BGP-Multipath Routing

in the Internet

Jie Li

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Computer Science

University College London

October 20, 2021

I, Jie Li, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

BGP-Multipath, or BGP-M, is a routing technique for balancing traffic load in the Internet. It enables a Border Gateway Protocol (BGP) border router to install multiple 'equally-good' paths to a destination prefix. While other multipath routing techniques are deployed at internal routers, BGP-M is deployed at border routers where traffic is shared on multiple border links between Autonomous Systems (ASes). Although there are a considerable number of research efforts on multipath routing, there is so far no dedicated measurement or study on BGP-M in the literature.

This thesis presents the first systematic study on BGP-M. I proposed a novel approach to inferring the deployment of BGP-M by querying Looking Glass (LG) servers. I conducted a detailed investigation on the deployment of BGP-M in the Internet. I also analysed BGP-M's routing properties based on traceroute measurements using RIPE Atlas probes. My research has revealed that BGP-M has already been used in the Internet. In particular, Hurricane Electric (AS6939), a Tier-1 network operator, has deployed BGP-M at border routers across its global network to hundreds of its neighbour ASes on both IPv4 and IPv6 Internet. My research has provided the state-of-the-art knowledge and insights in the deployment, configuration and operation of BGP-M. The data, methods and analysis introduced in this thesis can be immensely valuable to researchers, network operators and regulators who are interested in improving the performance and security of Internet routing. This work has raised awareness of BGP-M and may promote more deployment of BGP-M in future because BGP-M not only provides all benefits of multipath routing but also has distinct advantages in terms of flexibility, compatibility and transparency.

Impact Statement

The Internet has experienced a rapid increase of users and traffic volume. In the meantime, it has been suffering from a series of challenges, such as the routing delays, network attacks and link failures. To resolve these challenges, various techniques have been proposed. Among these techniques, BGP-Multipath (BGP-M) allows a border router to learn and install multiple 'equally-good' paths to the same destination prefix.

My research work is valuable to both the academia and the industry as the first systematic study on BGP-M. It fills the gap by providing the state-of-the-art knowledge on the deployment, the unique characteristics, and the routing properties of BGP-M. The novel knowledge helps to understand the Internet routing paths and locate the network problems. With increasing interest in my research, more and more researchers, technicians and managers become aware of BGP-M and its benefits. Therefore, BGP-M will be deployed by more networks and will help to improve the Internet routing performance.

The deployment of BGP-M is beneficial to the Internet routing performance because (1) it helps to achieve load balancing between two neighbouring ASes, (2) it is beneficial to routing performance by reducing congestion and improving network resilience against link failures and sudden traffic surge, (3) it can help network operators optimise their agreements with neighbour ASes, (4) it is easy to be implemented without modification to the BGP process, and (5) it relies on existing multiple border links without producing extra burdens to the infrastructure.

Acknowledgements

My most sincere thanks and appreciation go to my supervisor, Dr. Shi Zhou. He is very professional, ensuring that I am delivering accurate and valuable information to the community. He is very patient with all of my questions. He is very encouraging when I encountered problems and frustration. He is very caring about my daily life and wellbeing, especially during the COVID-19 pandemic.

Then, I shall express my gratitude to Dr. Vasileios Giotsas from Lancaster University, and Dr. Yangyang Wang from Tsinghua University. I have been working closely with them during my PhD. Their ideas, suggestions, comments, and inputs have played an important role in improving the quality of my works. I am also thankful to the examiners for my vivas, the anonymous reviewers for my papers, the professionals during my attendance to academic conferences, and the technicians from Hurricane Electric and RIPE. Their feedbacks were very helpful and valuable to my research.

Moreover, I am grateful to China Scholarship Council for the financial support with No. 201406060022, and University College London for the Faculty of Engineering Postgraduate Research Scholarship. Both scholarships made my dream come true to study in UCL and finish My PhD research.

My special thanks goes to my friends and my colleagues in UK, whose company and support made my life in UK full of laughter and happiness.

Finally, I thank my family for their endless support and love to me. It is easy to feel homesick for an overseas student. It is their support and love that makes me to be fully devoted to my research and finish my PhD.

Contents

Acronyms

APAR	Analytic and Probe-based Alias Resolver
APNIC	Asia Pacific Network Information Centre
APPLE	Alias Pruning by Path Length Estimation
AS	Autonomous System
ASN	AS Number
BGP	Border Gateway Protocol
BGP-M	BGP-Multipath
BGP-XM	Border Gateway Protocol-eXtended Multipath
BGPSEC	BGP Security
CAIDA	Center for Applied Internet Data Analysis
CPU	Central Processing Unit
DDoS	Distributed Denial of Service
DIMR	Disjoint Interdomain Multipath Routing
eBGP	external BGP
ECMP	Equal-Cost Multi-Path
EGP	Exterior Gateway Protocol
ETMP-BGP	Effective Tunnel-based Multi-Path BGP
Euro-IX	European Internet Exchange Association
HTTP	Hypertext Transfer Protocol
IANA	Internet Assigned Numbers Authority
iBGP	internal BGP
ICMP	Internet Control Message Protocol
IETF	Internet Engineering Task Force

IGP	Interior Gateway Protocol
IP	Internet Protocol
IRR	Internet Routing Registry
ISP	Internet Service provider
ITDK	Macroscopic Internet Topology Data Kit
IXP	Internet eXchange Point
LDoS	Low-rate Denial of Service
LG	Looking Glass
MAC	Medium Access Control
MBGP	Multi-path BGP
MCA	Multipath Classification Algorithm
MDA	Multipath Detection Algorithm
MED	Multi Exit Discriminator
MIDAR	Monotonic ID-Based Alias Resolution
MIRO	Multi-path Interdomain ROuting
ML	Machine Learning
MOAS	Multiple Origin ASes
MPLS	Multiprotocol Label Switching
MPTCP	Multi-Path TCP
MRAI	Mininum Routing Advertising Interval
OSPF	Open Shortest Path First
PCH	Packet Clearing House
PoP	Point-of-Presence
RAM	Random Access Memory
RIS	Routing Information Service
RIP	Routing Information Protocol
RIPE NCC	Réseaux IP Européens Network Coordination Centre
ROA	Route Origin Authorizations
RPKI	Resource Public Key Infrastructure
RPSL	Routing Policy Specification Language

RTT	Round Trip Time
SDN	Software-Defined Networking
TCP	Transport Control Protocol
TTL	Time-to-Live
TE	Traffic Engineering
UDP	User Datagram Protocol

Notations and Descriptions

Notation	Description
SrcIP	Source IP address
DstIP	Destination IP address
DstPrfx	Destination Prefix
NearAS	Nearside AS
NearBR	Nearside border router
NearIP	IP address of ingress interface of *NearBR*
FarAS	Farside AS
FarBR	Farside border router
FarIP	IP address of ingress interface of *FarBR*
BL	Border link between ASes

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Internet Routing

There is a dramatic increase of Internet users and traffic volume in the last two decades. The Internet faces a number of challenges, including routing delays, network attacks and link failures, to name just a few. These challenges can negatively affect the performance of Internet routing.

To tackle these challenges, researchers have proposed various methods and designs. Multipath routing is one of such efforts. It allows a router to use multiple paths to deliver traffic, which is called load balancing or traffic sharing. This traffic engineering technique has been widely used by network operators to improve their networks' routing performance in terms of reduced congestion and increased resilience and security.

There are a number of load balancing schemes for multipath routing, including per-packet load balancing, per-destination load balancing, per-session load balancing, per-application load balancing, and per-flow load balancing. Researchers have measured [41, 42, 47, 173] the deployment of multipath routing using traceroute. They showed that multipath routing has already been extensively deployed in the Internet, with more than 4 millions of cases observed [173]. These results indicate that multipath routing plays an increasingly important role for Internet routing.

1.2 BGP-Multipath (BGP-M)

BGP-Multipath (BGP-M) is a special type of multipath routing. Different from other multipath routing techniques deployed at internal routers, BGP-M is deployed at Border Gateway Protocol (BGP) border router connecting between different Autonomous Systems (ASes).

Although there are many research works and measurements on multipath routing due to its significant importance to Internet routing, there is no dedicated measurement or study on BGP-M in the literature. So far, BGP-M is largely overlooked by the research community.

My measurement results in this thesis will show, however, that BGP-M has already been deployed by several ASes, including large transit ASes and stub ASes. This is because BGP-M not only provides all benefits that multipath routing can achieve, including traffic load sharing and increased resilience to link failures, but also has many distinctive advantages.

Firstly, it is well known that there can be many border links between two ASes, in particular large ASes in the core of the Internet. These links usually have high bandwidths. Instead of using only one border link by default, BGP-M enables border routers to use multiple border links for traffic routing to a destination, and therefore achieve fuller usage of available links and bandwidth resources which are already there.

Secondly, BGP-M utilises the BGP mechanism. It is compatible to existing protocols and configurations. Indeed, network operators can activate BGP-M on a border router by changing only one parameter. Also the BGP mechanism can enable automatic reaction to network changes and disruptions by switching traffic to other border links.

Thirdly, BGP-M is transparent to other ASes, thus a network operator can deploy BGP-M independently and therefore receive all the benefits without needing any support or agreement from other operators.

I choose to study BGP-M because despite the above benefits, there is only very limited information or knowledge of it, and more importantly, there is no data on

it at all. It is necessary to conduct research on BGP-M to fill this gap, to provide knowledge to the research community and the industry for a better understanding about this technique. Such that, better agreements and solutions can be proposed to solve the existing problems in the Internet routing.

1.3 Research Questions

There are many challenges for a study on BGP-M. Firstly, there are only very limited technical documentations on BGP-M from router vendors and IETF working groups and a handful research papers that merely mentioned the possible existence of BGP-M. Secondly, there is no measurement data for BGP-M at all. Past measurements on multipath routing relied on traceroute, but it is difficult to use traceroute to discover the deployment of BGP-M because it would require complicated traceroute probings, ideally, from all ASes to all destinations. The real obstacle, however, is the inherent difficulty in mapping the border of a network [184]. There have been proposals to map AS borders. They are still not accurate enough due to the third-party IP address issue and the usage of layer-2 switching devices at AS borders. Hence, so far there is no accurate method or dataset for such mapping.

Given the urgent need for study on BGP-M and the challenges we face, this research aims to investigate the following research questions. Our objective is to produce a comprehensive understanding on deployment and properties of BGP-M.

- How to discover a deployment of BGP-M? We need to identify a trustworthy data source and a method to reveal reliable information on the deployment of BGP-M in the Internet.

- Whether and how widely has BGP-M been deployed in the Internet? To answer this question, we need to uncover which network operators have deployed BGP-M with which of their neighbour ASes. As the first measurement, we can focus on how to produce a summary picture of the deployment of BGP-M in the global Internet, instead of a complete measurement.

- How does an AS use and deploy BGP-M? After we have a general picture, we can focus on a typical AS and study its deployment of BGP-M. For example,

are there any patterns in the connectivity fabric of its deployment of BGP-M? Which types of neighbour ASes of the AS are deployed with BGP-M? What about the geographical properties of the deployment of BGP-M? Can BGP-M be deployed on both IPv4 and IPv6?

- What are the routing properties of BGP-M? For example, which load balancing schemes are used? Does BGP-M interfere with the routing in neighbour ASes? Are traffic equally allocated on border links with different bandwidths?

1.4 My Research Contributions

1.4.1 New Method to Infer the Deployment of BGP-M

I proposed the first method to discover and measure the deployment of BGP-M in the Internet. The method is based on queries to Looking Glass (LG) servers. The proposed method provides reliable results because LG servers can reveal the actual configuration information and routing tables installed at border routers. The LG data contain detailed and abundant information which can allow us to not only infer the deployment of BGP-M but also reveal the relevant configurations.

1.4.2 State-of-the-art Knowledge on Deployment of BGP-M in Internet

I provided a rich set of knowledge on the deployment of BGP-M in the Internet. Most importantly, it has been deployed by large transit ASes as well as stub ASes around the world, on both IPv4 and IPv6. On average, each BGP deployment can be used for traffic routing to more than 10 destination prefixes in the farside AS alone. I conducted an in-depth analysis on a Tier-1 network operator, Hurricane Electric (HE, AS6939). I observed that BGP-M is often deployed via IXPs. There is a tendency for HE to connect content provider networks with BGP-M. All of the identified BGP-M cases are deployed at Cisco routers or Juniper routers. HE has been actively and constantly adjusting and maintaining its deployment of BGP-M.

1.4.3 Understanding the Routing Properties of BGP-M

While I relied on LG data for discovering BGP-M, I used traceroute probings to obtain routing properties of known BGP-M cases. I studied the load balancing schemes of BGP-M at different types of routers for different types of traffic. Notably, I revealed that UDP packets and ICMP packets are often handled differently. I also investigated the routing delays on border links based on Round Trip Time (RTT) data.

Most of the material in this thesis are published or under review by IEEE conferences or journals (see Appendix B).

Chapter 2

Background

2.1 Basics of Internet Routing

This section introduces some basic concepts and background knowledge in Internet routing. Those who are familiar with these topics can skip this section.

2.1.1 End-to-end IP-level Routing

The Internet has developed for decades and consists of various kinds of nodes and links. The Internet processes a tremendous volume of traffic generated from the network users through daily activities like shopping, chatting, holding state-scale conferences, watching video streams and working remotely.

Routers are responsible to transmit the traffic from host to host based on routing protocols like Routing Information Protocol (RIP) [100], Open Shortest Path First (OSPF) [144] and Border Gateway Protocol (BGP) [154]. The information for transmission is encapsulated into a data packet along with its source and destination information. The source and destination information is stored as 32-bit address in IPv4 or 128-bit address in IPv6. A router uses a routing table to store the routing information for each routable destination IP. Destination IP addresses with the same next hop are merged into prefixes of common bits. Upon receiving a data packet, a router maps the destination IP address to a prefix with longest prefix matching, then checks whether there are any routes for the prefix in its routing table. If yes, the router delivers the packet via the best route; otherwise, the packet is dropped.

Normally, in Internet routing, there are multiple intermediate routers between

the source IP and the destination IP along the traffic direction, as shown in Figure 2.1. Each intermediate router receives the data packet via an ingress interface. The IP address of the ingress interface is denoted as an IP hop in the routing path. Each router has only one best route to the destination IP, so there is only one single best routing path between the source IP and the destination IP. For example, in Figure 2.1, the red line represents the best routing path from the source IP to the destination IP.

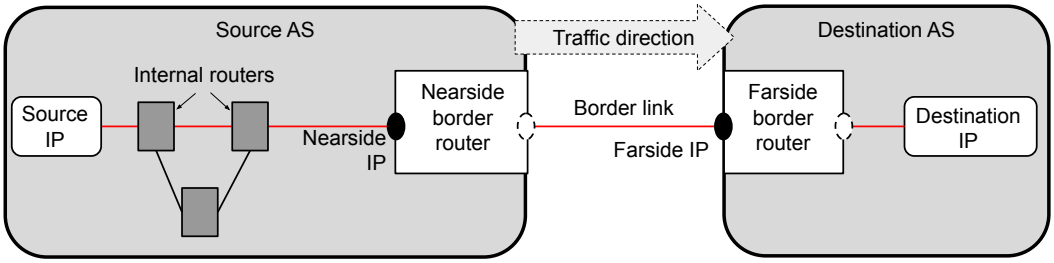


Figure 2.1: Illustrative example of Internet best-path routing, where the path with red links is used for routing between a source IP address and a destination IP address.

Researches on IP-level Internet routing have been conducted for decades, covering various topics, such as topology discovery [75, 176], path inference [130, 135] and IP geolocation [70, 80, 123, 139, 180].

2.1.2 Inter-domain Routing at Autonomous System (AS) Level

2.1.2.1 Autonomous System (AS)

The Internet consists of thousands of ASes, also called Internet domains. Each AS is allocated an AS Number (ASN) and a number of IP prefixes by the Internet Assigned Numbers Authority (IANA). An AS deploys the allocated IP prefixes to its routers and announces its prefixes to its neighbour ASes.

The Internet routing process is realised via inter-domain routing and intra-domain routing. For the inter-domain routing, ASes use a common External Gateway Protocol (EGP) to accomplish routing tasks, and the current de-facto protocol is BGP. Each AS has a number of routers implemented with BGP rules, and these routers are called BGP speakers. BGP speakers exchange BGP information with each other. For the intra-domain routing, the network operators of an AS apply an

Interior Gateway Protocol (IGP) to manage the network. Common IGPs include RIP and OSPF.

2.1.2.2 Border Gateway Protocol (BGP)

Two BGP speakers establish and maintain a BGP session via the exchange of different BGP messages. When the BGP messages are exchanged within an AS, it is called internal BGP (iBGP). When the BGP messages are exchanged between ASes, it is called external BGP (eBGP). After a BGP session is established, BGP speakers process the messages according to the BGP policies. *Import* policies determine whether a received route advertisement will be accepted or not. After a path to a prefix is accepted and stored in Routing Information Base (RIB), all the available paths to the same prefix are ranked, and the best path to the prefix will be selected and advertised to neighbour ASes. *Export* policies determine which neighbours can receive a route advertisement.

In BGP routing, BGP uses several attributes to rank the candidate paths to select the best path for traffic to a destination prefix. The attributes include Local Preference (`LocPref`), AS path, Origin, Multi Exit Discriminator (MED), eBGP/iBGP, IGP metric and router ID. For each attribute, there is a rule. When the candidate paths have different values for `LocPref`, the path with the highest `LocPref` value is selected. If multiple paths exist with the highest `LocPref` value, the path with the shortest AS path is selected. Otherwise, the other attributes are considered as tie-breakers according to the priorities given in Table 2.1.

Table 2.1: BGP best path selection algorithm

Priority	**Attribute**	**Preference Rule**
1	`LocPref`	Highest `LocPref`
2	AS Path	Shortest AS Path
3	Origin	Lowest Origin type (IGP $<$ EGP $<$ INCOMPLETE)
4	MED	Lowest MED
5	eBGP/iBGP	Prefer eBGP over iBGP paths
6	Metric	Lowest IGP metric
7	Router ID	Lowest Router ID

2.1.2.3 Border Router

A border router (or AS border router, or BGP border router) is located at the boundary of an AS with at least one interface connecting to an intra-domain router and at least one interface connecting to a border router in a neighbour AS. A border router is implemented with BGP. It can establish and maintain BGP sessions to exchange routing information with other ASes via BGP messages, and then update its routing table according to the network operator's policy configurations. For example, in Figure 2.1, there are two border routers, i.e., the nearside border router and the farside border router, located at the boundaries of the source AS and the destination AS, respectively.

2.1.2.4 Border Link

An inter-domain border link is a physical IP-level link connecting the border routers of two neighbouring ASes. As illustrated in Figure 2.1, depending on traffic direction, the border link starts from an egress interface of the nearside border router, and ends at an ingress interface of *FarBR*. Since the egress interface of the nearside border router is invisible in traceroute measurement, a border link is denoted by the IP addresses of the ingress interfaces of the two border routers, which can be identified as two consecutive IP addresses on a traceroute path that are mapped to the nearside AS and the farside AS. These IP addresses are called nearside IP and farside IP. For example, in Figure 2.1, the border link can be denoted by the pair of (Nearside IP, Farside IP).

2.1.2.5 AS-level Routing Paths

By default, there should be only one best path for a destination prefix installed in the routing table, which means there is only one AS-level path for a destination prefix. However, to improve the path utilisation and achieve load balancing, multiple AS-level routing paths have been used for inter-domain routing.

To use multiple AS-level paths, some methods modify the BGP best path selection process or use BGP update for multipath routing. Xu and Rexford [183] proposed Multi-path Interdomain ROuting (MIRO) where routers could learn de-

fault routes, and arbitrary domain pairs could negotiate to use additional paths. Fujinoki [77] presented Multi-path BGP (MBGP) to solve the problems caused by conventional BGP by dynamically utilising concurrent multiple BGP paths without routing loops. Beijnum [50] proposed to modify the BGP best path selection process by removing some tie breaking rules. The method proposed in [50] announced paths with longest AS_PATH to upstream ASes, and took actions to avoid compromising loop-freeness. Camacho *et al.* [57] presented Border Gateway Protocol-eXtended Multipath (BGP-XM), and the method merged into regular BGP updates information from paths which may even traverse different ASes.

Multiple AS-level routing paths can also be achieved by other approaches. Araújo *et al.* [44] studied the multipath routing with congestion charging, and focused on building a multipath routing architecture with existing congestion pricing models. Yin *et al.* [185] proposed Disjoint Interdomain Multipath Routing (DIMR) to help ASes discover two disjoint paths for each destination AS. Garcia Gomez *et al.* [78] introduced Effective Tunnel-based Multi-path BGP (ETMP-BGP) which used Software-Defined Networking (SDN) techniques to obtain whole control of tunnel-based multi-path BGP routing in terms of AS-level routing. Wang *et al.* [175] presented a route selection algorithm that calculated multiple paths based on geographical distance metric.

Several survey papers [120, 152, 167, 181] discussed the existing researches using multiple AS-level routing paths from different perspectives. As of this writing, there is no report on any of these methods being deployed in the Internet.

2.1.3 Challenges in Internet Routing

The Internet has experienced rapid increase of users and witnessed the expansion of traffic volume [62]. In the meantime, the Internet has been facing with various challenges, e.g. increased routing delays, various network attacks and link failures. These challenges can result in hindered routing performance and user experience. To tackle these challenges, a number of technical efforts have been proposed.

2.1.3.1 Routing Delay and Congestion

The increasing demand for high-bandwidth content (e.g. streaming video) can cause stress to the Internet, especially when the Internet traffic has entered the "Zettabyte Era" according to Cisco's white paper published in 2016 [61]. When the capacity of network resources are unable to handle the increased traffic, delays and congestion occur.

Generally, the traffic delay can be generated from various sources [54], e.g. structural delays, delays from the interaction between endpoints, delay along transmission paths, delays related to link capacities and intra-end-host delays. Higher delay and congestion can result in packet loss, hindered user experience, the loss of customers and the decrease of profit for Internet service providers (ISPs). For example, real-time games have their maximum allowed delay values [158] to keep users; and 100 milliseconds delay can bring a significant loss of sales to Akamai [186].

Many methods have been proposed to reduce routing delays. The survey paper by Briscoe *et al.* [54] has classified the relevant techniques based on the sources of routing delays. For example, multipath routing techniques like Equal-Cost Multi-Path (ECMP) can reduce the structural delay by sharing the capacity of multiple parallel links.

The techniques on machine learning (ML) have also been applied to control the Internet congestion [103]. ML has contributed in the aspects of traffic classification, traffic prediction, available bandwidth measurement and network topology discovery.

The techniques analysed in [54] and [103] ranged from link layer to application layer. Silva and Mota [166] narrowed down the techniques to those for reducing BGP routing convergence delay. Specifically, Silva and Mota [166] classified the techniques into five approaches, which are listed below with their advantages.

- Speeding up mainly relies on optimising Mininum Routing Advertising Interval (MRAI) timer to reduce the convergence time.
- Limiting path exploration addresses the root cause of BGP convergence delay

and eliminates inconsistency issue.

- Efficient policy configuration addresses inconsistency caused by policy conflicts and deals with routing oscillations.

- Multipath and multi-path forwarding speeds up convergence time by allowing routers to quickly select alternative paths without path exploration in RIB.

- Centralised control allows operators to manage their networks easily without management complexity.

The BGP-Multipath (BGP-M) studied in this thesis is a multipath routing technique. It uses multiple IP-level routing paths (learned via multiple border links) for traffic delivery, the delays on each border link will be reduced, and the risk of congestion is lower, compared to the routing via single border link.

2.1.3.2 Network Attacks

Different types of network attacks exist for various purposes and targets, including damage to the physical links, injection of malicious information to the data transmission and forgery of routing information [40, 141, 165]. These attacks to a network causes extra load to routers, instabilities and connectivity problems. Network attacks can also cause the crash of a network and unavailability of service for hours, resulting in loss of profit.

Moreover, the attacks are evolving [182]. For example, Distributed Denial of Service (DDoS) attacks have been damaging the Internet for more than 20 years, and it has evolved from large-traffic and high-rate attack to small-traffic and low-rate attacks.

These attacks also threat BGP inter-domain routing. Low-rate DoS (LDoS) attack can update the routing information of border routers repeatedly and decrease the performance of the border routers [182]. False IP prefix and false routes can be propagated into BGP messages and cause BGP traffic hijacking [165]. Route leak can cause interruption of Internet service, especially when non-customer route is advertised over a peer or a customer link [165].

The existing research efforts for coping with network attacks can be classified into two groups, i.e. detection and defense.

The survey paper by Al-Musawi *et al.* [40] studied the techniques for BGP anomaly detection. The authors firstly classified the existing BGP anomalies into four categories, i.e. direct intended anomaly, direct unintended anomaly, indirect anomaly and link failure. The direct intended anomalies are related to network attacks like prefix hijacking. The other anomalies are caused by misconfigurations or Internet components (e.g. Web servers). Then the authors grouped the BGP anomaly detection methods into five classes, i.e. time series analysis, machine learning, statistical pattern recognition, validation of BGP updates based on historical BGP data and reachability checks. The methods based on statistical pattern recognition, validation of BGP updates based on historical BGP data were able to detect direct intended anomalies.

Defense methods have been reviewed by Siddiqui *et al.* [165] and Mitseva *et al.* [141]. Siddiqui *et al.* [165] focused on the methods proposed by the IETF Secure Inter-Domain Routing Working Group. These methods include Resource Public Key Infrastructure (RPKI), Route Origin Authorisations (ROAs) and BGP Security (BGPSEC). The ROA targets prefix origin authorisation, the BGPSEC protocol addresses AS-Path validation and the RPKI facilitates ROA and BGPSEC in achieving their goals.

Mitseva *et al.* [141] reviewed the existing BGP security methods from both control plane and data plane. The methods from control plane aim to verify that incoming data is not removed, modified or replayed during the transmission. The methods from data plane rely on traceroute to check path consistency between the control plane and the data plane.

The BGP-M technique studied in this thesis can use as many as six border links to balance the traffic load. This increases the network's resilience to network attacks like DDoS, because when a border link is affected by these attacks, the other links can still be available for traffic delivery, thus avoid or lower the risk of damage caused by the attack.

2.1.3.3 Link Failures

Link failures have been reported as the most failures in Internet [60]. A link can fail for various reasons, such as destruction caused by natural disasters, cable cut caused by digger, power outage of facilities, configuration change or maintenance, congestion and attacks. Among these failures, physical link failures are dominant scenarios and those caused by natural forces are raising [60].

Networks should be able to deal with and recover from link failures to achieve stable services, network resilience and robustness [60]. The faster a network recovers from link failures, the less revenue it loses.

According to the survey paper by Al-Musawi *et al.* [40], link failures can be detected by methods based on machine learning, statistical pattern recognition, and reachability checks.

Chiesa *et al.* [60] provided a comprehensive survey on the fast recovery mechanisms on data plane dealing with the link failures. The methods were analysed on different layers. Specifically, link-layer methods were able to achieve quick recovery based on any of the four techniques: single spanning tree, multiple spanning trees, recovery tables and message flooding and deduplication. Multiprotocol Label Switching (MPLS) fast recovery were reported to improve network operation and performance in different failure scenarios. Techniques for the intra-domain link failures were based on IP fast reroute, shortest-path fast reroute and overlay-based reroute. The mechanism for the inter-domain fast reroute is rather simple by just advertising alternative paths, due to the challenges for ASes to cooperate or have control and/or visibility to each other. Fast recovery mechanisms provided by programmable networks are perhaps the most powerful ones because of their flexibilities.

The BGP-M technique studied in this thesis enables the routing between two ASes to be resilient to link failures. When one border link fails, the other border links can still be available for traffic delivery.

2.2 Multipath Routing

Network operators must control the distribution of traffic crossing their networks, in order to provide highly available and efficient services. Traffic engineering (TE) is referred to as the set of techniques and tools that operators utilise to this end [59].

Various techniques exist to achieve traffic engineering. Changing link weights and using loop-free next hops as static routes for traffic can help to achieve intra-domain TE [189]. Techniques to achieve inter-domain TE include selective advertisement, increasing AS path length, changing MED attribute [153], assigning different `LocPref` values to different outgoing links and advertising routes with BGP Communities [153, 160].

Multipath routing can also be used for traffic engineering and it is specifically used to balance the traffic between the same source and destination on multiple routes. These multiple routes are legitimate and lasting routes, and they are able to improve the routing performance, increase the throughput and the resource utilisation efficiency and decrease the latency.

Because of the emergence of multipath routing, researchers started to realise that "the traditional concept of a single network path between hosts no longer holds" [46,48,171]. For example, when multiple paths were observed, they were often considered as anomalies, possibly due to routing table misconfiguration [108]), link failures [64, 72, 150] or change of routing paths [39, 66, 156, 177, 179]. But as reported by Cunha *et al.* [67], the observed routing dynamics can be caused by the usage of multipath routing.

In general, the implementation of multipath routing requires a router to support ECMP algorithm [102]. The multiple paths used by this algorithm have equal cost for routing traffic to the destination. The cost can be in any form, depending on the routing protocol. The algorithm uses hash-threshold method to assign the equal-cost paths to the traffic flows.

Figure 2.2 illustrates multipath routing with an example. In this example, multiple routing paths are used for delivering the traffic between each source-destination pair of IP addresses, i.e. the pairs of (Source IP-1, Destination IP-1) and (Source

IP-2, Destination IP-2). The routing paths between Source IP-1 and Destination IP-1 form an intra-domain diamond within the Source AS; the routing paths between Source IP-2 and Destination IP-2 form an inter-domain diamond across the boundaries of the two neighbouring ASes relying on the two border links (Border link-2 and Border link-3).

