Università degli Studi di Cagliari
Dipartimento di Matematica e Informatica

Corso di Dottorato in

Informatica

Technologies, Routing Policies and Relationships between Autonomous Systems in Inter-domain routing

Tesi di Dottorato di

Massimiliano Picconi

Supervisor:

Prof. Gianni Fenu

Anno Accademico 2009/2010

to Alessia

Contents

List of Figures

List of Tables

Introduction

One of the most complex problems in computer networks is controlling the routing decisions in inter-domain routing.

Competing entities called Autonomous Systems (ASes), operated by many different Administrative Domains, must cooperate with each other in order to provide global Internet connectivity according to business and commercial agreements, using an inter-domain routing protocol able to apply local policies for selecting routing and distributing reachability information to each other.

A fully independent management provided by each AS makes the problem of controlling inter-domain routing difficult to solve, due to potentially conflicting policies essentially derived from the business agreements and competition between domains, and from the feature of the inter-domain routing protocol of advertising only a best path for a destination for performance and reliability reasons.

The goal of this thesis is a deeply exploration of the issues related to

routing decisions in inter-domain routing, with an analysis of the interconnection structure and the network hierarchy, the examination of the inter-domain routing protocol used to exchange network reachability information with other systems, the examination of the routing decision process between the entities according to the attributes and the policies and the study of the topology generators of the AS relationships, reviewing the most interesting proposals in this area, describing why these issues are difficult to solve, and proposing solutions allowing to better understand the routing process and optimally solve the trade-off of implementing a peering engagement between two Administrative Domains, against the extra cost that this solution represent.

More specifically, the objectives in this thesis are described as follows:

1. a deep analysis of the current inter-domain network model, and of the aspects related to the routing decision process inside a BGP speaker according to the policies and the filters over the updates. A new scheme for the Routing Process Model is proposed, and a formalization of a new and more complex routing process model inside a BPG speaker is made;

2. a deep analysis of the routing decision process between the Autonomous Systems, according to the attributes that administrators can apply to control the policies.

   Competing ASes must cooperate with each other in order to pro-

vide global Internet connectivity (cooperation), according to their business and commercial agreements (competition), using a routing process model to set their policy independently to each others (autonomy), and making all the manipulations allowed by the protocol (expressiveness), without any global coordination;

3. an analysis of the problem of controlling inter-domain routing, due to potentially conflicting policies derived from the business agreements and competition between domains, that can lead to routing anomalies such as instability, divergence in the update process, delays;

4. an analysis of the topology of the relationships and agreements between ASes to obtain a complete and accurate AS-level connectivity, and of the interconnection structure of the Administrative Domains, in order to understand the topological structure of the system;

5. an exploration of strategies for competitions between Administrative Domains to identify a Settlement Model from the business relationships. In particular, strategies for peering competitions are discussed;

6. an organization of the different aspects to be taken into account in a peering engagement. A methodology to structure these aspects is proposed to optimally solve the trade-off of implementing a peering engagement against the extra cost represented by this solution;

7. a formulation of the decision problem of maximization of the impor-

tance of the aspects and alternative options to be taken into account in ex-ante evaluations of a peering engagement, subject to their mutual relationship with budget constraints.

The problem has been expressed as an integer programming formulation, and a practical implementation of a decision maker framework called $XESS^2$ (eXtended EGP Support System) has been proposed, in order to find candidate solutions and to produce a synthetic conclusion on the allocation of budgets. It is able to make a comparative evaluation of alternative options through a numerical evaluation of a set of variables, and to find an optimal reduced set of solutions using a combinational optimization formulation and an integer programming formulation of the problem.

The thesis is organized in seven chapters.

Chapter 1 introduces the basics and the necessary background to understand the mechanisms and the techniques proposed along this thesis. The fundamental concepts of intra-domain and inter-domain routing, the network hierarchy and the types of interconnection between ASes are explained, including a description of the main features of the routing protocol and its policy-based nature. In particular, the Border Gateway Protocol (BGP), and the BGP message formats are deeply analyzed according to RFC 4271.

Chapter 2 covers the aspects of the routing decision process inside a BGP speaker, according to the attributes that administrators can apply to control the policies. In particular, mechanisms to run policies and filters over the updates, and to add or modify a route's path attribute before advertising it to a neighbor are analyzed.

In addition, the traditional CISCO's Routing Process Model is described, and a new and more complete Routing Process Model is proposed.

Finally, a formalization of the Decision Process with the definition of the *attribute function*, the *Degree of Preference function* and the *Route Selection function* is proposed.

Chapter 3 covers the aspects of the routing decision process between BGP speakers in inter-domain routing, exposing the major issues related to routing decisions through the review of the most interesting proposals in this area, and describing why these issues are difficult to solve.

In particular, issues related to BGP divergence derived from distributed conflicting routing policies, to the exploration of alternative paths when a path failure or routing policy changes occur, to well-known methodologies such as hot-potato routing or prefix hijacking, and in-band and out-of-band solutions are analyzed.

Chapter 4 exposes the major limitations of the research area related to the study of the AS relationships and the development of accurate topology

generators through the review of the most interesting papers in this area.

In particular, issues related to the commercial agreements model, to the interconnection structure model, focusing on the generation of synthetic AS topologies, and to the exporting policies model, are deeply discussed.

Chapter 5 introduces the analysis of the business relationships and the possible strategies for competitions between Administrative Domains, in order to identify a Settlement Model of the transactions.

Two types of business agreements between Administrative Domains to provide reachability are considered, related to *Transit services,* where one Administrative Domain provides reachability to all destinations in its routing table to its customers, and to *Peering services*, where Administrative Domains provide mutual reachability to a set of their routing table.

In particular, a deep analysis of the peering engagement is made, with the description of the types of peering interconnection and peering relationships.

Chapter 6 proposes a formulation of a methodology to structure the different aspects to be taken into account in a peering engagement, in order to optimally solve the trade-off of implementing a process of peering engagement against the extra cost that this solution represent.

In particular, a comparative analysis of the aspects and alternative options to be taken into account in ex-ante evaluations of a peering en-

gagement is proposed, and a decision maker called $XESS^2$ (eXtended EGP Support System) able to process the aspects and the alternative options, to find candidate solutions, and to produce a synthetic conclusion on the allocation of budgets in a peering engagement is explained.

$XESS^2$ is designed to help decision-makers to integrate the different options and to produce a single synthetic conclusion at the end of the evaluation through a combinational optimization formulation and an integer programming formulation of the problem.

Chapter 7 highlights the main conclusions of this thesis, focusing on the analysis of the issues related to the routing decision process in the interdomain routing generated by a lack of a global global coordination, demonstrating to be inaccurate and poorly effective in controlling and communicating the inter-domain decisions, and on the development of solutions aimed at optimally implementing of a peering engagement between multiple solutions with different monetary costs, proposing several areas of interest for future implementations of the work done in this thesis.

Chapter 1

Background

1.1 Terminology and Concepts

The Internet is decentralized collection of autonomous network, independent entities with its own policies, services, and customer targets. The performance of the communications over these entities depends on the routing process, that allows all networks to interconnect with each other directly or indirectly. The routing process determines how packets are treated and forwarded through network using a common IP addressing.

Each of these networks, known as *Autonomous System (AS)*, appears as a single coherent entity to other networks, with a common routing policy, and managed by a single administration authority. Each AS has the responsibility to route the traffic of a set of customer IP addresses,

and the scalability of the Internet routing infrastructure depends on the aggregation of IP addresses in contiguous blocks, called *prefixes*[1]. It is estimated that today's Internet is an interconnection of more than 26000 ASes [25].

RFC 4271 [126] defines the Autonomous System as "a set of routers under a single technical administration, using an interior gateway protocol (IGP) and common metrics to determine how to route packets within the AS, and using an inter-AS routing protocol to determine how to route packets to other ASes".

For the purpose of this work, each administration authority is called *Administrative Domain (AD)* as defined in RFC 1125 [37], as a set hosts and network resources that is governed by common policies[2]. Each AD may have the control of one or more ASes[3] as illustrated in Figure 1.1.

Examples of Administrative Domains range from universities and corporate networks to large Internet Service Providers (ISPs) such as AT&T.

[1]The term *prefix* indicates an aggregation of IP addresses in contiguous blocks consisting of a 32-bit IP address and a mask length (e.g., 193.43.2.0 255.255.255.0 or 193.43.2.0/24)

[2]RFC 1125 defines the Administrative Domain referred to the *Research Internet*, "the collection of government, university, and some private company, networks that are used by researchers to access shared computing resources (e.g., supercomputers), and for research related information exchange (e.g., distribution of software, technical documents, and email)" [37].

[3]The relationship between ASes inside an Administrative Domain are called *sibling relationships* when each AS export all of its routes to the other ASes [136] [53].

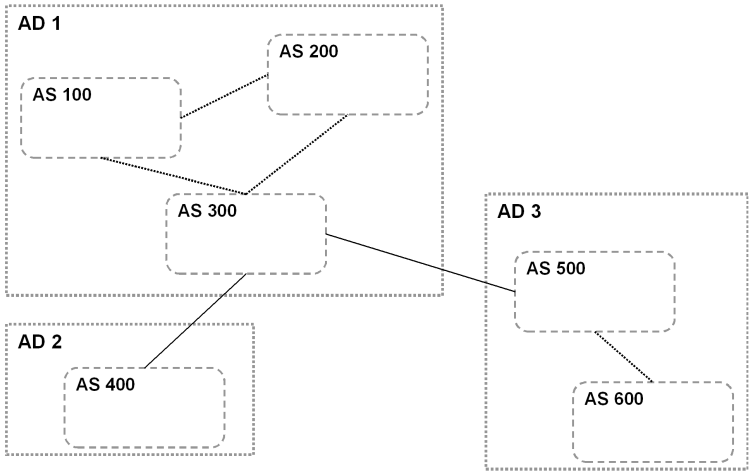

Figure 1.1: Administrative Domains and Autonomous Systems.

Each AS is represented by an identifying number, called *AS Number (ASN)*[4]. An ASN is a 16 bit number (65536 AS numbers), and some ASN are for private use or reserved.

The registration and the administration of IP and AS numbers is possible via an Internet Registry, as ARIN (an Internet registry that provides the WHOIS[5] lookup service in North America, South America, the Caribbean, and sub-Saharan Africa), RIPE NCC (which provides services for Europe, the Middle East, and parts of Africa), and APNIC (which provides services for Asia Pacific).

[4]There is no one-to-one relationship between AS numbers and ISPs.

[5]The WHOIS service provides information about each AS such as the name and address of the administrative domain that the AS belongs to.

Inter-AD routes are selected according to policy-related parameters (e.g., cost, access rights), in addition to the traditional parameters of connectivity and congestion. A *Policy Routing (PR)* is needed to navigate through the policy boundaries created by numerous interconnected ADs.

In addition, each AD has its own privileges and perspective of the network, and therefore it makes its own evaluation of legal and preferred routes. Today, there is little regulation, and each AD is free to decide where, how, and with whom to connect.

In the next sections, an brief explanation of intra-domain e inter-domain routing is made.

1.1.1 Routing Policies

An Autonomous System employs an *intra-domain (intra-AS)* routing to determine how to reach each customer path, and an *inter-domain (inter-AS)* routing to determine the reachability of paths in other ASes.

Intra-domain routing is usually optimized in accordance with the required technical demands, with the transmission of the prefixes towards their destination using algorithms able to find the best path to each destination.

In the intra-domain routing, routers run an Interior Gateway Protocol (IGP) such as Routing Information Protocol (RIP) [75], Open Shortest Path

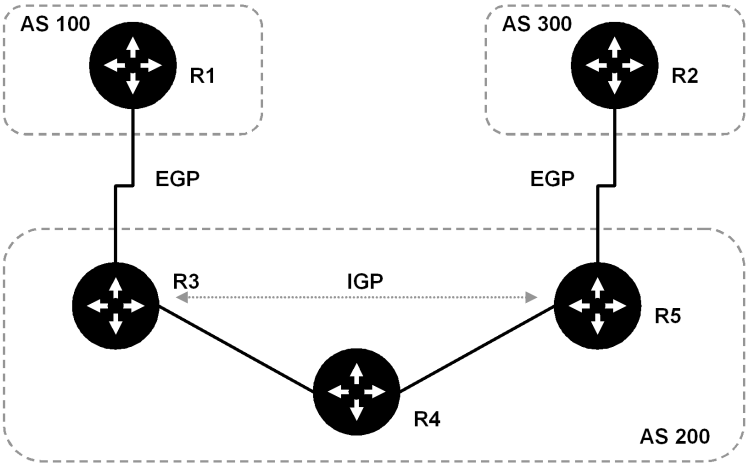

Figure 1.2: Autonomous Systems. EGP and IPG

First (OSPF) [116], or Intermediate System-to-Intermediate System (IS-IS) [79].

The interconnection between ASes is made via Exterior Gateway Protocols (EGPs), able to provide a more structured view of the Internet by segregating routing domains into separate administrations, and capable of solving the issues related to the scalability of IGP protocols in enterprise networks.

Inter-domain usually reflects political and business agreements between the networks and companies involved, imposing restrictions in traffic routing between ASes according to the routing policies of each domain involved in the routing process.

These policy-based metrics allow to override distance-based metrics in favor of policy concerns, enabling each AS to independently define its routing policies with little or no global coordination.

Inflated AS paths and a suboptimal routing are the results of these policies[6].

The current Internet *de facto* standard EGP is the *Border Gateway Protocol* Version 4 (BGP-4), defined in RFC 1771 [123] on March of 1995, and revised in RFC 4271 [126] on January of 2006. Figure 1.2 illustrates a scheme of a generic inter-domain scenario.

1.1.2 AS Interconnection

The interconnection between ASes can be applied via a single or via multiple connections (mainly used for load balancing reason and resilience).

The term *Multihomed AS* defines an AS which reaches a network outside its domain via multiple exit points belonging to a single entity or multiple entities.

Transit Traffic is defined as any traffic that has a source and destination outside an AS.

Three types of ASes are usually classified depending on the way they manage their transit traffic [124] [73]:

[6]Note that there is no one-to-one relationship between inter-domain routing and inter-AD routing. The inter-domain routing can be made inside and Administrative Domain.

1. *Single-Homed (Stub) AS*.

   A Single-Homed reaches networks outside its domain via a single exit point [Figure 1.3]. Most of the universities, the Internet Service Providers (ISPs) and Enterprise customers belong to this type of AS.

2. *Multihomed Nontransit AS*.

   A Multihomed Nontransit AS does not allow transit traffic to go through it [Figure 1.4]. It would only advertise the routes that has an origin or destination that belongs to the local AS and would not propagate routes learned from other ASes.

   In the scenario illustrated in Figure 1.4, AS 100 and AS 200 will learn X301 and X302, while AS 300 will learn X101, X102, X201 and X202. AS 300 advertises only its local routes X301 and X302.

3. *Multihomed Transit AS*.

   A Multihomed Transit AS allows transit traffic to go through it.

   In the scenario illustrated in Figure 1.5, AS 100 will learn X301, X302, X201 and X202 from AS300, AS 200 will learn X301, X302, X101 and X102 from AS 300, AS 300 will learn X101, X102, X201 and X202.

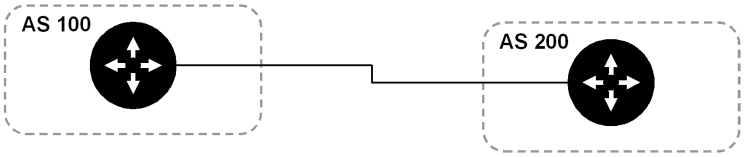

Figure 1.3: Single-Homed (Stub) AS

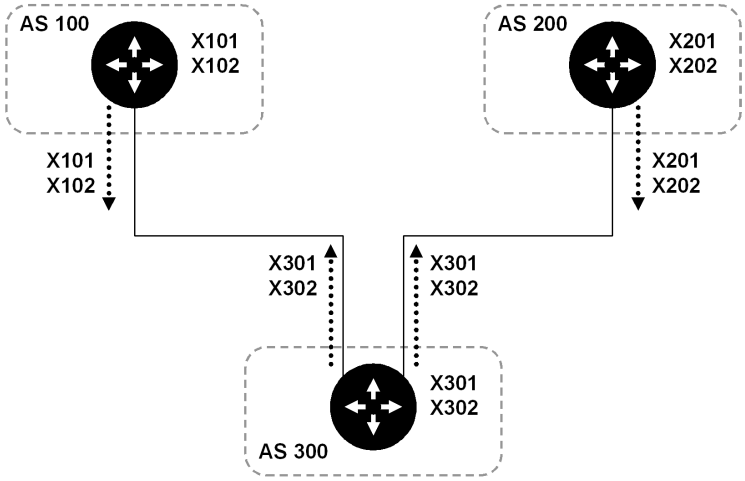

Figure 1.4: Multihomed Nontransit AS

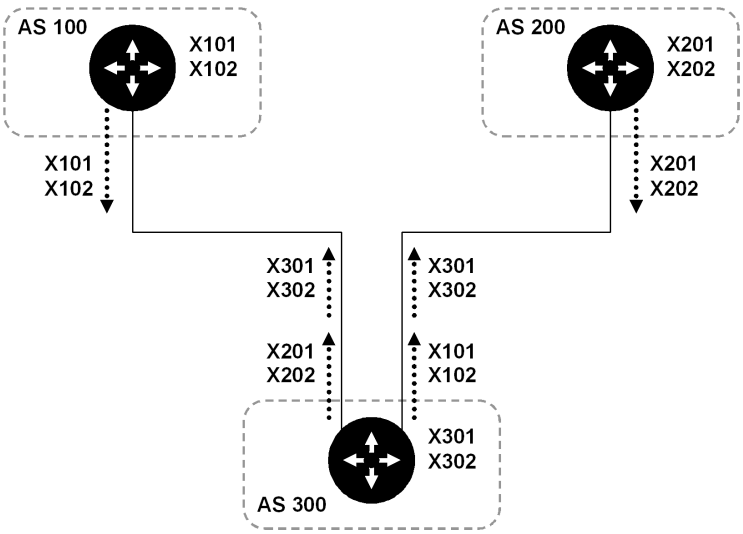

Figure 1.5: Multihomed Transit AS

1.1.3 The Network Hierarchy

Networks are often classified in Tier-1, Tier-2 and Tier-3 ISPs. Tier-1 ISPs are the largest ISPs interconnected with other Tier-1 ISPs via private peering[7]. They have the most direct control over the traffic that flows through their connections. However, they represent less than 0.1 % of the total number of ASes in the Internet [25]. Other ISPs are completely dependent on Tier-1 ISPs. Tier-1 ISPs are divided into Global and Regional Tier-1 ISPs. The key attribute of Global Tier-1 ISPs are:

- large size and scale;
- peering on more than one continent;
- no cost for the delivery of their traffic through similar-sized networks;
- access to the global Internet routing table via their peering relationships.

