protocol
  7.6 CAPWAP Future
References

PART II APPLICATION EXAMPLES OF SERVICE AUTOMATION
AND DYNAMIC RESOURCE PROVISIONING TECHNIQUES

8 Dynamic Enforcement of QoS Policies
  8.1 Introduction
    8.1.1 What is Quality of Service, Anyway?
    8.1.2 The Need for Service Level Specifications
  8.2 An Example
  8.3 Enforcing QoS Policies in Heterogeneous Environments
    8.3.1 SLS-inferred QoS Policy Enforcement Schemes
    8.3.2 Policy Rules for Configuring DiffServ Elements
References
9 Dynamic Enforcement of IP Traffic Engineering Policies
  9.1 Introduction 199
  9.2 Terminology Considerations 200
  9.3 Reference Model 201
  9.4 COPS Message Content
    9.4.1 Request Messages (REQ) 202
    9.4.2 Decision Messages (DEC) 203
    9.4.3 Report Messages (RPT) 203
  9.5 COPS-PR Usage of the IP TE Client-Type 204
  9.6 Scalability Considerations
    9.6.1 A Tentative Metric Taxonomy 205
    9.6.2 Reporting the Enforcement of IP Traffic Engineering Policy 206
  9.7 IP TE PIB Overview 206
  9.8 COPS Usage for IP TE Accounting Purposes 207
References 208

10 Automated Production of BGP/MPLS-based VPN Networks
  10.1 Introduction 211
  10.2 Approach 212
  10.3 Use of Policies to Define Rules 214
  10.4 Instantiation of IP VPN Information Model Classes 214
  10.5 Policy Components of an IP VPN Information Model
    10.5.1 Physical Components of an IP VPN Information Model 215
    10.5.2 Virtual Components of an IP VPN Information Model 217
    10.5.3 Inheritance Hierarchy 218
  10.6 Dynamic Production of IP VPN Services 221
  10.7 Context of a Multidomain Environment
    10.7.1 A Bit of Terminology 222
    10.7.2 Reference Model 223
  10.8 Possible Extensions of the VPN Model 224
References 224

11 Dynamic Enforcement of Security Policies in IP/MPLS Environments
  11.1 Enforcing Security Policies for Web-based Access Control 227
  11.2 Enforcing Security Policies in Companies with 802.1X 235
References 238

12 Future Challenges
  12.1 Introduction 239
    12.1.1 Current Issues with Configuration Procedures 239
    12.1.2 Towards Service-driven Configuration Policies 240
  12.2 Towards the Standardization of Dynamic Service Subscription
    and Negotiation Techniques 241
    12.2.1 Basic Motivation 241
    12.2.2 Commercial Framework 241
    12.2.3 A Service-oriented Architecture 242
Preface

Just remember the set of services offered by the Internet a few years ago – emails, web services, sometimes experimental voice services, over what used to be referred to as a ‘high-speed’ connection of a few hundred kbits/second! The Internet has gone through a profound transformation and has been evolving at an unprecedented rate compared with other industries, thus becoming the central component of all forms of communication: data (emails, web services, search engines, peer-to-peer, e-commerce, stock trading, etc.), voice but also video (TV broadcasting, videoconferencing).

New innovative applications and services will undoubtedly continue to emerge, and we are still at an early stage of what the Internet will be able to provide in the near future. With no doubt, the impact of the Internet on how people communicate around the world and access to information will continue to increase rapidly. New forms of communication will arise such as tele-presence, ubiquitous services and distributed gaming, and the Internet will ineluctably extend its reach to ‘objects’, which is sometimes referred to the ‘Internet of things’, with billions of objects interconnected with each other and new forms of machine-to-machine communication. This new era of services will lead to endless possibilities and opportunities in a variety of domains.

The offering of a wide range of new services has required the design of networking technologies in the form of sophisticated protocols and mechanisms based on open standards driven by the Internet Engineering Task Force (IETF). The non-proprietary nature of the Internet Protocol (IP) led to interoperable solutions, thus making the Internet a unique platform of innovation.