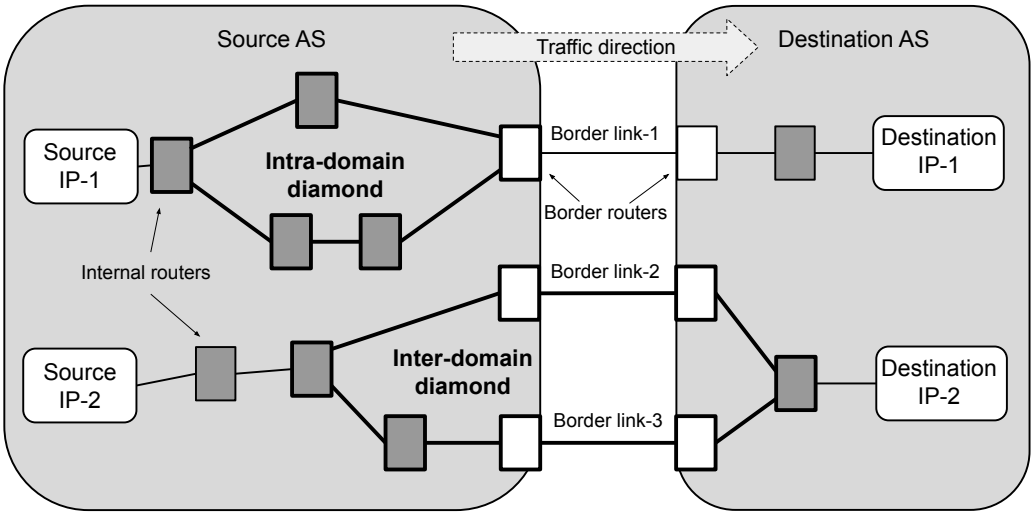


Figure 2.2: Illustrative example of multipath routing, where multiple routing paths are used between the same source and destination IPs – the paths may diverge and merge within the same AS forming an intra-domain 'diamond', or cross AS borders forming an inter-domain 'diamond'.

2.2.1 Measurement Efforts

In the area of measuring multipath routing, a line of well-known researches were conducted based on Paris traceroute [45], the concept of "diamond" shape load balancers [45] and Multipath Detection Algorithm (MDA) [46, 48].

2.2.1.1 Paris Traceroute and Multipath Detection Algorithm (MDA)

In 2006, Augustin *et al.* [45] proposed Paris traceroute to solve the problems caused by conventional traceroute. The problems include missing links or mis-identifying routing paths due to the existence of load balanced paths. Technically, Paris traceroute [45] maintains a constant flow identifier for probes it sends to a destination and is able to discover load balanced routing paths. Thereafter, Paris traceroute has been adopted widely as an improved variation of traceroute in the discovery of new links

and nodes, the detection of multipath routing and the topological characterisation of diamonds and load balancers.

Augustin *et al.* [46,48,171] proposed MDA and used Paris traceroute to detect load balanced paths. MDA [48] adjusted the number of probes to send to discover as many next hops as possible for each hop. They [46,48,171] also characterised the load balanced paths from traceroute measurement data. Their results showed that "the traditional concept of a single network path between hosts no longer holds". Veitch *et al.* [171] discussed how MDA used failure control to provide reliable discovery of multipath routes.

In 2018, Vermeulen *et al.* [174] proposed MDA-Lite, a lite-version of MDA with low failure probability based on an update to Paris traceroute. They [174] extended the tracing to multilevel multipath route tracing for a router-level view of multipath routes, and revealed that load balancing topologies had increased in size since 2016.

2.2.1.2 Measurements Based on Paris Traceroute and MDA

In the past decade, Paris traceroute with MDA has been used to measure the load balanced paths [41,42,47,173] on IPv4 and IPv6 Internet.

Augustin *et al.* [45] carried out measurement from one single source to 5,000 destination IP addresses. Later, Augustin *et al.* [46] conducted wider range of traceroute measurement from 15 source IPs to IPs in two destination lists. One destination list had 68,629 addresses and the other list had 500 addresses. In 2009, Augustin *et al.* [47] extended the measurement presented in 2007 [46]. Thus, the measurements introduced in [47] were from source IPs in two platforms to four lists of destination IPs to measure the multipath routing in the Internet.

Almeida *et al.* [42] proposed to use an IPv6 variation of Paris traceroute to measure and characterise load balancing on IPv6 Internet. They carried out measurement from 12 source IPs to around 52k IPv6 destination IP addresses. Recently, Almeida *et al.* [41] presented Multipath Classification Algorithm (MCA) to identify and classify load balancers on both IPv4 and IPv6 Internet. MCA used a different technique from MDA to determine the probes to send for each hop. MCA tried to

identify the set of bits in packet headers that a load balancer (i.e. a router) used for load balancing, and to identify the type of load balancing a load balancer performed. The measurement in [41] was carried out from 31 source IPs to over 19k IPv4 and over 16k IPv6 IP addresses, and different protocols (i.e. UDP, TCP and ICMP) were examined.

Vermeulen *et al.* [173] proposed Diamond-Miner (D-Miner), combining MDA and Yarrp (a high-speed randomized probing technique) [51] to make the scalable Internet-wide discovery of load balancing feasible. They [173] carried out traceroute measurement from 7 source IPs to over 14.4 million /24 prefixes.

2.2.1.3 Other Measurement Works

Aside from these works on actively detecting multipath routing based on Paris traceroute and MDA, other researches rely on existing traceroute data to study the multipath routing characteristics. Zhang and Perrig [191] utilised path availability history to reveal failure correlations, and proposed a path metric and selection scheme that is resilient to failure correlations to achieve multipath routing. Mok *et al.* [142] studied the load balancing behaviour on inter-domain links by YouTube with traceroute data. Iodice *et al.* [105] studied the periodical path changes by analysing the RIPE Atlas anchoring measurement data from 9,738 probes towards 258 anchors, which produced 101,715 active probe-anchor pairs.

2.2.2 Load Balancer and 'Diamond'

Section 2.2.1 has reviewed a number of researches and measurements on multipath routing in the Internet. This subsection summarises the results presented in these researches and shows the current status of deployment of multipath routing in the Internet.

Augustin *et al.* [45] observed 16,385 diamonds in the traces to 79% of the 5,000 destination IP addresses. 64% of the discovered diamonds were due to per-flow load balancing with most of the remainder explained by per-packet load balancing. Then, they reported in [48] that the routes to 1,525 (around 30%) of the 5,000 destinations (from one single source) in their measurements were affected by per-

flow load balancing, whereas per-packet load balancing affected less than 2% of the routes.

Later, Augustin *et al.* [46] reported that the paths between 70% of the 771,795 source-destination pairs traversed a per-destination load balancer. This percentage was 39% for per-flow load balancers and only 1.9% for per-packet load balancers. As reported by Augustin *et al.* [47], the results in 2007 remained the same as introduced in [46], and the measurement results in 2009 showed that 50% of the source–destination pairs traversed a per-flow load balancer (83% for per-destination and 1% for per-packet). Although no specific number of source-destination pairs was provided for the results in 2009, it was shown that per-flow and per-destination load balancing were still widely deployed in the Internet, and the application of per-packet load balancing had decreased.

Almeida *et al.* [42] reported that 74% of IPv6 routes traversed at least one load balancer. Per-destination load balancing was the most common on IPv6, and per-packet load balancing was more common on IPv6 than on IPv4. Moreover, 4% of IPv6 routers considered the `Traffic Class` and `Flow Label` header field for load balancing. Almeida *et al.* [41] further reported that on both IPv4 and IPv6, per-flow load balancers were observed more than per-destination load balancers with UDP and TCP, whereas per-flow load balancers were observed much less than per-destination load balancers with ICMP. Other load balancers like per-packet and per-application were rarely observed with all the protocols on both IPv4 and IPv6. Note that Almeida *et al.* [41,42] did not provide any specific number of load balancers in their measurements.

Vermeulen *et al.* [173] extracted 4,029,866 unique diamonds from the traceroute data. They reported that 64.7% of their traces towards all of the /24 prefixes contained at least one branching point, and 1.9% of branching points were per-packet load balancers.

Iodice *et al.* [105] reported that 36% of probe-anchor pairs experienced at least one periodicity, and a total amount of 186,403 periodicities were observed.

2.2.3 Summary

In summary, the existing works have advanced the researches on multipath routing in several ways. First, all of them have reported the wide prevalence of multipath routing because as many as over 4 million diamonds were observed from traceroute data. Second, they specifically demonstrated the wide usage of per-flow and per-destination load balancing in the Internet, as well as the rare usage of per-packet and per-application load balancing. Third, these papers show a general picture about the development of multipath routing in the past decade. Fourth, these researches and their results about the wide deployment of multipath routing suggest that multipath routing is an efficient and important technique for load balancing.

Most of the existing research works focused on multipath routing deployed at an internal router inside an AS. Some of them, such as Augustin *et al.* [47] and Almeida *et al.* [42], mentioned that a small part of the observed multipath routing might be related to BGP-Multipath (BGP-M). These works, however, provided no further detail. For example, Augustin *et al.* [47] suggested that most 'diamonds' are converged within a single AS and 'very few core networks enable BGP multipath capabilities in their routers'.

My research steps further based on these researches and focuses on BGP-M. Next section will introduce the details about BGP-M, including the definitions and notations, its difference from the above-mentioned multipath routing techniques, and the related works in the literature. My results in Chapters 4-6 will show how to measure the deployment of BGP-M, how an AS has deployed BGP-M and how BGP-M performs.

2.3 BGP-Multipath (BGP-M)

In a normal routing, when a border router has learned multiple routing paths to the same destination prefix, and it will select a single best path for the traffic delivery according to the BGP best path selection algorithm (see Table 2.1).

Because single-best path is prone to problems like link failure and network attacks, BGP-M has been supported by major router vendors on their border routers.

BGP-M is based on Equal-Cost Multi-Path (ECMP) function. ECMP allows a router to install multiple routing paths with equal cost for the traffic delivery to a same destination IP. The cost can be calculated with metrics like path length, link weight, etc. This helps to achieve higher link utilisation ratio and reduce the risk of congestion.

The support of ECMP allows a border router to route packets to the same destination prefix along multiple paths of equal cost [102]. When multiple 'equally-good' paths to the same destination prefix are learned from the *same* neighbour AS, instead of applying last-resort tie-breaker, an AS can use BGP-M to install more than one active paths to a corresponding destination. By 'equally-good', it means these paths have exactly the same values for the first six attributes in Table 2.1, i.e. `LocPref`, AS path, Origin, MED, eBGP/iBGP and IGP metric. Thus, these paths have equal cost to the destination prefix.

2.3.1 Definitions and Notations

Figure 2.3 illustrates the definitions and notations involved in my study of BGP-M. In the figure, the nearside AS (*NearAS*) is connected to the farside AS (*FarAS*) at the nearside border router (*NearBR*) via two border links, i.e., BL-1 and BL-2. In order for the *NearAS* to deploy BGP-M at the *NearBR*), the following conditions must be satisfied. (1) *NearBR* supports the ECMP function; (2) *NearBR* has multiple border links connecting to a same neighbour AS; (3) *NearBR* has learned from the same neighbour AS multiple routes via different border links, to a given destination prefix (*DstPrfx*); and (4) the multiple routes have equal values for the first six attributes in Table 2.1.

The conditions to deploy BGP-M are restrictive and the violation of these conditions will not be considered as BGP-M. For example, if the multiple paths to the same *DstPrfx* are learned from different neighbour ASes, these paths will have different values for AS path, not fulfilling the fourth condition.

If the above conditions are met (i.e. the routes learned over different paths are considered sufficiently equal), *NearAS* can deploy BGP-M at *NearBR* by installing the multiple routes in the routing table such that *NearBR* is configured to use these

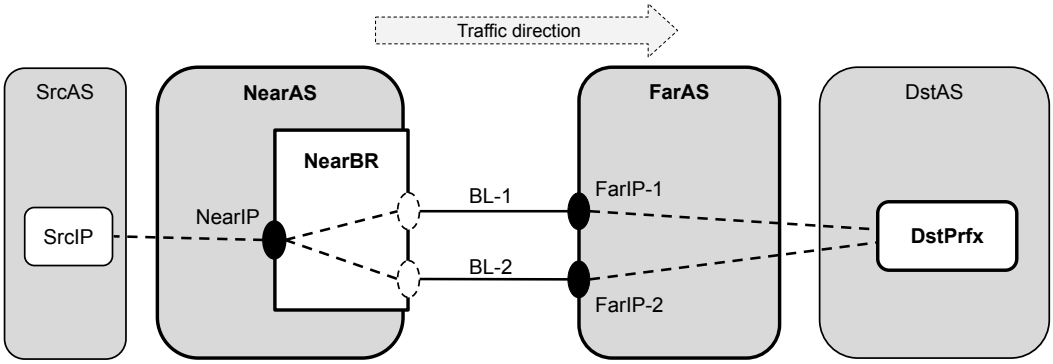


Figure 2.3: Illustrative example of BGP-Multipath (BGP-M), where the nearside border router (*NearBR*) uses multiple border links (BL-1 and BL-2) to share traffic flows to different IP addresses in the destination prefix (*DstPrfx*).

paths concurrently. Because all the relevant BGP attributes for the routes over different paths are the same, and *NearBR* still announces one route as the best route, there is no impact to BGP loop detection or other BGP processing [170].

By deploying BGP-M, a *NearBR* achieves the following two at the same time: (1) multiple paths to the same *DstPrfx*; and notably (2) a single, permanent AS-level path to each IP address in the *DstPrfx*. Thus, BGP-M is different from the terms of 'Multi-path BGP' in [77, 170] and 'Multipath BGP' in [50] that use multiple AS-level paths for inter-domain routing.

2.3.2 Difference from Other Multipath Routing Techniques

As a multipath routing technique, BGP-M can achieve load balancing by using different routing paths for the same destination. At the same time, BGP-M is profoundly different from other multipath routing techniques that are deployed at internal routers inside an AS.

Firstly, multipath routing techniques, e.g. load balancers and 'diamonds' observed in [41, 42, 47, 173], are implemented at internal routers. The multipath routing techniques are able to reveal the load balancing schemes used by the internal routers, instead of the specific configuration that is invisible to the public.

BGP-M is implemented at border routers between ASes. The information obtained from border routers reveal exactly how they are configured to achieve BGP-M load balancing.

Secondly, among the multipath routing techniques, the intra-domain diamonds merge within an AS and only use intra-domain links, and the inter-domain diamonds use inter-domain border links crossing AS boundaries. The links (both intra-domain links and inter-domain border links) used in these multipath routing techniques can have huge difference in geographical locations and bandwidths.

BGP-M uses border links to balance traffic load. These border links start from the same *NearBR* and are connected to one or multiple *FarBRs* in the *FarAS*. These border links used for BGP-M normally have high bandwidths to handle the high volume of inter-domain traffic.

Thirdly, the multipath routing techniques are implemented based on IGPs, e.g. OSPF. Each IGP has its own metrics and algorithm to rank the paths to each destination. Each AS has its distinctive configurations of IGP. The configurations are not visible to the public and difficult to be observed from traceroute data.

BGP-M is deployed based on the BGP mechanism. BGP is a de-facto EGP applied by ASes. Each AS uses the rules in Table 2.1 to rank the paths to each destination. An AS' BGP configuration on border router is often visible to the public and can be obtained from publicly available BGP dumps and queries to LG servers.

2.3.3 Related Works in the Literature

So far I only find a few technical documents and research papers in the literature that are related to BGP-M.

2.3.3.1 RFC2992: Analysis of an Equal-Cost Multi-Path Algorithm

The Request for Comments of RFC2992 [102] gives an analysis of one method for routers to decide which next-hop (path) to use in the sense of ECMP. The method is called hash-threshold. The router performs a hash over the packet header fields to select a key and then assigns the next-hops to unique regions in the key space. This document is focused on the method in terms of performance analysis, disruption and comparison to other methods. Thus it provides guidance to network operators to implement ECMP. The introduced method can be used for the deployment of

BGP-M. This RFC, however, did not provide further description on how to apply the method to BGP-M.

2.3.3.2 IETF Draft: Equal-Cost Multipath Considerations for BGP

This Internet Draft [118] by the Network Working Group of the Internet Engineering Task Force (IETF) is perhaps the most relevant working document to BGP-M. This draft describes the application of ECMP to BGP in different occasions. It introduces how to apply ECMP to multipath among eBGP-learned paths, among iBGP-learned paths and among eBGP and iBGP paths. Note that this Internet draft has already expired.

2.3.3.3 Router Vendor Documentations

BGP-M is today supported by most major router vendors, including Cisco [8], Juniper [18], and Huawei [13]. Specifically, both Juniper and Cisco routers support BGP-M in the term of BGP multipath [8, 18], Huawei routers support BGP-M in the form of BGP load balancing [13].

These router vendors have provided specific descriptions on how to configure their routers to achieve BGP-M. On Cisco routers, network operators can use the command of `maximum-paths` to configure the maximum number of equally-good paths used for BGP-M load balancing [8]. On Juniper routers, network operators can use the command of `set multipath` to activate BGP-M load balancing [18]. On Huawei routers, network operators can use the command `maximum load-balancing` to configure BGP-M [13].

The router vendors have defined the BGP-M (i.e., BGP multipath by Cisco and Juniper [8, 18], and BGP load balancing by Huawei [13]) and the conditions to deploy BGP-M. My research is based on the descriptions provided by these documentations.

2.3.3.4 Brief Discussions in Research Papers

Valera *et al.* [170] has a two-paragraph discussion on a scenario where a border router can potentially use multiple paths concurrently if they are 'sufficiently equal'. Apart from a brief discussion of such a possibility, there is no mention of any data

or observation in real network.

Augustin *et al.* [47] and Almeida *et al.* [42] have mentioned that a small portion of multipath routing can be related to or caused by the existence of BGP-M. However, they did not conduct any study to verify such hypothesis.

Mok *et al.* [142] studied YouTube's load balancing via border links connecting between different ASes. This work, however, is irrelevant to my research because the YouTube's load balancing does not use BGP mechanism.

2.4 Summary

This chapter has introduced the background of Internet routing, reviewed the researches on multipath routing, and presented the basics about BGP-M. First of all, BGP-M can be used to address the problems like routing congestion, network attacks, and link failures. Secondly, BGP-M has not been studied in the literature. The existing researches have extensively studied and measured the multipath routing in the Internet. Some of them [42,47] have mentioned the usage of BGP-M but did not explore further on BGP-M due to the difficulty on AS border mapping and the lack of suitable datasets. Aside from these researche papers, there are only the router vendor documentations on BGP-M. A thorough research on BGP-M will fill this gap and provide new knowledge on inter-domain routing and multipath routing.

Chapter 3

Internet Measurement Methods and Public Data Sources

Chapter 2 has introduced the background information on Internet routing, highlighted the importance of multipath routing, and discussed the little mention of BGP-M in the literature. This chapter will introduce the existing measurement methods and datasets in the Internet from the aspects of basics, typical projects and data sources, related researches. and how I use these datasets for my research. I will also discuss the existing problems and the state-of-the-art efforts.

3.1 Passive Measurement

Passive measurements allow researchers to study the Internet routing with BGP dumps, i.e. BGP routing tables and BGP updates. BGP dumps are continuously collected by BGP monitors from neighbour ASes, especially backbone ASes.

3.1.1 BGP Tables and Updates

BGP tables and updates provide BGP data from different perspectives. Specifically, BGP table provides records for each prefix the AS path, `LocPref`, MED, origin, etc. A BGP speaker makes routing decisions based on the information in its routing table, the BGP best path selection process.

When the routing information to a prefix changes, a BGP speaker sends a BGP update about this prefix to its neighbour BGP speaker, instead of sending the whole routing table. Thus, BGP updates provide the updated routing information for each

prefix. A piece of BGP update contains prefix, AS path, BGP Communities, etc. Upon receiving BGP update for a prefix, a BGP speaker updates its routing table.

The information contained in BGP dumps can be used for researches on various topics. The AS_PATH attribute can help generate as complete AS-level Internet topology as possible [51, 69, 114, 148, 162]; BGP Communities attributes [90] are decoded with information like AS relationship, routing policy and geolocations; the prefix and AS_PATH provide IP-to-AS mapping information [187].

So far, BGP data can be obtained from two major sources, i.e. RouteViews [34] and RIPE Routing Information Service (RIS) [30]. Besides, Orsini *et al.* [149] have introduced BGPStream, a software framework, to process large amounts of distributed and/or live BGP measurement data. The following subsections will focus on introducing RouteViews and RIPE RIS with details.

3.1.2 RouteViews

RouteViews is a well known project and source for archives of real-time BGP dumps about global routing system. RouteViews currently has 34 collectors distributed around the world (17 in USA, 4 in Brazil, 2 in Australia, 2 in South Africa, 1 in Chile, 1 in Ghana, 1 in Japan, 1 in Kenya, 1 in Philippines, 1 in Serbia, 1 in Singapore, 1 in Sweden and 1 in UK). It also provides the collected datasets in different formats and various tools to analyse the datasets.

Datasets provided by RouteViews have been applied to researches on a wide range of topics, including BGP Communities [90], AS-level topology discovery [51, 69, 114, 148, 162], AS relationship inference [86, 89, 95, 110, 111, 129], Internet routing [39, 108], IP-to-AS mapping [187], AS border mapping [136, 137], IP geolocation [117, 163, 180] and IXP peering links discovery [87, 145].

For my study on BGP-M, I extracted prefix-to-AS information from BGP updates published by RouteViews for different purposes. I used the AS_PATH extracted in data snapshot on 1/January/2020 to obtain neighbour ASes for each near-side AS. I also used the prefix-to-AS information to obtain the prefixes announced by neighbour ASes. The neighbour ASes and prefixes were used for identifying the deployment of BGP-M. Details about the data are provided in Chapter 4.

3.1.3 RIPE Routing Information Service (RIS)

RIPE RIS employs a globally distributed set of Remote Route Collectors, typically located at Internet Exchange Points (IXPs), to collect and store Internet routing data. Currently, RIPE RIS maintains 24 collectors (4 in USA, 3 in Switzerland, 2 in France, 2 in the Netherlands, 1 in Austria, 1 in Brazil, 1 in Germany, 1 in Italy, 1 in Japan, 1 in Romania, 1 in Russia, 1 in Singapore, 1 in South Africa, 1 in Spain, 1 in Sweden, 1 in UK and 1 in Uruguay).

The data provided by RIPE RIS has been widely used in researches like the impact of IXP in Internet routing [93], inter-domain routing [39, 43, 64, 72], AS path inference [169], AS relationship inference [84, 111, 129], AS boundary mapping [128, 136, 137], the deployment of IPv6 [109] and IP geolocation [70, 180]. Note that the datasets from RIPE RIS are not applied in my work yet but they are definitely valuable for improving the discovery of the deployment of BGP-M.

3.2 Active Measurement

3.2.1 Traceroute Probing

Active measurements rely on traceroute probing to measure the Internet for different research purposes. A source host sends data packets to a destination host with different values of Time-to-Live (TTL), starting from 1. When an intermediate router receives the data packet, it checks the value of TTL. If TTL == 1, the intermediate router sends back an ICMP Time Exceeded Message to the source with an IP address (normally the IP address of the interface receiving the probing) as the source of the ICMP message; otherwise, the value of TTL decreases by 1 and the intermediate router sends the packet to the next-hop router. The source host records the source IP in each ICMP Exceeded Message as the IP address at each hop. These recorded IP addresses form the traceroute path.

Traceroute data is currently the major source for researches on data-plane Internet routing, and there have been several well-known lines of researches. The first line is Internet topology discovery on different levels of granularities (e.g., [71, 114]). A second line of researches is IP-to-AS mapping, ranging from the

works in [132,133] to the currently available datasets provided by IP2Location [16], Team Cymru [32] and MaxMind [21]. A third line is on AS border mapping (e.g. [104, 184]) involving alias resolution (e.g. [97, 172]). The fourth line is the detection of load balancing behaviour, including the works on Paris traceroute [45], MDA [46,48], measurement on multipath routing [47], MDA-Lite [174], Diamond-Miner [173] and MCA [41].

Traceroute has its own drawbacks [135]. For example, routers can be configured with no response to traceroute probe for security concerns, and an unresponsive hop is represented by * in the traceroute path. Besides, a router may respond to a traceroute probe with a third party IP address. Thus, studying the problems related to traceroute or the shortcomings of traceroute is a fifth line of researches. For example, He *et al.* [99] deployed traceroute measurements to study the routing asymmetry for two networks on both AS- and router-level. Marchetta *et al.* [134] proposed an active probing technique based on the IP timestamp option to detect third-party IP addresses in traceroute paths. The problem of third-party IP address was then further studied by Luckie and claffy [126]. Giotsas *et al.* [83] presented techniques to identify out-of-date traceroutes without issuing any measurements.

Some of the above researches relied on self-deployed measurements (e.g. [45]), while some relied on existing traceroute data (e.g. [71, 114]).

Currently, traceroute data can be obtained from a number of existing projects and datasets. RIPE Atlas [157] and PlanetLab [27] are two widely used platforms that provide nodes/probes to users to conduct traceroute measurements. The difference between datasets provided by RIPE Atlas and PlanetLab was analysed in detail in [58]. Projects like iPlane [131], CAIDA Ark [5] and DIMES [161] have made their data publicly available. Besides, Looking Glass (LG) servers also provide command of traceroute for network diagnose purpose. Moreover, Paris traceroute [45, 48, 174], an enhanced version of traceroute, has been designed and embedded with the ability to discover load-balanced routing paths. The following subsections focus on introducing RIPE Atlas, iPlane, CAIDA Ark and DIMES and how I use the data provided by these projects. PlanetLab is not used for my study

because UCL is not validated for registration, making PlanetLab less convenient than the other platforms. LG server data will be introduced in Section 3.3.2.

3.2.2 RIPE Atlas

RIPE Atlas [157] is a platform developed by RIPE NCC (Réseaux IP Européens Network Coordination Centre), and users can carry out various kinds of measurement for researches. RIPE NCC staff [157] have introduced RIPE Atlas from various perspectives, including history and funding, use cases, overall design, data storage and measurements. This subsection provides basics about the traceroute measurement on RIPE Atlas.

RIPE Atlas provides traceroute data produced from user-defined measurement and anchoring measurement. User-defined measurement are defined by users registered on RIPE Atlas. For user-defined traceroute measurement, the sources are either RIPE Atlas probes or RIPE Atlas anchors. Probes are physical devices hosted by ASes, and can be obtained by either sponsoring RIPE Atlas or submitting applications. Anchors are enhanced probes with more measurement capacity. Any IP address can be the destination of a measurement.

As in June 2021, there are over 10,000 RIPE Atlas probes being actively connected to the Internet, covering 3,722 IPv4 ASNs and 1,656 IPv6 ASNs. There are 816 RIPE Atlas anchors hosted by 80 organisations.

Users can create measurements by either using the web interface provided by RIPE Atlas, shown in Figure 3.1, or using a tool that is able to send POST requests with a payload to RIPE Atlas. Users can set the properties for the measurement as needed. User-defined measurement data are free to download as long as the measurement owner set the data publicly available. The anchoring measurements are produced by RIPE Atlas. The ongoing measurements are performed by hundreds of anchors from the RIPE Atlas network. The data from anchoring measurements are free to download.

There have been a number of research works studying the measurement deployed via RIPE Atlas from different perspectives, including the interference between measurements [101], lessons learned from using RIPE Atlas [49], study on

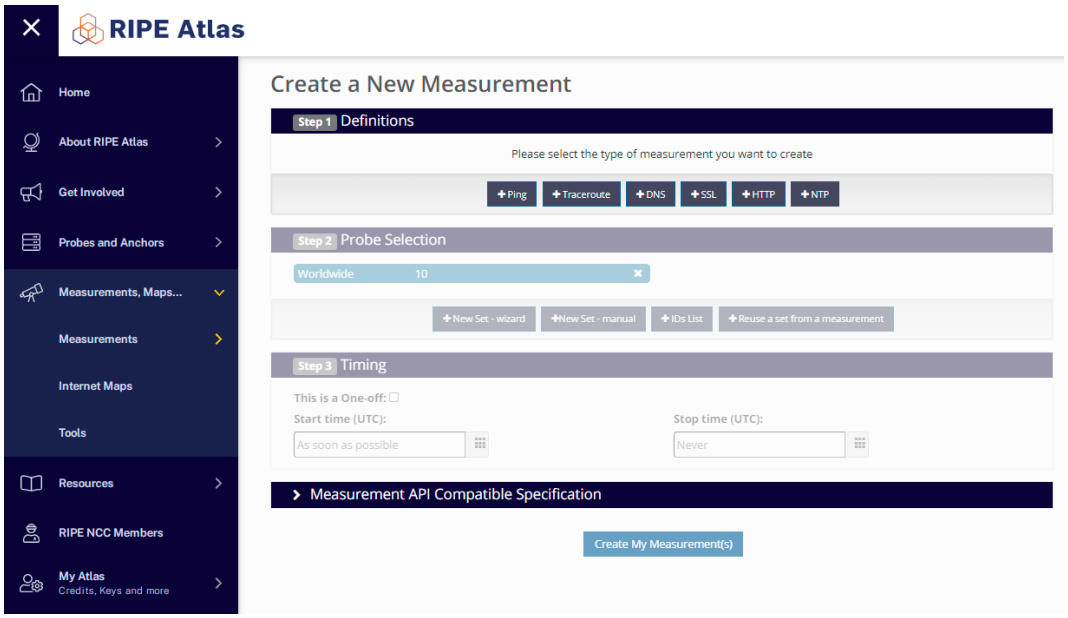


Figure 3.1: Web interface provided by RIPE Atlas to create a measurement.

the country-level interconnections between eyeball ASes [81], reverse path analysis [178], the periodicity of Internet routing path changes [105] and comparison between data sources [58].

