DISSERTATION

SYSTEMS FOR CHARACTERIZING INTERNET ROUTING

Submitted by

Anant Shah

Department of Computer Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2018

Doctoral Committee:

Advisor: Christos Papadopoulos

Shrideep Pallickara
Indrakshi Ray
Joseph Gersch
J. Rockey Luo
Randy Bush

ABSTRACT

SYSTEMS FOR CHARACTERIZING INTERNET ROUTING

Today the Internet plays a critical role in our lives; we rely on it for communication, business, and more recently, smart home operations. Users expect high performance and availability of the Internet. To meet such high demands, all Internet components including routing must operate at peak efficiency. However, events that hamper the routing system over the Internet are very common, causing millions of dollars of financial loss, traffic exposed to attacks, or even loss of national connectivity. Moreover, there is sparse real-time detection and reporting of such events for the public. A key challenge in addressing such issues is lack of methodology to study, evaluate and characterize Internet *connectivity*. While many networks operating autonomously have made the Internet robust, the complexity in understanding *how* users interconnect, interact and retrieve content has also increased. Characterizing how data is routed, measuring dependency on external networks, and fast outage detection has become very necessary using public measurement infrastructures and data sources.

From a regulatory standpoint, there is an immediate need for systems to detect and report routing events where a content provider's routing policies may run afoul of state policies. In this dissertation, we design, build and evaluate systems that leverage existing infrastructure and report routing events in near-real time. In particular, we focus on geographic routing anomalies i.e., detours, routing failure i.e., outages, and measuring structural changes in routing policies.

ACKNOWLEDGEMENTS

Looking back at my time as a PhD student, I realize how significantly it has transformed me. Every crest and trough during this process has been worth it. I see myself reaching out to hard problems and valuing the attention to details. But this change in myself and certainly this dissertation, would not have been possible without the support and contribution of many people.

I am sincerely thankful to my advisor, Christos Papadopoulos, for constant encouragement and direction. Early on in my PhD, Christos helped me to develop a focused view towards research. Christos helped me understand the academic process and how to use criticism and rejection rather than get demotivated by it. Going forward, I am certain that I will take a lot of what I have learned from him in my future endeavors.

I was fortunate to work in an environment that was fun but also filled with extremely talented teammates. I am grateful to the Network Security Research Group members Kaustubh Gadkari, Stephanie DiBenedetto, Han Zhang, Susmit Shannigrahi, Manaf Gharaibeh, Chengyu Fan for many insightful discussions and for sitting through numerous practice talks that I gave.

Each of my committee members has played an important role in my PhD journey. I am thankful to Shrideep Pallickara, Indrakshi Ray, Joe Gersch, Randy Bush, and Rockey Luo. Their feedback, support, and insightful comments have made this dissertation substantially better. Every conversation with my committee members has been extremely motivating and filled me with positivity.

I cannot thank enough for the contributions of several other researchers. In particular, I am greatly in debt to Romain Fontugne for many discussions, comments and working late nights (across timezones with the difference of more than 12 hours) to make this work possible. Romain has been a great mentor. He has the ability to keep a fine balance between work and play, a quality which I hope to achieve someday. Romain along with Kitamura-san made my stay in Tokyo extremely memorable. Randy Bush, Emile Aben, and Cristel Pelsser contributed heavily in making sure my work solved meaningful problems and provided several key insights about network operations in the wild that I was lacking. Dave Plonka and Arthur Berger at Akamai introduced

me to many new practical problems CDNs face and where I should direct my research efforts. Dave made my summer in Boston exciting with many stories about networking research, some of which admittedly even predate my birth! I was also fortunate to visit CAIDA at the University of California, San Diego. I thank Alberto Dainotti for hosting me and providing the opportunity to integrate some of my work into CAIDA's platform. I also thank Wim Bohm and partners for the award which partially supported this visit.

No measurement research is possible without data. Many researchers and volunteers make public data sources available which I have leveraged for my work. In particular, I would like to express my gratitude to University of Oregon RouteViews project which provided the BGP data along with BGPmon from Colorado State University. CAIDA's several Internet measurement datasets have also helped this work. Thousands of volunteers around the world host probes from RIPE NCC's project Atlas. Data from these probes have been the basis of systems I build in this dissertation. My deepest gratitude goes to these volunteers.

The support and love of my family is invaluable. My sister, Nikita Shah, and brother in law, Aashish Shah, helped me get comfortable in the United States. They are my home away from home. Talks with Aashish have immensely inspired me to think in unconventional ways. My father, Kiran Shah, has always motivated me to reach for the stars, work hard, and most importantly taught me to do the right thing while leading by example. My mother, Nilima Shah, made me the person I am today. She has relentlessly strived for my success and well being. She has the power to sense when I am feeling low and always finds a way to make me feel positive and ambitious. Simrik Neupane's love and encouragement kept me going strong and her positive outlook towards life has certainly rubbed off on me. I am thankful to her for believing in me more than what I did in myself. Friends make life more colorful. I have the pleasure of having friends that undoubtedly colored my life with numerous fun conversations, jokes, parties, and trips to explore Colorado. I thank all of them for keeping me young at heart.

The work in this dissertation was supported by several grants from the U.S. Department of Homeland Security. I thank them and appreciate their commitment to science and innovation.

DEDICATION

Thank you Mom and Dad.

TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENTS iii
LIST OF TABLES viii
LIST OF FIGURES ix

Chapter 1 Introduction 1
1.1 Thesis Statement and Contribution 4
1.2 Organization 5

Chapter 2 Background 6
2.1 Border Gateway Protocol 6
2.2 Control and Data Plane 7
2.3 Traffic Engineering 8

Chapter 3 Detecting Routing Detours 9
3.1 Data Sources 9
3.2 AS Geolocation 11
3.2.1 Prefix Geolocation 11
3.2.2 Infrastructure IP Geolocation 13
3.2.3 IXP Presence of an AS 13
3.2.4 AS to Country Set 14
3.3 Methodology 16
3.4 Validation 19
3.5 Results 24
3.5.1 Characterizing Detours 28
3.5.2 Transient and Flash Detours 31
3.6 Quantifying Geolocation Challenges 33
3.6.1 Evaluating Geolocation DB Accuracy 33
3.6.2 Active Measurements 36
3.7 Discussion 37
3.8 Summary 38

Chapter 4 Internet Outage Detection 40
4.1 Data Sources 40
4.2 Outage Detection 43
4.2.1 Burst Modeling 43
4.2.2 Sub Streams 45
4.2.3 Outage Reporting 46
4.2.4 Detection Results 47
4.3 Validation 49
4.3.1 Comparison to Traceroutes 49

4.3.2 Comparison to Trinocular . . . 50
4.4 Outage Characterization . . . 51
4.5 Case Studies . . . 55
4.5.1 Outages in TWC . . . 55
4.5.2 Outages in Amsterdam . . . 56
4.6 Outages in Control Plane . . . 56
4.7 Discussion . . . 58
4.8 Summary . . . 58

Chapter 5 Mapping the AS-Level Connectivity . . . 60
5.1 Betweenness Centrality . . . 61
5.2 Methodology . . . 63
5.3 AS Hegemony . . . 63
5.4 Results . . . 65
5.4.1 Country Level AS Graphs . . . 67
5.4.2 IPv4 and IPv6 Global AS Graphs . . . 67
5.4.3 Case Studies . . . 70
5.5 Discussion . . . 74
5.6 Summary . . . 74

Chapter 6 Internet Health Report . . . 75
6.1 Visualization . . . 75
6.2 API . . . 78

Chapter 7 Related Work . . . 79

Chapter 8 Future Work . . . 83

Chapter 9 Conclusions . . . 85

Bibliography . . . 87

LIST OF TABLES

3.1 Datasets: Detours Detection 10
3.2 Comparison of CAIDA's AS Rank with number of countries in AS Geolocation 14
3.3 Aggregate number of detours detected 25
3.4 Routes that may have peering relations 25
3.5 Top Detour Origin ASNs for all detoured paths 26
3.6 Top Detoured prefixes and corresponding percentages 27
3.7 Top Transient Detour Origin ASNs 32
3.8 Prefixes affected the most by transient detoured BGP paths 33
3.9 Some prefixes affected by flash detours 34