As a direct implication of the Internet becoming critical to our personal and professional lives, user expectation has become very high in terms of reliability, quality of service (QoS) and security. A network failure of a few minutes is now considered as unacceptable! Fast network failure detection and traffic rerouting mechanisms have been designed to find alternate paths in the network within the timeframe of a few milliseconds while maintaining path quality.

Fine granularity in terms of QoS is now a must: although some applications are inherently delay tolerant (e.g. asynchronous communications such as emails), other traffic types impose bounded delays, jitters and reliability constraints that require complex configuration tasks to engineer the network. QoS guarantees imply traffic classification at the edge of the network, sophisticated local forwarding techniques (multipriority scheduling and traffic discard) and traffic engineering.

The ability to effectively engineer the traffic within the network is now of the utmost importance and is known as a fairly difficult task for service providers considering the high
volume of varying traffic. Furthermore, service providers have to engineer the network carefully in order to meet the quality of services imposed by demanding applications while having to deal with resource constraints. Security has become a central component: user identification and authentication and protection against attacks of different forms, including denial of service (DoS) attacks, require the configuration of complex networking technologies. Last but not least, the ability to efficiently manage and monitor the network is an absolute requirement to check service level agreements, enforce policies, detect network faults and perform network troubleshooting to increase the network availability.

A considerable amount of attention has been paid to service automation, network provisioning and policy enforcement. Network technology designers have been actively working on various tools to effectively provision, configure and monitor the network with sophisticated network components so as to ensure the toll quality that the Internet is now delivering, far from the ‘best effort’ service of the early days of the Internet. These tasks are increasingly crucial and complex, considering the diversity of the set of services provided by the Internet and the scale at which such tasks must be performed, with hundreds of millions of end-users, hundreds of services and a very significant traffic growth.

This is the right book at the right time, and the authors are known for their deep level of expertise in this domain. The organization of the book is particularly well suited to the topic. The first part examines the protocols and architecture required for network provisioning and policy enforcement in IP/MPLS networks. However, a book on this key subject would not be thorough without a strong emphasis on issues of a practical nature, and this is what the second part of the book is about. A number of highly relevant examples are provided on QoS, traffic engineering and virtual private networks, ideally complementing the theory expounded in the first part of the book.

**JP Vasseur**

*Cisco Distinguished Engineer*

*Chair of the IETF Path Computation Element Working Group*
Acknowledgements

*Christian*  
To my wife Béatrice and my sons Pierre and Paul, with all my love

*Gilles*  
To my wife and my son

*Mohamed*  
To my parents and my wife, with all my love
Part I

Architectures and Protocols for Service Automation
1

Introduction

1.1 To Begin With

The Internet has become a privileged playground for the deployment of a wide range of value-added IP service offerings. These services rely upon the combination of complex yet advanced capabilities to forward the corresponding traffic with the desired level of quality, as per a set of policies (in terms of forwarding, routing, security, etc.) that have been defined by the service provider, and sometimes negotiated with the customers.

This is a book about techniques that allow the dynamic enforcement of such policies.

Before discussing the motivation for such a book and detailing its organization, this chapter begins with an introductory reminder about the basics of IP networks. A 30 000 ft overview of the Internet as we know it.

1.1.1 On IP Networks in General, and Routers in Particular

An IP network is a set of transmission and switching resources that process IP traffic. The IP traffic is composed of protocol data units (PDUs) (RFC 791 [1]), which are called datagrams. The transmission resources of an IP network rely upon various link-level transport technologies, such as asynchronous transfer mode (ATM), synchronous digital hierarchy (SDH), etc.