A Tier-2 ISP is any transit AS which is a customer of one or more Tier-1 ISPs. Often, they have lower-quality networks than Tier-1 ISPs and tend to establish peering relationships with other neighboring Tier-2 ISPs. A representative example of a Tier-2 ISP is a national service provider.

Tier-3 ISPs focus on local retail and consumer markets, with a coverage limited to a specific country or to subregions providing local access for end customers. They tend to have low-quality networks and access speed.

[7]Routing information and traffic exchange between AS occurs through a process called *peering*. More-detailed information is provided in the chapters thereafter.

1.2 Border Gateway Protocol

The performance of the communications over the Internet depends on the routing process that determines how packets are forwarded through network.

In the routing process, the routes to destinations can be injected manually in the router *static routing*. Whether a destination is active or not, the static routes remain in the routing table, and traffic is still sent toward the specified destination. The most stable, but less flexible configurations are based on static routing[8].

Another way of learning routes is via a routing protocol (*dynamic routing*). The next section discusses in detail the dynamic routing protocols for intra-domain and inter-domain routing.

1.2.1 Intra-domain Routing Algorithms

Routers inside an Autonomous System run Interior Gateway Protocols (IGPs) for intra-domain routing. All these routing protocols transmit IP packets towards their destination using algorithms able to find the best path to each destination.

Most intra-domain routing protocols used today are based on one of two types of routing algorithms: *distance vector* and *link-state* routing algo-

[8]Instead, the term *default routing* refers to a route used when a destination is unknown to the router.

rithms. A detailed discussion of distance vector and link-state algorithms is beyond the scope of this thesis. More-detailed information can be found in [73],[140] and [44].

A *Distance Vector* routing protocol includes a vector of distances associated with each destination prefix routing message. Each router separately computes the best path to each destination, and sends distance vectors to its neighbors, notifying them of the available path (and the corresponding metrics associated with the path) it has selected to reach the destination.

Distance vector protocols require that each node separately computes the best path to each destination. Every neighbor determines if a better path exists upon every message is receipted, and, in that case, it will update its routing table and will notify its neighbors of its selected paths. The cycle goes on until a *convergence* [9] towards a common topology is built. RIP [75] is a distance vector routing protocol.

A distance vector routing protocol has several drawbacks and limitations in the maintenance of large routing tables. It works on the basis of periodic updates and hold-down timers, translating into minutes in convergence time before the whole network detects a modification in the state of the network.

In a *Link-state* protocol, information elements (link states), which carry information about links and nodes, are exchanged by routers in the routing

[9] *Convergence* refers to the point in time at which the entire network becomes updated to the fact that a particular route has appeared, disappeared, or changed.

domain. There is no exchanging of routing tables. A flooding mechanism ensures the exchange of information related to the adjacent neighbors (including metric information associated with the connection).

Each routes uses these information to construct the network topology, and to build a tree of destination placing itself at the root, by applying the *Shortest Path First (SPF)* algorithm (Dijkstra Algorithm), in order to compute the shortest path to each destination.

Link-state algorithms provide fast convergence capabilities, and better routing scalability. However, several drawbacks and limitations have been associated with traditional link-state routing protocol handling on inter-domain routing.

Open Shortest Path First (OSPF) [116] is the commonly used link-state protocol.

1.2.2 Inter-domain Routing Algorithms

Routers outside an Autonomous System run Exterior Gateway Protocols (EGPs) for inter-domain routing.

A particular distance vector routing category called *path vector routing protocol* is defined. An additional mechanism referred to as the *path vector* is used to ensure a loopfree interdomain routing. A routing information carries a sequence of AS numbers that identifies the path of ASes that a network prefix has traversed. If an AS receives a routing information

containing its AS number, this route[10] is ignored.

Border Gateway Protocol (BGP) is a path vector routing protocol.

1.2.3 Border Gateway Protocol (BGP) concepts

In 1984, an Exterior Gateway Protocol called EGP was defined [108] in order to exchange reachability information between the backbone and the regional networks in NSFNET[11]. Due to several drawbacks and limitations (topology restrictions, inefficiency in dealing with routing loops) a new and more robust protocol called Border Gateway Protocol (BGP) [98] was defined in 1989 to ensure the interconnection and the reachability of different networks.

Today, *Border Gateway Protocol version 4 (BGP-4)* is the inter-domain routing protocol of choice on the Internet, in part because it efficiently handles route aggregation and propagation between domains.

Defined as inter-Autonomous System Routing protocol, as expressed in RFC 4271 [126] "The primary function of a BGP speaking system is to exchange network reachability information with other BGP systems.

[10] In this context, the term *route* represents a unit of information that pairs a destination with the attributes of a path to that destination. More-detailed information is provided in the sections thereafter.

[11] NSFNET was a network of multiple regional networks and peer networks (e.g., NASA Science Network) connected to a major backbone, born in 1985 and decommissioned in April 1995.

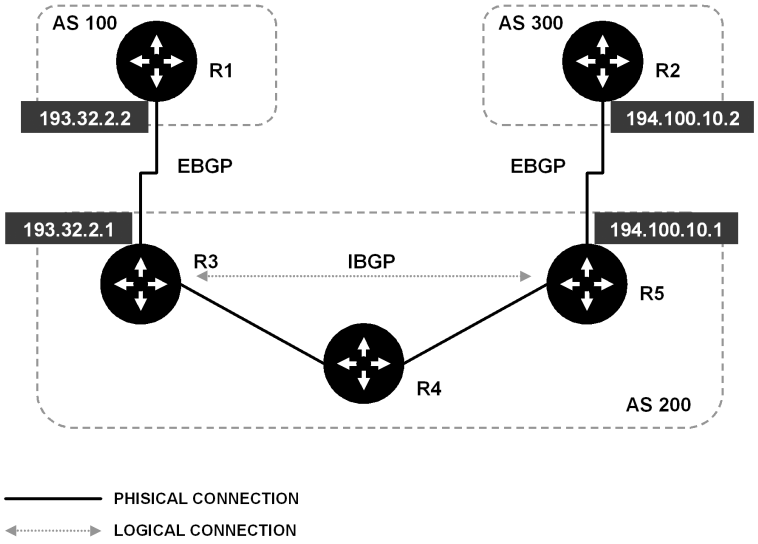

Figure 1.6: Internal and External BGP Implementations

This network reachability information includes information on the list of Autonomous Systems that reachability information traverses. This information is sufficient for constructing a graph of AS connectivity for this reachability from which routing loops may be pruned, and, at the AS level, some policy decisions may be enforced".

In addition, BGP-4 provides a set of mechanisms for supporting Classless Inter-Domain Routing (CIDR) [125], for the aggregation of a set of destinations in prefixes, and the aggregation of routes and AS paths.

BGP can be used inside an AS. BGP connections between routers inside an AS are called *Internal BGP (IBGP)*, while a peer connection between

routers in different ASes is referred to as *External BGP (EBGP)* [Figure 1.6].

Routing information exchanged via BGP supports only the *destination-based forwarding paradigm* (a router forwards a packet based solely on the destination address carried in the IP packet).

Routers that run a BGP routing process are referred to as *BGP speakers*. Two BGP speakers that exchange routing information are known as *neighbors* or *peers*. In Figure 1.6, routers R1 and R3 and routers R2 and R5 are BGP peers[12].

In the first phase, BGP neighbors exchange their full BGP routing tables. After the session has been established and the initial route exchange has occurred, only incremental updates are exchanged between BGP peers.

Information are injected into BGP dynamically depending on the status of the network. However, a statical injection of routes is possible, regardless of the status of the networks they identify. Today, the statical injection of information into BGP has proven to be the most effective method of ensuring route stability.

[12] **`neighbor remote-as`** command adds an entry to the BGP neighbor table. In Figure 1.6 the command on router R3 **`neighbor 194.100.10.1 remote-as 200`** indicates that a BGP peer session is to be established with the peer 194.100.10.1 in Autonomous System 200. So, also routers R3 and R5 are BGP peers.

1.2.4 BGP Message Formats

BGP messages are sent over TCP connections[13]. The maximum message size is 4096 bytes. The smallest message that may be sent consists of a BGP header without a data portion (19 bytes).

Each message has a fixed-size header. A message header format is a 16-byte *Marker* field, followed by a 2-byte *Length* field and a 1-byte *Type* field.

- The 16-byte *Marker* field is included for compatibility and it must be set to all ones[14] [126].
- The 2-byte *Length* field is used to indicate the total BGP message length, including the header. The value of the Length field must always be at least 19 and no greater than 4096.
- The 1-byte *Type* field indicates the type code of the message.
  - *OPEN (Type Code: 1)*
    It is the first message sent by each peer, after a TCP connection is established. A detailed information about the OPEN message format is provided in the RFC 4271 [126];

[13] See RFC4271 for details[126].

[14] The Marker field should be different if used as part of an authentication mechanism such as TCP MD5 Signature Option.

- *UPDATE (Type Code: 2)*

  Due to the importance of this message format, a detailed information is provided in the associated section thereafter;
- *NOTIFICATION (Type Code: 3)*

  A NOTIFICATION message is always sent whenever an error is detected. The BGP connection is closed immediately after it is sent;
- *KEEPALIVE (Type Code: 4)*

  KEEPALIVE messages are periodic messages exchanged between peers to ensure that the connection is kept alive. A KEEPALIVE message consists of only the message header and has a length of 19 bytes.

1.2.5 The UPDATE message

In a BGP routing process, only incremental updates are exchanged between BGP peers after the session has been established and the initial route exchange has occurred. This incremental update approach uses the type code 2 UPDATE message to exchange routing information between neighbors.

Basically, an UPDATE message contains a list of destinations that can be reached via a BGP speaker, and a set of path attributes containing information such as a list of ASes that the route has traversed and the degree of preference for a particular route.

The basic blocks of an UPDATE message consists of [Figure 1.7]:

- Unfeasible Routes;
- Network Layer Reachability Information (NLRI);
- Path Attributes.

These three blocks are discussed in the following subsections.

1.2.6 The UPDATE message. Unfeasible routes

The unfeasible routes represent the routes that are unreachable. When a route becomes unreachable, a BGP speaker informs its neighbors by removing the invalid route in the *withdrawn routes* field as 2-tuple format <length, prefix>.

An example of a BGP UPDATE message containing a set of withdrawn routes is shown below:

```
UPDATE Message
        Marker: 16 bytes
        Length: 35 bytes
        Type: UPDATE Message (2)
        Unfeasible routes length: 12 bytes
        Withdrawn routes:
            192.168.56.0/24
            192.168.20.0/24
            192.168.10.0/24
```

```
No.     Time        Source              Destination         Protocol Info
     31 75.287000   192.168.10.1        192.168.10.2        BGP   UPDATE Message

Frame 31 (100 bytes on wire, 100 bytes captured)
Cisco HDLC Internet Protocol,
Src: 192.168.10.1 (192.168.10.1), Dst: 192.168.10.2 (192.168.10.2)
Transmission Control Protocol,
Src Port: bgp (179), Dst Port: 29334 (29334), Seq: 65, Ack: 65, Len: 56
Border Gateway Protocol
    UPDATE Message
        Marker: 16 bytes
        Length: 56 bytes
        Type: UPDATE Message (2)
        Unfeasible routes length: 0 bytes
        Total path attribute length: 25 bytes
        Path attributes
            ORIGIN: INCOMPLETE (4 bytes)
            AS_PATH: 1 (7 bytes)
            NEXT_HOP: 192.168.10.1 (7 bytes)
            MULTI_EXIT_DISC: 0 (7 bytes)
         Network layer reachability information: 8 bytes
            192.168.56.0/24
            192.168.10.0/24
```

Figure 1.7: An UPDATE message

1.2.7 The UPDATE message. NLRI

The Network Layer Reachability Information (NLRI) indicates the networks being advertised in the form of a 2-tuple format <length, prefix> as shown below [Figure 1.7].

```
Network layer reachability information: 8 bytes
            192.168.56.0/24
            192.168.10.0/24
```

1.2.8 The UPDATE message. Path Attributes

Path Attributes describe the information related to given route.

In the following example the UPDATE message advertises the routes 192.168.50.4/30 and 192.168.1.0/24 (in the NLRI), via the ASes 3, 1, 2, 5 (AS_PATH).

```
UPDATE Message
        Marker: 16 bytes
        Length: 56 bytes
        Type: UPDATE Message (2)
        Unfeasible routes length: 0 bytes
        Total path attribute length: 24 bytes
        Path attributes
            ORIGIN: INCOMPLETE (4 bytes)
            AS_PATH: 3 1 2 5 (13 bytes)
            NEXT_HOP: 192.168.30.2 (7 bytes)
         Network layer reachability information: 8 bytes
            192.168.50.4/30
            192.168.1.0/24
```

In the routing decision process, the Path Attribute enables the degree of preference of a route, routing loops prevention, filtering, and enforcement of local and global routing policies.

From RFC 4271 [126] "A variable-length sequence of path attributes is present in every UPDATE message, except for an UPDATE message that carries only the withdrawn routes."

Each path attribute is a triple of variable length as follows:

$$< \textit{attribute type}, \textit{attribute length}, \textit{attribute value} >$$

The following subsections deeply describe the *Attribute Type* and its components.

Attribute Type

Attribute Type is a 16 bit field that consists of the following octets:

- *Attribute Flags* (1 octect);
- *Attribute Type Code* (1 octect).

Attribute Flags

Path attributes fall under two main categories: well-known attributes or optional attributes.

From RFC 1771 [123] "Well-known attributes must be recognized by all BGP implementations.

Some of these attributes are mandatory and must be included in every UPDATE message. Others are discretionary and may or may not be sent in a particular UPDATE message. All well-known attributes must be passed along (after proper updating, if necessary) to other BGP peers.".

- BIT 0.Optional bit.

  It defines whether the attribute is optional (if set to 1) or well-known (if set to 0). From RFC 1771: "Well-known attributes must be recognized by all BGP implementations. Some of these attributes are mandatory and must be included in every UPDATE message. Others are discretionary and may or may not be sent in a particular UPDATE message".

  If a well-known attribute is missing, a NOTIFICATION error is generated, and the session is closed. An example of a well-known mandatory attribute is the AS_PATH attribute. An example of a well-known discretionary attribute is LOCAL_PREF.

  "In addition to well-known attributes, each path may contain one or more optional attributes. It is not required or expected that all BGP implementations support all optional attributes. The handling of an unrecognized optional attribute is determined by the setting of the Transitive bit in the attribute flags octet.".

- BIT 1. Transitive bit.

  It defines whether an optional attribute is transitive (if set to 1) or non-transitive (if set to 0). From RFC 1771:

  - *Optional transitive*. If an optional attribute is not recognized by the BGP implementation and the flag is set, which indicates that the attribute is transitive, the BGP implementation should accept the attribute and pass it along to other BGP speakers.

    Paths with unrecognized transitive optional attributes should be accepted.

  - *Optional nontransitive*. When an optional attribute is not recognized and the transitive flag is not set, which means that the attribute is nontransitive, the attribute should be quietly ignored and not passed along to other BGP peers.

    Unrecognized non-transitive optional attributes must be quietly ignored and not passed along to other BGP peers.

- BIT 2. Partial bit.

  It defines whether the information contained in the optional transitive attribute is partial (1) or complete (0).

  From RFC 1771: "If a path with unrecognized transitive optional attribute is accepted and passed along to other BGP peers, then the unrecognized transitive optional attribute of that path must be passed

along with the path to other BGP peers with the Partial bit in the Attribute Flags octet set to 1.

If a path with recognized transitive optional attribute is accepted and passed along to other BGP peers and the Partial bit in the Attribute Flags octet is set to 1 by some previous AS, it is not set back to 0 by the current AS".

- BIT 3. Extended Length bit.

  It defines whether the Attribute Length is one octet (if set to 0) or two octets (if set to 1).

- BIT 4 - 7. UNUSED.

  They must be zero when sent and ignored when received.

Attribute Type Code

The Attribute Type Code octet contains the Attribute Type Code.

Table 1.1 shows some common attribute type code. An explanation of all the attribute types is beyond the scope of this work.

During the UPDATE process, when a BPG speaker have several routes to the same destination, only one of these routes for inclusion in its BGP routing table is selected.

Type Code	Attribute Name	Category
1	ORIGIN	Well-known mandatory
2	AS_PATH	Well-known mandatory
3	NEXT_HOP	Well-known mandatory
4	MULTI_EXIT_DISC	Optional nontransitive
5	LOCAL_PREF	Well-known discretionary

Table 1.1: Attribute Type Codes

The decision process associated to this mechanism is related to these attribute types, according to a tie-breaking algorithm[15] following the order explained below:

1. LOCAL_PREF
2. AS_PATH
3. ORIGIN
4. MULTI_EXIT_DISC (MED)
5. NEXT_HOP

These attributes are explained in the following subsections.

[15] see the following chapter for details.

LOCAL_PREF (TYPE CODE 5)

Well-known discretionary attribute

Usually used to set the exit point of an AS to reach a certain destination, the *local preference (LOCAL_PREF)* attribute is a degree of preference given to a route to compare it with other routes for the same destination.

The LOCAL_PREF attribute affects the BGP decision process, influencing BGP path selection to determine the best path for outbound traffic, but it is local to the Autonomous System and is exchanged between IBGP peers only.

If multiple paths for the same prefix are available, the path with the larger local preference value is preferred.

In the following example, the prefix 192.68.1.0/24 is preferred via IBGP (local preference is 300), even though the AS_PATH via EBGP is shorter[16].

```
ROUTERX1#show ip bgp
BGP table version is 85, local router ID is 193.43.2.254
Status codes: s suppressed, d damped, h history, * valid, > best,
i - internal Origin codes: i - IGP, e - EGP, ? - incomplete
Network          Next Hop     Metric  LocPrf  Weight    Path
*>i192.68.1.0    193.43.1.2           300     0         2 1 i
*                193.43.20.1          0       0         1 i
```

[16] Note that the letter *i* before the prefix 192.168.1.0 indicates an intra-domain prefix. The default local preference value is set to 100.

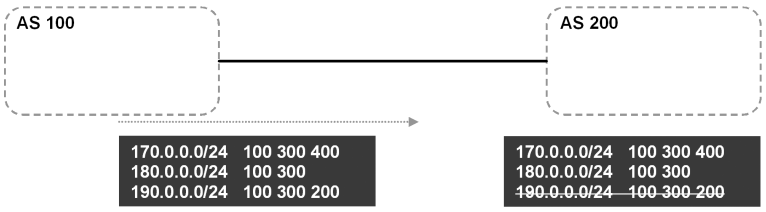

Figure 1.8: AS_Path loop detection

AS_PATH (TYPE CODE 2)

Well-known mandatory attribute

The AS_PATH attribute contains a sequence of autonomous system numbers that represent the path a route has traversed.

