BGP

from theory to practice

Flavio **LUCIANI**
Antonio **PRADO**
Tiziano**TOFONI**

NAMEX
ROMA IXP

REISS
ROMOLI

CONTENTS

FOREWORD

Back in 2011, Reiss Romoli published the first edition of the book "*BGP: From Theory to Practice*", written by Tiziano Tofoni. Many years have gone by since then, and the author of the first edition deemed it necessary for it to undergo an extensive review, since, in the meantime, BGP – although quite consolidated – underwent its own evolution. New techniques significantly improved its security and convergence speed aspects – the two Achilles's heels of the first BGP versions.

Despite our different professional journeys, we all have a common denominator in BGP. According to our 'vision', BGP is the standard protocol without which the entire Internet would not be possible. And this has been proven over the years, since BGP has gained such consensus that it has become the most important protocol for IP networks – the true supporting structure of the "Internet ecosystem".

BGP is based on simple, yet effective concepts, which allow for an extremely flexible use of this protocol. Although it was born and designed as an inter-domain routing protocol, today BGP is broadly employed also in other fields, such as:

- In modern public Service Provider networks, where it plays a key role in the overall routing architecture, because – thanks to its proven scalability – it has turned out to be a very efficient tool also to distribute external routing information within the network.
- In MPLS services control plane;
- For "painless" IPv4 to IPv6 migration, without major impacts on the backbone of the Service Providers;
- As private network access protocol to Service Provider networks;
- As IGP in big Data Centers, where it acts as routing protocol on the underlay network, and as transfer for several kinds of information on the overlay network.

Rather than an actual routing protocol in the traditional sense, BGP is routing policy application protocol. Indeed, in its definition, the protocol designers did not focus on some of the typical aspects of standard routing protocols, such as convergence speed and security. Rather, they focused on making the exchange of large quantities of IP prefixes scalable – and they certainly succeed in doing so, if we consider that today, in the routers used by large IP networks, BGP can manage the exchange of routing information related to almost one million IP prefixes.

All this has driven us to follow BGP's evolution up close, and to spread its knowledge to a vast audience of insiders. This is how the idea of writing a second edition of the original book came about. Following the spirit of the first edition, this edition also pursued the goal of combining theory and practice, and tried not to be only a (debatable) presentation of the standard. This is why, apart from explaining in detail and with many examples how the protocol works and its role in IP networks and in the entire Internet ecosystem, the book also includes many practical application examples, resulting from many years of experience.

In the way it is conceived, the book requires solid notions on TCP/IP architecture, and on IP routing fundamentals in particular. Moreover, since it covers several configuration aspects, both

in Cisco (IOS/IOS-XE/IOS-XR) and in Juniper (JUNOS) environments, it also requires a basic knowledge of these Operating Systems. Nevertheless, we firmly believe that being knowledgeable about a specific Operating System is not so important, once the basic concept behind the protocol and its services have been acquired. Jumping from one technology to another is just a question of learning the basic commands, and understanding how the protocol was implemented by that specific developer, with this last aspect being crucial in machine TCL scenarios.
In general, this is an upper-intermediate level book, while the notions on BGP can be read both by readers with basic knowledge who wish to deepen its concepts, and by those with no understanding of this standard. It is addressed to the wide audience of Internetworking experts, both on the Service Provider network and on the private network side (see all institutions such as Banks, Industries, Public Administrations, many of which have Corporate Networks based on the IP/MPLS backbone).
We hope that reading it will help, apart from understanding the standard's theoretical-practical fundamentals, also to grasp the importance of an intensive use of BGP in IP networks.

Flavio Luciani
Antonio Prado
Tiziano Tofoni

PRESENTATION

I was born in 1963 and I've always been attracted to technology; telecommunications have always fascinated me. I've always been curious about stories.
One of my favourite stories about telecommunications was about the Reiss Romoli Graduate School.
It talked about a graduated school founded in 1976, under the initiative of the STET Group, and then passed on to Telecom Italia, devoted to post-graduate education of young minds who would be sent for months to L'Aquila, in a campus equipped with all amenities (labs, swimming pool, gym, library), and there they were trained to become the new ranks of engineers and executives of the former monopoly.
I met many of these former young people, trained at the Reiss Romoli, who, over the years, went on to cover strategic positions in Tim and in other Italian ISP, and I've listened to their many stories. Stories of a top-level school.
Stories of young people who studied hard, and who were also young people trying to enjoy that experience in the best way possible, and so they held "harmless breakouts" at night from this barrack-school that sometimes did not agree with their age.
Nice stories.
It was rumoured they had very good, and very passionate professors.
To me, the school and its professors were sort of legends, because I never met them. L'Aquila is a city I'm familiar with. I lived there for a while, when I was working at the Physics Labs underneath the Gran Sasso mountains. For some reason, I never visited the campus of the Reiss Romoli School in L'Aquila.
Then, one day, Flavio Luciani, our CTO at Namex, talked to me about meeting one of these mythical professors, Tiziano Tofoni, and told me about the possibility of working with him.
During an ITNOG event in Bologna, I met Tiziano and "*BGP: From Theory to Practice*", the book he wrote and published in 2011, and discovered it was a stable book in our community. Many of the technicians and engineers employed in Italian ISPs were formed by that book.
During the conference breaks, we came up with the idea that saw us collaborate all these years – train the employees of Namex-partner ISPs, through the great experience of the Reiss Romoli School.
The matter was that many ISPs connecting to Namex needed to train their newly hires and hold update courses for existing staff. Finding these courses on the market was certainly no easy task.
There weren't many companies offering training on such specific topics, like the one ISPs are interested in (BGP, MPLS, DNS, IPv6, etc.), on the market; and even less of them were able to offer a quality equalling that of the Reiss Romoli School. What's more, the cost was very high, especially for smaller ISPs.
There were other lunches after that event. I had the chance to see the old campus at the Reiss Romoli School, even if only from the outside, since it is no longer open.
We decided to found the Namex *School Of Advanced Networking*, with the motto "Training Course for ISPs made by ISPs".

It was 2019.
The first SOAN catalogue started with the classes that were part of the Reiss Romoli catalogue (3 days, and, in some cases, exam + final certification), which we decided to enrich with contributions/workshops by our CTO Flavio Luciani and by experts/friends from Namex-member ISPs, such as Antonio Prado – a benchmark of the Italian ISP and PA community, whom I met many years earlier, when he was working for one of Namex-member ISPs.
We decided to offer the classes free of charge, using part of the revenue from the services offered by Namex to the ISPs. Tiziano's book on BGP was the reference book of the most popular class – that on BGP.
It was a success. Since then, Namex has provided, in all editions, over 50 classes, training hundreds of people, and – more importantly – it fostered a moment of aggregation and debate between the people that "deal with the Internet" in Italy.
It's something I'm especially proud of, and I hope it will go down in Namex history (small in absolute terms, yet so big for us living it).
The cherry on top is this new edition of the BGP book, wanted by "Admiral" Tiziano, with his First Officers Antonio and Flavio. Namex has enthusiastically joined the sponsorship request, right from the start, counting on the fact that it can continue to be a reference for all those professionals dealing with interdomain interconnection, and with the Internet ecosystem in general, for many years to come.
A big thank you to Antonio, Flavio and Tiziano, who worked on this new edition.
A thank you to the Reiss Romoli School, editor of the book, which has trained Italian telecommunication professionals for decades, keeping the quality level very high.
And lastly, a thank you to Namex-member ISPs, which, with their feedback, prompted us to start the Namex *School Of Advanced Networking* and to improve it, year after year.

Maurizio Goretti
Namex CEO

ACKNOWLEDGEMENTS

This book is to me a moment of professional and – above all – personal growth. And it wouldn't be so, without my two travel companions and friends, Tiziano Tofoni and Antonio Prado, whom I want to thank from the bottom of my heart. To my family, for their love and patience. To my father, who would be proud of me.

Flavio Luciani

I would love to thank hundreds of people, because I've learned something from each and every one of them, during my career. The first people I want to thank are my friends Flavio and Tiziano, with whom I shared this experience (and, I hope, many more to come), and then Mauro Angiolillo, my inseparable sparring partner, and Professor Fabio Fioravanti, for his precious advice. Lastly, my parents, for listening to me, my children who always bring me down to earth, and Belinda, my wife, who has always supported me.

Antonio Prado

During the creation of this book, I benefited from the help of many people, who I'm proud to call my Friends (with a capital F), and to whom I want to offer my deepest gratitude.
I also wish to thank the many Friends of Italian ISPs with whom I had interesting debates on the role of BGP and of its actual applications in the networks of ISPs.
Last but not least, as in every book I ever wrote, I want to thank the two women of the house, Vicky and Fiammix (Maria Vittoria and Fiammetta). Without their evening tiredness (which allowed me to focus on my work), and without their patience in bearing my constant absent-mindedness, this book probably would have never seen the light of day. I dedicate this work to them.

Tiziano Tofoni

The authors wish to thank Reiss Romoli srl, editor of the book, Namex CEO Maurizio Goretti, for the enthusiasm with which he joined the project, and Belinda Menzietti, for her patient and professional editing.

Last but certainly not least, a big thank you to Simone Morandini for the immense contribution provided in the revision of this book.

Those who fall for practice without science are like the helmsman who enters a ship without a rudder or compass who never has certainty where he is going. Always practice must be built on top of good theory.