The switching resources of an IP network are called ‘routers’. IP routers are in charge of processing each IP datagram, as per the following chronology:

- Upon receipt of a datagram, the router analyzes the contents of the destination address field of the datagram. This allows the router to identify the output interface through which the IP datagram will be forwarded, according to the contents of the forwarding information base, or FIB. An FIB of an IP router is typically composed of a set of {next hop; IP network} associations. The first member of these associations corresponds to the interface identifier of the next router capable of processing the datagram whose

---

Service Automation and Dynamic Provisioning Techniques in IP/MPLS Environments  C. Jacquenet, G. Bourdon and M. Boucadair  © 2008 John Wiley & Sons Ltd
destination address field corresponds to the IP network (expressed as an IP address) which is the second member of the pair.

- The analysis of the FIB allows the router to perform the switching features that will direct the datagram to the appropriate output interface through which the next hop router’s interface identified in the aforementioned pair can be reached.
- Then the router performs the forwarding task which will actually transmit the datagram over the selected output interface.

Thus the forwarding of an IP datagram relies upon the hop-by-hop paradigm owing to the systematic identification of the next router on the path towards the final destination [2–4]. Note also that Postel [1] also mentions the source routing mode, where the path to be followed by IP datagrams can either be partially (‘loose source routing’) or fully (‘strict source routing’) defined by the source that sends the IP datagram.

An FIB of an IP router is fed by information that comes from the use of a routing process, which can be either static or dynamic. In the case of static routing, the set of paths towards destination prefixes is manually configured on every router of the network.

In the case of a dynamic routing process, the FIB is dynamically fed by information that is stored and maintained in a specific table – the routing information base (RIB). There are at least as many RIB databases as routing protocols activated on the IP router.

The IP routers, which are operated by a globally unique administrative entity within the Internet community, form an autonomous system (AS) (see Figure 1.1) or border gateway protocol (BGP) domain (RFC 4271 [5]). From a typological standpoint, an AS is composed of a set of routers, thus yielding the distinction between the inner of an AS and the outer of an AS. The outer of an AS is the rest of the Internet.

![Figure 1.1](image-url) The Internet organized into autonomous systems
1.1.2 On the Usefulness of Dynamic Routing Protocols in IP Networks

The deployment of IP networks of large scale (such as those that compose today’s Internet) has rapidly led to the necessity of using dynamic routing protocols, so that routers might determine as efficiently as possible (that is, as fast as possible) the best route to reach a given destination (such an efficiency can be qualified in terms of convergence time).

Protocol convergence can be defined as the time it takes for a routing protocol to compute, select, install and disseminate the routing information [that is, the required information to reach a (set of) destination prefix(es)] at the scale of a region, be it an OSPF area or a BGP domain. That is, for a given destination prefix, a converged state is reached when information regarding this prefix has been added/modified or withdrawn in all relevant databases of the routers in the region. Traffic for a ‘converged’ prefix should be forwarded consistently inside the region.

As a matter of fact, static routing reveals itself as being incompatible with the number of IP networks that currently compose the Internet, because the static feeding of the FIB databases (which may therefore contain tens of thousands of entries, as per http://bgp.potaroo.net/) is a tedious task that may obviously impact upon the forwarding efficiency of such IP networks, because of network failures or congestion occurrences. Indeed, static routing leads to ‘frozen’ network architectures, which cannot adapt easily to the aforementioned events, unlike dynamic routing.

Dynamic routing protocols therefore allow routers to dynamically exchange network reachability information. Such information is stored in the RIB bases of these routers (as mentioned above) and is dynamically refreshed. The organization of the Internet into multiple autonomous systems yields the following routing protocol classification:

- dynamic routing protocols making it possible to exchange reachability information about networks that are part of the autonomous system: such protocols are called interior gateway protocols, or IGP;
- dynamic routing protocols making it possible to exchange reachability information about networks that are outside the autonomous system: such protocols are called exterior gateway protocols, or EGP.

Figure 1.2 depicts such a classification. Note that the white arrow of the figure should not be understood as a limitation of EGP exchanges that would be restricted to inter-AS communications. As a matter of fact, there are also BGP exchanges within domains.