When an AS originates a route to external BPG neighbors, it adds it own AS number. Each AS, which receives that routes and passes it on to others neighbors, adds its own AS number to the list.

The mechanism of adding an AS number to the beginning of the list is called *prepending*. The final list represents all the AS numbers that a route has traversed.

This ensures a loop-free topology on the Internet [Figure 1.8]. A route is not accepted by an AS if the prefix has the AS in its AS_PATH attribute.

In the next example, BGP prefers the second one as being the "best" path due to its lower AS_PATH length.

```
ROUTERX1#show ip bgp
BGP table version is 87, local router ID is 193.43.2.223
Status codes: s suppressed, d damped, h history, * valid, > best,
i - internal Origin codes: i - IGP, e - EGP, ? - incomplete
Network          Next Hop        Metric  LocPrf  Weight  Path
* 192.68.1.0     192.68.4.3                      0       2 1 i
*>i              193.43.122.1    0       100     0       1 i
```

The mechanism of the AS_Path manipulation is commonly used to change a router's process decision. This technique will be described in the following chapter.

ORIGIN (TYPE CODE 1)

Well-known mandatory attribute

The ORIGIN attribute is used to establish a preference ranking among multiple routes in the decision-making process.

The data octet can assume the following values [126]:

- 0: IGP. Network Layer Reachability Information is interior to the originating AS;
- 1: EGP. Network Layer Reachability Information learned via the EGP protocol [RFC904] [108];
- 2: INCOMPLETE. Network Layer Reachability Information learned by some other means.

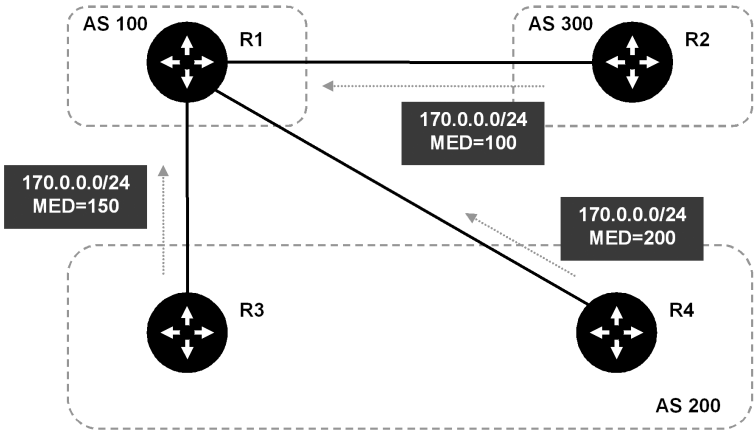

Figure 1.9: MULTI_EXIT_DISC

MULTI_EXIT_DISC (TYPE CODE 4)

Optional nontransitive attribute

In the BPG speaker's decision process, the MULTI_EXIT_DISC (MED) attributes may be used to make a comparison between multiple paths from external neighbors of the same AS[17] [Figure 1.9].

This is useful when a customer has multiple connections to the same provider, and it can be used for traffic balancing by both providers and customers.

[17]MED attributes from different ASes are not comparable. The MED usually gives information of the AS's internal topology, routing policies, and routing protocol

Some features of this attribute are described as follows:

- unlike LOCAL_PREF, the MED attribute is exchanged between ASes;
- a MED attribute is received by an AS;
- a lower MULTI_EXIT_DISC value is preferred over a higher MED value;
- a MED received by an AS does not leave the AS. When BGP passes the routing update to another AS, the MED is reset to 0[18];
- MEDs are not always accepted by peers;
- when the route is originated by the AS itself, MED value is generally set to the internal IGP metric of the route.

In Figure 1.9, router R1 is receiving routing updates about 170.0.0.0/24 from R2 (MED=100), R3 (MED=150) and R4 (MED=200). From AS 200, R1 will prefer the R3 route to reach 170.0.0.0/24 because router R3 is advertising a lower MED value (MED=150).

The comparison between MED values between different ASes in not generally possible[19].

[18] Unless the outgoing MED is explicitly set to a specific value.

[19] **`bgp always-compare-med`** command is used to compare MEDs coming from different ASes. In this case, route to reach 170.0.0.0/24 coming router R2 will be preferred.

NEXT_HOP (TYPE CODE 3)

well-known mandatory attribute

It defines the IP address of the border router that should be used as the next hop to the destinations listed in the NLRI.

In the next chapter, the routing decision process between BGP and its peers will be explained, according to routing policies and filtering mechanisms over the updates.

Chapter 2

Routing Decision Process

This chapter covers the aspects of the routing decision process inside a BGP speaker in order to cooperate with other BPG speakers to provide global Internet connectivity (*cooperation*), according to business and commercial agreements (*competition*), using a routing process model to set their policy independently to each others (*autonomy*), and making all the manipulations allowed by the protocol (*expressiveness*), without any global coordination.

As seen in the previous chapter, routes[1] are advertised between BGP neighbors using UPDATE messages.

Each BGP speaker applies policies and filters over the updates, and may add or modify a route's path attribute before advertising it to a peer through a mechanism known as *route filtering and attribute manipulation*, and described in section 2.2.1.

In addition, in case of multiple routes to the same destination, a *ranking mechanism* is used to choose the best route to advertise to other neighbors.

In order to accomplish this process, a separate BGP Routing Table from IP Routing Table is required. The IP Routing Table is the final routing decision, and contains the routes learned from BGP peers and valid local routes originated inside the AS.

In the next section, the traditional CISCO's Routing Process Model is described.

In the final section of this chapter, a new and more complete scheme of the Routing Process Model is proposed according to RFC 4271 with the examination of the *input policy engine*, the *decision process*, and the *output policy engine* inside a BGP speaker. In addition, the formalization of the decision process related to the route selection mechanism is proposed.

[1] As specified in RFC 4271 [126], "A *route* is defined as a unit of information that pairs a destination with the attributes of a path to that destination".

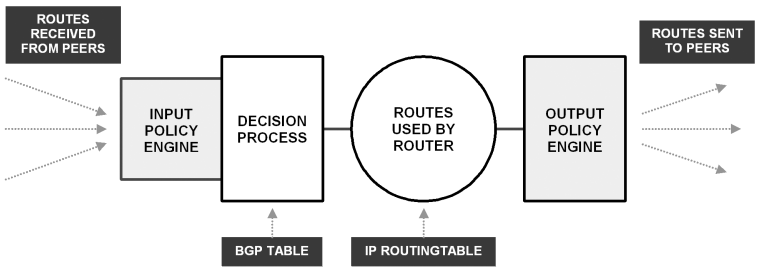

Figure 2.1: CISCO's Routing Process Overview

2.1 CISCO's Routing Process Model

CISCO's routing process model [73] involves the following components [Figure 2.1]:

- a pool of routes that the router receives from its peers;
- an *input policy engine* that can filter the routes or manipulate their attributes;
- a *decision process* that decides which routes the router itself will use;
- a pool of routes that the router itself uses;
- an *output policy engine* that can filter the routes or manipulate their attributes;
- a pool of routes that the router advertises to other peers.

2.1.1 The BGP Routing Table

The BGP routing table consists of three parts:

- *Adj-Routing Information Base IN (Adj-RIBs-In);*
- *Local Routing Information Base (Loc-RIB);*
- *Adj-Routing Information Base OUT (Adj-RIBs-Out).*

The following subsections describe these three parts and its roles.

Adj-Routing Information Base IN (Adj-RIBs-In)

The Adj-RIBs-In contains unprocessed routing information that has been advertised to the local BGP speaker by its peers [126].

Routes that are received from other BGP speakers are present in the Adj-RIBs-In.

A *Route Filtering* (based on different parameters, such as IP prefixes or AS_PATH) and *Attribute Manipulation* (in order to influence route decision process) might be applied by the operator via an Input Policy Engine[2].

A filter in an incoming prefix indicates that BGP does not want to reach that destination via that peer, or a better LOCAL_PREF value indicates that BGP prefers the prefix from a specific peer.

[2]Route Filtering and Path Manipulation are discussed in the next sections.

Local Routing Information Base (Loc-RIB)

The Loc-RIB contains the routes that have been selected by the local BGP speaker's Decision Process [126].

These route become candidates for the placement in the IP Routing Table and the advertisement to other neighbors.

The Loc-RIB contains only the preferred routes that have been selected as the best path to each available destination.

Adj-Routing Information Base OUT (Adj-RIBs-Out)

The Adj-RIBs-Out contains the routes for advertisement to specific peers by means of the local speaker's UPDATE messages [126].

It stores routing information that the BGP speaker has selected for advertisement to the neighbors. Likewise the Input Policy Engine, an Output Police Engine may be used to apply a Route Filtering and Attribute Manipulation before sending the UPDATE message.

The set of *routes advertised to peers* consists of those routes that successfully pass through the Output Policy Engine and are advertised to the BGP neighbors.

2.2 A new scheme for Routing Process Model

In this section, a new and more complete Routing Process Model is proposed according to RFC 4271 (Figure 2.2).

This routing process model involves the following components:

- a pool of routes that the router receives from its peers and stored in the Adj-RIBs-In (I.UPDATE 1,...N in Figure 2.2);

- an *Input Policy Engine* that filters the routes or manipulate their attributes taken from the local *Policy Information Base (PIB)*, through different mechanisms and policies such as *Route Identification and Filtering*, *Route Authorization* and *Attributes Manipulation* described in following paragraphs;

- a *decision process* able to select which routes the router itself will use, via an individual application of a degree of preference to each route (see the next sections for details), and the choice of the route with the highest degree of preference (the PREFERRED ROUTES in Figure 2.2);

- a pool of *preferred routes* selected as the best path to each available destination contained in the Loc-RIB. These routes are candidates for the placement in the IP Routing Table, to be used locally by the router;

- a selection of routes contained in the Loc-RIB for advertisement to other BPG peers, stored in the Adj-RIBs-Out;
- an *Output Policy Engine* that can filter the routes or manipulate their attributes as seen in the Input Policy Engine, and according to the policies in the PIB);
- a pool of routes that the router advertises to other peers via UPDATE messages (O.UPDATE1,....M in Figure 2.2).

2.2.1 Route Filtering and Attribute Manipulation

Route Filtering and Attribute Manipulation involves three actions [Figure 2.3]:

- *Route Identification and Filtering;*
- *Route Authorization;*
- *Attributes Manipulation.*

In the example of Figure 2.4, the following commands:

```
neighbor 193.32.2.2 prefix-list 1 out
ip prefix-list 1 seq 5 deny 170.0.0.0/24
```

prevents R3 from propagating prefix 170.0.0.0/24 to AS 100.

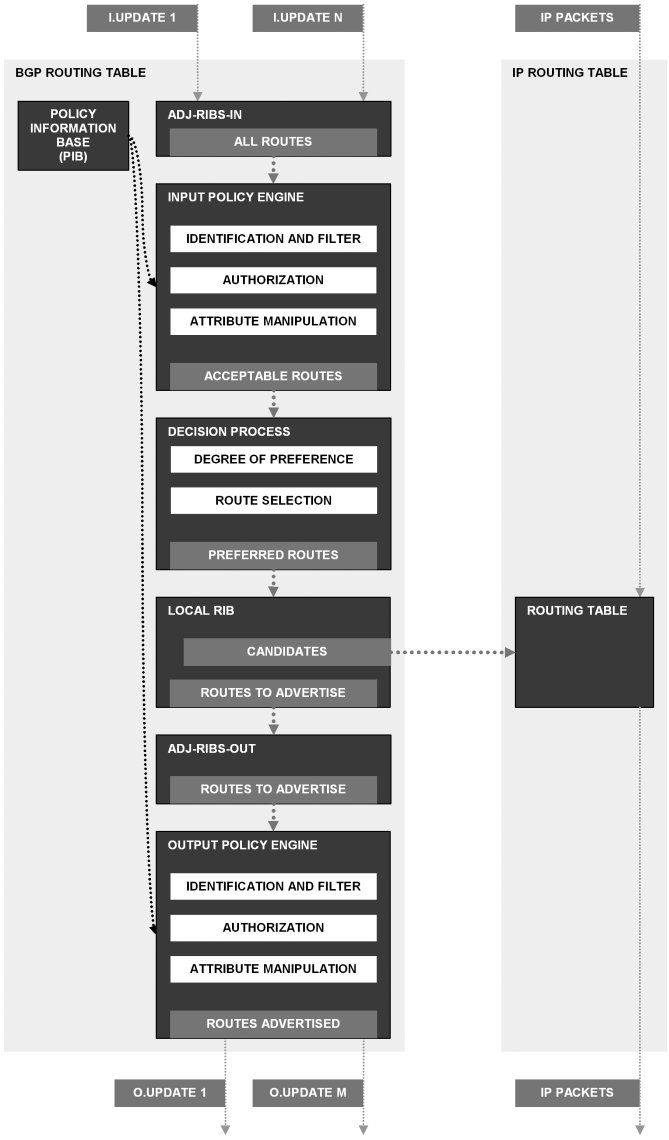

Figure 2.2: A new scheme for Routing Process Model

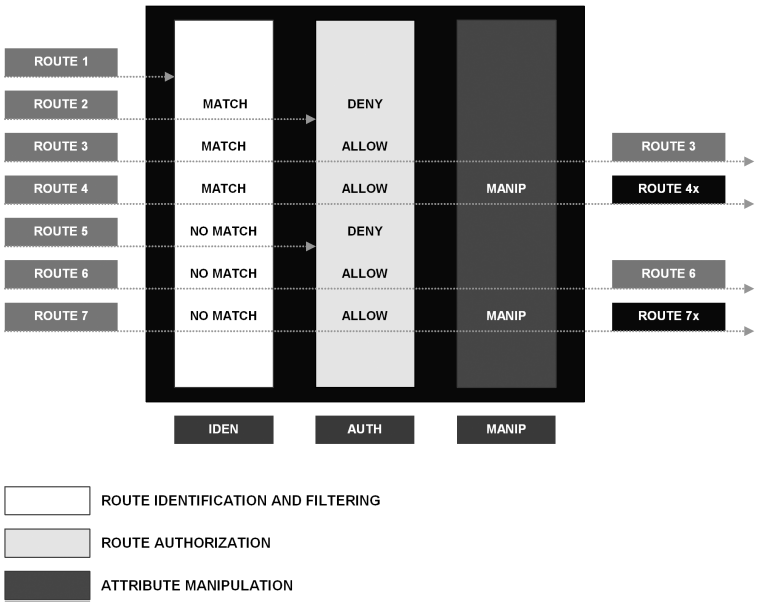

Figure 2.3: Route Filtering and Manipulation Process

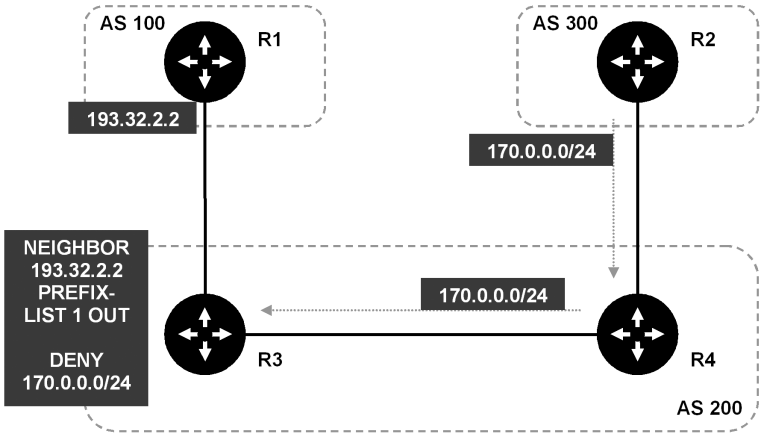

Figure 2.4: Example of filtering routes based on the NLRI

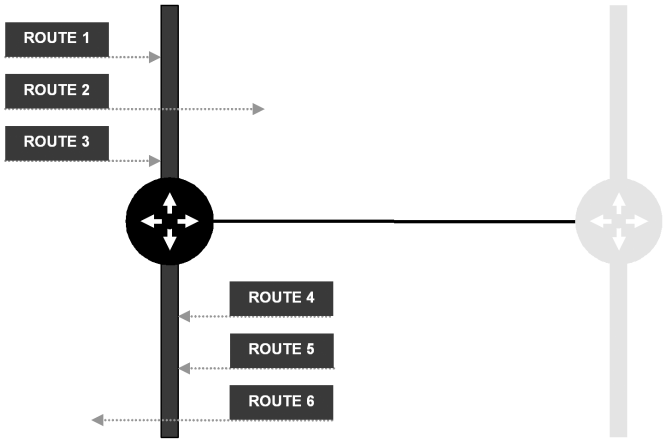

Figure 2.5: Inbound and Outbound Route Identification and Differentiation

Route Identification and Filtering

The mechanism of route identification and differentiation of routes exchanged by BGP peers may be used in the updates received from other peers (ROUTE4, ROUTE5 and ROUTE6 in Figure 2.5) or in updates advertised to other peers (ROUTE1, ROUTE2 and ROUTE3 in Figure 2.5).

If a route remains unidentified, the route is discarded (ROUTE1 in Figure 2.3).

The mechanism of route identification and differentiation is based on different criteria. The most common way of identifying routes is based on NLRI and the AS_PATH.

Route Authorization

A mechanism of permitting or denying the identified and differentiated routes.

If a route is denied, that route is discarded (ROUTE2 and ROUTE5 in Figure 2.3).

If a route is permitted, it can be accepted "as is" (ROUTE3 and ROUTE6 in Figure 2.3), or is submitted for attributes manipulation (ROUTE4 and ROUTE7 in Figure 2.3).

Attributes Manipulation

A permitted route submitted for attributes manipulation may have its attributes changed to affect the decision process for the identification of the best routes to a destination (ROUTE4 and ROUTE7 in Figure 2.3).

A common example of attribute manipulation is called *AS_PATH Manipulation*. After LOCAL_PREF attribute, AS_PATH attribute is the preferred attribute type in order in order of attribute preference in the routing decision process. Carrier operators may use the manipulation of AS_PATH attribute in order to influence interdomain traffic trajectory by including dummy AS_PATH entries.

This AS_PATH manipulation is made by prepending AS numbers at the beginning of an AS_PATH in order to have a longer path length.

In the next example ROUTERX1 received an UPDATE that changed

router's decision about reaching 192.68.1.0/24.

By prepending two extra AS numbers, the preferred path is via 193.43.1.2 instead the internal path via 193.43.20.1.

```
ROUTERX1#show ip bgp
BGP table version is 44, local router ID is 193.43.2.254
Status codes: s suppressed, d damped, h history, * valid, > best,
i - internal Origin codes: i - IGP, e - EGP, ? - incomplete
Network          Next Hop     Metric  LocPrf  Weight    Path
*>192.68.1.0   193.43.1.2         0      0              2 1 i
* i            193.43.20.1        0      0              1 1 1 i
```

2.2.2 Formalization of the Decision Process

The Decision Process is a crucial phase, and is responsible for the information stored in the Loc-RIB.