I have analysed anchoring measurement data on RIPE Atlas during my PhD. The collected anchoring measurement data was from 1 December 2018 to 31 May 2019, containing traceroute paths from 10,623 probes to 366 anchors around the world (with in total 16,323 source-target pairs). Each source-destination pair was probed every 15 minutes, with the default settings like ICMP-based [130] message and the variation 16 of Paris-traceroute [45]. The data was used for analysis on the usage of inter-domain border links, and is not introduced in the following chapters.

My research during PhD also relies a lot on the traceroute data obtained via user-defined measurements on RIPE Atlas. I applied for a RIPE Atlas probe and embedded it into UCL's network system without compromising UCL's routine network operation. Hosting a probe allowed me to gain credits and conduct user-defined measurements. With the user-defined measurements, I understood the complexity of inter-domain routing in terms of the usage of border links. More importantly, I sent tracreoute probings to specific destination IPs from the probes hosted by several ASes, and studied how border links were allocated and how border links performed

in BGP-M cases. More details about the traceroute measurements and the results about BGP-M will be introduced in Chapter 6.

3.2.3 iPlane

iPlane [131] is a scalable service providing accurate predictions of Internet path performance in terms of latency, bandwidth, capacity and loss rates. It builds a structural model of the Internet and clusters the interfaces belonging to the same Point-of-Presence (PoP) and interfaces within geographically nearby portions of the same AS.

iPlane has published the datasets collected between 2006 and 2016. The published datasets by iPlane have been applied for researches on path performance analysis [178], topology discovery [113, 114, 164, 190], analysis of Internet RTT [117], IP geolocation [65, 163], alias resolution [92, 97], comparative analysis among topology datasets [58], path latency through IXPs [38], etc.

During my PhD research, I analysed the traceroute data provided by iPlane to study the PoP-level Internet routing and learned that inter-domain routing is very complicated because two ASes can be connected with each other at multiple PoPs.

3.2.4 CAIDA Archipelago (Ark)

CAIDA Archipelago (Ark) [5] is a globally distributed measurement platform deployed and maintained by CAIDA since 2007. CAIDA Ark distributes hardware measurement nodes to improve the public view of the global Internet. The goals of the Ark infrastructure are to: reduce the effort needed to develop and deploy sophisticated large-scale measurements, and provide a step toward a community-oriented measurement infrastructure on a security-hardened distributed platform. So far, Ark has more than 200 monitors, globally distributed among business, commercial, educational, research, infrastructure, and residential networks.

Currently, traceroute measurements on Ark are conducted using Scamper [125], an open-source packet prober for active measurements. Traceroute measurements based on Scamper and Ark project have been widely used in researches on Internet routing, covering topics like topology discovery [71, 126, 134, 190], AS

border mapping [128], topology discovery on IXP [145], IXPs' impact on Internet routing [37, 93], IPv6 Internet routing [42, 52, 109] and load balancing [42, 142].

I have run traceroute measurement with Scamper from May of 2019 to March 2021, from a UCL host to 920 RIPE Atlas probes (IPv4) belonging in the Top-50 ASes according to CAIDA AS rank data [4], repeatedly every 30 minutes. The measurement was conducted for the purpose of study on the usage of inter-domain border links in terms of multipath routing. However, because of the limited scale, I did not apply it to my study on multipath routing and BGP-M. Therefore, this work does not provide further details about this measurement. I expect to reveal interesting observations from the data in the future.

3.2.5 DIMES

DIMES [161] is a distributed measurement infrastructure for the Internet. DIMES relies on traceroute measurement from over 5,000 agents within more than 570 ASes to study the Internet topology from various levels of granularities, i.e., AS level, PoP level and router level. The project was launched in September 2004, and it stopped updating data in April 2012.

I have also analysed DIMES data. However, DIMES only provided topological data (i.e, links, nodes) instead of traceroute data, making it not suitable for researches on multipath routing and BGP-M. Moreover, the website (`www.netdimes.org`) stopped updating datasets in April 2012, and has been unavailable to visit for a while (still not available as of this writing).

3.3 Other Data Sources

3.3.1 CAIDA

CAIDA has published a variety of datasets on Internet routing to the public. Aside from the traceroute data produced by Ark, Internet Toplogy Data Kit (ITDK) [20], AS Relationship data [3], AS rank data [4], AS-to-Organisation data [6] and IXP data [7] are also popular datasets. They have been widely used in researches like AS relationship inference [84, 111, 129], AS border mapping [136, 137, 184], load balancing behaviour [142], and IP geolocation [79]. These datasets are still being up-

dated as indicated by the "Ongoing" status on the website of CAIDA Datasets [10].

The CAIDA datasets used during my PhD research are listed below with how I used them.

- CAIDA Ark: I have run traceroute measurement with CAIDA Ark as introduced in Section 3.2.4.
- AS relationship data: used for my study on the usage of border links.
- AS-to-organisation data: used for my study on the usage of border links.
- ITDK: the interface-to-router data provided in ITDK was used for my study on the usage of border links.
- Customer cone data: used for my study on the usage of border links, and my study on BGP-M (see details in Chapter 5).
- AS rank data: used for my study on BGP-M, see details in Chapters 4-5.

3.3.2 Looking Glass (LG) Servers

Looking Glass (LG) servers offer an option to collect data for both active measurement and passive measurement. Many network operators host LG servers, and an LG server can provide Web-based interfaces to allow non-privileged execution of network commands at one or more border routers for network measurement and diagnosis [114]. The commands include `ping`, `traceroute`, `nslookup`, `whois`, `show ip bgp summary`, `show ip bgp (IPv4 address)` and some of them support IPv6 querying. An updated list of Looking Glasses is provided by `http://www.traceroute.org/`. Moreover, PeeringDB [25] and BGP Looking Glass Database [2] also provide information on Looking Glasses.

The data from LG servers are used in researches like discovery of IXP peering links [36, 55, 87, 88, 146], AS-level topology discovery [114, 148], AS border mapping [146] and IP-geolocation [123]. Moreover, the Periscope platform was proposed [82] to unify LG servers with publicly accessible querying API and to support on-demand measurements.

Aside from the value of Looking Glass, Bruno *et al.* [56] highlighted the drawbacks of LG data, including Reverse Cross-Channel Scripting and web flaws (e.g., exposed routers credentials). The authors also outlined a threat model, reviewed the Looking Glass software, and performed experiments to confirm their findings.

I relied on queries to LG servers to infer the deployment of BGP-M in the Internet. So far, I have discovered that BGP-M has been deployed by 12 ASes, on both IPv4 and IPv6 Internet. Chapters 4-5 provide detailed information on my measurement method using LG server data and the results.

3.3.3 Datesets on Internet eXchange Point (IXP)

Currently, there are three widely used sources for datasets on IXP, i.e., PeeringDB [25], Packet Clearing House (PCH) [24] and European Internet Exchange Association (Euro-IX) [11].

I firstly introduce the basics about these data sources and then review the researches based on these data sources. PeeringDB provides a freely available, user-maintained, database of networks and the go-to location for interconnection data. The database facilitates the global interconnection of networks at IXPs, data centers and other interconnection facilities. It is non-profit and promoted by volunteers. PCH helps build and support IXPs for more than 20 years. It maintains the global directory of IXPs, and publishes statistics about the IXPs' use and growth. Euro-IX [11] was formed to develop, strengthen and improve the IXP community. It now has 71 member IXPs and provides information about 681 IXPs, involving 16,798 ASNs.

To the best of my knowledge, the three sources are used as complementary to each other in most of the related researches. Klöti *et al.* [115] did cross-comparison of the IXP datasets from PeeringDB [25], Euro-IX [11] and PCH [24]. The analysis covered aspects of linking IXPs, geographical distribution, facilities, IXP status, IXP participants and the completeness of IXP participant data. Therefore, the research in [115] provided guidance for choosing IXP datasets. Gregori [93] collected a list of contacts from these sources for the study on the impact of IXPs. Nomikos and Dimitropoulos [145] used the IXP data from PeeringDB and PCH

to help `traIXroute` to detect IXPs in end-to-end traceroute paths. Nomikos *et al.* [146] examined the IXP data from these three sources and other sources for the study of remote peering.

Sometimes only the data from PeeringDB is used. For example, Lodhi *et al.* [124] studied the data provided by PeeringDB. With BGP data, the paper demonstrated that the PeeringDB membership is representative of Internet business types and geography of participants, and the data by PeeringDB is up-to-date.

In my work on BGP-M, I obtained [88] a list of IXPs and their prefixes based on the data provided by PeeringDB [25] in January 2020, to check whether a BGP-M deployment was deployed via IXP or not. I also referred to PeeringDB for bandwidth information of border links. Detailed results are introduced in Chapters 5 and 6. Datasets from PCH and Euro-IX will be applied as future work.

3.3.4 IP Geolocation Datasets

The geolocation datasets used in Internet measurement are mostly IP-to-geolocation data. Some researches (e.g. [79]) focus on router geolocation. In this subsection, I will use IP geolocation datasets to represent all the geolocation datasets for convenience.

IP-to-geolocation mapping provides the geolocations of IP addresses and helps us to know the geographical locations where the traffic enters or exits an AS. Several researches were proposed to map IP addresses to their geolocations. For example, Mátray *et al.* [139] proposed `Spotter` to geolocalise IP addresses, by measuring delays between IP addresses and vantage points and applying a probabilistic delay-distance model. Lee *et al.* [119] produced an IP geolocation DB with crowdsourcing Internet broadband performance measurements tagged with the location. They [119] also reported that the low accuracy of databases like MaxMind was caused by selecting a single representative location for a large IP block. Giotsas *et al.* [83] proposed a ping-based method to geolocate IP addresses. The method in [83] selected a vantage point that was mostly likely to have a direct path to a target IP, and pinged the target IP from the vantage point. The selection of vantage point in [83] involved DNS data, AS relationship data, PeeringDB, RIPE Atlas

probes and LG data.

Aside from the methods to actively map IP geolocations, some researches focus on evaluating the accuracy of the existing IP geolocation datasets. Poese *et al.* [151] compared several geolocation databases, highlighted their limitations, and showed the overly fined granularity could claim country-level accuracy, but not city-level. The analysis conducted by Gharaibeh *et al.* [79] showed the accuracy of several data sources for router geolocation, and found that the datasets were inaccurate at neither country-level nor city-level. Du *et al.* [70] introduced the single-radius engine of RIPE IPmap [29], and evaluated the method on IP-geolocation. The result suggested IPmap achieves around 80% city-level accuracy. Livadariu *et al.* [123] investigated the accuracy of several IP geolocation datasets (i.e. MaxMind, IP2Location, IPmap and HLOC [159]) with end-to-end traceroute paths.

Despite the reported low-accuracy on city-level, IP geolocation datasets (e.g. MaxMind, IP2Location, NetAcuity [22], DNS, PeeringDB and RTT-based data) have been widely applied in the researches on Internet routing. The topics include inference of multilateral peering links [88] and complex AS relationship [84], remote peering [146], AS-level Internet map [94], AS border and co-location facility mapping [116, 143, 147], Internet routing [65, 72, 73] and prefix-level geolocation [180].

During my PhD research, I have used the IP geolocation data from sources like IP2Location, MaxMind and IPmap to study the inter-domain border links inferred from the RIPE Atlas anchoring measurement data. As for my study on BGP-Multipath, the geolocation information about each border router is directly provided by LG server data, with detailed information shown in Chapters 4-6.

3.3.5 Internet Routing Registry (IRR) Data

IRR includes some world-wide routing policy databases, which are individually operated by organisations using Routing Policy Specification Language (RPSL). IRR can be used for network troubleshooting, route filtering and validation and can be queried using the WHOIS protocol. Typical IRR databases are the RIPE and the APNIC (Asia Pacific Network Information Centre) WHOIS databases. IRR

data has been used in researches like Internet topology discovery (e.g. [113]), AS relationship inference (e.g. [85, 88, 129]) and IXP interconnectivity (e.g. [98]).

During my PhD research, I studied IRR data, gained knowledge about ASes' routing policies, and learned the complex connectivity between ASes. The IRR data is not used in my study on BGP-M, and is not introduced in the following chapters.

3.4 Discussion

3.4.1 Accuracy vs Completeness

As reviewed in previous sections, various datasets have been published for researches on Internet routing. However, it is challenging for a dataset to be both complete and accurate.

Firstly, as reported in some researches (e.g. [94, 96, 115]), the datasets on Internet routing share the same shortcoming of incompleteness because the vantage points are in limited number for the collection of data, and the scale of the Internet is immensely large. Researchers often need to combine datasets from various sources for a rather complete picture about the Internet.

Secondly, because the Internet is changing and updating rapidly, the datasets need to be updated frequently to keep fresh and accurate. However, it is difficult for researchers to always obtain the up-to-date datasets. Moreover, it is also difficult to validate a dataset with convincing and sufficient ground truth data, especially when the ground truth data should be obtained via requests or surveys to network operators. Therefore, to yield better accuracy, it is very common that multiple datasets are applied in a research for cross-validation [84, 111, 136, 180]. In this situation, the accuracy is often of top priority. Thus, the completeness is sacrificed because different datasets have varying coverage and only the data shared in common are accurate.

In my work, I rely on LG server data to infer the deployment of BGP-M. My method provides accurate results because LG servers provide direct access to border routers' routing tables including reliable information on the deployment of BGP-M. Because only a part of the prefixes announced by neighbour ASes are queried,

prefixes announced by remote ASes are not queried, and LG servers that support the commands used in my work are in limited number, my measurement is not complete yet. There are definitely many more BGP-M cases out there. They will be discovered as a part of future work.

3.4.2 IP-to-AS Mapping

IP-to-AS mapping has attracted the researchers' attention for more than a decade. It relies on prefix-to-AS information announced in BGP routing tables or updates. For example, Mao *et al.* [132, 133] proposed several methods for accurate IP-to-AS mapping with iterative matching. Later, Zhang *et al.* [187, 188] proposed to refine the IP-to-AS mapping with the IP address granularity and their method allowed IP addresses in the same prefix to be mapped to different ASes.

To some extent, the accuracy of IP-to-AS mapping depends on the $(IP\,prefix, its\,origin\,AS)$ pairs. Khan *et al.* [113] carried out comparative analysis between IRR data and BGP data on $(IP\,prefix, its\,origin\,AS)$ pairs. The results suggested the quality of IRR data depends on several factors.

Apart from these methods, some data sources provide datasets for direct reference. These data sources include IP2Location [16], Team Cymru [32] and MaxMind [21]. I have used the IP-to-AS mapping data from these datasets to identify the inter-domain border links from RIPE Atlas anchoring measurement data.

3.4.3 AS Border Mapping

As reported in the literature (e.g. [126, 134]), it is not accurate enough to simply rely on the existing methods and datasets on IP-to-AS mapping for AS border mapping. Therefore, a lot of methods have been proposed for higher accuracy.

Huffaker *et al.* [104] collected data from large-scale traceroute measurements, used alias resolution techniques, and developed a heuristics to assign routers to ASes, producing an AS-router dual graph. Giotsas *et al.* [85] introduced the Constrained Facility Search algorithm to map IP connectivity to PoPs. Nur and Tozal [147] presented the cross-AS topology maps and defined the cross-border interfaces to study relevant topological properties. Motamedi *et al.* [143] presented

the mi^2 (mapping Internet interconnections) algorithm and improved PoP mapping through more accurate identification of inter-domain borders.

In recent years a number of border mapping techniques have found that such border identification can lead to inaccurate mapping since ASes may number their interfaces of border routers with IPs of neighbour ASes [128, 136, 137]. Among these techniques, bdrmap [128] accurately inferred the boundaries between a given network and four networks with validation against ground-truth data. MAP-IT [137] made use of the existing traceroute traces and public BGP data to infer the AS boundaries between distant peering ASes. bdrmapIT [136] combined bdrmap and MAP-IT, achieved better coverage and accuracy than either of the previous methods. However, as stated in a recent work [184], bdrmapIT and MAP-IT are not able to map AS borders where layer-2 switching fabrics are employed at the network borders, and bdrmap can lead to inconsistent inference results.

During my PhD research, I have used bdrmapIT to locate the AS borders for the study on the usage of inter-domain border links. Due to the potential false positives from bdrmapIT, I relied on LG server data to obtain AS border information for the inference of the deployment of BGP-M in Chapters 4-5.

3.4.4 IP Alias Resolution

A number of researches have been conducted to improve the accuracy of IP alias resolution, and to increase the reliability of existing AS border mapping methods. Gunes and Sarac [97] firstly presented studies on the impact of alias resolution on topology measurement studies, and then introduced an alias resolution approach called analytic and probe-based alias resolver (APAR), which used common IP address assignment scheme to infer IP alias. Spinelli *et al.* [168] presented ALIAS-CLUSTER, a learning-based methodology to disambiguate router aliases using only observed traceroute measurements. Keys *et al.* [112] proposed Monotonic ID-Based Alias Resolution (MIDAR) to provide ID comparison test based on monotonicity. MIDAR integrated multiple probing methods, multiple vantage points and a sliding-window probe scheduling algorithm to increase scalability to millions of IP addresses. Grailet and Donnet [91] introduced a generic methodology to conduct

efficient and scalable alias resolution, which combined the space search reduction of TreeNET [92] with a fingerprinting process to assess the feasibility of several alias resolution methods, using a small, fixed amount of probes. Vermeulen *et al.* [172] presented an alias resolution tool called Limited Ltd. The method exploited ICMP rate limiting, extracted features from the probe reply loss traces, and used a machine learning classifier to designate pairs of interfaces as aliases. Marder [138] presented Alias Pruning by Path Length Estimation (APPLE) to filter potential router aliases seen in traceroute by comparing the reply path length from each address to a distributed set of vantage points.

During my PhD research, I have used the publicly available Vela MIDAR API provided by CAIDA [35], to identify the inter-domain border links from the RIPE Atlas anchoring measurement data. IP alias resolution is not involved in my research on BGP-M, thus not introduced in the following chapters.

As reviewed in previous subsections, researches have been extensively conducted to resolve challenges in regard of accuracy and completeness, IP-to-AS mapping, AS border mapping and IP alias resolution. Indeed, the proposed methods have improved a lot in performance, and provided the research community better understanding about the inter-domain routing. However, these methods are not sufficient enough. Moreover, challenges like Multiple Origin ASes (MOAS) [106,192] and third-party addresses [126,134] still exist. It requires more advanced techniques or variations of traceroute to resolve these problems and to benefit the study on inter-domain routing.

3.5 Datasets Used in This Research

In this research I will collect my own dataset, including query data from LG servers (see Chapters 4-5) and traceroute on RIPE Atlas (see Chapter 6). In addition, I will use other datasets that are publicly available online. All the datasets used in this research are listed below.

- BGP data and IP geolocation data from LG servers. I searched the existing LG servers in the Internet, queried them, obtained the BGP routing tables on

border routers to infer the deployment of BGP-M. I also extracted geolocation information for border routers with the data provided by LG servers.

- BGP data from RouteViews. I extracted prefix-to-AS and AS paths from the BGP update data provided by RouteViews to obtain the neighbour ASes of an AS and the prefixes announced by neighbour ASes, for infering the deployment of BGP-M.

- AS rank and customer cone size data from CAIDA. I used the AS rank and customer cone size data from CAIDA to study the relation between BGP-M deployment and ASes' sizes.

- IXP data from PeeringDB. I used the IXP data provided by PeeringDB to study the connectivity fabrics of BGP-M deployment.

- Traceroute measurement data on RIPE Atlas. I conducted traceroute measurement to study the routing properties of BGP-M.

Chapter 4

Measurement of the Deployment of BGP-M

This chapter firstly introduces BGP-M deployment and BGP-M case, which describe the deployment of BGP-M from different levels. I also present the challenges during the measurement. Then I propose the measurement method based on Looking Glass server data. This chapter ends with discussions on the LG-based method. The measurement results will be analysed in Chapter 5.

4.1 BGP-M Deployment and BGP-M Case

4.1.1 3-tuple of a BGP-M Deployment

I denote a *BGP-M deployment* as a 3-tuple.

$$< NearAS, NearBR, FarAS >$$

These three values are sufficient to uniquely describe which AS (*NearAS*) has deployed BGP-M, at which border router (*NearBR*), and to which neighbour AS (*FarAS*) the relevant border links are connected to.

When studying a BGP-M deployment, we only need to focus on traffic routing between the two neighbouring ASes, i.e. traffic exiting *NearAS* and entering *FarAS*. The source (*SrcIP*) of the traffic can be outside of *NearAS*, and the destination (*Dst-Prfx*) of the traffic can be outside of *FarAS* – indeed they can be anywhere on the

Internet as long as the traffic traverses through *NearAS* and *FarAS* via *NearBR*.

If an AS deploys BGP-M at different *NearBRs* to the same *FarAS*; or it deploys BGP-M at the same *NearBR* to different *FarASes*, these are considered as different BGP-M deployments as they have different 3-tuples.

The values of *SrcIP*, *NearIP*, *BLs* and *FarIPs* are not included in the tuple to avoid redundancy.

If *NearAS* and *FarAS* are connected via an IXP, the 3-tuple does not need to include the IXP because IXP is 'transparent' in BGP routing [107] and therefore the existence of IXP does not affect the function and the deployment of BGP-M.

4.1.2 4-tuple of a BGP-M Case

The definition of *BGP-M deployment* (with the same *NearAS*, *NearBR* and *FarAS*) describes the deployment of BGP-M with focus on the interconnections between the two neighbouring ASes. This section introduces *BGP-M case*, to describe the traffic flows that can benefit from the deployment of BGP-M.

A *BGP-M case* is defined as the usage of a given BGP-M deployment for routing toward a given prefix. For each BGP-M deployment, an AS can implement many different *BGP-M cases* for different destination prefixes (*DstPrfxes*).

A BGP-M case can be uniquely denoted as the following 4-tuple.

$$< NearAS, NearBR, FarAS, DstPrfx >$$

The tuple includes *NearAS*, *NearBR* and *FarAS* for a given BGP-M deployment; and the tuple also includes the value of *DstPrfx* to identify the exact destination prefix to which traffic routing can benefit from the BGP-M deployment.

4.2 Challenges in Measuring BGP-M

In the past, traceroute data were used to study multipath routing deployed at intra-domain routers [41, 42, 47, 173], where specific traceroute tools were designed and deployed and large amounts of data were collected. In theory, traceroute with UDP packets has the potential to discover BGP-M deployed at border routers, but there

are a number of challenges.

One challenge is that we will need to design a traceroute probe specially customised for discovering BGP-M. Then, without any prior knowledge, we will have to deploy the traceroute probe in as many ASes as possible; and from each probe, we will have to run traceroute to as many different destination prefixes in as many other ASes as possible.

The second and largest challenge is the lack of sound tools or datasets for AS border mapping – despite more than a decade of research effort. If we use traceroute data to discover BGP-M, we must be able to accurately determine the border of an AS on a traceroute path, so that we can credibly identify border router and border links. A recent study [184] shows that existing efforts on IP-to-AS mapping and AS border mapping [85, 104, 128, 136, 137, 143, 147] still cannot avoid erroneous results. For example, on IP-to-AS mapping, the same prefix can be announced by different ASes, making it difficult to map IP addresses and this can generate false positives on border link. Moreover, third party IP addresses can be used to respond to traceroute probings, causing wrong mappings of AS borders.

4.3 My Measurement Method Based on LG Data

Section 4.2 has discussed the challenges to discover BGP-M from traceroute data. To alleviate these challenges, I utilise data from LG servers to discover the deployment of BGP-M. An LG server provides Web-based interfaces to allow non-privileged execution of network commands at one or more border routers for network measurement and diagnosis [114]. These commands provide direct access to the BGP configuration and routing tables of border routers beyond what is propagated through BGP updates collected by RouteViews [34] and RIPE RIS [30]. The BGP configuration and routing table contains direct information on BGP-M deployment on a border router. Thus, LG servers are able to provide reliable information on BGP-M deployment.

Moreover, routing tables on border routers are configured directly to prefixes. Only one query is sufficient to learn whether BGP-M is deployed at the border router

to a neighbour AS for a destination prefix. Compared to traceroute measurement, queries to LG servers require less resources and is much simpler.

Different commands can be used to check whether BGP-M is deployed or not on a border router [8, 13, 18]. When BGP-M is deployed at Cisco routers, the command of `show ip bgp` will show and label the multiple routes for a destination prefix with `multipath` and `external`. When BGP-M is deployed at Juniper routers, the command of `show route <ip address> detail` will show and label the multiple routes for a destination prefix with `State: <Active Ext>` and `Accepted Multipath`. When BGP-M is deployed at Huawei routers, the command of `display bgp routing-table` will show and label the multiple paths for a destination prefix with `select` and `external`.

4.3.1 Searching for LG Servers

I have used three sources to obtain LG server information, which are BGP Looking Glass Database [2], PeeringDB [25, 26] and `http://www.traceroute.org/` [33]. I searched the information in these sources and compiled a list of ASes with LG servers. The list contains 2,709 ASes and Table 4.1 lists the information about these ASes, ranked by CAIDA [4]. Among these ASes, 1,434 ASes' LG servers were accessible, and 527 ASes had accessible LG servers and supported the `routes` command (e.g. `show ip bgp routes detail <IP address>`) which was needed for my method [121].

For clarity, in the following I call a border router to which I send LG query a 'nearside border router' (*NearBR*); and an AS that owns and manages the nearside border router a 'nearside AS' (*NearAS*).

Table 4.1: The 2,709 ASes with LG servers, ranked by CAIDA [4].

	Number of ASes in AS rank groups				
	Top 1–10	11–50	51–200	> 200	Total
With known LG URL	10	31	82	2,586	2,709
With accessible URL	8	20	62	1,344	1,434
Support `routes` command	4	11	36	476	527
With BGP-M deployment	1	3	0	8	12

4.3.2 Obtaining Lists of Neighbour ASes

For each *NearAS*, I obtained a list of neighbour ASes that were connected with each *NearBR* within this *NearAS*.

Among the 527 ASes with `routes` command available, 184 ASes provided `summary` command (e.g. `show ip bgp summary`). The `summary` command enables me to not only find the neighbour ASes, but also identify those neighbour ASes that are connected to the *NearBR* via multiple border links, which is required for BGP-M deployment as clarified in Section 2.3.1.

Figure 4.1 shows an example table returned by the command `show ip bgp summary` from `core1.tor1.he.net` (`tor1`), a border router of Hurricane Electric (HE, AS6939). The table lists the AS numbers (ASNs) of the BGP neighbours and the IP addresses of the interfaces through which the BGP session is established. Some neighbour ASes, such as AS19752, AS21834 and AS22616 highlighted in red boxes, are connected via multiple neighbour addresses. This means they are connected to `tor1` via multiple border links and they are potential neighbour ASes with deployment of BGP-M. These neighbour ASes will be queried for the identification of BGP-M.

The `summary` command helps to narrow down the list of neighbour ASes to be queried in two aspects. The first one is that within a *NearAS*, some border routers have neighbour ASes connected via multiple border links, while some border routers do not. The second one is that even for a given *NearBR*, not all of the neighbour ASes are connected to the *NearBR* via multiple border links. Therefore, for *NearASes* with `summary` command, I only need to query the neighbour ASes connected to *NearBRs* via multiple border links.

The other 343 ASes among the 527 ASes did not provide the `summary` command, so I am unable to obtain the neighbour ASes connected to each *NearBR*. Alternatively, I relied on the BGP RIB data provided by RouteViews [34]. The snapshot of RouteViews data was on 1/January/2020. Each RIB entry includes the AS path to a prefix. I extracted the ASes next to the *NearAS* in any RIB AS path as neighbour ASes. These neighbour ASes were queried from each *NearBR* within the

core1.tor1.he.net> show ip bgp summary

Local AS Number	6939
Number of Neighbors Configured	247, 229 up
Number of Routes Installed	3298347 (326536353 bytes)
Number of Routes Advertised	68081406 (4658509 entries) (223608432 bytes)
Number of Attribute Entries	680924 (61283160 bytes)

Neighbor Address	ASN	State	Time	Rt:Accepted
198.32.181.61	19551	ESTAB	483d20h26m	67
198.32.181.46	19752	ESTAB	570d 6h 5m	72
206.108.34.48	19752	ESTAB	47d 9h 8m	72
206.108.34.73	20161	ESTAB	92d22h 0m	105
206.108.34.164	20365	CONN	250d 5h53m	0
206.108.35.117	20473	ESTAB	92d21h56m	55
206.108.34.24	20940	ESTAB	74d12h54m	45
216.66.14.42	21513	ESTAB	260d 6h17m	17
206.108.34.102	21724	ESTAB	92d22h 0m	58
206.108.34.184	21834	CONN	185d21h56m	0
206.108.34.233	21834	CONN	305d22h40m	0
206.108.34.31	21949	ESTAB	56d 0h47m	147
209.51.168.70	22264	ESTAB	329d 4h 9m	1
206.108.35.7	22616	ESTAB	92d21h59m	3
206.108.35.8	22616	ESTAB	9d17h 3m	3
198.32.181.38	22634	ESTAB	403d 1h 4m	8

Figure 4.1: An example of LG response to the command `show ip bgp summary`. Each red box highlights an example of a neighbour AS with multiple neighbour addresses.

NearAS. This inevitably increased the number of queries to *NearBR* and the overall overhead. But this is necessary for a complete measurement on the deployment of BGP-M.

4.3.3 Retrieving Routing Tables

For each *NearBR*, I retrieved its routing table information using the `routes` command. I take `show ip bgp routes detail <IP address>` for illustration. To determine the parameter `IP address`, I obtained the prefixes announced by each neighbour AS with data provided by RouteViews [34]. The snapshot of RouteViews data was on 1/January/2020.

According to the principle of BGP, a border router only learns and announces routes to prefixes. Queries to all the IP addresses in the same destination prefix should return the same routing table. Hence, I only queried one IP address in each prefix, and set the parameter `IP address` as `X.Y.Z.1` for IPv4 (or `X:Y:Z::1` for IPv6) for each prefix in a neighbour AS.