These dynamic routing protocols use a specific algorithm whose calculation process takes into account one or several parameters which are often called metrics. These metrics are used by the routing algorithm to enforce a routing policy when the administrator of an IP network has the ability to actually define (and possibly modify) the values of such metrics.

Among the most commonly used metrics, one can cite:

- the number of routers (hop count metric) to cross before reaching a given destination [the fewer the routers, the better will be the route, whatever the characteristics of the links (in terms of speed, among others) that interconnect the routers];
- the cost metric, the meaning of which is broader than the previous hop count metric, and which generally reflects a weight assigned to an interface, a transmission link, the crossing of an autonomous system or a combination of these components.
The nature of the routing algorithms yields another typological effort, which consists in distinguishing the following:

- Routing protocols using algorithms based upon distance-vector calculation. Such an algorithm is generally inspired by the Bellman–Ford probabilistic calculation.
- Routing protocols using algorithms that take into account the state of the links interconnecting the routers. Such routing protocols are called ‘link-state’ routing protocols, and their algorithms are generally based upon the use of the Dijkstra probabilistic calculation.

Table 1.1 provides a summary of the principal IGP-specific characteristics of both distance-vector and link-state routing algorithms.

The very first IGP to be specified, standardized, developed and implemented by router vendors was the routing information protocol (RIP) (RFC 1058 [6], RFC 2453 [7]) back in 1984. The route selection process of RIP relies upon the use of a distance-vector calculation, directly inspired from the Bellman–Ford algorithm.
An example of a link-state routing protocol is the open shortest path first (OSPF) protocol (RFC 2328 [8]), which is supported by most of the routers on the market.

### 1.1.3 On the Inability of an IGP to Address Interdomain Communication Needs

The organization of the Internet into autonomous systems does not necessarily justify the aforementioned IGP/EGP typology, since the network reachability information exchange between autonomous systems is primarily based upon the use of a dynamic routing protocol, whatever this protocol might be (static routing between ASs is not an option, for the reasons mentioned in Section 1.1.2).

Therefore, why not use an IGP protocol to exchange network reachability information between autonomous systems? Here is a couple of reasons:

1. A router that activates a distance-vector routing protocol advertizes to its neighbors the whole set of networks it can reach. This information is displayed as a vector list that includes the cost of the path associated with each network. Each router of the network builds its own RIB database according to the information contained in these vector lists.
but this information does not provide any clue concerning the identity of the routers and the networks that have to be crossed before reaching a given destination. This may present some difficulty when exchanging such reachability information between autonomous systems:

- The distance-vector routing protocol states that all the routers running it have a common understanding of the metric that allows them to select a next hop rather than another. This common understanding may not be the case for routers belonging to different autonomous systems.
- The routing policy that has been defined within an autonomous system might be such that communication with specific autonomous systems is forbidden (e.g. for exchanging specific network reachability information). A distance-vector routing protocol has no means to reflect such filtering capabilities in the vector lists it can propagate.

2. A router that activates a link-state routing protocol advertizes network reachability information which is partly composed of the costs associated with the links that connect the router to adjacent networks, so that each of these routers has the ability to build up a complete image of the network topology. This advertisement mechanism relies upon the use of a flooding capability, which may encounter some scalability issues when considering communication between autonomous systems:

- The autonomous systems do not necessarily have a common understanding of the metrics that are used to compute a shortest path, so that the topological information that is maintained by the routers may be dramatically different from one autonomous system to another.
- The aforementioned flooding capability of a link-state protocol can rapidly become incompatible with networks of large scale (in terms of the number of routers composing a given domain), especially when considering the traffic volume associated with the broadcasting of network reachability information.

The basic motivation that yielded the specification, the standardization and the development of routing protocols of the EGP type was based upon the following information: since the metrics used by IGP routing protocols can be understood differently by routers belonging to different autonomous systems, the network reachability information to be exchanged between autonomous systems should rely upon other metrics.