In this section, the formalization of the decision process related to the route selection mechanism is proposed, which involves the following items:

- *Candidates*. The routes that are candidates for the placement in the IP Routing Table, to be used locally by the BPG peer;
- *Routes to Advertise*. A selection of routes or advertisement to other BPG peers.

We introduce the Decision Process model which is able to define a pool of loop-free and feasible routes through two different phases[3]:

- *Degree of Preference*;
- *Route Selection*.

Degree of Preference

RFC 4271 defines the approach for the computation of a *degree of preference* for each route received via an UPDATE message as described as follows:

- if the route is learned from an internal peer, the value of DoP will be the LOCAL_PREF attribute, or the local system computes the degree of preference of the route based on preconfigured policy information;

- if the route is learned from an external peer, the local BGP speaker computes the degree of preference based on preconfigured policy information.

According to the set of tie-breaking criteria proposed in RFC 4271 and shown in Figure 2.6, for a given route r, we define the *attribute function* $\alpha_{\tau,r}$ [eq. 2.1] which describes the importance of the attribute τ taken from the

[3]RFC 4271 uses three phases to explain this mechanism.

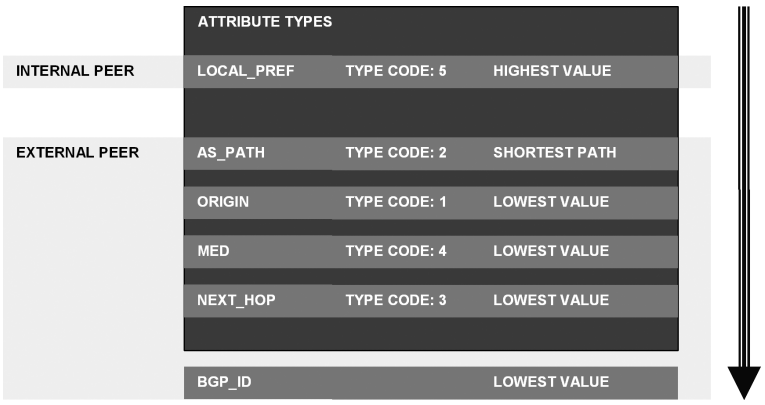

Figure 2.6: Tie-Breaking Criterion

Type Code Attribute of the UPDATE message for the route r as follows:

$$\alpha_{\tau,r} = \begin{cases} 101 & \text{if } \tau = 3 \text{ (NEXT\_HOP)} \\ 102 & \text{if } \tau = 4 \text{ (MED)} \\ 103 & \text{if } \tau = 1 \text{ (ORIGIN)} \\ 104 & \text{if } \tau = 2 \text{ (AS\_PATH)} \\ 105 & \text{if } \tau = 5 \text{ (LOCAL\_PREF)} \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

The tie-breaking algorithm[4]:

1. considers only the routes with the smallest number of AS numbers in their AS_PATH attributes;

2. considers only the routes with the lowest ORIGIN attribute values;

3. considers only the routes with lowest MED attribute values from the same neighboring AS;

4. considers only the routes with lowest *interior cost*, determined by calculating the metric to the NEXT_HOP for the route using the routing table;

5. considers only the route with lowest BGP Identifier value.

At this point, we introduce a formal way to describe the degree of preference by defining a function *Degree of Preference* $DoP_{\alpha,r} \in \mathbb{N}$ [eq. 2.2] which takes the value of the attribute function $\alpha_{\tau,r}$ for a given route r as

[4]A tie-breaking algorithm which differs from the algorithm proposed in this work is proposed in RFC4271. In RFC 4271, firstly all equally preferable routes to the same destination are considered, and then selected routes are removed from consideration. The algorithm terminates as soon as only one route remains in consideration. See [126] for details.

argument and returns:

$$DoP_{\alpha,r} = \begin{cases} > 0 & \text{the degree of preference for the route } r \\ = 0 & \text{the route is ineligible to be installed} \end{cases} \tag{2.2}$$

As example, a $DoP_{105,192.167.20.0/24} = 220$ is a Degree of Preference for the prefix 192.167.20.0/24 using a LOCAL_PREF attribute.

If another $DoP_{105,192.167.20.0/24} = 250$ exists in another UPDATE message, the highest value of DoP is considered (due to the fact that the tie-breaking algorithm considers only the routes with the highest LOCAL_PREF attribute values [Figure 2.6]).

Route Selection

Broadly speaking, when a BGP speaker has several routes to the same destination, it can select only one of these routes for inclusion in the Loc-RIB.

This process is called *Route Selection* and can be summarized in :

1. the application of the *DoP* to each feasible route *r*;
2. the choice of the highest *DoP* value for this destination.

Let $r_1, r_2, ..., r_n \ \forall n \in \mathbb{N}$ represent the n routes for the same destination d. We define α^*_d [eq. 2.3] the highest value of the attribute function for the

set of n routes for the same destination d, and r^* a route for a destination d with α_d^* as attribute function value:

$$\alpha_d^* = \left[max \sum_{k=1}^{n} \alpha_k \right]_d \ni \exists \left[DoP_{\alpha^*, r^*} \right]_d > 0 \tag{2.3}$$

The condition $\left[DoP_{\alpha^*, r^*} \right]_d > 0$ ensures that at least one route with α_d^* is eligible to be installed in the Loc-RIB.

At this point, let m the number of m routes for each destination d with α_d^* as attribute function value.

We can identify a *Route Selection function* $RouteSel_d$ [eq. 2.4] for a given destination d as follows:

$$RouteSel_d = \left[max \sum_{k=1}^{m} DoP_{\alpha_d^*, r_k} \right]_d \tag{2.4}$$

At the end of the process, all the chosen routes for the a set of destinations D are installed into Loc-RIB as described in eq. 2.5.

$$DecisionProcess = \sum_{d=1}^{D} RouteSel_d = \sum_{d=1}^{D} \left[max \sum_{k=1}^{m} DoP_{\alpha_d^*, r_k} \right]_d \tag{2.5}$$

2.3 Conclusion

After a brief introduction on the routing process model and the well-known manipulating techniques, a new scheme for the Routing Process Model is proposed, and a formalization of a new and more complex routing process model inside a BPG speaker is made.

Besides, a formalization of the problem through a "route selection function" defined as the maximum value of the "degree of preference function" built from all the available routes for a given destination is proposed.

In the next chapter, an examination of the issues related to routing decisions in the inter-domain routing is made, with a review of the most interesting proposals in this area.

Chapter 3

Routing Decision in the Inter-AS routing

One of the most complex problems in computer networks is controlling the routing decisions of the ASes in the interdomain routing.

BGP was designed as a protocol to apply diverse local policies for selecting routes and distribute reachability information to other ASes.

The fully independent management provided by each AS domain makes the problem of controlling inter-domain routing, due to potentially conflicting policies derived from the business agreements and competition between domains, that can lead to routing instability, demonstrating to be inaccurate and poorly effective in controlling and communicating the inter-domain decisions.

In addition, recent studies made by Uhlig et al. reveal that the AS-paths show large variations over time, because they are present in the BGP routing tables for a few minutes [143].

This chapter examines some of the many issues related to routing decisions in ASes, reviewing the most interesting proposals in this area and describing why these issues are difficult to solve.

3.1 BGP divergence

The lack of a global global coordination between ASes, with a distributed conflicting policy-based routing system, creates routing anomalies such as divergence in the update process to the exchange of routing information within an AS, and producing endless streams of routing updates unrelated to changes in topology or policy [70] [71] [144] [145].

BGP allows each autonomous system to independently formulate its routing policies to override distance metrics and enabling each AS to independently define its routing policies with no global coordination.

A divergence anomaly occurs when BGP routers permanently fail to obtain a stable path to reach a destination.

A natural approach to the route convergence problem requires a *global coordination*, using a repository of routing policies. Internet Routing Registry (IRR) [78] was born in order to solve the lack of global coordination, but at the moment this solution is not practicable because a global coor-

dination does not ensure a convergence in presence of link failures or a policy change.

Besides, the registry is update via voluntary basis, and the Administrative Domains may be unwilling to reveal their policies; so the information are incomplete and out of date.

Several research have studied route convergence under the presence of a global knowledge of the routing policies and topology, by using simple ring topologies [144], or by focusing on LOCAL_PREF and AS_PATH attributes [71], but they proves only negative results [144] [71] [64] [7].

The solution proposed are divided into methodologies that extend the capabilities of BGP [68] [26] [27] [16], restrict the types of routing policies adopted by the ASes [56] [85] [86] , use a global coordination to avoid routing conflicting policies [64], or detect conflicts at runtime [69].

In particular, Varadhan et al. [144] [145] first define the concept of *safety*, observing that routing policies can cause BGP to diverge, and affirming that only the policies based on a mechanism of shortest path routing or next-hop are safety.

Griffin et al. [68] were the first to present the causes of this problem, developing and approach to detect and resolve divergence by suppressing routes that contain cycles, and generating a stable routing in unstable BGP configurations (it is called *stable paths problem*). However they do not consider the effects of the filtering process.

Other works [16] extend the original Simple Path-Vector Protocol used

in [68] [26] [27].

Gao et al. [56] believe that convergence without requiring a coordination without other ASes is possible by restricting the set of policies that each AS can apply. They formalize the notion of a *stable state* where no AS would change its routes and a *safe BGP system* that is guaranteed to converge to a stable state.

In addition, they enunciate a guideline for choosing routing policies based on a set of constraints called *Gao-Rexford Constraints*[1], over a set of commercial relationships between ASes divided into customer-provider, and peer-to-peer and backup relationships[2] [6] [77].

Govindan et al. [64] [63] propose a routing architecture where ASes coordinate their policies using a standardized object oriented language for specifying routing policies called Routing Policy Specification Language (RPSL) [7].

Tradeswoman et al [85] [86] in their papers propose an architecture for policy based networks that involves semantically tagging packets, using a semantically highly extensible language called OWL [106] instead of RPSL [7].

Jaggard et al. [81] propose a set global conditions to guarantee safety

[1]Gao-Rexford Constrains are based on a set of assumptions and theorems. As example they assume that peer-to-peer relationships satisfy the condition that there is no cycle in the graph that represents the topology of a BGP system.

[2]These commercial relationships will be discussed thereafter.

of routing systems based only of next-hop preferences, and a propose algorithms to check these global conditions.

However, all these methodologies present several drawbacks [107], due to the extension to the BPG protocol to carry additional information, to an additional computation in the analysis of routing updates on every router involved to be detected and identified at runtime, to protocol overhead, to an incomplete analysis of the relationship between all the actors involved in the process, to the need of an analysis of the all routing policies to verify that they do not contain policy conflicts.

3.2 BGP convergence time

In BGP, when a path failure or routing policy changes occur, BGP peers explore alternative paths to find new paths before selecting a new path or declaring the unreachability to a destination, using a large amount of BPG advertisements, in order to achieve a new steady state (convergence). This is called *path exploration*. Path exploration should happen as fast as possible.

Several studies have shown that a long time period may elapse before the whole network eventually converges to a final decision. The convergence time is very slow (tens of seconds) [67] [89] [90] [102], leading to severe performance problems in data delivery.

The techniques used to measure the path exploration and the convergence time are based on passive measurements [91] [92] [147] [127] [43], with the study of the instability on the prefixes, or on active measurements [89] [102] [90], using a small number of beacon sites to control the events.

Different methodologies are proposed to reduce BPG convergence time using different techniques, such as algorithms to force the quick distribution of bad news (in order to have a control of the number of messages exchanged during the convergence) [18], or modifying the BGP protocol (in order to of carrying additional information inside the BPG message) [20] [119].

None of the mechanisms is able to accomplish the objectives.

3.3 AS mechanisms

Different mechanisms are developed by different types of ASes in the inter-AS routing.

Transit multihomed AS uses a well-known methodology of reaching a destination preferring eBGP over iBGP in the decision process called *hot potato routing*, that causes routing instabilities across the boundaries of the ASes caused by the lack of coordination between the policies of the domains [3].

Akella et al. [5] have shown that stub ASes use mechanisms in order to operate in short timescales via multiple connections. In this way they

ensure an improvement of the performance containing costs.

However, this kind of techniques could create important problems in the reliability of the inter-AS routing system if used in a massive way.

3.4 Prefix Hijacking

Prefix hijacking is a technique used by an AS to originate a prefix it does not own. This false route may appear more attractive to some ASes than the actual route to that prefix. Thus, these deceived ASes might choose this false route as the best route and send packets to the false origin.

This AS path forgery is treated as a dangerous trick used by attackers to threat network security, but no positive light for the solution is brought by the research [12] [155] [94] [96].

The existing efforts in the area may be divided into mechanisms of hijack prevention (based on cryptographic authentications) [84] [137] [134], and mechanism schemes of hijack detection [93] [121] [82].

All these solutions require heavy changes to all router implementations, or require a public key infrastructure.

In addition, the hijack detection mechanisms provide only the hijack detection but not the hijack correction[3].

[3] An inter-domain routing architecture of hijack detection and correction called MIRO is proposed in [151].

3.5 Out-Of-Band Solutions

Non-BGP-based techniques have arisen because the Administrative Domains remain cautious about modifying BGP. This kind of techniques are often called Out-of-Band Solutions.

Different types of approaches are proposed with the use of DNS-based optimizers rely on Network Address Translation [110] [72] [5] (but traffic control is unfeasible for medium and large ASes because this solutions are not scalable) or with the use of *Internet Route Controllers (IRCs)*, indipendent intelligent devices able to control the routing decision process inside the multihomed stub domains.

IRC solutions improve the end-to-end performance of inter-domain routing [4], but they are not applicable to large transit ASes.

In addition, they are standalone solutions, so a cooperation between devices is not possible, as the study of a global effect of the decision process in the whole network.

Besides, Gao et al. [56] show that persistent oscillations can occur in a competitive environment causing significant performance degradations.

3.6 QoS capabilites

The request of services such as VoIP requires mechanisms to offer services similar to the differentiated services inside the intra-domain routing, with

the creation of different levels of QoS for network services [111].

This set of mechanisms is called QoSR (QoS Routing), but BPG has no inbuilt QoSR capabilities [30].

This problem is discussed in many papers and several solutions are proposed [31] [150] using in-band or out-of-band solutions, but all of them present strong limitations and are not appealing to become deployed in practice.

3.7 Completeness

The reconstruction of the AS topology is an active area of research, but building a complete set of links between Autonomous Systems, in order to obtaining an accurate AS-level connectivity has proven difficult [95] [24] [148] [23], due to the inter-domain decentralized architecture.

BGP uses only one route as the best path, if a router receives multiple advertisements for the same destination. BGP uses UPDATE messages to propagate only the best paths between peers, according to the routing process and the policy mechanism[4].

As a result, a set of peer links are invisible to the observation and the AS topology remains incomplete.

This is called *completeness* problem.

[4]Some recent works [146] propose methods to advertise multiple routes for the same destination to peers.

Obtaining a complete and accurate topology of the relationships and agreements between ASes remains one of the most active areas of research [122] [21] [154] [40] [100] [6] [101].

Govindan et al. [65] recovered the traces of BGP updates and inferred topological results and route stability results.

Faloutsos et al. [40] defined three power-laws inferred from the study of a dataset of routing table information taken from three AS network topology instances. This kind of approach is used also in [101].

Therefore, the quality of the currently used AS maps has remained by and large unknown.

3.8 AS Relationship Inference

The generation of an accurate synthetic AS topology of a BGP system to model the interconnection structure of the Administrative Domains is another interesting research area.

In order to understand the topological structure of the BGP system, a global hierarchical structure is inducted by the commercial relationships.

In the next chapter, an exhaustive review of the most interesting papers in this research area is made, and methodologies for the generation of an accurate synthetic AS topology are explained.

Chapter 4

AS Relationships

The research area referred to the study of AS relationships and to the development of accurate topology generators is essential for producing realistic simulation studies of protocols and network architectures, to reduce misconfiguration or to debug router configuration files, for the identification of potential erroneous routes or to plan for future contractual agreements [42] [88].

AS relationships have a profound influence on how traffic flows through the Internet. Internet topology does not provide enough information, but despite the volume of research in this area, current topology synthetic generators fail to capture an inherent aspect of the AS topology.

A link between ASes is established when a contractual agreement to exchange traffic is made between them.

In general, the ASes tend to treat their agreements as proprietary information, so collecting the complete set of inter-domain links has proven difficult. In the absence of a global registry, the AS-level structure of the Internet is typically inferred from analysis of routing data.

In addition it is well known that AS paths in the Internet are longer than the shortest path [57] [139] [138][135].

In order to conducting accurate and realistic simulation studies, an evaluation of modeling AS relationships occurs [34] [117].

In this chapter a methodology for the generation of an accurate synthetic AS topology is made through the review of the most interesting papers in this area, and starting with the study of the available Data Set.

4.1 Data Set

A *Data Set* contains information related to several aspects of the interconnection structure of the Internet topology. For this reason, different types of Data Set are often used to infer relationships between ASes. Data Set can be divided into:

- Internet Registries;
- IXP Data;
- BGP Table Dumps;
- Traceroute Data.

4.1.1 Internet Registries

As seen in the previous chapter, the natural approach to the route convergence problem requires a *global coordination*, using a repository of routing policies.

Internet Registries are distributed databases containing information related to the AS administration or AS number allocations, such as ARIN [10] or RIR (Regional Internet Registries) [129].

In addition, Internet Routing Registry (IRR) [78] was created as a repository of inter-AS connections and routing policies - and to perform consistency checking on the registered information - that use the standard language Routing Policy Specification Language (RPSL) [105] [7].

But several impediments to the global coordination were born:

- a global coordination does not ensure a convergence in presence of link failures or a policy change;
- Internet Routing Registry is update via voluntary basis;
- contractual agreements between Administrative Domains are in general proprietary;
- Administrative Domains may be unwilling to reveal their policies.

For these reasons the information are incomplete and out of date [21] [28], and, in general do not imply anything about how ASes relate to each other.

4.1.2 IXP Data

Internet Exchange Points (IXPs) or Network Access Point (NAPs) are infrastructures that enable physical connectivity between their member networks through a shared medium, such as a FDDI ring, an ATM switch, a Gigabit Ethernet. Table 5.1 shows a list of Internet Exchange Points [80].

Physical connectivity between member does not imply a connectivity between ASes (reachability).

In fact, only the relationships between ASes via the negotiation of contractual agreements ensure the exchange of traffic between them.

The lists of IXPs are available with the names of participants in same cases [39] [120] [118]. But, since the information are input on a voluntary basis, they are incomplete and outdated.

