IMT Institute for Advanced Studies, Lucca

Lucca, Italy

An Analysis of the Completeness of the Internet AS-level Topology Discovered by Route Collectors

PhD Program in Computer Science and Engineering

XXVI Cycle

By

Luca Sani

2014

The dissertation of Luca Sani is approved.

Program Coordinator: Prof. Rocco De Nicola, IMT - Institute for Advanced Studies, Lucca

Supervisor: Prof. Luciano Lenzini, University of Pisa, Italy

Supervisor: Ing. Enrico Gregori, Institute of Informatics and Telematics - National Research Council of Italy, Pisa, Italy

Tutor: Prof. Alberto Lluch Lafuente, IMT - Institute for Advanced Studies, Lucca, Italy

The dissertation of Luca Sani has been reviewed by:

Prof. Giuseppe Di Battista, University of Roma 3, Roma, Italy

Prof. Christos Papadopolous, Colorado State University, Fort Collins (CO), USA

,

IMT Institute for Advanced Studies, Lucca

2014

To my Family.

Contents

List of Figures

List of Tables

Acknowledgements

First of all, I'd like to thank to all the *Isolario* team for making the work place a better place to live: thanks to Alessandro Improta, Pietro J. Giardina, Alessandro Pischedda, Lorenzo Rossi and Andriano Faggiani (even though it is part of the Portolan team). Special thanks to prof. Luciano Lenzini and Dr. Enrico Gregori for their supervision during my PhD and for having made this adventure possible. Thanks also to all the people I met during my period abroad in Colorado, in particular prof. Dan Massey, prof. Christos Papadopolous, Dr. Catherine Olschanowsky, Anant and Daniel.

Vita

Apr 08, 1986 Born, Barga (Lucca), Italy

Sep 26, 2008 B.Sc. Degree in Computer Engineering
Final mark: 103/110
University of Pisa, Pisa

Dec 17, 2010 M.Sc. Degree in Computer Engineering
Final mark: 110/110
University of Pisa, Pisa

Mar 2011 Enrolled as PhD student at IMT - Institute for Advanced Studies, Lucca

Jan 27, 2013 (Six months) Visiting research scholar at Colorado State University, Department of Computer Science under the supervision of Dr. Dan Massey, to work on BGPmon (BGP monitoring system).

Publications

1. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, Luca Sani, "BGP and Inter-AS Economic Relationships," *IFIP TC-6 Networking*, vol. 2, pp. 54-67, Valencia, Spain, May 9-13th, 2011.
2. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, Luca Sani, "Inferring Geography from BGP Raw Data," in *IEEE INFOCOM NetSciCom*, pp. 208-213, Orlando, U.S.A., March 30th 2012.
3. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, Luca Sani, "Selecting new BGP Feeders to Address the Incompleteness of the Internet AS-level Graph," in *Informatica Quantitativa (InfQ) Workshop*, Lucca, Italy, July 5-6th 2012.
4. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, Luca Sani, "On the Incompleteness of the AS-level graph: a Novel Methodology for BGP Route Collector Placement," in *Internet Measurement Conference (IMC)*, pp. 253-264, Boston, U.S.A., November 14-16th 2012.
5. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, Luca Sani, "Discovering the geographic properties of the Internet AS-level topology," in *Networking Science*, Volume 3, Issue 1-4, pp 34-42, December 2013.
6. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, Luca Sani, "Improving the Reliability of Inter-AS Economic Inferences Through a Hygiene Phase on BGP Data ," in *Computer Networks*, Volume 62, pp 197-207, April 2014.
7. Adriano Faggiani, Pietro G. Giardina, Enrico Gregori, Alessandro Improta, Valerio Luconi, Luciano Lenzini, Alessandro Pischedda, Lorenzo Rossi, Luca Sani, "When Traceroute Met BGP... How to Reveal Hidden Internet AS-level Connectivity with Portolan and Isolario ," in *IEEE INFOCOM*, Toronto, Canada, April 29th-May 2th, 2014.
8. Adriano Faggiani, Enrico Gregori, Alessandro Improta, Luciano Lenzini, Valerio Luconi, Luca Sani, "A Study on Traceroute Potentiality in Revealing the Internet AS-level Topology ," in *IFIP TC-6 Networking*, Trondheim, Norway, June 2-4 2014.
9. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, Luca Sani, "A Novel Methodology to Address the Internet AS-level Data Incompleteness ," in *IEEE/ACM Transactions on Networking*, 2014 *(To appear)*

Presentations

1. "Inferring Geography from BGP Raw Data," at *IEEE INFOCOM NetSciCom*, Orlando, USA, March 30th, 2012.
2. "Selecting new BGP Feeders to Address the Incompleteness of the Internet AS-level Graph," at *Informatica Quantitativa (InfQ) Workshop*, Lucca, Italy, July 5-6th 2012.
3. "Isolario Project," at *Internet Measurement Conference (IMC), Feedback Session*, Boston, USA, 2012, November 14-16th.
4. "When Traceroute Met BGP... How to Reveal Hidden Internet AS-level Connectivity with Portolan and Isolario," at *IEEE INFOCOM 2014*, Toronto, Canada, April 29th-May 2th, 2014 (Demo presentation together with Adriano Faggiani).

Abstract

In the last decade many studies have exploited the BGP data provided by route collector projects to infer an Internet AS-level topology to perform several analyses, from discovering its graph properties to assessing its impact on the effectiveness of worm-containment strategies. Nevertheless, the topology that can be extracted from this data is far from being complete, i.e. from this data it is not possible to infer all the AS connections which actually exist among ASes.

This thesis analyses the available data and investigates the contribution of route collectors in terms of AS-level connectivity by taking into account economic and geographic characteristics of the Internet AS-level ecosystem. By leveraging on a new metric, named *p2c-distance*, this analysis shows that the largest amount of ASes currently connected to a route collector belongs to the Internet core, thus the collected data is highly biased and is missing a lot of connections established in the Internet periphery. To address this problem, it should be increased the amount of ASes participating to a route collector project. To this end, this thesis describes how to improve the coverage of route collectors by means of an optimization problem based on the *p2c-distance* metric, which solution quantifies the minimum number of ASes that should join a route collector in order to obtain an Internet AS-level topology as complete as possible. The results show that route collectors are rarely connected to the selected ASes, highlighting that much effort is needed to devise an ideal route collector infrastructure that would be able to capture a complete view of the Internet.

These analyses require the ability to infer the economic relationships which rule the exchange of BGP messages between each pair of connected ASes of the topology. Existing economic tagging algorithms do not take properly into account that BGP data has to be purged from *spurious routes*, usually caused by router misconfigurations on BGP border routers and which shows up during the BGP path exploration phenomenon. In this thesis an economic tagging algorithm which is able to get rid of these spurious routes is described. This algorithm leverages on robust statistical concepts, rather than on debatable time thresholds and questionable graph metrics. The analyses provided in this thesis are further refined considering the geographical distribution of ASes, which in the global AS-level topology are all considered as a single node. A global analysis could lead to misleading results since an AS connection may hide multiple connections located in different geographic regions, possibly regulated by different economic relationships. From this analysis are indeed highlighted peculiar characteristics of regional topologies previously unrevealed, and it is showed that these considerations also affect the estimated number of feeder ASes needed to improve the completeness of the global AS-level topology.

Chapter 1

Introduction

The Internet is a complex system that evolved over the last few decades from a small network confined to the U.S. (i.e. ARPANET, 1969) to the current worldwide network of networks. It now consists of a plethora of networks, grouped under the administrative control of about 40,000 Autonomous Systems (ASes). An Autonomous System "*is a connected group of one or more IP prefixes run by one or more network operators which has a* single *and* clearly *defined routing policy*" [RFCc]. Each AS is uniquely identified by an AS number (ASN) [IAN].

ASes belong to many different kind of organizations, e.g. research institutions, universities, Internet Service Providers (ISPs), Content Delivery Networks (CDNs), public institutions, private companies. More in general, an AS may belong to whichever organization owns a sufficiently complex network which inter-domain routing management meets the requirements stated in [RFCc]. Table 1 reports some examples of ASes together with their owner[1].

As a result, the Internet can be viewed as an *ecosystem* composed by autonomous players which compete and co-operate each other in order to guarantee the global connectivity to end users. Traffic generated by end users flows through the Internet crossing *routes* that ASes build and

[1] An organization may own more than one AS, however an AS usually belongs only to one organization.

Organization	ASN	Website
China United Telecom	9800	http://www.chinaunicom.com.cn
European Commission	42848	http://ec.europa.eu
Facebook-Inc	32934	https://www.facebook.com
First National Bank of Omaha	14888	https://www.firstnational.com
GEANT European Backbone	20965	http://www.geant.net
Holy See	8978	http://www.vatican.va
Level 3 Communications, Inc.	3356	htpp://www.level3.com
Registry of ccTLD it	2597	http://www.nic.it
Telstra	1221	http://www.telstra.com.au

Table 1: Organizations and ASes

maintain exchanging Border Gateway Protocol (BGP) messages upon establishment of dedicated BGP sessions (see Fig. 1).

As any other complex system, also the Internet ecosystem can be described by means of mathematical tools. One of the most suitable tools is a *graph*, in which each node represents an Internet player and each edge represents an interaction among two players. Over time different level of abstractions have been taken into consideration. Notable examples are the *IP interface-level* graph – in which each node represents a router interface and each edge represents a link between two interfaces; the *router-level graph*, in which each node represents a router and each edge represents one or more link between two routers; *the AS-level*, in which each node represents an AS and each edge represents one or more

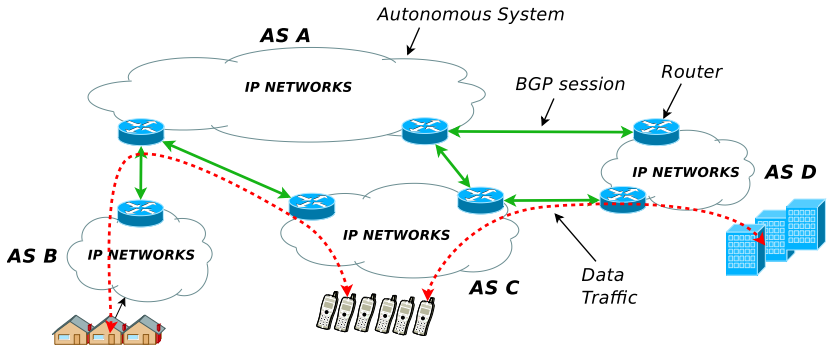


Figure 1: Internet ecosystem (example)

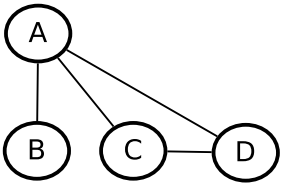


Figure 2: Internet AS-level graph example

BGP sessions between pair of ASes. Each level of abstraction has its own pros and cons and may be the most suitable depending on the type of analysis that has to be performed.

In this work, the chosen level of abstraction is the AS-level graph, which represents the so-called *Internet AS-level topology*. For example Fig. 2 depicts the AS-level topology of the Internet ecosystem depicted in Fig. 1. This work will be specifically focused on the analysis of the data from which the AS-level topology is usually inferred.

1.1 Inferring the Internet AS-level topology

The Internet AS-level topology has been the subject of many different kinds of studies, from the statistical analysis of its graph properties [FFF99, CCG+02, SARK02a] to the development of evolutionary models [CJW06, OZZ07, PSV04] passing through the analysis of its resilience in response to targeted attacks [DAL+05, LOZZ07]. Due to the distributed nature of the Internet, there not exists trusted third-party repositories containing an up-to-date available Internet AS-level topology that can be downloaded. The only available repositories are the Internet Routing Registries. However it is still difficult to distinguish fresh and complete connectivity information from stale or mistaken one, since they are manually maintained on a voluntary basis [CGJ+04]. Researchers, thus, tried to infer the AS-level topology by exploiting *collateral effects* of the BGP protocol. Specifically, a particular attribute of BGP messages – the `AS_PATH` – can be used to extract AS connectivity information. To infer the AS-level topology researchers exploited BGP data made available by route collec-

tor (RC) projects such as Route Views [Rou] and RIPE-RIS [RIS], which collects BGP data from routers belonging to ASes willing to participate. Despite the purpose of these projects was *not* to collect and provide data to infer the Internet AS-level topology – for example the homepage of Route Views states that "*Route Views project was originally motivated by interest on the part of operators in determining how the global routing system viewed* their *prefixes and/or AS space*" – researchers started to infer an Internet AS-level topology from these data and used this topology as the basis for their research studies, without concerning too much on its completeness [FFF99, WAD09].

Recently there have been efforts to analyse the (in)completeness of data obtained through BGP RC projects [OPW+10, RWM+11a], however there have not been any study to quantify it. In this thesis it will be proposed a methodology to quantify the completeness of the Internet AS-level topology captured by RC projects and to identify which ASes should join them in order to increase its completeness.

1.2 Thesis contribution

The contribution of this thesis is threefold. Firstly, BGP data currently gathered by well-known RC projects is analyzed, highlighting and explaining the causes of their incompleteness. It is shown that the current view of the Internet is limited due to the small number of ASes that are *effectively* feeding the RCs. It is also *biased* due to the nature of the feeding ASes, which are mostly managed by worldwide ISPs. This top-down view means that a large set of connections established among small ASes – i.e. elements of the Internet periphery – cannot be discovered [OPW+10, HSFK09, CR06]. Unlike other approaches [OPW+10], in this thesis the level of incompleteness of BGP data is analysed and quantified by relying only on public RC data. To do this, it is introduced a new metric, named *p2c-distance*, which takes into account both the presence of BGP decision processes and BGP export policies crossed by `UPDATE` messages before reaching a RC.

Secondly, it is designed a methodology to overcome the large amount

of incompleteness highlighted. By exploiting a tailored Minimum Set Cover (MSC) problem based on the inter-AS p2c-distance, this methodology selects the minimum number of ASes that should provide their default-free full routing table to the RCs. Although MSC problems have been proved to be NP-complete [GJ90], the provided methodology exploits the concepts of dominance and essentiality [McC56, Qui55, Qui59] to reduce the size of the problem, which can finally solved via an exhaustive search. In addition, the methodology provides a ranking list of ASes in order to understand to what extent the coverage can be improved with a limited number of new feeding ASes. To achieve this, it is solved a tailored Maximum Coverage (MC) problem using the classic greedy approach [CSRL01], considering only the coverage of the elements part of at least one MSC solution. Furthermore, it is analysed the current status of the RC coverage by identifying how many ASes in the optimal solutions are currently connected and how many new ASes should be connected in each region. It is also shown that by doubling the number of feeding ASes, the coverage would improve greatly.

Thirdly, the solutions obtained by applying the methodology is analysed on the global topology and five regional topologies of the Internet, showing the impact that geographical peculiarities have on the selection of the optimal set of ASes. The main reason behind the introduction of geographic topologies is that in the global topology each AS is represented by *one* node, and one or more BGP sessions between two ASes are represented by *one* link. Thus the analysis of the global topology underestimates the number of feeders needed to achieve the full coverage of the Internet core, as it is shown later in this work.

The analyses described so far would be however impossible to develop without a proper methodology to characterize the Internet from an economic perspective, e.g. the computation of p2c-distances requires the knowledge of economic relationships between ASes. For this reason, in this thesis it is also developed an economic tagging algorithm to infer an economic tagged AS-level topology, i.e. an AS-level topology in which each link is labeled with a tag representing the economic relationships between each pair of connected ASes. The most common types

of economic relationships are provider-to-customer (p2c), customer-to-provider (c2p), peer-to-peer (p2p) and sibling-to-sibling (s2s) [Gao01b]. An AS announces to its customers and siblings the routes obtained by its peers, customers, providers and siblings, while it announces to its providers and peers only the routes related to its customers. Moreover, in order to perform the geographical analysis, an algorithm to infer regional topologies from BGP data is presented.

1.3 Internet AS-level topology: *cui prodest?*

A legitimate question that could rise in reading works on the Internet AS-level topology is about its usefulness, i.e. *who* could exploit the Internet AS-level topology and *why*. The knowledge of the Internet AS-level topology, together with its economic and geographical characterization, could be useful for many different kinds of users. For example:

Scientific users could exploit this data to increase the scientific knowledge about the Internet AS-level ecosystem. The topology graph could be used to extract particular patterns in the connectivity (e.g. k-cliques, communities) which could be used as a starting point to build mathematical models describing the evolution of the Internet AS-level topology [AAG+14]. Moreover, economic and geographical topologies could be used to analyse how data packets are actually routed on the Internet. This knowledge would allow, for example, to design and test new protocols (e.g. multicast protocols [RTY+00]) or to build overlay networks aware of the underlying topology (e.g. peer-to-peer networks [RRW10]).

Network operators could exploit the dataset to select new peers or providers based on the connectivity that candidate ASes shows in the topology thus improving the service offered to customers. Moreover, a network administrator could use this knowledge to analyse the role that its AS plays in the Internet AS-level ecosystem. Operators running a content distribution network (CDN) may take advantage of economic and geographic topologies to select the most suitable places in which server replicas should be deployed [Bas03].

Internet governance could exploit the dataset to assess the resilience

of the Internet AS-level in response to targeted attacks (e.g. fiber cuts, DDoS and prefix hijacking [LOZZ07]), catastrophic events [WZMS07] (e.g. 9/11, earthquakes), censorship [DSA+11, KFR09, XMH11] etc. Moreover, this knowledge could be used to identify critical failures points in the Internet structure which may be crucial for political, economical, commercial and strategical purposes [ENI11].

1.4 Thesis organization

The rest of this work is organized as follows. Chapter 2 describes the RC projects that will be taken into consideration in this work, the data they provide and an analysis of its completeness. Chapter 3 describes the optimization problem which solution is the minimum number of feeders needed to achieve the full coverage of the Internet core and the optimization problem to devise a ranking of the candidate feeders. Chapters 4 and 5 describe respectively the methodology to infer economic relationships among ASes and regional AS-level topologies, which results are used to compute the solution of the optimization problems described in Chapter 3. Chapter 6 shows the results of the completeness analysis obtained solving the MSC and MC problems. Chapter 7 introduces Isolario, a new type route collector project which offers services in change of BGP data and which aims to attract the large amount of feeders needed to improve the completeness of the AS-level topology. Finally, Chapter 8 concludes the work.

1.5 Thesis Material

This thesis is based on several co-authored papers listed in page xvi. In detail, Chapters 2, 3 and 6 are based on [3, 4, 9], Chapter 4 is based on [1, 6]. Chapter 5 is based on [2, 5], Chapter 7 is in part based on [7].

Chapter 2

Route Collectors, Feeders and their contribution

2.1 Route Collectors

A route collector is a device which collects routing data by establishing BGP sessions with co-operating ASes (Fig. 3), hereafter called *feeders*. Typically a route collector runs a routing software suite (e.g. Quagga [Qua]) which handles a BGP session with each of its feeder' router. A feeder router thus perceives the RC as another BGP neighbor to which UPDATE messages are sent whenever a change occurs in its BGP routing table, following the rules of the BGP protocol. In other words, a RC receives and collects the *best routes* selected by each of its feeder routers to reach the Internet destinations. For example in Fig. 3 the RC R receives the route that AS3549 selected as the best on February 09, 2012 at 08:08:47 to reach the destination `212.77.0.0/19`. This means that all the traffic outgoing from AS3549 and destined to an IP address inside the subnet `212.77.0.0/19` will cross AS137 before to reach the destination, which is inside AS8978. As soon as AS3549 will change its best route to reach that subnet, the RC R will receive an UPDATE message containing the new best route or the withdrawal of the current best route.