Thus, a router belonging to autonomous system A would advertize to autonomous systems B, C, etc., the networks it can reach, including the autonomous systems that have to be crossed to reach such networks. This very basic concept is used by EGP routing protocols, and it is called ‘path-vector routing’.

An EGP routing protocol has the following characteristics:

- The information exchanged between routers that belong to different autonomous systems does not contain any clues about the use of a specific metric, or the value of any cost.
- The information exchanged between routers that belong to different autonomous systems describe a set of routes towards a set of destination prefixes. The description of such routes includes (but is not necessarily limited to) the number and the identity of the autonomous systems that have to be crossed to reach the destination networks.
The latter characteristic allows a router to enforce a routing policy that has been defined by the administrator of an autonomous system, so that, for example, this router could decide to avoid using a specific route because this route traverses autonomous systems whose degree of reliability is incompatible with the sensitive nature of the traffic that could use this route.

The forwarding of IP traffic over the Internet implies the crossing of several autonomous systems, thus yielding the activation of an EGP routing protocol. The BGP-4 (border gateway protocol version 4) protocol (RFC 4271 [9]) is currently the EGP that has been deployed over the Internet. The BGP protocol has arisen from the experience acquired during the very first stages of Internet deployment, especially through the deployment of the NSFNET (National Science Foundation NETwork), owing to the specification and the implementation of the exterior gateway protocol (EGP) (RFC 904 [10], RFC 1092 [11], RFC 1093 [12]).

1.1.4 On the BGP-4 Protocol

The principal feature of a BGP-4-enabled router consists in exchanging reachability information about IP networks (aka IP destination prefixes) with other BGP-4-enabled routers. Such information includes the list of the autonomous systems that have been crossed, and it is sufficiently specific for it to be possible to build up an AS connectivity graph from this information.

This AS connectivity graph will help BGP-4-enabled routers in avoiding routing loops (which result in the development of IP network-killing ‘black holes’), and it will also help in enforcing the routing policies that have been defined by the AS administrator.

The BGP protocol relies upon transmission control protocol (TCP) port 179 (RFC 793 [13]) – a transport layer-specific protocol that supports fragmentation, retransmission, acknowledgement and sequencing capabilities.

The BGP communication between two routers can be briefly described according to the following chronology:

- The BGP routers establish a TCP connection between themselves by exchanging messages that aim to open this connection, then confirming the parameters that characterize this connection.
- Once the TCP connection has been established, the very first exchange of (reachability) information is composed of the overall contents of the BGP table maintained by each peer.
- Then, information is exchanged on a dynamic basis. This information actually represents specific advertisements every time the contents of one or the other BGP tables have changed. Since the BGP-4 protocol does not impose a periodic update of the global contents of the BGP routing table, each router must keep the current version of the global contents of all the BGP routing tables of the routers with which it has established a connection.

Specific messages are exchanged on a regular basis, so as to keep the BGP connection active, whereas notifications are sent in response to a transmission error or, more generally, under specific conditions. The receipt of a notification results in the BGP communication breakdown between the two BGP peers, but such a breakdown is smoothed by the TCP
protocol, which waits for the end of the ongoing data transmission before effectively shutting down the connection. Although the BGP-4 protocol is a routing protocol of the EGP type, routers that belong to the same autonomous system have the ability to establish BGP connections between themselves as well, which yields the following typology:

- The connections that are established between BGP routers belonging to different autonomous systems are called ‘external sessions’. Such connections are often named ‘external BGP’ or ‘eBGP’ connections.
- The connections that are established between BGP routers belonging to the same autonomous system are called ‘internal sessions’. Such connections are often named ‘internal BGP’ or ‘iBGP’ connections.

iBGP connections are justified by the will to provide (to the BGP routers belonging to the same autonomous system) as consistent a view of the outside world as possible. Likewise, an IGP protocol provides a homogeneous view of the internal routes within an autonomous system.