However, most IXPs publish the subnet prefixes they use keeping reverse DNS entries for the assigned IP addresses of each IXP participant inside the IXP subnet [62].

A method to infer IXP participant was proposed by He et al. [74], but all of the methods do not accurately convert router paths to corresponding AS paths, with the generation of false and inflated paths.

4.1.3 BGP Table Dumps

In the absence of a global registry, the AS-level structure of the Internet is inferred from analysis of routing data, taken from BGP table dumps or

taken from traceroute data.

In the BPG table dump, BGP forwarding tables and routing updates are passively listened by data collectors [53] [65] [58]. University of Oregon RouteViews server [130], RIPE-RIS [128], Abilene [142], Geant [59], has been used for the creation of the set of inferred links between the ASes for the generation of a synthetic AS topology.

Gao [53] presents heuristic algorithms for inferring the relationships from BGP routing tables, based on the fact that a provider is larger than its customers and two peers are of comparable size, considering the relationship between neighboring ASes as an inherent aspect of the inter-domain routing structure.

Subramanian et al.[136] propose a methodology for combining data from multiple vantage points in order to construct a more complete view of the AS relationships and to network topology. It is an approach that differs from the global coordination, based form a partial view of the Internet topology, the analysis of AS paths from multiple locations, and considering the commercial relationships between ASes.

4.1.4 Traceroute Data

Traceroute command provides a view of the path from a source to a destination host. A set of monitors send periodic UDP or ICMP packets to a set of IP addresses, and convert router paths to AS paths [133] [131] [99] [19].

The process of conversion of router paths into AS paths may introduce false AS links as shown in [103] [76] [22].

4.2 Modeling the Commercial Agreements

ASes have the responsibility for carrying traffic to and from a set of prefixes. BGP allows each AS to choose its own administrative policy in selecting routes and propagating reachability information to others.

The negotiation of contractual commercial agreements between Administrative Domains ensures the exchange of traffic between ASes, but the routing policies are constrained by these AS relationships.

Routing policies are often manually configured in BPG routers by Administrative Domain operators.

Several efforts are made to classify these commercial agreements [2] [77] [53] [6].

Awduche et al. [2] define two types of peering relationships called *customer peering* and *non-customer peering*.

In *customer peering*, an Administrative Domain provides transit service to its customers for a fee, and routing their in-bound and out-bound traffic.

In *non-customer peering*, an Administrative Domain provides non-transit service to other ASes on the basis of bilateral agreements.

Houston [77] defines a set of commercial agreements as follows:

Customer-Provider agreements

- a customer pays its provider for connectivity to the rest of the Internet;
- a provider transits traffic for its customers;
- a customer does not transit traffic between its providers;

Peer-Peer agreements

- peers exchange traffic between their customers free of charge.

Mutual-Transit agreements

- the Administrative Domains provide connectivity to the rest fo the Internet for each other. This is a typical agreement between two small Administrative Domains located close to each other that cannot afford additional Internet services for better connectivity.

Backup Relationship

- a backup connectivity to the Internet for Administrative Domains in the event of a connection failure.

Even if this classification does not capture all the possible commercial agreements between ASes, it is used as reference point by several authors [53] [136] [34].

Dimitropoulos et al. [34] affirm that ASes prefer customer routes over routes through peers or providers, because ASes do not have to pay for sending traffic to a customer and tend to avoid congestion at peering exchange points.

4.3 Modeling the Interconnection Structure

A large number of works have focused on the generation of synthetic AS topologies, in order to model the interconnection structure of the Administrative Domains.

These network topologies have modeled the network structure as a *graph* to represent the relationships and the interactions between the ASes.

In order to understand the topological structure of the AS connectivity graph, a global hierarchical structure is inducted by the commercial relationships. The study of the inference of the type of relationships between interconnected ASes based on a collected data set (which, in general, are not part of the AS connectivity data) and their routing policies, is often called the *Type of Relationship (ToR) problem*.

Some early studies consider network as a random structure or a structured network as abstract undirected graphs, missing the different types of node relationships, inducing an unrealistic model of the interconnection structure [149].

Siamwalla et al. [132] and Govindan et al. [66] presents two heuristic

methods to discover routing adjacencies using traceroute command.

Faloutsos et al. [40] and Magoni et al. [101] describe a set of average properties of the AS network from distributions (called *power-laws*), concerning degree, distance, number of shortest paths or trees taken from a very limited set of instances of BGP data[1] [101].

Three power-laws are defined by Faloutsos et al. [40], and five power-laws by Magoni et al. [101], to give a view of the current AS network topology as well as a view of its on-going evolution, and to model the AS network as accurately as possible.

But, all of them are *empirical laws*, inferred from a reduced set of data, which generate an incomplete and unrealistic model of the interconnection structure[2].

All the aforementioned work do not present an explicit notion of AS an hierarchical structure of the topology network.

A new, and more realistic view of interconnection structure as an *hiearchical* and *structural* network topology, with the explicit notion of AS relationships in the topology characterization is presented in several work [65] [53] [35] [153] [34] [17].

Govidan et al. [65] define the *degree* of AS as the number of ASes that are

[1] Only six instances.

[2] As example, two of these empirical laws are: "(ASs growth). Currently, the number of ASs in the AS network increases by 45% each year" or "(Connection growth). Currently, the number of BGP connections in the AS network increases by 53% each year".

its neighbors. The degree of an AS is used as an heuristic in determining the size of the AS and to classify ASes into four levels of hierarchy.

Dimitropoulos et al. [34] define the *customer-degree* d_{p2c} of an AS as the number of its customers, the *provider-degree* d_{c2p} as the number of its providers and the *peer-degree* d_{p2p} as the number of its peers, in order to capture their distribution in the ASes[3].

Gao [53] proposes an hierarchical and structured AS graph representation $G = (V, E)$, where the node set V consists of ASes and the edge set E consists of the set of relationships between the ASes, based on a set of commercial agreements.

This graph is represented as a partially directed graph, called *annotated graph*, where its edges are classified into:

- *provider-to-customer* $u \rightarrow v$

  AS u is a provider of AS v if u transits traffic for v and v does not transit traffic for u. The edge (u, v) is directed from u to v[4].

[3] As example, they affirm that large Tier-1 ASes typically have a large d_{p2c}, zero d_{c2p}, small d_{p2p}.

[4] Note that, according to Gao [53], the direction of the node is from provider to customer *provider* $\rightarrow$ *customer*.

Several authors propose the direction of the node from customer to provider *customer* $\rightarrow$ *provider* because the traffic flows from customer to provider. This may generate confusion because the provider is a higher-level entity than the customer in this structured and hierarchical topology.

- *customer-to-provider* $u \leftarrow v$

  AS u is a customer of AS v if u does not transit traffic for v and v transits traffic for u. The edge (u, v) is directed from v to u.

- *peer-to-peer* $u \cdots v$

  ASes u and v have a peering relationship if u does not transit traffic for v and v does not transit traffic for u. The edge (u, v) is undirected.

- *sibling-to-sibling* $u \cdots v$

  ASes u and v have a sibling relationship if u transits traffic for v and v transits traffic for u. The edge (u, v) is undirected.

Because provider-customer relationships are considered asymmetric, and peer-to-peer and sibling-to-sibling relationships are considered symmetric, the edges in the graph between providers and customer are directed while the edges between siblings and peers are undirected.

The relationship between the ASes does not correspond to their commercial agreements.

In addition, a definition of a *valley-free* is made as: "After traversing a provider-to-customer or peer-to-peer edge, the AS path cannot traverse a customer-to-provider or peer-to-peer edge".

In addition, the traffic flows from customer to provider only in certain conditions. See section 1.1.2 for more details about AS interconnection.

Therefore, an heuristic algorithm is proposed assuming that a provider has a larger size than its customer, and the size of an AS is proportional to its degree in the AS level connectivity graph. The experimental results [29] indicate that 90% of the links in the Route-Views database are of type customer-provider, 8% are of type peer-peer, and 1.5% are of type sibling-sibling.

For this reason, sibling-to-sibling relationships are not often taken into account in the modeling of the interconnection structure.

An analysis of the properties of the annotated graphs obtained with this heuristic algorithm is provided by Ge et al. [58].

The topology presented by Gao, and the previously described Gao-Rexford Constraints [56], have been used in several other works [77] [136] [56] [29] [33] [41] [96] [6] [55], becoming the reference point of the set of AS relationships.

A formal definition of the *Type of Relationship (TOR)* problem as a maximization problem is described by Subramanian et al. [136] as follows:

They denote an edge from a customer to a provider with a -1, an edge from one peer to another with a 0, and edge from a provider to a customer with a $+1$.

If every AS obeys the customer, peer, and provider export policies, then every advertised path belongs to one of these two types for some $M, N \geq 0$:

- Type-1: $-1, \ldots (Ntimes), +1, \ldots (Mtimes)$
- Type-2: $-1, \ldots (Ntimes), 0, +1, \ldots (Mtimes)$

The first stage of a Type-1 path contains only customer-provider links and the second stage contains only provider-customer links.

The Type-2 captures all paths which traverse exactly one peering link.

In order to solve the ToR problem, a structure of partial views of the AS graph as seen from different locations is considered. Each partial view, called *vantage point*, is taken from the routing table of a BPG speaker.

Then, each vantage point is combined with the others. This technique is called Internet Hierarchy from Multiple Vantage Points.

Formally, given an undirected graph $G = (V, E)$, with vertex set V and edge set E and a set of paths P, label the edges in E as either -1, 0 or $+1$ to maximize the number of valid paths in P. G represents the entire Internet topology, and P consists of all paths seen from the various vantage points.

4.4 Modeling the Exporting Policies

The relationships between ASes are translated into policies for exporting route advertisements via BGP sessions. Each AS defines its export policies according to its agreements with their neighbors.

Alaettinoglu [6] defines four types of exporting policies. This set of policies used in several other work [77] [53] [136] [34] and considered the

standard set of exporting policies, are described as follows:

Exporting to a provider

- in exchanging routing information with a provider, an AS can export local routes;
- in exchanging routing information with a provider, an AS can export routes of its customers;
- in exchanging routing information with a provider, an AS usually does not export routes learned from its providers;
- in exchanging routing information with a provider, an AS usually does not export routes learned from its peers.

Exporting to a customer

- in exchanging routing information with a customer, an AS can export local routes;
- in exchanging routing information with a customer, an AS can export routes of its customers;
- in exchanging routing information with a customer, an AS can export routes learned from its providers;
- in exchanging routing information with a customer, an AS can export routes learned from its peers;
- in exchanging routing information with a customer, an AS can export routes learned from its sibling ASes [34].

Exporting to a peer

- in exchanging routing information with a peer, an AS can export local routes;
- in exchanging routing information with a peer, an AS can export routes of its customers;
- in exchanging routing information with a peer, an AS usually does not export routes learned from its providers;
- in exchanging routing information with a peer, an AS usually does not export routes learned from its peers.

Exporting to a sibling

- in exchanging routing information with a peer, an AS can export local routes;
- in exchanging routing information with a peer, an AS can export routes of its customers;
- in exchanging routing information with a peer, an AS can export routes learned from other providers;
- in exchanging routing information with a peer, an AS can export routes learned from other peers;
- in exchanging routing information with a peer, an AS can export routes learned from other sibling ASes [34].

4.5 Issues

Generally speaking, all the relationships and the topologies shown in the previous sections present several important drawbacks, and do not capture the global hierarchical structure.

Firstly, the hierarchical and structured AS representation of the interconnection structure of the Administrative Domains, and the classification of the AS relationships into provider-to-customer, customer-to-provider, peer-to-peer and sibling-to-siblings edges, does not reflect a real business agreement between the Administrative Domains.

A study of the real commercial relationships between Administrative Domains will be made thereafter, but just right now we can affirm that a one-to-one relationship between business agreement and an AS relationship is not possible. So, the rules in the export policies previously described are not completely responding to the real enviroment.

In addition, we will see that several types of agreements between ASes are possible with the generation of confusing rules and approaches. As example, a public peering agreement between to ASes can be considered as a provider-customer relationship in modeling of the AS relationships.

Besides, the Type of Relationships problem does not consider the hierarchical structure of the AS graph, and accepts cyclic structures as solutions.

In this way a solution with a cycle such as:

$$\overleftarrow{\textit{nationalprovider} \rightarrow \textit{regionalprovider} \rightarrow \textit{localprovider}}$$

is acceptable, in contradiction with the assignment that a customer-provider relationship does not contain cycles and, in general, with the Gao-Rexford Constraints[5] [56]. There are hundreds of similar real-world examples [41] [34] .

Another drawback is related to the affirmation that ASes prefer customer routes over routes through peers or providers, because ASes do not have to pay for sending traffic to a customer [34]. This is true only to some AS interconnection types (see section 1.1.2 for more details about AS interconnection). If the *Exporting to a provider policies* previously described are true, and, in particular, the sentence "*in exchanging routing information with a provider, an AS usually does not export routes learned from its peers and its providers*" is true, then the customer may hide to a provider a set of routes learned from its providers and peers, hijacking provider's traffic towards other connections.

Finally, two ASes may have a peer indirectly through an intermediate AS [56].

In summary, starting from a data set, and according to the synthetic

[5]Several variants are proposed to the Type of Relationship problem as a maximization of the number of paths keeping the directed graph acyclic, with no applicable results [29] [87].

topologies shown in the previous sections, a mechanism of *reverse lookup* - in order to have a clear and a realistic view of the interconnection structure of the network - is not actually possible.

In addition, a process of verification of the results related to topology inference is difficult without a complete and accurate repository of the relationships between ASes, which are actually considered sensitive information by Administrative Domains.

In the next chapter, an analysis of the business relationships and the possible strategies for competitions between Administrative Domains is made.

Chapter 5

Toward a more realistic model

As general assumptions, Internet is not considered as a well ordered provider-client hierarchy, but a no-ordered subset of interconnections, driven by business agreements, and where performance is not the first scope.

For these reason, business relationships reflect both packet flow and money flow.

In this chapter, a deep analysis of these business relationships to identify the Settlement Model, and an exploration of possible strategies for competitions between Administrative Domain is made.

5.1 Analysis of the Settlement Model

A value of the traffic and money flows between Administrative Domains must be given, in order to identify a *Settlement Model* of these transactions.

The Settlement Model in the Public Switched Telephone Network (PSTN) is easy to identify. A call between two users $U_1 \rightarrow U_2$ is made via a provider P_1. In general, P_1 belongs to a set n of providers $P_1 \rightarrow P_2 \rightarrow \ldots \rightarrow P_n$. In this model, U_1 pays P_1 for this end-to-end service. The contractual agreements between P_1 and P_2 , P_2 and P_3, $\ldots$, P_{n-1} and P_n ensure a periodically and balanced money flow between all the actors involved.

In this way, each traffic flow and each money flow is identified, with each transaction with a measurable value.

The analysis of the Settlement Model in the business relationships between Administrative Domains reveals that an identification through bidirectional and measurable transactions of traffic and money flows is not possible. Each individual IP packet can be considered as an individual 'transaction', but a per-packet charging is not a a practicable way (as example, packets can be lost).

Indeed, an analysis of costs and benefits related to the type of services provisioned and managed by the Administrative Domains is necessary. Prices for services can vary between Administrative Domains, even for the same services, and depend on several factors such as physical network topology, redundancy, types of interconnection with other networks, sub-

scription ratios (available capacity:utilized capacity)[1], demarcation point issues[2].

In the analysis of this Settlement Model, two types of business agreements between Administrative Domains are considered:

- *Transit*; where one Administrative Domain provides reachability to all destinations in its routing table to its customers;
- *Peering*; where Administrative Domains provide mutual reachability to a set of their routing table.

An analysis of these two models is made in the following sections.

5.2 Transit

One Administrative Domain provides reachability to all destinations in its routing table to its customers.

Transit services are generally sold in the form of Ethernet connections with different port speeds (10 Mbps, 100 Mbps, 1 Gbps, 10 Gbps port

[1]Subscription ratios vary based on the product being offered, typically from 4:1 ratios (no more than four links for each backbone connection) to 10:1 ratios. Higher subscription ratios ensures higher percentages of network bottlenecks and congestions.

[2]A d*emarcation point* is the boundary between the network and responsibilities of each Administrative Domain. Demarcation points are defined down to the cables and connectors to ensure that no disagreements occur in case of equipment or network problems.

speed)[3].

A common practice to determine the traffic volume on a connection for billing purposes over a calendar month is the *95% model for traffic measurement* as follows [115]:

- every five minutes two measurements (transmission and reception) of the total traffic passed over the port since the last measurement are made;
- at the end of the calendar month all measurements (the highest of transmit and receive) of that month, generally 8.640[4] measurements per direction are lined up and sorted from high to low;
- the highest 5% of the measurements ($5\% * 8.640 = 432$) is dropped;
- the next highest measurement defines the 95% traffic level on which the billing for that month is based.

In this way a burst traffic of (432 * 5 minutes / 60 =) ± 36 hours per month does not affect on the monthly payment.

[3] The carrier can limit the port speed to a lower value in some cases. In order to avoid problems related to traffic bursts and congestion, the port speed is usually chosen by considering an average traffic usage lower than 75% of the maximum port speed. A port buffer is available but when the port buffer is full, traffic will be dropped.

[4] 12 (measurements per hour and direction) * 24 (hours) * 30 (days) = 8.640 measurements.

A *Traffic commitment level* is one of the most commonly forms of payments in a traffic business agreement. All traffic up to this traffic commitment level is included in the fixed monthly price, and any traffic over the commitment level is charged based on an agreed burst fee. So, using 6 Mbps in a 10 Mbps commitment there is no any additional monthly fee to pay, while using 16 Mbps there is an extra fee of 6 Mbps of burst traffic.

5.3 Peering

Peering allows the exchange of routes and traffic limited to both networks and their respective customers, and does not include transit routing to non-customer networks.

Administrative Domains provide mutual reachability to a set of their routing table. These destinations are generally customers, and reached via zero cost peering links in many cases.

Peering reduces the traffic flow sent to its providers, saving operational costs. Besides, additional equipment and management costs are required.

Peering can be realized through private interconnections (*private peering interconnection*) or public interconnection (*public peering interconnection*), through *bilateral peering relationships* or *multilateral peering relationships*.

5.3.1 Private Peering Interconnection

A *private peering interconnection* is a private interconnection made in general through a dedicated point-to-point network cable. Private peering allows a direct control over the traffic flows, but the cost for buying, maintaining and managing the equipment, and for each interconnection, can be very high. For each AS to connect, a cable is required.

5.3.2 Public Peering Interconnection

A *public peering interconnection* is a interconnection implemented through a physical public infrastructure called *Internet eXchange Point (IXP)*, using virtual circuits in a Switching Mesh.

In this case, only one cable to connect with the Internet Exchange Point is required, reducing the cost for buying, maintaining and managing the equipment. However the available bandwidth capacity between any two participants can be limited.