(Leonardo da Vinci)

1 – INTRODUCTION

BGP (Border Gateway Protocol) was created as a standard EGP (Exterior Gateway Protocol) protocol, that is, it was developed to exchange routing information between different Autonomous Systems (ASes). The version currently used is number 4, defined in RFC 1771 – *A Border Gateway Protocol* 4 (*BGP*-4), March 1995, rewritten with the same title in January 2006 as RFC 4271. Its main characteristics are the following:

- it is a Path Vector routing protocol, which means it is conceptually similar to a Distance Vector protocol, although with hops measured in terms of numbers of ASes, instead of number of routers;
- it supports CIDR (Classless Inter Domain Routing);
- it determines optimal paths through a very complex selection process, based on metrics of different kinds;
- due to the presence of different types of metrics, it allows the creation of routing policies both for outbound and inbound traffic in the AS;
- it allows a reliable exchange of routing information, achieved through TCP connections;
- updates are event-driven.

All those features make BGP a routing policy application protocol, rather than an actual routing protocol. In fact, protocol designers did not consider some of the typical aspects present in IGP routing protocols when defining it, such as, for instance, speed of convergence, load balancing, etc. Instead, they focused on making the management of large quantities of IP prefixes scalable – and succeeded in doing so, if we consider that today, in routers installed in big IP networks, BGP is capable of managing the exchange of routing information related to hundreds of thousands of IP prefixes. In this regard, there's an interesting statement by Yakov Rekhter, who, together with Kirk Lougheed, may be considered the father of BGP:

"Kirk Lougheed and myself's goal was to build a routing protocol able to convey 1000 routes and not fall into pieces. If you think the total routes being today in the Internet, we pushed the envelope a bit."

The purpose of this chapter is explaining some of the key definitions to understand BGP, and, above all, define an operating model that will be constantly referenced in the next chapters.

1.1 HISTORICAL NOTES

In the early days of the Internet, what is today known as the "network of networks" was actually a single network – ARPANET, developed at the end of the 1960s – and its satellite extension – SATNET, developed in the mid-1970s. Routers – which were called gateways back then – exchanged routing information through a single Distance Vector protocol known as GGP (Gateway-to-Gateway Protocol), then evolved into RIP (Routing Information Protocol), which remained for many years the only Internet routing protocol.

As the number of users and nodes grew, it became evident that adopting a model without any kind of hierarchy was not scalable. A single routing protocol was not enough to manage the network's

complexity. Eric Rosen – who worked as Engineer at Bolt Beranek and Newman Inc. at the time – pointed out (in RFC 827 – *Exterior Gateway Protocol (EGP)*, October 1982) the flat model's scalability issues, and the need to adopt a hierarchical model, dividing the Internet into a set of ASes, i.e. networks managed by the same administration. One of the ASes – called Core AS – comprised ARPANET and SATNET, and worked as the Internet's Backbone. All the other ASes – called stub AS – were connected by one or more routers (Exterior Gateways) to the Core AS. Generally speaking, communication between stub ASes occurred through the Core AS. The exchange of routing information between ASes was delegated to a new protocol, standardized in RFC 827.
However, in a matter of years, due to the continuous growth of the Internet, EGP revealed many limitations, essentially linked to the fact that it had been designed based on the Internet Hub-and-Spoke model (all the stub ASes (Spoke) connected to a single Core AS (Hub)). Among them:

- the lack of loop avoidance mechanisms;
- the fact that only classful routing was supported (RFC 1817 – *CIDR and Classful routing*, August 1995);
- the fact that the routing information communication mechanism was based, as in Distance Vector protocols, on periodic transmission of the entire IP routing table to the nearest neighbors (the timeframe was set to 2 minutes);
- the impossibility to define inbound and/or outbound traffic management policies in an AS.

In January 1989, at the 12th IETF Meeting in Austin, Texas, Yakov Rekhter and Kirk Lougheed – Head Researcher at IBM's T.J. Watson Research Center the former, and Engineer at Cisco Systems the latter – sat down at a table (according to "The Packet" newsletter, Volume 1, Number 2, published in Winter of 1989 by Cisco Systems, Leonard Bosack, Cisco co-founder, was also with them) and laid the foundations for a new inter-AS routing protocol – Border Gateway Protocol (BGP) – jokingly called The Two-Napkin Protocol, because BGP's initial project was drawn on some restaurant napkins (which, some say, were even soiled with ketchup!).
Lougheed himself shed some light on this episode, by stating: *"After I wrote up the notes on two napkins, I made a second copy, also on napkins. I gave Yakov that first copy and took the second copy for myself. Apparently Yakov made photocopies of his napkins and these photocopies are the origin of the images you've seen. Having no sense of history, I discarded my napkins at some point."*
Photocopies of the drawings contained in those napkins are now displayed on the walls of the *Routing Protocol Development* department, in Cisco Systems headquarters in Santa Clara, CA. They are shown in Figure 1.1.
In a short while, the first practical implementation of BGP was created from this draft, and then the first standard, defined in RFC 1105 – *A Border Gateway Protocol* (*BGP*), June 1989, written by Yakov Rekhter and Kirk Lougheed.

NOTE: In Lougheed's words: *"Once back home I started drafting what eventually became RFC 1105. Yakov and I passed that document back and forth while developing and refining our own implementations in a classic iterative process. By the time RFC 1105 was published there were two implementations, my Cisco router implementation and Yakov's IBM router implementation on the NSFnet backbone. We naturally tested for interoperability. I think the gated implementation came out after RFC 1105."*

The standardization process involved several stages, which led to the definition of a second and third version, published in RFC 1163 of June 1990 and in RFC 1267 of October 1991, respectively. Version 4 – the final one – was published in March 1995, in RFC 1771 – *A Border Gateway Protocol* 4 (*BGP*-4), written by Y. Rekhter and T. Li, who worked for Cisco Systems. RFC 1771 was rewritten in January 2006 by the same authors, alongside S. Hares, in RFC 4271.
From Rekhter and Lougheed's napkins was born the most important protocol of today's Internet, the one that glues the "network of networks" together. BGP-4 is *de facto* the standard protocol, universally employed as inter-domain routing protocol. Over the years, it underwent continuous updates, which enhanced its functions, stability and scalability.
Despite being created as an inter-domain routing protocol, over time, BGP expanded its field of application, and today it is employed:

- in modern public networks of big ISPs (Internet Service Provider), where it plays a key role in the overall routing architecture, because – thanks to its proven scalability – it has turned out to be a very efficient tool also to distribute IP prefixes external to the AS inside an AS;
- in the control plane of VPN services based on the BGP/MPLS model;
- as an access protocol of private networks (Enterprise networks) to ISP networks.

Without fear of contradiction, we can say that BGP is the most important routing protocol for IP networks.