A BGP route (i.e. the reachability information that is transmitted within the context of the establishment of a BGP connection) is made up of the association of an IP prefix and the attributes of the path towards the destination identified by this prefix. Upon receipt of such information, the router will store it in the BGP routing table, which is actually made up of three distinct tables:

- The Adj-RIB-In table, which stores all the advertised routes received by a BGP peer. This information will be exploited by the BGP decision process.
- The Adj-RIB-Out table, which stores all the routes that will be advertised by a BGP peer. These are the routes that have been selected by the BGP decision process.
- The Loc-RIB table, which stores all the routes that will be taken into consideration by the BGP decision process. Among these routes there will be those that are stored in the Adj-RIB-Out.

The distinction between these three tables is motivated by the BGP route selection process. In practice, most of the BGP-4 implementations use a single BGP routing table, which will be indexed appropriately according to the above-mentioned typology.

1.1.5 The Rise of MPLS

The hop-by-hop IP routing paradigm of the old days of the Internet (as introduced in Section 1.1.1) is being questioned by the multiprotocol label switching (MPLS) technique (RFC 3031 [14]). MPLS is a switching technique that allows the enforcement of a consistent forwarding policy at the scale of a flow, where a flow can be defined as a set of IP datagrams that share at least one common characteristic, such as the destination address.

In this case, all the IP datagrams of a given flow [designated as a forwarding equivalence class (FEC) in the MPLS terminology] will be conveyed over the very same path, which is called a label switched path (LSP) (see Figure 1.3).

MPLS switching principles rely upon the content of a specific field of the MPLS header, which is called the label. Labels are the primary information used by MPLS-enabled routers
to forward traffic over LSP paths. MPLS has been defined so that it can be used whatever the underlying transport technology, or whatever the network layer-specific communication protocol, such as IP. The MPLS forwarding scheme is depicted in Figure 1.4.

The MPLS forwarding scheme relies upon the maintenance of label tables, called label information bases (LIBs). To forward an incoming MPLS packet, the MPLS-enabled router will check its LIB to determine the outbound interface as well as the outgoing label to use, based upon the information about the incoming interface as well as the incoming label. As per the example provided by Figure 1.4:

- Router A of the figure, which does not support MPLS forwarding capabilities, is connected to (or has the knowledge of) networks N1 and N2, which can be reached through its Ethernet 0 (E0) interface. Table 1.2 is an excerpt from its FIB, which basically lists the network prefix, the outgoing interface and the associated next hop router.
- The black arrow in Figure 1.4 suggests that an ordinary routing update (by means of a dynamic routing protocol, such as OSPF), advertizes the routes to the MPLS-enabled router [or label switch router (LSR) in the MPLS terminology], which is directly connected to router A.
- Using the label distribution protocol (LDP) (RFC 3036 [15]), router 1 selects an unused label [label 3 in the example provided by the excerpt of its label information base below (Table 1.3)] and advertizes it to the upstream neighbor. The hyphen in the ‘Label’ column of Table 1.3 denotes that all labels will be popped (or removed) when forwarding the
packet to router A, which is not MPLS capable. Thus, an MPLS packet received on the serial 1 interface with label 3 is to be forwarded out through the serial 0 interface with no label, as far as LSR 1 which is directly connected to router A is concerned. The white arrow in Figure 1.4 (between router 1 and router 2) denotes the LDP communication that indicates the use of label 3 to the upstream LSR 2.

LSR 1 has learned routes that lead to N1 and N2 network prefixes. It advertizes such routes upstream. When LDP information is received, router 1 records the use of label 3 on the outgoing interface serial 0 for the two prefixes mentioned previously. It then allocates label 16 on the serial 1 interface for this FEC and uses LDP to communicate this information.

**Table 1.2** Excerpt from the forwarding information base of router A (as per Figure 1.4)

<table>
<thead>
<tr>
<th>Network</th>
<th>Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>E0</td>
</tr>
<tr>
<td>N2</td>
<td>E0</td>
</tr>
</tbody>
</table>
to the upstream LSR. Thus, when label 16 is received on serial 1, it is replaced with label 3 and the MPLS packet is sent out through serial 0, as per Table 1.4.