The physical interconnection (Layer-1 and Layer-2 ISO/OSI model) with a IXP does not ensure reachability to other ASes. An IXP provides only a physical connectivity among all participants. Traffic can be exchanged after a negotiation of peering agreements with other Administrative Domains. It is up to individual networks to decide with whom to establish BGP sessions.

Name	Country	Members	Update
Equinix [38]	USA, Europe, Asia	491	2010-09-10
AMS-IX [8]	Amsterdam (Netherlands)	365	2010-09-15
DE-CIX [32]	Frankfurt (Germany)	353	2010-08-15
LINX [97]	London (UK)	332	2010-07-31
MSK-IX [112]	Moscow (Russia)	304	2010-07-31
NL-ix [114]	Amsterdam (Netherlands)	240	2010-03-11

Table 5.1: List of Internet eXchange Points by members

However, the payload peering traffic over the Internet Exchange port is often free of charge.

Internet Exchange Points are classified by the number of members, traffic volume and amount of routes. Table 5.1 shows a list of Internet Exchange Points by members [80].

Table 5.2 shows an example of pricing of an interconnection with two different IXPs in Netherland, NL-ix and AMS-IX [115].

In this example, the demarcation point of the service is the port on the NL-ix or AMS-IX switch on the datacenter. The patchcable from customer's equipment to the IXP is not included.

	Port Speed	100 Mbps$^{\alpha}$	1 Gbps$^{\beta}$	10 Gbps$^{\gamma}$
NL-ix	Initial	250 €	500 €	1000 €
	Monthly recurring	100 €	350 €	1000 €
AMS-IX	Initial	0 €	0 €	0 €
	Monthly recurring	500 €	1000 €	2500 €

$^{\alpha}$ RJ45/UTP connector

$^{\beta}$ Multimode fiber SX/850nm or singlemode fiber LX/1310 nm (+ 500 €initial fee)

$^{\gamma}$ Singlemode fiber LX/1310 nm

Table 5.2: Pricing of an Interconnection with NL-ix and AMS-IX

5.3.3 Bilateral Peering Relationship

In a *Bilateral Peering Relationship*, members will peer on a one-to-one basis through a private peering interconnection or a public peering interconnection. A bilateral peering relationship is commonly made via Layer-2 direct switching, or Layer-2 virtual circuits in a Switching Mesh.

A bilateral peering relationship allows each member to select a preferred path to a given destination.

The bilateral *peering agreement* is the formal and signed document that formalizes the relationship between the parties, including routing policy, the settlement character of the traffic exchange, the duration of thee agreements, technical best practices, and so on.

Table 5.3 shows an example of pricing of a bilateral peering through Open Peering [115], a bilateral *payed peering* interconnection with the implementation and maintenance of a full set (up to 350) or a subset (top-25) of peers.

Strenght:

- private peering;
- granular policy control;
- easy monitoring control and troubleshooting;
- total control over legal contract and technical agreements.

Weakness:

- technically complex;
- one BGP session per neighbor;
- management of multiple legal contracts and technical agreements;
- high cost.

5.3.4 Multilateral Peering Relationship

In a *Multilateral Peering Relationship*, members will only peer with a *Exchange Route*. The Exchange Route announces the members routes to all peers and is able to select a preferred path to a given destination, imposing transit policies.

Type of Peering		Price
Bilateral - Full	Initial	12000 €
	Monthly recurring	1200 €
Bilateral - Top 25	Initial	3000 €
	Monthly recurring	300 €

Table 5.3: Pricing of a bilateral peering with Open Peering with a full bilateral peering or a top-25 largest networks on an Internet Exchange Point

A Multilateral Peering Relationship can be made through a private peering interconnection with the Exchange Route, or a public peering interconnection . In this case the Exchange Route is an Internet Exchange Point.

Also in this case, a multilateral *peering agreement*, the formal and signed document that formalizes the relationship between the parties, is required.

In the case of a mutual peering relationship over an Internet Exchange Point, a compliance with an Acceptable Use Policy is required. An *Acceptable Use Policy (AUP)* is a standand formal and signed document containing terms and conditions of use, technical agreements of use, to improve the efficiency of routing and the general connectivity.

Examples of AUP are "Technical Standards and Policy for Subscribers to LAP & MAE-LA" [141] and the "Multi-Lateral Peering Agreement (MLPA)" [113].

In these AUP are defined standards, policies, rules to respect and obligations.

Here some rules:

- exchange of routes will be performed using BGP4 [126];
- subscribers will route prefixes that are a maximum prefix length of 24 bits. Aggregation of routing information where possible is required;
- subscribers will make use of a unique Autonomous System Number assigned by a suitable registration authority;
- routing policy must be published in the Internet Routing Registry [78];
- subscribers are obligated to advertise all its customers' routes to all other participants and to accept the customer's routes advertised by other participants;
- subscribers are not obligated to announce routes obtained from its Bilateral Peering Agreements;
- subscribers are not obligated to provide transit to other subscribers;
- the agreement is implemented by each subscriber on a best-effort basis.

In general, monetary settlements are not required, and there is no installation fee and recurring fee, but the costs related to hardware and connection to the Internet Exchange Point are not covered by the AUP.

Multilateral Peering	**MLPA Registry**	**MLPA Routing**
Initial	Free	Free
Monthly recurring	Free	Free

Table 5.4: Pricing of a multilateral peering with MLPA Registry and Routing services

.

Table 5.4 shows an example of pricing of a multilateral peering with MLPA Registry and Routing services. MLPA Registry and Routing services are free of charge and not for profit services of Open Peering [115] and provided on a time-permitting basis without 24*7 support and, theoretically, they can be terminated at any point in time.

Strenght:

- technically easy;
- management of a single legal contract;
- cost.

Weakness:

- no private peering;
- lack of policy control;
- complex monitoring control and troubleshooting;
- additional AS_PATH.

5.3.5 Strategies for Peering Competitions

Broadly speaking, peering guarantees short and fast connections, improving performance and resilience, reducing bottlenecks and dependences on transit providers. It also reduces the costs of delivering traffic, and can increase money flows by new contractual agreements with customers due to a more appealing status.

In addition, peering is free in general[5].

In a zero cost bilateral peering, no money is payed for traffic exchanged and the costs for the infrastructure are shared.

This implies equal benefit to all the actors involved. So, peering is an appealing solution.

For these reasons, many Administrative Domains have very selective peering engagements, because peering consumes resources and requires continuous efforts in maintenance [9] [11].

Besides, small Administrative Domains have in general poor services, and an unbalanced traffic flow between large and small peers is an unwilling solution.

In addition, it is possible that substantial part of the traffic flow in a small peer is from a larger one, impacting on the quality of the service offered to other peers, or not allowing new contractual agreements with

[5]It is not true. Often peering is made using forms of contractual agreements called *payed peering*.

other ASes.

For this reason, the large peers tend to have strict peering policies, demanding strict maximum subscription ratios or peering in geographically distribute locations.

In the next chapter, a formulation of a methodology to structure the different aspects to be taken into account in a peering engagement is proposed, in order to optimally solve the trade-off of implementing a peering engagement against the extra cost that this solution represent.

Chapter 6

Ex-Ante Evaluations of Peering Engagements

A key problem to be faced by Administrative Domains is how to optimally solve the trade-off of implementing a peering engagement against the extra cost that this solution represent.

An estimation of the additional income due to the peering engagement is required.

As seen an Administrative Domain may choose between multiple solutions with different monetary costs.

In this chapter, the formulation of a methodology to structure the different aspects to be taken into account in a peering engagement, and efficiently solve the decision problem of maximization of the importance

of these aspects, subject to a mutual relationship between the involved aspects and budget constraints is explained.

This study makes the following contributions:

- a comparative analysis of the aspects and alternative options to be taken into account in ex-ante evaluations of a peering engagement is explained;

- a decision maker called *XESS*^{2} *(eXtended EGP Support System)* able to process the aspects and the alternative options in a peering engagement, in order to find candidate solutions in a fast and high efficient way, and to produce a synthetic conclusion on the allocation of budgets and on the enhancements of effectiveness of the services is proposed.

  XESS^2 is the second revision of *XESS (eXtended E-Learning Support System)* [49] [48] [50] [45] [46] [47], a Decision Support System able to make a comparative evaluation of alternative options in an e-Learning solution through a numerical evaluation of the variables and the selection of the best possible solution, using a combinational optimization formulation and an integer programming formulation of the problem [1] [13] [83] [109].

6.1 Modeling Peering Engagements

In this section an analysis of the aspects and alternative options to be taken into account in ex-ante evaluations of a peering engagement is explained, to formulate an exhaustive model of the real environment, to be used by the decision maker to find candidate solutions, and to produce synthetic conclusion on the allocation of budgets. So, this section is intended to focus primarily on the definition of the problem properly, in order to be as exhaustive as possible.

In the next subsections a description of all options is made.

6.1.1 Category *Equipment*

This category is related to the purchase of equipment and the related costs of the human resources required to make the infrastructure operational.

BGP Router

A router which supports the BGP4 protocol is required, and pricing is defined by the router *class* and by using refurbished routers. A *refurbished router* is an used router, which have completely been updated and tested.

Table 6.1 describes the available options. For each Router Class, two options related to a new router or a refurbished router are available.

Option	Router Class	Description
1-2	100% CAM Router	A full routing table is contained in the router's Content Addressable Memory (CAM). Each packet is forwarded without using the CPU.
3-4	CAM Cache Router	A partial routing table is contained in the router's CAM.
5-6	Appliance Router	Based on standard PC hardware components, and running a custom OS and routing software. The performance is largely limited by the performance of the CPU.
7-8	Software Router	Based on standard PC hardware, and an open source Operating System and open source routing software. The main advantages of software routers is their low cost, despite to their performance (limited by the performance of the CPU) and stability.

Table 6.1: Router Options

Each option is intended with a standard configuration included of:

- Chassis;
- Power Supply;
- Slot cover;
- Power Cable;
- Documentation.

Additional options are considered in Table 6.2.

Hardware Setup and Support

Two types of hardware setup and support services are considered [Table 6.3].

Hardware support services are considered as monthly costs with different support hours and max response time. For refurbished equipment, they are often calculated as percentage of the total refurbished prices.

BGP Support

Three type of BGP support service are considered [Table 6.4]. The support is done only via email, phone, or remote access. No on-site support is included. Generally, pricing is divided into initial costs and monthly costs, and a maximum amount of management or support hours per month is considered.

Option	Additional Router Options	Description
9	Chassis	Chassis, with more empty interface and power supply slots.
10	Additional Power Supply	Additional Power Supply.
11	Interfaces & Management	UTP Ethernet Blades, or SX multiMode Fiber Blade with Management Blades.
12	Lasers	Lasers.
13	Patches	Singlemode patches or multimode patches.
14	Accessories	Slot cover for Power Supply slot or for empty Interface Slot, Flash Disk for management modules.
15	Other	Other options.

Table 6.2: Router Additional Options

Option	Hardware Setup	Description
16	Router Setup 1	Setup of BGP sessions, to transit providers and peers.
17	Router Setup 2	Setup of BGP sessions, to transit providers and peers. Installation or upgrade of the Operating System. Configuration of ethernet interfaces, VLAN's, interface IP addresses and static routes.
18	Support Service 1	Support Hours: 8*5 , Max Response time (MRT): Next Business Day.
19	Support Service 2	Support Hours: 24*7,Max Response time (MRT): 4 hours.

Table 6.3: Hardware Setup and Support

Course

In order to have a technical knowledge of routing, extra fees can be payed for courses [Table 6.5]. They are generally divided into a theory part (routing, addressing, BGP route mechanisms, policies, tools, issues and troubleshooting), and a practical workshop part (router setup and configuration of transit and peering, path and attribute manipulation, filtering and security).

6.1.2 Category *Addresses*

AS Number

A globally unique identification (AS number) is required in exchanging exterior routing information with other networks [Table 6.6]. In the European region, the AS number is assigned by the RIPE Network Coordination Center (NCC), and needs to be registered and maintained in the RIPE database.

IP Space

An IP Space (a set of globally unique IP addresses) is required to be able to route traffic between Administrative Domains. A minimum block of 256 IP addresses is required. Pricing depends on the IP Space width [Table 6.7].

Option	BGP Support	Description
20	BGP Service 1	Support Hours: 8*5, Max Response Time: Next Business Day.
21	BGP Service 2	Support hours: 24*7, Max Response Time: 4 hours.
22	BGP Service 3	Support hours: 24*7, Max Response Time: 1 hour.

Table 6.4: BGP Support Services

Option	Course	Description
23	BGP Course	A course divided into a theory part and a workshop part with a maximum number of attendants per course.

Table 6.5: Course

Option	AS Numbers	Description
24	AS Number	The Registration and the maintenance of an AS Number as an yearly recurring fee.

Table 6.6: AS Number Registration

6.1.3 Category *Connectivity and Rackspace*

Generally, the Demarcation Point of a service is a port of the network device on a Datacenter. So, the network connectivity with the Datacenter [Table 6.9] and the rack collocation in the Datacenter (Rackspace) [Table 6.8] are considered in this category.

6.1.4 Category *Peering*

According to Section 5.3, private and public peering interconnections through bilateral or multilateral peering relationships are considered [Table 6.10].

Pricing are divided into initial costs and monthly recurring costs.

The physical interconnection with a IXP does not ensure reachability to other ASes, but only a physical connectivity among all subscribers. Traffic is exchanged after a negotiation of peering agreements with other Administrative Domains.

So physical interconnection and peering agreements with a peer must be considered in each option.

Option	IP Space	Description
25	256+	IP Space of 256 IP addresses and multiple IP Spaces. Pricing is generally divided into a price for the first 256 addresses block, and a price for any extra 256 addresses block.
26	2048	An IP Space of maximum 2048 IP addresses.
27	4096	An IP Space of maximum 4096 IP addresses.
28	8192	An IP Space of maximum 8192 IP addresses.

Table 6.7: IP Space

Option	RackSpace	Description
29	Rackspace 1	A rackspace for partial rack (10U,11U,14U)
30	Rackspace 2	A rackspace for full height rack (42U)

Table 6.8: Rackspace

Option	Connectivity	Description
31	100 Mbps	100 Mbps connection based on 100Base-TX standard, Cat5e UTP cable for a maximum distance of 100 meter.
32	1 Gbps MM	1 Gbps connection based on 1000Base-X or 1000Base-SX (850 nm) standard over multi mode fiber (with 62.5/125 μm core/cladding diameter) using SC connectors, for a maximum distance of 550 meter.
33	1 Gbps SM	1 Gbps connection based on 1000Base-LX (1310 nm) standard over single mode fiber, for a maximum distance of 10 - 25 Km.
34	10 Gbps	10 Gbps connection based on 10GBase-LR (1310 nm) standard over single mode fiber (with 8-10/125 μm core/cladding diameter), for a maximum distance between 10 and 25 Km, depending on the cable quality and loss specifications.

Table 6.9: Connectivity

Option	Peering	Description
35	2-Private	Bilateral Peering through a private peering interconnection (See section 5.3.3 for details).
36	2-Public-F	Bilateral Peering through a full public bilateral peering interconnection (See section 5.3.3 for details). As example, Open Peering [115] [Table 5.3] considers the implementation and maintenance up to 350 potential peers.
37	2-Public-S	Bilateral Peering through an interconnection based on a subset of peers. (See section 5.3.3 and Table 5.3).for details).
38	2-Public-IXP	Bilateral Peering interconnection implemented through an IXP (See section 5.3.2 for details). Only the physical interconnection with the IXP is considered.
39	M-Private	A Multilateral Peering with a Exchange Route (See section 5.3.4 for details). Only the physical interconnection with the Exchange is considered.
40	Mu-Public	A Multilateral Peering with an IXP(See section 5.3.4 for details). Only the physical interconnection with the IXP is considered.

Table 6.10: Peering Relationship

Option (ω)	*Name*	*Cost* $c(\omega)$	*Importance* $\lambda(\omega)$
Option 1	100% CAM Router New	40000 €	80%
Option 3	CAM Cache Router New	20000 €	50%
Option 4	CAM Cache Router Refurbished	5000 €	50%
Option 23	BGP Course	500 €	20%

Table 6.11: Examples of cost values and importance values related to some options

6.2 Problem Formulation

This section introduces a set of assumptions, and a formulation of the mathematical model is made.

Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a set of finite n options to be taken into account in ex-ante evaluations of a peering engagement.

A cost $c(\omega_i)$ and an importance $\lambda(\omega_i)$ are associated with each option $\omega_i \in \Omega$ [Table 6.11]. CAPEX and OPEX are included into each cost $c(\omega_i)$[1].

[1] An analysis of the cost and the importance of each option is behind the scope of this work.

Let $\chi(\omega_i)$ be a binary decision variable of the option ω_i with the following properties:

$$\chi(\omega_i) = \begin{cases} 1 & \text{if the option } \omega_i \text{ is chosen} \\ 0 & \text{otherwise} \end{cases} \tag{6.1}$$

Finally, let C be an assigned maximum budget to respect.