4.1 Datasets: Outage Detection 41
4.2 Outages reported by Disco vs. Trinocular 49

LIST OF FIGURES

2.1 Example showing BGP announcement 7

3.1 Flowchart: AS-to-Country mapping creation 12
3.2 CDF: Number of countries in AS geolocation 14
3.3 Example showing direction of BGP announcement and direction of observed detour . . 17
3.4 Example showing peering of ASes and RouteViews peer 19
3.5 Data plane measurements: Example showing selection of RIPE Atlas probes and target IPs 20
3.6 Example showing mapping from traceroute to AS Path 21
3.7 Validation Results: Live traceroutes using RIPE Atlas 23
3.8 Top Detour on May 2^{nd} 2016: Detected using `Netra`, visualization using `OpenIPMap`. Dotted arrow represents multiple hops and solid arrow represents direct hop. 24
3.9 Total number of definite detours per day in January 2016 25
3.10 Average number of detours per country 27
3.11 Flap Rate vs DC for US, RU and BR prefixes 29
3.12 Flap Rate vs DC for Non US, RU and BR prefixes 30
3.13 Persistence of definite detoured paths as seen by all peers 30
3.14 Distribution of detour duration 31
3.15 Average number of transient detours per country 32
3.16 Distribution of ASes that originated a transient detour. The top 4 Detour Origin ASes account for 50% of all transient detours 33
3.17 Distribution of prefixes that experienced a transient detour. About 30 prefixes account for 50% of all transient detours 34
3.18 Comparison of Maxmind Free DB and NetAcuity with ground truth. 35
3.19 CDF showing performance of `OpenIPMap` in comparison to ground truth and other DBs. 35
3.20 CDF showing performance of `AtlasCBG` in comparison to other databases. 37

4.1 Example showing raw connected probe counts and using country sub-stream of results obtained from the burst model with disconnect events reported on June 7th, 2016. . . . 42
4.2 Number of reported events for different threshold values. 47
4.3 Distribution of Average Velocity Ratio for normal and outage durations. 49
4.4 Number of outages with complete, incomplete, or no traceroute. 52
4.5 CDF: Ratio of unique faulty hops to the number of probes, low ratio indicates incomplete traceroutes ending at the same hop during an outage. 52
4.6 Distribution of the percentage of traceroute with forwarding loops per outage. 54
4.7 TWC outage: Probe counts and corresponding burst levels. 55
4.8 Amsterdam power outage. Probes detected using geoProximate sub-streams (red) and connected probes (green). 57
4.9 A timeline of BGP updates and data plane outage during hurricane Sandy 57

5.1 Comparison of Betweenness Centrality (BC) and AS hegemony with a toy example and BGP data. 61
5.2 KL divergence between AS hegemony scores obtained with $\alpha = 0.49$ and different values of α (global graph on 2017/12/15 with rv2, LINX, rrc00, and rrc10 collectors). 66
5.3 AS hegemony for paths toward countries 67
5.4 Distribution of AS hegemony for all ASes in the global graph. 68
5.5 AS hegemony for Tier-1 ISPs from 2004 to 2017 (global graph, IPv4). 68
5.6 Distribution of AS hegemony for Google and Akamai local graphs. Same color scale as Fig. 5.4. 71
5.7 AS hegemony for nodes in F-root (AS3557) and B-root (AS394353) local graphs from 15th January to 15th September 2017. 73

6.1 Internet Health Report 75
6.2 Outage detection reporting example 76
6.3 TraceMon Integration 77
6.4 AS Hegemony for Github 78
6.5 API example 78

Chapter 1

Introduction

Billions of users rely on the Internet for day-to-day activities. In the past decade, there has been a paradigm shift in how businesses operate, in an online-first or even online-only model, drastically increasing the dependency of the global economy on the Internet. According to a 2016 study by Boston Consultancy Group (BCG), the Internet-related economy will be worth $4.2 trillion worldwide. If considered as a national economy, it would be the fifth largest economy only after the US, China, India, and Japan. As the growth of the Internet continues, the Internet user-base is expected to increase by more than a billion by the year 2020. In context of the global economy, Internet's contribution will be 7.1% of that of G20 nations [1]. The increase in the number of users is accompanied by an increase in connection speeds, faster devices, and new platforms to interact with each other leveraging the Internet. Augmented/virtual reality, smart cities, Internet of things, autonomous vehicles, and many more upcoming platforms depend on the Internet to be available and reliable. As with any complex system, the Internet too has many moving parts and a comprehensive study of their *connectivity* becomes imperative.

Connectivity is often misinterpreted as a measure of Internet penetration in a region. While the number of users does play a role here, in this context the right question to ask is *if* users can connect, *how* do they reach the content. In such case, the focus is on routing. Internet routing plays a crucial role in overall user experience. Anomalies and failures in routing are commonplace on the Internet leading to increased latency and in some cases, complete loss of connectivity. In this work, we define connectivity as loss or degradation (unexpected change) in users' ability to reach the content.

Modern content delivery networks try to solve this problem by providing redundancy across geographically distributed servers such that most users can fetch the content from a near-by server rather than fetching it from the actual publisher of the content. This makes the studying of routing and how networks are connected all the more important. CDN providers (and in turn their users)

need measurement methods to understand how the content is routed, which networks are reachable and keep track of networks they rely on to be able to reach the clients. Moreover, with recent revelations about potential surveillance by foreign states, more and more nations, ISPs, and users are becoming aware of how their content over the Internet is routed [2].

We argue that there is a need for systems that can detect and characterize routing events that may impact performance, routing failures, dependencies between networks and make such information available to the research community in a near-real time fashion.

One example is routing detours that have been commonly observed on the Internet, for example, cities located in the African continent communicating via an external exchange point in Europe [3]. Many autonomous systems are also multinational, which means that routes traversing the AS may cross international boundaries. We define an international detour (detour for short) as a path that originates in an AS located in one country, traverses an AS located in a different country and returns to an AS in the original country. There have also been suspicious cases of detours. In November 2013, the Internet intelligence company Renesys (now owned by Dyn+Oracle) published an online article detailing an attack they called Targeted Internet Traffic Misdirection [4]. Using `Traceroute` data they discovered three paths that suffered a man-in-the-middle (MITM) attack. One path originated from and was destined to organizations in Denver, CO, after passing through Iceland, prompting concern and uncomfortable discussions with ISP customers.

Each of these anecdotes, while interesting in its own right, does not address the broader question about how prevalent such detours are, their dynamics and impact. Characterizing detours is important to several players: (a) as a tool for network engineers trying to diagnose problems; (b) policy makers aiming at adhering to potential national communication policies mandating that all intra-country communication be confined within national boundaries, (c) entrepreneurs looking for opportunities to deploy new infrastructure in sparsely covered geographical areas such as Africa, or (d) privacy-conscious states trying to minimize the amount of internal communication traversing different jurisdictions. In this dissertation we develop methodology and a system to detect detours to monitor the Internet routing system in near real-time and produce alerts. Network operators can

not only appear informed about the incident, but also may be able to take action in peer selection in response to the alerts. Finally, longitudinal analysis of detours can give us insight into how routing and network infrastructure evolve over time.

Routing failures, i.e., outages in the Internet also have large economic as well as social impact. It is well known that during the uprising against the government in Libya, the government imposed Internet censorship and entire country lost connectivity to the outside world [5]. A similar event occurred in Syria [6]. In many cases, however, routing failures are not intended and are a result of misconfiguration or hardware issues. It is common news that a routing misconfiguration at an ISP caused global or national Internet slow-down [7–9]. To detect such events many research projects throw millions of packets all over the Internet with the goal of detecting connectivity losses. While providing good outage survey datasets, they neglect to exploit existing data and infrastructure to alert in near-real time and characterize such outages. In this dissertation we aim to mine already available data, generating no new traffic, in order to measure connectivity issues and then compare this technique to other efforts. We present an outage detection system which uses stable, long-running TCP connections from wide-spread infrastructure, much of which is behind NATs and other devices that would otherwise block measurements to detect outages.