As a proper BGP router, RCs maintain a Routing Information Base

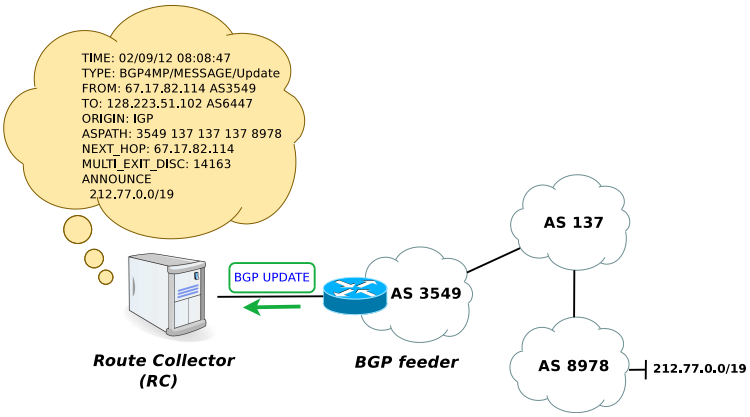


Figure 3: Route Collector (RC) example

(RIB) table [RFCb], which contains the best routes selected by all its feeders. Usually, the routing software running on a RC can be configured to periodically dump the content of its RIB – producing the so called *RIB snapshots* – as well as sequences of BGP update messages received over time. For example a route collector could be configured to produce RIB snapshots every two hours and to dump the sequence of UPDATE messages received every five minutes. Note that, over a fixed time interval, all routes present in a RIB snapshot are also present in the sequences of received BGP UPDATEs, while the vice versa is not necessarily true, e.g. a route could be announced and withdrawn between two RIB snapshots. Starting from a RIB snapshot, the subsequent UPDATE messages can be used to re-create the BGP session evolution of each feeder router without any loss of data.

A concrete example of these files can be found e.g. on `http://archive.routeviews.org/route-views.kixp/bgpdata/`, which is the web page from which BGP data dumped by the RC `route-views.kixp` (maintained by RouteViews [Rou] in Nairobi, Kenya) can be downloaded.

2.2 MRT data

Typically, both RIB snapshots and UPDATE messages are stored in files according to the Multi-Threaded Routing Toolkit (MRT) format [MRT], further compressed with the `bzip2` or the `gzip` algorithm. Tools like `libbgpdump` [lib] or `pybgpdump` [pyb] can be used to easily convert the content of these files in a human readable format. Each MRT file contains multiple entries – i.e. either a route in a RIB snapshot or a BGP UPDATE message – together with an header containing metadata about the entry itself, which is needed to properly use the information. For this work, the most significative fields of MRT metadata are three:

Peer IP address (*ipfrom*) and **AS peer (*asfrom*)**: they are respectively the IP address and the AS number of the remote-end of the BGP session from which the information is received. For example in Fig. 3 the *ipfrom* and the *asfrom* of the message are respectively `67.17.82.114` and `3549`. Note that the *ipfrom* and *asfrom* may not coincide with the actual *source* of the information. This happens – for example – in the case of a route server, since the BGP session from which the BGP data is received is established with the route server, but the actual source of the information are the clients of the route server itself [Cis, Iea13]. In practice, this means that the *ipfrom* may differ from the `NEXT_HOP` field of BGP UPDATEs.

Timestamp: this field stores the instant – expressed in Unix Time[1] – in which the message was received by the RC. This is useful in order to compute the lifespan of a route. Consider for example the two BGP UPDATE messages shown in Fig. 4. Those messages comes from the same BGP session, i.e. they have the same *asfrom* and *ipfrom*. The first advertises to the RC a route toward the destination `198.32.67.0/24` whereas the second withdraws such a route. This means that, this specific route, lasted one minute and two seconds. So, given a reference interval, it is possible to compute the lifespan of a route by summing all the intervals in which the route was present.

[1] The number of seconds that have elapsed since 00:00:00 UTC, Thursday, 1 January 1970.

TIME: 02/10/14 10:23:01
TYPE: BGP4MP/MESSAGE/Update
FROM: 196.223.21.66 AS4558
TO: 196.223.21.126 AS6447
ORIGIN: IGP
ASPATH: 4558 33770 21280 36948
NEXT_HOP: 196.223.21.81
ANNOUNCE
198.32.67.0/24

TIME: 02/10/14 10:24:03
TYPE: BGP4MP/MESSAGE/Update
FROM: 196.223.21.66 AS4558
TO: 196.223.21.126 AS6447
WITHDRAW
198.32.67.0/24

Figure 4: Announcement and withdrawn

2.3 BGP Route Collector Projects

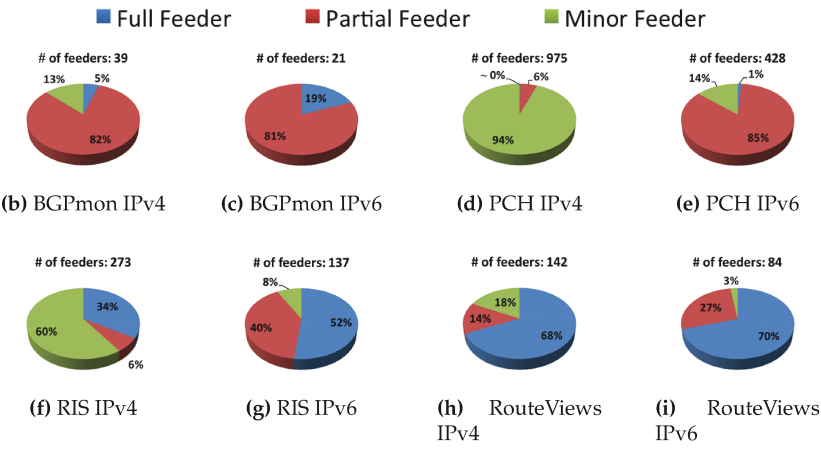
There are four main projects which deployed RC that collect BGP data and make it publicly available: RouteViews, RIPE-RIS, PCH and BGPmon.

RouteViews [Rou] was conceived in 1997 at the University of Oregon as a tool for Internet operators to obtain real-time information on the global routing system from the perspectives of several different backbones and locations.

RIS (Routing Information System) [RIS] was developed by Réseaux IP Européens (RIPE) Network Coordination Centre (NCC), the European Internet Registry, which collects and stores Internet routing data from several locations around the globe deployed on the largest IXPs.

PCH (Packet Clearing House) [PCH] is a non-profit research institute that supports operations and analysis in the areas of Internet traffic exchange, routing economics and global network development. Since July 2010, PCH has made BGP data available on its website, collected by several route collectors deployed on distinct IXPs.

BGPmon (BGP Monitoring System) [BGP] was set up at the Colorado State University to monitor BGP UPDATE messages and routing tables in real-time, and since August 2012 has made parsed Multi-threaded Routing Toolkit (MRT) data publicly available in XML format. Note that actually BGPmon is not a RC project like the previouses. Instead, it is a framework software able to collect BGP messages from BGP peers, i.e. it can replace the Quagga software. This means that BGPmon could be



(b) BGPmon IPv4 **(c)** BGPmon IPv6 **(d)** PCH IPv4 **(e)** PCH IPv6

(f) RIS IPv4 **(g)** RIS IPv6 **(h)** RouteViews IPv4 **(i)** RouteViews IPv6

Figure 5: Feeder classification per project

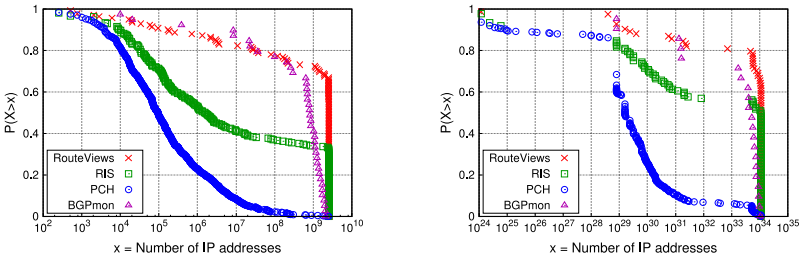
used by other RC projects on their devices to collect BGP data and – actually – it is being used by some RCs of RouteViews[2]. However, since BGPmon is also being used on some Colorado State University devices to collect BGP data from different peers, here it is considered a RC project like the others.

Hereafter in this thesis are analysed BGP routing information provided by these projects during February 2014. In this time interval, the number of active route collectors which provided BGP data in MRT format was 1 for BGPmon, 69 for PCH, 13 for RIS and 13 for RouteViews.

2.4 Feeder contribution analysis

Although apparently counter-intuitive, several feeders do not provide any relevant routing information to the RCs. To better quantify the total contribution of feeders, they are subdivided on the basis of the amount

[2] http://www.routeviews.org/bgpmon-routeviews.html



(a) IPv4 space **(b)** IPv6 space

Figure 6: CCDF of the amount of IP space announced by each feeder

Feeder type	IPv4	IPv6
Minor	$A_{IPv4} \leq 2^{24}$	$A_{IPv6} \leq 2^{96}$
Partial	$2^{24} < A_{IPv4} < (S_{IPv4} * 75)/100$	$2^{96} < A_{IPv6} < (S_{IPv6} * 75)/100$
Full	$A_{IPv4} \geq (S_{IPv4} * 75)/100$	$A_{IPv6} \geq (S_{IPv6} * 75)/100$

Table 2: Feeder classification considering the amount of IP space announced (A) to the RC with respect to the total IP space (S)

of IP space[3] that each feeder advertised to the RCs, as reported in Table 2: *minor feeders* announce an IPv4 space smaller than a single /8 IPv4 subnet or an IPv6 space smaller than a single /32 IPv6 subnet[4], *full feeders* announce an IPv4 (IPv6) space *close* to the full Internet IPv4 (IPv6) space currently advertised, while *partial feeders* include those ASes in between. For the sake of simplicity, in this work the full Internet IPv4 (IPv6) space is considered to be composed by the IPv4 (IPv6) addresses collected by RCs, and an AS is considered to be a full feeder if it announces more than 75% of the full Internet IPv4 (IPv6) space.

Following this subdivision – as shown in Fig. 5 – are found two IPv4 full feeders for BGPmon, three for PCH, 92 for RIS and 97 for RouteViews, and there are four IPv6 full feeder for BGPmon, five for PCH, 71

[3]The IP space is computed considering only *non-overlapping* subnets, i.e. those subnets that are not included in any other subnet

[4]http://www.ripe.net/internet-coordination/press-centre/understanding-ip-addressing

for RIS, and 59 for RouteViews. Together they make up a set of 167 different IPv4 full feeders – i.e. 14.74% of the total number of IPv4 feeders – and 119 different IPv6 full feeders – i.e. 21.80% out of the total number of IPv6 feeders. The big picture regarding feeder class distribution is depicted in Figures 6a and 6b, which show respectively the amount of non-overlapping IPv4 and IPv6 addresses announced from the feeders of each project. The number of full feeders – identified by the height of the vertical tail of the CCDFs – represents the clear majority only in the RouteViews project. On the other hand there are very few full feeders owned by PCH compared with its total number of feeders (cf. Figures 5d and 5e). In addition, it must be considered that only 94 ASes announce to the RCs both the IPv4 and the IPv6 full routing table. This behavior can be partially explained by the nature of the feeders themselves. There are still ASes on the Internet that have not deployed IPv6 on their networks yet, thus their routers are unable to announce any IPv6 route to the RCs. However, only seven ASes out of the 73 that announce only an IPv4 full routing table to the RC infrastructure do not announce any IPv6 prefix towards the Internet. This means that most of these AS networks are actually IPv6-capable. Probably for any technical or commercial reasons some of the feeders are just interested in announcing their IPv4 (IPv6) reachability, and limit the amount of information concerning IPv6 (IPv4) reachability. This can be proved by the presence of several IPv4 (IPv6) full feeders in the minor/partial set of IPv6 (IPv4) feeders. In detail, 24 out of 73 IPv4 full feeders appear either as IPv6 minor or partial feeders, while 20 out of 25 IPv6 full feeders appear either as IPv4 minor or partial

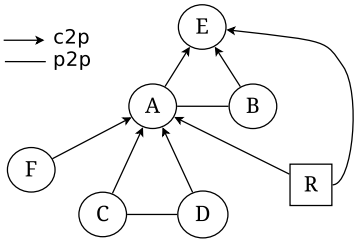
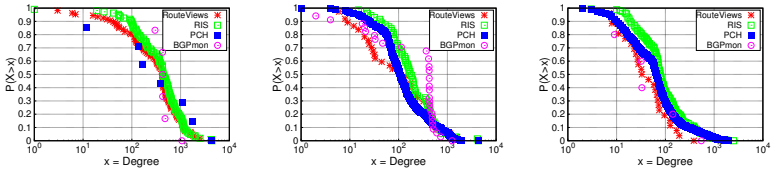


Figure 7: Connectivity scenario I



(a) Full feeders **(b)** Partial feeders **(c)** Minor feeders

Figure 8: CCDF of the node degree distribution per class of feeders

feeders. However, given the low number of ASes that are either part of the IPv4 or the IPv6 set of full feeders, the set of full feeders used in this work from now on is considered to be as made up of the union of the two sets, which consists of 192 different ASes (16.81% out of the total number of different feeders).

One feature that can be used to gain a deeper insight into these different classes of feeders is their node degree distribution (Fig. 8). This is computed on the union of the AS-level topologies inferred from BGPmon, PCH, RIS and RouteViews datasets. In each project the full feeder set mainly consists of ASes that have developed a large number of BGP connections, which is typical of transit ISPs. To confirm this, the nature of these ASes has been analysed by browsing their websites and parsing their entries in the regional IRRs, and the vast majority of full feeders were shown be large ISPs. Furthermore, from this analysis results that 14 of the 18 ASes listed as being provider-free[5] are currently feeding the RCs. Since the vast majority of full feeders are large ISPs, the AS-level view of the Internet extracted from RCs is more likely to represent the Internet viewed by some of the most important ISPs in the world rather than the complete Internet AS-level topology. In fact, a view of the Internet from the top of the AS hierarchy is not able to discover a large number of connections. Due to BGP export policies, a RC connected with ASes that are part of the top of the hierarchy does not reveal all the p2p

[5] A list of provider-free ASes is available at [Wik]. The list used in this work was collected on February 15, 2014.

	BGPmon	PCH	RIS	RouteViews
# of nodes	47,024	47,015	47,164	47,167
# of edges	113,917	105,453	177,617	159,574
# of common edges	96,504			
Union				
# of nodes	47,246			
# of edges	209,456			

Table 3: AS-level topology characteristics (February 2014)

connections that are established at the lower levels. On the other hand, the lower in the hierarchy the feeder is located, the greater the chances to gather information about an AS path involving a previously hidden p2p connection. Consider for example Fig. 7. In this case, if the RC `R` is connected to AS `E` located at the top of the economic hierarchy, it cannot reveal either the p2p connection between `A` and `B`, or the p2p connection between `C` and `D`. On the other hand, if `R` is connected to `A`, it can reveal the p2p connection between `A` and `B`, but not between `C` and `D`. In addition, it is fundamental that the RCs establish a c2p relationship with their feeders, playing the role of customer. Otherwise, even if the RC is connected to `A`, the connection `(A,B)` will not be revealed. A real example of the importance of obtaining the full routing table from BGP feeders located in the lowest part of the Internet hierarchy is represented by PCH. This data source is potentially extremely useful for discovering hidden AS connections, since its RCs are deployed on 65 different IXPs and connected to 979 ASes. This is about three times the total number of feeders of RIS (289) and RouteViews (149) – many of which have small node degree value (Fig. 8), which is a *rough* indication of their location at the bottom of the Internet hierarchy. Note however that, since minor feeders announce only partial routing information to the route collector, an analysis of their degree distribution does not provide as reliable results as those inferred from the degree distribution of the full feeders, thus this latter conclusion about the nature of PCH feeders must be taken with a grain of salt.

Anyway, as shown in Table 3, the number of AS connections[6] de-

[6]The AS-level topology is extracted from AS paths gathered by the respective projects and cleaned as described in Section 2 of [GIL+11].

tected by PCH and not discovered by RouteViews and RIS is extremely low, since 96,504 connections out of the total 105,453 connections discovered – i.e. 91.51% of the total number of PCH connections – have already been revealed by the other projects. This happens because PCH establishes only p2p connections with its BGP feeders except with its providers, i.e. its RCs obtain only routes concerning prefixes owned directly by their feeders or announced by their feeder customers. Consequently, it is likely that almost every connection found by PCH represents a p2c (c2p) economic agreement, thus failing to discover all those p2p connections that represent the largest set of hidden connections [HSFK09, CR06, AKW09a] and greatly limiting the topology discovery potential of PCH RCs.

A deeper insight into the amount of information provided by each feeder can be gained by analysing the difference between the direct node degree and the inner node degree (Fig. 9). The *direct* node degree of a feeder X is defined as the cardinality of the set of its neighbors that are discovered using only BGP data directly announced by X to a RC, and the *inner* node degree of X is defined as the cardinality of the set of its neighbors that are discovered using BGP data announced by every feeder but X. A similar approach was proposed in [DCDC12], but with a different purpose. With this metric it is possible to differentiate between two different classes of feeder behaviors: a) ASes that only announce a partial view of the Internet (degree difference < 0) such as those ASes that consider the RCs as peers and not as customers, and b) ASes that contribute with connections not contained in any AS path announced by other feeders (degree difference > 0), such as p2p and p2c connections that are hidden from other feeders due to the effects of BGP export policies crossed during the propagation of the routing information. The first class is typical of minor feeders, whose connectivity is mostly discovered via other feeders. On the other hand, the second class is typical of full feeders, which typically introduce previously undiscovered AS connections. However, some of the full feeders partially hide their connectivity despite advertising their full routing table to at least one RC. This phenomenon happens in about 20% of the full feeders (see negative values of

degree difference in Fig. 9) and is caused by the BGP decision process on the feeder side. Depending on the policies established among ASes and on technical decisions, some direct connections may not be announced to the RC. For example, as highlighted above, some ASes may decide to announce only their IPv4 (IPv6) full routing table, and not their IPv6 (IPv4) full routing table to the RC, possibly hiding some connections. However it is possible that the those hidden connections are included in routes advertised to other neighbors, then propagated on the Internet and finally detected by some RC connected to another feeder. Other slight exceptions are related to minor feeders with a positive degree difference (less than 5%). These feeders hide part of their p2c connectivity from the RCs to which they are not connected because of the cross effect of their multi-homed nature and of multiple BGP decision processes crossed by `UPDATE` messages. More in detail, the presence of multiple BGP decision processes along the AS path may limit the completeness of the AS-level topology collected, since each AS Border Router (ASBR) selects and announces only the best route per-destination [RFCb] to their neighbors. In summary, the information that a feeder announces to the RC is the re-

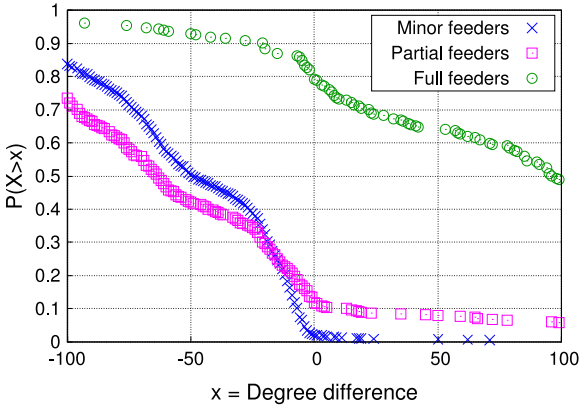


Figure 9: CCDF of the degree difference of feeders

Region	Feeders		Full feeders	
	IPv4	*IPv6*	*IPv4*	*IPv6*
Africa (AF)	32	8	2	2
Asia-Pacific (AP)	124	75	24	22
Europe (EU)	727	383	99	79
Latin America (LA)	53	54	29	21
North America (NA)	356	130	61	37
World (W)	1,132	545	167	119

Table 4: Geolocation of feeders

sult of its BGP decision process which, in turn, is fed only with routes that are the result of the BGP decision processes of its neighboring ASes, and so on. Each BGP decision process, from an AS-level measurement perspective, is a route filter which can potentially reduce the AS-level connectivity information received from each RC. As a consequence, the higher the distance of an AS from the RCs, the higher the number of BGP decision processes crossed and, thus, the probability that one (or more) of them will filter out some AS connections.

2.4.1 Geographical coverage

The BGP data incompleteness is even stronger if analysed from a geographical perspective. Table 4 details the total number of feeders as well as the number that supply the full routing table to any of the RCs. To perform this analysis, the IP address of each feeder has been geolocated using the Maxmind GeoIPLite database [Max] and considering the world as being subdivided into five macro-regions: Africa, Asia-Pacific (i.e. Asia and Oceania), Europe, Latin America (the Caribbean, Central America, Mexico and South America) and North America. Table 4 highlights that most full feeders are located either in Europe or in North America. Interestingly, in Africa only three full feeders can be found (i.e. the union of IPv4 and IPv6 Africans full feeders), even though Africa hosts two RouteViews RCs and three PCH RCs. This means that every inference regarding the African part of the Internet is mostly obtained through views located in different regions. Thus, some relevant characteristics of the African Internet may be hidden from the current RCs, e.g. most of African p2p connectivity. This is not only a problem regarding

Africa, in fact the number of feeders in other regions is also low, compared with the total number of ASes of the Internet.