I only query /24 prefixes for IPv4 and /48 prefixes for IPv6 for simplicity,

which account for 56.5% and 45.8% in the RouteViews data for IPv4 and IPv6, respectively, as shown in Table 4.2.

Table 4.2: Number and proportion of prefixes with different lengths in the BGP data provided by RouteViews, on IPv4 and IPv6.

Prefixes on IPv4			Prefixes on IPv6		
Length	Number	Proportion	Length	Number	Proportion
24	524,203	56.5%	48	43,646	45.8%
22	111,606	12.0%	32	13,699	14.4%
23	95,012	10.2%	44	5,127	5.4%
20	45,329	5.0%	40	4,513	4.7%
32	40,361	4.3%	36	3,565	3.7%
Other	111,471	12.0%	Other	24,772	26.0%
Total	927,982	100.0%	Total	91,912	100.0%

4.3.4 Identifying the Deployment of BGP-M

Figure 4.2 shows an example response to the command of `show ip bgp routes detail <IP address>` from the border router `tor1` of HE. The output is provided in both table format and raw format. `tor1` has learned four paths from the next hop IP addresses 198.32.181.46, 206.108.34.48, 198,170.18.29, and 196.32.181.50, for the traffic to the destination prefix (142.46.150.0/24). Among these paths, three paths are learned via external BGP (i.e., labelled with "E" in Figure 4.2), and the other path is via internal BGP (i.e., labelled with "I" in Figure 4.2). Three paths have the same AS path, containing only AS19752 (the neighbour AS and destination AS), and the other path contains two ASes.

Among the four paths, the paths learned from 198.32.181.46 and 206.108.34.48 (i.e. via two border links) are labelled with "M" (for "Multipath", see the first column in Figure 4.2), suggesting that that they have the same values for several attributes for Metric (with value of 0), LocPrf (i.e., `LocPref`, with value of 100), Path (with value of 19752), Origin (with value of IGP), and MED (with value of 0, see the raw format in Figure 4.2). Moreover, these two paths are both labelled with "E" (for "eBGP"), meaning they are learned via external BGP. towards the same destination prefix (142.46.150.0/24) in the neighbour AS of Hydro One Telecom Inc. (AS19752). Thus, the two paths are equal-cost multipath. This is the ground-

core1.tor1.he.net> show ip bgp routes detail 142.46.150.1

Matching Routes	6
Status Codes	**A** - Aggregate **B** - Best **b** - Not Install Best **C** - Confederation eBGP **D** - Damped **E** - eBGP **H** - History **I** - iBGP **L** - Local **M** - Multipath **m** - Not Installed Multipath **S** - Suppressed **F** - Filtered **s** - Stale **x** - Best-External

Status	Network	Next Hop	Metric	LocPrf	Weight	Path	Origin	ROA
BMEx	142.46.150.0/24	198.32.181.46	0	100	0	19752	IGP	-
ME	142.46.150.0/24	206.108.34.48	0	100	0	19752	IGP	-
ME	142.46.150.0/24	206.108.34.48	0	100	0	19752	IGP	-
ME	142.46.150.0/24	206.108.34.48	0	100	0	19752	IGP	-
I	142.46.150.0/24	198.179.18.29	80	100	0	19752	IGP	-
E	142.46.150.0/24	198.32.181.50	0	100	0	6327, 19752	IGP	-

Last Update 51d6h11m58s ago (2 paths installed)

```
core1.tor1.he.net> show ip bgp routes detail 142.46.150.1
  Number of BGP Routes matching display condition : 6
       S:SUPPRESSED F:FILTERED s:STALE x:BEST-EXTERNAL
1       Prefix: 142.46.150.0/24, Rx path-id:0x00000000, Tx path-id:0x00030001, rank:0x00000001, Status: BMEx,
         NEXT_HOP: 198.32.181.46, Metric: 0, Learned from Peer: 198.32.181.46 (19752)
          LOCAL_PREF: 100,  MED: 0,  ORIGIN: igp,  Weight: 0,  GROUP_BEST: 1
         AS_PATH: 19752
            COMMUNITIES: 6939:7036 6939:8124 6939:9001
2       Prefix: 142.46.150.0/24, Rx path-id:0x00000000, Tx path-id:0x002b0001, rank:0x00000002, Status: ME,
         NEXT_HOP: 206.108.34.48, Metric: 0, Learned from Peer: 206.108.34.48 (19752)
          LOCAL_PREF: 100,  MED: 0,  ORIGIN: igp,  Weight: 0,  GROUP_BEST: 0
         AS_PATH: 19752
            COMMUNITIES: 6939:7036 6939:8124 6939:9001
3       Prefix: 142.46.150.0/24, Rx path-id:0x00000000, Tx path-id:0x00000000, rank:0x00000003, Status: ME,
         NEXT_HOP: 206.108.34.48, Metric: 0, Learned from Peer: 206.108.35.253 (11670)
          LOCAL_PREF: 100,  MED: 0,  ORIGIN: igp,  Weight: 0,  GROUP_BEST: 0
         AS_PATH: 19752
            COMMUNITIES: 6939:7036 6939:8124 6939:9001
4       Prefix: 142.46.150.0/24, Rx path-id:0x00000000, Tx path-id:0x00000000, rank:0x00000004, Status: ME,
         NEXT_HOP: 206.108.34.48, Metric: 0, Learned from Peer: 206.108.35.254 (11670)
          LOCAL_PREF: 100,  MED: 0,  ORIGIN: igp,  Weight: 0,  GROUP_BEST: 0
         AS_PATH: 19752
            COMMUNITIES: 6939:7036 6939:8124 6939:9001
5       Prefix: 142.46.150.0/24, Rx path-id:0x00000000, Tx path-id:0x01100001, rank:0x00000005, Status: I,
         NEXT_HOP: 198.179.18.29, Metric: 80, Learned from Peer: 216.218.252.193 (6939)
          LOCAL_PREF: 100,  MED: 0,  ORIGIN: igp,  Weight: 0,  GROUP_BEST: 0
         AS_PATH: 19752
            COMMUNITIES: 6939:1111 6939:7066 6939:8124 6939:9001
6       Prefix: 142.46.150.0/24, Rx path-id:0x00000000, Tx path-id:0x00040001, rank:0x00000006, Status: E,
         NEXT_HOP: 198.32.181.50, Metric: 0, Learned from Peer: 198.32.181.50 (6327)
          LOCAL_PREF: 100,  MED: 0,  ORIGIN: igp,  Weight: 0,  GROUP_BEST: 1
         AS_PATH: 6327 19752
            COMMUNITIES: 6939:7036 6939:8124 6939:9001
       Last update to IP routing table: 51d5h52m56s, 2 path(s) installed:
# Entry cached for another 60 seconds.
```

Figure 4.2: An example of LG response to the command of `show ip bgp routes detail <IP address>`, from the border router `core1.tor1.he.net` in Hurricane Electric. Both the table format and the raw format are provided.

truth evidence that Hurricane Electric has deployed BGP-M at `tor1` to Hydro One (AS19752). This BGP-M deployment is denoted as <HE, `tor1`, Hydro One>. Moreover, the response also reveals the BGP-M case of <HE, `tor1`, Hydro One, 142.46.150.0/24>.

Note that in the output, the path via 206.108.34.48 appears three times. The raw format shows that this path is actually learned from three different neighbour devices. According to PeeringDB [25], the IP addresses of 206.108.35.253 and 206.108.35.254 belong to the route servers of TorIX (with AS number AS11670). This means that both route servers advertise the same path (via 206.108.34.48) to HE, and `tor1` uses 206.108.34.48 as the next hop, instead of using 206.108.35.253

and 206.108.35.254. This confirms that IXP is 'transparent' in BGP routing.

If the query to a prefix in a neighbour AS reveals the deployment of BGP-M, a BGP-M case are recorded. If this is the first BGP-M case, a BGP-M deployment is therefore identified.

After all prefixes in a neighbour AS are queried, if no BGP-M deployment is identified, the query goes to another neighbour AS in the list. This does not indicate that BGP-M is not deployed at the border router to the neighbour AS, because I only query /24 (or /48) prefixes in the neighbour AS.

When all the obtained neighbour ASes are queried for a *NearBR*, the query goes to the next *NearBR*. When all the *NearBRs* of a *NearAS* are queried, the query goes to another *NearAS*.

4.4 Discussions

4.4.1 Advantages

The method proposed above can be used for the measurement of BGP-M in the Internet with LG server data. It has several advantages.

- The results are reliable. LG servers provide BGP routing tables obtained from border routers. These information directly reflect the configuration of network operators about BGP-M. Thus, the obtained data are ground-truth about BGP-M deployment.

- The method is simple to realise because it only requires access to the LG server of an AS and the major operation in the method is to query LG servers. Over 90% of the queries can be sent via automatic HTTP (Hypertext Transfer Protocol) requests.

- The method is resource-friendly. The process does not require complicated computation and operations on computers. The queries were sent and processed at three machines, which have 2-core CPU (Central Processing Unit) and 8GB RAM (Random Access Memory), 2-core CPU and 16GB RAM, and 16-core CPU and 64GB RAM, respectively.

- The method can obtain rich information on BGP-M. The LG server data not only provides values of *NearAS*, *NearBR*, *FarAS* and *DstPrfx*, but also other information such as *FarIPs*, the status codes, and how long BGP-M has been activated since last update. Most of these information are not available from traceroute data.

4.4.2 Datasets

The proposed method relied on datasets from LG servers and RouteViews. In my measurement on BGP-M, the queries to LG servers were between January 2020 to July 2021, and the datasets from RouteViews were from the snapshot collected on 1/January/2020. LG server data provides the direct information on the deployment of BGP-M. Datasets from RouteViews provide the lists of neighbour ASes and the prefixes announced by neighbour ASes for each nearside AS to be studied.

The datasets from RouteViews is fixed in my measurement because the aim of this study is to provide a general picture on the deployment of BGP-M in the Internet. First of all, using changing datasets from RouteViews requires filtering inconsistent datasets, which may produce little new knowledge but much unnecessary overhead in my measurement. Besides, this allows me to study the change of BGP-M deployment with focus on LG servers, i.e., how nearside ASes' configuration changes, and this is reflected in my study presented in Section 5.3.2.

4.4.3 Limitations

The method is based on queries to LG servers. As analysed in Table 4.1, only 2,709 ASes in the Internet provide LG servers, and 527 ASes in them support `routes` command. Thus, the vantage points to the Internet may not be sufficient enough to reflect the whole picture about BGP-M deployment in the Internet.

Aside from the LG servers, the limited numbers of queried prefixes will also affect the measurement result. I only queried the /24 prefixes on IPv4 and /48 prefixes on IPv6, and not all of the prefixes are queried in my measurement. Therefore, for these 12 ASes, there are more BGP-M deployments to be discovered.

Chapter 5

Analysis on the Deployment of BGP-M in the Internet

Based on the method in Chapter 4, I have queried all accessible LG servers. Because there has been no prior-knowledge provided in the literature, an expected discovery is that BGP-M has been widely deployed in the Internet.

My results show that BGP-M has been deployed by 12 ASes in the Internet. This is a tiny portion in the ASes in the global Internet. This observation is caused by reasons like limited number of accessible LG servers and limited numbers of queried prefixes.

My results also show that different ASes can exhibit different behaviours in their deployment of BGP-M. This is possibly related to their roles in Internet routing. For example, transit ASes like Hurricane Electric requires complex connectivity and routing with many neighbour ASes while stub ASes only need connection to their provider ASes, causing huge difference in the numbers of BGP-M deployments deployed by them.

Despite the limitations on the results, the analysis in this chapter aims to provide knowledge on the deployment of BGP-M, especially on how a large transit AS has deployed BGP-M with its neighbour ASes. I will firstly summarise the BGP-M deployment in the whole Internet, on both IPv4 and IPv6 Internet. Then I will analyse the deployment of BGP-M by Hurricane Electric (HE, AS6939) from the perspectives of BGP-M deployment and BGP-M case. I will also briefly introduce

the results about the other 11 ASes and discuss the measurement results. Most of the analysis in this chapter is about HE because it is a major Internet service provider and has the largest number of BGP-M deployments in my results.

5.1 Deployment of BGP-M in the Internet

As shown in Table 4.1, 527 ASes have accessible LG servers and provide the `routes` command. I have sent over 1.5 million queries to these LG servers, and discovered BGP-M deployed by 12 ASes on both IPv4 and IPv6 Internet.

Table 5.1 lists information of these ASes, including AS number, short AS name, full AS name and AS rank. Table 5.2 lists the statistics about the BGP-M deployments for each of these ASes on both IPv4 and IPv6. Their BGP-M deployments are analysed from three perspectives: connectivity fabrics (IXP, Direct, and Hybrid), the border routers and the neighbour ASes. AS number and AS rank are not listed in Table 5.2. These ASes are ordered in Tables 6.1 and 6.2 according to their AS ranks. The AS rank information was provided by CAIDA's AS rank data on 1/January/2020 [4]. There was no AS rank information for TechCom (AS196965).

Table 5.1: Basic information about the 12 ASes with BGP-M deployment

AS number	Short AS name	Full AS name	AS rank
6939	HE	Hurricane Electric LLC	7
9002	RETN	RETN Limited	13
3216	Vimpelcom	PJSC Vimpelcom	25
20764	RASCOM	CJSC RASCOM	30
8647	AS-T2012	LLC TELEMIST 2012	1264
22691	ISPnet	ISPnet	2337
52201	TCTEL	OOO Suntel	3788
12303	ISZT	Council of Hungarian Internet Providers	4104
328112	LBSD	Linux-Based-Systems-Design-AS	6339
48972	BetterBe	BetterBe B.V.	35096
131713	IDNIC	PT Sano Komunikasi	45081
196965	TechCom	TechCom	–

The 12 ASes listed in Table 5.2 can be roughly divided into two groups, i.e. large transit ASes and stub ASes, according to their AS ranks. Large transit ASes include HE (AS6939), RETN (AS9002), Vimpelcom (AS3216) and RASCOM (AS20764). Stub ASes include the remained 8 ASes.

Table 5.2: Statistics about the ASes with BGP-M deployment in the Internet

AS	Number of BGP-M deployments				Number of border routers			Number of neighbour ASes		
name	Total	IXP	Direct	Hybrid	Total	w/ BGP-M	Ratio	Total	w/ BGP-M	Ratio
IPv4										
HE	1,088	1,006	68	14	112	69	61.6%	5,868	611	10.4%
RETN	155	87	65	3	130	51	39.2%	1,547	108	7.0%
Vimpelcom	2	0	2	0	16	2	12.5%	770	2	0.3%
RASCOM	27	23	4	0	27	6	22.2%	858	23	2.7%
ISPnet	3	0	3	0	7	1	14.3%	24	3	12.5%
TCTEL	1	0	1	0	1	1	100.0%	11	1	9.1%
ISZT	2	2	0	0	2	1	50.0%	59	2	3.4%
LBSD	13	0	2	11	2	1	50.0%	29	13	44.9%
BetterBe	2	2	0	0	4	2	50.0%	9	1	11.1%
IDNIC	1	1	0	0	5	1	20.0%	10	1	10.0%
TechCom	24	24	0	0	2	2	100.0%	36	15	41.7%
IPv6										
HE	300	266	14	20	112	35	31.3%	3,880	146	3.8%
RETN	45	25	18	2	130	24	18.5%	926	23	2.5%
AS-T2012	2	2	0	0	1	1	100.0%	46	2	4.3%
LBSD	6	6	0	0	2	1	50.0%	28	6	21.4%
BetterBe	2	2	0	0	4	2	50.0%	6	1	16.7%
IDNIC	1	1	0	0	5	1	20.0%	5	1	20.0%

Overall, I have identified 1,674 unique BGP-M deployments, with 1,318 deployments on the IPv4 network and 356 deployments on IPv6. As shown in Table 5.2, most of the BGP-M deployments are deployed by the large transit ASes, especially by HE (AS6939) and RETN (AS9002). A plausible explanation is that large transit ASes have more complicated connectivity fabrics than the stub ASes. Thus, large transit ASes require more load balancing for inter-domain traffic. Moreover, the observation of BGP-M deployments on these 12 ASes suggests that BGP-M is indeed helpful to the inter-domain routing for both large transit ASes and stub ASes. That is to say, BGP-M can be deployed by an AS regardless of its scale, as long as it requires load balancing with its neighbour ASes.

These ASes are a small portion in the existing ASes, but their deployment of BGP-M can provide some basic knowledge or insights on the deployment of BGP-M in the world.

5.2 BGP-M Deployments by Hurricane Electric (HE, AS6939)

As shown in Table 5.2, the most notable AS in the inference result is HE (AS6939), a Tier-1 network, ranked 7th in the Internet [4]. As a major Internet service provider, HE had 112 border routers, neighbouring with 5,868 ASes on IPv4 and neighbouring with 3,880 ASes on IPv6 in January of 2020. It is remarkable that, as shown in Table 5.2, HE has already extensively implemented *at least* 1,088 BGP-M deployments at 69 border routers to prefixes in 611 of its neighbour ASes on IPv4, and implemented *at least* 300 BGP-M deployments at 35 border routers to 146 neighbour ASes on IPv6. Thus, this section focuses on Hurricane Electric and studies its deployment of BGP-M.

5.2.1 Variety in Connectivity Fabrics

This section analyses the BGP-M deployments deployed by HE from the perspective of connectivity fabrics, aiming to study the relation between connectivity fabrics and the deployment of BGP-M.

Among the 1,088 BGP-M deployments on IPv4, 911 deployments are via 2 links, 92 deployments are via 3 links and 85 deployments are via 4 links. And among the 300 BGP-M deployments on IPv6, 248 deployments are via 2 links, 33 deployments are via 3 links and 19 deployments are via 4 links.

I relied on the data from PeeringDB [25] to identify whether a BGP-M deployment was via IXP, or direct links, or hybrid links. I obtained [88] a list of IXPs and the prefixes belonging to them in January 2020. I searched the list for the *FarIPs* in each deployment. If all the *FarIPs* in a BGP-M deployment belong to IXPs, this deployment is identified as via IXPs (IXP); if none of the *FarIPs* belong to IXPs, this deployment is identified as via direct links (Direct); otherwise, this deployment is identified as via hybrid links (Hybrid).

Among the 1,088 IPv4 deployments, 1,006 (92.5%) deployments are via IXPs, 68 deployments are via direct links, and 14 deployments are via hybrid links. Among the 300 IPv6 deployments, 266 (88.9%) deployments are via IXP, 14 de-

ployments are via direct links, and 20 deployments are via hybrid links. This indicates that various connectivity fabrics can be used for the deployment of BGP-M, and IXPs play a vital role in HE's BGP-M deployment.

Note that I only used the IXP data from one data source (i.e. PeeringDB), instead of multiple sources, so some BGP-M deployments via direct links or via hybrid links might be actually via IXP. Thus, IXPs might be more important in BGP-M deployment than I observed.

5.2.2 Global Distribution of Border Routers with BGP-M

This section studies HE's deployment of BGP-M from the perspective of border routers, in order to understand how geographical locations affect the deployment of BGP-M.

Hurricane Electric's LG server covered 112 border routers distributed around the world, based on the geo-locations extracted from the router names as given by the LG server.

Table 5.3 lists the geographical distribution of the border routers. As can be seen, most (95 in total) of the border routers are located in North America and Europe, among which 60 border routers have been implemented with BGP-M on IPv4 and 28 border routers with BGP-M on IPv6. Although there are only a few border routers located in Asia and other parts of the world, a large portion of them have been implemented with BGP-M. This suggests that the global distribution of HE's border routers enable its extensive deployment of BGP-M.

Figure 5.1 plots the number of neighbour ASes in triangle and the number of neighbour ASes with BGP-M deployment in square at each of HE's 112 border routers on IPv4 and IPv6. Y-aix in the figure is plotted on log scale. The border routers are ordered by the number of IPv4 neighbour ASes.

In general, the number of neighbour ASes with BGP-M (in square) does not follow the trend of neighbour ASes (in triangle) on either IPv4 and IPv6. This suggests the deployment of BGP-M at each border router is not determined by the number of neighbour ASes, but determined by the actual requirement for load balancing.

Table 5.3: Geographical distribution of Hurricane Electric's border routers.

	Number of border routers	with BGP-M deployments (IPv4, IPv6)
North America	**55**	**33, 19**
United States	47	26, 15
Canada	8	7, 4
Europe	**40**	**27, 9**
Germany	5	4, 0
United Kingdom	3	2, 0
France	2	2, 0
Other	30	19, 9
Asia	**6**	**4, 4**
Other	**11**	**5, 3**
Total	**112**	**69, 35**

On IPv4, HE has deployed BGP-M at the border router `par2` to the largest number (78) of neighbour ASes. Among the 10 border routers with the largest numbers of BGP-M deployments, the top-4 border routers are in Europe, followed by 3 border routers in Asia, 1 border router is in Africa and 2 border routers are in North America.

On IPv6, HE has deployed BGP-M at `ams1` to the largest number 38 neighbour ASes. The 10 border routers with the largest numbers of BGP-M deployments include 1 border router in Europe, 3 border routers in Asia and 6 border routers in North America.

My queries to HE's 112 border routers involve the commands of `show ip bgp summary` and `show ip bgp routes detail <IP address>`. These commands are both in Cisco-style. Moreover, the descriptions and format of the responses to the commands are also in Cisco-style. Therefore, I infer that HE has deployed BGP-M at Cisco routers. Identifying the vendors for the routers deployed with BGP-M is helpful to study the routing properties of BGP-M as introduced in Chapter 6.

5.2.3 Diversity of Neighbour ASes

This section studies HE's BGP-M deployment from the perspective of neighbour ASes, aiming to find any relation between the deployment of BGP-M and neighbour

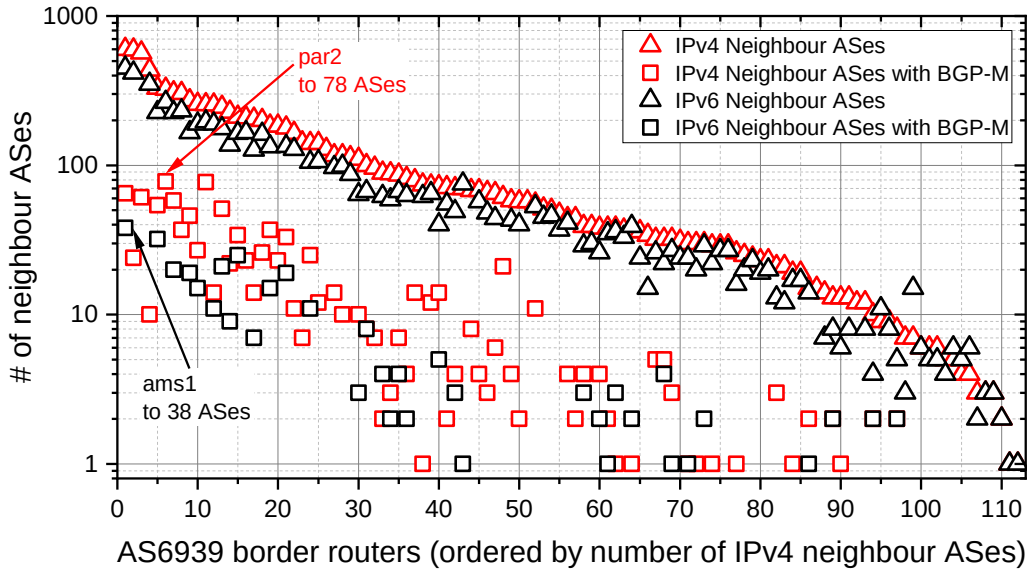


Figure 5.1: List of 112 border routers of Hurricane Electric (AS6939). The border routers are ordered by the number of connected IPv4 neighbour ASes. Plots in triangle indicates the number of neighbour ASes, and plots in square indicates the number of neighbour ASes with BGP-M deployment.

ASes. The identified BGP-M deployments by HE are deployed to 611 neighbour ASes on IPv4 and to 146 neighbour ASes on IPv6.

5.2.3.1 On IPv4

Figure 5.2 plots the neighbour ASes deployed with BGP-M by HE on IPv4, ordered by their customer cone sizes [4] (in red). These ASes are in four groups according to their customer cone sizes, with the number of ASes in each group as 30, 80, 59 and 439. 3 ASes were missing in the plot due to the lack of information in the AS rank data snapshot [4]. The plot also shows for each neighbour AS the Total number of BGP-M deployments in large circle, and the number of BGP-M deployments via IXP in small circle.

The plot shows four interesting observations. Firstly, since the customer cone size determines an AS' rank [4], the plot shows that Hurricane Electric has deployed BGP-M extensively to neighbour ASes among different rank groups, suggesting that the reason for the deployment of BGP-M to a neighbour AS is whether the traffic to the neighbour AS requires load balancing, instead of the scale of the neighbour AS. Secondly, Yahoo! (AS10310), a content provider network with customer cone

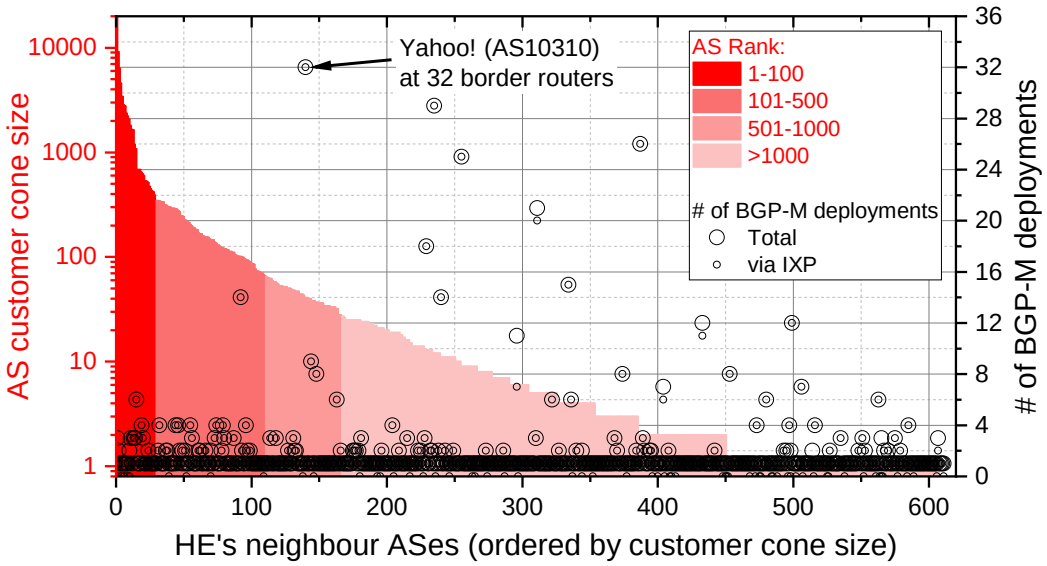


Figure 5.2: Hurricane Electric (HE, AS6939)'s neighbour ASes with BGP-M deployment on IPv4. The neighbour ASes are ordered by their customer cone sizes (y axis on the left in red colour). Also shown is the total number of HE border routers with BGP-M deployment (y axis on the right in black colour) to each neighbour AS and the number of border routers with BGP-M to a neighbour AS via an IXP.

size of 41 and AS rank of 747, is deployed with BGP-M by HE at as many as 32 border routers. Thirdly, small & medium ASes (with customer cone size $<$100) are more likely to be deployed with BGP-M at multiple border routers, suggesting Hurricane Electric has deployed richer and more complex connections to small & medium ASes than to top-rank ASes. Fourthly, IXPs are widely involved in Hurricane Electric's BGP-M deployment. For many neighbour ASes, all of their BGP-M deployments are connected via IXP(s). It is possible that HE's heavy reliance on IXP is a reason why I have observed so many BGP-M deployments with small & medium ASes.

Table 5.4 lists the 10 highest ranked neighbour ASes with BGP-M deployment by HE on IPv4. As can be seen, these ASes are deployed with BGP-M at only a few ($<=$ 3) border routers.

Table 5.5 lists the 10 neighbour ASes with the largest numbers of BGP-M deployments deployed by HE on IPv4. Although these ASes are not highly ranked, most of them are well-known content provider networks. And eight of them are so-

Table 5.4: The 10 highest ranked HE's neighbour ASes with BGP-M deployments – IPv4.

CAIDA's AS rank	AS number	Customer cone size	AS name	# of BGP-M deployments
2	1299	32,929	Telia Company	3
9	6461	9,175	Zayo Bandwidth	2
13	9002	6,374	RETN	1
15	4637	4,548	Telstra	1
20	12389	3,425	PJSC Rostelecom	1
24	7922	2,820	Comcast Cable	1
25	3216	2,777	Vimpelcom	1
27	9498	2,361	Bharti Airtel	1
29	6830	2,218	Liberty Global	1
30	20764	2,073	RASCOM	2

Telstra: Telstra International Limited

called hyper-giant ASes with wide geographical coverage, large port capacity and large traffic volume [53], which are Yahoo! (AS10310), Cloudflare (AS13335), Apple (AS714), MicroSoft (AS8075), Twitch (AS46489), Amazon (AS16509), Google (AS15169) and Twitter (AS13414).

A comparison between Table 5.4 and Table 5.5 highlights the difference between top-rank ASes and hyper-giant ASes (with low ranks) in terms of the requirement for BGP-M deployment. BGP-M is more needed and useful for routing with content providers, where load balancing can be crucial for delivery of large traffic volume.

5.2.3.2 On IPv6

Figure 5.3 plots the neighbour ASes deployed with BGP-M by HE on IPv6. These ASes are in four groups according to their customer cone sizes, with numbers in the groups being 12, 25, 14 and 89. 6 neighbour ASes are not shown in the plot due to the lack of AS rank data. The plot in the figure shows similar observations to Figure 5.2. Those observations are (1) the size of a neighbour AS is not the reason for it to be deployed with BGP-M, but whether the traffic to it requires load balancing; (2) Yahoo! (AS10310) is deployed with BGP-M at the largest number of (16) border routers; (3) small & medium ASes are more likely to be deployed with BGP-M at multiple border routers; and (4) IXPs are widely involved in HE's

Table 5.5: Ten neighbour ASes of Hurricane Electric (AS6939) with the largest numbers of BGP-M deployments.