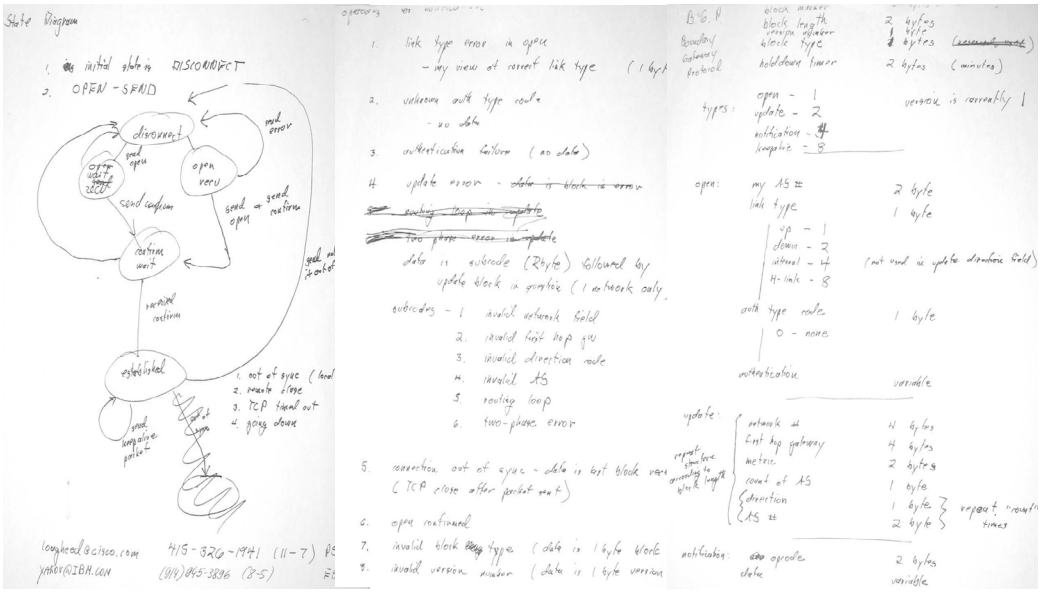
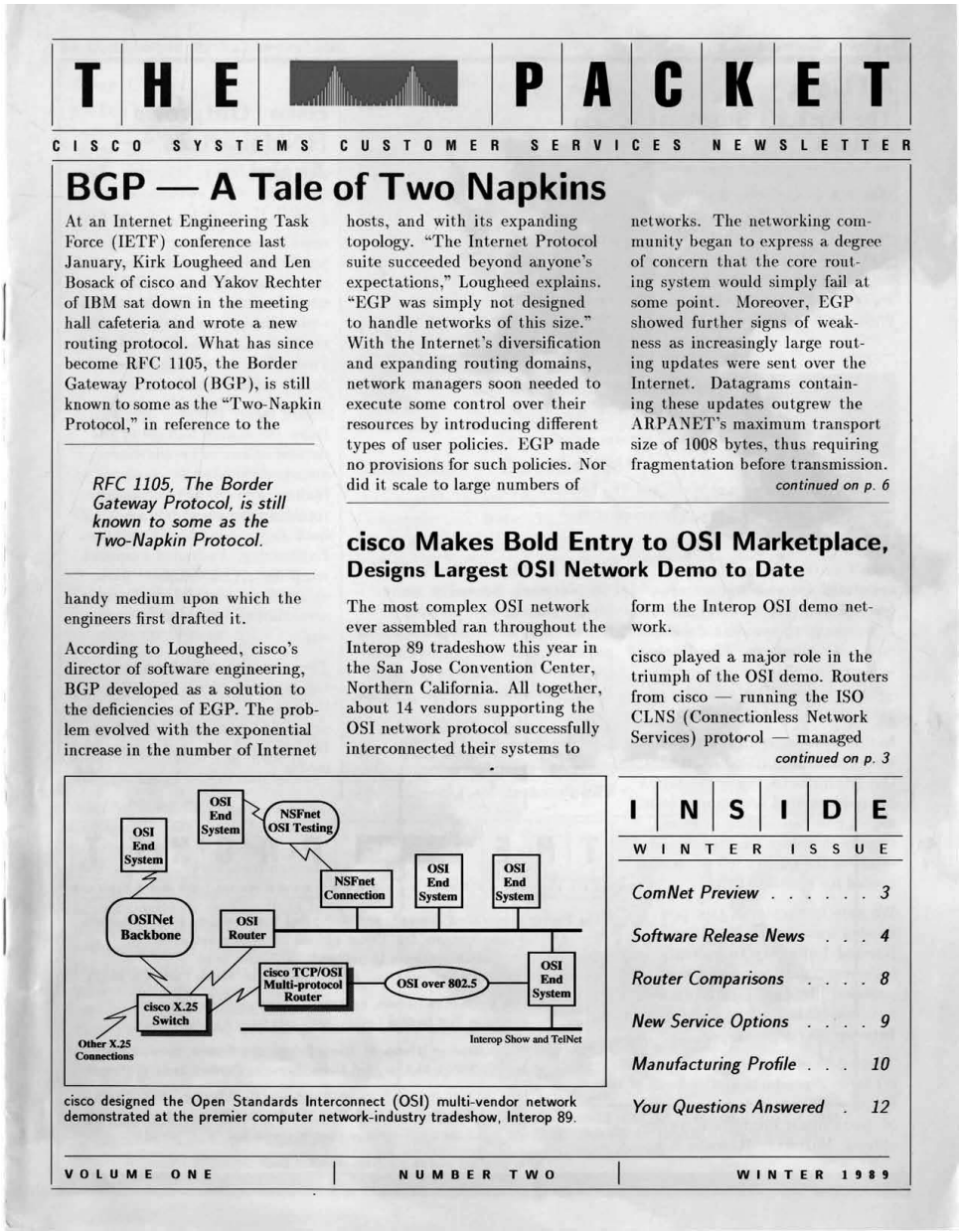


Figure 1.1 a – The *Two-Napkins Protocol*



THE PACKET

CISCO SYSTEMS CUSTOMER SERVICES NEWSLETTER

BGP — A Tale of Two Napkins

At an Internet Engineering Task Force (IETF) conference last January, Kirk Lougheed and Len Bosack of cisco and Yakov Rechter of IBM sat down in the meeting hall cafeteria and wrote a new routing protocol. What has since become RFC 1105, the Border Gateway Protocol (BGP), is still known to some as the "Two-Napkin Protocol," in reference to the handy medium upon which the engineers first drafted it.

RFC 1105, The Border Gateway Protocol, is still known to some as the Two-Napkin Protocol.

According to Lougheed, cisco's director of software engineering, BGP developed as a solution to the deficiencies of EGP. The problem evolved with the exponential increase in the number of Internet hosts, and with its expanding topology. "The Internet Protocol suite succeeded beyond anyone's expectations," Lougheed explains. "EGP was simply not designed to handle networks of this size." With the Internet's diversification and expanding routing domains, network managers soon needed to execute some control over their resources by introducing different types of user policies. EGP made no provisions for such policies. Nor did it scale to large numbers of networks. The networking community began to express a degree of concern that the core routing system would simply fail at some point. Moreover, EGP showed further signs of weakness as increasingly large routing updates were sent over the Internet. Datagrams containing these updates outgrew the ARPANET's maximum transport size of 1008 bytes, thus requiring fragmentation before transmission.

continued on p. 6

cisco Makes Bold Entry to OSI Marketplace, Designs Largest OSI Network Demo to Date

The most complex OSI network ever assembled ran throughout the Interop 89 tradeshow this year in the San Jose Convention Center, Northern California. All together, about 14 vendors supporting the OSI network protocol successfully interconnected their systems to form the Interop OSI demo network.

cisco played a major role in the triumph of the OSI demo. Routers from cisco — running the ISO CLNS (Connectionless Network Services) protocol — managed

continued on p. 3

cisco designed the Open Standards Interconnect (OSI) multi-vendor network demonstrated at the premier computer network-industry tradeshow, Interop 89.

INSIDE

WINTER ISSUE

VOLUME ONE | NUMBER TWO | WINTER 1989

Figure 1.1 b – *The Packet*.

Through the eyes of a modern provider, we must admit that BGP, despite its best intents, was born with an original sin: it assumes that all Internet networks are reliable secure.
The fact that it has been created before security (in a broad sense) became an issue, has marked its development since the 1990s. This aspect – which we will explore in depth in Chapter 10, when we talk about security – can be easily summarized in Internet expert Randy Bush's scathing yet spot-on line: *"You're in Hackerville here on the Internet. Period. All of this stuff lacks formal discipline. It's paint and spackle"*.

1.2 DEFINING AN AUTONOMOUS SYSTEM

An Autonomous System (AS) is a set of routers managed by a single entity which usually (but not necessarily!) employs a single, internal IGP (Interior Gateway Protocol).
From a technical standpoint, we can find its definition in RFC 1930 – *Guidelines for creation, selection, and registration of an Autonomous System (AS)*, March 1996, that reads:
"*An autonomous system is a group of one or more IP prefixes, managed by one or more network providers, with a UNIQUE and WELL-DEFINED routing policy.*"

NOTE: IGP are routing protocols used within an AS. The most used protocols in today's enterprise and ISP networks are OSPF and IS-IS Link State protocols. IGPs now considered obsolete are RIP and EIGRP; the latter was a Cisco proprietary protocol, subsequently standardized (RFC 7868).

From the outside world, an AS is seen as a single entity identified by a number, coded at 16 or 32 bits, and assigned by five RIRs (Regional Internet Registries): RIPE (Europe, Western Asia and former URSS), APNIC (Asia-Pacific Area: Central Asia, South-east Asia, Indo-China, Oceania), ARIN (North America, Atlantic Islands), LACNIC (Central-South America, Caribbean), AfriNIC (Africa).
The exchange of routing information between ASes occurs through protocols from the EGP (Exterior Gateway Protocol) family, which today, in practice, consists only of BGP. Figure 1.2 below shows the relationship between IGP, EGP and ASes.

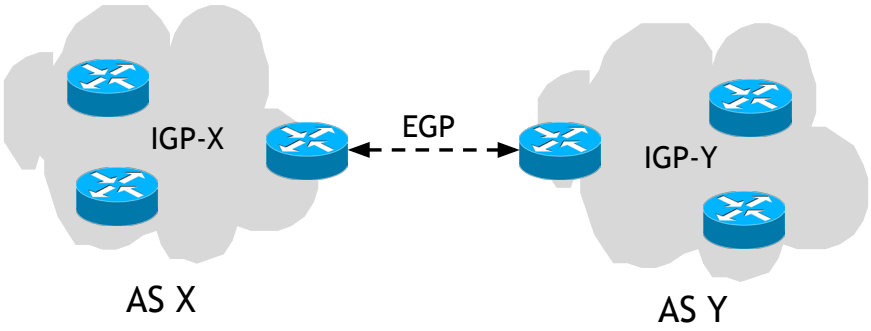


Figure 1.2 – Relationship between IGP and EGP and ASes.

Until 2007, the AS numbers available were only those taken from a 16-bit space that fell within the 0 - 65535 interval. However, only those between 1 and 64511 could be publicly assigned (and not every one of them, see the next note). Values between 64512 and 65534 cannot be assigned to public ASes (i.e. directly on the Internet), and are reserved for private use. The last value, 65535, is reserved (RFC 7300 – *Reservation of Last Autonomous System (AS) Numbers*, July 2014). Generally speaking, they are assigned by ISPs to customers that use BGP as an access protocol to their network. Value AS=0 is reserved to certain BGP security aspects (see RFC 7607 – *Codification of AS 0 Processing*, August 2015), covered in Chapter 10.