The goal of the problem is to find an optimal reduced set of solutions $\Omega^* \subseteq \{\omega_1, \ldots, \omega_n\}$, which can be formally stated as the following integer programming formulation (Table 6.12 introduces the notation used):

$$\textit{maximize} : \sum_{i=1}^{n} \lambda(\omega_i)\chi(\omega_i) \tag{6.2}$$

$$\textit{subject to} : \sum_{i=1}^{n} c(\omega_i)\chi(\omega_i) \leq C \tag{6.3}$$

$$\chi(\omega_i) \in \{0, 1\} \qquad \forall i \in \{1, \ldots, n\} \tag{6.4}$$

$$\sum_{r=R_0}^{R_{max}} \chi(\omega_r) = 1 \qquad \forall R_0, R_{max} \in \{1, \ldots, n\} \tag{6.5}$$

$$\sum_{s=S_0}^{S_{max}} \chi(\omega_s) \leq 1 \qquad \forall S_0, S_{max} \in \{1, \ldots, n\} \tag{6.6}$$

$$\sum_{u=U_0}^{U_{max}} \chi(\omega_u) \leq 1 \qquad \forall U_0, U_{max} \in \{1, \ldots, n\} \tag{6.7}$$

$$\sum_{v=V_0}^{V_{max}} \chi(\omega_v) \leq 1 \qquad \forall V_0, V_{max} \in \{1, \ldots, n\} \tag{6.8}$$

$$\sum_{j=J_0}^{J_{max}} \chi(\omega_j) \leq 1 \qquad \forall J_0, J_{max} \in \{1, \ldots, n\} \tag{6.9}$$

$$\sum_{k=K_0}^{K_{max}} \chi(\omega_k) \leq 1 \qquad \forall K_0, K_{max} \in \{1, \ldots, n\} \tag{6.10}$$

$$\sum_{m=M_0}^{M_{max}} \chi(\omega_m) \leq 1 \qquad \forall M_0, M_{max} \in \{1, \ldots, n\} \tag{6.11}$$

$$\sum_{q=K_0}^{M_{max}} \chi(\omega_q) \neq 1 \qquad \forall K_0, M_{max} \in \{1, \ldots, n\} \tag{6.12}$$

$$\sum_{p=P_0}^{P_{max}} \chi(\omega_p) \leq 1 \qquad \forall P_0, P_{max} \in \{1, \ldots, n\} \tag{6.13}$$

Symbol	Desription
Ω	The set of finite options to be taken into account in ex-ante evaluations of a peering engagement
n	Total number of the options to be taken into account in the analysis
$c(\omega_i)$	Cost of the option ω_i
$\lambda(\omega_i)$	Importance of the option ω_i
$\chi(\omega_i)$	Decision variable $\{0, 1\}$ depending if the ω_i is chosen or not
C	Total admissible cost of the peering engagement
$\{\omega_{R_0}, \ldots, \omega_{R_{max}}\}$	Set of candidate routers
$\{\omega_{S_0}, \ldots, \omega_{S_{max}}\}$	Set of candidate options for setup of BGP sessions
$\{\omega_{U_0}, \ldots, \omega_{U_{max}}\}$	Set of candidate options for Hardware Support Services
$\{\omega_{V_0}, \ldots, \omega_{V_{max}}\}$	Set of candidate options for BGP Support Services
$\{\omega_{J_0}, \ldots, \omega_{J_{max}}\}$	Set of candidate options for IP Space
$\{\omega_{K_0}, \ldots, \omega_{K_{max}}\}$	Set of candidate options for a rackspace.
$\{\omega_{M_0}, \ldots, \omega_{M_{max}}\}$	Set of candidate options for a network connectivity with the Datacenter.
$\{\omega_{P_0}, \ldots, \omega_{P_{max}}\}$	Set of candidate peering agreements

Table 6.12: Notation

Expression (6.2) represents the object function. Expression (6.3) ensures that the total cost of the chosen set of options does not exceed the total budget C. Expression (6.4) ensures that each $\chi(\omega_i)$ is either 0 or 1. Expression (6.5) assures that at least one router ω_{r*} is chosen from a set of $\{\omega_{R_0}, \ldots, \omega_{R_{max}}\}$ candidates.

Expression (6.6) assures that at most one setup of BGP sessions ω_{s*} is selected from a set of $\{\omega_{S_0}, \ldots, \omega_{S_{max}}\}$ choices. Expression (6.7) ensures that at most one Hardware Support Service ω_{u*} is selected from a set of $\{\omega_{U_0}, \ldots, \omega_{U_{max}}\}$ alternatives. Expression (6.8) ensures that at most one BGP Support Service ω_{v*} is chosen from a set of $\{\omega_{V_0}, \ldots, \omega_{V_{max}}\}$ choices.

Expression (6.9) assures that at most one IP Space ω_{j*} is chosen from a set of $\{\omega_{J_0}, \ldots, \omega_{J_{max}}\}$ alternatives.

Expressions (6.10), (6.11) and (6.12) ensure that a network connectivity with the Datacenter ω_{m*} and a rack collocation in the Datacenter (Rackspace) ω_{k*} are simultaneously considered.

Finally, expression (6.13) assures that at most one peering agreement ω_{p*} is selected from a set of $\{\omega_{P_0}, \ldots, \omega_{P_{max}}\}$ alternatives.

The solution of the above integer programming formulation gives the optimal combination of the options Ω^* to be considered in the analysis of a peering engagement from a set of candidate solutions.

The problem in (6.2) is a special case of the Knapsack Problem [83] [1] [13].

6.3 Practical Implementation

In order to solve the decision problem of maximization of the importance of the aspects related to peering engagements, subject to a mutual relationship between the involved aspects and budget constraints, a practical implementation called *XESS*^{2} *(eXtended EGP Support System)* has been constructed.

Both *XESS*^{2} (eXtended EGP Support System) and *XESS* (eXtended E-leaning Support System) use a common framework [49] [48] [50] [45] [46], which is able to make a comparative evaluation of alternative options through a numerical evaluation of a set of variables, and to find an optimal reduced set of solutions Ω^* using a combinational optimization formulation and an integer programming formulation of the problem [1] [13] [83] [109].

This framework is designed to help decision-makers to integrate the different options and to produce a single synthetic conclusion at the end of the evaluation, aiding the stakeholders in understanding and choosing the best possible solution when the result is not obvious.

The framework exhibits a four-tier architecture:

1. *a model of the real environment*, intended to focus primarily on the definition of the problem properly, in order to be exhaustive as possible.

   In this case, the model is based on an analysis and identification

of the options to be taken into account in ex-ante evaluations of a peering engagement. The options are grouped into four categories (Equipment, Addresses, Connectivity and Rackspace, Peering) and several subcategories, as described in section 6.1. These categories represent only a logical organization of the options, in order to have a clear view of all the choices through a subdivision of the different aspects to be evaluated in logical groups. Therefore they don't affect the final results of the solution algorithm.

A cost $c(\omega_i)$ and an importance $\lambda(\omega_i)$ are associated with each option $\omega_i \in \Omega$ [Table 6.11].

2. *a model of the mathematical correlations and dependences* between the options.

   An analysis of the mathematical correlations and dependences between the options is made, in order to make a numerical evaluation of the benefits in the combined use of the options.

3. *a method for the solution of the integer programming problem* shown in the section 6.2.

   Many solution techniques for this class of problems have been proposed in the literature [14] [15] [83] [104] [61] [36], including linear programming relaxations (with the conversion of the integer problem into a standard linear programming problem), branch and bound

techniques, and heuristic methods (providing suboptimal but acceptable solutions to the integer programming problem).

XESS^2 uses the branch-and-bound method for the solution of the integer programming problem shown in the section 6.2., finding an exact solution to the mathematical formulation of the model using an exact solution approach, and with a reasonable amount of time [1] [13].

For each solution, an *objective function value* that represents the "quality" of the solution (expressed as a quantitative value) is given.

4. *an interface for the decision-makers*.

   A client-server WEB application that provides the interface between *XESS*^{2} and the decision-makers, using a common WEB browser via local network or via internet connection.

The series of simulations performed in [49] [48] [50] [45] [46] provide supporting evidence of the quality of the solutions that the framework used in *XESS* and *XESS*^{2} is capable of finding.

Chapter 7

Conclusions

This thesis has studied the routing decision process in inter-domain routing, with special focus on the analysis of the issues related to the routing decision and to the development of accurate topology generators, and on the development of solutions aimed at optimally implementing of a peering engagement between multiple solutions with different monetary costs.

We have deeply analyzed the BGP protocol and the traditional routing process model, in order to define a new and more complete routing process model, and the formalization of the problem through the *Route Selection function* defined as the maximum value of the degree of preference of all the available routes for a given destination.

We have discussed about routing decisions in the interdomain routing,

which present several issues generated by a lack of a global global coordination between ASes, and demonstrating to be inaccurate and poorly effective in controlling and communicating the inter-domain decisions.

We have shown that all the efforts and the most interesting proposals in this area of research present several drawbacks difficult to solve.

In addition, another active area of research related to the reconstruction of the AS topology, with the building a complete set of links between Autonomous Systems, in order to obtaining an accurate AS-level connectivity has proven difficult, and the AS topology remains by and large incomplete.

We have also discussed about the development of accurate topology generators, which are essential for producing realistic simulation studies of protocols and network architectures, to reduce misconfiguration or to debug router configuration files, and to planning for future contractual agreements, because AS relationships have a profound influence on traffic flows. We have shown that, despite the volume of research in this area, current topology synthetic generators fail to capture an inherent aspect of the AS topology.

The hierarchical and structured AS representation of the interconnection structure of the Administrative Domains, and the current classification of the AS relationships do not reflect the real business agreements between the Administrative Domains, and many considerations are made in order to provide a more accurate and realistic view of the real interconnection structure.

For these reasons, an analysis of the business relationships and the possible strategies for competitions between Administrative Domains has studied, to define standards, policies, rules and obligations between the actors involved.

Finally, we have formulated and efficiently solved the problem the decision problem of maximization of the importance of the different aspects to be taken into account in a process of peering engagement, subject to a mutual relationship between the involved aspects and budget constraints.

A real and complete model of peering engagement is explained, with the definition of 40 variables related to the identification of 40 options grouped into four categories (Equipment, Addresses, Connectivity and Rackspace, Peering) and several subcategories.

The problem is formulated as a integer programming formulation and a practical implementation of a framework called $XESS^2$ (eXtended EGP Support System), which is able to make a comparative evaluation of alternative options through a numerical evaluation of a set of variables, and to find an optimal reduced set of solutions using a combinational optimization formulation and an integer programming formulation of the problem, has been proposed. Extensive experiments and simulations performed in several previous works provide supporting evidence of the quality of the solutions that XESS is capable of finding.

The most promising outcome of this part of this work is that the contributions can be applied in other problems as *XESS* and $XESS^2$ demonstrate.

In particular, our proposals can be applied in all the environments where constrained problems considering maximum cost vs. alternative options are critical.

Appendix A

Publications

- G.Fenu, M.Picconi,
  Identification of the Variables in an E-Learning Platform Using a Cost-Benefit Analysis and an Automatic Decision-Making Tool
  in Collection of A. Respicio, et al. (Eds.) Bridging the Socio-technical Gap in Decision Support Systems. Challenges for the Next Decade. Frontiers in Artificial Intelligence and Applications. IOS Press, pp. 473-484, 2010

- G.Fenu, M.Picconi,
  An Optimized Cost-Benefit Analysis for the Evaluation in E-Learning Services
  F. Zavoral et al. (Eds.), Networked Digital Technologies, Second International Conference, NDT 2010, Prague, Czech Republic, July 7-9,

2010. Proceedings, Part II ,Springer, CCIS 88, pp. 215-225, 2010.

- G.Fenu, M.Picconi,
  A Cost-Benefit Comparison of E-Learning Solutions
  Article in International Journal Of Information Studies Volume 2 Issue 1 January 2010, 2010.

- G.Fenu, M.Picconi, S. Surcis,
  XESS - Extended E-learning Support System
  In Proceedings of NDT 2009, The First International Conference on "Networked Digital Technologies", technically co-sponsored by IEEE Communication Society, VSB-Technical University of Ostrava, Czech Republic, 28-31 July 2009, pp. 165-170, 2009.

- G.Fenu, M.Picconi, S. Surcis,
  XESS, Sistema di Supporto alle Decisioni per la valutazione di soluzioni e-Learning
  in Proceedings of DIDAMATICA 2009, University of Trento, Faculty of Economics, 2009.

- G.Fenu, M.Picconi, S. Surcis,
  Metodiche progettuali per Personal Learning Environments in ambiente Grid e Cloud Computing
  in Proceedings of DIDAMATICA 2009, University of Trento, Faculty of Economics, 2009.

- F.M.Aymerich, G.Fenu, M.Picconi, A.Crisponi, S.Cugia,
  An approach to a PDA based system for network's remote device management
  in Proceedings of 3rd International Conference on Networking ICN'04, sponsored by IEEE, Universite de Haute-Alsace, Proceedings of 3rd IEEE International Conference on Networking (ICN), French Caribbean, Feb 29 - Mar 4, pp. 439-442, 2004.

- F.M.Aymerich, G.Fenu, M.Picconi, A.Crisponi, S.Cugia,
  Infrastruttura di e-Learning con Virtual Classroom Booking
  in Proceedings of AICA 2003 - I costi dell'ignoranza e il valore della conoscenza nella societa' dell'informazione, Atti del Congresso, 15-17 Settembre 2003, pp. 171-174,2003.

- G.Fenu, A.Crisponi, S.Cugia, M.Picconi,
  A Wireless Based System for an interactive approach to medical parameters exchange
  in Proceedings of IMAGE - e-Learning, Understanding, Information Retrieval, Medical, Proceedings of the First International Workshop, Cagliari, Italy, 9-10 June 2003, pp.200-210, 2003.

Bibliography

[1] R.K. Ahuja, T.L. Magnanti, J.B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, 1993.

[2] D.O. Awduche, J. Agogbua, J. McManus, *An Approach to Optimal Peering Between Autonomous Systems in the Internet*, in Proceedings of International Conference on Computer Communications and Networks (IC3N'98), Lafayette, Louisiana, October 1998.

[3] S. Agarwal, A. Nucci, S. Bhattacharyya, *Controlling Hot Potatoes in Intradomain Traffic Engineering*, SPRINT ATL Research Report RR04-ATL-070677, July 2004.

[4] A. Akella, J. Pang, B. Maggs, S. Seshan and A. Shaikh, *A Comparison of Overlay Routing and Multihoming Route Control*, in Proceedings of ACM SIGCOMM, Portland, USA, August 2004.

[5] A. Akella, S Seshan, A. Shaikh, *Multihoming Performance Benefits: An Experimental Evaluation of Practical Enterprise Strategies*, USENIX Annual Technical Conference 2004, Boston, MA, USA.

[6] C. Alaettinoglu, *Scalable router configuration for the Internet*, in Proceedings of IEEE IC3N, October 1996.

[7] C. Alaettinoglu, C. Villamizar, E. Gerich, D. Kessens, D. Meyer, T. Bates, D. Karrenberg, M. Terpstra, *Routing Policy Specification Language (RPSL)*, Internet Engineering Task Force: RFC 2622, June 1999.

[8] AMS-IX (Amsterdam, Netherlands), http://www.ams-ix.net/.

[9] *AOL peering requirements*, http://www.atdn.net/settlement free int.shtml.

[10] ARIN, http://www.arin.net/whois/arinwhois.html.

[11] *AT&T peering requirements*, http://www.corp.att.com/peering/.

[12] H. Ballani, P. Francis,X. Zhang, *A study of prefix hijacking and interception in the Internet* in Proceedings of ACM SIGCOMM, August 2007.

[13] M. Bazaraa, J. Jarvis, H. Sherali, *Linear Programming and Network Flows*, Wiley, 1994.

[14] C. Bazgan, H. Hugot, D. Vanderpooten, *An efficient implementation for the 0-1 multi-objective knapsack problem*, in WEA, pages 406-419, 2007.

[15] C. Bazgan, H. Hugot, D. Vanderpooten, *Solving efficiently the 0-1 multi-objective knapsack problem*, Computers & Operations Research, 36(1), pp. 260-279, 2009.

[16] U.Bornhauser, P.Martini *A Divergence Analysis in Autonomous Systems using Full-mesh iBGG*, in Proceedings of Communication Networks and Services Research Conference, 2008.

[17] *BRITE Intemet Topology Generator*, http://cs-www.bu.edu/brite/.

[18] A. Bremler-Barr, Y. Afek, S. Schwarz, *Improved BGP Convergence via Ghost Flushing*, in Proceedings of IEEE INFOCOM, 2003.

[19] CAIDA, http://www.caida.org.

[20] J. Chandrashekar, Z. Duan, Z. L. Zhang, J. Krasky, *Limiting path exploration in BGP*, in Proceedings of INFCOM, Miami, USA, 2005.

[21] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, W. Willinger. *Towards capturing representative AS-level Internet topologies*, Elsevier Computer Networks Journal, 2004.

[22] H. Chang, S. Jamin, W. Willinger, *Inferring AS-level Internet topology from router-level path traces*, in Proceedings of SPIE ITCom, 2001.

[23] H. Chang, S. Jamin, W. Willinger, *Internet connectivity at the AS-level: an optimization-driven modeling approach*, in Proceedings of ACM SIGCOMM MoMeTools workshop, 2003.

[24] H. Chang, S. Jamin, W. Willinger, *To peer or not to peer: modeling the evolution of the Internet's AS-level topology*, in Proceedings of IEEE INFOCOM, 2006.

[25] *CIDR report*, August 2007, http://www.cidr-report.org/.

[26] J. A. Cobb, M. G. Gouda, R. Musunuri, A Stabilizing Solution to the Stable Paths Problem, in Proceedings of Symposium of Self-Stabilizing Systems, Springer-Verlag LNCS, vol. 2704, pp. 169-183, 2003.

[27] J. A. Cobb, R. Musunuri, *Convergence of Interdomain Routing*, in Proceedings of IEEE GLOBECOM, pp. 1353-1358, 2004.

[28] R. Cohen, D. Raz, *The internet dark matter - on the missing links in the as connectivity map*, in Proceedings of IEEE INFOCOM, 2006.

[29] R. Cohen, D. Raz, *Acyclic Type of Relationships Between Autonomous Systems*, in Proceedings of the IEEE INFOCOM 2007, pp. 1334-1342, 2007.

[30] E. Crawley, R. Nair, B. Rajagopalan, H. Sandick, *A Framework for QoS-based Routing in the Internet*, IETF RFC 2386, August 1998.

[31] G. Cristallo, C. Jacquenet, *An Approach to Inter-domain Traffic Engineering*, in Proceedings of the XVIII World Telecommunications Congress (WTC2002), France, September 2002.

[32] *DE-CIX*, Frankfurt am Main (Germany), https://www.de-cix.net/.

[33] G. Di Battista, T. Erlebach, A. Hall, M. Patrignani, M. Pizzonia, T. Schank, *Computing the Types of the Relationships Between Autonomous Systems*, IEEE/ACM Transactions on Networking, vol. 15, no. 2, April 2007.

[34] X. Dimitropoulos, G. Riley, *Modeling Autonomous-System Relationships*, in Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation (PADS'06), 2006.

[35] M. Doar, *A better model for generating test networks*, in IEEE GLOBECOM, 1996.

[36] J.J. Dujmovic, *A Method for Evaluation and Selection of Complex Hardware and Software System*, San Francisco: Department of Computer Science, San Francisco State University, 1996.

[37] D. Estrin, *Policy requirements for inter administrative domain routing*, IETF, RFC 1125, November 1989.

[38] *Equinix Exchange (United States, Europe, Asia - Pacific)*, http://ix.equinix.com/peeringstats/.

[39] *European Internet exchange association*, http://www.euro-ix.net.

[40] M. Faloutsos, P. Faloutsos, C. Faloutsos, *On power-law relationships of the internet topology*, in Proceedings of ACM SIGCOMM, 1999.

[41] N. Feamster, R. Johari, H. Balakrishnan, *Implications of Autonomy for the Expressiveness of Policy Routing*, IEEE/ACM transactions on networking, vol. 15, no. 6, December 2007.

[42] A. Feldmann, J. Rexford, *IP network configuration for intradomain traffic engineering*, IEEE Network, Sept./Oct. 2001, pp. 46-57, 2001.

[43] A. Feldmann, O. Maennel, Z.M. Mao, A. Berger, B. Maggs, *Locating internet routing instabilities*, in Proceedings of ACM SIGCOMM, 2004.