Networks connected to the Internet inherently rely on other Autonomous Systems (ASes) for routing data. To determine the path of ASes to go from one network to another, the Internet relies solely on the Border Gateway Protocol (BGP). Computed AS paths are the result of an involved process that considers various peering policies set by each connected AS. BGP only exposes paths that are selected by border routers at ASes, concealing many peering policies and the exact routing process. However, as the connectivity of a network depends greatly on the connectivity of other ASes (provider AS, provider's provider, etc.), operators need to clearly understand ASes that are crucial for their connectivity. Identifying these AS interdependencies facilitates decisions for deployments, routing decisions, and connectivity troubleshooting [10].

We aim at estimating the AS interdependencies from BGP data. We devise a methodology that models AS interconnections as a graph and measures AS centrality, that is the likelihood of an

AS to lie on paths between two other ASes. We identify shortcomings of classical graph metrics such as the centrality metric Betweenness Centrality (BC), when used with BGP data. From these observations, we employ a robust method to estimate AS centrality that we call AS hegemony.

1.1 Thesis Statement and Contribution

Current Internet routing analysis faces a number of challenges. For example, there are many data sources focusing on a specific feature of Internet measurement such as routing tables, traceroute measurements, peering relationships, geolocation of routers, etc. However, such data is produced independently at each source and is not designed to overlap. While there are many projects that produce a large amount of measurement data, there is sparse near-real time reporting to the public. Network operators often turn to outage mailing lists to find information about routing failures and customers often use social media-based platforms like downdetector.com, which are not very reliable and are of not much use to research community. Commercial tools do monitor such routing events for paying customers, but there is a lack of systems that make detection, analysis, and reporting tools available to a larger set of the community.

We believe it is necessary to develop systems that can monitor and characterize Internet routing using the available data sources and present characterization at a global scale. Given the scarcity of free public reporting of routing events and plethora of Internet measurement research projects we make the following thesis statement:

"Developing systems for near-real time Internet routing analysis at a global scale is possible using current public measurement infrastructure"

In this dissertation we make the following contributions:

1. **Mapping geographic presence of autonomous systems**

    - Design and develop methodology to use various geolocation sources and produce AS geolocation mapping

2. **Detecting international routing detours using control plane**

- First tool to detect routing detours using control plane data
- Develop metrics and classify detours

3. **Detecting bursty outages in the Internet**

   - First light-weight outage detector that utilizes existing TCP connections
   - Use existing traceroutes to detect outage characteristics like forwarding loops
   - Predict fault location using previous knowledge of successful traceroutes

4. **Measure Inter-dependence of autonomous systems**

   - Novel and robust methodology to evaluate degree of dependence on other networks
   - Enable monitoring of global AS topology using public sources

5. **Provide access to the results and datasets to public via RESTful API**

1.2 Organization

This dissertation is organized as follows. In Chapter 2 we present an overview of the Border Gateway Protocol (BGP) that makes inter domain routing over the Internet possible; here we also provide information on public Internet measurement infrastructures RouteViews [11] and RIPE RIS and Atlas [12]. In Chapter 3 we present our methodology to detect routing detours and their characterization. We describe outage detection in Chapter 4. In Chapter 5 we present the new AS dependency analysis, AS Hegemony, show it's advantages and analyze popular networks as case studies. In Chapter 6 we detail how the culmination of work presented in this dissertation is being exposed to larger research and operations community via a public portal, Internet Health Report. In Chapter 7 we present related work and highlight previous efforts in the direction similar to ours and point out key areas where this dissertation differs from them. Finally, in Chapters 8 and 9 we discuss open areas of research for future directions and conclude respectively.

Chapter 2

Background

In this Chapter, we present information about the protocols, terminologies, measurement systems used throughout the dissertation.

2.1 Border Gateway Protocol

The Internet comprises of numerous inter-connected networks. A collection of such sub-networks under one organization is referred to as an Autonomous System (AS). Each autonomous system owns a set of prefixes. Prefixes are an aggregate representation of a block of IP addresses that are used by routers to appropriately forward the packets they receive. Interconnected ASes announce the set of prefixes they own using the Border Gateway Protocol (BGP). Once an AS receives an announcement from neighboring AS it processes it by applying import filters. These filters evaluate if the received route to a particular AS is the best one received so far, does it meet organization's routing policies such as customer-provider relations, peering relations, etc. If it does match import filters, the router adds the received route in its routing table.

The routing announcements along with the prefix also include 'AS path'. Once a route is accepted by an AS, if it also matches the export policies, the AS prepends itself in the AS path and announce reachability to the prefix to its neighbors. An example of this process is shown in Figure 2.1. The routing information for various prefixes at a router of an AS is called Routing Information Base (RIB).

This exchange of routing information is referred to as **control-plane**. Information in this plane is only intended for the routers and not the end hosts (e.g. users). The user generated data comes under the **data-plane**. This includes data packets generated by active measurement tools `Traceroute` and `Ping` as well. Both these planes belong to the networking layer in the OSI model and the said labeling is only conceptual.

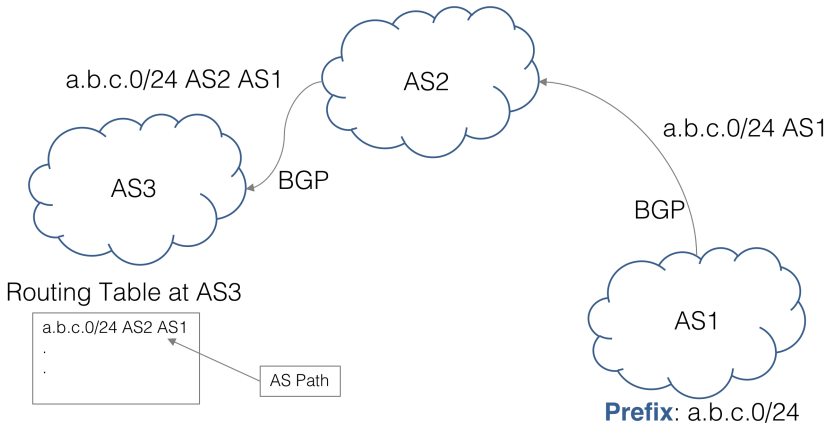


Figure 2.1: Example showing BGP announcement

Gathering data from both planes is important for Internet routing analysis. Thanks to a number of funded projects, such rich data sources and measurement platforms are already available publicly. We elaborate on the source of data for each plane next.

2.2 Control and Data Plane

RouteViews Project: The University of Oregon provides the biggest public repository of BGP announcements (updates) and RIBs under the RouteViews Project [11]. A binary RIB dump from more than 300 routers from around the world is obtained every 2 hours.

RIPE Routing Information System (RIS): Similar to RouteViews, RIPE NCC, the European Internet Registry, provides RIB dumps every 8 hours under their RIPE RIS project.

RIPE Atlas [12] is a global deployment of "probes" (tiny Linux machines) that continuously perform measurements such as pings and traceroutes to root servers and other special Atlas probes called Anchors. There are more than 10K probes in 178 countries. These probes also allow users to run measurements to desired destinations. Results of these measurements and probe availability metadata are available to users both as a historical dataset and a live stream.

2.3 Traffic Engineering

Traffic engineering plays a big role in the context of topics discussed in this dissertation. In particular, the way BGP and in turn Internet routing functions is heavily influenced by traffic engineering decision network operators make. Broadly, relationships between ASes can be divided into three types: Transit, Peering or Sibling. In a transit relationship, client AS pays its transit (provider) AS for connectivity. In a peering relationship, two ASes (usually with a similar volume of traffic) agree to exchange traffic between each other without any payment. In some cases, the relation is of a sibling, where ASes that belong to the same parent organization share traffic which is usually governed by internal policies. Moreover, Internet Exchange Points (IXPs) play a crucial role in how ASes peer with each other. IXPs are large physical locations where ASes interconnect. Some IXPs are free while some follow a paid peering format. Finally, relationships between ASes while logically one link, in reality, are physical interconnects in different metros, and capacity of those links becomes very important while steering Internet traffic. This complex framework of choosing how to steer traffic that takes into account free (peering) relations, link capacity, user experience, etc., comprises traffic engineering.