NOTE: Values between 64496 and 64511 cannot be assigned to a public AS; they are reserved to documentation (RFC 5398), and will be used extensively in this textbook. Number 112 is also unavailable (RFC 7534 – *AS112 Nameserver Operations*, May 2015), as it has been destined to the special purpose of hosting anycast instances for authoritative DNS servers for reverse resolutions of IPv4 and IPv6 address spaces that cannot be routed on the Internet (for instance, those described in RFC 1918 – *Address Allocation for Private Internets*, February 1996, but not

only those). AS number 23456 (better known as AS_TRANS) cannot be freely assigned, because it is used to facilitate communications between a router that doesn't support the 32-bit AS notation, and one that uses a 32-bit AS (RFC 6793 – *BGP Support for Four-Octet Autonomous System (AS) Number Space*, December 2012). More details on this in Annex A.1.

The limited availability of public AS numbers led to a 32-bit expansion of the AS number, standardized in RFC 4893 – *BGP Support for Four-octet AS Number Space*, May 2007.

NOTE: AS representation was done through a 16-bit number, and then through a 32-bit number (e.g., the AS number 65551 can be represented as 65551 in the asplain format or as 1.15 in the asdot+ format), as regulated by RFC 5396 – *Textual Representation of Autonomous System (AS) Numbers,* December 2008. For further details, see Annex A.1.

Even in the 32-bit space, ASes have special use reserved numbers. Indeed, the 65536-65551 interval is reserved to documentation, see RFC 5398 – *Autonomous System (AS) Number Reservation for Documentation Use*, December 2008, the 4200000000-4294967294 interval (approx. 95 million ASes) is for private use, see RFC 6996 – *Autonomous System (AS) Reservation for Private Use*, July 2013, and the last, number 4294967295, is reserved to possible future uses, see RFC 7300 – *Reservation of Last Autonomous System (AS) Numbers*, July 2014.
Based on their outgoing connection and on how transit traffic is processed, ASes can be classified into:

- single-homed ASes: characterized by a single (stub AS) or redundant connection toward only one other AS;
- multi-homed ASes: characterized by more than one connection toward several ASes.

NOTE: In the literature, definitions are not always concordant. Indeed, very often, redundant connectivity toward a single AS is also defined as multi-homed. For the sake of clarity, in this textbook, we prefer to use the term home for an AS, hence our classification .

1.2.1 Single-homed AS

A single-homed AS is characterized by a single connection, or, as it generally occurs in practice, a redundant connection toward another AS. In case of single connection, we talk about stub AS.
A typical example of stub AS is a private network AS, or a small ISP connecting to the network of a larger ISP, through a single connection. A stub AS with a single connection toward the ISP does not need to know all the Internet prefixes. In fact, with a single connection pointing outside (see Figure 1.3), reachability of the prefixes outside the AS can be guaranteed through a simple default route on the access router connected to the ISP's network.
Therefore, in theory, a stub AS doesn't need to use BGP on its access router to exchange routing information with the ISP to which it is connected. Some stub ASes use BGP anyway as access protocol, even in similar situations, to compensate, for instance, for Level 2 network deficiencies (e.g. access via an Ethernet network, where convergence may be slow).

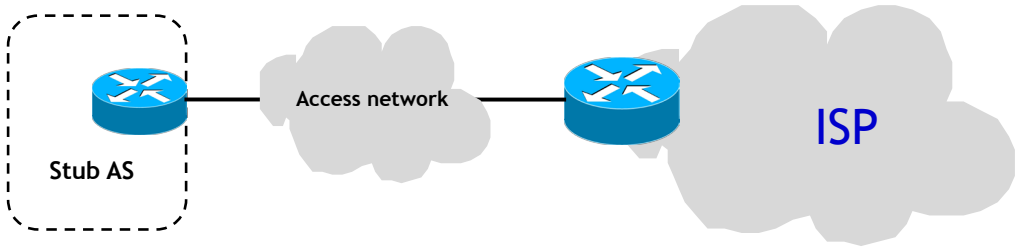


Figure 1.3 – Single-homed AS with single connection (stub AS).

In case of redundant connections (fault-tolerant), in order to optimize the stub AS inbound/outbound traffic, it is convenient to use BGP. We will go over those aspects in Chapter 7.

1.2.2 Multi-homed AS

As we were saying, typical examples of a multi-homed AS include private network ASes, or small ISPs, that connect to the network of two or more public ISPs for reliability reasons, or ISPs that exchange IP routing information with more than one AS (see Figure 1.4). The exchange of routing information between ASes occurs through BGP, or rather, as we will see in the next chapter, through BGP sessions.

We can identify two types of multi-homed AS:

- multi-homed Transit ASes: they allow exchange of traffic between different ASes, using their own resources as transit;
- multi-homed Non-Transit ASes: they do not allow external traffic to transit on the AS.

NOTE: For an AS, transit traffic is defined as the set of IP packets with both source and destination addresses that do not belong to any of the IP subnets used by the AS.

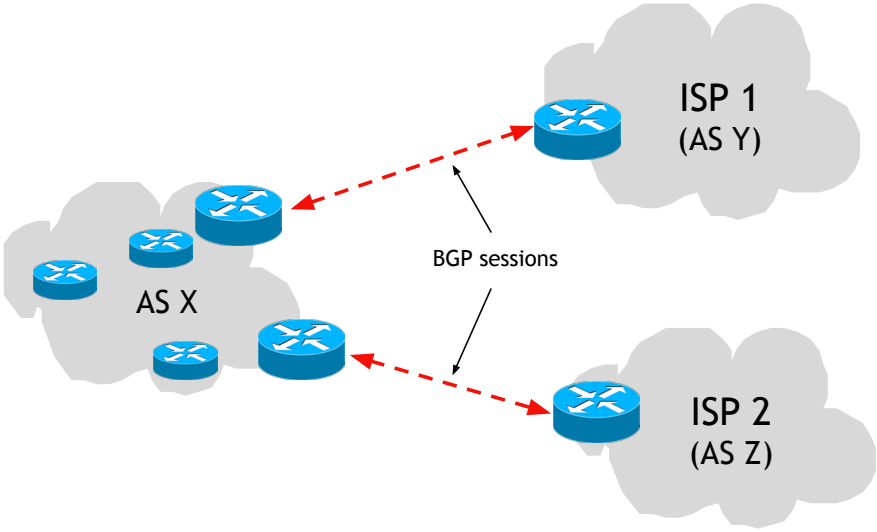


Figure 1.4 – Multi-homed AS.

A multi-homed AS becomes a transit AS when it propagates BGP advertisements of the prefixes received by other ASes. For instance, in Figure 1.5, the three ASes X, Y and Z have prefixes (P1, P2), (P3, P4) and (P5, P6), respectively. AS X receives information on how to reach prefixes (P3, P4) from AS Y, through BGP advertisements. If AS X propagates the information received from AS Y toward AS Z, AS X automatically becomes a transit AS for IP traffic from AS Z toward AS Y's prefixes (P3, P4). In the same way, if the prefixes (P5, P6) AS X receives from AS Z were propagated toward AS Y, AS X would automatically become a transit AS for IP traffic from AS Y toward AS Z's prefixes (P5, P6).
In order to prevent a multi-homed AS to become a transit AS, it is sufficient that it does not propagate the advertisement from other ASes. In particular, a multi-homed non-transit AS only announces its own prefixes outbound. For instance, in Figure 1.5, AS X, to prevent becoming a transit AS for traffic exchanged by AS Y and Z, should only announce its own prefixes, and avoid propagating (as in the example above) the prefixes received from AS Y and Z.

NOTE: It is worth mentioning that possible BGP configuration errors on the network of a multi-homed AS could cause unpleasant situations. Indeed, if an AS involuntarily acts as a transit AS for other units, it would mean that it is surrendering part of its Internet access bandwidth. On the other hand, if an AS voluntarily exploits the configuration error of another AS, it would engage in unethical behavior. We will go back to this in Chapter 10.

In multi-homed ASes, it is a good practice to use BGP for its loop prevention and routing policy definition properties. Those aspects will be treated further in Chapter 7.

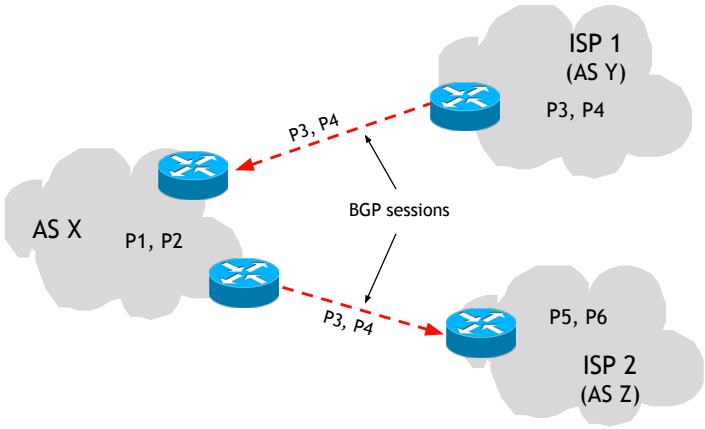


Figure 1.5 – Multi-homed Transit AS.