[44] G. Fenu, *Livello Rete: Principi ed Architetture*, First Edition, CUEC, University Press Informatica, January 2005.

[45] G.Fenu, M.Picconi, S. Surcis, *XESS - Extended E-learning Support System*, In Proceedings of NDT 2009, The First International Conference on "Networked Digital Technologies", technically co-sponsored by IEEE Communication Society, VSB-Technical University of Ostrava, Czech Republic, 28-31 July 2009, pp. 165-170, 2009.

[46] G.Fenu, M.Picconi, S. Surcis, *XESS, Sistema di Supporto alle Decisioni per la valutazione di soluzioni e-Learning*, in Proceedings of DIDAMATICA 2009, University of Trento, Faculty of Economics, 2009.

[47] G.Fenu, M.Picconi, S. Surcis, *Metodiche progettuali per Personal Learning Environments in ambiente Grid e Cloud Computing*, in Proceedings of DIDAMATICA 2009, University of Trento, Faculty of Economics, 2009.

[48] G.Fenu, M.Picconi, *A Cost-Benefit Comparison of E-Learning Solutions*, Article in International Journal Of Information Studies Volume 2 Issue 1 January 2010, 2010.

[49] G.Fenu, M.Picconi, *Identification of the Variables in an E-Learning Platform Using a Cost-Benefit Analysis and an Automatic Decision-Making Tool*, in Collection of A. Respicio, et al. (Eds.) Bridging the Socio-technical Gap in Decision Support Systems. Challenges for the Next Decade. Frontiers in Artificial Intelligence and Applications. IOS Press, pp. 473-484, 2010

[50] G.Fenu, M.Picconi, *An Optimized Cost-Benefit Analysis for the Evaluation in E-Learning Services*, F. Zavoral et al. (Eds.), Networked Digital Technologies, Second International Conference, NDT 2010, Prague, Czech Republic, July 7-9, 2010. Proceedings, Part II ,Springer, CCIS 88, pp. 215-225, 2010.

[51] *Future INternet Design (FIND)*,
http://www.nets-find.net/.

[52] Future Internet, *The Future Networked Society: A white paper from the EIFFEL Think-Tank*,
http://future-internet.eu/.

[53] L. Gao, *On inferring automonous system relationships in the Internet*, IEEE/ACM Trans. Networking, vol. 9, no. 6, December 2001.

[54] R. Gao, C. Dovrolis, E.W. Zegura, *Avoiding Oscillations due to Intelligent Route Control Systems*, in Proceedings of INFOCOM 2006, Barcelona, Spain, April 2006.

[55] L. Gao, T.G. Griffin, J. Rexford, *Inherently safe backup routing with BGP*, in Proceedings of IEEE INFOCOM, Apr. 2001, pp. 547-556, 2001.

[56] L. Gao, J. Rexford, *Stable Internet Routing without Global Coordination*, IEEE/ACM Trans. Net., vol. 9, no. 6, pp. 681-692, 2001.

[57] L. Gao, F. Wang. *The extent of AS path inflation by routing policies*, in IEEE Global Internet Symposium, 2002.

[58] Z. Ge, D. Figueiredo, S. Jaiwal, L. Gao, *On the hierarchical structure of the logical Internet graph*, in Proceedings of SPIE ITCOM, August 2001.

[59] *Geant2 looking glass*, http://stats.geant2.net/lg/.

[60] *Global Environment for Network Innovations (GENI)*, http://www.nets-find.net.

[61] D.K. Goldenberg, L. Qiu, H. Xie, Y. R. Yang, Y. Zhang, Optimizing cost and performance for multihoming, in Proceedings of ACM SIGCOMM, August 2004.

[62] *Good practices in Internet exchange points*, http://www.pch.net/resources/papers/ix-documentation-bcp/ix-documentationbcp-v14en.pdf.

[63] R. Govindan, C. Alaettinoglu, K. Varadhan, D. Estrin, *Route servers for inter-domain routing*, Computer Networks and ISDN Systems, vol. 30, pp. 1157-1174, 1998.

[64] R. Govindan, C. Alaettinoglu, G. Eddy, D. Kessens, S. Kumar, W. Lee, *An Architecture for Stable, Analyzable Internet Routing*, IEEE Network, vol. 13, no. 1, Jan./Feb. 1999, pp. 29-35, 1999.

[65] R. Govidan, A. Reddy, *An analysis of internet inter-domain topology and route stability*, in Proceedings of IEEE lnfocom' 97, Kobe, Japan, April 1997.

[66] R. Govindan, H. Tangmunarunkit, *Heuristics for Internet map discovery*, in Proceedings of IEEE INFOCOM, vol. 3, Mar. 2000, pp. 1371-1380, 2000.

[67] T. Griffin, B. Presmore, *An Experimental Analysis of BGP convergence time*, in Proceedings of IEEE ICNP, November 2001.

[68] T.G. Griffin, F.B. Shepherd, G. Wilfong, *Policy Disputes in Path Vector Protocols*, in Proceedings of IEEE ICNP, pp. 21-30, 1999.

[69] T.G. Griffin, F.B. Shepherd, G. Wilfong, *A Safe Path Vector Protocol*, in Proceedings of INFOCOM 2000, pp. 490-499, 2000.

[70] T.G. Griffin, F.B. Shepherd, G. Wilfong, *The Stable Paths Problem and Interdomain Routing*, in IEEE/ACM Transactions on Networking, Volume 10, Issue 2, April 2002, pp 232-243, 2002.

[71] T.G. Griffin, G.T. Wilfong, *An analysis of BGP convergence properties*, in Proceedings of SIGCOMM, Cambridge, MA, August 1999, pp. 277-288, 1999.

[72] F. Guo, J. Chen, W. Li, C., T. Chiueh, *Experiences in Building a Multihoming Load Balancing System*, INFOCOM 2004, Hong Kong, China, March 2004.

[73] S. Halabi, *Internet Routing Architectures*, 2nd Edition, Cisco Press, ISBN 1587054353.

[74] Y. He, G. Siganos, M. Faloutsos, S. V. Krishnamurthy, *A systematic framework for unearthing the missing links: measurements and impact*, in Proceedings of NSDI, 2007.

[75] C. Hendrick, *Routing information protocol*, IETF, RFC 1058, 1988.

[76] Y. Hyun, A. Broido, K.C. Claffy, *On third-party addresses in traceroute paths*, in Proceedings of Passive and Active Measurement Workshop (PAM), 2003.

[77] G. Huston, *Interconnection, peering, and settlements*, Proceedings of INET, June 1999.

[78] *IRR - Internet Routing Registry*, http://www.irr.net/.

[79] *IS-IS. Intermediate System-to-Intermediate System*, http://www.ietf.org/html.charters/isis-charter.html.

[80] *Internet eXchange Points. List of IXPs by size*, http://en.wikipedia.org/wiki/List_of_Internet_Exchange_Points_by_-size.

[81] A.D. Jaggard, V. Ramachandran, *Robustness of class-based path vector systems*, in IEEE International Conference of Network Protocols (ICNP), Berlin, Germany, Oct. 2004, pp. 84-93, 2004.

[82] J. Karlin, S. Forrest, J. Rexford, Pretty good bgp: Protecting bgp by cautiously selecting routes. Technical Report TR-CS-2005-37, University of New Mexico, October 2005.

[83] H. Kellerer, U. Pferschy, D. Pisinger, *Knapsack Problems* Springer, Berlin, Germany, 2004.

[84] S. Kent, C. Lynn, K. Seo, Secure Border Gateway Protocol. IEEE Journal of Selected Areas in Communications, 18(4), 2000.

[85] S. B. Kodeswaran, A. Joshi, *Content and context aware networking using semantic tagging*, in Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06). Washington, DC, USA: IEEE Computer Society, p. 77, 2000.

[86] S. B. Kodeswaran, O. Ratsimor, A. Joshi, F. Perich, *Utilizing Semantic Tags for Policy Based Networking*, in Proceedings of Globecom 2007, 2007.

[87] S. Kosub, M.G. Maab, H. Taubig, *Acyclic type-of-relationship problems on the internet*, in Proceedings of the 3rd Workshop on Combinatorial and Algorithmic Aspects of Networking (CAAN'2006), pp. 98-111, 2006.

[88] C. Labovitz, A. Ahuja, F. Jahanian, *Experimental study of Internet stability and wide-area network failures*, in Proceedings of International Symposium of Fault-Tolerant Computing, June 1999, pp. 278-285, 1999.

[89] C. Labovitz, A. Ahuja, A. Bose, F. Jahanian, *Delayed Internet routing convergence*, in Proceedings of ACM SIGCOMM, 2000.

[90] C. Labovitz, A. Ahuja, R. Wattenhofer, S. Venkatachary, *The impact of Internet policy and topology on delayed routing convergence*, in Proceedings of IEEE INFOCOM, April 2001.

[91] C. Labovitz, G. Malan, F. Jahanian, *Internet Routing Instability*, in Proceedings of ACM SIGCOMM, September 1997.

[92] C. Labovitz, R. Malan, F. Jahanian *Origins of Internet routing instability*, in Proceedings of IEEE INFOCOM, 1999.

[93] M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, L. Zhang, *PHAS: A prefix hijack alert system*, in 15th USENIX Security Symposium, 2006.

[94] M. Lad, R. Oliveira, B. Zhang, L. Zhang, *Understanding the impact of BGP prefix hijacks*, ACM SIGCOMM Poster, 2006.

[95] L. Li, D. Alderson, W. Willinger, J, Doyle, *A first-principles approach to understanding the Internet's router-level topology*, in Proceedings of ACM SIGCOMM, 2004.

[96] L. Li, C. Chen, *Exploring Possible Strategies for Competitions between Autonomous Systems*, in Proceedings of IEEE International Conference on Communications, ICC 2008, Beijing, China, 19-23 May 2008. IEEE, 2008.

[97] LINX (London, United Kingdom), http://www.linx.net/.

[98] K. Lougheed, Y. Rekhter, *Border Gateway Protocol*, IETF, RRFC 1105, June 1989.

[99] H. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, A. Venkataramani. *iPlane: an information plane for distributed services*, in Proceedings of OSDI, 2006.

[100] P. Mahadevan, D. Krioukov, K. Fall, A, Vahdat *Systematic topology analysis and generation using degree correlations*, in Proceedings of ACM SIGCOMM, 2006.

[101] D.Magoni, J.J. Pansiot, *Analysis of the Autonomous System Network Topology*, ACM SIGCOMM - Computer Communication Review, Volume 31, Issue 3, pp. 26-37, 2001.

[102] Z.M. Mao, R. Bush, T. Griffin, M. Roughan, *BGP beacons*, in ACM SIGCOMM Internet Measurement Conference (IMC), 2003.

[103] Z.M. Mao, J. Rexford, J. Wang, R.H. Katz, *Towards an accurate AS-level traceroute tool*, in Proceedings of ACM SIGCOMM, 2003.

[104] G. Mavrotas, D. Diakoulaki, *A branch and bound algorithm for mixed zero-one multiple objective linear programming*, European Journal of Operational Research, N. 107: pp. 530-541, 1998.

[105] D. Meyer, J. Schmitz, C. Orange, M. Prior, C. Alaettinoglu, *Using RPSL in Practice*, IETF RFC 2650, August 1999.

[106] D. L. McGuinness, F. van Harmelen, OWL web ontology language overview, W3C Recommendation 10 February 2004, Tech. Rep., 2004. http://www.w3.org/TR/owl-features/.

[107] D. McPherson, V. Gill, D. Walton, A. Retana. *Border Gateway Protocol (BGP) Persistent Route Oscillation Condition*, IETF, RFC 3345, August 2002.

[108] D. Mills, *Exterior Gateway Protocol Formal Specification*, IETF, RFC 904, April 1984.

[109] P. Mirchandrani, R. Francis, *Discrete Location Theory*, John Wiley and Sons, 1990.

[110] P. Morrissey, *Mapping Out the Best Route*, Network Computing - Manhasset NY - Vol. 14:(25), pp. 47-55, 2003.

[111] M. Morrow, V. Sharma, T. D. Nadeau, L. Andersson, *Challenges in Enabling Interprovider Service Quality in the Internet*, IEEE Communications Magazine, vol. 43, no. 6, June 2005.

[112] MSK-IX (Moscow, Russia) http://www.msk-ix.ru/.

[113] *Multi-Lateral Peering Agreement (MPLA)*, http://www.openpeering.nl/mlpa.pdf.

[114] *NL-ix (Amsterdam, Netherlands)*, http://www.nl-ix.net/.

[115] *OpenPeering*, http://www.openpeering.nl/prices.shtml.

[116] *OSPF Version 2*, IETF, RFC 1583, http://www.isi.edu/in-notes/rfc1583.txt.

[117] C. Palmer, G. Steffan, *Generating network topologies that obey power laws*, In GLOBECOM, November 2000.

[118] *Packet clearing house IXP directory* , http://www.pch.net/ixpdir/Main.pl.

[119] D. Pei, M. Azuma, D. Massey, L. Zhang, *BGP-RCN: improving BGP convergence through root cause notification*, Computer Networks, Volume 48, Issue 2, pp 175-194, 2005.

[120] *PeeringDB* , http://www.peeringdb.com/.

[121] S. Qiu, F. Monrose, A. Terzis, P. McDaniel, *Efficient techniques for detecting false origin advertisements in inter-domain routing*, In Second workshop on Secure Network Protocols (NPSec), 2006.

[122] D. Raz, R.Cohen, *The Internet dark matter: on the missing links in the AS connectivity map.* In Proceedings of IEEE INFOCOM, 2006.

[123] Y. Rekhter, T. Li, *A Border Gateway Protocol 4 (BGP-4)*, IETF, RFC 1771, March 1995.

[124] Y. Rekhter, P. Gross, *Application of the Border Gateway Protocol in the Internet*, IETF, RFC 1772, March 1995.

[125] Y. Rekhter, T. Li, *An Architecture for IP Address Allocation with CIDR*, IETF, RFC 1518, September 1993.

[126] Y. Rekhter, T. Li, *A Border Gateway Protocol 4 (BGP-4)*, IETF, RFC 4271, January 2006.

[127] J. Rexford, J. Wang, Z. Xiao, Y.Zhang, *BGP routing stability of popular destinations*, In ACM SIGCOMM Internet Measurement Workshop (IMW), 2002.

[128] *RIPE routing information service project*, http://www.ripe.net/.

[129] *RIR - Regional Internet Registry data*, ftp://www.ripe.net/pub/stats.

[130] *RouteViews routing table archive*, http://www.routeviews.org/.

[131] Y. Shavitt, E. Shir, *DIMES: Let the Internet measure itself*, ACM SIGCOMM Computer Comm. Review (CCR), 2005.

[132] R. Siamwalla, R. Sharma, S. Keshav, *Discovering Internet topology*, Cornell Univ., Ithaca, NY, Tech. Rep., May 1999.

[133] *Skitter AS adjacency list*, http://www.caida.org/tools/measurement/skitter/as_adjacencies.xml.

[134] S.W. Smith, M. Zhao, D. Nicol, *Aggregated path authentication for efficient bgp security*, In 12th ACM Conference on Computer and Communications Security (CCS), November 2005.

[135] N. Spring, R. Mahajan, T. Anderson, *Quantifying the causes of path inflation*, In ACM SIGCOMM, 2003.

[136] L. Subramanian, S. Agarwal, J. Rexford, R. Katz, *Characterizing the Internet Hierarchy from Multiple Vantage Points*, INFOCOM 2002,New York, NY, USA, June 2002.

[137] L. Subramanian, V. Roth, I. Stoica, S. Shenker, R. Katz, *Listen and whisper: Security mechanisms for bgp*, In Proceedings of ACM NDSI 2004, March 2004, 2004.

[138] H. Tangmunarunkit, R. Govindan, S. Shenker. *Internet path inflation due to policy routing*, In SPIE ITCom, 2001.

[139] H. Tangmunarunkit, R. Govindan, S. Shenker, D. Estrin, *The impact of routing policy on Internet paths*, In IEEE INFOCOM, 2001.

[140] A.S. Tanenbaum, *Computer Networks*, Third Edition Prentice Hall International, 2001.

[141] *Technical Standards and Policy for Subscribers to LAP & MAE-LA*, http://www.isi.edu/div7/mla/.Tech-Stds.html.

[142] *The Abilene Observatory Data Collections*, http://abilene.internet2.edu/observatory/data-collections.html.

[143] S. Uhlig, V. Magnin, O. Bonaventure, C. Rapier, L. Deri, *Implications of the Topological Properties of Internet Traffic on Traffic Engineering*, Proceedings of the 19th ACM Symposium on Applied Computing, Special Track on Computer Networks, Nicosia, Cyprus, March 2004, 2004.

[144] K. Varadhan, R. Govindan, D. Estrin, *Persistent route oscillations in inter-domain routing*, ISI technical report 96-631, USC/Information Sciences Institute, 1996.

[145] K. Varadhan, R. Govindan, D. Estrin, *Persistent route oscillations in inter-domain routing*, Computer Networks, vol. 32, no. 1, pp. 1-16, 2000.

[146] D. Walton, A. Retana, E. Chen, *Advertisement of Multiple Paths in BGP* Internet draft, draft-walton-bgp-add-paths-04.txt, work in progress, August 2005.

[147] L. Wang, X. Zhao, D. Pei, R. Bush, D. Massey, A. Mankin, S.F. Wu, L. Zhang, *Observation and analysis of BGP behavior under stress*, In ACM SIGCOMM Internet Measurement Workshop (IMW), 2002.

[148] X. Wang, D. Loguinov, *Wealth-based evolution model for the Internet ASlevel topology*, In Proceedings of IEEE INFOCOM, 2006.

[149] B.M. Waxman, *Routing of multipoint connections*, IEEE JSAC, 1988.

[150] L. Xiao, K. Lui, J. Wang, K. Nahrstedt, *QoS extensions to BGP*, ICNP2002, November 2002.

[151] Wen Xu, J. Rexford. *Miro: multi-path interdomain routing*, In SIGCOMM, pages 171-182, 2006.

[152] M. Yannuzzi, X. Masip-Bruin, O. Bonaventure, *Open Issues in Interdomain Routing: A Survey*, IEEE Network, Vol. 19, No. 6, November/December 2005, 2005.

[153] E. W. Zegura, K. Calvert, S. Bhattacharjee, *How to model an internetwork*, In IEEE INFOCOM, 1996.

[154] B. Zhang, R. Liu, D. Massey, L. Zhang. *Collecting the Internet AS-level topology*, ACM SIGCOMM Computer Comm. Review (CCR), 2005.

[155] C. Zheng, L, Ji, D. Pei et al, *A light-weight distributed Scheme for detecting IP prefix hijacks in real-time*, in Proceedings of ACM SIGCOMM, August 2007.