Chapter 3

Completeness analysis

3.1 A novel methodology to deal with BGP data incompleteness

Given the high level of incompleteness of the BGP data shown in Section 2.4 (page 12), the first step to infer an Internet AS-level topology closer to reality is to introduce a larger number of feeders that announce their full routing tables. One of the biggest obstacles is the vast number of ASes that make up the Internet. Obtaining routing data from each one would require the participation of thousands of administrators in a project that might not be appealing to many of them, and would be practically unfeasible. In this section, it is designed a methodology that enables to identify the minimal set of ASes that need to become full feeders in order to gather as much BGP data as possible about the ASes part of the core of the Internet, i.e. the non-stub ASes. The same methodology also provides an additional ranking list of candidate feeders to enable any RC project to identify which ASes are the best candidates to maximize their coverage with a limited amount of resources.

3.1.1 A new metric: p2c-distance

As already highlighted in Section 2.4, the set of routes announced by an AS to a neighbor depends on the economic relationship regulating their BGP session. Given the BGP export policies related to each economic relationship [Gao01b], a necessary – but not sufficient – condition for a RC to reveal the full connectivity of a given AS is that at least one AS path should exist that is made up exclusively of p2c connections from that AS to the RC. This because only customers in a p2c connection are able to obtain routes towards every Internet destination. It is also preferable that routing data arrives at a RC crossing the lowest number of p2c connections as possible, in order to limit the filtering effects of BGP decision processes.

On the basis of the two conditions hypothesized above, here is defined a new metric which is able to quantify the amount of ASes whose connectivity can be fully discovered by the current set of RCs. The *p2c-distance* of one AS `X` from another AS `Y` is the minimum number of consecutive p2c connections that connect `X` to `Y` in the considered economic topology or, likewise, the minimum number of consecutive c2p connections that connect `Y` to `X`. The p2c-distance is not defined if AS `Y` is not in the *customer cone*[1] of `X`. Note that an s2s connection is considered both as a p2c and c2p connection between the sibling ASes. If this metric is computed in terms of the p2c-distances of ASes towards the RCs, it can identify whether the RC has at least one possibility to discover the full connectivity of an AS and, at the same time, quantifies the the amount of BGP decision processes crossed by any `UPDATE` message to reach a RC. It can also reveal which part of the Internet is well-monitored and which part is still a dark zone. Note that this metric still relies on an inference made on the AS-level topology, thus it may be inaccurate due to the incompleteness of data shown in Section 2.4. However, in [LHD+13] it has been proved that most of the economic algorithms developed so far are quite reliable in inferring p2c connections. This is because most of the algorithms rely on the presence of a tangible proof in AS paths to infer

[1] The customer cone of an AS X is defined as the set of ASes that X can reach using p2c links [LHD+13].

p2c relationships (e.g. provider-free AS presence [OPW+10]), while p2p connections are usually inferred by exclusion.

To better understand how this metric works, consider the connectivity scenario depicted in Fig. 7 (page 14). In this example, the RC `R` has a p2c-distance of 1 from AS `A` and `E`, while the p2c-distance from `B`, `C`, `D` and `F` is not defined. This means that `R` has the possibility of revealing all p2p connections established by `A` and `E`. On the other hand, it also means that `R` is not able to reveal the p2p connectivity of `B`, `C`, `D` and `F` in any way, thus `R` will not reveal the connection `(C, D)` in any AS path. Nevertheless, `R` can discover the p2c (c2p) connectivity of each AS in the scenario.

3.1.2 Feeder selection

Given the definition of p2c-distance, a complete view of the Internet can only be obtained by connecting a RC to each stub AS, as already concluded in [OPW+10]. Stub ASes are ASes that are typically managed by local access providers – which provide connectivity to end users but not to other ASes – and organizations that do not have the Internet transit as part of their core business (e.g. banks and car manufacturers), and appear in BGP data as the right-most element in every `AS PATH` attribute that involves them. Due to the nature of their organizations, these ASes tend to be customers in inter-AS economic relationships, representing a perfect starting point to minimize the p2c-distance of every AS that make up the Internet. However, since p2c connections are already discovered by RCs connected to the top of the hierarchy [OPW+10], most BGP data collected from a hypothetical ideal RC infrastructure connected to each stub AS would be redundant. Moreover, since it is not possible to infer a priori which stub AS is actually interested in establishing p2p connections, it is impossible to reduce the number of new feeders required to obtain full Internet AS-level connectivity. This means that, based on February 2014 data, a BGP connection with each of the 38,820 stub ASes (out of 47,246 total ASes) is required. This makes this approach unfeasible. A good trade-off between the possibility of discov-

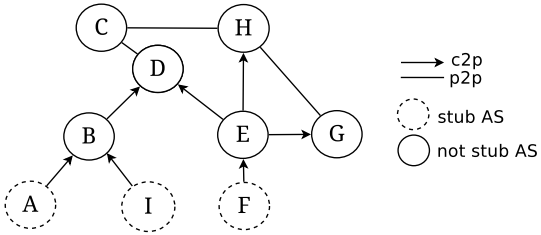


Figure 10: Connectivity scenario II

ering hidden p2p connections and the feasibility of obtaining such data is represented by those ASes that are actually interested in deploying p2p connections to improve the quality of their services, i.e. non-stub ASes [DKF+07]. The lack of interest of stub ASes in establishing p2p connections is highlighted by how few (only 7.17%) participate in at least one IXP[2], where ASes typically interconnect with settlement-free p2p connections to reduce the amount of traffic directed at their providers (see [HSFK09, AKW09a, GILO10, ACF+12] and [CRS12] for more details on IXPs). The aim of this work is to select new feeders such that each non-stub AS has a finite and bounded p2c distance from the RCs in order to minimize the effects of BGP filters and, consequently, to increase the possibility of revealing the hidden p2p connectivity of the actual core of the Internet. stub ASes

3.1.3 Identifying the feeders

The first part of the methodology focuses on identifying the minimum number of full feeders required to obtain a complete[3] view of the Internet core. This can be modeled as an MSC problem that can be described through the following Integer Linear Programming (ILP) formulation:

[2]The set of ASes that participate in at least one IXP was collected by downloading and parsing the participant list web page of 242 IXPs that were found to be active February 15th, 2014. This kind of data has been collected once per month since February 2012 and is publicly available at [Iso].

[3]Note that in this methodology it is assumed that every AS in the Internet has a full routing table and it is willing to advertise it to the RCs.

$$\text{Minimize} \qquad \left(\sum_{AS_i \in \mathcal{U}} x_{AS_i} \right) \tag{3.1}$$

subject to

$$\sum_{AS_j : AS_n \in S^{(d)}_{AS_j}} x_{AS_j} \geq 1 \qquad \forall AS_n \in \mathcal{N} \tag{3.2}$$

$$x_{AS_i} \in \{0, 1\}, \qquad \forall AS_i \in \mathcal{U} \tag{3.3}$$

where $\mathcal{U} = \{AS_1, AS_2, \ldots, AS_n\}$ is the set of all ASes making up the Internet, $\mathcal{N} \subset \mathcal{U}$ is the set of non-stub ASes and $S^{(d)}_{AS_i}$ represents the *covering set* of AS_i at fixed p2c-distance d, i.e. the set of ASes in $\mathcal{N}$ that have a p2c-distance value of at most d from AS_i. The goal of this MSC problem is, given d, to obtain the minimum number of elements of $\mathcal{S}^{(d)}$ such that their union is $\mathcal{N}$ or, in other words, to select the minimum number of feeders from $\mathcal{U}$ such that the p2c-distance of any non-stub AS from at least one feeder is at most d. The parameter $d \geq 0$ defines the maximum number of BGP decision processes[4] that the `UPDATE` messages generated by each non-stub AS will traverse before reaching a potential feeder and, thus, indicates the number of filters encountered that can cause loss of information. Note that AS_i belongs to $S^{(d)}_{AS_i}$ for any fixed d. Furthermore, x_{AS_i} is 1 if $S^{(d)}_{AS_i}$ is part of the final solution, 0 otherwise. In other words, the problem is to minimize the number of ASes (3.1) required to cover each non-stub AS (3.2) and, thus, there is at least a chance to discover its complete connectivity from a RC. Note also that imposing $d = 0$ implies that the solution is composed of the entire set of non-stub ASes. The larger the value of d, the heavier the filtering effects introduced by BGP decision processes, however the smaller the number of required BGP feeders and, thus, the number of required BGP connections. To better understand the problem, consider the scenario depicted in Fig. 10. In this example, $\mathcal{U} = \{A, B, C, D, E, F, G, H, I\}$

[4]The number of BGP decision processes encountered before reaching a RC is $d+1$, since feeders introduce an additional BGP decision process before announcing BGP data to the RCs.

and $\mathcal{N} = \{B, C, D, E, G, H\}$. Thus are computed $S_A^{(1)} = S_I^{(1)} = \{B\}$, $S_B^{(1)} = \{B, D\}$, $S_C^{(1)} = \{C\}$, $S_D^{(1)} = \{D\}$, $S_E^{(1)} = \{E, D, G, H\}$, $S_F^{(1)} = \{E\}$, $S_G^{(1)} = \{G\}$, $S_H^{(1)} = \{H\}$. One of the optimal solutions to cover every non-stub AS is $\{S_B^{(1)}, S_C^{(1)}, S_E^{(1)}\}$, i.e. ASes `B`, `C` and `E` should be selected as new feeders.

3.1.4 Solving the MSC problem

MSC problems are known to be NP-hard [Hoc97], however that does not mean that it is always practically impossible to obtain their optimal solution [Woe03]. Several techniques to reduce the problem into smaller problems were developed in the last century in various research fields, and can be particularly effective depending on the nature of the problem. In this case, it is particularly effective to apply iteratively the concepts of *essentiality* and *dominance* described in the Quine-McCluskey procedure ([McC56, Qui55, Qui59]) on the covering matrix related to the described MSC problem. The *covering matrix* of a given MSC problem is the Boolean matrix in which each row represents a covering element and each column represents an element that has to be covered. A generic element (i, j) of that matrix thus contains 1 if the element placed at row i covers the element placed at column j, and is 0 otherwise. In this work, each row in the covering matrix represents an AS in $\mathcal{U}$, and each column represents an AS in $\mathcal{N}$. Consequently, its size is $|\mathcal{U}| \times |\mathcal{N}|$, and each element (i, j) in the matrix contains 1 if AS i covers AS j, i.e. $S_{AS_i}^{(d)}$ contains AS j. The size of this matrix can be then reduced by applying the following techniques:

Essentiality. *A row in the covering matrix M is defined as essential iff it is the only row covering a given element. Consequently, an AS is essential iff its covering set contains a non-stub AS covered only by the AS itself. More formally, $AS_x \in \mathcal{U}$ is essential iff $\exists AS_n \in S_{AS_x}^{(d)} : AS_n \in (S_{AS_x}^{(d)} \setminus \bigcup_{y \neq x} S_{AS_y}^{(d)})$*

Dominance. *A row of the covering matrix M is defined as dominated by another row iff every element covered by the considered row is also covered by the dominating row. Consequently, AS_x dominates AS_y iff the non-stub ASes*

covered by AS_x are also covered by AS_y. Formally, given $x, y \in rows(M)$, x dominates y iff $S^{(d)}_{AS_y} \subseteq S^{(d)}_{AS_x}$, and $rows(M)$ is the set of rows of matrix M.

Using essentiality and dominance it is possible to devise a technique to retrieve an optimal solution for the MSC problem. The procedure consists of the following four phases: *phase a*) Selection of essential covering sets, *phase b*) Deletion of dominated covering sets, *phase c*) Exhaustive search on the remaining covering matrix, and *phase d*) Selection of ASes that are part of at least one optimal solution. Details of the procedure are depicted in Fig. 11.

The aim of phase *a*) of the procedure is to find *essential* ASes, that are required as part of the final solution (lines 6–9). Indeed, whenever an AS X is found to be essential, this means that X is the only AS to cover at least one non-stub that is required to be covered, and has to be part of the solution. Every AS contained in $S^{(d)}_X$ then can be considered as covered (line 9), thus reducing the columns of the covering matrix. Then, in phase *b*), the procedure tries to find if any AS is dominated by another AS (lines 10–18). In this case, the row corresponding to the dominated AS is deleted from the covering matrix. Note that an AS is considered to be dominated by another AS even if the two ASes have the same covering set. In this case, one of the two ASes is randomly chosen to continue the procedure. Every dominated AS is maintained in a separate set together with the related dominating AS in order to check their presence in at least one optimal solution during the final phase. Essentiality and dominance foster each other, since essentiality reduces the columns of the matrix on the basis of the available rows, whereas dominance reduces the rows of the covering matrix on the basis of the available columns. For this reason phases *a*) and *b*) are iteratively applied until a cyclic core is found [Cou94] or, in other words, the size of the covering matrix cannot be reduced any further 4).

In this work, essentiality and dominance were particularly effective due to the economic nature of the Internet and the way the covering matrices are populated. The Internet is mainly made up of by organizations whose economic markets are not driven by Internet traffic transit services

```
Input: Covering matrix M of size |𝒰| × |𝒩|

  O = null matrix of size |𝒰| × |𝒩|
  while M ≠ O
    O = M
    foreach row r ∈ M
      if AS_r is essential
        insert AS_r in 𝒫
    delete from M each column c covered by AS_r
    foreach row r ∈ M
      foreach row s ∈ M ∧ s ≠ r
        if AS_r dominates AS_q
          delete AS_q from M
          record <AS_r, AS_q> in 𝒯
        else if AS_q dominates AS_r
          delete AS_r from M
          record <AS_q, AS_r> in 𝒯
          break

  if M is not empty
    partition M in a block matrix
    foreach block b_i ∈ M
      compute 𝒫_{b_i} applying exhaustive search on b_i
      add 𝒫_{b_{i_0}} to 𝒫
      insert 𝒫_{b_{i_{j>0}}} in ℐ

  insert 𝒫 in ℐ
  foreach <dominating AS_r, dominated AS_q> in 𝒯
    if S_𝒫 = S_{𝒫 − AS_r + AS_q}
      insert AS_q in ℐ

  Output:
    solution set 𝒫
    set of elements in a possible optimal solution ℐ
```

Figure 11: MSC reduction procedure

[GIL+11] and which are typically located in a single country [Nor11]. Moreover, national Internet markets are typically closed [Nor11] and, as a consequence, the number of providers present in each country is limited. The (relatively) poor heterogeneity of choices that an AS located in a single country has in choosing its providers means that several ASes have a common set of providers, and thus several rows in the covering matrix are similar, which often leads to dominating/dominated pair of

rows. This large similarity between rows and the low density of the p2c-distance covering matrices allows to reduce the problem greatly, often finding an optimal solution by just applying essentiality and dominance iteratively. Consider for example the scenarios related to February 2014. In this scenario, the original covering matrix sizes were $47,246 \times 8,426$, and the matrix densities[5] were 0.0003 ($d = 1$), 0.0052 ($d = 2$) and 0.042 ($d = 3$). The reduction technique just depicted shrank the covering matrices respectively to a cyclic core of 40×40 ($d = 1$), 9×13 ($d = 2$) and 33×49 ($d = 3$).

If a non-empty cyclic core is found, the problem could be solved via an exhaustive search or by applying a branch-and-bound technique. In the scenario of February 2014, this would entail computing $\sum_{k=1}^{40} \binom{40}{k}$ comparisons. However, it is still possible to reduce the problem (phase *c*) by exploiting the Internet regionality characteristic [Nor11] to partition the remaining covering matrices into diagonal blocks (line 21) by opportunely permuting their rows and columns using standard mathematical techniques. This enables the remaining covering problem to be divided into disjoint and smaller subproblems that can be solved independently [OC01] via an exhaustive search (lines 22–24), computing for each block b_i the set of every possible solution $\mathcal{P}_{b_i}$. The final optimal solution is then represented by the union of the solution found so far and one of the solutions found per block. Consider again the February 2014 scenario and focus on $d = 1$, which showed the largest cyclic core. In this case, the cyclic core is decomposed into nine sub-matrices – each composed by elements located in the same country – whose maximum size was 8×8, which can be solved with a maximum of $\sum_{k=1}^{8} \binom{8}{k}$ operations via an exhaustive search. Given the results obtained, there is no need to further complicate the procedure with additional reducing steps, although it is still possible to reduce the size of cyclic cores using Gimpel's technique [Gim65].

Finally, in phase *d*) it is checked whether any of the dominated elements recorded during the procedure could be part of at least one opti-

[5]The *matrix density* is defined as the ratio of non-zero entries to total number of entries in a matrix.

mal solution (lines 28–30) using the following lemma and corollary:

Lemma 1. *An element AS_q that is dominated by an element AS_r – with $AS_r \in \mathcal{P}$ and not part of the cyclic core – is part of at least one optimal solution only if the solution $\mathcal{P}^*$ obtained by swapping AS_q with AS_r in $\mathcal{P}$ is still an optimal solution, i.e. $|S_{\mathcal{P}}| = |S_{\mathcal{P}^*}|$, where $\mathcal{P}^* = \mathcal{P} - AS_r + AS_q$.*

Proof. By hypothesis, if $|S_{\mathcal{P}}| \neq |S_{\mathcal{P}^*}|$, the space of elements covered uniquely by AS_r in $\mathcal{P}$ that made AS_r essential in the MSC procedure – i.e. $S_{\mathcal{P}} \setminus S_{\mathcal{P}-AS_r}$ – cannot be fully covered by AS_q. Now, consider the solution $\mathcal{P}'$ obtained by forcing AS_q in the initial solution set and by applying the MSC procedure once again. Since the MSC procedure is devised to retrieve an optimal solution ([McC56, Qui59]), it is also able to retrieve the solution with the minimal cardinality that involves AS_q. Since AS_q was found to be dominated by AS_r in the original MSC procedure, this means that the space of elements $S_{AS_q} \setminus S_{AS_r}$ was found to be covered by essential elements. Thus, even forcing AS_q in solution these elements are still going to be selected as essentials by the MSC procedure, since S_{AS_q} does not include the elements that characterize their essentiality. In other words, the space of elements $S_{AS_q} \setminus S_{AS_r}$ is covered redundantly and does not justify the presence of AS_q in an optimal solution. The remaining covering space $S_{AS_q} \cap S_{AS_r}$, by hypothesis, covers the space of elements $S_{\mathcal{P}} \setminus S_{\mathcal{P}-AS_r}$ only partially, which has to be covered by an additional AS, i.e. AS_r itself or one of the elements that were found to be dominated during the original MSC procedure. Thus, the cardinality of $\mathcal{P}'$ has to be larger than $\mathcal{P}$, since there is an element (AS_q) that does not uniquely cover any space of elements in $\mathcal{N}$. □

Corollary 1. *An element AS_q that is dominated by an element AS_r – with AS_r part of block b_i of the cyclic core – is part of at least one optimal solution only if there is at least one solution $\mathcal{P}_{b_{i_j}}$ found via an exhaustive search on b_i where the solution $\mathcal{P}^*_{b_{i_j}}$ obtained by swapping AS_q from $\mathcal{P}_{b_{i_j}}$ with its dominating element AS_r is still an optimal solution, i.e. $|S_{\mathcal{P}_{b_{i_j}}}| = |S^*_{\mathcal{P}_{b_{i_j}}}|$, where $\mathcal{P}^*_{b_{i_j}} = \mathcal{P}_{b_{i_j}} - AS_r + AS_q$ and $0 \leq j < |\mathcal{P}_{b_i}|$.*

As a result of this procedure, the algorithm provides: i) a set $\mathcal{P}$ of ASes made up of the set of ASes that were inserted into the solution during phase a) and of one of the solutions per each block found $\mathcal{P}_{b_i}$, and ii) a set $\mathcal{I}$ of ASes containing every AS that can be part of at least one optimal solution.

```
Input:
Covering matrix M of size |I| × |N|
---------------------------------------------------
  C = ∅
  while (C ≠ N)
    AS_r = row covering the largest number of columns
    append AS_r to R
    foreach row s ∈ M
      if S_{AS_q} = S_{AS_r}
        insert AS_q in R as alternative to AS_r
    delete from M each column covered by AS_r
    C = C ∪ S_{AS_r}
---------------------------------------------------
  Output:
    Ranking list R
```

Figure 12: Greedy heuristic

3.1.5 Ranking the candidates

The solution to the MSC problem on its own provides only a quantification of the number of feeders required to obtain an optimal coverage of the Internet core, but does not help much in understanding to what extent the coverage would be improved with just a limited set of feeders. This is because the aim of the MSC problem is to completely cover the set of non-stub ASes, and at each step the ASes are selected due to the *essentiality* concept described earlier, rather than choosing the ASes that would maximize the partial coverage. The results obtained from the MSC problem have thus a theoretical importance, since they enable the current coverage of the RCs to be quantified and to identify which ASes should join them. At the same time these results have a rather low practical importance, since it is almost impossible to connect each feeder found to a RC. In the second part of the methodology it is devised a procedure to rank each AS found to be part of at least one optimal solution of the MSC problem, i.e. belonging to $\mathcal{I}$, to provide any RC project with ordered list of ASes that should join them in order to maximize their coverage with limited resources.