1.3 BGP AND THE INTERNET ECOSYSTEM

The entire Internet can be modeled as a flow graph, where nodes consist of ASes, and connections between nodes consist of BGP sessions (Figure 1.6). BGP sessions are logical connections between routers, on which routing information is exchanged. This information, suitably propagated between all ASes, allows reaching all system devices (hosts), and therefore the entire great ocean of information on the Internet.
Routing information comprises pairs of the following kind: <IP prefix, prefix length>. For the sake of simplicity, in this textbook we will refer to those pairs simply as IP prefixes, using the classic notation "IP prefix/IP prefix length" (e.g. 203.0.113/24). Every AS injects a set of IP prefixes – that is, IP address blocks generally assigned by a RIR to the AS administrator (or at least, that's how it should be) – into the system.

NOTE: Observing the hierarchy when managing numerical resources (hierarchy that from IANA proceeds to RIRs, and from those to NIR/LIR) is essential for Internet operation. In fact, it is worth remembering that, failing to observe the hierarchy can cause service interruptions in particular areas of the Internet, at best, and, in worst case scenarios, it can unleash illicit behavior that constitutes punishable crimes.

Logically announced prefixes within an AS are automatically propagated (unless routers are instructed to behave otherwise) on the different BGP sessions, following the rules described in Paragraph 2.1. Propagation can be seen as a selective flooding mechanism, through which IP prefixes are spread to all ASes of the Internet system.

In order for the system to work, it is not necessary to spread all IP prefixes to all the routers on the Internet. The type of prefixes to spread and their recipients basically depend on the AS topology and function. For instance, as mentioned earlier in Paragraph 1.2., it is not necessary to spread all the IP prefixes of the Internet world to the routers of a stub AS. Because of the way the stub AS is connected to the AS graph, it just requires a default route that allows it to reach one AS, which in turn is capable of reaching all the Internet prefixes, through a given path.

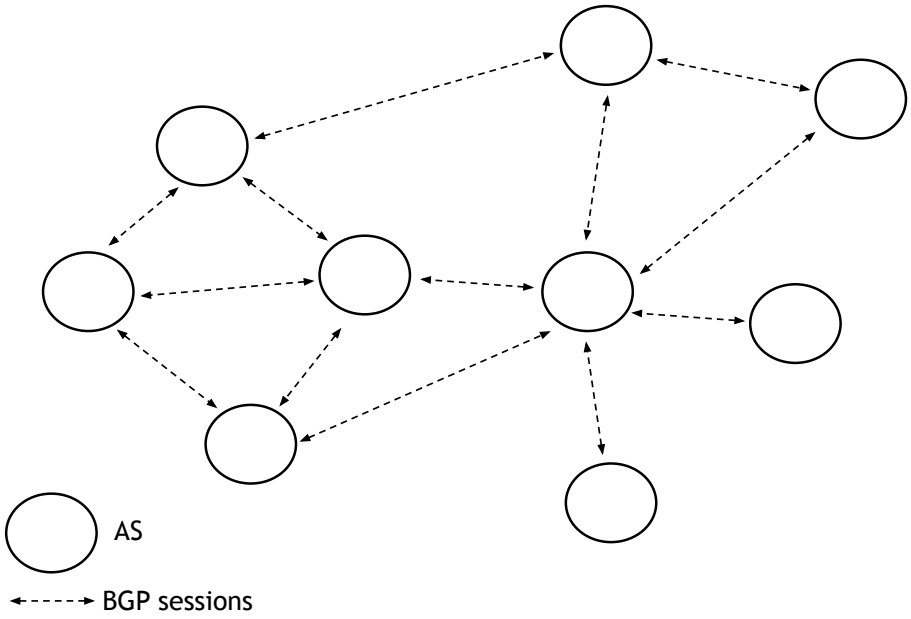


Figure 1.6 – Internet model.

1.3.1 Relationships between ISPs: peering and transit

One of the main issues of the Internet – already known in the days of the old telephone world – is how to connect all devices (PCs, smartphones, tablets, servers, etc.) that have a public IP address (i.e., directly connected on the Internet) scattered throughout cyberspace.
According to Martin Libicki – cybersecurity expert – cyberspace consists of three layers:

- physical: the physical components of cyberspace (underwater cables, antennas, satellites and optical fibers, etc.);
- syntactic: protocols, rules and natural properties governing the operation and interaction between the different physical components of cyberspace;
- semantic: the result of the interaction between the first two levels is what gives meaning to the processes of the underlying levels, ensuring their operation.

Obviously, it is inconceivable that each ISP, when reaching all the different devices connected to the networks of the other ISPs, is directly connected to all other ISPs worldwide. There must necessarily be an interconnecting mechanism, with different networks acting as transit for the other networks. Interconnection can be direct or indirect (transit), through one or more networks that accept to transport traffic.
There are two kinds of interconnection agreements between ISPs:

- **Peering**: two or more ISPs interconnect directly to one another, to exchange traffic between their clients. This is often done without interconnection or traffic charges (in the literature, these are called settlement-free agreements). Please note that peering is a non-transitive relationship, i.e., if ISP-A has a peering agreement with ISP-B, and ISP-B has a peering agreement with ISP-C, this does not imply that ISP-A has a peering agreement with ISP-C. Peering agreements are exclusively between two ISPs (bilateral). What's more, in such a situation, ISP-A cannot use ISP-B as transit to exchange traffic with ISP-C.
- **Transit**: an ISP accepts to transport the traffic originated by an ISP and directed to another ISP. Since no ISP directly connects to all the other ISPs, an ISP providing a transit service will deliver part of the traffic indirectly through one or more transit ISPs. The transit service provider usually receives economic compensation for the service.

Figure 1.7 below, shows the difference between peering and transit relationships.

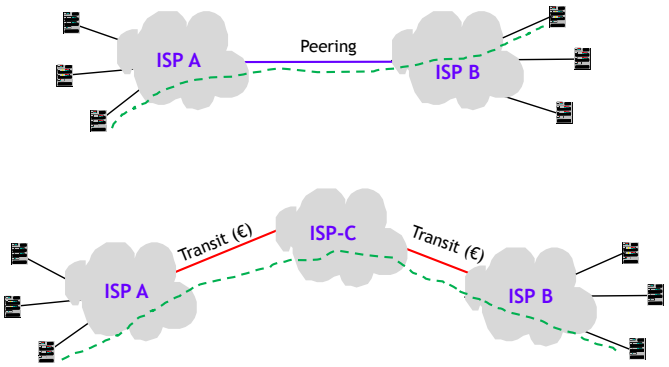


Figure 1.7 – Relationships between ISPs: peering and transit.

1.3.2 Internet eXchange Point (IXP)

An Internet Exchange Point (IXP), traditionally known as NAP (Network or Neutral Access Point), is a physical infrastructure that allows different ASes to exchange Internet traffic between one another.

NOTE: Whether they are commercial or not, in general, European IXPs are managed neutrally with respect to their participants. If a participating ISP or carrier or content network owned and managed the IXP, potential conflicts of interests could arise. Neutrality is the reason for the success of many big Northern European IXPs.

By promoting AS interconnection through peering agreements that are usually free of charge (at least until the power relations between the ASes involved are balanced), an IXP allows ASes to save part of the bandwidth they buy from their Upstream Providers, with efficiency and reliability gains.

NOTE: An Upstream Provider is generally a big transit ISP that provides access to the Internet to a local ISP or content network.

The main purpose of an IXP is allowing ISP networks to connect to one another directly, rather than making traffic pass through one or more external Upstream Providers. This offers the following advantages:

- **Speed**: a direct connection between two ASes, without intermediate passages, minimizes the latency of the packets crossing them, improving network performance, especially toward all real-time interactive or content applications.
- **Efficiency**: diversifying the connections of an Internet provider toward the rest of the ISPs, allows for greater routing control (by enhancing local Internet connectivity and security), increased network infrastructure redundancy, and therefore a higher number of possible paths toward a given destination.
- **Cost**: fixed costs related to being associated with an IXP (including interconnection costs toward the IXP data center and toward the Fabric) are generally lower (per exchange bandwidth unit) compared to Internet transit costs. Most of the times, peering agreements between participants to an IXP take place free of charge, which makes access to the Internet cheaper, and therefore available to a larger number of end users in a certain country or region (think about developing economies).

The typical infrastructure of an IXP – also called IXP Fabric – consists of one or more switches to which different participants connect their routers. In addition, it might include servers through which the IXP provider offers additional services to its participants (e.g. aggregate and target AS traffic statistics), as well as services that allow for correct Internet operation: e.g., anycast replicas of root name servers and analysis tools such as Routing Information Services (RIS), managed by Regional Internet Registries and hosted precisely inside the IXP infrastructures.

NOTE: The most used switching technology in IXPs has switched from ATM (very popular in the 1990s) to Ethernet. Some IXP migrated to more scalable solutions, such as the IP Fabric with VXLAN transport and EVPN control plane.