IPv4				IPv6			
AS name	AS number	AS rank	# of BGP-M deployments	AS name	AS number	AS rank	# of BGP-M deployments
Yahoo!	10310	747	32	Yahoo!	10310	747	16
Cloudflare	13335	1845	29	Cloudflare	13335	1845	15
Apple	714	6385	26	Google	15169	1743	12
MicroSoft	8075	2288	25	Apple	714	6385	12
Twitch	46489	33522	25	MicroSoft	8075	2288	11
Fastly	54113	38523	25	Fastly	54113	38523	11
Amazon	16509	3560	21	Amazon	16509	3560	10
Google	15169	1743	18	Verizon	15133	3172	9
Twitter	13414	4119	15	WoodyNet	42	1931	7
WoodyNet	42	1931	14	Limelight	22822	344	6

BGP-M deployment.

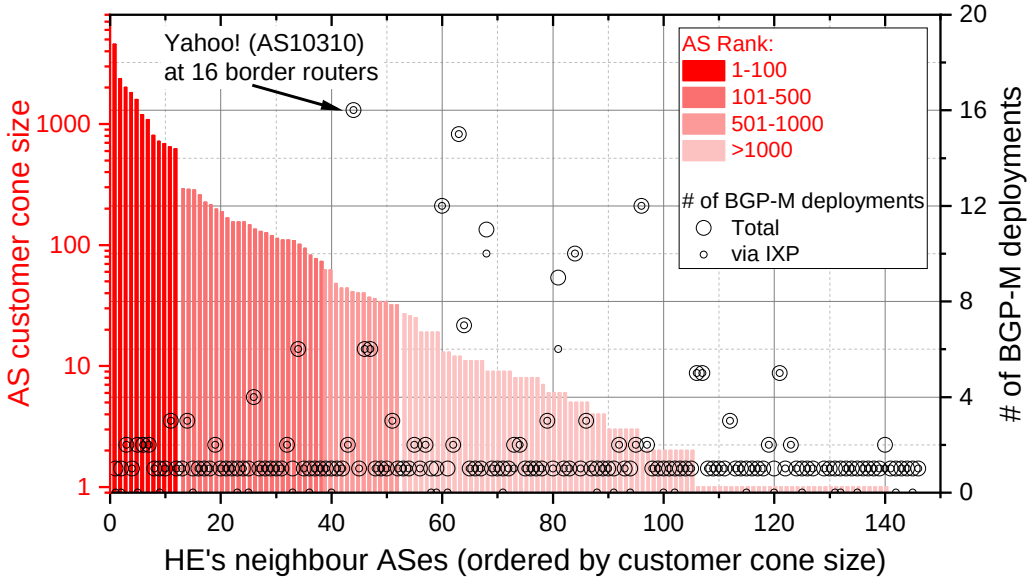


Figure 5.3: Hurricane Electric (HE, AS6939)'s neighbour ASes with BGP-M deployment on IPv6.

Table 5.6 lists the 10 highest ranked neighbour ASes with BGP-M deployments deployed by HE on IPv6. As can be seen, these ASes are deployed with BGP-M at very small number ($<= 3$) of border routers. Table 5.5 lists the 10 neighbour ASes with the largest numbers of BGP-M deployments on IPv6. The lists for IPv6 and for IPv4 are similar to each other by sharing 8 ASes in common. The differences

are the change of order for the 8 shared ASes, and Twitch (AS46489) and Twitter (AS13414) in the IPv4 list being replaced by Verizon (AS15133) and Limelight (AS22822) in the IPv6 list. A comparison between Table 5.5 and Table 5.6 again highlights that the deployment of BGP-M is more useful for routing with content providers.

Table 5.6: The 10 highest ranked HE's neighbour ASes with BGP-M deployments – IPv6.

CAIDA's AS rank	AS number	Customer cone size	AS name	# of BGP-M deployments
15	4637	4,548	Telstra	1
27	9498	2,361	Bharti Airtel	1
32	52320	2,005	GlobeNet	2
36	8359	1,810	MTS PJSC	1
40	4826	1,593	Vocus	2
48	41095	1,190	IPTP LTD	1
51	8220	1,083	COLT	2
57	4230	805	CLARO S.A.	1
65	5588	686	GTSCE	1
69	3303	647	Swisscom	3

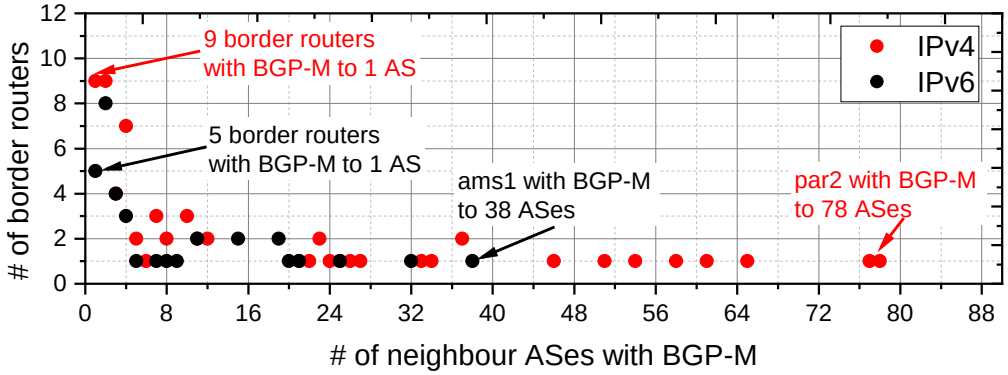
Telstra: Telstra International Limited
GlobeNet: GlobeNet Cabos Submarinos Colombia, S.A.S.
Vocus: Vocus Communications
COLT: COLT Technology Services Group Limited
GTSCE: T-Mobile Czech Republic a.s.
Swisscom: Swisscom (Schweiz) AG

On one hand, HE's BGP-M deployments on IPv6 Internet is much less than that on IPv4 Internet. This may be caused by two reasons. The first reason is that IPv6 is still under the process of deployment. The second reason is that IPv6 indeed requires less BGP-M than IPv4 Internet. On the other hand, the results on IPv4 and IPv6 show similar trends and observations. For example, HE has deployed more BGP-M to content provider networks than transit ASes, and IXP has played an important role in HE's deployment of BGP-M.

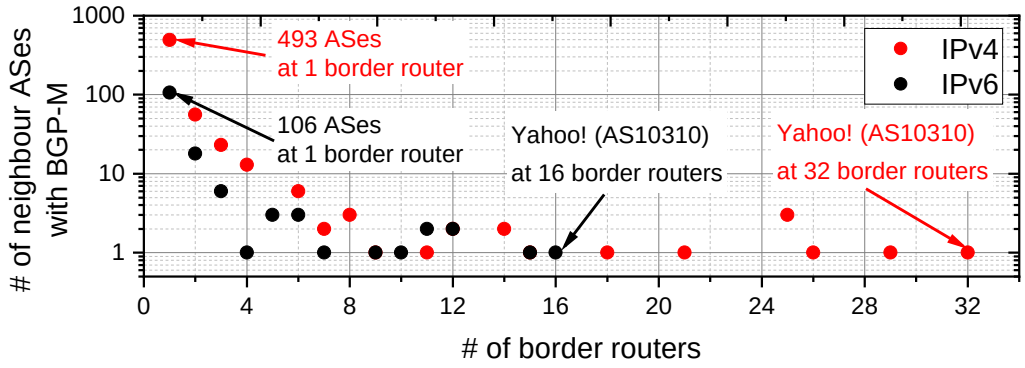
5.2.4 Relation between Border Routers and Neighbour ASes

Figure 5.4 shows the relation between the number of border routers and the number of neighbour ASes with BGP-M deployment within HE on IPv4 and IPv6. The

figures describe the deployment of BGP-M deployed by HE from different angles for a better and clearer understanding.



(a) Number of border routers as a function of the number of neighbour ASes.



(b) Number of neighbour ASes as a function of the number of border routers.

Figure 5.4: Relation between the number of border routers and the number of neighbour ASes with BGP-M deployment in Hurricane Electric.

Figure 5.4(a) plots the number of border routers as a function of the number of neighbour ASes with BGP-M. On IPv4, HE has deployed BGP-M at 9 border routers to only one neighbour AS, while it deployed BGP-M at the other 60 border routers to at least two neighbour ASes. The border router `par2` is deployed with BGP-M to the largest number of (78) neighbour ASes. On IPv6, HE deployed BGP-M at 5 border routers to only one neighbour AS, while it deployed BGP-M at the other 30 border routers to at least two neighbour ASes. The border router `ams1` is deployed with BGP-M to the largest number of (38) neighbour ASes.

Figure 5.4(b) plots the number of neighbour ASes with BGP-M as a function of the number of border routers. On IPv4, HE has deployed BGP-M to 493 neighbour

ASes at only one border router, and to the other 118 neighbour ASes at at least 2 border routers. HE has deployed BGP-M with Yahoo! (AS10310) at 32 border routers, accounting for 87% of the border routers connected to Yahoo! (AS10310). On IPv6, HE has deployed BGP-M to 106 neighbour ASes at only one border router, and to the other 40 neighbour ASes at at least 2 border routers. HE deployed BGP-M to Yahoo! (AS10310) at 16 border routers, accounting for 47% of the border routers connected to Yahoo! (AS10310).

5.3 BGP-M Cases Deployed by HE

The previous section has analysed HE's BGP-M deployments. As defined in Section 4.1, each 3-tuple BGP-M deployment at a border router can support many 4-tuple BGP-M cases for routing towards different destination prefixes. This section studies the BGP-M cases deployed by HE from two perspectives, i.e. the identified BGP-M cases and the change of BGP-M cases.

5.3.1 Identified BGP-M Cases Deployed by HE

Basically, a BGP-M deployment is identified by discovering a BGP-M case to any prefix in the *FarAS*; whereas in this section, I aim to discover BGP-M cases to as many prefixes in the *FarAS* as possible.

As the aim of this study is to infer the relation between BGP-M case and BGP-M deployment, e.g. how many BGP-M cases are there for each BGP-M deployment, here I measure BGP-M cases deployed by HE to prefixes in its neighbour ASes on IPv4 only.

For each of the 1,088 BGP-M deployments by HE on the IPv4 Internet, I send queries to the relevant border routers for all prefixes in the relevant farside ASes, and in total I have discovered 12,642 BGP-M cases relevant to the 1,088 deployments. This suggests that on average, each BGP-M deployment has been used for traffic routing to more than 10 destination prefixes in the farside AS alone.

Table 5.7 lists the 5 neighbour ASes of HE with the largest numbers of BGP-M cases. These neighbour ASes are all hyper-giant ASes and four of them (i.e., except for Cloudflare (AS13335)) are content provider networks. This confirms

the finding in Section 5.2.3 that BGP-M tends to be used for routing with content provider networks, where load balancing can be crucial for delivery of large traffic volumes. Table 5.7 also lists the 5 border routers with the largest numbers of BGP-M cases, suggesting the dense connectivity fabrics of HE and its heavy reliance on BGP-M load balancing at these border routers.

Table 5.7: The 5 neighbour ASes and the 5 border routers with the largest numbers of BGP-M cases of HE.

HE's neighbour AS			HE's border router	
AS number	AS name	# of BGP-M cases	Router name	# of BGP-M cases
714	Apple	3,684	`sto1`	683
15169	Google	3,500	`ams1`	659
16509	Amazon	1,701	`mia1`	650
13335	Cloudflare	1,103	`jnb1`	645
8075	MicroSoft	352	`tyo1`	590

Table 5.8 lists the 10 BGP-M deployments with the largest numbers of BGP-M cases. These deployments are all to Google (AS15169), suggesting Google has dense connectivity fabrics and it relies heavily on BGP-M load balancing for traffic coming from HE.

Table 5.8: The 10 BGP-M deployments of HE with the largest numbers of BGP-M cases, where AS15169 is Google.

	BGP-M deployment	# of cases		BGP-M deployment	# of cases
1	<AS6939, `jnb1`, AS15169>	382	6	<AS6939, `mil2`, AS15169>	228
2	<AS6939, `ams1`, AS15169>	238	7	<AS6939, `tyo1`, AS15169>	227
3	<AS6939, `mrs1`, AS15169>	238	8	<AS6939, `mia1`, AS15169>	223
4	<AS6939, `waw1`, AS15169>	233	9	<AS6939, `zrh3`, AS15169>	215
5	<AS6939, `sto1`, AS15169>	230	10	<AS6939, `kbp1`, AS15169>	213

5.3.2 Change of BGP-M Cases by HE

LG server's response to the command of `show ip bgp routes detail <IP address>` contains rich details on a BGP-M case, including the time lapse since the routing table has been last updated (see Figure 4.2) from which I can derive the precise time when the routing table became valid. Thus it is possible to monitor any change of a BGP-M case by re-querying the *DstPrfx* at a later time.

When measuring BGP-M deployments, I queried prefixes announced by the farside AS, one by one, until the first BGP-M case is identified. Thus, for example, the 1,088 BGP-M deployments by HE correspond to 1,088 BGP-M cases. The 1,088 BGP-M cases by HE on IPv4 were discovered in January to May in 2020, and the 300 BGP-M cases by HE on IPv6 were discovered in July to October 2020. In July 2021, I revisited the 1,388 BGP-M cases by HE by re-querying the *DstPrfx* in each of these BGP-M cases. The results are shown in Table 5.9.

Table 5.9: Revisit of the BGP-M cases deployed by HE on IPv4 and IPv6 Internet

	IPv4	IPv6
2020 measurement dates	Jan-May 2020	July-Oct 2020
Total # of BGP-M cases revisited	1,088	300
2021 measurement date	July 2021	July 2021
# of remaining cases	692	218
Exactly same as before	632	204
With different BLs	27	7
With more BLs	33	7
# of disappeared cases	396	82
NearBR 'Not existing'	13	0
'No routes' for *DstPrfx*	143	12
Status without 'M' (multipath)	109	25
Status without 'E' (eBGP)	102	32
Via different *FarAS*	29	13

For the 1,088 BGP-M cases by HE on the IPv4 Internet, 632 (or 58%) of the cases remained exactly the same; and 60 cases had replaced or additional border links, which, according to my definition, were still of the same BGP-M cases as they were deployed at the same border routers via border links to the same farside AS to the same destination prefixes. I also observed that 396 (or 36%) cases were disappeared since my 2020 measurement. For example, LG queries suggested some nearside border routers were 'Not existing' anymore, and some routes were not labelled as 'M' (i.e. multipath) anymore. A small number of cases were observed with farside ASes that were different from those observed in my 2020 measurement, making them different or new BGP-M cases. I observed similar results for the BGP-M cases on IPv6.

These observations suggest that HE has been carefully and actively maintaining and rearranging its BGP-M cases. Some of the changes may occur due to network changes, and others were likely to achieve more optimal configuration in order to better utilise the benefits of BGP-M load balancing.

5.4 Deployment of BGP-M by Other Network Operators

5.4.1 RETN (AS9002)

As shown in Table 5.2, RETN (AS9002) is the second highest ranked AS (13th) and it also has the second largest number of BGP-M deployments identified.

The commands used to query RETN's border routers are `show bgp summary` and `show route detail protocol bgp table <IP address>`, both of which are in Juniper-style. The responses to the commands are also in Juniper-style. Thus, it can be inferred that RETN has deployed BGP-M at Juniper routers.

Although using routers produced by different manufactures, RETN's BGP-M deployments show similar properties to those of HE. For example, (1) IXPs are important in the deployment of BGP-M; (2) BGP-M are heavily used by the *NearAS* for load balancing with content providers; (3) BGP-M is deployed in flexible ways; and (4) the deployment of BGP-M on IPv4 and IPv6 Internet reveal similar observations.

5.4.2 Other Lower-Ranked ASes

As shown in Table 5.2, the other 10 ASes have much fewer BGP-M deployments. It is notable that even very small ASes can deploy BGP-M – although with much simpler connectivity fabrics. IXPs are commonly involved in BGP-M deployments by large or small ASes.

Regarding the type of border routers, I can infer from LG commands used for querying that Vimpelcom (AS3216) and RASCOM (AS20764) use both Cisco and Juniper routers; ISPnet (AS22691), TCTEL (AS52201), ISZT (AS12303), LBSD

(AS328112), BetterBe (AS48972), IDNIC (AS131713) and TechCom (AS196965) use Cisco routers only; and AS-T2012 (AS8647) uses Juniper routers only. Although Huawei is a major producer of routers, we have identified no BGP-M deployment at any Huawei router.

5.5 Discussion

This chapter has analysed the results obtained from the over 1.5 million queries to the LG servers. The process of sending queries and analysing the results requires a lot of manual work. Firstly, even though over 90% of the queries can be sent via automatic requests, I need to set intervals between two consecutive requests to avoid being detected as malicious attacks. Secondly, the other 10% queries should be sent manually. Thirdly, because each LG server has its own format in presenting its routing table, I need to manually check the output and look for patterns to conduct automatic processing.

Despite the manual work, my results revealed that BGP-M has been deployed in the Internet by 12 ASes. Among the 12 ASes, HE has 1,088 BGP-M deployments at more than 60% of its border routers to over 600 of its neighbour ASes on the IPv4 Internet. IXPs have played an important role in HE's deployment of BGP-M. HE's globally distributed border routers enables its extensive deployment of BGP-M. HE tends to deploy more BGP-M to content provider networks than to transit ASes because the former requires load balancing more than the latter. HE has also been actively maintaining its deployment of BGP-M. The BGP-M deployments deployed by the other 11 ASes are much less than those by HE, but they reveal that the deployment of BGP-M is not determined by the scale of an AS, but its requirement of load balancing with neighbour ASes.

The measurement of BGP-M cases reveals that on average, a given BGP-M deployment can be used for traffic routing to more than 10 destination prefixes in the farside AS. Moreover, these observations confirm that BGP-M deployments are extensively utilised by HE for load balancing, especially for traffic to content provider networks.

My measurement is limited in two aspects. First of all, the study on these ASes is still not complete because only a part of the prefixes are queried and only the prefixes announced by neighbour ASes are queried. A part of future work is to query all the prefixes announced by both the neighbour ASes and remote ASes because a BGP-M deployment can be used for prefixes belonging to both neighbour ASes and remote ASes.

Secondly, my measurement only reveals deployment of BGP-M by a small number of ASes. One reason is that I only had access to a limited number of ASes' LG servers. It is possible that more BGP-M deployments will be discovered if more ASes provide access to their LG servers. Another reason is that the ASes with LG servers only account for a small portion in the existing AS numbers. A possible direction is to rely on traceroute data to infer wider range of deployment of BGP-M.

Chapter 6

Study of BGP-M Routing Properties Based on Traceroute Probing

Chapters 4-5 described my method to measure BGP-M and analysed how network operators deploy BGP-M with their neighbour ASes. This chapter steps further by presenting my study on the routing properties of BGP-M using traceroute probing on RIPE Atlas. The studied routing properties include load balancing schemes and routing delays. The analysis aims to provide insights in the performance of BGP-M in load balancing.

Note that I have described the challenges to measure BGP-M with traceroute data in Section 4.2. The challenges exist in terms of discovering *unknown* BGP-M deployments and BGP-M cases. Relying on the queries to LG servers, I have obtained the *static* knowledge about 1,674 BGP-M deployments and over 12k BGP-M cases, including the owner (*NearAS*), the location (*NearBR*), the neighbour AS (*FarAS*), the destination (*DstPrfx*) and the border links, etc. Thus, I can send traceroute probings to the *DstPrfxes* in the *known* BGP-M cases and study their *dynamic* routing properties. To be more specific, the traceroute measurements and analysis in this chapter can help to confirm whether these identified BGP-M cases can reveal the expected properties of BGP-M in terms of load balancing and routing delays.

6.1 My Traceroute Probing on RIPE Atlas

As introduced in Section 3.2, there are a number of projects for traceroute measurements, including RIPE Atlas [157], CAIDA Ark [5] and iPlane [131]. Among these projects, I used RIPE Atlas as the platform for my work because RIPE Atlas had publicly accessible traceroute probes in 5 of the 12 ASes where I identified BGP-M cases. These 5 ASes are HE (AS6939), VimpelCom (AS3216), RASCOM (AS20764), ISZT (AS12303) and BetterBe (AS48972), which had 3, 3, 4, 2 and 1 RIPE Atlas probes, respectively.

For each BGP-M case identified as <*NearAS, NearBR, FarAS, DstPrfx*>, I sent traceroute probings from available RIPE Atlas probes in the *NearAS* to all IP addresses from `X.Y.Z.1` to `X.Y.Z.254` of *DstPrfx* on IPv4, or the 254 IP addresses from `X:Y:Z::1` to `X:Y:Z::fe` of *DstPrfx* on IPv6. Each *DstPrfx* was probed once with ICMP packets and once with UDP packets, aiming to filter out the cases whose *NearBR* and border links were not observed in the traceroute paths. ICMP packets represent the traffic with control message in the Internet and UDP packets represent the normal traffic. I also tried TCP packets but they all failed in revealing *NearBRs* and border links. The other settings remained as default on RIPE Atlas, such as Paris traceroute variation 16 [45] and 3 packets for probing to each *DstIP*. Note that I only probed the prefixes used in the identification of each BGP-M deployment, because it is difficult to probe all the prefixes in the over 12k BGP-M cases. Thus, the total number of BGP-M cases studied in this chapter is 1,423.

For each BGP-M case, I check whether the traceroute paths sent from a probe to IP addresses in the *DstPrfx* actually traverse the *NearBR* where the BGP-M case is deployed. If the traceroute paths traverse elsewhere, the paths are discarded. Below is the procedure used to process each traceroute path.

(1) Obtain the list of ending points of border links, i.e. *FarIPs*, which are given as the 'Next Hop IPs' in the routing table returned by the `routes` command (see Figure 4.2).

(2) For each traceroute path, check if any *FarIP* appears in the traceroute path. If

yes, go to (3); otherwise, discard this traceroute path.

(3) Use the DNS Chain service provided by RIPEstat Data API [31] to obtain the router name of the predecessor IP address of the *FarIP* by using the link of `https://stat.ripe.net/data/dns-chain/data.json?resource=<IPaddress>`. If the router is *NearBR*, this IP is labelled as *NearIP* and the process for this traceroute path finishes; otherwise, discard this traceroute path.

For most BGP-M cases, when the *FarIPs* in a BGP-M case are observed, the *NearBR* in this case is traversed, and Step (3) can be ignored. However, Step (3) is still necessary because different *NearBRs* can be deployed with BGP-M to the same neighbour AS via the same *FarIPs*. I have observed such BGP-M cases. The two cases deployed by BetterBe (AS48972) on IPv4 share the same *FarIPs*. Specifically, these cases are <BetterBe, `BGP01`, Previder, 84.241.176.0/24> and <BetterBe, `BGP02`, Previder, 84.241.176.0/24> (Previder is AS20847). The shared *FarIPs* are 193.108.98.241 and 193.108.98.245. This suggests that the two *NearBRs* are both connected to the *FarBR* at the same IXP via layer-2 switching devices. Traceroute paths to the *DstPrfx* (84.241.176.0/24) traverse the *FarIPs*. The predecessor IP of the *FarIPs* is 95.130.232.2. I used Step (3) and learned that the router name of 95.130.232.2 is `bgp01.as48972.net`, suggesting that `BGP01` is traversed by the traceroute paths. Such that, the *NearBR* is located and the *NearIP* is 95.130.232.2.

In this study, I set a standard for traceroute measurement. That is, I will only consider traceroute measurement of a BGP-M case if I am able to obtain traceroute data to at least 250 of the 254 IP addresses in the destination prefix and they traverse the relevant *NearBR* and *BLs*. BGP-M cases whose traceroute paths do not fulfill this standard will not be studied for the purpose of reliability.

From the traceroute measurement, I obtained results for 89 cases that fulfill the standard. I will carry out more specific traceroute measurements to study the load balancing schemes and the routing delays on border links in BGP-M cases. The following sections will provided detailed analysis with specific case studies.

6.2 BGP-M Load Balancing Schemes

BGP-M uses multiple border links to balance the traffic to a destination prefix. This section investigates the load balancing schemes used by BGP-M with the traceroute data.

In the area of multipath routing, multiple load balancing schemes have been observed, such as per-flow, per-destination, per-packet and per-application [41]. The configuration documentations in Cisco introduced the scheme of per-session load balancing [8]. These five schemes have different features.

- Per-packet load balancing routes each packet via a path. Different packets are routed via different paths.
- Per-destination load balancing routes the packets to the same destination IP via the same path.
- Per-session load balancing routes the packets for the same pair of source IP address and destination IP address via the same path.
- Per-application load balancing routes the packets with the same transport port numbers via the same path.
- Per-flow load balancing routes the packets with the same flow identifier via the same path. A flow identifier includes the source and destination IP addresses, source and destination ports and protocol.

6.2.1 Schemes Supported by Router Vendors

Different router vendors have different settings and support different algorithms in their routers to achieve different kinds of load balancing.

6.2.1.1 Cisco Routers

Cisco routers support four algorithms, i.e. universal algorithm, include-ports algorithm, tunnel algorithm and original algorithm. Network operators can set their routers with different algorithms according to their needs.

- Universal algorithm is the default algorithm on Cisco routers. It uses hash function and achieves per-session load balancing.
- Include-ports algorithm achieves per-flow load balancing by using the source and destination ports as part of the load-balancing decision.
- Tunnel algorithm achieves per-packet load balancing using the round-robin method.
- Original algorithm produces distortions in load balancing across multiple routers. Original algorithm is not recommended for usage.

Most BGP-M deployments identified in Chapter 5 involve Cisco routers, including all BGP-M deployments by HE. Moreover, all the cases fulfilling the standard are deployed on Cisco routers.

6.2.1.2 Juniper Routers

By default, Juniper routers use hash algorithm to achieve per-flow load balancing, with the algorithm only including Layer 3 information. Network operators can set the hash algorithm to include both Layer 3 and Layer 4 information. Juniper routers handle ICMP packets differently from other packets, because checksum field in ICMP message makes each ping packet a separate "flow" [17].

Aside from per-flow load balancing, Juniper routers also support per-prefix load balancing and per-packet load balancing [19]. Per-prefix load balancing routes the packets to the same destination prefix via the same path. Per-packet load balancing can cause packet reordering and is therefore recommended only if the applications absorb reordering.

I observe that a small number of BGP-M deployments identified in Chapter 5 involve Juniper routers, such as 1 BGP-M deployment by RASCOM (AS20764) on IPv4 and 2 deployments by AS-T2012 (AS8647) on IPv6.

6.2.1.3 Huawei Routers

Huawei routers by default use hash algorithm and consider the 5-tuple information (i.e. source IP, destination IP, protocol, source port and destination port) for per-flow

load balancing [14].

Huawei routers also support network operators to configure per-packet load balancing with either round-robin mode or random mode.

I do not observe Huawei router involved in any BGP-M deployment in this study.

6.2.2 Case Study: <HE, tyo1, NII, 160.18.2.0/24>

This subsection studies a BGP-M case to illustrate the load balancing used in BGP-M. This case is <HE, tyo1, NII, 160.18.2.0/24>. NII is short for National Institute of Information (AS2907).

For this case, I sent traceroute probes with UDP packets and ICMP packets from two source IPs (i.e. two RIPE Atlas probes located inside HE) at three time points to each of the 254 destination IPs. Specifically, at 10:00am (GMT) and 10:15am (GMT), I sent traceroute probes from SrcIP-1 with UDP packets, to study the load balancing scheme for UDP packets at different time points. At 10:30am (GMT), I sent traceroute probes to each of the 254 destination IPs from SrcIP-1 and SrcIP-2 with ICMP packets, to study the load balancing scheme for ICMP packets from different source IPs.

Figure 6.1 shows the topology map of this case, extracted from the measurement. In this case, the *NearAS* is connected to the *FarAS* via two border links at the *NearBR*. Traffic from the two sources arrived at the *NearBR* (called tyo1) at two different ingress interfaces, i.e. NearIP-1 and NearIP-2. Traffic from each source was shared on the two border links, i.e. BL-1 and BL-2. According to the IXP data from PeeringDB [25], FarIP-1 and FarIP-2 belong to JPIX TOKYO and JPNAP Tokyo, respectively.