This result can be obtained by using a greedy algorithm [CSRL01] to

solve a particular Maximum Coverage (MC) problem restricted to the elements found to be part of $\mathcal{I}$, i.e. the rows of the covering matrix of this MC are the rows corresponding to the ASes belonging to set $\mathcal{I}$. The MC problem can be described through the following ILP formulation for every given $k > 0$:

$$\text{Maximize} \qquad \left(\sum_{AS_j \in \mathcal{N}} y_{AS_j} \right) \tag{3.4}$$

subject to

$$\sum_{AS_i \in \mathcal{I}} x_{AS_i} \leq k \tag{3.5}$$

$$\sum_{AS_i \in \mathcal{I} \wedge AS_j \in S_{AS_i}} x_{AS_i} \geq y_{AS_j}, \qquad \forall AS_j \in \mathcal{N} \tag{3.6}$$

$$y_{AS_j} \in \{0, 1\}, \qquad \forall AS_j \in \mathcal{N} \tag{3.7}$$

$$x_{AS_i} \in \{0, 1\}, \qquad \forall AS_i \in \mathcal{U} \tag{3.8}$$

where $\mathcal{U}$, $\mathcal{N}$, $\mathcal{I}$ and $S^{(d)}_{AS_i}$ represent respectively the set of ASes, the set of non-stub ASes, the set of ASes part of at least one MSC optimal solution and the covering set of AS_i – x_{AS_i} is 1 if $S^{(d)}_{AS_i}$ is part of the MC problem solution, 0 otherwise, and y_{AS_j} is 1 if AS_j is part of the final coverage, 0 otherwise.

An *exact* solution $\mathcal{E}_k$ of this MC problem is the set of k ASes (3.5) that covers the largest set of non-stubs (3.4) chosen from the output of the MSC problem (3.6). However, exact solutions cannot be used to retrieve the desired AS ranking list, since $\mathcal{E}_k$ may differ from $\mathcal{E}_{k-1}$ by more than one element, and an AS that is part of $\mathcal{E}_{k-1}$ may no longer be part of $\mathcal{E}_k$.

Instead of looking for exact solutions, the greedy heuristic and the concept of dominance described above are exploited to obtain an approximate solution $\mathcal{G}_k$ which can be interpreted as the first k ASes that should be added to the RCs on the basis of their potential contribution to the coverage. In detail, for a given k, the greedy heuristic consists in k steps, selecting the AS at each step that covers the maximum number of non-stubs currently uncovered. This means that $\mathcal{G}_k$ is obtained by adding to

$\mathcal{G}_{k-1}$ the AS selected by the heuristic at step k. It must also be noted that this approach is proved to be *at least* a $(1 - \frac{1}{e}) \approx 0.632$ approximation of $\mathcal{E}_k$ [Hoc97].

Figure 12 depicts the pseudo-code of the greedy algorithm contextualized into the framework of this work. Given the set $\mathcal{I}$ of ASes, the heuristic selects at each step the AS that covers the largest number of non-stubs and adds it in the ranking list $\mathcal{R}$ (lines 6-7). Then, it searches for alternatives to the selected AS, i.e. ASes covering the same set of non-stub ASes currently covered by the selected AS (line 10). Finally, the remaining covering sets are updated by deleting the non-stub ASes covered by the selected AS (line 11). Instead of stopping when a given bound k is reached, as described in the classical MC formulation, the proposed algorithm stops when *all* the non-stubs ASes are (line 5). Note also that, each time an AS is added to the ranking list, it is possible to keep track of the percentage of non-stubs covered, thus the ranking list can also be used to identify which ASes would need to become feeders in order to cover a given percentage of non-stub ASes.

3.2 Methodology limitations

The proposed methodology assumes that the selected feeders can always provide a default-free BGP flow to the RCs. This is not always the case. Several ASes obtain (only) a default route from their providers due to the limited capacities either because of their routing infrastructure or because they do not need anything more specific. Those ASes are not physically capable of being full feeders. In addition, the methodology assumes that the routing table announced by an AS is representative of the BGP routing inside the whole feeding AS. Some organizations however split their AS into different routing islands for traffic engineering purposes [GIL+11], often related to geographical constraints (as it has been outlined in Chapter 5). There are special cases though. Particularly significant is the case of AS 3557, which manages the f-root nameserver. This AS owns at least 50 independent networks all over the world and announcing the same set of prefixes, which appears as one network thanks

to the anycast technique [fro]. Each of these networks then establishes a BGP session with different providers depending on the region considered. This means that, even if AS 3557 is perceived as a node located in five different regions, five different BGP sessions would not be enough to discover its full connectivity. It should rather be considered as the composition of many different and independent ASes, each of which should be a feeder of an RC to get the full coverage of AS 3557 providers. All of these must be taken into account while using the results of the methodology proposed. In any case it is possible to re-compute the results by explicitly excluding those feeders which, for any reason, are not able to provide the default-free BGP flow(s) required to cover its providers to the RCs.

Chapter 4

Economic Tagging

4.1 Economic relationships

In this thesis, the ability to infer economic relationships among ASes will be exploited to compute p2c-distances and, then, to obtain results of the MSC problem described in Section 3.1.3 (page 24). However, their knowledge is also important outside of this scope, since they allow to shed light about the role that each AS is playing in the Internet ecosystem.

This is technically implemented by applying outbound route filters described via BGP import/export policies. Despite the large number of possible economic agreements, inter-AS relationships can still be categorized by three main classes on the basis of the set of routes that each AS announces to the other: provider-to-customer (p2c) – or customer-to-provider (c2p) – peer-to-peer (p2p) and sibling-to-sibling (s2s) [Gao01b]. In p2c and c2p relationships, the provider announces to the customer the routes required to reach each Internet destination, by selecting them from the routes that the provider obtained from its customers, providers and peers (if any) plus the routes owned by the provider itself. The customer, on their part, announces to the provider only those routes related to the customer's own IP prefixes and routes obtained from its customers (if any). In p2p relationships, each AS announces to the other peer the routes related to its own IP prefixes and the routes obtained from its cus-

tomers, typically free-of-charge. Finally, in s2s relationships each AS acts like a provider by announcing its full routing table. This level of granularity is typically considered as consistent with reality [RWM+11b], although there are some exceptions [MFM+06] mainly caused by policies established on a geographical basis [Nor11]. For example ASes may agree to establish a *partial transit*, in which the provider propagates the customer routes only a particular region of the world or to a particular set of its neighbors[LHD+13].

The existence of these types of relationships implies that the largest amount of graph paths that can be extracted from the Internet AS-level topology do not exist in reality. This means that the bare undirected topology cannot be used to accurately study or model the real routing behavior of the Internet. Thus, knowledge of inter-AS economic relationships plays a fundamental role, and any realistic Internet AS-level analysis has to take these relationships into account. The common approach is to transform the undirected AS-level topology into an economic AS-level topology where each edge is tagged with a proper economic label which reflects the type of relationship existing between the involved ASes. Despite (or due to) their key role, details regarding inter-AS economic relationships are not usually publicly available, and researchers have needed to develop heuristics to infer them.

The first work on tagging the Internet AS-level topology was [Gao01a], which proposed applying a heuristic on public BGP routing information to infer economic relationships between ASes. The heuristic was based on the fact that routes that two ASes exchange must reflect the economic relationship between them, and that a provider typically has a larger node degree[1] than its customers, while two peers typically have comparable degree. In [Gao01a] it was also proved that if *all* ASes respect the export policies imposed by such economic relationships, then the AS path in any BGP routing table must be *valley-free*, i.e. after traversing a p2c or p2p, the AS path cannot traverse a c2p or p2p. Later, [SARK02b] formulated the problem of assigning a tag to each connection as an optimiza-

[1]The node degree of a vertex of a graph is the number of edges incident to the vertex. In this context case, the degree indicates the number of BGP neighbors of an AS.

tion problem, the Type of Relationships (ToR) problem, using the number of *valley-free* paths as an objective function and proposed a heuristic to solve it. The ToR problem has been proven to be NP-complete [BPP03, EHS02], thus several authors have proposed different heuristics and enhancements to resolve it, for example [BPP03, EHS02, DKH+05, KMT06].

Other interesting approaches in the tagging issue were developed in [XG04] and in [OPW+10]. The algorithm proposed in [XG04] started from a partial set of information about the relationships between ASes, inferred using the BGP COMMUNITY attribute and from a set of information gathered through the IRR databases in order to obtain an entire set of AS relationships. However there is not a standard in using the BGP COMMUNITY attribute that could lead to a systematic method to extract useful information, and data available in IRRs have no guarantees regarding completeness and freshness. The algorithm proposed in [OPW+10] was based on the fact that BGP monitors at the top of the routing hierarchy, i.e. monitors at Tier-1 ASes, are able to reveal all the downstream p2c over time, assuming that routes follow a no-valley policy.

The main problem in inferring AS-relationships from BGP data collected by route collectors is that, besides its incompleteness, these data contains several *spurious* entries caused by router misconfigurations [MWA02]. These entries show up during BGP path exploration phenomena [OZP+06] and can potentially affect the accuracy of the inferences drawn. Despite some of the heuristics used to try to minimize the impact of these routes by limiting the impact of short-lived routes [GIL+11, OPW+10], no definitive solution has been developed yet. In this section it is proposed a methodology to fill this gap by introducing a preliminary data hygiene phase where spurious routes are identified and purged from the BGP data available. In addition, a a threshold-free tagging heuristic is designed, where cleaned BGP data are used to tag each connection with an economic label. The data hygiene phase is designed to remove spurious routes which may affect the accuracy of the inter-AS economic inferences. Consequently, it is not aimed at removing all general spuriousness in BGP data.

The rest of this chapter is organized as follows. Sections 4.2 describes the causes and the extent of *spurious* routes that may affect the correctness of AS-relationships inference. In Sections 4.3 a methodology aimed to remove those spurious entries is proposed and an economic tagging algorithm to infer AS-relationships from cleaned AS paths is designed. Section 4.4 presents the results.

4.2 BGP misconfigurations and inferences

Entries in BGP data contain AS paths announced on the Internet, however they still cannot be used in their raw form in all types of inference. Inter-AS economic relationship inferences suffer from a particular class of entries – hereafter *spurious* entries – which can lead to wrong results. These entries are typically caused by BGP export policy misconfigurations [MWA02] and show up during BGP path exploration phenomena [OZP+06], which occurs when an ASBR receives a withdrawn message. In detail, before declaring the destination network unreachable, an ASBR tries to use several different routes available in its RIB table for which a withdrawn message has not yet been received, very likely due to time propagation delays. As a consequence, during the lifetime of this process several `UPDATE` messages that contain routes not commonly used are generated and propagated. These may also include messages containing routes that violate the business relationship agreement between ASes – i.e. violate the valley-free principle described in [Gao01b] – typically due to human errors in defining BGP export policies on ASBRs [GIL+11, MWA02, OZP+06]. These spurious routes are then propagated through the Internet up to the route collectors and, thus, appear in public datasets.

Consider for example the scenario depicted in Figure 13, and assume that i) each AS selects the route with the shortest path to reach P, ii) each link has the same propagation delay, and iii) `C` applies a prefix-based outbound filtering, i.e. `C` announces to its providers every route to reach its customer networks [GIL+11, MWA02], irrespectively which neighbor the route was received from. For example in this scenario, `C`

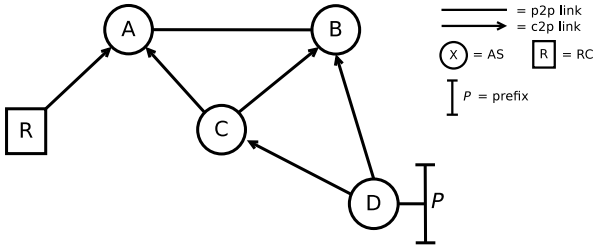


Figure 13: Path exploration example scenario

preferred path to reach prefix P is through its customer D, over the AS paths A−B−D and B−D received from its providers A and B respectively. Note however that paths A−B−D and B−D are still stored in C RIB table as alternative routes [RFCb]. Now suppose that D sends a withdrawn message containing P to both B and C due to a network failure. C will then remove from its RIB the AS path D to reach P and will search for alternatives before informing its neighbors that P is unreachable. Due to time propagation delays, C has not yet received any withdrawn message from either A or B concerning P. Thus C will select the best route from the alternatives stored in its RIB table, i.e. the route with AS path B−D. The same thing happens on B, where the chosen route is the route with AS path C−D. Since C performs an outbound filtering implemented as described above, it announces to A and B the route C−B−D to reach P, while B announces to A and C the route B−C−D. Note that the route announced by C to A is clearly in contrast with the p2c agreement signed with B. As a consequence, A may decide to use this route (e.g. if A applies the commonly used prefer-customer policy [GGR01]) and then announce to the RC R the spurious route A−C−B−D. As soon as C and B realize that loop-free[2] routes to reach P do not exist, they advertise a withdrawn message towards A. Finally, since A has no other alternatives to reach P, it will declare that the destination is unreachable, which will be recorded by R.

In other words, R sees that C is transiting traffic between its providers

[2]Due to the BGP loop avoidance mechanism, every route that contains the local AS number in the AS path attribute should not participate in the best route selection mechanism [RFCb].

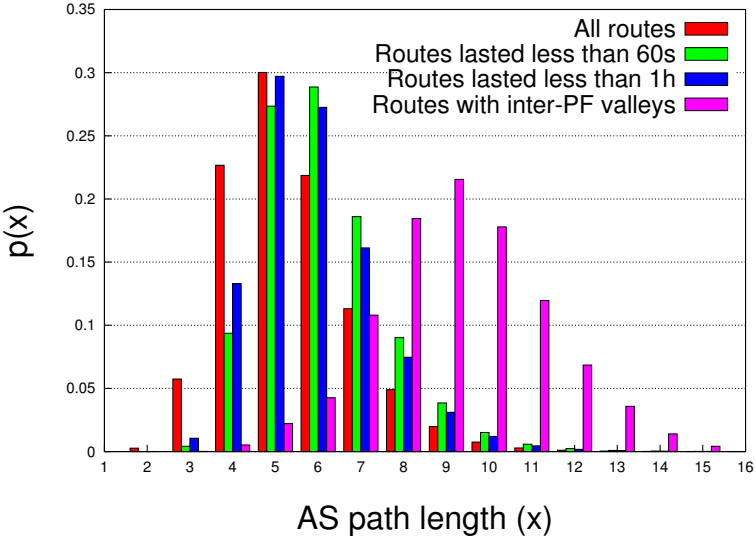


Figure 14: AS path length distribution

A and B for a short time since the network P will be withdrawn from the Internet at the end of the convergence of BGP protocol. As a result, several economic inferences made using BGP public data are potentially wrong, since AS paths contained in spurious entries interfere with those obtained from stable AS paths. Note that these spurious routes do not actually affect the network performances of the AS owner of the misconfigured ASBR due to their ephemeral nature, thus it is very likely that several misconfigurations go unnoticed by network operators. Likewise, it is unlikely (but still possible) that an AS unconsciously maintains a BGP misconfiguration for a long period of time, since an AS which transits non-planned traffic for at least one of its providers (or peers) will perceive that its network performance is degraded.

The presence of spurious entries is well-known. The common solution it is to consider every transient route as a potential source of problem for the inferences, thus applying heuristics that limit their effect on the final tagging results ([Gao01b, GIL+11, OPW+10]). This approach is not entirely correct, since not all transient routes carry no-valley-free AS paths. For example, backup routes may have a relatively short lifespan, but they carry perfectly legitimate AS paths from which previously unnoticed p2c relationships between pair of ASes may be inferred. One of the most evident characteristics which differentiates a spurious entry from a normal transient route is related to its AS path length. To prove this, Figure 14 depicts the Probability Distribution Function (PDF) of the AS path length[3] of four sets of routes, consisting of: *a*) every route, *b*) routes that lasted less than 60 seconds, *c*) routes that lasted less than one hour and *d*) routes containing easily detectable no-valley-free AS paths. Note that the four sets are not disjoint, e.g. routes in set *b* are also included in set *a* and *c*. To identify the routes belonging to set *d*, the a priori knowledge regarding the provider-free property owned by a limited set of ASes, i.e. those ASes that do not need to buy transit from any other AS to reach each Internet destination, has been exploited. A list of these ASes can be found on Wikipedia [Wik], whose validity has been discussed in [GIL+11] and [HG12]. Then, every AS path that includes two (or more) provider-free ASes not directly connected to each other has been selected, i.e. every AS path in which a third AS transited traffic for one of the provider-free ASes. Hereafter these routes are referred as *inter-PF* routes. As can be seen from Figure 14, inter-PF routes have a greater probability of having longer AS paths than routes in the other sets. This is not the general behavior of short-lived routes (sets *b* and *c*), which can be generated for other reasons (e.g. backup connections, route flapping, traffic engineering issues) and which typically carry perfectly legitimate AS paths. In addition routes belonging to set *d* are usually short-lived with respect to the total set of routes, as shown in the Complementary Cumulative Distribution Function (CCDF) depicted in

[3]Multiple consecutive ASes in every AS path (consequence of AS prepending) have been removed

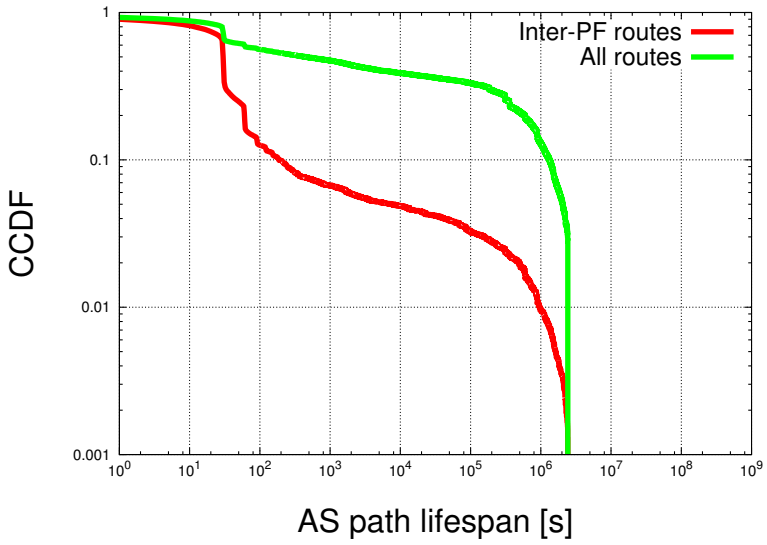


Figure 15: AS path lifespan CCDF

Figure 15, thus confirming their extremely transient nature. The largest amount of *long-lived* routes belonging to this set can still be considered as legitimate, since it largely consists of routes in which provider-free ASes are separated by a third AS which can be traced back to organizations that own one of the provider-free ASes involved. Indeed, worldwide ISPs often own more than one AS number, both due to traffic engineering (e.g. Verizon Communications manages AS 701 for North-American routing, AS 702 for European, Middle Eastern and African routing and AS 703 for Asian and Pacific region routing) and because of mergers and acquisitions [CHKW10]. Thus, it seems reasonable to *assume* that spurious routes can be identified by their abnormal AS path length and by their short-lived appearance in BGP datasets.