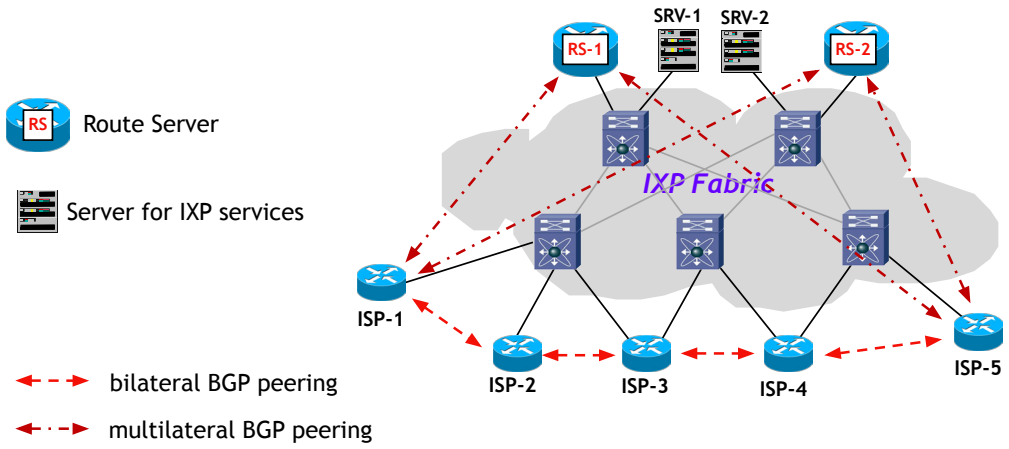


Figure 1.8 – IXP infrastructure.

Through BGP, routers establish peering agreements that allow ISPs to exchange Internet traffic. Peering agreements are called bilateral, when they are established directly between ISPs. There might be numerous bilateral agreements, therefore, in order to reduce them, IXPs make special devices called Route Servers available, which are used to reflect BGP advertisements from one ISP to all the other ISPs. So an ISP, instead of managing N-1 peerings, where N is the total quantity of ISPs within the IXP it wants to exchange routing information with, just needs a BGP peering toward a Route Server, or rather, by redundancy, two BGP peerings toward two Route Servers, to drastically reduce the number of BGP peerings. In this case, we talk about multilateral agreements. As mentioned earlier, the main purpose of an IXP is providing a physical infrastructure, through which network providers can interconnect and exchange traffic, gaining common advantages. Over the years, the role of IXPs has changed. Through an increasingly broader and diversified participation, they expanded their services from simple peering to facilitating a market where participants can purchase different services they need from other participants (e.g., DDoS mitigation services as well as access services). On one hand, this new nature is a strength for IXPs, which can attract more and more providers (which consider participation as a new business opportunity); on the other hand, participants can benefit from a broader service range.

1.3.3 ISP Classification

The ISP classification generally accepted among networking professionals was first drafted in *An analysis of internet inter-domain topology and route stability*, published in 1997 by Ramesh Govindan and Anoop Reddy. Then, a more structured definition can be found in Geoff Huston's article, *Interconnection, Peering, and Settlements* of 1999:

- **First level** (**Tier-1**): An AS (usually, though not necessarily, an ISP) capable of connecting to the entire Internet, without purchasing transit services from other ISPs (transit free). Tier-1 ISPs generally have peering agreements between them, and do not use the default route. Therefore, this group of ASes (very few, less than 20 worldwide) is often called Default Free Zone (DFZ).
- **Second level** (**Tier-2**): a network communicating with the other networks, by purchasing at least two IP transits to reach the entire Internet. Possible examples are big national ISP networks. Tier-2 ISPs too, generally have peering agreements between them. Usually, peering

between Tier-2 ISPs occurs through direct private connections (PNI – Private Network Interconnection), often achieved using the passive interconnection infrastructure (MMR, Meet-Me-Room). In this case, we talk about private peering, as opposed to public peering, where ISPs connect using the IXP Fabric.

- **Third level** (**Tier-3**): a network that must necessarily purchase a right to transit from other networks (at least two) to reach the Internet. Usually, Tier-3 ISPs work within a limited territory (usually a region), and are very aggressive in their pricing policies. Their clients generally include retail or small businesses. Usually, Tier-3 establish peering agreements with Content Providers; interconnection often occurs within the IXP, and prefixes are exchanged through Route Servers (Content Providers, with their PoP scattered throughout the world, try to limit the number of BGP sessions, by enabling bilateral sessions only above a certain traffic threshold).

NOTE: The need to have at least two transit relations with other independent systems is one of the requirements a RIR may request in order to assign an AS number. See document RIPE-679 – *Autonomous System (AS) Number Assignment Policies*, March 2017, that reads: *"A network must be multi-homed in order to qualify for an AS Number."*. The RIR established practice consists in deeming even the subscription of a transit agreement with two different independent systems sufficient, without necessarily requiring the interconnections to be operational.

There are many reasons why network professionals use level hierarchy to describe the network, the most important being a greater understanding of the political and economic reasons behind a network, based on how and with whom it communicates.

NOTE: We should specify that the classification suggested and used herein, only aims at showing what we observe every day in the Internet ecosystem. This means that, since the Network is perpetually changing, an independent system can shift from one category to the other, based on the expansion and growth policies it applies.

1.4 BASIC OPERATION

BGP's basic operation is very simple, and, in some ways, it resembles the old RIP. Each router, after establishing some sort of relationship with another router – belonging to the same AS or not – informs it of the IP prefixes it can reach, by linking to each prefix a distance measured in terms of number of ASes it needs to cross, to reach the AS originating the prefix.
The relationship that a router establishes with another router is called BGP session. Inside a BGP session, routing information is exchanged through special protocol messages, called BGP UPDATE, to which a set of BGP attributes may be associated. Some of these attributes are mandatory, while others are optional (see Chapter 2), and they have different functions, the most important being the definition of suitable traffic management policies.
UPDATE messages can also be used to notify the need to eliminate (withdraw) the advertisements of a previously sent prefix to the other routers, due to the impossibility of finding available paths toward those prefixes. This BGP behavior would lead to network instability, due to the route flapping phenomenon – close advertisements and withdrawals of the same IP prefix, due to close up/down transitions of a BGP session. Route flapping causes a significant engagement of the routers' CPU on the entire network, due to the propagation of UPDATE messages. If not properly regulated, this phenomenon could lead to serious operating issues on the entire Internet. In Chapter 8, we will see how the Route Flap Damping mechanism (although this topic generated much discussion among experts) helps to mitigate the instability caused by route flapping.

Even the aggregation of IP prefixes – which we'll see in Chapter 5 – can be a valid tool to ensure stability.
Figure 1.9 below summarizes the basic operation described herein. In the figure, the BGP table is a memory area where the router keeps the BGP advertisements received from each neighbor with whom it has established a BGP session.

1.4.1 BGP as Path Vector protocol

Generally, IP routing protocols fall within two categories: *Distance Vector* and *Link State*. In Distance Vector protocols, the routing process announces the "Distance Vector" to its neighbors, comprising a set of elements, where each element is a <IP prefix, distance> pair. The IP prefix is a reachable prefix (i.e., there is a path to reach it in the IP routing table), and the distance is the minimum cost to reach it, calculated by each type of routing protocol in the relevant manner. For instance, RIP – the first Distance Vector protocol ever developed – uses the Hop Count metrics to calculate the minimum distance, meaning it measures the distance in terms of number of routers to cross to reach the one where the IP prefix is located, by applying the Bellman-Ford algorithm, formulated in the mid-1950s. On the other hand, Link State protocols employ a more sophisticated criterion to determine the optimal paths. First, they determine the network topology, through a relevant message exchange, then the metric related to each connection, and lastly, based on the SPF (Shortest Path First) algorithm – created in 1959 by Dutch mathematician Edsger W. Dijkstra – they calculate the minimum cost path. The two most important examples of routing protocols based on the Link State algorithm are OSPF – Open Shortest Path First (version 2 in RFC 2328; and version 3 in RFC 5340, which supports IPv6) and IS-IS – Intermediate System to Intermediate System, in ISO/IEC 10589:2002.

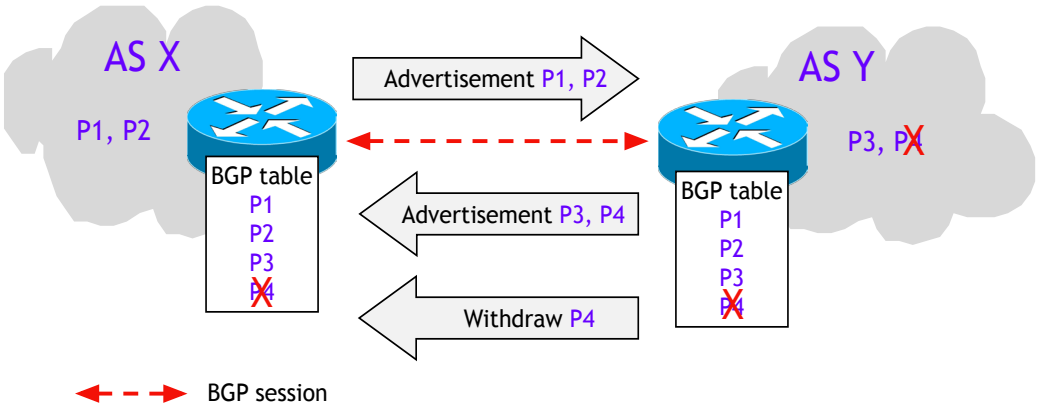


Figure 1.9 – Basic BGP operation.

Even though its operation resembles a Distance Vector protocol, BGP does not actually belong to the Distance Vector type, nor to the Link State type. It belongs to the Path Vector type. The reason behind this, is the presence of a special attribute – the Path Vector (which, as we will see, is actually called AS_PATH) – a sorted list (Vector) of the AS crossed by the BGP advertisements. Path Vector indicates the path to cross to reach a prefix, in terms of ASes. For instance, in Figure 1.10 below, prefix 192.0.2/24, belonging to AS 64503, will be announced to AS 64501 both by AS 64502 and by AS 64505. In the first advertisement, the Path Vector is [64502 64503], while in the second is [64505 64504 64503]. This is why BGP is referred to as a Path Vector protocol.