Figures 6.2(a) and 6.2(b) show the routing maps based on UDP packets. Figures 6.3(a) and 6.3(b) show the routing maps based on ICMP packets. The routing maps illustrate how the traffic is allocated on the border links. Each routing map shows the SrcIP, NearIP, FarIPs, IPs within FarAS and DstIPs. The DstIPs are listed in increasing order from X.Y.Z.1 to X.Y.Z.254. The blue lines represent the traffic allocated on BL-1 (denoted by FarIP-1), and the red lines represent the traffic allo-

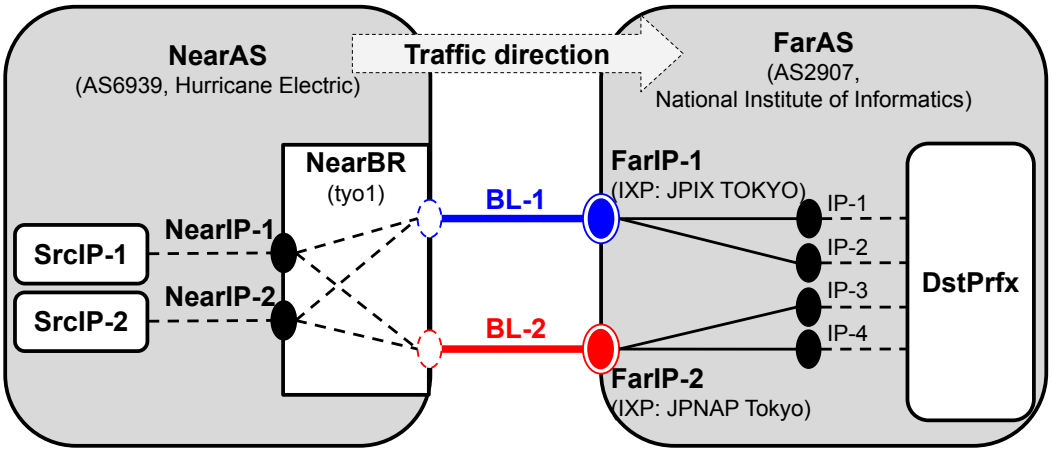


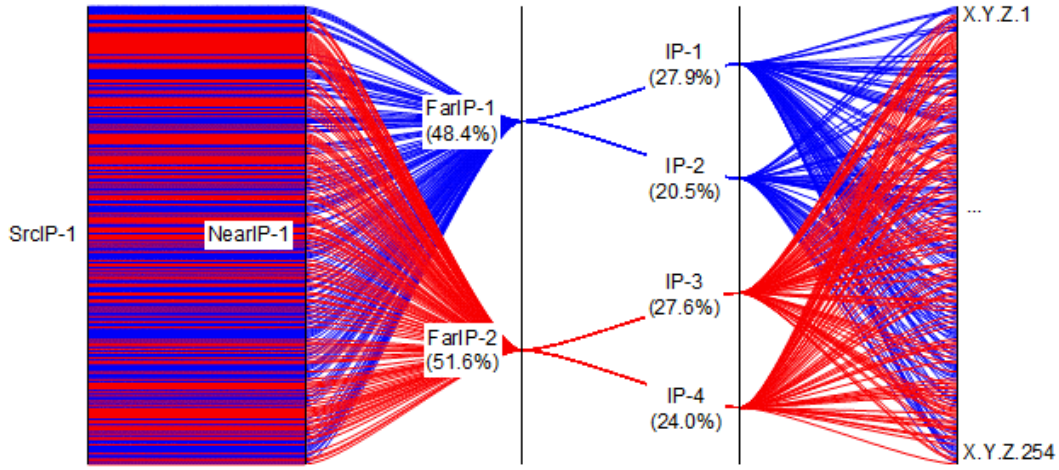
Figure 6.1: Topology map for BGP-M case <HE, `tyo1`, NII, 160.18.2.0/24>.

cated on BL-2 (denoted by FarIP-2). A percentage represent the portion of DstIPs, the traffic to whom is allocated on a link. For example, in Figure 6.2(a), 48.4% under FarIP-1 represents that the traffic to 48.4% of the DstIPs are allocated on BL-1, and 27.9% under IP-1 represents that the traffic to 27.9% of the DstIPs are allocated on BL-1. Here are some observations.

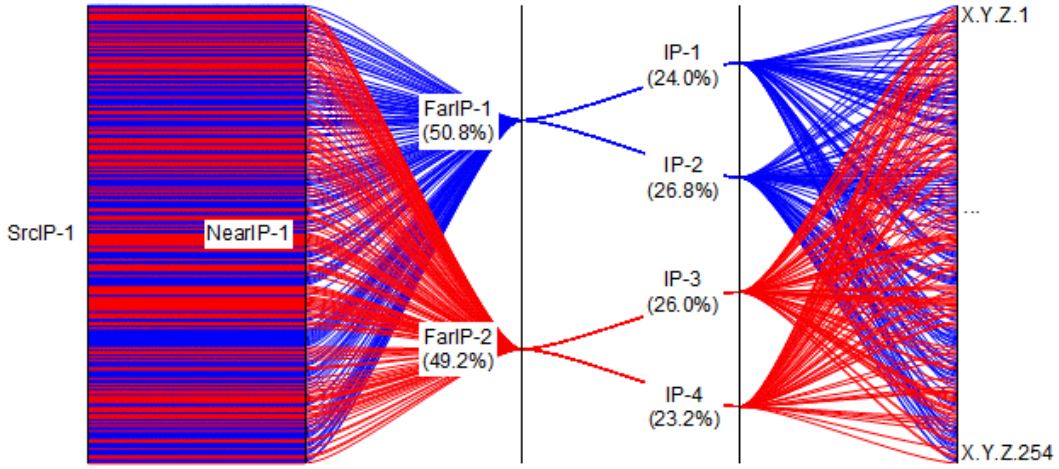
Firstly, all the four routing maps show that packets to the IP addresses in the destination prefix are always equally shared on the two border links, which, as expected, shows BGP-M provides load balancing at the level of destination prefix.

Secondly, on the routing maps in Figures 6.2(a) and 6.2(b) based on UDP packets, the packets to different destination IPs are randomly allocated on the two border links, and the allocations vary at different time points. This is the feature of load balancing based on the include-ports algorithm [9], which considers IP addresses and port numbers of source and destination. While hash function is sensitive to any change of bits in the identifiers, the UDP packets sent at different time points have not only different destination IPs, but also different port numbers.

Thirdly, on the routing maps in Figures 6.3(a) and 6.3(b), the ICMP packets are allocated on the two border links in a regular way: packets to 4 consecutive destination IPs are allocated on one border link, and the next 4 on the other border link; then the pattern repeats alternately. This suggests (1) the Cisco router is configured to conduct per-session load balancing for ICMP traffic using the so-called universal



(a) Routing map from SrcIP-1 with UDP at **10:00am (GMT)**

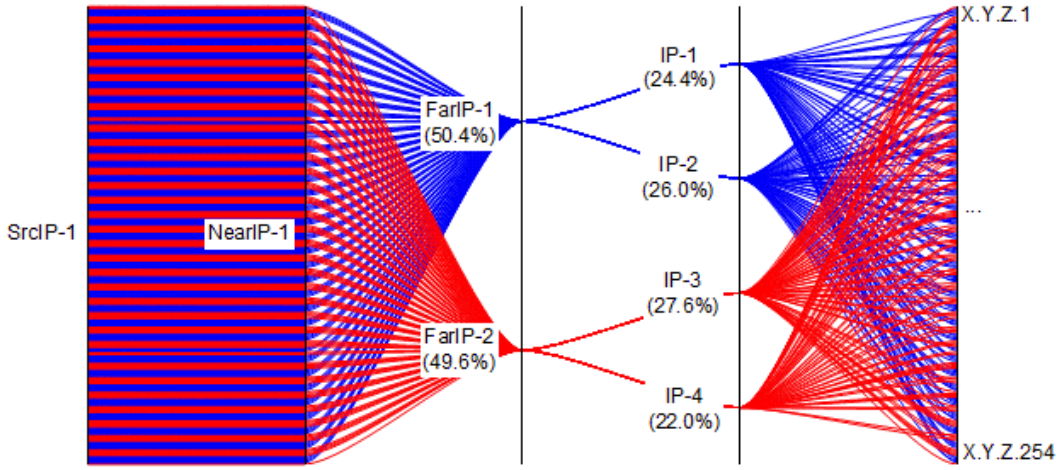


(b) Routing map from SrcIP-1 with UDP at **10:15am (GMT)**

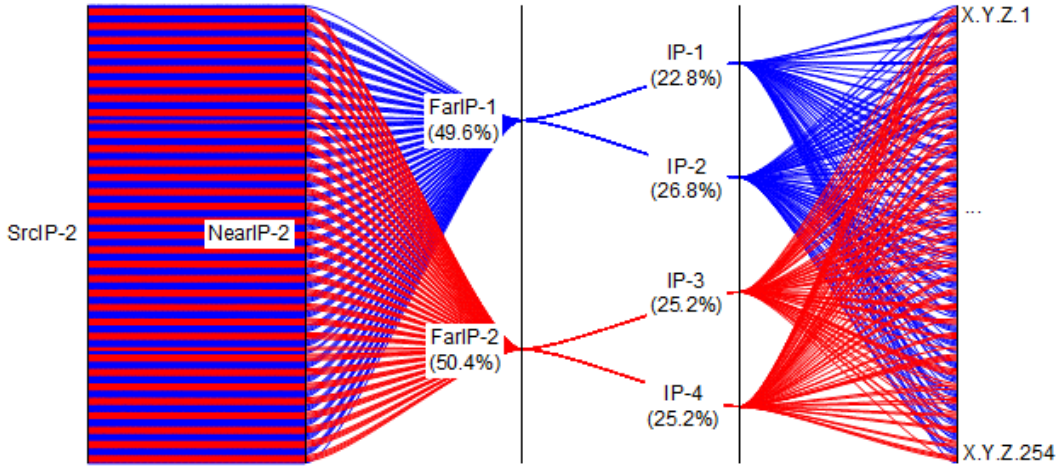
Figure 6.2: Routing maps for BGP-M case <HE, `tyo1`, NII, 160.18.2.0/24>. The routing maps are probed from the same source (SrcIP-1 at 209.51.186.5) using UDP packets at different times (i.e. Time Point 1 and Time Point 2).

algorithm [9] which considers only source and destination addresses; and (2) only a part of the destination IP address is considered [41].

Closer inspection reveals that the BGP-M allocation patterns in the two routing maps are exactly opposite to each other, i.e. destination IPs allocated to BL-1 in Figure 6.3(a) are allocated to BL-2 in Figure 6.3(b), and vice versa. This is because the routing maps are based on packets sent from different source addresses. Indeed, due to the universal algorithm, there are only two possible allocation patterns for ICMP packets from any sources to IP addresses in a destination prefix.



(a) Routing map from **SrcIP-1** with ICMP at 10:30am (GMT)



(b) Routing map from **SrcIP-2** with ICMP at 10:30am (GMT)

Figure 6.3: Routing maps for BGP-M case <HE, `tyo1`, NII, 160.18.2.0/24>. The routing maps are probed from different sources (SrcIP-1 at 209.51.186.5 and SrcIP-2 at 65.19.151.10) using ICMP packets at the same time.

These observations reveal the expected properties of BGP-M implemented on Cisco routers. For example, Cisco routers implement load balancing for UDP and ICMP traffic differently. Since most real traffic flows are TCP or UDP, traceroute measurements should be conducted with UDP packets for the true picture of BGP-M load balancing. Note that because all the cases fulfilling the standard are deployed at Cisco routers, these observations can be generalised to the BGP-M implementation on Cisco routers.

6.3 Delays on Border Links

To understand the effectiveness and performance of load balancing by BGP-M, here I study traffic delay on BGP-M border links based on traceroute measurements using ICMP and UDP packets.

6.3.1 Round Trip Time (RTT)

Round Trip Time (RTT) is the time a signal is sent from the source to the destination plus the time an acknowledgement signal from the destination is received by the source. An RTT value is the round trip transmission time. RTT values are provided by ping command and embedded to traceroute, thus they can help diagnose network problems. For example, a sudden increase of RTT values at an IP hop and the successor hops often suggests the occurrence of network congestion.

The delay between two hosts have four major sources: processing delay, queuing delay, serialisation delay and transmission delay [23, 140]. Processing delay is the time spent at a router or switch to process the packet header. Queuing delay is the time spent by a data packet to wait in the queue to be transmitted. Serialisation delay is the time spent to transmit a data packet onto the link. Transmission delay is the time spent being transmitted on the transmission media.

Processing delay and serialisation delay can be ignored because of the existing high-speed routers and high-bandwidth links. Queuing delay is related to the traffic volume to be handled by a router and this can cause a network congestion and severe delays. Transmission delay is related to the length of a link, or the geographical distance between two routers.

Network delays based on RTT values can be affected by other factors. For example, multipath routing can cause different paths being traversed by traceroute paths and varying RTT values for the same hop [46]. ICMP packets with low priority at routers cause larger RTT values than other packets.

RTT has been widely used to study the Internet routing delay and congestion. The first kind of research is to study the network delays in Colocation Facilities [116] or in a region like Africa [76]. Second, the changes of RTTs can reveal the root cause of path changes [108], or correlate with the BGP routing

changes [155, 156] and topological changes [73]. Third, RTT values are helpful for geolocation-related studies [63, 117], because higher RTT values often correlate with long-distance packets transmission. Fourth, RTT values can not only be predicted with various systems [140], but also be used to predict the routing dynamics [179]. Fifth, the time sequence (or time series) of RTT has been examined with different models [68, 74, 127] to infer the inter-domain congestion.

6.3.2 Calculation of Link Delay

From each traceroute path to a destination IP, I obtained the RTT value at each IP hop, which is the median value of the RTT values provided in the traceroute path. Then I calculated the *delay* on a border link, which is the difference between the RTT values of *NearIP* and *FarIP* of the border link. Note that the calculated delay is still round trip delay. Then I plot the delay distributions from various perspectives and focus on the 25th, 50th and 75th percentiles in the distributions for reliability.

6.3.3 Case Study: <HE, `hkg1`, Akamai, 23.67.36.0/24>

This subsection takes the BGP-M case of <HE, `hkg1`, Akamai, 23.67.36.0/24> to study the routing delays of BGP-M. I sent traceroute with ICMP packets and UDP packets from a RIPE Atlas probe to each of 254 IP addresses in the destination prefix at 15-minute intervals for 3 days from 00:00 (GMT+08:00, local time in Hong Kong) on 16/June/2021. For each destination IP, the traceroute probe with UDP packet is sent one-minute after the traceroute probe with ICMP packet, aiming to make the two types of traffic as close to each other as possible and avoid interference. I used default RIPE Atlas traceroute settings.

As shown by the topology map in Figure 6.4, this BGP-M case is deployed by HE at its border router `core1.hkg1.he.net` (`hkg1`, with *NearIP* 184.105.64.129) to the *DstPrfx* of 23.67.36.0/24 in a neighbour AS called Akamai (AS20940). Two *FarIPs* are observed in this case, which are 103.247.139.17 (i.e. FarIP-1 for BL-1) with bandwidth of 10G and 123.255.91.169 (i.e. FarIP-2 for BL-2) with bandwidth of 100G. FarIP-1 belongs to the IXP of AMS-IX Hong Kong, and FarIP-2 belongs to the IXP of HKIX. The bandwidths of the border links are

obtained from the public peering data for Akamai (AS20940) at PeeringDB [25] updated on 10/July/2021.

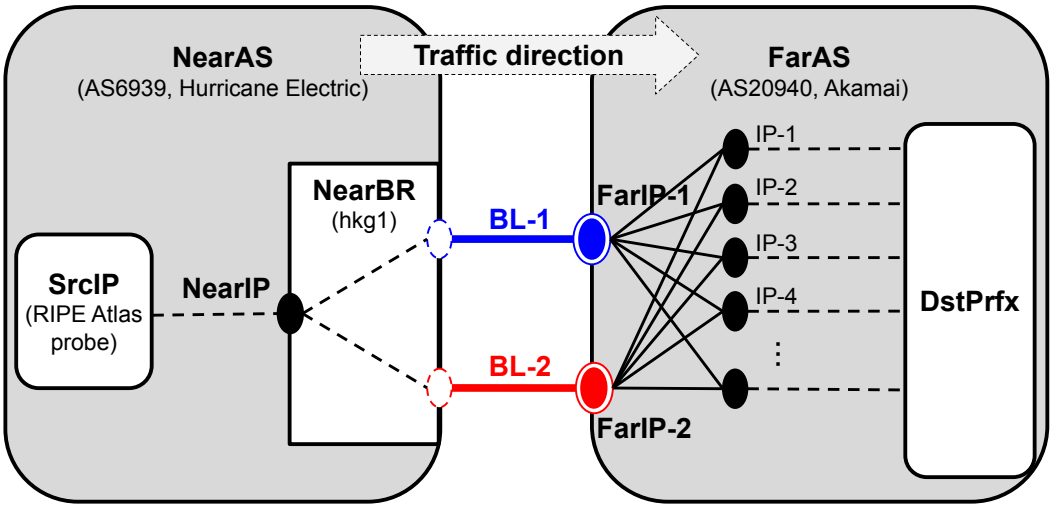


Figure 6.4: Topology map for BGP-M case <HE, `hkg1`, Akamai, 23.67.36.0/24>. FarIP-1 belongs to the IXP of AMS-IX Hong Kong, and FarIP-2 belongs to the IXP of HKIX.

In this case, HE and Akamai are connected with each other via IXPs located in Hong Kong. Both *NearBR* and *FarBR*(s) are located in Hong Kong. Although I do not have the specific locations of these routers, the geographical distance between two neighbour routers is estimated to be less than 30 km, according to the information of HE's PoPs [15] and Akamai's peering facilities published on PeeringDB [25]. Thus, the transmission delay is negligible, and the link delay now mainly measures the queuing delay at *NearBR*. Therefore, the link delay reflects the level of traffic congestion for each of the border links and can be considered as an indicator for routing performance.

As observed and explained in Section 6.2.2, when using ICMP packets, traceroutes for the same pair of source and destination IPs at different time points always go through the same border link due to per-session load balancing considering only source and destination addresses; whereas for UDP packets, traceroutes for the same pair of source and destination IPs at different time points are allocated to any of the two border links randomly due to per-flow load balancing considering IP addresses and port numbers of source and destination.

6.3.3.1 Frequency Distribution of Link Delays

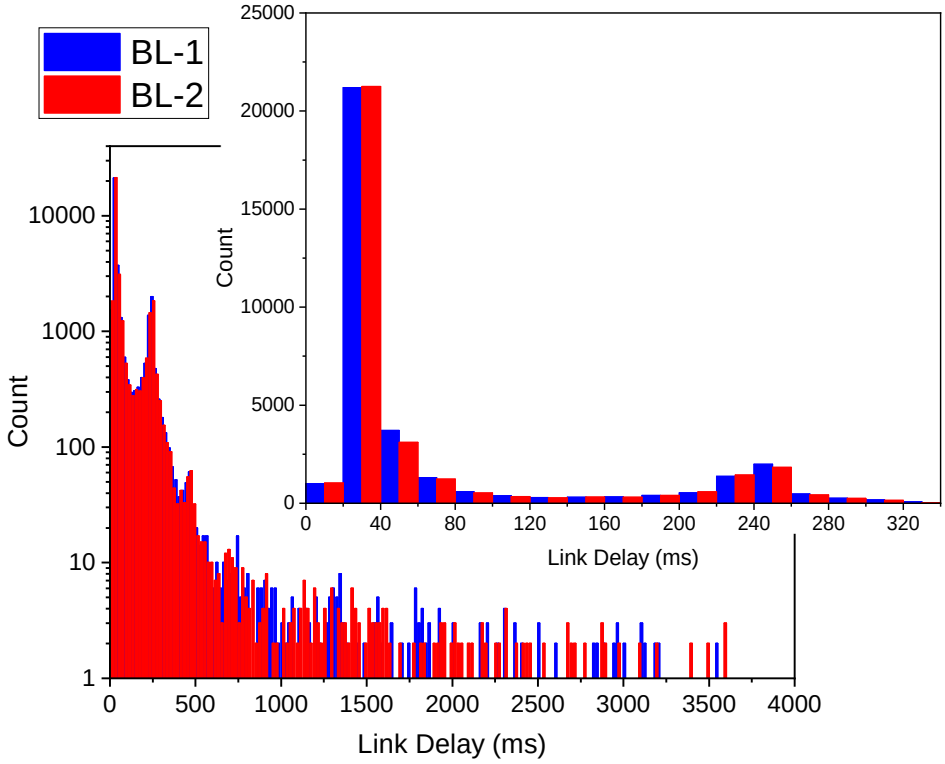
Figures 6.5(a) and 6.5(b) plot the delays on each border link at all the time points for all the *DstIPs* with ICMP packets and UDP packets, separately. The negative delay values account for less than 1.3% of all the delay values and they are neglected. A possible reason for the negative delays is the clocks on the *NearBR* and *FarBR* are not synchronised. Detailed reasons will be studied in the future.

Each figure is plotted with 20ms bin size for X-axis. The lower-left plot is on log scale for Y-axis. The upper-right inset plot is on linear scale for Y-axis and shows the delay values between 0ms and 340ms for X-axis that account for over 95.1% of all the values.

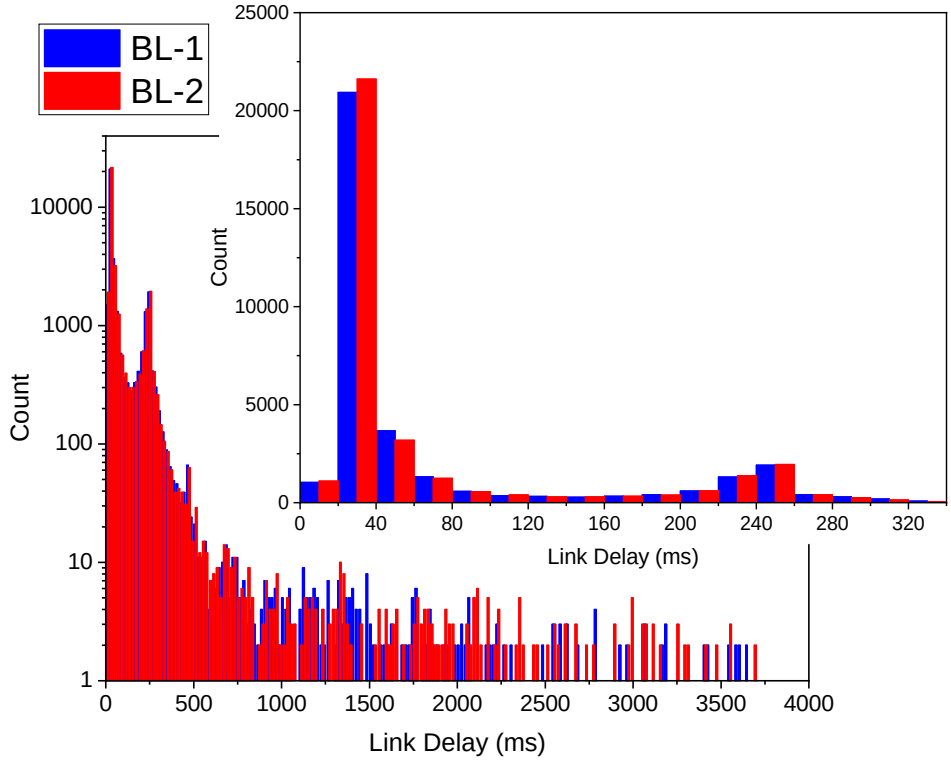
Figure 6.5 shows that both border links experienced stable and sound routing performance because the delay values are between (20ms, 40ms) at most of the times, accounting for 58.4%∼68.3% of the delays values for the border links with the two protocols. Although larger delays (those more than a few seconds) are observed, they are very rare. This indicates that transit on these border links were mostly free of congestion.

The congestion-free transit can be resulted by multiple factors. Firstly, it can be related to the deployment of BGP-M. BGP-M is designed to share the traffic on multiple border links and reduce the delays on single border link. We can imagine that if BGP-M was not deployed and all the traffic to the DstPrfx travelled on one single border link, more high-value delays would be observed, and this congestion-free transit might not even be observed.

Secondly, it is possibly related to the large bandwidths and relatively small traffic volumes on the two border links. The bandwidths of the two border links were: 10G for BL-1 (with FarIP-1 103.247.139.17) and 100G for BL-2 (with FarIP-2 123.255.91.169). The bandwidth information was obtained from PeeringDB, where the public peering data for Akamai (AS20940) was last updated on 10/July/2021. I also obtained from Akamai's technical report [28] that Akamai at Hong Kong (where the *FarBRs* were located) had average traffic volume of 21.9 Mbps and peak volume of 129.5 Mbps in Q1 2017. Although the report was four years ago, today's



(a) **ICMP:** Link delay distribution



(b) **UDP:** Link delay distribution

Figure 6.5: Distribution of delays on two border links of the BGP-M case <HE, `hkg1`, Akamai, 23.67.36.0/24> measured by traceroute in 3 days with 15-minute interval.

traffic volume is likely to remain well below the bandwidths of the border links.

6.3.3.2 Temporal Distribution of Link Delays

Figures 6.6(a) and 6.6(b) plot the 25th, 50th (i.e., median) and the 75th percentiles of delays in ascending order for traffic to *all* the *DstIPs* allocated on each border link over the 3-day measurement. Both figures use log-scale for Y-axis.

As can be seen, the two border links experienced similar trends of link delays for all the three curves with both ICMP packets and UDP packets. At most of the time, both links are very stable with the median values being around 30ms. Moreover, the 25th percentile curve is very close to the median curve. These observations, along with high density of delays in (20ms, 40ms), confirming the congestion-free transit. The vibration of the 75th percentile curve indicates the change of traffic volume for some *DstIPs* causing unstable queuing delays, corresponding to the delay values over 200ms in Figure 6.5.

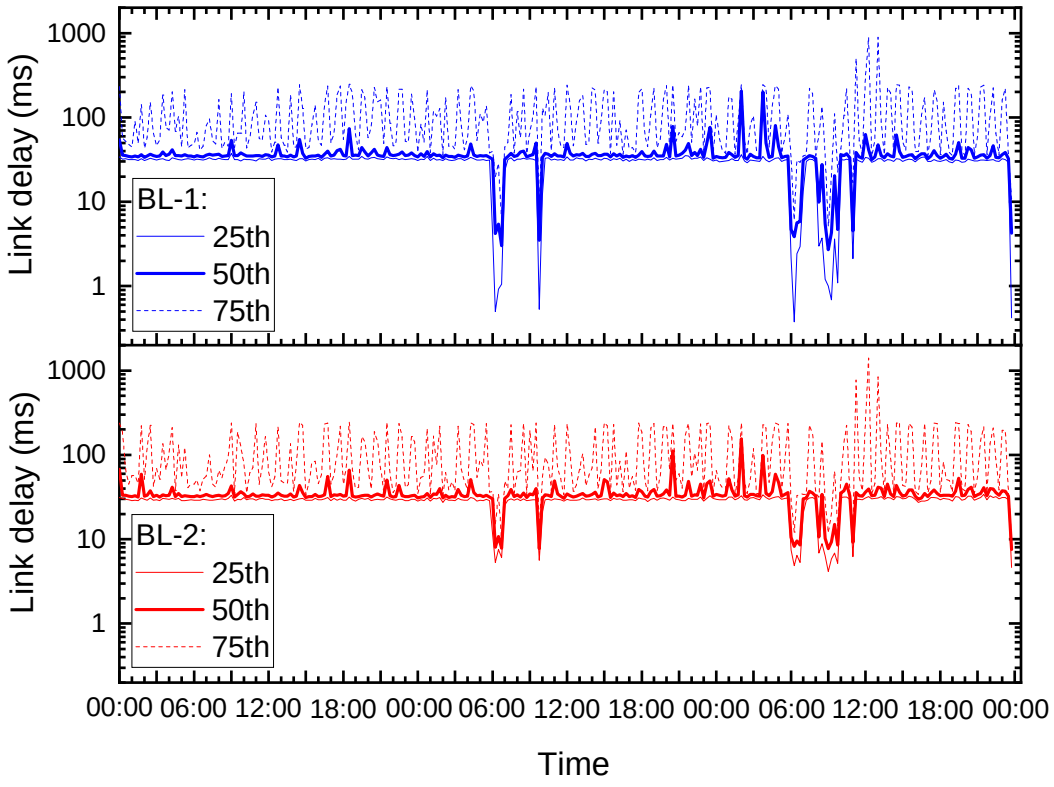
6.3.3.3 Distribution of Link Delays over Destination IPs

Figures 6.7(a) and 6.7(b) plot the 25th, 50th (i.e., median) and the 75th percentiles of delays in ascending order at *all* the time points on the border links allocated for traffic to each *DstIP*. The figures use log scale for Y-axis.

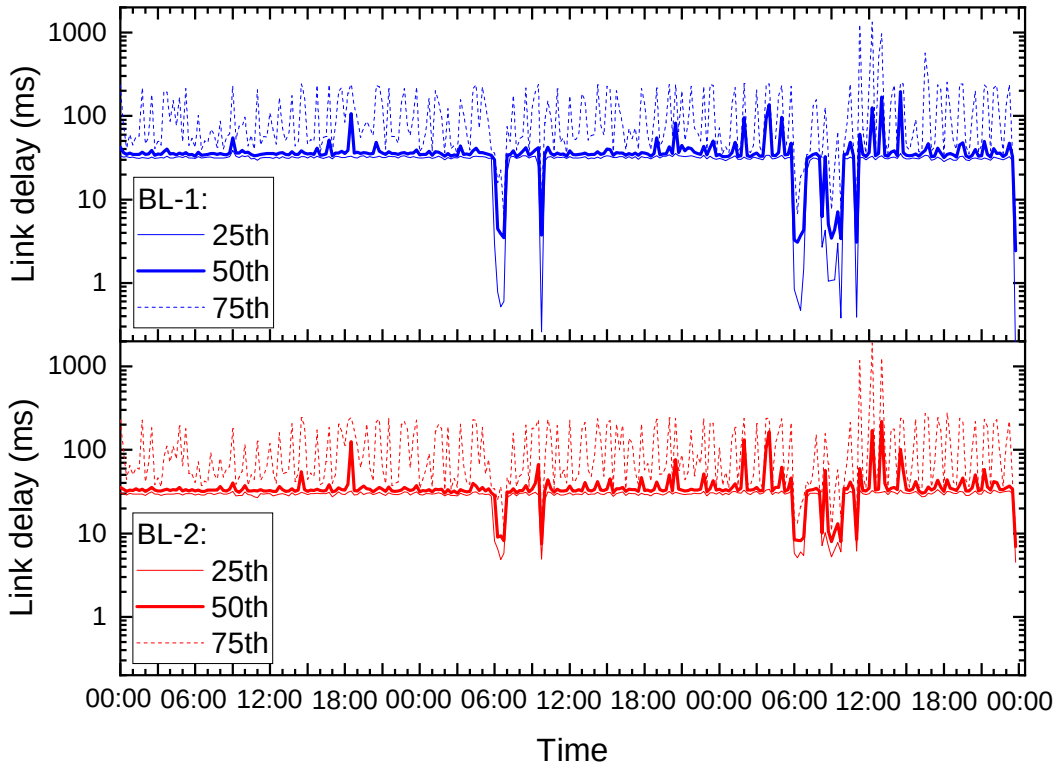
As previously explained, ICMP packets to destination IPs are equally allocated to the two border links in exactly the same way at every time point. This is confirmed by Figure 6.7(a). The statistics for BL-1 show link delays to only 128 destination IPs, each of which is calculated from 288 measurements (= 3 days $\times$ 24 hours $\times$ 4 times/h); whereas the statistics for BL-2 show link delays to 125 different destination IPs. Note that there was no border link observed for traffic to 23.67.36.1 during the measurement.