While ASes make decisions to optimize their traffic delivery, their decisions impact the global Internet routing system. Apart from attacks, even misconfigurations can lead to global Internet routing events. For example, in 2015, a route leak from Malaysia telecom cause a global slowdown in Level3's network [13]. Several apps and services, including credit card websites in North America suffered. It is, therefore, crucial to keep in mind that given the complex relations of ASes, routing over the Internet is subject to many potential threats some of which could even occur across the world and still impact your region.

With this background, we begin our discussion to design, build, and evaluate systems to characterize Internet routing with the goal of understanding its impact on connectivity. In the next chapter, we focus on geographically anomalous paths i.e., detours, that degrade user experience and even expose data to potential threats.

Chapter 3

Detecting Routing Detours

There are currently no requirements (technical or otherwise) that routing paths must be contained within national boundaries. Indeed, some paths experience *international detours*, i.e., originate in one country, cross international boundaries and return to the same country. In most cases these are sensible traffic engineering or peering decisions at ISPs that serve multiple countries. In some cases such detours may be suspicious. Characterizing international detours is useful to a number of players: (a) network engineers trying to diagnose persistent problems, (b) policy makers aiming at adhering to certain national communication policies, (c) entrepreneurs looking for opportunities to deploy new networks, or (d) privacy-conscious states trying to minimize the amount of internal communication traversing different jurisdictions. To detect detours we introduce `Netra`, a tool that reports detours in near-real time from more than 30 countries and can launch additional measurement if desired. To show detection capability of `Netra`, we also characterize international detours in the Internet on historical data for the month of January 2016.

3.1 Data Sources

We use variety of data sources to perform AS geolocation, BGP RIBS for detour detection and Traceroutes from RIPE Atlas for detour validation. In Table 3.1 we list different datasets with their usage and relevant information about each. Our sampling rate is 3 RIBs per day (one every eight hours, as provided by RIPE RIS) for a total of 38,688 RIBs from 416 peers. This spans 30 countries, which amounts to about 55GB of compressed MRT data. We acknowledge that 30 countries do not necessarily represent global scale, but our scope is limited by placement of peers that provide BGP feeds. We used all v4 peers in our analysis.

For geolocation of IP addresses we use MaxMind GeoLite City DB [14]. We treat end user IPs and infrastructure IPs differently since MaxMind is known to be more accurate for eye-ball networks only. To gather the list of infrastructure IPs we used list of routers from CAIDA Ark tracer-

Table 3.1: Datasets: Detours Detection

Name	Usage	Date	Sources	Info
BGP	AS Geolocation; Detour Detection	2016-01	RouteViews, RIPE RIS	38,688 RIBS, 416 peers, 30 countries, 55GB
Infrastructure IP List	AS Geolocation	2016-01 to 2016-03	CAIDA Ark, iPlane, OpenIPMap, RIPE Atlas Measurements	3M Router IPs
Infrastructure IPs to AS Mapping	Infrastructure IP geolocation	2015-08	CAIDA ITDK, iPlane	6.6M IP to AS mappings
AS to IXP Mapping	AS Geolocation	2016-01 to 2016-03	IXP websites, PeeringDB, PCH	368 IXP websites crawled
AS Relationship	Filtering peered paths from detection	2016-01	CAIDA AS Relationship	482,657 distinct relationships
Traceroute	Detour Validation	2016-05-01	RIPE Atlas	Used by `Netra`, 163 traceroutes
MaxMind	Prefix Geolocation; Detour Validation	2016-01, 2016-03	MaxMind GeoLite City (free and paid)	Paid version used only for geolocating infrastructure IPs and detour validation

outes [15], OpenIPMap [16], iPlane [17] and RIPE Atlas built-in measurements and the anchoring measurements. The built-in measurements use all the RIPE Atlas probes and the destinations are root servers. The anchoring measurements are from 400 Atlas probes to other 189 Atlas anchors. These infrastructure IPs are then mapped to AS using IP to AS mappings from CAIDA ITDK [18], iPlane or longest prefix match.

In addition to BGP sources, we use AS-to-IXP mapping to estimate presence of an AS in a country. We gather AS to IXP mappings from Packet Clearing House (PCH) [19], PeeringDB [20] and by crawling 368 IXP websites that make their participant list public. Finally, we use CAIDA AS Relationship datasets [21] to eliminate false positives from detours detected. In Section 3.2 we provide more details on how these datasets are used in AS geolocation along with a flowchart (Figure 3.1).

3.2 AS Geolocation

Currently the AS geographic information is limited. Regional Internet Registries (RIRs) provide location information as where the ASes are registered but not where they are announcing prefixes or where their infrastructures are located. Knight et al. [22] have gathered public data from network operators to build a collection of geolocated topologies of about 140 networks. A richer geolocation knowledge of ASes is however needed for large scale studies. This is a complex problem; there are variety of ASes (edge, transit, content providers, etc.) that may have infrastructure internationally and/or presence at IXPs which needs to be taken into consideration. Moreover, datasets that geolocate infrastructure IPs and provide IXP mappings are not available easily and belong to unrelated sources which causes disparity in time of data capture impacting usability. We provide a careful and exhaustive method to estimate geolocation of ASes, present key insights about the dataset and also make it open for community access and feedback.

We are interested in country level geolocation. We define AS geolocation as presence of an AS in a country. An AS can have presence in multiple countries, especially ASes that belong to large providers. We detect the presence of an AS in country A if it :

1. Announces a prefix that geolocates to A or
2. Has infrastructure IPs that geolocate to A or
3. Has a presence at an IXP in A.

In Figure 3.1 we show a flowchart detailing AS geolocation processes. There are 3 main steps as described above. In next sections we elaborate on each.

3.2.1 Prefix Geolocation

We begin by geolocating all advertised BGP prefixes by an AS. It is possible that during our analysis in January 2016 some AS erroneously announced prefixes that it did not own. Therefore we perform a simple filtering; we trust an AS to be owner of a prefix if it announced the prefix for at least 15 days in our dataset. We assume most mistakes or hijacks will be less than this duration.