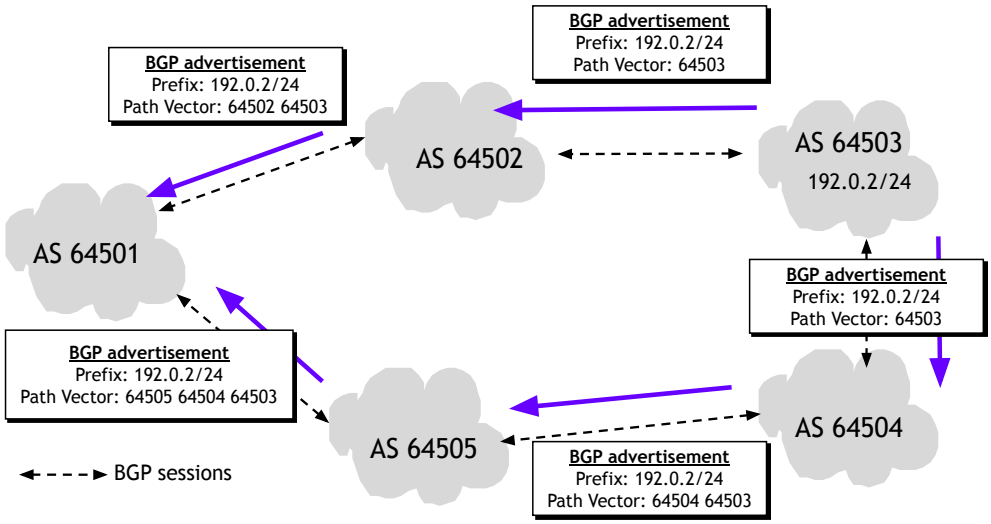


Figure 1.10 – The Path Vector attribute.

We will see later on how the Path Vector plays an essential role in BGP operation.
From the basic operation description above, we can notice once again how BGP is conceptually very simple, and maybe this explains also why it is so flexible. The complex part of BGP concerns one of the essential aspects it was designed for, that is its routing policy definition applications. And this is surely the most important and interesting aspect in this protocol.

1.4.2 Selecting the best path

Generally, as shown in the example in Figure 1.10, a router receives advertisements of the same prefix from different sessions. The advertisements of all prefixes exchanged are stored in certain memory areas, linked to each BGP session (see Section 1.4.3), and then subject to possible handling through BGP attributes and/or by applying filtering policies. The BGP process within each router chooses the best path among all the advertisements of the same prefix; the best path (and only that) is propagated on the BGP sessions, following specific rules that we will see in Paragraph 2.1.
The best path is selected according to a well-established and sorted sequence of choices, based on different metrics or advertisement properties. This sorted sequence of choices is called the selection process, and it generates a unique best path, in the end. For instance, referring to Figure 1.10, the routers of AS 64501 have two possible paths to reach prefix 192.0.2/24: they can transit through AS 64502 or through AS 64505 and AS 64504. Which path should they choose? Intuitively, we should choose the shortest path. Since BGP does not give us any information on the AS internal layout, we can define the shortest path as the one that crosses the fewest ASes. According to this criterion, the best path would be the one using AS 64502 as transit.
However, are we sure that it wouldn't be more convenient for AS 64501 to transit through AS 64505? Some of the reasons could be a higher bandwidth available, more advantageous business agreements, and so on. Through the selection process (more on this in Paragraph 2.5), BGP makes a series of metrics available, to manage the selection and choose the best path for the AS administrator's requirements.

1.4.3 BGP Process Model

The BGP process within a router can be modeled as shown in Figure 1.11 below, and it comprises the following macro-blocks:

- **Adj-RIB-in**: memory areas linked to each BGP session, where advertisements received from BGP sessions are stored (through UPDATE messages).
- **Input Policy Engine**: a set of inbound routing policies applied to the advertisements received, comprising filters on the advertisements and/or BGP attribute manipulation.
- **BGP selection process**: it chooses, among the advertisements of the same prefix accepted by inbound routing policies, the best path to reach the prefix.
- **Loc-RIB**: a table that contains the best paths.
- **Output Policy Engine**: a set of outbound routing policies applied to the best paths to propagate on the other BGP sessions, comprising filters on the advertisements and/or BGP attribute manipulation.
- **Adj-RIB-out**: memory areas linked to each BGP session, where the outbound advertisements to be propagated on BGP sessions are stored.

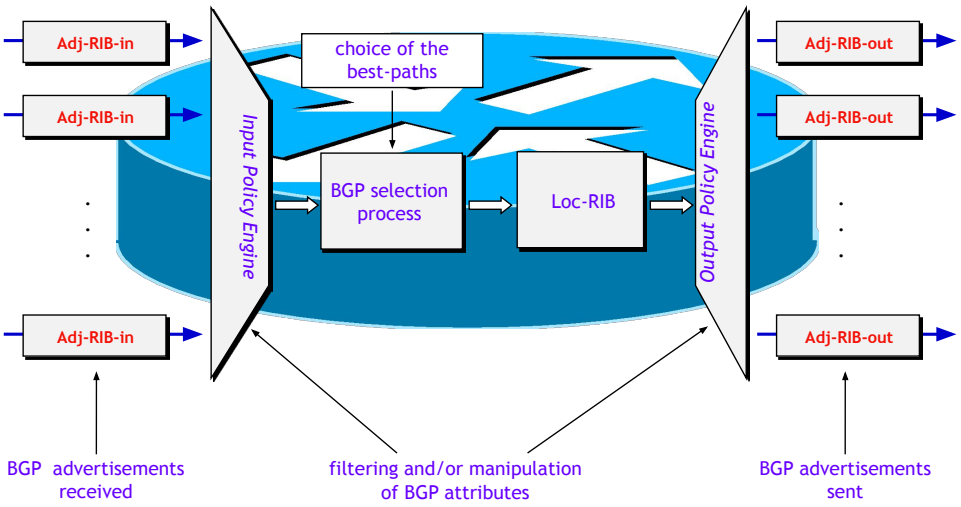


Figure 1.11 – BGP Process Model.

The partition of the memory used by the BGP process in Adj-RIB-in, Loc-RIB and Adj-RIB-out is purely logical, and it is the one described in standard BGP documents (RFC 1771 and RFC 4271). In practical implementations, each device manufacturer manages the memory areas to be assigned to the BGP process as they best see fit. Hereinafter, for the sake of simplicity, we will define the BGP table as the set of advertisements deemed valid for the selection process, i.e., the set of all BGP advertisements received that have been processed by inbound routing policies.

NOTE: The best paths determined by the BGP selection process are not necessarily installed in the IP routing table (RIB, Routing Information Base) and then transferred to the FIB (Forwarding Information Base) to be used in traffic forwarding. Indeed, if the same prefix is announced to

the router via BGP and also in other ways (dynamic routing protocol, static routing, directly connected prefix), the router chooses which advertisement to install in the RIB, based on a level of preference assigned by the router to each routing protocol. Please be aware that the level of preference (known in Cisco documents as administrative distance, or in Juniper documents as preference value), is a number assigned locally by the router to each routing protocol, and expressing a level of preference for the protocol. For instance, let's assume that a prefix Pfx is announced to the router both by BGP and by OSPF. Which of the two protocols does the router consider more reliable? In other words, what information <*Pfx*, *Next-Hop*> will be installed in the RIB, the one announced by OSPF or the one announced by BGP? The rule used by all manufacturers is preferring the advertisement of the protocol with the lowest preference level. Note that preference level values are assigned by manufacturers according to different logics, and so it is common to see completely different numbers, which can be varied through a configuration, if needed.

The complex part in the practical implementation of BGP, consists of two routing policy blocks, which are the core of the protocol. In Section 1.4.4 below, we will see what routing policy means, and throughout the textbook we will go over the tools available to implement them, along with several practical applications.

1.4.4 Routing Policies

One of the main reasons behind BGP's success as a routing protocol of the networks based on the TCP/IP architecture, is the option of creating very flexible routing policies that meet (almost) all the network administrators' needs.
A routing policy defines the rules adopted by an AS to manage inbound and outbound traffic, and the BGP advertisement acceptance and sending rules.
There are basically two tools to define a routing policy:

- filtering of BGP advertisements;
- BGP attribute manipulation.

Filtering allows a router to choose what BGP advertisements to accept and/or propagate. Based on the application direction, filtering can be of two types:

- *Inbound*: Allows choosing the advertisements to accept and to reject, between all the BGP advertisements received on the different BGP sessions. The advertisements accepted take part in the selection process. Vice versa, the advertisements rejected are deemed invalid for the selection process.
- *Outbound*: Allows choosing, between all the best paths determined, which ones to propagate to the routers with active BGP sessions.

Filtering can also be used to choose which prefixes to redistribute from an IGP protocol to BGP. Examples of filters are:

- rejecting all advertisements from IP prefixes with too big a mask (e.g., longer than 24 bits);
- not propagating the advertisements of IP prefixes received from a specific AS to other ASes (e.g., to prevent an AS from becoming a transit AS);
- rejecting all advertisements from prefixes that cannot be routed on the Internet (e.g., private IP prefixes from RFC 1918, Martian List, etc.).