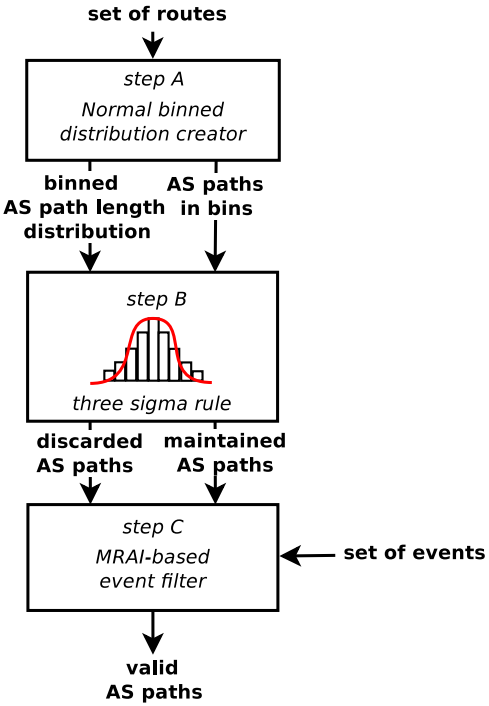


Figure 16: Data hygiene phase filters

4.3 Towards spuriousness-free inferences

To date, economic tagging algorithms have tried to mitigate the effects of spurious routes by using thresholds, either on the number of entries used to infer transit relationships [Gao01b] or on the lifespan of each route, e.g. cutting off every route lasting less than two days [OPW+10] or whose lifespan is not comparable with long-lasting lifespans [GIL+11]. In both cases, the main problem is how to select a suitable value for the threshold. This section describes a methodology which, rather than relying on arbitrary and debatable thresholds, identifies spurious entries by analyzing the data characteristics themselves and purging them from the initial dataset before inferences are drawn.

```
1  Input: feeder f, destination d, routes_{f,d}
2  ---------------------------------------------------------------
3  foreach r in routes_{f,d}
4      as_path = r.as_path
5      as_path = remove_prepending(as_path)
6      as_path_length = length(as_path)
7      length_as_paths[as_path_length].insert(as_path)
8      length_distr[as_path_length] += lifespan(as_path)
9
10 left_edge = right_edge = compute_mode(distr)
11 ranked_length_distr = sort_by_lifespan_desc(length_distr)
12 for(i = 1; i < size(ranked_length_distr); i++)
13     bin_size = right_edge − left_edge +1
14     binned_distr_{f,d} = length_in_bins = ∅
15     foreach (length, lifespan) in length_distr
16         if(length < left_edge)
17             bin = (length−left_edge)/bin_size)
18             if((length−left_edge)%bin_size)
19                 bin−−
20         if(length > right_edge)
21             bin = (length−right_edge)/bin_size)
22             if((length−right_edge)%bin_size)
23                 bin++
24         binned_distr_{f,d}[bin] += lifespan
25         as_paths_in_bins_{f,d}[bin].insert(length_as_paths[length])
26
27     1st_quartile = compute_1st_quartile(binned_distr_{f,d})
28     3rd_quartile = compute_3rd_quartile(binned_distr_{f,d})
29     if(1st_quartile == 3rd_quartile)
30         break
31     if(ranked_length_distr[i] < left_edge)
32         left_edge = ranked_length_distr[i]
33     if(ranked_length_distr[i] > right_edge)
34         right_edge = ranked_length_distr[i]
35     bin_size = right_edge − left_edge
36
37 ---------------------------------------------------------------
38 Output: binned_distr_{f,d}, as_paths_in_bins_{f,d}
```

Figure 17: Step a) Binned AS path length distribution creation

4.3.1 Preliminary data hygiene phase

Section 4.2 highlighted that spurious routes have a short lifespan and abnormally long AS paths. To identify candidate spurious routes, a three-step filter as been devised, (see Fig. 16) exploiting the following consid-

```
1  Input: f, d, binned_distr_{f,d}, as_paths_in_bins_{f,d}
2  ---------------------------------------------------------
3  μ_{f,d} = compute_mean(binned_distr_{f,d})
4  σ_{f,d} = compute_std_dev(binned_distr_{f,d})
5  foreach bin in binned_distr_{f,d}
6      if bin > μ_{f,d} + σ_{f,d}
7          D_{f,d}.insert(as_paths_in_bins_{f,d}[bin])
8      else
9          M_{f,d}.insert(as_paths_in_bins_{f,d}[bin])
10 ---------------------------------------------------------
11 Output: discarded AS paths D_{f,d}, mantained AS paths M_{f,d}
```

Figure 18: Step b) Three-sigma rule filtering

eration: an AS tends to select predominant routes to proficiently reach a destination during a month [RWXZ02, FRBM07] and, thus, predominant *AS path lengths* to reach a destination. It is thus possible to apply data binning to the time-based AS path length distribution in order to retrieve a binned normal distribution centered around the mode of the original distribution. As a consequence, AS paths that do not fall within the normal binned behavior can be marked as outliers by applying the three-sigma rule, commonly used to identify outliers in normal distributions [LF04]. More in detail, step *a*) of the algorithm (Fig. 17) aims to create the AS path length distribution experienced throughout the entire month by each pair `<feeder IP f, destination p>` and to compute the proper bin value to consider this distribution as a well-approximated normal distribution. To do this, the RIB dynamics related to the pair `<f,p>` are recreated. Then the *AS path length* experienced between the first and the last announcement of `p` every second is sampled, i.e. it is assigned the total amount of time (in seconds) to each length found in which a route with an AS path with that length is found to be active for the pair `<f,p>` during the sampling period (lines 3-8). Then the correct size of the bin is computed and the distribution just computed is transformed into a binned normal distribution centered on the mode of the original distribution, i.e. bin zero contains *at least* the predominant length of the distribution. The size of the bin is determined by the *left_edge* and *right_edge*, which are initially set equal to the distribution

mode (line 10). Then, each length value is assigned to the appropriate bin (lines 15-25) and it is checked if the values of the first quartile and of the third quartile of the binned distribution are the same (line 29). In this case, the mode of the *binned* distribution predominates over the other bin values and the binned distribution can be considered as normal with a good approximation. Otherwise, the procedure is repeated increasing the bin size by including the next longest-lasting length and updating the value of the left or right edge accordingly (lines 32-35).

In step *b)* (Fig. 18), the three-sigma rule [LF04] is applied on the binned distribution that has just been created, considering every AS path included in a bin whose value is greater than $\mu_{f,d} + 3\sigma_{f,d}$ as an outlier, where $\mu_{f,d}$ and $\sigma_{f,d}$ are the temporal mean and the standard deviation respectively of the binned distribution created during step *a*). $\mu_{f,d}$ and $\sigma_{f,d}$ is computed as follows:

$$\mu_{f,d} = \frac{\sum_{i=1}^{N_{f,d}} b_i \cdot w_i}{\sum_{i=1}^{N_{f,d}} w_i} \tag{4.1}$$

$$\sigma_{f,d} = \sqrt{\frac{\sum_{i=1}^{N_{f,d}} w_i \cdot (b_i - \mu_{f,d})^2}{\sum_{i=1}^{N_{f,d}} w_i}} \tag{4.2}$$

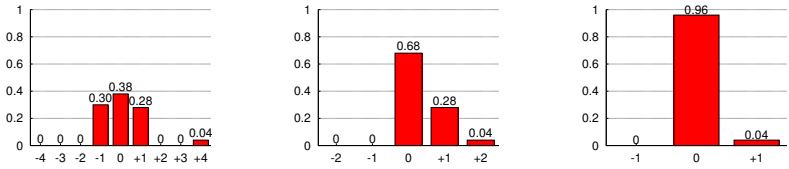
where $N_{f,d}$ is the number of bins, and b_i and w_i are the value of the i-th bin and its temporal weight, respectively. Note that $\mu_{f,d} + 3\sigma_{f,d}$ is still a threshold, but its value is not arbitrary. Rather, it has both a real and statistical significance, since it depends on the AS path lengths that `f` uses to reach `d`. As a result of this filtering step, each AS path is classified as *dropped* or *maintained*. However some AS paths generated during a path exploration phenomenon may still not have been correctly discarded, due to their limited AS path length. Step *c)* (Fig. 19) aims to identify these remaining routes by checking whether any route was announced less than thirty seconds earlier than any spurious route identified in step *b*). In such cases, the closest precedent route is flagged as part of the path exploration phenomenon (line 9) and checks are made once again starting from this route, until no more routes are found in the time interval analysed. Every AS path passing through this last filter is declared as

```
1  Input: f, d, routes_{f,d}, D_{f,d}, M_{f,d}
2  ---------------------------------------------------------
3  foreach r in routes_{f,d}
4      if (r.as_path ∈ D_{f,d})
5          curr_r = r
6          while(exists previous route)
7              prev_r = previous_route(curr_r)
8              if (birth(curr_r) − birthbirth(prev_r) ≤ MRAI)
9                  to_check.remove(prev_r.as_path)
10             else
11                 break
12             curr_r = prev_r
13
14         V.insert(to_check)
15         to_check = ∅
16     else
17         to_check.insert(r.as_path)
18
19 V.insert(to_check)
20
21 ---------------------------------------------------------
22 Output: valid AS paths V
```

Figure 19: Step c) MRAI-based event filtering

valid and can be used to draw economic inferences. Note that thirty seconds is not an arbitrary value, but instead reflects a protocol operational standard value, since it is the the default value of the *MinRouteAdvertisementIntervalTimer* (MRAI timer) [RFCb]. This parameter is the minimum time interval that should elapse between consecutive `UPDATE` messages for a given route sent by an ASBR to its neighbor. It is usually used to limit the amount of announced routes during BGP transients to improve BGP routing convergence ([LABJ01, SKM06, PZMZ06]). Note that network operators typically leave this parameter at the default value, but some use a lower value to reduce the convergence times even further [FSR11].

To better understand the methodology, consider as an example the evolution of the routes related to pair <$\overline{d}$, $\overline{f}$> (Table 5), where $\overline{d}$ = `2a03:2040::/32` and $\overline{f}$ = `(2001:43f8:1f0::29, AS 6968)`, which has been found to involve at least one inter-PF route. The steps



(a) Bin size = 1 **(b)** Bin size = 2 **(c)** Bin size = 3

Figure 20: Binned distribution creation related to routes collected for <$\overline{\text{d}}$, $\overline{\text{f}}$>

AS path length	lifespan [s]	final bin	status
5	220,858	0	valid
6	282,158	0	valid
7	211,402	0	valid
10	26,774	+1	dropped
$\mu_{binned} = 0.036$ $\sigma_{binned} = 0.03$ $\mu_{binned} + 3\sigma_{binned} = 0.09$			

Table 5: AS path length distribution of routes related to <$\overline{\text{d}}$, $\overline{\text{f}}$>

involved are depicted in Figure 20. Firstly, the correct bin value it is computed in order to consider the distribution as normal. The distribution is initially centered around its mode value (*length*=6) and analysed with a bin size equal to one. Since the distribution does not show a predominant bin, the bin size is enlarged to two, i.e. the difference between the first (*length*=6) and the second (*length*=5) modes plus one. The resulting distribution, although more peaked, still does not satisfy the condition of normality required. Thus, the bin is further enlarged (*bin size*=3) to also include the third mode (*length*=7). The distribution is now extremely peaked, and step *b*) cuts off all the AS paths included in the final $+1$ bin, including the inter-PFs.

4.3.2 Economic inference phase

Now that a methodology able to eliminate most spurious routes is available, to infer inter-AS economic relationships there is the need to have an appropriate algorithm that can exploit the filtered routes. In this thesis,

```
1  Input: set of valid AS paths V
2  ---------------------------------------------------------
3  foreach p in V
4      foreach connection [A,B] in p
5          if (T1 ∈ T_list follows [A,B] in p)
6              F[A,B].insert(c2p)
7          if (T1 ∈ T_list precedes [A,B] in p)
8              F[A,B].insert(p2c)
9          if (does not exist any T1 ∈ T_list )
10             F[A,B].insert(p2p)
11 ---------------------------------------------------------
12 Output: set of sets of economic tags F
```

Figure 21: Step a) of the economic tagging algorithm

```
1  Input: set of tags F produced by step a)
2  ---------------------------------------------------------
3  foreach t in F[A,B]
4      if (exists P[A,B])
5          P[A,B] = merge(P[A,B], t)
6      else if (exists P[B,A])
7          P[A,B] = merge(P[B,A], t)
8      else
9          P[A,B] = t
10 ---------------------------------------------------------
11 Output: set of preliminary tags P
```

Figure 22: Step b) of the economic tagging algorithm

	p2c	**p2p**	**c2p**	**s2s**
p2c	p2c	p2c	s2s	s2s
c2p	s2s	c2p	c2p	s2s
p2p	p2c	p2p	c2p	s2s
s2s	s2s	s2s	s2s	s2s

Figure 23: Merging rules

will be exploited the techniques described in [GIL+11] and [OPW+10] due to their strong bonds with collected raw data and few hypotheses and assumptions supporting the heuristic. The original algorithms rely on the valley free property of AS paths [Gao01b] and on a priori knowledge of a set of provider-free ASes [Wik]. The valley-free property implies that in a given AS path: *a*) at most one single p2p connection exists, *b*) no c2p connections can follow a p2p connection, and *c*) no c2p connections can follow a p2c connection. By also considering the provider-free property of a well-identifiable set of ASes, this also means that *d*) any connection that appears before a provider-free AS can be considered a

```
1  Input: set of preliminary tags 𝒫 produced by step b) of the algorithm
2  ---------------------------------------------------------------------
3  foreach p in AS path tagged with 𝒫
4      foreach direct connection [A,B] in p
5          if (𝒫[A,B] == p2p)
6              if([A, B] precedes any c2p ∈ 𝒫 in p)
7                  if(exists(𝒯[B,A]))
8                      𝒯[B,A] = s2s
9                  else
10                     𝒯[A,B] = (𝒯[A,B] == s2s) ? s2s : c2p
11             if ([A, B] follows any p2c ∈ 𝒫 in p)
12                 if(exists(𝒯[B,A]))
13                     𝒯[B,A] = s2s
14                 else
15                     𝒯[A,B] = (𝒯[A,B] == s2s) ? s2s : p2c
16
17 foreach [A, B] in 𝒫
18     if (!exists 𝒯[A,B])
19         𝒯[A,B] = 𝒫[A,B]
20
21 ---------------------------------------------------------------------
22 Output: set of tags 𝒯
```

Figure 24: Step c) of the economic tagging algorithm (enhancement step)

c2p, and *e*) any connection that appears after a provider-free AS can be considered a p2c. Note that connections between provider-free ASes are considered p2p by definition. The main difference between the two algorithms is in how they handle spurious routes. Since the preliminary step described in Section 4.3 entails handling problematic routes, the basic algorithm is the same as the one in [OPW+10] without the two-day time threshold filtering. Likewise it is the same as the algorithm proposed in [GIL+11] but with $N_{MAG} = \infty$, i.e. with no time threshold. In addition, in the following it will be introduced also an enhancement step, which enables the quality of the inferences drawn to be refined further.

However, for the sake of completeness, Figures 21 ,22 and Fig. 23 reports an adapted version of original economic tagging algorithm described in [GIL+11], i.e. the basic algorithm after which the enhancement step will be applied. The first step of the algorithm (Fig. 21) takes in input the set of valid AS paths produced by the data-hygiene phase

(cf. Figures 16 and 24), and assigns at least one economic tag to each connection. In particular, whenever a connection between two ASes precedes an AS belonging to T_{list} (i.e. a provider-free AS), it is tagged as c2p. On the other hand, if it follows a provider-free AS it is tagged as p2c. Note that a connection may appear in more than one AS path, and the economic relationship that is inferred by analysing one path may differ from the economic relationship inferred for the same connection by analysing another path. In the second step (Fig. 22), all the economic tags found for each connection are merged to obtain one single tag, according to the merging rules reported in Table. 23. Extensive explanation about this table can be found in [GIL+11], however the basic rationale is that if two ASes results having both a p2p and a p2c connection then the result is that the two ASes have a p2c connection, since the export policies associated with the p2c relationship are a superset of the export policies associated with a p2p relationship (and the same rationale applies for s2s with respect to p2c and p2p relationships).

The enhancement (Fig. 24) exploits the economic tags drawn by the original algorithm to infer additional p2c (or c2p) relationships *also* from those AS paths that do not contain any provider-free ASes. In the original algorithm, these AS paths led only to simple p2p connections, which were potentially overwritten by p2c (c2p) relationships whenever a path was found containing a provider-free AS and the proof that one of the two ASes was providing transit to the other. However, despite the lack of provider-free ASes, these AS paths *may* still contain useful information to infer further transit relationships. Connections inferred to be p2c (or c2p) may appear because they were found in at least one other AS path containing a provider-free AS. This information can be exploited to force these AS paths to comply with the valley-free property, by turning some p2p relationships present in AS paths without provider-free ASes into c2p (or p2c) relationships by exploiting the same rationale as the original algorithm. If a set of p2p connections precedes a c2p connection, each is converted into a c2p. Likewise, if any p2p connection follows a p2c connection, each is converted into a p2c. In both cases, in order to detect an s2s connection the presence of other inferences made on different AS

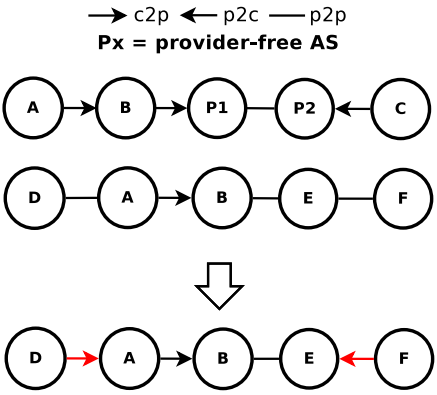


Figure 25: Example of application of the enhanced step

paths must be taken into account.

The key rationale behind this algorithm is that a p2c label is tangible proof that an AS announced networks previously received from one of its peers or providers to another AS, while a p2p label only proves that the ASes exchanged routing information related to each others' customers. Thus, if multiple different tags are found for the same connection, a p2p label is overruled by a p2c label, while contrasting p2c and c2p labels lead to an s2s label, i.e. each AS is a provider of the other. For example, consider the situation in Figure 25. Despite the fact that the AS path `D-A-B-E-F` does not contain any provider-free AS, it is still possible to infer that `D` is customer of `A`, and that `F` is customer of `E`. This because, from the AS path `A-B-P1-P2-C`, there is proof that `A` and `E` announced routing information to them which was related to their providers.

4.4 Results

The methodology described in Section 4.3 has been applied to the routes extracted from BGP data provided by BGPmon [BGP], PCH [PCH], RIS [RIS] and RouteViews [Rou] over February 2014. By applying the data hygiene

	p2c	p2p	s2s	Total
with spuriousness	110,441 (52.75%)	95,221 (45.48%)	3714 (1.77%)	209,376
without spuriousness	107,563 (52.99%)	93,508 (46.06%)	1925 (0.95%)	202,996

Table 6: Impact of spuriousness on tagging algorithm results

phase on this set of routes the results highlight that 42.24% of the total 47 million different AS paths collected appear only in spurious routes. The data hygiene phase was able to eliminate 91.62% of the inter-PF routes that led existing tagging algorithms to inexistent s2s relationships. Table 6 compares the results obtained using every AS path available and only the AS paths purged of spurious routes. Despite the large number of AS paths discarded, their effect on economic inferences is rather limited. This can be explained by the fact that only a limited set of ASes are misconfigured and contribute to creating no-valley-free AS paths. In addition, over the last few years several techniques have been proposed to reduce the propagation of spurious routes [HRA10] to limit the BGP traffic volume due to excessive UPDATE messages. However, the percentage of changed tags is not negligible (5.08%), and cannot be ignored. It should also be noted that 6380 connections out of 209,376 (3.05%) were found to appear *only* in spurious routes, and due to their ephemeral nature were not tagged. Note also that the increasing amount of p2p connections together with the simultaneous decrease in p2c connections means that several p2c inferences were based solely on spurious routes, and are thus likely to be wrong. Due to the nature of the proposed tagging algorithm, a p2c relationship is inferred only when a tangible proof is found of such relationship in the AS paths, while every other connection is considered to be a p2p. Table 7 highlights that the enhancement step enables us to infer an amount of p2c connections that is typically larger than other algorithms, still maintaining the spuriousness-free characteristic introduced by the data hygiene phase. An arbitrary time threshold is not a good solution. In fact the algorithm proposed by Oliveira et al. [OPW+10] misses a large amount of p2c exclusively due to their two-day threshold. This also happens in the results obtained with the most conservative N_{MAG} values of Gregori et al.'s algorithm [GIL+11], i.e. small

		p2c	p2p	s2s
Gregori et al. [GIL+11]	$N_{MAG} = 1$	98,505	109,076	1795
	$N_{MAG} = 2$	99,837	107,437	2102
	$N_{MAG} = 3$	101,171	105,911	2294
	$N_{MAG} = 4$	103,286	103,496	2594
	$N_{MAG} = 5$	106,423	99,443	3510
	$N_{MAG} = 6$	106,728	99,065	3583
	$N_{MAG} = 7$	106,785	98,962	3629
	$N_{MAG} = \infty$	106,928	98,770	3678
Oliveira et al. [OPW+10]		96,233	84,946	1796
Proposed tagging algorithm		107,563	93,508	1925

Table 7: Comparison of the results of economic tagging algorithms

values of N_{MAG} that represent the strictest transient filters. On the other hand, by using less conservative N_{MAG} values, some of the missing p2c increase, but also the number of wrongly inferred s2s increases due to the presence of short-lived no-valley-free AS paths. In particular, the larger the value of N_{MAG}, the larger the number of inter-PF AS paths (as shown in Section 4.2) that also participate in the s2s generation.