By comparison, UDP packets to destination IPs are equally, but randomly, allocated to the two border links, and allocation changes randomly at every time point. This is confirmed by Figure 6.7(b). The statistics for both BL-1 and BL-2 show link delays to all of 253 destination IPs, each of which is calculated from measurements at about 144 (half of the 288) time points.

Figure 6.7 shows that both links experienced similar delays in terms of *DstIPs*.

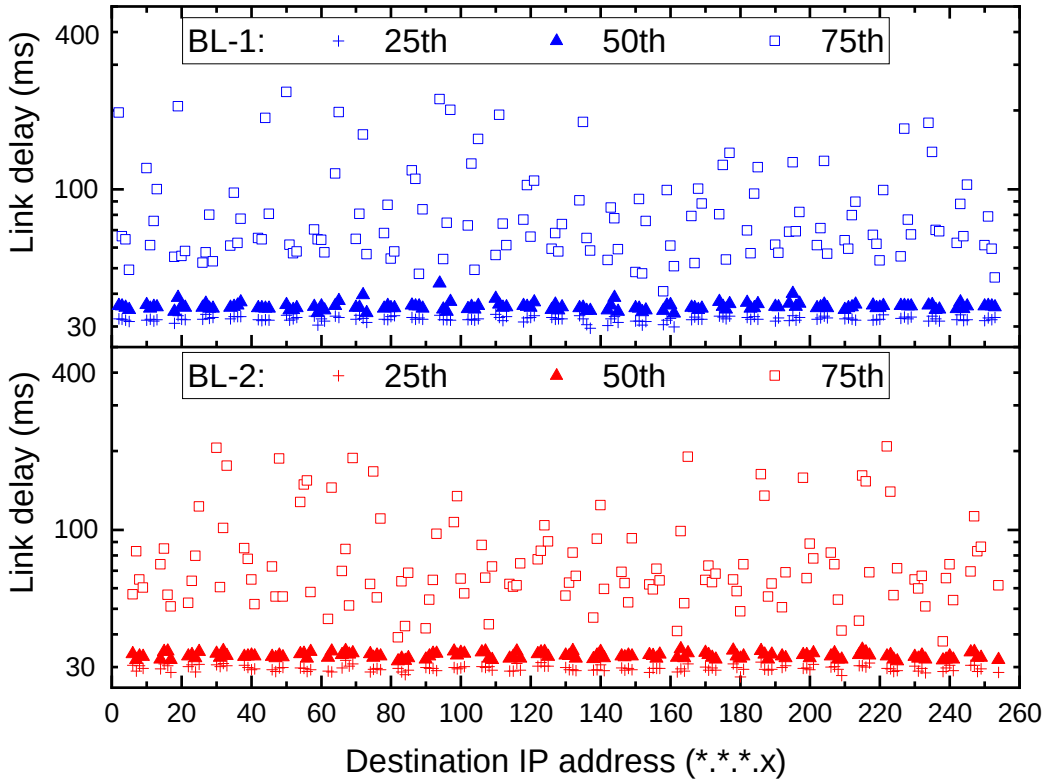


(a) **ICMP:** Link delay over 3 days

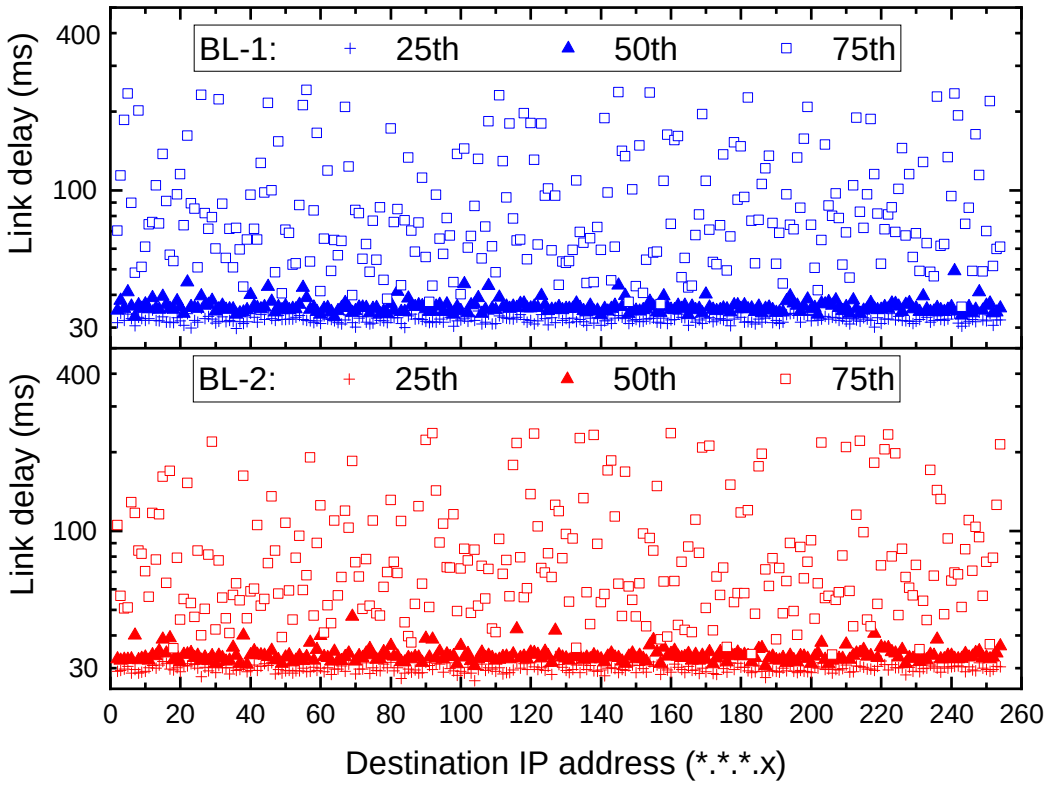


(b) **UDP:** Link delay over 3 days

Figure 6.6: Delay on two border links of the BGP-M case <HE, `hkg1`, Akamai, 23.67.36.0/24> measured by traceroute over the 3 days.



(a) **ICMP:** Link delay (on log scale) for each destination IP



(b) **UDP:** Link delay (on log scale) for each destination IP

Figure 6.7: Delay on two border links of the BGP-M case <HE, `hkg1`, Akamai, 23.67.36.0/24> measured by traceroute for each *DstIP* during the 3 days.

Again, the 25th percentile plot and the median plot are very close to each other while the 75th percentile curve is very unstable. These confirm the findings in Figure 6.6. Moreover, UDP packets experienced more severe vibrations than ICMP packets. This indicates that UDP packets and ICMP packets are handled differently. As UDP packet is more suitable than ICMP packet to imitate the real traffic, this observation may reveal the real picture about the delays on both border links.

These observations are as expected because it is exactly what BGP-M is designed to achieve as a load balancing technique, to fully utilise the available routing facility and capacity. Although this is only one case study, the observations can be generalised to the other BGP-M cases deployed on the routers by all the router vendors because load balancing with reduced routing delays is the basic function of BGP-M. It may also bring other benefits such as reduced congestion (if it happens); and diverse routes for increased flexibility and security.

6.4 Discussion

This chapter analysed BGP-M from the perspective of routing properties, i.e. whether BGP-M can provide expected routing performance. To achieve the purpose, this chapter described the traceroute measurement to the known BGP-M cases, in order to reveal the routing properties of BGP-M. The results indicate that BGP-M is able to provide load balancing for different types of traffic on Cisco routers, provide congestion-free transit, and reduce the traffic delay on border links. These are exactly what BGP-M is designed for. The study in this chapter is a complimentary study to the work in previous chapters in BGP-M with traceroute measurements.

The measurement and analysis in this chapter is limited in scale because only some case studies are provided. Traceorute measurements for more cases are needed to confirm the observations presented in this chapter and provide a generalised conclusion. Moreover, traceroute has its limits on revealing more routing performance compared to real-time traffic data, and this will be explored as a future direction of work.

Chapter 7

Conclusion

7.1 PhD Research Achievements

I have obtained several achievements on studying BGP-Multipath, including introducing novel measurement method and new datasets and revealing novel knowledge and insights in BGP-M. These achievements are valuable for the researchers and the network operators to better understand BGP-M and conduct further researches.

7.1.1 New Measurement Methods and Datasets

- I proposed a novel method to discover the deployment of BGP-M in the Internet relying on queries to LG servers. LG servers provide direct and reliable information on AS borders, resolving the problem of AS border mapping that traceroute data suffers from. The LG server data provides the actual configuration on border routers, and contains rich information on BGP-M deployment. The discovered BGP-M deployments and BGP-M cases are therefore all ground-truth. This method based on LG data answers the first question asked in Section 1.3.

- My work produced new datasets helpful to the research community for the study on BGP-M. The datasets include the LG output data from multiple ASes and the traceroute measurement data on RIPE Atlas. The data obtained from LG servers are available at GitHub [12]. The traceroute measurement IDs are also provided at GitHub [12] and readers can freely download the measurement data on RIPE Atlas through these measurement IDs.

7.1.2 Knowledge and Insights in BGP-M

My work has brought to the academia and the industry a rich set of novel knowledge and insights in BGP-M, and answered the second to the fourth questions asked in Section 1.3. These knowledge and insights are summarised and listed below.

- BGP-M has been deployed by 12 ASes in the Internet. These 12 ASes include both large transit ASes and stub ASes. They have deployed BGP-M with their neighbour ASes. Compared to the large number of ASes in the Internet, 12 ASes account for a small portion. There might be more ASes that have deployed BGP-M with their neighbour ASes.

- Hurricane Electric (HE), a large transit AS, has extensively deployed BGP-M with hundreds of its neighbour ASes covering different AS rank groups at its globally distributed border routers. HE has been also actively maintaining its deployment of BGP-M.

- IXPs play an important role in the deployment of BGP-M. It has been involved in around 90% of the BGP-M deployments deployed by HE.

- The neighbour ASes of HE can have different requirements for BGP-M. Those neighbour ASes which are content delivery are deployed with more BGP-M more than those neighbour ASes which are transit ASes.

- HE has deployed BGP-M in a flexible way to suit its needs. Different neighbour ASes can be deployed with BGP-M at different sets of border routers, and different border routers can be deployed with BGP-M to different sets of neighbour ASes.

- HE has deployed BGP-M on IPv6 Internet similarly to its deployment of BGP-M on IPv4 Internet. Similar observations are revealed from the results on IPv4 and IPv6 Internet regarding HE's deployment of BGP-M.

- A BGP deployment can be used for traffic routing to hundreds of destination prefixes in the farside AS. The top ten BGP-M deployments of HE are all used for traffic delivery to over 200 destination prefixes.

- Cisco routers are configured with with universal algorithm to achieve per-session load balancing for BGP-M by default. ICMP packets and UDP packets are handled differently for load balancing.

- BGP-M helps network operators to fully utilise the border links between ASes and achieve optimal routing performance as a load balancing technique.

7.2 Making a Case for BGP-M

7.2.1 Advantages and Benefits of BGP-M

As a load balancing technique, BGP-M provides balanced traffic and enhanced routing performance, and offers a number of unique advantages and benefits.

7.2.1.1 Wide Availability for Implementation

Both hardware and software requirements for BGP-M deployment are already widely available in the Internet.

Firstly, there is a wide presence of multiple border links between ASes in the Internet, where more than one border links are connecting from a border router of an AS to border router(s) of a neighbour AS. Such multiple border links commonly exist, especially among core ASes or between core and peripheral ASes.

Secondly, most border routers provided by major router vendors, such as Cisco, Juniper and Huawei, already support BGP-M load balancing, which is an integral part of their design and function.

This means BGP-M can be readily implemented by network operators to their neighbour ASes without changing or upgrading their infrastructure or agreements.

7.2.1.2 Easy Implementation

The implementation of BGP-M is rather simple and straightforward. For example, the minimum action required on a Cisco border router is to activate BGP-M by setting a single parameter `maximum-paths` from its default value 1 to the number of (different) paths for a given *DstPrfx* [8]. There is literally no additional cost for BGP-M implementation.

7.2.1.3 Independent, Flexible and Transparent Deployment

Although the technique is called BGP-M and it follows BGP's best path selection process, network operators do not need to alter their BGP process to deploy BGP-M as the load balancing will still follow exactly the same AS-level path as before. As such, network operators can freely and independently implement or remove BGP-M without informing or obtaining new agreement from their neighbour ASes. Network operators can deploy, revise and cancel BGP-M for any selections of destination prefixes in any neighbour or remote ASes.

BGP-M deployed at an AS' border routers has no interference to any other multipath routing techniques implemented within or outside of the AS, or to any traffic engineering configurations elsewhere. For example, as shown by my analysis in Sections 5.2.1, there is no impact on BGP-M load balancing whether the border links connect to *FarAS* directly or via IXPs.

Basically, BGP-M deployed at a border router is transparent to other parties participating in the relevant traffic routing, which gives network operators flexibility and convenience.

7.2.1.4 Benefits of BGP-M Load Balancing

The benefits of load balancing gained from BGP-M deployment is no less than any other multipath routing techniques.

First of all, BGP-M is implemented at border routers on the BGP mechanism. This enables easy deployment of BGP-M between ASes. Moreover, border routers can react to network changes without affecting the BGP routing. For example, when a border link for BGP-M fails, the traffic to the destinations can still be traversed via the available border link(s). Thus, BGP-M helps improve the network resilience and security.

Secondly, BGP-M uses multiple border links for load balancing. This can increase the usage ratio of border links and reduce risk of congestion in face of traffic surges. It can also improve routing path diversity, which can be useful for network resilience and security. Moreover, border links between core ASes and their neighbour ASes are valuable resources which are common, already deployed. Many

border links have high bandwidths and might even cost much. The fuller utilisation of these border links by BGP-M can be an effective and economic option for ASes to improve their routing performance without extra investment in infrastructure.

Thirdly, a network operator benefits from the deployment of BGP-M regardless of whether or how many other networks have implemented the technique. The more deployment, the more benefit. And such benefits are likely to be mutually beneficial to not only the AS that deploys BGP-M but also its neighbour ASes. This is proved by my observations on the wide deployment of BGP-M between HE and its neighbour ASes (especially those hyper-giant ASes).

7.2.2 Awareness and Promotion of BGP-M

My research is of great value by promoting BGP-M to both the research community and the network operators.

Firstly, my research is the first systematic study on BGP-M and fills the gap in this area. I proposed a novel method to infer the BGP-M deployment and revealed rich and state-of-the-art knowledge on BGP-M. These new knowledge are helpful to the research community to understand BGP-M and multipath routing. I have presented my results in several academic conferences and published my results, and many researchers have shown their interests in this topic. Further researches on BGP-M are also of interest to the academia. For example, how to achieve large-scale measurement of BGP-M with traceroute data?

Secondly, my research has helped network operators become aware of this load balancing technique, which can potentially benefit their networks hugely. I have contacted with HE's network operators and technicians from AMS-IX. They have also expressed their interests in BGP-M deployment.

Thirdly, with more and more researchers, technicians and regulators become aware of BGP-M, the real achievement of my research in practice will depend on how the technique is promoted for adoption and deployment in more networks. Thanks to the wide range of benefits of BGP-M that my research has revealed, in particular its readiness to be deployed and effectiveness for load balancing, I am confident that network operators would be interested in deploying more BGP-M

following this research.

7.2.3 Potential for Future Deployment of BGP-M

Based on my research presented in previous chapters, I estimate that the actual scale of the deployment of BGP-M is very large and there is immense potential for wider range of deployment.

First of all, in my measurements shown in Chapters 5-6, I have shown the deployment of BGP-M by large transit ASes and stub ASes, on both IPv4 and IPv6. In particular, HE, a large transit AS, has extensively deployed BGP-M with its neighbour ASes at its globally distributed networks. This suggests the vast benefit and potential of BGP-M.

Moreover, as shown in Section 5.3.1, the extension of my measurement to all the /24 prefixes has revealed that each BGP-M deployment can be used for traffic routing to hundreds of destination prefixes in the farside AS. This suggests that when the queries are extended to all the prefixes announced by neighbour ASes, more BGP-M cases will be discovered.

Furthermore, prefixes announced by remote ASes are not queried yet. Querying these remote ASes will result in many more BGP-M cases and produce a rather complete picture about the deployment of BGP-M based on my method.

In addition, my method has studied a rather small number of ASes. The 2,709 ASes with LG servers only account for a small portion (less than 2.5%) of the 110,589 allocated ASNs (with both 16-bit and 32-bit) [1]. ASes with inaccessible LG servers should be studied. There should be more ASes that have deployed BGP-M.

Nevertheless, based on my data and analysis so far, I estimate that the scale of existing deployment of BGP-M is still far smaller than the intra-domain multipath routing, of which millions of cases [173] have been uncovered throughout the Internet. Besides, the wide availability of border links between ASes and the benefits of BGP-M indicate the possibility for large-scale deployment of BGP-M. Therefore, there is an immense scope for future deployment of BGP-M by more ASes to more destinations.

7.3 Future Works

Here I envisage two directions for future research works.

Fuller Measurement of BGP-M Deployment In this research, I used the method proposed in Section 4.3 to measure the deployment of BGP-M. The method relied on LG server data and produced reliable and ground-truth results. The measurement was not complete as analysed in Sections 4.4 and 5.5, but this research has achieved its goal by answering the questions asked in Section 1.3 and providing state-of-the-art knowledge on the deployment and properties of BGP-M.

Future researches may conduct a fuller measurement on the deployment of BGP-M. First of all, like in the extended measurement, all the prefixes announced by neighbour ASes should be queried. This will increase the number of BGP-M cases as shown in Section 5.3.1. Secondly, the prefixes announced by remote ASes should be queried, because the *DstPrfx* can be anywhere on the Internet for a BGP-M deployment. Considering the huge number of remote ASes, this will improve the completeness of the measurement to a higher level. Thirdly, more ASes with LG servers should be studied. The ASes studied in this research account for a small portion in the allocated ASNs in the Internet. It requires more effort to study those ASes with inaccessible LG servers.

When LG servers are still not sufficient, traceoute data can be used to measure the deployment of BGP-M. Recent works on multipath routing (e.g. [41, 173]) have shown the feasibility to send traceroute probings to multiple (even all) IP addresses in a prefix (say /24). This makes it possible to infer the deployment of BGP-M with large-scale traceroute measurement. Moreover, the high-frequency probing techniques (e.g. [51]) are potential to improve the efficiency of inference. The key problem is how to achieve accurate AS border mapping, which is still a challenge on Internet routing [184]. Furthermore, it is necessary to carefully design the inference algorithm to control the overhead.

Analysis on BGP-M Performance Using Real Traffic Data Chapter 6 has introduced my study and the results about the routing properties of BGP-M using traceroute measurement data. Traceroute data has its limitation on studying the

performance in two aspects. First, it is difficult to conduct long-lasting traceroute measurement to capture the performance difference before and after BGP-M is deployed, due to the difficulty to predict the activation and deactivation of BGP-M. Second, traceroute data can only reveal a part of the discovered BGP-M cases due to the limited resources in the existing traceroute platforms.

To cope with the above limitations of traceroute data, a future work is to use real traffic data to study the BGP-M performance. Real traffic data makes it possible to analyse the traffic before and after BGP-M is deployed, and to study to what extent BGP-M deployment can indeed improve the routing performance between two neighbouring ASes. Moreover, researchers can study more BGP-M cases, and investigate whether and how BGP-M benefits the Internet routing via the comparison between different BGP-M cases.

This might be challenging due to the difficulty for a third-party researcher to obtain the traffic data from network operators, because network operators are often reluctant to share their traffic data for privacy and security concerns. This is why I did not use real traffic data for my work. Therefore, it requires both the research community and the network operators to cooperate for a solution, which is potentially beneficial to network operators to improve their network routing performance.

Appendix A

ASes Studied in This Work

AS number	Organisation name	AS rank*	Customer cone size*	Country/Territory
1299	Telia Company AB	2	32,929	Sweden
6939	Hurricane Electric LLC	7	16.047	United States
6461	Zayo Bandwidth	9	9,175	United States
9002	RETN Limited	13	6,374	United Kingdom
4637	Telstra International Limited	15	4,548	Hong Kong
12389	PJSC Rostelecom	20	3,425	Russia
7922	Comcast Cable	24	2,820	United States
3216	PJSC Vimpelcom	25	2,777	Russia
9498	Bharti Aritel Limited	27	2,361	India
6830	Liberty Global B.V.	29	2,218	Netherlands
20764	CJSC RASCOM	30	2,073	Russia
52320	GlobeNet Cabos Submarinos Colombia, S.A.S.	32	2,005	Colombia
8359	MTS PJSC	36	1,810	Russia
4826	Vocus Communications	40	1,593	Australia
41095	IPTP LTD	48	1,190	United Kingdom
8220	COLT Technology Services Group Limited	51	1,083	United Kingdom
4230	CLARO S.A.	57	805	Brazil

5588	T-Mobile Czech Republic a.s.	65	686	Czechia
3303	Swisscom (Schweiz) AG	69	647	Switzerland
22822	Limelight Networks, Inc.	344	101	United States
2907	National Institute of Informatics	356	96	Japan
10310	Yahoo!	747	41	United States
19752	Hydro One Telecom Inc.	880	33	Canada
24971	Master Internet s.r.o.	1195	22	Czechia
8647	LLC TELEMIST 2012	1264	20	Ukraine
15169	Google LLC	1743	13	United States
13335	Cloudflare, Inc.	1845	12	United States
42	WoodyNet	1931	11	United States
20940	Akamai International B.V.	1998	11	Netherlands
8075	Microsoft Corporation	2288	9	United States
22691	ISPnet, Inc.	2337	9	United States
44679	INVITE Systems SRL	2885	7	Romania
20847	Previder B.V.	3123	6	Netherlands
15133	MCI Communications Services, Inc. d/b/a Verizon Business	3172	6	United States
16509	Amazon.com	3560	5	United States
52201	OOO Suntel	3788	5	Russia
12303	Council of Hungarian Internet Providers	4104	4	Hungary
13414	Twitter Inc.	4119	4	United States
328112	Linux Based Systems Design SA (Pty) Ltd	6339	3	South Africa
714	Apple Inc.	6385	2	United States
26667	The Rubicon Project, Inc.	22076	1	United States
46489	Twitch Interactive Inc.	33522	1	United States
48564	IP Vision A/S	34831	1	Denmark
48972	BetterBe B.V.	35096	1	Netherlands

54113	Fastly	38523	1	United States
131713	PT Sano Komunikasi	45081	1	Indonesia
196965	TechCom	–	–	–

* The information on AS rank and customer cone size are provided by CAIDA AS Rank data [4].

Appendix B

Invited Talks and Publications

- Jie Li. Group of Border Links (GBL) Used in Internet Multipath Routing (Invited talk). RIPE Meeting 79. Rotterdam, the Netherlands, October 2019.
- Jie Li and Vasileios Giotsas and Shi Zhou, "Anatomy of Multipath BGP Deployment in a Large ISP Network," in Proceedings of 4th Network Traffic Measurement and Analysis Conference (TMA Conference), 2020, 9 pages, http://arxiv.org/abs/2012.07730.
- Jie Li and Shi Zhou and Vasileios Giotsas, "Performance Analysis of Multipath BGP," in Proceedings of IEEE International Conference on Computer Communications Workshops (INFOCOM WKSHPS): Global Internet, 2021, 6 pages, https://arxiv.org/abs/2103.07683.
- Jie Li and Vasileios Giotsas and Yangyang Wang and Shi Zhou, "BGP-M Routing in the Internet,"Journal article (under review), 15 pages, https://arxiv.org/abs/2107.10938.

Bibliography

[1] The 32-bit AS number report.
http://www.potaroo.net/tools/asn32/ (August 2020)

[2] BGP Looking Glass Databases.
http://www.bgplookingglass.com/ (January 2020)

[3] The CAIDA AS Relationships Dataset.
https://www.caida.org/data/as-relationships/

[4] CAIDA AS Rank.
http://as-rank.caida.org/ (January 2020)

[5] CAIDA: Archipelago (Ark) Measurement Infrastructure.
http://www.caida.org/projects/ark/ (December 2018)

[6] The CAIDA UCSD AS to Organization Mapping Dataset. <20200101>
http://www.caida.org/data/as_organizations.xml

[7] The CAIDA UCSD IXPs Dataset.
http://www.caida.org/data/ixps.xml

[8] CISCO, "IP Routing: BGP Configuration Guide - BGP Multipath Load Sharing for Both eBGP and iBGP in an MPLS-VPN",
https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/iproute_bgp/configuration/xe-16/irg-xe-16-book/bgp-multipath-load-sharing-for-both-ebgp-and-ibgp-in.html

[9] CISCO, “IP Switching Cisco Express Forwarding Configuration Guide, Cisco IOS XE Release 3S”,
https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/ipswitch_cef/configuration/xe-3s/isw-cef-xe-3s-book/isw-cef-load-balancing.html#GUID-D8A86BB9-FCA8-48CA-881D-153F4383728D

[10] Data Overview - CAIDA.
https://www.caida.org/data/

[11] European Internet Exchange Association.
https://www.euro-ix.net/

[12] GitHub-jieliucl/BGP-M.
https://github.com/jieliucl/BGP-M

[13] Huawei, “Example for Configuring BGP Load Balancing,” Configuration Guide - IP Unicast Routing,
https://support.huawei.com/enterprise/en/doc/EDOC1000178324/fd6029a9/example-for-configuring-bgp-load-balancing

[14] Huawei, “Load Balancing Hash Algorithms,”
https://support.huawei.com/enterprise/en/doc/EDOC1100086965

[15] IP Transit - Hurricane Electric Internet Services.
https://www.he.net/ip_transit.html

[16] IP2Location Lite.
https://lite.ip2location.com/

[17] Juniper Networks, “Configuring Per-Packet Load Balancing”,
https://www.juniper.net/documentation/us/en/software/junos/sampling-forwarding-monitoring/

topics/concept/policy-configuring-per-packet-load-balancing.html

[18] Juniper Networks, "Examples: Configuring BGP Multipath - TechLibrary", https://www.juniper.net/documentation/en_US/junos/topics/topic-map/bgp-multipath.html

[19] Juniper Networks, "Understanding the Algorithm Used to Load Balance Traffic on MX Series Routers — Traffic Sampling, Forwarding, and Monitoring User Guide", https://www.juniper.net/documentation/us/en/software/junos/sampling-forwarding-monitoring/topics/concept/hash-computation-mpcs-understanding.html

[20] Macroscopic Internet Topology Data Kit (ITDK). https://www.caida.org/data/internet-topology-data-kit/

[21] MaxMind: IP Geolocation and Online Fraud Prevention. https://www.maxmind.com

[22] NetAcuity. https://www.digitalelement.com/solutions/ip-location-targeting/netacuity/

[23] Network delay - Wikipedia. https://en.wikipedia.org/wiki/Network_delay

[24] Packet Clearing House - Internet Exchange Directory. https://www.pch.net/ixp/dir

[25] PeeringDB. https://www.peeringdb.com/

[26] PeeringDB API Documentation. https://www.peeringdb.com/apidocs/ (January 2020)

[27] PlanetLab: An open platform for developing, deploying, and accessing planetary-scale services.
`http://www.planet-lab.org/`

[28] Q1 2017 State of the Internet Security Report | Akamai.
`https://www.akamai.com/us/en/multimedia/documents/state-of-the-internet/q1-2017-state-of-the-internet-connectivity-report.pdf`

[29] RIPE IPmap: A Collaborative Approach to Mapping Internet Infrastructure.
`https://ipmap.ripe.net/` (March 2019).

[30] Ripe routing information service.
`http://www.ripe.net/ris`

[31] RIPEstat Data API.
`https://stat.ripe.net/docs/data_api#whois`

[32] Team Cymru.
`http://www.team-cymru.com`

[33] traceroute.org.
`http://traceroute.org/` (January 2021)

[34] University of Oregon Route Views Project.
`http://www.routeviews.org/` (January 2020)

[35] Vela MIDAR API - CAIDA.
`https://www.caida.org/projects/ark/vela/midar-api/`

[36] Ager B., Chatzis, N., Feldmann A., Sarrar N., Uhlig S., and Willinger W. Anatomy of a Large European IXP. In *ACM SIGCOMM'12*, pp. 163–174.

[37] Ahmad M. Z., and Guha R. Impact of Internet Exchange Points on Internet Topology Evolution. In *IEEE LCN'10*. pp. 332–335.

[38] Ahmad M. Z., and Guha R. A Tale of Nine Internet Exchange Points: Studying Path Latencies Through Major Regional IXPs. In *IEEE LCN'12*, pp. 618–625.

[39] Ahmed N., and Sarac K. An Experimental Study on Inter-domain Routing Dynamics Using IP-level Path Traces. In *IEEE ICN'15*, pp. 510—517.

[40] Al-Musawi B., Branch P., and Armitage G. BGP Anomaly Detection Techniques: A Survey. *IEEE Commun. Surveys Tuts.* 11, 1 (2017), 377–396.

[41] Almeida R., Cunha Í., Teixeira R., Veitch D., and Diot C. Classification of Load Balancing in the Internet. In *IEEE INFOCOM'20*, pp. 1987–1996.

[42] Almeida R., Morais O., Fazzion E., Guedes D., Meira Jr W., and Cunha Í. A Characterization of Load Balancing on the IPv6 Internet. In *PAM'17*, pp. 242–254.