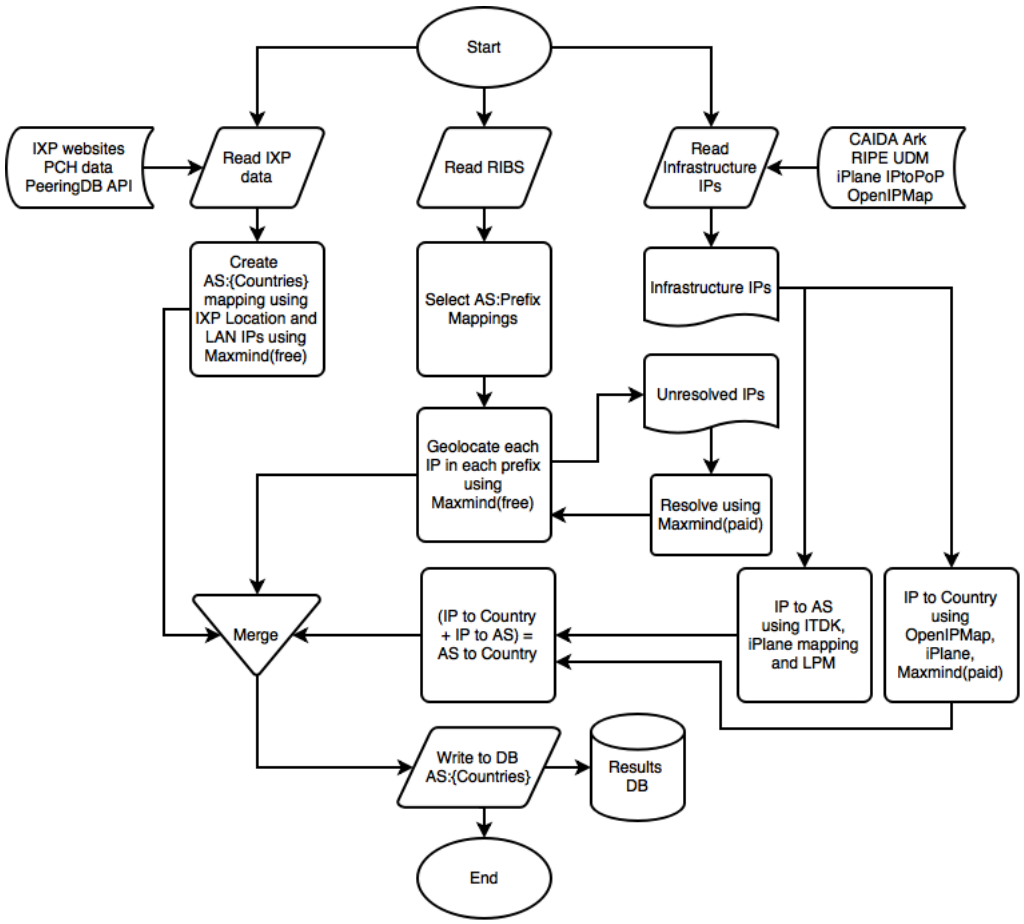


Figure 3.1: Flowchart: AS-to-Country mapping creation

Next, to map a BGP prefix to a country we geolocate each IP in the prefix using MaxMind-free. MaxMind could not geolocate 3.8M IPs. We could successfully geolocate 1.48M of these IPs with MaxMind-paid, we could not use remaining 2.32M IPs. Now we use the union of IP geolocation sets to get the BGP prefix geolocation. Due to the 2.32M IPs not geolocating even with paid version of MaxMind, 614 BGP prefixes could not be geolocated. For the remaining 610,722 BGP prefixes[1] which were geolocated we observe that more than 99% geolocated to single country. We note that 328,398 BGP prefixes were /24s. When BGP prefixes map to more than one country, the average size of the set was 2.9 countries. Finally, we perform union of geolocation sets of all BGP prefixes that an AS announces to create 1^{st} AS to country set.

[1]We use BGP prefixes 'as is' from the RIBs and do not perform any prefix aggregation. For example, if both /8 and /9 blocks of a prefix were seen in RIBs of the same or different peers, they are treated as 2 separate prefixes.

3.2.2 Infrastructure IP Geolocation

As mentioned previously, we treat infrastructure IP addresses separately. Router geolocation is known to be inaccurate [23]. Therefore for these IPs we want to create country geolocation set as large as possible. We populate list of router IPs from CAIDA Ark Traceroutes, iPlane IP to PoP mappings, OpenIPMap and RIPE built in measurements. Our list included 3M router IPs. This is the 'Read Infrastructure IPs' step shown in flowchart Figure 3.1. To geolocate each router IP we look at country location provided by iPlane, OpenIPMap[2] and Maxmind-paid and perform a union to give a set of countries. Next step is to map these routers to ASes. IP to AS is a challenging problem and active area of research. We use the best datasets available to create these mappings. Both CAIDA ITDK and iPlane datasets provide IP to AS mappings using the methodology described in [24]. For cases where either of these datasets fail to provide IP to AS mapping, we perform longest prefix match on the global routing table and map the IP to the AS announcing the longest matching prefix. Lastly, we combine IP to Country and IP to AS mappings to give 2^{nd} AS to country set.

3.2.3 IXP Presence of an AS

We extract presence of ASes at different IXPs and add the geolocation of IXP to the AS geolocation. As shown in 'Read IXP data' step in Figure 3.1, we use 3 sources of AS to IXP mappings. First, we crawl 368 IXP websites and extract their corresponding participants. Next, we use PeeringDB 2.0 API [20] and lastly, we use dataset from Packet Clearing House (PCH) that lists participants at IXPs that PCH is also a part of. We then combine geolocation obtained from these IXP sources to obtain 3^{rd} AS to country set. We acknowledge that IXP mappings from websites, PCH and PeeringDB might not be updated regularly and hence lead to mapping of an AS to a country that it does not have a presence in. Note that this will lead to false negative (not false positives) in detour detection, a trade-off we make to error on safe side.

[2]We use cases where confidence level for router geolocation is higher than 90%.

3.2.4 AS to Country Set

Finally, we map an AS to a set of countries by taking a union of all the 3 steps above. This is the merge step in Figure 3.1. The distribution of AS geolocation is shown in Figure 3.2. Perhaps surprisingly, only about 11.6% ASes out of a total of 52,984 geolocated to multiple countries. We believe that this is the result of a practice where most organizations use a different AS number in different countries. If an AS does geolocate to multiple countries we use the set of all countries in our analysis. We could not geolocate 24 ASes because none of their BGP prefixes could be geolocated, no infrastructure IP from our set mapped to it nor did we find its IXP presence in public datasets. These ASes on an average announced only 2 to 3 BGP prefixes.

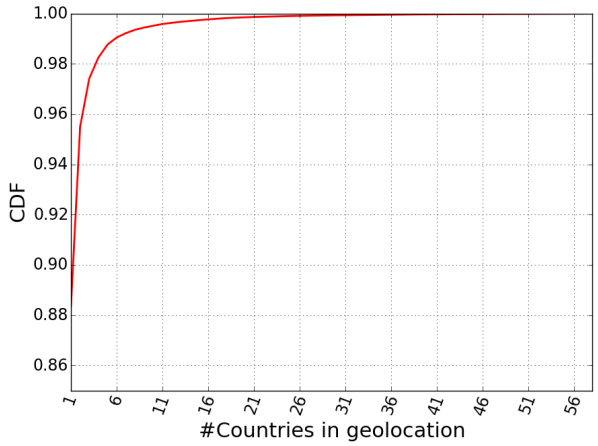


Figure 3.2: CDF: Number of countries in AS geolocation

Table 3.2: Comparison of CAIDA's AS Rank with number of countries in AS Geolocation

AS Rank	ASN	Customer Cone Size	AS Name	#Countries
1	3356	24,553	Level 3 Communications	63
2	174	17,891	Cogent Communications	58
3	3257	16,963	Tinet Spa	34
998	25394	18	MK Netzdienste GmbH Co. KG	2
999	6724	18	Strato AG	4
1000	52925	18	Ascenty DataCenters Locacao e Servicos LTDA	2

Comparison with CAIDA's AS Rank

Although our end goal is to detect detours, these geolocation results provide interesting insights. To understand more about which ASes geolocate to more than one country we use CAIDA's AS Rank [25]. This dataset gives higher ranks to ASes that have large customer cones. Intuitively, ASes with higher rank should resolve to many countries due to their wider presence. Table 3.2 shows ASes with their CAIDA AS rank and corresponding number of countries the AS geolocated to for top 3 and bottom 3 in the first 1000 ranked ASes. As expected, we see that ASes which have large presence with many customers across the world geolocate to large number of countries and low rank ASes with smaller customer cones geolocate to fewer countries.

RESTful API access

This is the first study to create an exhaustive mapping of AS geolocation which accounts for IXP relationships, we refer to this dataset and methodology as `ASMap` and make it available as a service. Our geolocation data can be accessed easily by issuing a simple web request. For example to get list of countries AS12145 has presence in:

```
$curl http://geoinfo.bgpmon.io/201704/asn_country/12145
```

The API returns JSON object:

```
{"ASNLocation": "{'US'}", "ASN": "12145"}
```

More information on using them and contributing to the dataset can be found at `http://geoinfo.bgpmon.io`. Researchers and network operators can also contribute with their ground truth data.