Chapter 5

Geography tagging

5.1 Introduction

The Internet AS-level topology is usually analysed at a worldwide level of detail. Every inference found for an AS is extrapolated from the global set of AS paths gathered from RCs, independently of the geographic location of the ASes. This approach is useful when the Internet is analyzed at a very coarse level. However, it may be misleading if the analysis is more focused on a specific geographical region. The risk is that the particular characteristics that the Internet has in that region may be lost.

Typically each AS has a particular role and specific economic behavior in each region of the world where it is present, which strictly depend on the connectivity and performance that it can provide for its customers in that region. For example, an intercontinental AS may be widespread in its home region—in terms of the number of connections and services offered—while it may be not competitive outside that region. This different level of pervasiveness may lead the same AS to establishing economic relationships in those regions with different criteria. This also implies that an AS connection that has been identified in a global analysis may hide multiple connections located in different geographic regions, each with its own characteristics. Despite that, research on the Internet AS-level analysis has considered ASes as homogeneous entities, each with a

global set of metrics and characteristics, regardless of their heterogeneity.

This chapter will show a methodology which – starting from the geolocation of ASes – is able to infer geographic information from AS paths and to analyze local properties of the Internet, with a special focus on AS topology properties and economic relationships. Specifically, it will be introduced a methodology which is able to geolocate AS paths and, then, to infer continental and geographical topologies, which will be analysed both from a statistical and economical point of view.

The economic analysis will be carried out by adapting the algorithm described in Chapter 4 to deal with geographic AS paths. It will be shown that the Internet actually consists of regional and independent ecosystems, which differ greatly in terms of both topological and economic properties and introduce particular characteristics that are hidden in a global-level analysis. These differences should be taken into serious consideration by all research based on the AS-level topology of the Internet - e.g. Internet modeling evolutions, protocol analyses, worm spread analyses - in order to rely on a more realistic structure of the Internet, instead of on a coarse and potentially misleading representation.

The rest of this chapter is organized as follows. Section 5.2 details the process of geolocating each AS from raw BGP data. Section 5.3 introduces the methodology aimed at producing geographically tagged AS paths. Section 5.4 shows the graph properties found by the undirected analysis of the regional topologies extracted from geographically tagged AS paths. Section 5.5 shows how to modify the economic tagging algorithm described in Chapter 4. Finally, Section 5.6 presents the economic analyses of economic regional topologies extracted using the economic tagging algorithm.

5.2 AS geolocation

Knowledge regarding the geographic range of an AS is one of the fundamental parameters for decisions concerning the establishment of a settlement-free peering or a transit type of relationship between ASes. Several Tier-1 (T1) ASes include in their peering requirements at least one geographic

constraint for candidate peers that need to be fulfilled. Just to name a few, AT&T[1] requires a list of the countries served by the candidate peer in the peering request submission; Verizon[2] requires a minimum number of served countries in the region where the peering is requested and that candidates own a "geographically-dispersed network"; and Telia-Sonera[3] requires that the candidate peer is present and able to exchange traffic and to be interconnected in a minimum number of cities in two out of three regions (Europe, North America and/or Asia Pacific/Oceania). The formal definition of AS – given in [RFCc] – is used as a starting point to perform the geolocation:

> *"An AS is a connected group of one or more IP prefixes run by one or more network operators which has a single and clearly defined routing policy."*

Given this definition, it is straightforward that an AS is geolocated if its own prefixes are geolocated. The list of all the prefixes advertised by a given AS can be collected by parsing the BGP raw data provided by RC projects. Each prefix can be geolocated in turn by geolocating each IP address inside it, using one of the IP geolocation databases available [SZ11]. Consider a generic route `x.y.z.0/24-A B C D`. It is possible to claim that the last element of the AS path[4] owns at least a network – and thus it may or may not be present – in the region(s) where the prefix is geolocated. This approach is correct for any given geographic region (e.g. countries, continents) iff the granularity of the geolocation tool is fine enough and iff the route does not carry the `AGGREGATOR` and `ATOMIC_AGGREGATE` attribute. The `AGGREGATOR` is an optional transitive attribute and the `ATOMIC_AGGREGATE` is a well-known discretionary attribute of the BGP protocol and may be included in UPDATE messages by a BGP speaker which performs route aggregation. If one of these attributes is present, it is possible that part of the real AS path

[1] http://www.corp.att.com/peering/

[2] http://www.verizonbusiness.com/terms/peering/

[3] http://www.teliasoneraic.com/dms/teliasoneraic/Documents/tsic-sfi-010.pdf

[4] The UPDATE message is originated by the rightmost AS of the path.

is missing, hidden by the aggregating router. Optionally, the aggregating router could set the AS_SET attribute, which contains the *unordered* set of ASes the route traversed before the aggregation occurred. Either the AS_SET is present or not, it is not possible to state that the considered prefix belongs to the last element of the AS path, but additional confirmation is needed from the whois service provided by the Internet Routing Registries [5]: the prefix is considered to belong to the last AS of the AS path if that AS is the owner of the announced prefix also according to the whois response. For example consider the route shown in Fig. 26 – which is the textual representation in MRT data of a route collected by the RC rrc12 of RIPE RIS – where the prefix is entirely geolocated in Europe. Given the presence of the AGGREGATOR attribute, whois service has to be queried. Since the response state that the prefix belongs to AS 2597, it is possible to conclude that AS 2597 is present in Europe. In the following details about the Internet infrastructure at continental level will be showed, by exploiting the Maxmind GeoIPLite database[6], which has been proved to be reliable at both level [PUK+11]. In the continental level analysis, the Internet has been divided in five macro regions: Africa (AF), Asia-Pacific (Asia and Oceania – AP), Europe (EU), Latin America (the Caribbean, Central America, Mexico and South America – LA) and North America (Bermuda, Canada, Greenland, Saint Pierre and Miquelon, and U.S.A. – NA).

5.3 Introduction of geography in BGP data

Geolocation of ASes by itself is not enough to extract geographic information from AS paths. An AS can have a geographic range that spread across multiple regions, thus it is not possible to infer where each AS connection forming an AS path is located. To overcome this problem it is proposed a three-step algorithm which, based on the geolocation of each AS, is able to geolocate each AS connection of the AS paths.

[5] A list of available whois locations can be found at http://www.irr.net/
[6] http://dev.maxmind.com/geoip/legacy/geolite/

```
TIME: 10/01/11 08:00:06
TYPE: TABLE_DUMP_V2/IPV4_UNICAST
PREFIX: 192.12.193.0/24
SEQUENCE: 241676
FROM: 80.81.192.98 AS9189
ORIGINATED: 08/17/11 00:23:52
ORIGIN: IGP
ASPATH: 9189 8422 3356 2597
NEXT_HOP: 80.81.192.98
MULTI_EXIT_DISC: 100
AGGREGATOR: AS2597 217.29.66.79
COMMUNITY: 9189:1003 9189:1102
```

Figure 26: Textual representation of a route in MRT format

Inference of enhanced routes from BGP raw data. In this step it is obtained an enhanced route – defined as the triplet `{SOURCE, ASPATH, DESTINATION}` – for each route available in the BGP data. `SOURCE` is the region where the BGP AS Border Router that announced that route to the RC is located and can be obtained by geolocating its IP address (i.e. its *ipfrom*, cfr. Section 2.2, pag. 10). `ASPATH` is the content of the homonym BGP attribute. `DESTINATION` is the region where the prefix announced is located. Since a prefix could be geolocated in more than one geographic region, more than one enhanced route could be created from a single route, one for each region where the destination is found to be located. Consider the route reported in Fig. 26. Both the IP address of the BGP speaker (`80.81.192.98`) and the prefix announced (`192.12.193.0/24`) are located in Europe, thus leading to the enhanced route `{EU, 9189 8422 3356 2597, EU}`.

Detection of Single Region Located Transit Points (SRLTPs) in each enhanced route. In this step it is extracted from each enhanced route the set of SRLTPs, i.e. regional intermediate points where the traffic needs to flow. The `SOURCE` and the `DESTINATION` of each enhanced route are by definition part of this set, since they are both geolocated in a single region. This set also includes two classes of ASes that can be found in the `ASPATH`. The first class is represented by ASes that own prefixes only located in a single region, i.e. ASes that do not own an inter-regional

infrastructure. Another class is represented by those ASes that have a single region in common with a neighboring AS. The basic idea is that typically ASes follow a regional principle to route traffic. Inter-regional ASes tend to subdivide their ASes into different areas by exploiting the features of BGP or IGP protocols such as OSPF and IS-IS in order to maintain traffic as regional as possible to maximize the performance. Thus, an inter-regional AS avoid using its infrastructure when it is not needed, by representing a SRLTP under certain circumstances. For example consider the enhanced route extracted from the route shown in Fig. 26. Geolocating each AS using the methodology described in Section 5.2, the result is that ASes 9189, 8422 and 2597 are located only in Europe, while 3356 is located in every continent. Since also `DESTINATION` is located in Europe, each AS represents a SRLTP.

Inference of Geographic AS paths. In this step it is exploited the information just extracted to geolocate the connections of the AS path of each enhanced route. Following the same regional principle introduced above, inter-regional ASes are typically interconnected on every location where they are co-located, trying to maintain inter-AS traffic as regional as possible. This means that, for example, if the traffic flows from a source to a destination both located in region R through ASes located in R, the traffic is very likely to remain confined in R even if these ASes are co-located in other regions. By exploiting these considerations and the SRLTPs identified in the previous step it is possible to complete the geographic tagging algorithm, which is presented in Fig. 27. The algorithm aims to create a set of geolocated AS connections from each enhanced route. Each enhanced route together with its set of geolocated connections hereafter is defined as a Geographic AS path. To achieve this, it is required to analyze each AS in the ASPATH and check whether the connection with its neighboring ASes can be established in the considered region. At the beginning, `SOURCE` is considered as the initial region (line 2), and all those AS connections that may belong to the considered region (line 18) are added to the set of AS connections located in that region to the set of AS connections located in that region. The consid-

```
1  foreach enhanced route R
2      region = SOURCE;
3      for(i=0; i < length(ASPATH); i++)
4          if(AS_i ∈ SRLTP && region ∉ regions(AS_i))
5              region = regions(AS_i);
6              for(j = i; j > 0; j−−)
7                  if(region ∈ regions(AS_j))
8                      add(AS_{j−1}, AS_j) to GEO_PATH(Region);
9                  else
10                     break;
11         elsif(region ∈ regions(AS_i))
12             add(AS_i, AS_{i−1}) to GEO_PATH(Region);
13         else
14             i = index of next SRLTP;
15             region = regions(AS_i);
16             for(j = i; j > 0; j−−)
17                 if(region ∈ regions(AS_j))
18                     add(AS_{j−1}, AS_j) to GEO_PATH(Region);
19                 else
20                     break;
21     if (region ≠DESTINATION)
22         region =DESTINATION;
23         for(j = length(ASPATH); j > 0; j−−)
24             if(region ∈ regions(AS_j))
25                 add(AS_{j−1}, AS_j) to GEO_PATH(Region);
26             else
27                 break;
```

Figure 27: Geographic tagging algorithm

ered region will be changed if an SRLTP (line 4) or if a multi-regional AS (line 13) not located in that region is found. In this last case the change is preceded by a jump to the next SRLTP (line 14). The output of the geographic tagging algorithm is composed by the set of Geographic AS paths, each inferred from the related enhanced routes. Considering the route in Fig. 26 and its characteristics shown so far, it is inferred that the full AS path is located in Europe.

5.4 Undirected graph analyses

The algorithm to obtain Geographic AS paths described in section 5.3 has been applied on BGP routes gathered by Route Views, RIS, PCH and BGPmon RCs during February 2014. Then, from each geographic AS path have been extracted the geolocated connections to create regional AS-level topologies, finding that 46,806 out of 47,246 ASes appear in at least one regional topology. The missing ASes are in 59 out of 441 cases due to missing IP prefixes in the Maxmind database. In the remaining cases the ASes, although being geolocated, do not appear in any regional topology because they do not share any region with the neighboring ASes in any AS path in which they appear. This may be due to the use of BGP multi-hop sessions – where the ASes are actually located in different regions – or due to the mistaken/partial geolocation of prefix(es) by the Maxmind database. In both cases the geographic tagging algorithm is not able to infer where the connection is geolocated. For the same reason, the geographic tagging algorithm is not able to assign a region to 7,447 connections out of 209,456. The first important result is that the continental topologies extracted overlap only slightly, as highlighted by the Jaccard similarity indices[7] computed between pairs of continents for nodes (J_{nodes}) and connections (J_{edges}) and reported in Table 8.

This poor overlap is confirmed by the fact that only about 4.51% of ASes are located in more than one region and only about 1.02% in more than two. This evidence show that the Internet consists of regional ecosystems interconnected by just a very small number of inter-regional ASes. These ASes guarantee full Internet reachability, since only one AS owning an inter-regional IP network infrastructure can handle inter-regional traffic. This poor overlap also means that the regional principle has been applied by the algorithm has only on a small set of connections and thus the largest part of connections is correctly geolocated.

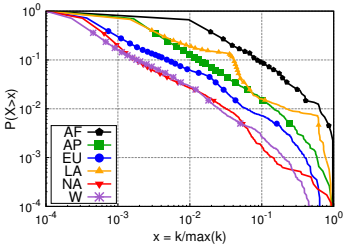
Further evidence regarding the differences between regional topologies is provided by the graph properties summarized in Table 9 and by

[7]The Jaccard similarity index is a measure that allows to quantify the similarity between pair of sets and is defined as $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$.

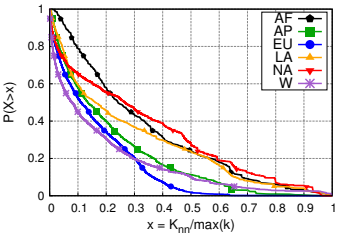
the CCDFs of the normalized node degree k and the normalized average neighbor degree $\frac{k_{NN}}{\max(k)}$ depicted in Figures 28a and 28b respectively. The sizes of the topologies differ greatly in terms of nodes, reflecting the different degrees of economical and technological development of the regions. It is particularly interesting to compare North American and European topologies, which have a quite similar number of nodes but differ significantly in terms of edges. The CCDF of the node degree shows that the European region is more densely connected than the North American region where, on the other hand, there are ASes with a larger degree and where the number of ASes with a small degree is higher. This suggests quite a hierarchical structure in North America versus a flatter structure in Europe. This is confirmed by the CCDF of the normalized Average Neighbor Degree, which shows that in Europe, ASes tend to connect to ASes with a similar degree, while in North America they tend to connect to ASes with a very large degree. The differences between these ecosystems reflect the Internet's historical evolution in the respective regions. In North America, especially in the U.S., a small set of large ISPs (e.g. AT&T, Centurylink and Verizon Communications) provide connectivity to all the states. In Europe on the other hand, each country is typically characterized by the presence of a national telco (e.g. Deutsche Telekom and Telecom Italia) which usually own a large part of the national Internet infrastructure, and by the presence of at least one Internet eXchange Point (IXP) that encouraged the establishment of settlement-free peering connections among local ISPs. More details on the role of IXPs in the development of the Internet can be found in [GILO10, AKW09b] and [XDZC04].

	Africa	**Asia Pacific**	**Europe**	**Latin America**	**North America**
Africa	-,-	0.009,0.012	0.006,0.004	0.011,0.008	0.011,0.008
Asia Pacific	0.009,0.012	-,-	0.020,0.029	0.011,0.016	0.031,0.073
Europe	0.006,0.004	0.020,0.029	-,-	0.007,0.006	0.030,0.085
Latin America	0.011,0.008	0.011,0.016	0.007,0.006	-,-	0.017,0.019
North America	0.006,0.007	0.031,0.073	0.030,0.085	0.017,0.019	-,-

Table 8: Jaccard similarities indices $J = (J_{nodes}, J_{edges})$



(a) Normalized node degree



(b) Normalized average neighbor degree

Figure 28: CCDF of the Normalized Degree and Normalized Average Neighbor Degree per continent

	Africa	Asia Pacific	Europe	Latin America	North America
Nodes	886	7607	19,981	3472	17,449
Edges	2222	23,359	121,175	18,834	59,303
$\langle k \rangle$	5.02	6.14	12.13	10.85	6.80
max k	103	606	2715	675	3115

$\langle . \rangle$ = average, k = degree

Table 9: Regional topology statistics

5.5 Geography and inter-AS business relationships

The analysis of the undirected graph of the Internet highlighted just part of the extreme complexity of the Internet ecosystem, since it completely lacks the real business model of each element. As discussed in Chapter 4, the Internet consists of a network managed by thousand of different organizations. Some of these organizations, e.g. small or large ISPs, live on the sale of Internet transit to other organizations, others, such as CDNs and search engines, just aim to offer content to end users. Most organizations, though, just care about Internet connectivity. In order to highlight that heterogeneity and to get a better insight into the Internet, in the fol-

lowing it is adapted the economic tagging algorithm shown in Chapter 4 to deal with Geographic AS paths, in order to obtain economic tagged *regional* AS-level topologies.

The original algorithm operates in two main parts: *i)* Removal of spurious routes and *ii)* computation of economic tags on the set of AS paths belonging to the cleaned routes. In this case, the first part is adapted to produce cleaned *enhanced routes*. This can be done since a route can be mapped to one or more enhanced routes (cfr. Section 5.3). Thus,given the set of spurious routes, it can be converted to the set of spurious enhanced routes, which can be removed from the whole set of enhanced routes, leading to the set of cleaned enhanced routes. Then, the set of cleaned enhanced routes is transformed in cleaned geographic AS paths according to the the methodology shown in Fig. 27. Finally, the second part of the original algorithm – i.e. the computation of economic tags – is adapted to deal with those cleaned Geographic AS paths instead of plain AS paths. To achieve this behavior, the step *a)* of the economic tagging algorithm (see Fig. 21) must record also the region in which each tag is found and, then, during step *b)* and *c)* (see Figures 22 and 24) different economic tags related to the same connection might affect each other only if they belong to the same region.

5.6 Economic analyses

The application of the enhanced economic tagging algorithm to the sets of geographic AS paths allows deeper insights of each regional ecosystem which reveal the real nature of the regional differences that were only deduced from the undirected analysis of the Internet.