Manipulation of BGP attributes allows one to change the value of BGP attributes, according to one's needs. Having control over BGP attribute values allows conditioning the choices of the best path selection process, and therefore defining suitable AS inbound and/or outbound traffic management policies.
There are many different routing policies, with obvious applications in case of multi-homed ASes or stub ASes with redundant connections. For stub ASes with single connection, it makes little sense to speak about routing policies, except for aspects relating to advertisements filtering, since stub ASes have no alternative ways to manage inbound and/or outbound traffic. Some examples of routing policies are:

- sending/receiving traffic using a first-choice AS (primary AS), and, in case of loss of connectivity toward it, using a backup AS;
- balancing inbound and/or outbound traffic between two or more paths;
- choosing the most convenient paths, based on the prefix. For instance, in a primary/backup configuration, it may be convenient to make traffic toward local prefixes of backup ASes pass through a backup connection.

The practical implementation of routing policies requires complex configurations that use specific tools made available by the BGP implementations of the different manufacturers. We will go over the tools made available by Cisco and Juniper platform later in this book.
Figure 1.12 shows an example of a routing policy application to check the advertisements received/propagated, and affect the results of the selection process.

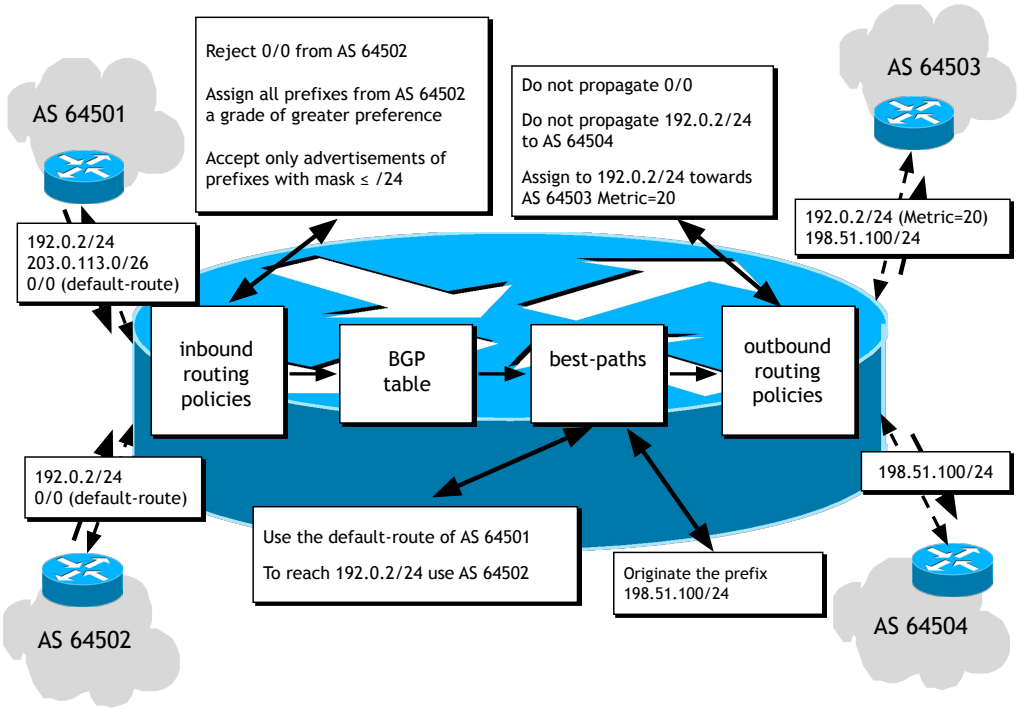


Figure 1.12 – Example of inbound/outbound routing policies.

By applying the inbound routing policies specified in the figure, we obtain the following results:

- the default route advertisement sent by AS 64502 is rejected, while the one sent by AS 64501 is accepted;
- the advertisement of prefix 203.0.113/26 sent by AS 64501 is rejected, because the prefix mask is too big (greater than /24);
- the advertisements of prefix 192.0.2/24 sent by AS 64501 and 64502 are both accepted, however, the advertisements coming from AS 64502 is assigned a greater preference value (through the special BGP attribute Local Preference, which we'll see in the next chapter).

The advertisements accepted are stored, along with their attributes, in the BGP table, and the selection process is applied to them, with the following results:

- best path for prefix 192.0.2/24: AS 64502;
- best path for default route: AS 64501.

NOTE: The BGP table is a set of BGP advertisements, with the purpose of providing information on how to reach the networks of the different autonomous systems. We should highlight that each BGP advertisement in the table is linked to the AS that originated it on the Internet.

The BGP table also includes advertisements of prefix 198.51.100/24, which becomes the best path, as it is the only one present.
Lastly, let's consider best path propagation only in BGP sessions toward AS 64503 and AS 64504. Before being propagated, prefixes undergo the outbound routing policies, with the following results:

- the best path of prefix 192.0.2/24 is propagated only to AS 64503 with metric 20 (through the special BGP attribute MED, which we will see in the next chapter);
- the best path of prefix 198.51.100/24 is propagated to both AS 64503 and 64504;
- the best path of the default route undergoes an output filtering process and is not propagated (although it remains in the BGP table).

This basic example shows that there can be many different routing policies, and, if they are defined well, they can meet (almost) all network administrators' needs.

SUMMARY

In this opening chapter, we described the role BGP plays in the Internet ecosystem, and its manifold applications in Service Provider networks; then, we paved the way for a conceptual model on which the BGP process operation depends. Everything we presented here will be explored further in the next chapters.

Worth remembering:

1. BGP's role in the Internet ecosystem and its applications in the services offered by Service Providers (e.g. BGP/MPLS services).
2. The concept of Autonomous System and its numbering method. In addition, you should remember the AS classification into single-homed (stub AS) and multi-homed, and the further division of the latter into transit AS and non-transit AS.
3. BGP's basic operation, BGP sessions and the Path Vector protocol. In particular, the BGP UPDATE message exchange, and the best path selection.
4. The operating model described in Figure 1.11, which will be the logic behind the entire textbook.
5. And last but not least, what is perhaps BGP's greatest value, the option of creating routing policies on AS inbound and outbound traffic.

BGP from theory to practice

The book describes the main aspects of BGP (*Border Gateway Protocol*) and its most important practical applications, such as traffic management policies, scalable architectures, stability and security mechanisms, and its role in big Enterprises and ISP networks. For each topic covered, also the implementation aspects in Cisco and Juniper technologies are included, essential to fully understand the main mechanisms of the protocol and its applications. The central idea behind the book is combining theory and practice, to avoid turning it only into a (debatable) presentation of the standard. This is why, apart from explaining in detail and with many examples how the protocol works, and its role in IP networks, the book also includes many practical application suggestions, resulting from many decades of experience.

Main topics covered:
Fundamentals (AS, sessions, messages, attributes, etc.) - Prefix aggregation and filtering - Inbound/outbound routing policies - Scalability, stability and security aspects - Convergence functions - BGP in Enterprise and Service Provider networks - The role of BGP in MPLS services - Basic and advanced implementation aspects in Cisco IOS XE/XR and Juniper JUNOS - BGP security aspects - Best implementation practices.

Flavio Luciani was born in Rome in 1981 and he graduated Computer Engineering at the Roma Tre University in 2005. Since 2008, he is part of the Namex team - the Internet eXchange Point in Rome - first as member of the technical staff, and, since 2020, as Chief Technology Officer.
He is currently involved in several Internet Community initiatives. He works with the RIPE NCC association, with the European inter-exchange point association EURO-IX, and he is a member of the Steering Committee of the initiative promoted by the Internet Society (ISOC), Mutually Agreed Norms for Routing Security. Through workshops, courses and feature articles, he promotes greater awareness on routing security.

Antonio Prado, after an educational background in Humanities Studies, lands a PhD in Computer Science. He has been active in the Internet industry since 1995, and, after a long experience as CTO for several telecommunication operators, he works in the Italian Public Administration, where he deals with digital infrastructures and digital transition. As journalist, he spreads knowledge on how the Internet works.
He has been holding classes in public institutions (Schools and Universities) and private organizations (ISOC, Reiss Romoli School) for over twenty years. Ambassador and member of MANRS advisory group, and committed every day to supporting the development of the Web in Italy, and to spreading awareness of the Internet Governance, through articles and conferences.

Tiziano Tofoni graduated in Engineering at the University of Padua, and obtained a Master's Degree in Mathematical Statistics at the *Florida State University, Tallahassee*, Florida (USA). He started his career as scholar at the Istituto di Dinamica dei Sistemi e Bioingegneria of the CNR, in Padua, and as Teaching Assistant at the Statistics Department of the *Florida State University*.
Then, he joined the staff of the G. Reiss Romoli Post-Graduate School (which was then part of the Telecom Italian Group), where he worked in the Traffic Engineering and Technology sector for the IP networks of Service Providers.
He held several courses at the University of L'Aquila, and he is the author of the first edition of the book "*BGP: From Theory to Practice*" and of the book "*MPLS Services"* (Ed. Reiss Romoli). He is member of Namex Technical Committee, honorary member of ITNOG Board, and Chairman of Reiss Romoli srl.

NAMEX ROMA IXP

REISS ROMOLI

€ 60.00