Use Cases

`ASMap` is beneficial for a wide range of studies. For example, to understand the relationships between countries' policies and traffic routing [26] or to estimate the dependency of a country on foreign ASes for international connectivity [27]. There is an increasing need to analyze geographic paths taken by local traffic either to avoid transnational detours through potential surveillance states [2, 28] or reduce latencies [3]. Detecting such cases in BGP data is eased using `ASMap` by geolocating possible countries within the AS path. Our results also complement tools like AS

Watch [29] (an AS reputation system), for example, to study which countries host more malicious ASes and if the location of such ASes changes drastically over a period of time. The geographical expansion of ASes over time can also be efficiently studied. In Chapter 5 we will use `ASMap` to gather ASes that have presence in a country to understand interdependency of networks between countries.

Detection capabilities of such systems will only improve with increasing quality of data and fixing corner cases of geolocation inaccuracies. For example, contribution from SINET (AS2907) revealed an extra mapping to Canada where it does not have a presence, a false positive due to inaccurate PoP geolocation.

3.3 Methodology

In this section, we focus on our methodology to detect international routing detours. We use the `ASMap` to detect detours in the control plane. RouteViews and RIS make RIBs from peering routers available every few hours, the foot print of control plane data is much smaller than data plane measurements. Using control data only for detection makes near-real time detour detection possible. We geolocate all the ASes along the AS path. To analyze the AS path, we provide the following definitions:

We define a path as having a detour if the origin and destination is country *'A'* but the path unambiguously includes some other country *'B'*. Note that this approach examines paths where the prefix origin AS and the AS where the peer is located are in the same country. To analyze the AS path, we provide the following definitions:

- **Prefix Origin**: The AS that announces the BGP prefix.
- **Detour Origin AS**: The AS that starts a detour in country *'A'* and diverts the path to foreign country *'B'*.
- **Detour Origin Country**: The country where we approximate location of Detour Origin AS, country *'A'*.

- **Detour Destination AS**: The AS in foreign country *'B'*.
- **Detour Return AS**: The AS where detour returns back in country *'A'*.

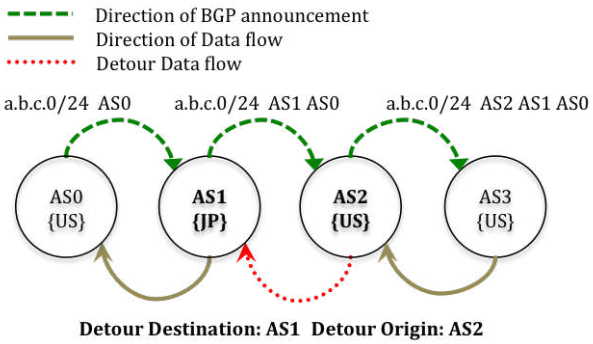


Figure 3.3: Example showing direction of BGP announcement and direction of observed detour

Figure 3.3 illustrates detours. *AS0* announces prefix *a.b.c.0/24* to AS1, AS2 and AS3. AS1 geolocates to JP whereas AS3, AS2 and AS0 are in the US. In this case, data traversing from AS3 to AS0 will contain a detour from AS2 (Detour Origin) to AS1 (Detour Destination). We do not include sub-paths in our analysis; other portions of the path that may experience a detour. For example, in path AS1{US}-AS2{IN}-AS3{CN}-AS4{IN}-AS5{US}, we only count the detour US-IN-US. We do not count the detour IN-CN-IN.

There are some cases where we need to approximate detour origin and country. In a path such as AS1{US}-AS2{US,BR}-AS3{CN}-AS4{US}. We resolve the uncertainty of the detour origin by assuming that it starts in AS2, since there is a likely path to AS2 from AS1 through the US and AS2 starts the detour from US, not BR. We do not characterize ***possible*** detours. For example, a path that geolocates to {US}-{US,IN}-{US} may in fact stay within the US and never visit India. In this work we only focus on paths that contain ***definite*** detours, such as {US}-{IN}-{US} or {US}-{IN,CN}-{US}. Again, we re-emphasize that in this work we only look at paths that confidently start and end in the same country; paths like {US,BR}-{IN}-{US} or {US}-{IN}-{BR} are not considered. We discard paths where we see an AS whose geolocation is unknown and a detour is not certain. For example, paths like AS1{US}-AS2{}-AS3{US} are discarded. However, if we see the detour occurring before the AS that could not be geolocated we do count it as a valid

detour i.e., in AS1{US}-AS2{BR}-AS3{US}-AS4{}-AS5{US}, AS4 does not have geolocation information but the US-BR-US detour occurred earlier. We treat this path as definite detour. We note that in addition to geolocation accuracy there is also some ambiguity about exact country boundaries. Some territories and relationships are currently disputed between multiple authorities and no worldwide consensus exists. For example, Hong-Kong and the People's Republic of China could be considered one or two entities. Hong-Kong is affiliated with China but it is a charter city and has its own independent constitution and judiciary system. For our analysis, we left the resolution of boundaries and countries to the MaxMind database. With this particular example, Hong-Kong and China are treated as two separate entities. MaxMind follows ISO 3166 country codes. In some cases the geolocation from MaxMind is ambiguous: 'A1:Anonymous Proxy', 'A2: Satellite Provider', 'O1: Other Country', 'EU: Europe', 'AP: Asia/Pacific'. We discard detours caused by these ambiguous codes, such as {DE}-{EU}-{DE}.

Filtering Peered AS paths

It is possible that the detour origin and the detour return ASes have a peering relationship (for example, Figure 3.4) and in reality traffic was not detoured at all. This, however, is hard to determine with certainty since peering relations and policies are not public. What we can do is provide an upper bound on how many detours may be eliminated due to peering. To detect such cases we use CAIDA's AS relationship dataset [21]. This dataset provides information of provider to provider (p2p) and provider to customer (p2c) relationship between ASes. We count cases where p2p link might be used, i.e., data originates from the peer itself or from a downstream customer. In case of p2c link we assume this link is always chosen. We eliminate such paths from our analysis and revisit this issue in the next section summarizing the peering relationships in Table 3.4.

Multi-Origin Prefixes

Some prefixes are announced by more than one ASes. We do not eliminate such cases. So, if a prefix *a.b.c.0/24* is seen in RIBs of 2 peers with AS paths 'X Y Z' and 'P Q R' then we treat each path as independent and detect detour if it fits above mentioned criteria of starting and ending

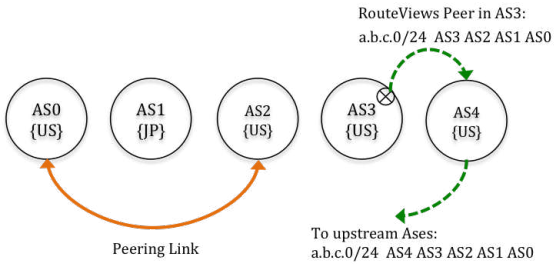


Figure 3.4: Example showing peering of ASes and RouteViews peer

in the same country. In our geolocation dataset we observed 7,579 prefixes of multi-origin (7,247 originated from 2 ASes). Out of these 6,104 suffered a detour. Motivation to not eliminate these prefixes is as follows: Network operators of such prefixes might want to re-evaluate their decisions especially if the ASes originating the prefix are in different countries. This might be a cause of high latency.

3.4 Validation

In this section we validate detours in near real time using *traceroutes* from RIPE Atlas probes. Our validation comprises of four steps:

1. Run `Netra` with live BGP feeds from 416 peers to detect detours.
2. When a detour is detected, run corresponding traceroutes (from same country and same AS) using RIPE Atlas.
3. Check if the traceroute and detour see similar AS path.
4. Validate using traceroute IP hops and RTT.

Data Plane Measurements

We ran `Netra` from May 2^{nd} 2016 noon to midnight (using BGP feeds from 416 peers). When a detour was detected in control plane we selected RIPE Atlas probes in the same country and same AS which we detected detour from and ran traceroute (ICMP Paris-traceroute [30]) to IP addresses in the detoured prefix. The methodology to run data plane measurements is shown in Figure 3.5.