The most relevant characteristic highlighted by the economic analy-

	Africa	Asia Pacific	Europe	Latin America	North America
P2C	1605	17,068	47,643	7876	38,707
P2P	525	5576	68,173	9999	18,497
S2S	53	439	880	262	514

Table 10: Economic *regional* topologies

sis is the large proliferation of potential p2p connections in the European ecosystems (see Table 10). This is in contrast with the peering behaviors of other regions, where the amount of p2c connections is larger (around 70% of the total) than the amount of p2p connections (with a slightly exception for the Latin American ecosystem). Together with the conclusions drawn in Section 5.4, this allows to understand the real nature of the flat structure of the European Internet ecosystem. Indeed, this joint analysis shows that Europe is rich in small/medium transit providers that, in addition to offering transit to end-users and stub ASes, tend to establish settlement-free p2p connections among them. The establishment of these BGP connections allows the ISPs to minimize the amount of traffic directed to their provider, thus reducing their transit costs. The proliferation of these small/medium providers is also the reason for the development in Europe of largely-populated IXPs (e.g. AMS-IX, LINX and DE-CIX) which in turn facilitated the establishment of settlement-free relationships among ASes, helping to create the large amount of p2p connections just described.

Another regional feature is highlighted by tag changes. In Table 11 are summarized the most relevant tag changes from the worldwide to the regional scenarios, i.e. peering (p2p) to transit (p2c, c2p, s2s) and vice versa. Although the number of these connections may look not relevant at a first glance (around 4-10% of the total in each region), it should be considered that these tags are referred to AS connections that compose the core of the region. This is highlighted by the large number of tag changes that involve only non-stub ASes and that involve only multi-regional ASes. Most of the changes consists in shifting from transit to peering connections, showing that an AS may establish multiple economic relationships with another AS, depending on the location of the interconnection. This means that inter-regional providers may decide to establish regional p2p connections – exchanging only routes of regional customers – in those region where their pervasiveness is similar, while they may decide to establish a p2c agreement elsewhere. The presence of this regional type of relationship is confirmed by the existence of BGP communities dedicated to regional peers in some of the largest ASes, but

	Africa	Asia Pacific	Europe	Latin America	North America
Tag changes	132	572	2065	148	1439
Peering to transit	21	78	335	38	216
Transit to peering	104	457	1,631	93	1160
Among multi-reg. ASes	114	261	554	83	1133
Among non-stub ASes	111	376	1559	133	1298

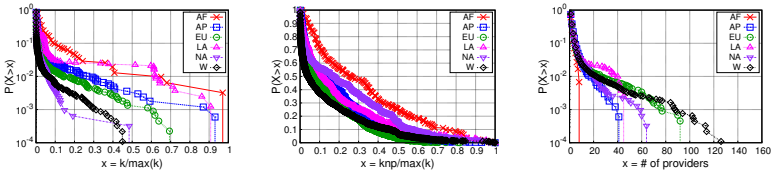
Table 11: Economic relationships changes from global to regional topologies

our methodology is still too coarse-grained to detect their presence and to distinguish it from the general peering, where the ASes exchange the routes of all the customers. Note that the total number of tag changes (first row in table Table 11) it is not the sum of the classified tag changes (rows from 2 to 5), since a change may belong to multiple classes. For example a tag could change from peering to transit and could involve multi-regional ASes.

Chapter 6

Towards an ideal RC infrastructure

In this chapter the MSC and MC problem described in Chapter 3 are applied on BGP data collected during Februay 2014. The computation of the results exploits the economic tagging algorithm described in Chapter 4 and the geographic tagging algorithm described in Chapter 5. In detail, hereafter are shown the results obtained by solving the MSC and MC problems on the global topology of the Internet, referred to as *World (W)*, and to five regional topologies: *Africa (AF)*, *Asia-Pacific (AP)*, *Europe (EU)*, *Latin America (LA)*, *North America (NA)*. The global economic topology is the one inferred in Chapter 4 (cf. Table 7, page 54). The regional economic topologies are the ones inferred in Chapter 5 (cf. Table 10, page 65). For the sake of readiness, the characteristics of these economic topologies are summarized in Table 12. Afterward, it are analysed the candidate feeders selected and identified their peculiar characteristics. Finally, the coverage of the current feeders is compared with the ideal set from the results of the methodology, and it will be shown how much the coverage of the RCs would improve if new feeders were chosen wisely using the provided ranking list.



(a) Normalized node degree

(b) Normalized average provider degree

(c) Number of providers

Figure 29: CCDF of node properties of candidate feeders

6.1 Global vs regional analysis

Firstly, the p2c-distances for each of the available topologies are computed, as required by the MSC procedure. Note that these values can also be used to highlight which zones of the Internet are poorly captured by the RCs, thus providing further proof of the incompleteness of the current collected topologies. To prove this, it has been analysed the p2c-distances of each non-stub AS from the current set of full feeders (see Table 13). Note that the p2c-distance of ASes that cannot reach

	AS-level topology					
	AF	*AP*	*EU*	*LA*	*NA*	*W*
ASes	886	7607	19,981	7876	17,449	47,246
Non-stub ASes	288	1662	3921	861	2820	8426
p2c	1605	17,068	47,643	7876	38,707	107,563
p2p	525	5576	68,173	9999	18,497	93,508
s2s	53	439	880	262	514	1925

Table 12: Economic AS-level topology summary

p2c-distance	**# non-stub ASes**					
	AF	*AP*	*EU*	*LA*	*NA*	*W*
0	2	24	108	21	53	168
1	23	178	512	140	297	1004
2	43	373	934	244	542	1902
3	49	496	1169	239	657	2472
> 3	171	591	1198	217	1271	2880

Table 13: Regional distribution of p2c-distances of non-stub ASes from current full feeders

Table 14: MSC procedure results

Region	$\lvert\mathcal{P}\rvert$ ($\lvert\mathcal{I}\rvert$)		
	$d = 1$	$d = 2$	$d = 3$
AF	161 (308)	139 (263)	134 (250)
AP	843 (1684)	711 (1468)	677 (1352)
EU	2143 (4454)	1865 (4121)	1801 (3869)
LA	427 (811)	352 (702)	337 (648)
NA	1595 (3229)	1424 (4351)	1389 (2818)
W	4344 (9211)	3674 (8415)	3529 (7897)

Table 15: Number of current feeders included in the set of elements candidates to be part of at least one optimal solution

	# of current feeders $\in \mathcal{I}$ (% out of the total feeder class)					
	AF	*AP*	*EU*	*LA*	*NA*	*W*
Full	1 (50%)	13 (52%)	54 (49.09%)	15 (60%)	27 (50.00%)	86 (44.79%)
Partial	6 (75%)	26 (49.05%)	88 (30.87%)	10 (30.55%)	48 (44.44%)	139 (34.32%)
Minor	13 (44.82%)	36 (38.70%)	179 (29.68%)	4 (12.12%)	94 (34.94%)	290 (32.36%)

any full feeder using only p2c connections is considered to be ∞. For $d = 0$ the number of non-stubs covered is equal to the number of full feeders which are non-stub. Most ASes are currently either too far from the set of full feeders or cannot be reached by any full feeders via c2p connections alone, thus potentially representing hideouts for AS connectivity which need further investigation. Now that the p2c-distances had been calculated, it is possible to apply the methodology to each of the economic topologies available. The results are summarized in Table 14, which shows the cardinality of the solution set $\mathcal{P}$ and, in round brackets, the cardinality of the set of ASes that can be part of a solution ($\mathcal{I}$) for each topology. In each geographical scenario, the number of feeders required is significantly smaller than the number of non-stub ASes (cf. Tables 12 and 14). More importantly, the sum of feeders required by regional scenarios is greater than the number of those required by the *World* scenario. This result was expected since the complete capture of the connectivity of an AS with a large geographical range may entail deploying multiple feeders around the world. Inter-regional ASes typically follow a regional principle to route their traffic, in order to maximize their performance and minimize latency ([GIL+11, Nor11]). To do this, they tend to subdivide their ASes into different routing areas by exploiting the features

of Interior Gateway Protocols (IGPs) such as OSPF and IS-IS and set up connections that can only be exploited in regional traffic routing. A total of 1073 out of 8426 non-stub ASes are present in more than one single regional topology and thus may fit this description.

6.2 Candidate feeder analysis

Hereafter will be outlined the characteristics of the candidate feeders found by applying the methodology with $d = 1$, which represents the best trade-off between AS-level connectivity discovery and the number of BGP feeders required. With $d = 1$, the positioning algorithm finds the set of feeders required to obtain BGP routing data filtered by at most two BGP decision processes from each non-stub AS: the source AS and the feeder itself. In this way, it can be exploited the multihomed nature of several ASes to lower the cardinality of the final solution. The characteristics of the feeders obtained by applying the methodology with other values of d are available at [Iso].

Table 16 and Figure 29 show the most relevant characteristics of these ASes. Figure 29 depicts a) the degree distribution, b) the *average provider degree* (k_{np}) distribution, where k_{np} is computed for each AS as the average degree of its providers, and c) the number of providers per candidate feeder. Note that the degree and the k_{np} distribution have been normalized with the maximum value of the node degree k found in the related region to allow full-scope analysis and to trace the characteristics of a typical candidate feeder. The most relevant classes of ASes found among candidate feeders are: a) stub ASes (see Table 16), b) ASes that have set

Table 16: Characteristics of candidate feeders

Region	# of ASes $\in \mathcal{I}$ (% out of $\lvert\mathcal{I}\rvert$)	
	On IXPs	*Stubs*
AF	42 (13.63%)	138 (44.80%)
AP	484 (28.74%)	808 (47.98%)
EU	2379 (53.41%)	2241 (50.31%)
LA	327 (40.32%)	340 (41.92%)
NA	528 (16.35%)	1591 (49.27%)
W	3894 (42.47%)	4691 (50.92%)

up a small number of BGP connections (see Fig. 29a), and c) ASes that have chosen a rather small number (see Fig. 29c) of small-medium ISPs (see low-medium values of the normalized k_{np} in Fig. 29b) to be their providers. Only a small percentage of these ASes are present on at least one IXP (see Table 16), which implies that their typical interconnectivity behavior is to not establish public peering with other ASes. Thus, it is possible to conclude that the typical AS that should become a feeder of the current RCs is a small multi-homed AS, which has set up multiple connections with different regional providers to guarantee its route diversity and increase the reliability of its reachability. This is not surprising since these ASes are likely to be located at the bottom of the hierarchy and, thanks to multi-homing, can cover several non-stub ASes at once.

6.3 Current status of the coverage of RCs

In the following it will be analysed how many of the current feeders are present in the ideal set of candidates found by the methodology. Their distribution for each region is shown in Table 15, as well as the percentage of the total number of feeders in the region that fall into this category. The main result is that only a small percentage of the current full feeders are part of an optimal solution in any of the topologies analysed. This is a direct consequence of their position in the Internet hierarchy, as already shown in Section 2.4. These ASes are not likely to have a large number of providers, thus their contribution is limited. In terms of minor and partial feeders, only a few can be found in the set of candidates, thus highlighting that, in terms of p2c-distance, only a few of them are placed in an optimal position and would be useful in a topological discovery perspective even if their contribution was total. It is also interesting to work out how many feeders would have to be added to the current RCs in order to improve the quality of the data. To determine these values, the methodology illustrated in Section 3.1.3 (page 24) have been slighthly modified by considering the current set of full feeders $\mathcal{F}$ as part of the initial set of solution $\mathcal{P}$, and considering the number of additional feeders as $n = |\mathcal{P}| - |\mathcal{F}|$. The results for each geographical region are reported in

Table 17: Additional (full) feeders required in each region

Region	$n = \lvert\mathcal{P}\rvert - \lvert\mathcal{F}\rvert$		
	$d = 1$	$d = 2$	$d = 3$
AF	160	139	134
AP	829	701	673
EU	2048	1829	1775
LA	412	341	328
NA	1568	1409	1376
W	4256	3626	3,500

Table 17. A comparison of the number of additional ASes required and the number of non-stub ASes (see Table 12) reveals that the methodology covers every non-stub AS with a number of new feeders which is about 50-60% of the number of non-stub ASes in each region. Finally, if new full feeders were chosen according to the ranking list obtained, it would then be possible to capture a *more complete* AS-level view of the Internet with a limited amount of new elements and, consequently, with limited costs. This can be seen by analysing Fig. 30, which shows the percentage of non-stub ASes covered in each region by introducing the ASes selected by each step k of the greedy algorithm in the set of full feeders. As can be seen from Table 18, this is also proved by the fact that, just by doubling the number of full feeders in each region, it is also possible to double the coverage of the non-stub ASes. The detailed ranking list of ASes per region can be found in [Iso]. Note that the methodology extracts the optimal solution for the input data provided, thus, if a new feeder is introduced, the solution may no longer be optimal. In any case, it still represents an upper bound to the number of additional full feeders needed. In fact, introducing new data may add previously hidden connections and may lead the tagging algorithm exploited to infer a greater number of p2c connections. These new connections may change the p2c-distance of several ASes that might be reached by exploiting a lower number of feeders. However, it would still be possible to apply the methodology once again to the new data to obtain a new optimal value.

Table 18: Coverage improvements by doubling the number of full feeders ($d = 1$)

Region	Current number of full feeders	# Not stub ASes covered (percentage)	
		Current status	*Doubling Full Feeders*
AF	2	25 (8.68%)	43 (14.93%)
AP	25	188 (11.31%)	387 (23.29%)
EU	110	570 (14.54%)	1158 (29.53%)
LA	25	150 (17.42%)	292 (33.91%)
NA	54	321 (11.38%)	706 (25.03%)
W	192	1078 (12.79%)	2337 (27.74%)

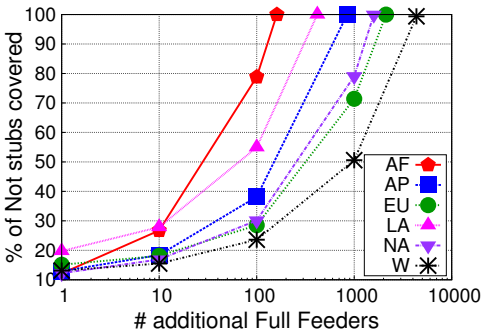


Figure 30: MC greedy algorithm results ($d = 1$)

Chapter 7

Isolario

As a consequence of this view from the top of the hierarchy, a large amount of connections established among ASes placed in the lower levels of the Internet are missing, and cannot be revealed by current BGP route collectors. The only opportunity to reveal these connections would be to increase drastically the amount of ASes located in the Internet periphery that are feeding the route collecting infrastructure. However, most of the administrators of these ASes are not interested in joining the current route collector projects just to advertise their network reachability or for mere altruism. Isolario is a route collector project based on the *do ut des* principle, that aims at persuading the administrators of ASes owned by small-medium organizations to share their full routing table by offering useful services in return.

The goal of Isolario is to improve the knowledge about the AS-level ecosystem of the Internet by increasing the amount of ASes from which BGP data is collected, hereafter *feeders*. To stimulate AS administrators to join the route collecting infrastructure, Isolario provides them *real-time* monitoring and alerting services on the health status of their own BGP routing system. With these services, a feeder administrator would be able to detect, monitor and analyse pathological events affecting its BGP routing system, like route flapping, prefix hijacking and loss of reachability, without increasing the computational load on the router or intro-

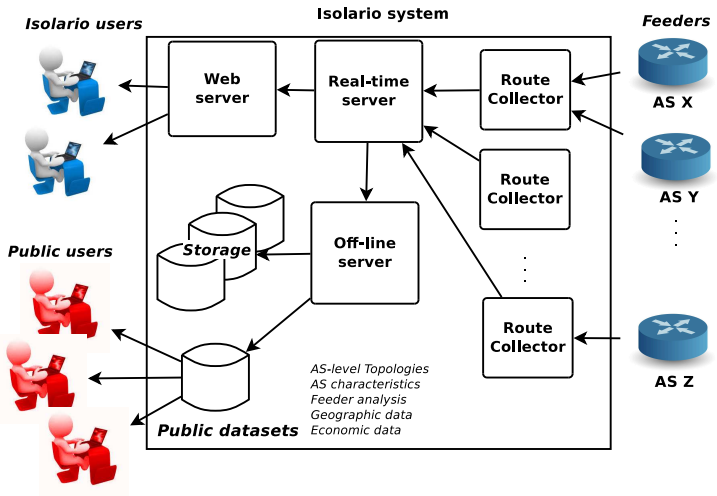


Figure 31: Isolario infrastructure

ducing third-party software. In order to join the Isolario project, feeders are required to establish (at least) one BGP session with one of Isolario route collectors and to announce us their default-free full routing table. The BGP data collected from feeders is used both to provide the real-time services and to improve the quality and completeness of the Internet AS-level ecosystem. Upon agreement with the feeder, BGP data will be made publicly available in MRT format (RFC 6396) to the research community, similarly to what is done by existing route collecting projects. The Isolario system (Fig. 31) is composed by *a*) a set of Route Collectors, which establish BGP sessions with external feeders, *b*) a real-time server dedicated to analyse in real-time BGP messages coming from each feeder, *c*) an off-line server dedicated to periodically analyse collected BGP data, *d*) a set of servers dedicated to store information and results of computations, and *e*) a web server to offer Isolario real-time services to the Isolario users and the results of the analyses concerning the Internet AS-level ecosystem to the public users (`http://www.isolario.it`).

7.1 Real-time services

Isolario main feature is the provisioning of real-time services to the users feeding the route collectors. The real-time feature is obtained in two steps. First, each incoming BGP flow is directly parsed, filtered and redirected towards dedicated modules which implement the services. Then, the result is provided to Isolario users through an HTML5 website which exploits the *WebSocket* protocol (RFC 6455) to update web pages only when new data related to the client become available, without additional polling traffic being generated. Hereafter there are introduced some of the Isolario services that will be available to the Isolario users.

7.1.1 Routing table viewer

This service enables a network administrator to monitor in real-time the routes that its AS is using to reach a set of Internet destinations. To do that, a dedicated module feeded by the incoming BGP flow of the feeder filters the messages on the basis of the portion of the Internet space under analysis. The user is firstly required to select at least one subnet to monitor. Then, the real-time evolution of routes towards the IP networks selected will appear on the user browser, as well as related real-time statistics and logs concerning announcements, changes and withdrawals. This information could be obtained also by exploiting the BGP Monitoring Protocol, however this protocol is not supported by all BGP routers and requires extra hardware functioning as a monitoring station. Alternatively, an administrator could exploit the Command Line Interface of its router through screen-scraping, but this approach is clearly inefficient and does not guarantee that all route changes are captured.

7.1.2 Route flap detector

Route flapping is a rapid sequence of route state changes [RFCd] that is computationally expensive for routers. Currently, network administrators use Route Flap Damping mechanism [RFCa] on their routers to mitigate problems caused by this pathological behavior. However, this

mechanism can significantly increase the convergence times of relatively stable routes [MGVK02] and the selection of its operational parameters is heavily under discussion [PMM+11]. This service detects route flapping while it is occurring through real-time analysis of BGP flows provided by feeders, by highlighting those feeder destinations affected by at least N events in T seconds, where N and T are parameters customizable by the user. This service allows a user to early detect the occurrence of such events enabling her/him to take the opportune countermeasures.

7.1.3 My subnet reachability

This service enables users to check in real-time the routes that other feeders use to reach her/his own AS networks. This is extremely useful, since no AS can obtain this information by itself. So far, an AS administrator could develop something similar is to rely either on periodic analyses of looking glass servers, or on the periodic dumps of BGP flows provided by RouteViews and RIS. However, both approaches completely lose the advantages of a real-time analysis. Another opportunity would be to exploit the real-time flow of BGP updates offered by BGPmon [OWS+13], however the user is required to develop its own application. This service shows also to the user information about reachability of her/his own networks from an IP-level perspective, thanks to the collaboration with Portolan [GLLV13, FGL+12]. Portolan is an active measurement infrastructure based on crowdsourcing and on mobile phones, which currently counts more than 200 probing smartphones capable to perform traceroute and ping measurements. Isolario triggers traceroute and ping measurements from Portolan probing-smartphones through the Google Cloud Messaging service, and provides to the user IP-traces and latency between his AS networks and the ASes hosting at least one Portolan agent.

7.1.4 Alerting services

Alerting services will warn an user when something anomalous is happening on its inter-domain reachability. For example, they can advertise

an user when a prefix hijack of its networks occurs, or when its infrastructure is not able to reach a particular destination (e.g. a specific website, a subnet). The alerting system sends an email to the user when it detects the anomalous event and, then, after a customizable amount of time, it will compile and send a report to allow the user to inspect in deep the routing events occurred near to the trigger event.