[43] Anwar R., Niaz H., Choffnes D., Cunha I., Gill P., and Katz-Bassett E. Investigating Interdomain Routing Policies in the Wild. In *ACM IMC'15*, pp. 71–77.

[44] Araújo J. T., Rio M., and Pavlou G. Towards Cost-aware Multipath Routing. In *IFIP AIMS'09*, pp. 207-–210.

[45] Augustin B., Cuvellier X., Orgogozo B., Viger F., Friedman T., Latapy M., Magnien C., and Teixeira R. Avoiding Traceroute Anomalies with Paris Traceroute. In *ACM IMC'06*, pp. 153—158.

[46] Augustin B., Friedman T., and Teixeira R. Measuring Load-balanced Paths in the Internet. In *ACM IMC'07*, pp. 149–160.

[47] Augustin B., Friedman T., and Teixeira R. Measuring Multipath Routing in the Internet. *IEEE/ACM Trans. Netw.* 19, 3 (2011), 830—840.

[48] Augustin B., Friedman T., and Teixeira R. Multipath Tracing with Paris Traceroute. In *IEEE E2EMON'07*, pp. 1–8.

[49] Bajpai V., Eravuchira S. J., and Schönwälder J. Lessons Learned from Using the RIPE Atlas Platform for Measurement Research. *ACM SIGCOMM Comput. Commun. Rev.* 45, 3 (2015), 35–42.

[50] van Beijnum I., Crowcroft J., Valera F., and Bagnulo M. Loop-freeness in Multipath BGP through Propagating the Longest Path. In *IEEE INFOCOM'09*, 6 pages.

[51] Beverly R. Yarrp'ing the Internet: Randomized High-Speed Active Topology Discovery. In *ACM IMC'16*, pp. 413–420.

[52] Beverly R., Luckie M., Mosley L., and Claffy K. Measuring and Characterizing IPv6 Router Availability. In *PAM'15*, pp. 123–135.

[53] Böttger T., Cuadrado F., and Uhlig S. Looking for Hypergiants in PeeringDB. *ACM SIGCOMM Comput. Commun. Rev.* 48, 3 (2018), 13–19.

[54] Briscoe B., Brunstrom A., Petlund A., Hayes D., Ros D., Tsang I.-J., Gjessing S., Fairhurst G., Griwodz C., and Welzl M. Reducing Internet Latency: A Survey of Techniques and Their Merits. *IEEE Commun. Surveys Tuts.* 18, 3 (2016), 2149–2196.

[55] Brito S. H. B., Santos M. A. S., dos Reis Fontes R., Perez D. A. L., and Rothenberg C. E. Dissecting the Largest National Ecosystem of Public Internet eXchange Points in Brazil. In *PAM'16*, pp. 333–345.

[56] Bruno L., Graziano M., Balzarotti D., and Francillon A. Through the Looking-Glass, and What Eve Found There. In *USENIX WOOT'14*, 8 pages.

[57] Camacho J. M., García-Martínez A., Bagnulo M., and Valera F. BGP-XM: BGP eXtended Multipath for Transit Autonomous Systems. *Comput. Netw.* 57 (2013), 954–975.

[58] Canbaz M. A., Thom J., and Gunes M. H. Comparative Analysis of Internet Topology Data Sets. In *INFOCOM WKSHPS'17*, pp. 635–640.

[59] Cardona J. C., Francois P., and Lucente P. Collection and Analysis of Data for Inter-domain Traffic Engineering. In *CSBC'14*, 10 pages.

[60] Chiesa M., Kamisiński A., Rak J., Rétvári G., and Schimid S. A Survey of Fast Recovery Mechanisms in the Data Plane. TechRxiv. Preprint (2020), 48 pages. https://doi.org/10.36227/techrxiv.12367508.v2

[61] Cisco, The Zettabyte Era: Trends and Analysis, *White Paper*, July 2016.

[62] Cisco, Cisco Annual Internet Report (2018–2023) White Paper, *White Paper*, March 2020.

[63] Clark D., Bauer S., Lehr W., claffy kc, Dhamdhere A., Huffaker B., and Luckie M. Measurement and Analysis of Internet Interconnection and Congestion. In *TPRC'14*, 16 pages.

[64] Comarela G., Gürsun G., and Crovella M. Studying Interdomain Routing over Long Timescales. In *ACM IMC'13*, pp. 227—234.

[65] Csoma A., Gulyás A., and Toka L. On Measuring the Geographic Diversity of Internet Routes. *IEEE Commun. Mag.* 55, 5 (2017), 192–197.

[66] Cunha Í., Teixeira R., Veitch D., and Diot C. DTRACK: A System to Predict and Track Internet Path Changes. *IEEE/ACM Trans. Netw.* 22, 4 (2014), 1025—1038.

[67] Cunha Í., Teixeira R., and Diot C. Measuring and Characterizing End-to-End Route Dynamics in the Presence of Load Balancing. In *PAM'11*, pp. 235–244.

[68] Dhamdhere A., Clark D. D., Gamero-Garrido A., Luckie M., Mok R. K. P., Akiwate G., Gogia K., Bajpai V., Snoeren A. C., and Claffy kc. Inferring Persistent Interdomain Congestion. In *ACM SIGCOMM'18*, pp. 1–15.

[69] Dhamdhere A., and Dovrolis C. Twelve Years in the Evolution of the Internet Ecosystem. *IEEE/ACM Trans. Netw.* 19, 5 (2011), 1420–1433.

[70] Du B., Candela M., Huffaker B., Snoeren A. C., and claffy kc. RIPE IPmap Active Geolocation: Mechanism and Performance Evaluation. *ACM SIGCOMM Comput. Commun. Rev.* 50, 2 (2020), 4–10.

[71] Faggiani A., Gregori E., Improta A., Lenzini L., Luconi V., and Sani L. A Study on Traceroute Potentiality in Revealing the Internet AS-level Topology. In *IFIP Networking'14*, pp. 1-–9.

[72] Fanou R., Francois P., and Aben E. On the Diversity of Interdomain Routing in Africa. In *PAM'15*, pp. 41-–54.

[73] Fanou R., Francois P., Aben E., Mwangi M., Goburdhan N., and Valera F. Four Years Tracking Unrevealed Topological Changes in the African Interdomain. *Comput. Netw.* 106 (2017), 117–135.

[74] Fanou R., Valera F., and Dhamdhere A. Investigating the Causes of Congestion on the African IXP Substrate. In *ACM IMC'17*, pp. 57–63.

[75] Fei G., Ye J., Wen S., and Hu G. Network Topology Inference Using Higher-Order Statistical Characteristics of End-to-End Measured Delays. *IEEE Access* 8 (2020), 59960–59975.

[76] Formoso A., Chavula J., Phokeer A., Sathiaseelan A., and Tyson G. Deep Diving into Africa's Inter-Country Latencies. In *IEEE INFOCOM'18*, pp. 2231–2239.

[77] Fujinoki H. Multi-Path BGP (MBGP): A Solution for Improving Network Bandwidth Utilization and Defense against Link Failures in Inter-Domain Routing. In *IEEE ICON'08*, pp. 1—6.

[78] Garcia Gomez J. L., Wang R., Chen M.-H., and Chou C.-F. ETMP-BGP: Effective Tunnel-based Multi-Path BGP Routing Using Software-Defined Networking. In *IEEE SMC'17*, pp. 420—425.

[79] Gharaibeh M., Shah A., Huffaker B., Zhang H., Ensaif R., and Papadopoulos C. A Look at Router Geolocation in Public and Commercial Databases. In *ACM IMC'17*, pp. 463–469.

[80] Gharaibeh M., Zhang H., Papadopoulos C., and Heidemann J. Assessing Co-Locality of IP Blocks. In *IEEE INFOCOM WKSHPS'16*, 6 pages.

[81] Gigis P., Kotronis V., Aben E., Strowes S. D., and Dimitropoulos X. Characterizing User-to-User Connectivity with RIPE Atlas. In *ACM ANRW'17*, pp. 4–6.

[82] Giotsas V., Dhamdhere A., and claffy kc. Periscope: Unifying Looking Glass Querying. In *PAM'16*, pp. 177–189.

[83] Giotsas V., Koch T., Fazzion F., Cunha Í., Calder M., Madhyastha H. V., and Katz-Bassett E. Reduce, Reuse, Recycle: Repurposing Existing Measurements to Identify Stale Traceroutes. In *ACM IMC'20*, pp. 247–265.

[84] Giotsas V., Luckie M., Huffaker B., and claffy k. Inferring Complex AS Relationships. In *ACM IMC'14*, pp. 23-–30.

[85] Giotsas V., Smaragdakis G., Huffaker B., Luckie M., and claffy kc. Mapping Peering Interconnections to a Facility. In *ACM CoNEXT'15*, pp. 1–13.

[86] Giotsas V., and Zhou S. Detecting and Assessing the Hybrid IPv4/IPv6 AS Relationships. In *ACM SIGCOMM'11*, pp. 424-–425.

[87] Giotsas V., and Zhou S. Improving the Discovery of IXP Peering Links through Passive BGP Measurements. In *IEEE INFOCOM WKSHPS'13*, pp. 121-–126.

[88] Giotsas V., Zhou S., Luckie M., and claffy kc. Inferring Multilateral Peering. In *ACM CoNEXT'13*, pp. 247–258.

[89] Giotsas V., and Zhou S. Inferring Internet AS Relationships Based on BGP Routing Policies. http://arxiv.org/abs/1106.2417.

[90] Giotsas V., and Zhou S. Valley-free Violation in Internet Routing - Analysis Based on BGP Community Data. In *IEEE ICC'12*, pp. 1193–1197.

[91] Grailet J.-F., and Donnet B. Towards a Renewed Alias Resolution with Space Search Reduction and IP Fingerprinting. In *TMA'17*, 9 pages.

[92] Grailet J.-F., Tarissan F., and Donnet B. TreeNET: Discovering and Connecting Subnets. In *TMA'16*, 8 pages.

[93] Gregori E., Improta A., Lenzini L., and Orsini C. The Impact of IXPs on the AS-level Topology Structure of the Internet. *Comput. Commun.* 34 (2011), 68–82.

[94] Gregori E., Improta A., Lenzini L., Rossi L., and Sani L. On the Incompleteness of the AS-level Graph: A Novel Methodology for BGP Route Collector Placement. In *ACM IMC'12*, pp. 253–264.

[95] Gregori E., Improta A., Lenzini L., Rossi L., and Sani L. Improving the Reliability of Inter-AS Economic Inferences through a Hygiene Phase on BGP Data. *Comput. Netw.* 62 (2014), 197–207.

[96] Gregori E., Improta A., Lenzini L., Rossi L., and Sani L. A Novel Methodology to Address the Internet AS-Level Data Incompleteness. *IEEE/ACM Trans. Netw.* 23, 4 (2015), 1314–1327.

[97] Gunes M. H., and Sarac K. Resolving IP Aliases in Building Traceroute-Based Internet Maps. *IEEE/ACM Trans. Netw.* 17, 6 (2009), 1738–1751.

[98] Gupta A., Calder M., Feamster N., Chetty M., Calandro E., and Katz-Bassett E. Peering at the Internet's Frontier: A First Look at ISP Interconnectivity in Africa. In *PAM'14*, pp. 204–213.

[99] He Y., Faloutsos M., Krishnamurthy S., and Huffaker B. On Routing Asymmetry in the Internet. In *IEEE GLOBECOM'05*, pp. 904–909.

[100] Hedrick C. Routing Information Protocol, RFC 1058, June 1988.

[101] Holterbach T., Pelsser C., Bush R., and Vanbever L. Quantifying Interference between Measurements on the RIPE Atlas Platform. In *ACM IMC'15*, pp. 437–443.

[102] Hopps C. Analysis of an Equal-Cost Multi-Path Algorithm. RFC 2992, November 2000.

[103] Huang H., Zhu X. Bi J. Cao W. and Zhang X. Machine Learning for Broad-Sensed Internet Congestion Control and Avoidance: A Comprehensive Survey. *IEEE Access* 9 (2021), 31525–31545.

[104] Huffaker B., Dhamdhere A., Fomenkov M., and claffy K. Toward Topology Dualism: Improving the Accuracy of AS Annotations for Routers. In *PAM'10*, pp. 101–110.

[105] Iodice M., Candela M., and Di Battista G. Periodic Path Changes in RIPE Atlas. *IEEE Access* 7 (2019), 65518—65526.

[106] Jacquemart Q., Urvoy-Keller G., and Biersack E. A Longitudinal Study of BGP MOAS Prefixes. In *TMA'14*, pp. 127–138.

[107] Jasinska E., Hilliard N., Raszuk R., and Bakker N. Internet Exchange BGP Route Server, RFC 7947, September 2016.

[108] Javed U., Cunha I., Choffnes D. R., Katz-Bassett E., Anderson T., and Krishnamurthy A. PoiRoot: Investigating the Root Cause of Interdomain Path Changes. *ACM SIGCOMM Comput. Commun. Rev.* 40, 4 (2013), 183–194.

[109] Jia S., Luckie M., Huffaker B., Elmokashfi A., Aben E., Claffy K., and Dhamdhere A. Tracking the Deployment of IPv6: Topology, Routing and Performance. *Comput. Netw.* 165 (2019), 15 pages.

[110] Jin Z., Shi X., Yang Y., Yin X., Wang Z., and Wu J. TopoScope: Recover AS Relationships from Fragmentary Observations. In *ACM IMC'20*, pp. 266–280.

[111] Jin Y., Scott C., Dhamdhere A., Giotsas V., Krishnamurthy A., and Shenker S. Stable and Practical AS Relationship Inference with ProbLink. In *USENIX NSDI'19*, pp. 581–597.

[112] Keys K., Hyun Y., Luckie M., and Claffy K. Internet-Scale IPv4 Alias Resolution with MIDAR. *IEEE/ACM Trans. Netw.* 21, 2 (2013), 383–399.

[113] Khan A., Kim H.-C., Kwon T. T., and Choi Y. A Comparative Study on IP Prefixes and Their Origin Ases in BGP and the IRR. *ACM SIGCOMM Comput. Commun. Rev.* 43, 3 (2013), 17–24.

[114] Khan A., Kwon T. T., Kim H.-C., and Choi Y. AS-level Topology Collection through Looking Glass Servers. In *ACM IMC'13*, pp. 235–241.

[115] Klöti R., Ager B., Kotronis V., Nomikos G., and Dimitropoulos X. A Comparative Look into Public IXP Datasets. *ACM SIGCOMM Comput. Commun. Rev.* 46, 1 (2016), 22–29.

[116] Kotronis V., Nomikos G., Manassakis L., Mavrommatis D., and Dimitropoulos X. Shortcuts through Colocation Facilities. In *ACM IMC'17*, 470–476.

[117] Landa R., Araújo J. T., Clegg R. G., Mykoniati E., Griffin D., and Rio M. The Large-Scale Geography of Internet Round Trip Times. In *IFIP Networking'13*, 9 pages.

[118] Lapukhov P. Equal-cost Multipath Considerations for BGP, Internet Engineering Task Force. Network Working Group Internet Draft. https://tools.ietf.org/id/draft-lapukhov-bgp-ecmp-considerations-02.html. July 2019.

[119] Lee Y., Park H., and Lee. Y. IP Geolocation with a Crowd-sourcing Broadband Performance Tool. *ACM SIGCOMM Comput. Commun. Rev.* 46, 1 (2016), 13–20.

[120] Li M., Lukyanenko A., Ou Z., Ylä-Jääski A., Tarkoma S., Coudron M., and Secci S. Multipath Transmission for the Internet: A Survey. *IEEE Commun. Surveys Tuts.* 18, 4 (2016), 2887–2925.

[121] Li J., Giotsas V., and Zhou S. Anatomy of Multipath BGP Deployment in a Large ISP Network. In *TMA'20*, 9 pages. https://arxiv.org/abs/2012.07730.

[122] Li J., Zhou S., and Giotsas V. Performance Analysis of Multipath BGP. In *IEEE INFOCOM WKSHPS'21*, 15 pages. https://arxiv.org/abs/2103.07683.

[123] Livadariu I., Dreibholz T., Al-Selwi A. S., Bryhni H., Lysne O., Bjørnstad S., and Elmokashfi A. On the Accuracy of Country-Level IP Geolocation. In *ACM ANRW'20*, pp. 67–73.

[124] Lodhi A., Larson N., Dhamdhere A., Dovrolis C., and claffy kc. Using PeeringDB to Understand the Peering Ecosystem. *ACM SIGCOMM Comput. Commun. Rev.* 44, 2 (2014), 21–27.

[125] Luckie M. Scamper: A Scalable and Extensible Packet Prober for Active Measurement of the Internet. In *ACM IMC'10*, pp. 239–245.

[126] Luckie M., and claffy kc. A Second Look at Detecting Third-Party Addresses in Traceroute Traces with the IP Timestamp Option. In *PAM'14*, pp. 46–55.

[127] Luckie M., Dhamdhere A., Clark D., Huffaker B., and claffy kc. Challenges in inferring Internet Interdomain Congestion. In *ACM IMC'14*, pp. 15–21.

[128] Luckie M., Dhamdhere A., Huffaker B., Clark D., and claffy kc. bdrmap: Inference of Borders between IP Networks. In *ACM IMC'16*, pp. 381—396.

[129] Luckie M., Huffaker B., Dhamdhere A., Giotsas V., and claffy kc. AS Relationships, Customer Cones, and Validation. In *ACM IMC'13*, pp. 243—256.

[130] Luckie M., Hyun Y., and Huffaker B. Traceroute Probe Method and Forward IP Path Inference. In *ACM IMC'08*, pp. 311—324.

[131] Madhyastha H. V., Isdal T., Piatek M., Dixon C., Anderson T., Krishnamurthy A., and Venkataramani A. iPlane: An Information Plane for Distributed Services. In *USENIX OSDI'06*, pp. 367-–380.

[132] Mao Z. M., Johnson D., Rexford J., Wang J., and Katz R. Scalable and Accurate Identification of AS-Level Forwarding Paths. In *IEEE INFOCOM'04*, pp. 1605–1615.

[133] Mao Z. M., Rexford J., Wang J., and Katz R. Towards an Accurate AS-Level Traceroute Tool. In *ACM SIGCOMM'03*, pp. 365–378.

[134] Marchetta P., de Donato W., and Pescapé A. Detecting Third-Party Addresses in Traceroute Traces with IP Timestamp Option. In *PAM'13*, pp. 21–30.

[135] Marchetta P., Montieri A., Persico V., Pescapé A., Cunha Í., and Katz-Bassett E. How and How Much Traceroute Confuses Our Understanding of Network Paths. In *IEEE LANMAN'16*, 7 pages.

[136] Marder A., Luckie M., Dhamdhere A., Huffaker B., claffy kc, and Smith J. M. Pushing the Boundaries with bdrmapIT: Mapping Router Ownership at Internet Scale. In *ACM IMC'18*, pp. 56-–69.

[137] Marder A., and Smith J. M. MAP-IT: Multipass Accurate Passive Inferences from Traceroute. In *ACM IMC'16*, pp. 397-–411.

[138] Marder A. APPLE: Alias Pruning by Path Length Estimation. In *PAM'20*, pp. 249–263.

[139] Mátray P., Hága P., Laki S., Csabai I., and Vattay G. On the Network Geography of the Internet. In *IEEE INFOCOM'11*, pp. 126–130.

[140] Mirkoviv D., Armitage G., and Branch P. A Survey of Round Trip Time Prediction Systems. *IEEE Commun. Surveys Tuts.* 20, 3 (2018), 1758–1776.

[141] Mitseva A., Panchenko A. and Engel T., The State of Affairs in BGP Security: A Survey of Attacks and Defenses. *Comput. Commun.* 124 (2018), 45–60.

[142] Mok R. K. P., Bajpai V., Dhamdhere A., and Claffy K. C. Revealing the Load-Balancing Behavior of YouTube Traffic on Interdomain Links. In *PAM'18*, pp. 228—240.

[143] Motamedi R., Yeganeh B., Chandrasekaran B., Rejaie R., Maggs B. M., and Willinger W. On Mapping the Interconnections in Today's Internet. *IEEE/ACM Trans. Netw.* 27, 5 (2019), 2056–2070.

[144] Moy J. OSPF Version 2. RFC 2178, July 1997.

[145] Nomikos G., and Dimitropoulos X. traIXroute: Detecting IXPs in Traceroute Paths. In *PAM'16*, pp. 346–358.

[146] Nomikos G., Kotronis V., Sermpezis P., Gigis P., Manassakis L., Dietzel C., Konstantaras S., Dimitropoulos X., and Giotsas V. O Peer, Where Art Thou? Uncovering Remote Peering Interconnections at IXPs. In *ACM IMC'18*, pp. 265–278.

[147] Nur A. Y., and Tozal M. E. Cross-AS (X-AS) Internet Topology Mapping. *Comput. Netw.* 132 (2018), 53-–67.

[148] Oliveira R., Pei D., Willinger W., Zhang B., and Zhang L. The (In)Completeness of the Observed Internet AS-level Structure. *IEEE/ACM Trans. Netw.* 18, 1 (2010), 109–122.

[149] Orsini C., King A., Giordano D., Giotsas V., and Dainotti A. BGPStream: A Software Framework for Live and Historical BGP Data Analysis. In *ACM IMC'16*, pp. 429–444.

[150] Paxson V. End-to-End Routing Behavior in the Internet. *IEEE/ACM Trans. Netw.* 5, 5 (1997), 601-–615.

[151] Poese I., Uhlig S., Kaafar M. A., Donnet B., and Gueye B. IP Geolocation Databases: Unreliable? *ACM SIGCOMM Comput. Commun. Rev.* 41, 2 (2011), 53-–56.

[152] Qadir J., Ali A., Yau K.-L. A., Sathiaseelan A., and Crowcroft J. Exploiting the Power of Multiplicity: A Holistic Survey of Network-Layer Multipath. *IEEE Commun. Surv. Tutor.* 17, 4 (2015), 2176-–2213.

[153] Quoitin B., Pelsser C., Swinnen L., Bonaventure O., and Uhlig S. Interdomain Traffic Engineering with BGP. *IEEE Commun. Mag.* 41, 5 (2003), 122–128.

[154] Rekhter Y., Li T., and Hares S. A Border Gateway Protocol 4 (BGP-4), RFC 4271, January 2006.

[155] Rimondini M., Squarcella C., and Di Battista G. From BGP to RTT and Beyond: Matching BGP Routing Changes and Network Delay Variations with an Eye on Traceroute Paths. https://arxiv.org/abs/1309.0632.

[156] Rimondini M., Squarcella C., and Di Battista G. Towards an Automated Investigation of the Impact of BGP Routing Changes on Network Delay Variations. In *PAM'14*, pp. 193-–203.

[157] RIPE NCC Staff. RIPE Atlas: A Global Internet Measurement Network. *Internet Protoc. J.* 18, 3 (2015), 2—26.

[158] Saldan J. Delay Limits for Real-time Services. *IETF Draft*, 2016.

[159] Scheitle Q., Gasser O., Sattler P., and Carle G. HLOC: Hints-Based Geolocation Leveraging Multiple Measurement Frameworks. In *TMA'17*, 9 pages.

[160] Shao W., Devienne F., Iannone L., and Rougier J.-L. On the Use of BGP Communities for Fine-grained Inbound Traffic Engineering. 9 pages, https://arxiv.org/abs/1511.08336.

[161] Shavitt Y., and Shir E. DIMES: Let the Internet Measure Itself. *SIGCOMM Comput. Commun. Rev.* 35, 5 (2005), 71–74.

[162] Shavitt Y., and Weinsberg U. Quantifying the Importance of Vantage Point Distribution in Internet Topology Mapping. *IEEE J. Sel. Areas Commun.* 29, 9 (2011), 1837–1847.

[163] Shavitt Y., and Zilberman N. Improving IP Geolocation by Crawling the Internet PoP Level Graph. In *IFIP Networking'13*, 9 pages.

[164] Shavitt Y., and Zilberman N. Internet PoP Level Maps. In *DTMA'13*, pp. 82–103.

[165] Siddiqui M. S., Montero D., Serral-Gracià R., Masip-Bruin X., and Yannuzzi M. A Survey on the Recent Efforts of the Internet Standardization Body for Securing Inter-domain Routing. *Comput. Netw.* 80 (2015), 1–26.

[166] da Silva R. B., and Mota E. S. A Survey on Approaches to Reduce BGP Interdomain Routing Convergence Delay on the Internet. *IEEE Commun. Surv. Tutor.* 19, 4 (2017), 2949–2984.

[167] Singh S. K., Das T., and Jukan A. A Survey on Internet Multipath Routing and Provisioning. *IEEE Commun. Surv. Tutor.* 17, 4 (2015), 2157—2175.

[168] Spinelli L., Crovella M., and Eriksson B. AliasCluster: A Lightweight Approach to Interface Disambiguation. In *IEEE GI'13*, pp. 127–132.

[169] Tao N., Chen X., and Fu X. AS Path Inference: From Complex Network Perspective. In *IFIP Networking'15*, 9 pages.

[170] Valera F., van Beijnum I., García-Martínez A., and Bagnulo M. Multi-path BGP: Motivations and Solutions. In *Next-Generation Intenet Architectures and Protocols*, Ramamurthy B., Rouskas G. N., and Sivalingam K. M. (Ed.) Cambrige, UK: Cambridge Univ. Press, 2011.

[171] Veitch D., Augustin B., Teixeira R., and Friedman T. Failure Control in Multipath Route Tracing. In *IEEE INFOCOM'09*, pp. 1395–1403.

[172] Vermeulen K., Ljuma B., Addanki V., Gouel M., Fourmaux O., Friedman T., and Rejaie R. Alias Resolution Based on ICMP Rate Limiting. In *PAM'20*, pp. 231–248.

[173] Vermeulen K., Rohrer J. P., Beverly R., Fourmaux O., and Friedman T. Diamond-Miner: Comprehensive Discovery of the Internet's Topology Diamonds. In *USENIX NSDI'20*, pp. 479–493.

[174] Vermeulen K., Strowes S. D., Fourmaux O., and Friedman T. Multilevel MDA-Lite Paris Traceroute. In *ACM IMC'18*, pp. 29—42.

[175] Wang J., Bigham J., and Phillips C. A Geographical Proximity Aware Multipath Routing Mechanism for Resilient Networking. *IEEE Commun. Lett.* 21, 7 (2017), 1533–1536.

[176] Wang Z., Jin C., and Jamin S. Network Maps beyond Connectivity. In *IEEE GLOBECOM'05*, pp. 458-–462.

[177] Wassermann S., Casas P., and Donnet B. Machine Learning based Prediction of Internet Path Dynamics. In *ACM CoNEXT'16*, 3 pages.

[178] Wassermann S., Casas P., Donnet B., Leduc G., and Mellia M. On the Analysis of Internet Paths with DisNETPerf, a Distributed Paths Performance Analyzer. In *IEEE LCNW'16*, pp. 72–79.

[179] Wassermann S., Casas P., Cuvelier T., and Donnet B. NETPerfTrace – Predicting Internet Path Dynamics and Performance with Machine Learning. In *ACM Big-DAMA'17*, pp. 31-–36.

[180] Winter P., Padmanabhan R., King A., and Dainotti A. Geo-locating BGP Prefixes. In *IFIP TMA'19*, pp. 9–16.

[181] Wójcik R., Domżał J., Duliński Z., Rzym G., Kamisiński A., Gawłowicz P., Jurkiewicz P., Rząsa J., Stankiewicz R., and Wajda K. A Survey on Methods to Provide Interdomain Multipath Transmissions, *Comput. Netw.* 108 (2016), 233—259.

[182] Wu Z., Li W., Liu L., and Yue M. Low-Rate DoS Attacks, Detection, Defense, and Challenges: A Survey, *IEEE Access* 8 (2020), 43920–43943.

[183] Xu W., and Rexford J. MIRO: Multi-path Interdomain ROuting. In *ACM SIGCOMM'06*, pp. 171—182.

[184] Yeganeh B., Durairajan R., Rejaie R., and Willinger W. How Cloud Traffic Goes Hiding: A Study of Amazon's Peering Fabric. In *ACM IMC'19*, pp. 202–216.

[185] Yin X., Wu D., Wang Z., Shi X., and Wu J. DIMR: Disjoint Interdomain Multipath Routing. *Comput. Netw.* 91 (2015), 356–375.

[186] Young J., and Barth, T. Web Performance Analytics Show Even 100-millisecond Delays Can Impact Customer Engagement and Online Revenue, *Akamai Online Retail Performance Report*, 2017.

[187] Zhang B., Bi J., Wang Y., Zhang Y., and Wu J. Refining IP-to-AS Mappings for AS-level Traceroute. In *IEEE ICCCN'13*, 7 pages.

[188] Zhang B., Bi J., Wang Y., Zhang Y., and Wu J. Revisiting IP-to-AS Mapping for AS-level Traceroute. In *ACM CoNEXT'11*, 2 pages.

[189] Zhang B., Bi J., Wu J., and Baker F. CTE: Cost-Effective Intra-domain Traffic Engineering. In *ACM SIGCOMM'14*, pp. 115–116.

[190] Zhang Y., Oliveira R., Wang Y., Su S., Zhang B., Bi J., Zhang H., and Zhang L. A Framework to Quantify the Pitfalls of Using Traceroute in AS-Level Topology Measurement. *IEEE J. Sel. Areas Commun.* 29, 9 (2011), 1822—1836.

[191] Zhang X., and Perrig A. Correlation-Resilient Path Selection in Multi-Path Routing. In *IEEE GLOBECOM'10*, pp. 1—6.

[192] Zhao X., Pei D., Wang L., Massey D., Mankin A., Wu S. F., and Zhang L. An Analysis of BGP Multiple Origin AS (MOAS) Conflicts. In *ACM IMW'01*, pp. 31–35.