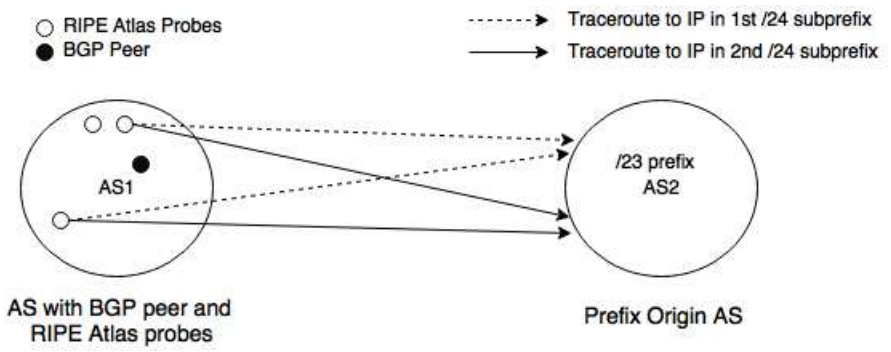


Figure 3.5: Data plane measurements: Example showing selection of RIPE Atlas probes and target IPs

There are a few cases where more than two Atlas probes are present in selected AS; in this case we selected 2 probes that are geographically farthest from each other. By doing this we aimed to account for cases where routes seen from geographically distant vantage points within the same AS are different. To select target IPs from detoured prefix, we break the prefix into its constituent /24s and randomly select an IP from each /24. For example, in a /23 prefix we select 2 IPs belonging to different /24s. By doing this we account for cases where a large prefix, even though in the same country, has different connectivity via different upstream provider. During this live run we detected 6,175 detours. Out of these 5,787 were unique detours ({peer,prefix,aspath} tuple).

Selecting Congruent Paths

Only 72 peers saw the 6,175 detours and the 72 peers belong to only 63 ASes. From these 63 ASes we then select ASes that also have active RIPE Atlas probes; there were only 10 ASes that both saw a detour and host a RIPE Atlas probe. 169 detours were seen from these 10 ASes corresponding to 6 countries: {Brazil, Italy, Norway, Russia, United States, South Africa}. From the 169 traceroutes we initiated to detoured prefixes, we discard 6 traceroutes where less than 3 hops responded since drawing detour conclusion from these is not possible. Finally, we are left with 163 traceroutes that can be used for validation. We acknowledge that 163 is not a very large number for validation purposes. However, running `Netra` for more hours does not necessarily increase the number of usable traceroutes for validation by a lot, we are limited by the number of ASes that have RIPE Atlas probes which also see a detour and detour-origin and detour-destination have no peering.

In total we detected 85 prefixes (corresponding to the 163 traceroutes) that suffered a detour that was visible from an AS which has RIPE Atlas probes. Note that some detoured prefixes were larger than /24, so we traceroute multiple IPs within it as explained in Section 3.4. The validation methodology is stated in Algorithm 1. As previous work [31] has pointed out, we found many cases where AS path seen in control plane and AS path seen in data plane do not match. However, these paths can still show detour if the detour origin AS and the detour destination AS are still present in the traceroute observed AS path. We call such AS paths *congruent*. More specifically, we consider the detoured AS path congruent only if detour origin AS and detour return AS both are present in the traceroute-observed AS path in the same order (detour origin first). For example, if an AS path '`A B C D E`' in control plane changed to '`A X B C E`' in data plane where '`B`' was detour origin and '`C`' was detour destination, we consider it as a congruent path. To resolve traceroute path to AS path we used CAIDA ITDK and iPlane IP to AS mappings and in cases where no match was found we use longest prefix match on the global routing table for the hop IP. Then we map the longest prefix match to the AS that originated it. As stated in Section 3.2.1 we believe an AS to own a prefix only if it announced it for more than 15 days in our one month dataset. While this is not an accurate method to determine ownership of a prefix, we believe most hijacks will be fixed within this window. The methodology to map traceroute to AS path is shown in Figure 3.6. Out of all the IPs we saw in 163 traceroutes, only 44 could be mapped to an AS using the IP to AS datasets. All other IPs were mapped using longest prefix match.

We observed 113 congruent AS paths. This includes 3 cases, insertions, deletions and mix of both. We borrow nomenclature of these paths from [31]. We saw 73 deletions, 29 insertions, 4 mix of insertion and deletions. The remaining 7 AS paths were exact matches. Note that these insertions and deletions occurred only for ASes that were not involved in the detour.

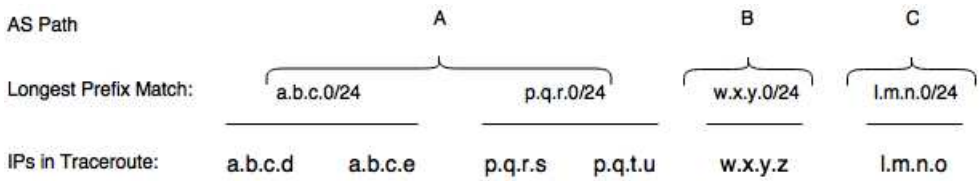


Figure 3.6: Example showing mapping from traceroute to AS Path

Algorithm 1 `Netra` Validation

```
procedure validateASPath
    aspath ← AS Path from Traceroute
    doas ← Detour Origin AS from Netra
    ddas ← Detour Destination AS from Netra
    if doas,ddas in aspath then
        if doas before ddas in aspath then
            Return True
procedure validateIPHops
    ipHops ← IP hops from Traceroute
    ipHopCountries ← MaxMind-paid
    if ipHopCountries show detour then
        detourDestTR ← Dest. from traceroute
        detourDestNetra ← Dest. from Netra
        if detourDestTR in detourDestNetra then
            Return True
procedure validateRTTs
    hopRTTs ← RTTs from Traceroute
    if hopRTTs show magnitudeJump then
        Return True
procedure MAIN
    loop: Each Detected Detour
    if validateASPath then
        validateIPHops
        validateRTTs
```

Validation Results

Now we validate detours detected by our methodology by comparing it with detours seen in data plane. For the 113 congruent AS paths, we evaluate if a data plane detour was seen. We chose to perform two tests. First, we resolve IPs observed in the hops of traceroute to country level geolocation using Maxmind-paid. We detect data plane detour if a path traversed foreign country and returned. We make sure that country visited (detour destination country) in data plane is present in the set of destination countries expected for this particular detour by `Netra`. We do this filtering to avoid false positives like: `Netra` detected detour {US}–{GB,DE}–{US} and traceroute detected detour {US}–{IT}–{US}. Although still a detour, since it was not accurately captured we count it as a miss. However, no such case was found. Second, we validate using RTT measurements. We detect RTT based detour if a hop in the traceroute showed increase in RTT

by an order of magnitude (at least 10 times increase). The results of this analysis are shown in Figure 3.7. We observed accuracy of about 85% (97 out of 113) in country-wise method and 90% (102 out of 113) by RTT measurements. The overlap between these two different tests was also large. 88 detours were detected in both (77.8%).

We investigate further the 9 detours that were seen in country-wise method but not in RTT. These detours covered small geographic area; 4 from Italy to France, 2 Norway to Sweden, 2 from Brazil to US and 1 from Russia to Sweden. RTTs between these countries have been previously reported to be low. Next we investigate 14 cases which were captured in RTT measurements but not in country-wise method. All of these do cross international boundaries. For 12 of these cases, due to large number of traceroute hops (especially towards the end of the traceroute) not responding we don't see the route returning to the origin country, hence not detected by country-wise method. We attribute remaining 2 cases as false positives due to inaccurate AS geolocation. In Figure 3.8 we provide a visualization of the most common detour we observed from Russia. Only visualization is done using OpenIPMap.