Chapter 8

Conclusions

Route collectors are extremely valuable for researchers, as they are the most reliable source of information regarding the inter-AS infrastructure of the Internet. However, they collect BGP data only from a small set of ASes, thus limiting the quality of the inferences that can be drawn from their analysis. In addition, the feeders that contribute with their full routing tables are typically large ASes such as provider-free and worldwide ISPs. This means that the current vision of RCs cannot capture any of the p2p connections established by small or medium-sized ASes (Chapter 2). Studies on the structure of the Internet topology need to be fully aware of the great extent of data incompleteness, since a topological analysis of the Internet as viewed from these monitors is like analysing a road map of a given country where the highways are known, but most of the secondary roads are not shown! To quote Sir Arthur Conan Doyle, *"It is a capital mistake to theorize before you have all the evidence. It biases the judgment"*. A solution to such incompleteness is to increase the number of feeders. In this work has been proposed a systematic methodology based on the p2c-distance metric in order to *i*) infer the minimum number of feeders to maximize the number of ASes belonging to the Internet core whose AS-level connectivity can be revealed, and *ii*) retrieve a ranking list of these candidate feeders, to understand the improvements – in terms of coverage – that can be obtained with a limited amount of re-

sources (Chapter 3. In order to solve these problems, they have been also introduced an algorithm to infer economic tagging algorithm from a spurious-free set of AS paths (Chapter 4) and a methodology to infer geographical AS-level topologies (Chapter 5).

The result is that the number of feeders should be drastically increased to decrease the incompleteness of the AS-level topology, However, it is extremely hard to convince any of them to provide their routing information and to participate in any RC. One of the main causes is that AS administrators are not stimulated enough to join, since no direct service is offered in exchange for their voluntary participation. Thus, it is not hard to understand that RCs are currently used mostly by large ISPs, which see in them a free opportunity to advertise their network reachability. For this reason, the Isolario project (Chapter 7) could be valuable, since it offers services based on the real-time analysis of inter-domain routing from different points of view in return for full routing tables, following the *do ut des* principle. Such services can be valuable for many ASes, ranging from local ISPs to CDNs, and would encourage their participation. Another alternative to improve the amount of routing data available is to exploit tools based on active probes. Some of the traceroute-based projects developed so far [CCP+09, Por, DIM, MMD+11] are able to bypass the reluctance in disclosing the routing information of AS owners by placing agents directly on user applications and, thus, obtaining data that would not otherwise be collected [FGI+14].

References

[AAG+14] Giovanni Accongiagioco, Eitan Altman, Enrico Gregori, Luciano Lenzini, et al. A game theoretical study of peering vs transit in the internet. In *Sixth IEEE International Workshop on Network Science for Communication Networks (NetSciCom 2014)(NetSciCom 2014)*, 2014.

[ACF+12] Bernhard Ager, Nikolaos Chatzis, Anja Feldmann, Nadi Sarrar, Steve Uhlig, and Walter Willinger. Anatomy of a Large European IXP. In *Proc. ACM SIGCOMM*, pages 163–174, 2012.

[AKW09a] Brice Augustin, Balachander Krishnamurthy, and Walter Willinger. IXPs: mapped? In *IMC*, pages 336–349, 2009.

[AKW09b] Brice Augustin, Balachander Krishnamurthy, and Walter Willinger. IXPs: mapped? In *IMC '09*, pages 336–349, 2009.

[Bas03] Hierarchy-aware algorithms for CDN proxy placement in the internet. *Computer Communications*, 26(3):251 – 263, 2003.

[BGP] BGP Monitoring System. http://bgpmon.netsec.colostate.edu/.

[BPP03] G. Di Battista, M. Patrignani, and M. Pizzonia. Computing the types of the relationships between autonomous systems. *IEEE INFOCOM*, 2003.

[CCG+02] Qian Chen, Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott Shenker, and Walter Willinger. The origin of power-laws in internet topologies revisited. In *INFOCOM*, 2002.

[CCP+09] Kai Chen, David R. Choffnes, Rahul Potharaju, Yan Chen, Fabian E. Bustamante, Dan Pei, and Yao Zhao. Where the Sidewalk Ends: Extending the Internet AS Graph Using Traceroutes from P2P Users. In *ACM CoNEXT '09*, pages 217–228, 2009.

[CGJ+04] Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott Shenker, and Walter Willinger. Towards Capturing Representative AS-level Internet Topologies. *Computer Networks*, 44(6):737–755, 2004.

[CHKW10] Xue Cai, John Heidemann, Balachander Krishnamurthy, and Walter Willinger. Towards an AS-to-Organization Map. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement (IMC '10)*, pages 199–205, 2010.

[Cis] Configuring a BGP Route Server. http://www.cisco.com/c/en/us/td/docs/ios/ios_xe/iproute_bgp/configuration/guide/2_xe/irg_xe_book/irg_route_server_xe.pdf. (working on March 2014).

[CJW06] Hyunseok Chang, Sugih Jamin, and Walter Willinger. To peer or not to peer: Modeling the evolution of the internet's as-level topology. 1001, 2006.

[Cou94] Olivier Coudert. Two-level Logic Minimization: an Overview. *Integration of the VLSI Journal*, 17(2):97–140, 1994.

[CR06] Rami Cohen and Danny Raz. The Internet Dark Matter - on the Missing Links in the AS Connectivity Map. In *Proc. IEEE INFOCOM*, pages 1–12, 2006.

[CRS12] Juan Camilo Cardona Restrepo and Rade Stanojevic. A History of an Internet Exchange Point. *ACM SIGCOMM Comput. Commun. Rev.*, 42(2):58–64, 2012.

[CSRL01] Thomas H. Cormen, Clifford Stein, Ronald L. Rivest, and Charles E. Leiserson. *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition, 2001.

[DAL+05] John C Doyle, David L Alderson, Lun Li, Steven Low, Matthew Roughan, Stanislav Shalunov, Reiko Tanaka, and Walter Willinger. The "robust yet fragile" nature of the internet. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41):14497–14502, 2005.

[DCDC12] Amogh Dhamdhere, Himalatha Cherukuru, Constantine Dovrolis, and KC Claffy. Measuring The Evolution of Internet Peering Agreements. In *Proc. IFIP-TC6 NETWORKING*, volume 2, pages 136–148, 2012.

[DIM] Distributed Internet MEasurement System. http://www.netdimes.org/new/.

[DKF+07] Xenofontas Dimitropoulos, Dmitri Krioukov, Marina Fomenkov, Bradley Huffaker, Young Hyun, Kimberley C. Claffy, and George Riley. AS Relationships: Inference and Validation. *ACM SIGCOMM Comput. Commun. Rev.*, 37(1):29–40, 2007.

[DKH+05] X. Dimitropoulos, D. Krioukov, B. Huffaker, kc claffy, and G. Riley. Inferring AS relationships: dead end or lively beginning? *4th Workshop on Efficient and Experimental Algorithms (WEA)*, 2005.

[DSA+11] Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 1–18. ACM, 2011.

[EHS02] T. Erlebach, A. Hall, and T. Schank. Classifying customer-provider relationships in the internet. *TIK-Report*, (145), July 2002.

[ENI11] ENISA. Inter-x: Resilience of the internet interconnection ecosystem. *Technical Report, Full Report*, Apr 2011. Available at http://www.enisa.europa.eu/activities/Resilience-and-CIIP/critical-infrastructure-and-services/inter-x/interx/report.

[FFF99] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *SIGCOMM Comput. Commun. Rev.*, 29(4):251–262, August 1999.

[FGI+14] Adriano Faggiani, Enrico Gregori, Alessandro Improta, Luciano Lenzini, Valerio Luconi, and Luca Sani. A Study on Traceroute Potentiality in Revealing the Internet AS-level Topology. In *Proc. of IFIP-TC6 NETWORKING*, 2014.

[FGL+12] Adriano Faggiani, Enrico Gregori, Luciano Lenzini, Simone Mainardi, and Alessio Vecchio. On the Feasibility of Measuring the Internet through Smartphone-based Crowdsourcing. In *Proc. of IEEE WiOpt*, pages 318–323, 2012.

[FRBM07] Ashley Flavel, Matthew Roughan, Nigel Bean, and Olaf Maennel. Modeling BGP Table Fluctuations. In *Proceedings of the 20th International Teletraffic Conference on Managing Traffic Performance in Converged Networks (ITC20 '07)*, pages 141–153, 2007.

[fro] F-root. http://www.isc.org/f-root/.

[FSR11] Alex Fabrikant, Umar Syed, and Jennifer Rexford. There's something about MRAI: Timing diversity can exponentially worsen BGP convergence. In *Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM '11)*, pages 2966–2974, 2011.

[Gao01a] L. Gao. On inferring autonomous system relationships in the internet. *IEEE/ACM TRANSACTIONS ON NETWORKING*, 9(6):733–745, December 2001.

[Gao01b] Lixin Gao. On Inferring Autonomous System Relationships in the Internet. *IEEE/ACM Transactions on Networking*, 9:733–745, 2001.

[GGR01] Lixin Gao, T.G. Griffin, and J. Rexford. Inherently Safe Backup Routing with BGP. In *Proceedings of the 20th IEEE International Conference on Computer Communications (INFOCOM '11)*, volume 1, pages 547–556, 2001.

[GIL+11] Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, and Luca Sani. BGP and Inter-AS Economic Relationships. In *Proc. IFIP-TC6 NETWORKING*, volume 2, pages 54–67, 2011.

[GILO10] Enrico Gregori, Alessandro Improta, Luciano Lenzini, and Chiara Orsini. The Impact of IXPs on the AS-level Topology Structure of the Internet. *Comput. Commun.*, 34(1):68–82, 2010.

[Gim65] James F. Gimpel. A Reduction Technique for Prime Implicant Tables. *IEEE Trans. Electronic Computers*, 14(4):535–541, 1965.

[GJ90] Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1990.

[GLLV13] Enrico Gregori, Luciano Lenzini, Valerio Luconi, and Alessio Vecchio. Sensing the Internet through crowdsourcing. In *Proc. of IEEE PerMoby*, pages 248–254, 2013.

[HG12] Syed Hasan and Sergey Gorinsky. Obscure Giants: Detecting the Provider-Free ASes. In *Proceedings of the 11th International IFIP TC-6 Conference on Networking (NETWORKING '12)*, volume 2, pages 149–160, 2012.

[Hoc97] Dorit S. Hochbaum. *Approximation Algorithms for NP-hard Problems*. PWS Publishing Co., Boston, MA, USA, 1997.

[HRA10] Geoff Huston, Mattia Rossi, and Grenville Armitage. A Technique for Reducing BGP Update Announcements Through Path Exploration Damping. *IEEE Journal on Selected Areas in Communications*, 28(8):1271–1286, 2010.

[HSFK09] Y. He, G. Sigano, M. Faloutsos, and S. Krishnamurthy. Lord of the links: A framework for discovering missing links in the internet topology. *IEEE/ACM TRANSACTIONS ON NETWORKING*, 17(2), April 2009.

[IAN] Autonomous system (as) numbers. https://www.iana.org/assignments/as-numbers/as-numbers.xml.

[Iea13] Kunihiro Ishiguro et al. Configuring Quagga as a Route Server. In *Quagga. A routing software package for TCP/IP networks*, pages 75–86. January 2013. http://www.nongnu.org/quagga/docs/quagga.pdf.

[Iso] Isolario project. http://www.isolario.it.

[KFR09] Josh Karlin, Stephanie Forrest, and Jennifer Rexford. Nation-state routing: Censorship, wiretapping, and bgp. *arXiv preprint arXiv:0903.3218*, 2009.

[KMT06] S. Kosub, M. G. Maaß, and H. Taubig. Acyclic type-of-relationship problems on the internet. *CAAN'06, LNCS #4235, pp. 98–111. Springer*, 2006.

[LABJ01] Craig Labovitz, Abha Ahuja, Abhijit Bose, and Farnam Jahanian. Delayed Internet Routing Convergence. *IEEE/ACM Transactions on Networking*, 9(3):293–306, 2001.

[LF04] R. Larson and E. Farber. *Elementary Statistics: Picturing the World*. Prentice Hall College Div, 5th edition, 2004.

[LHD+13] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, and kc Claffy. AS Relationships, Customer Cones, and Validation. In *Proc. ACM SIGCOMM IMC*, pages 243–256, 2013.

[lib] libbgpdump. https://bitbucket.org/ripencc/bgpdump/wiki/Home.

[LOZZ07] Mohit Lad, Ricardo Oliveira, Beichuan Zhang, and Lixia Zhang. Understanding resiliency of internet topology against prefix hijack attacks. In *Dependable Systems and Networks, 2007. DSN'07. 37th Annual IEEE/IFIP International Conference on*, pages 368–377. IEEE, 2007.

[Max] Maxmind GeoIPLite database. http://www.maxmind.com/app/geoip_country.

[McC56] Edward J. McCluskey. Minimization of Boolean Functions. *Bell System Technical Journal*, 35(6):1417–1444, 1956.

[MFM+06] Wolfgang Mühlbauer, Anja Feldmann, Olaf Maennel, Matthew Roughan, and Steve Uhlig. Building an as-topology model that captures route diversity. In *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '06)*, pages 195–206, 2006.

[MGVK02] Zhuoqing Morley Mao, Ramesh Govindan, George Varghese, and Randy H. Katz. Route Flap Damping Exacerbates Internet Routing Convergence. *SIGCOMM Comput. Commun. Rev.*, 32(4):221–233, 2002.

[MMD+11] P. Marchetta, P. Mérindol, B. Donnet, A. Pescapé, and J.-J. Pansiot. Topology discovery at the router level: a new hybrid tool targeting ISP networks. *IEEE JSAC, Special Issue on Measurement of Internet Topologies*, 29(6), October 2011.

[MRT] Multi-threaded routing toolkit (mrt) routing information export format. http://tools.ietf.org/html/rfc6396.

[MWA02] Ratul Mahajan, David Wetherall, and Tom Anderson. Understanding BGP Misconfiguration. In *Proceedings of the 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '02)*, pages 3–16, 2002.

[Nor11] William B. Norton. *The Internet Peering Playbook: Connecting to the Core of the Internet*. DrPeering Press, Palo Alto, CA, 2011.

[OC01] T. Sasao O. Coudert. *Two-Level Logic Minimization, Logic Synthesis and Verification*. Kluwer Academic Publishers, 2001.

[OPW+10] Ricardo Oliveira, Dan Pei, Walter Willinger, Beichuan Zhang, and Lixia Zhang. The (in)completeness of the Observed Internet AS-level Structure. *IEEE/ACM TON*, 18(1):109–122, 2010.

[OWS+13] Catherine Olschanowsky, Lawrence M. Weikum, Jason Smith, Christos Papadopoulos, and Dan Massey. Delivering Diverse BGP Data in Real-time and Through Multi-Format Archiving. In *Proc. of IEEE HTS*, 2013.

[OZP+06] Ricardo Oliveira, Beichuan Zhang, Dan Pei, Rafit Izhak-Ratzin, and Lixia Zhang. Quantifying Path Exploration in the Internet. In *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement (IMC '06)*, pages 269–282, 2006.

[OZZ07] Ricardo V Oliveira, Beichuan Zhang, and Lixia Zhang. Observing the evolution of internet as topology. *ACM SIGCOMM Computer Communication Review*, 37(4):313–324, 2007.

[PCH] Packet Clearing House. http://www.pch.net.

[PMM+11] Cristel Pelsser, Olaf Maennel, Pradosh Mohapatra, Randy Bush, and Keyur Patel. Route Flap Damping Made Usable. In *Proc. of PAM*, pages 143–152, 2011.

[Por] University of Pisa Portolan project. http://portolan.iet.unipi.it.

[PSV04] Romualdo Pastor-Satorras and Alessandro Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press, New York, NY, USA, 2004.

[PUK+11] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. Ip geolocation databases: Unreliable? *ACM SIGCOMM Computer Communication Review*, 41(2):53–56, 2011.

[pyb] pybgpdump. https://jon.oberheide.org/pybgpdump/.

[PZMZ06] Dan Pei, Beichuan Zhang, Daniel Massey, and Lixia Zhang. An Analysis of Convergence Delay in Path Vector Routing Protocols. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 50(3):398–421, 2006.

[Qua] Quagga routing software suite. http://www.nongnu.org/quagga/.

[Qui55] Willard V. Quine. A Way to Simplify Truth Functions. *The American Mathematical Monthly*, 62(9):627–631, 1955.

[Qui59] Willard V. Quine. On Cores and Prime Implicants of Truth Functions. *The American Mathematical Monthly*, 66(9):755–760, 1959.

[RFCa] Bgp route flap damping. http://tools.ietf.org/html/rfc2439.

[RFCb] A border gateway protocol 4 (bgp-4). http://tools.ietf.org/html/rfc4271.

[RFCc] Guidelines for creation, selection, and registration of an autonomous system (as). http://tools.ietf.org/html/rfc1930.

[RFCd] Terminology for benchmarking bgp device convergence in the control plane. http://tools.ietf.org/html/rfc4098.

[RIS] RIPE NCC Routing Information Service. http://www.ripe.net/data-tools/stats/ris/routing-information-service.

[Rou] University of Oregon Route Views Project. http://www.routeviews.org.

[RRW10] Amir Hassan Rasti, Reza Rejaie, and Walter Willinger. Characterizing the global impact of p2p overlays on the as-level underlay. In *Passive and Active Measurement*, pages 1–10. Springer, 2010.

[RTY+00] Pavlin Radoslavov, Hongsuda Tangmunarunkit, Haobo Yu, Ramesh Govindan, Scott Shenker, and Deborah Estrin. On characterizing network topologies and analyzing their impact on protocol design. Technical report, 2000.

[RWM+11a] M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush. 10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems. *IEEE JSAC*, 29(9):1810–1821, 2011.

[RWM+11b] Matthew Roughan, Walter Willinger, Olaf Maennel, Debbie Perouli, and Randy Bush. 10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems. *IEEE Journal on Selected Areas in Communications*, 29(9):1810–1821, 2011.

[RWXZ02] Jennifer Rexford, Jia Wang, Zhen Xiao, and Yin Zhang. BGP routing stability of popular destinations. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement (IMW '02)*, pages 197–202, 2002.

[SARK02a] L. Subramanian, S. Agarwal, J. Rexford, and R.H. Katz. Characterizing the internet hierarchy from multiple vantage points. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 618–627 vol.2, 2002.

[SARK02b] Lakshminarayanan Subramanian, Sharad Agarwal, Jennifer Rexford, and Randy H. Katz. Characterizing the Internet hierarchy from multiple vantage points. In *IEEE INFOCOM*, volume 2, pages 618–627, 2002.

[SKM06] Amit Sahoo, Krishna Kant, and Prasant Mohapatra. Characterization of BGP Recovery Time under Large-Scale Failures. In *Proceedings of the IEEE International Conference on Communications (ICC '06)*, volume 2, pages 949–954, 2006.

[SZ11] Yuval Shavitt and Noa Zilberman. A geolocation databases study. *Selected Areas in Communications, IEEE Journal on*, 29(10):2044–2056, 2011.

[WAD09] Walter Willinger, David Alderson, and John C Doyle. *Mathematics and the internet: A source of enormous confusion and great potential*. Defense Technical Information Center, 2009.

[Wik] Wikipedia - Tier 1 network. http://en.wikipedia.org/wiki/Tier_1_network.

[Woe03] Gerhard J. Woeginger. Exact Algorithms for NP-Hard Problems: A Survey. *Combinatorial Optimization - Eureka, You Shrink! - LNCS*, 2570:185–207, 2003.

[WZMS07] Jian Wu, Ying Zhang, Z Morley Mao, and Kang G Shin. Internet routing resilience to failures: analysis and implications. In *Proceedings of the 2007 ACM CoNEXT conference*, page 25. ACM, 2007.

[XDZC04] Kuai Xu, Zhenhai Duan, Zhi-Li Zhang, and Jaideep Chandrashekar. On properties of Internet exchange points and their impact on AS topology and relationship. In *IFIP-TC6 Networking*, pages 284–295, 2004.

[XG04] J. Xia and L. Gao. On the evaluation of as relationship inferences. *IEEE GLOBECOM*, 2004.

[XMH11] Xueyang Xu, Z Morley Mao, and J Alex Halderman. Internet censorship in china: Where does the filtering occur? In *Passive and Active Measurement*, pages 133–142. Springer, 2011.