Note that there will be no labels received by router B (and sent by router 4 in the figure), since the top router B is not an LSR, as illustrated by its routing table (no labels are maintained in this table). The label switched path (LSP) is now established.

Note also that MPLS labels can be encoded as the virtual path identifier/virtual channel identifier (VPI/VCI) information of an ATM cell, as the data link connection identifier (DLCI) information of a frame, in the sense of the frame relay technology, but also as 20-byte long information encoded in the 4-byte encoded MPLS header associated with each IP PDU, as depicted in Figure 1.5.

<table>
<thead>
<tr>
<th>Label</th>
<th>EXP bits</th>
<th>Stack</th>
<th>Time To Live (TTL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 bits</td>
<td>3 bits</td>
<td>1 bit</td>
<td>8 bits</td>
</tr>
</tbody>
</table>

Figure 1.5 The MPLS header

MPLS capabilities are now supported by most of the router vendors of the market, and the technique is gaining more and more popularity among service providers and network operators, as the need for traffic engineering capabilities emerges. Traffic engineering is the ability to (dynamically) compute and select paths whose characteristics comply with requirements of different kinds: the need to make sure that a given traffic will be conveyed by a unique path (potentially secured), e.g. for security purposes, or the need for minimum transit delays, packet loss rates, etc.

MPLS-based traffic engineering capabilities can be seen as some of the elementary components of a global quality of service (QoS) policy.

1.2 Context and Motivation of this Book

IP service offerings (ranging from access to the Internet to more advanced services such as TV broadcasting or videoconferencing) are provisioned owing to the combined activation of different yet complex capabilities, which not only require a high level of technical expertise but also result in the organization of complex management tasks.
1.2.1 Classifying Capabilities

As stated above, IP services are provided by means of a set of elementary capabilities that are activated in different regions and devices of an IP/MPLS network infrastructure. These capabilities can be organized as follows:

- **Architectural** capabilities, which are the cornerstones for the design and enforcement of addressing, forwarding and routing policies. Such policies aim to convey service-specific traffic in an efficient manner, e.g. according to the respective requirements and constraints that may have been (dynamically) negotiated between the customer and the service provider.
- **Quality of Service** (QoS) capabilities, as briefly introduced in Section 1.6.
- **Security** capabilities, which include (but are not necessarily limited to):
  - the user and device identification and authentication means;
  - the protection capabilities that preserve any participating device from any kind of malicious attacks, including (distributed) denial of service (DDOS) attacks;
  - the means to preserve the confidentiality of (some of) the traffic that will be conveyed by the IP network infrastructure;
  - the means to protect users and sites from any kind of malicious attack that may be relayed by the IP/MPLS network infrastructure;
  - the functions that are used to check whether a peering entity is entitled to announce routing information or not, and also the features that provide some guarantees as far as the preservation of the integrity (and validity) of such (routing) information is concerned.
- **Management** capabilities, composed of fault, configuration, accounting, performance and security (FCAPS) features. Monitoring tools are also associated with such features. They are used for analysis of statistical information that aims to reflect how efficiently a given service is provided and a given policy is enforced.

1.2.2 Services and Policies

The management tasks that are performed to provision and operate an IP network or a set of IP service offerings can be grouped into several policies that define what capabilities should be activated, and how they should be used (that is, the specification of the relevant configuration parameters).

Policies can relate to a specific service [e.g. the forwarding policy to be enforced at the scale of a BGP domain to convey voice over IP (VoIP) traffic with the relevant level of quality], or can be defined whatever the nature of the service offerings (e.g. the BGP routing policy to be enforced within a domain).

The design and the enforcement of a given policy must therefore address a set of elementary questions, as follows:

- **Why**? This is what this book is about – the need for policies to facilitate the automation of sometimes tedious management tasks (configuration of routers to support different services, identification of the users entitled to access a service, etc.) that need to be