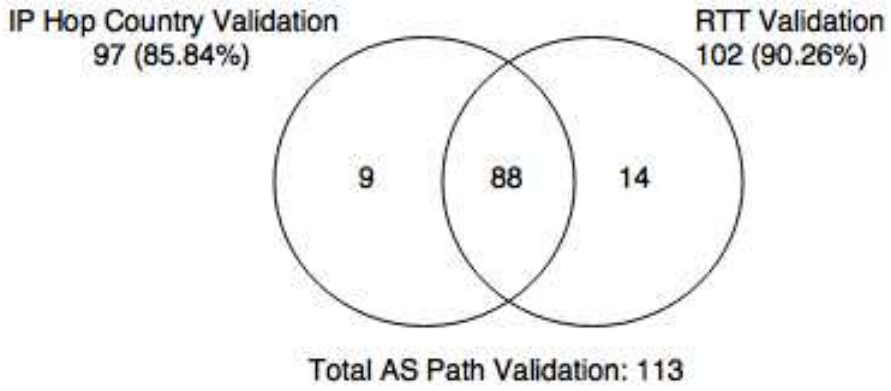


Figure 3.7: Validation Results: Live traceroutes using RIPE Atlas

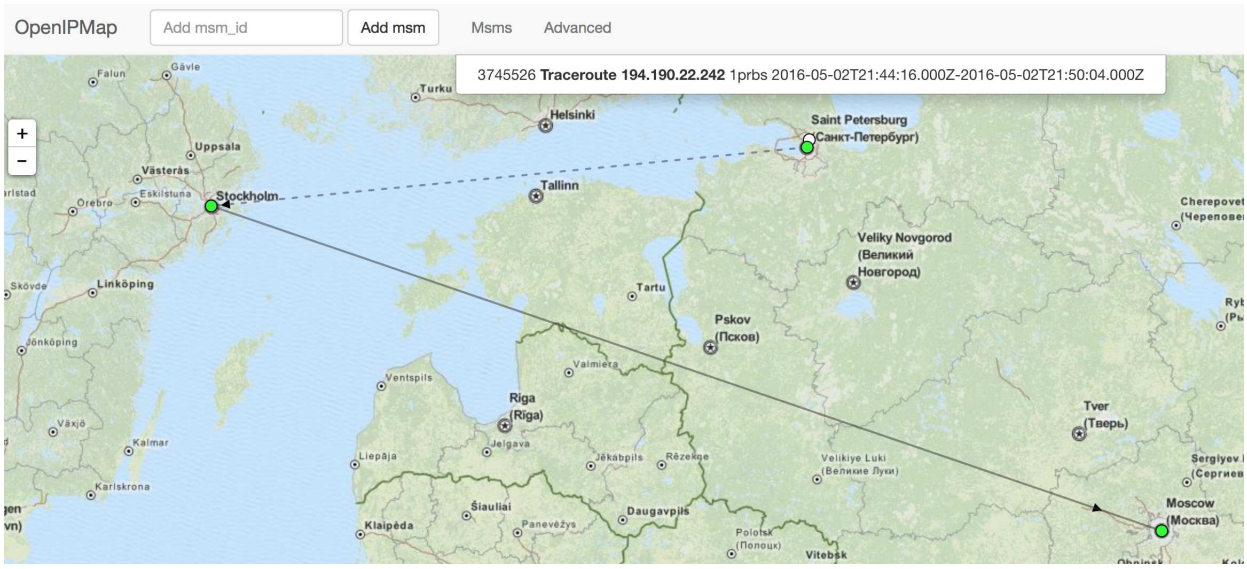


Figure 3.8: Top Detour on May 2^{nd} 2016: Detected using `Netra`, visualization using `OpenIPMap`. Dotted arrow represents multiple hops and solid arrow represents direct hop.

Validation Discussion

We show that large percentage of detours seen in control plane are accurately reflected in data plane as well. The main challenge is AS paths in both data plane and control plane don't agree in about 30% cases. We note that this could be an artifact of Atlas probes connected differently than the peers which provide BGP feeds. It is, however, possible to learn common AS insertions and deletions over a period of time and evolve detection capabilities.

3.5 Results

In this section we quantify detours detected in January 2016. First, in Section 3.5 we present an overview of all the detours detected in our dataset. In Section 3.5.1 we define metrics and classify detours based on their stability and availability. In Section 3.5.2 we focus on transient detours.

Aggregate Results

We begin by characterizing aggregate results, namely all detours seen by all peers; in other words, we count an incident every time an AS path appears in a RIB of any peer that contains a detour. Many of these incidents are duplicates. Therefore in addition to the total we also present

the number of unique detours. As expected, we observe that detours are not generally common. Also, not all peers see a detour. Only 79 peers, out of 416, saw one or more detours. Table 3.3 details the number of detours seen. We analyzed about 14 billion RIB entries and about 544K entries showed a detour; out of theses only 18.9K were unique (most detours re-appear during the month). Figure 3.9 shows the number of detours for each day in January 2016. On an average we find about 17.5K detoured entries per day.

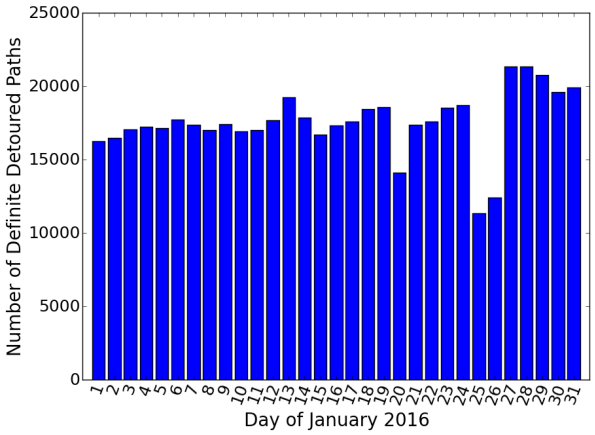


Figure 3.9: Total number of definite detours per day in January 2016

Table 3.3: Aggregate number of detours detected

#Total RIB Entries	#Detoured Entries	#Unique Detours
14,366,653,046	544,484	18,995

Table 3.4: Routes that may have peering relations

#Total Detours without filtering peered paths	#Detours with possible peering	%
659,569	115,085	17.4%

Next we examine the visibility of detours, where we observe an uneven distribution among ASes. Just 9 ASes originate more than 50% of the detours. Similarly, some prefixes experience

Table 3.5: Top Detour Origin ASNs for all detoured paths

Top Detour Origin AS	Total %	Frequent Detour Destination AS	% to frequent destination
3356 (Level 3 Communications,BR)	8.39%	32787 (Prolexic-Technologies DDoS Mitigation Network)	30.99%
12956 (Telefonica International Wholesale Services,BR)	5.74%	262182 Media Networks Latin America	46.33%
6939 (Hurricane Electric,US)	4.99%	45932 (Net Sys International Limited)	15.9%

detours more than others. 132 prefixes experienced more than 50% of the total detours. Looking at the average length of a detour, we see that a detour visits 1 to 2 foreign ASes before returning to its origin country.

Impact of Peering

We now estimate the effect of peering links on detours. Specifically, we are interested in cases where a peering relationship exists between the *Detour Origin AS* and the *Detour Return AS* as described previously using CAIDA AS relationship dataset. If such a link exists, it is possible that traffic traverses that link instead of the detour. Table 3.4 shows the number of detours between ASes that also have peering relations compared to total number of detours without filtering peered paths. We find that 17.4% of the detours are avoided due to peering relations. We do not count these as detours in our analysis.

Top Detour Origins and Prefixes

To understand more about the nature of these detours, we focus on the origin and destination ASes. In Table 3.5 we show the common detour origins and country where the AS was approximated to origin the detour from. Next is the percentage of detours out of the total that started from given origin. Following the percentage, is the most frequent destination that was visited from the origin, and lastly is the percentage of detours that went to most common destination from the said origin. We observe that most commonly these were access provider ASes. Similarly, in Table 3.6 we show top impacted prefixes.

Table 3.6: Top Detoured prefixes and corresponding percentages

Prefix Affected	Total %	Frequent Detour Destination AS	% to frequent destination
199.253.181.0/24 (Internet Systems Consortium,US)	0.51%	766 (Entidad Publica Empresarial Red)	100%
167.220.28.0/23 (Microsoft,US)	0.51%	6584 (Microsoft Corp)	100%
199.6.5.0/24 (Internet Systems Consortium,US)	0.51%	766 (Entidad Publica Empresarial Red)	77.11%

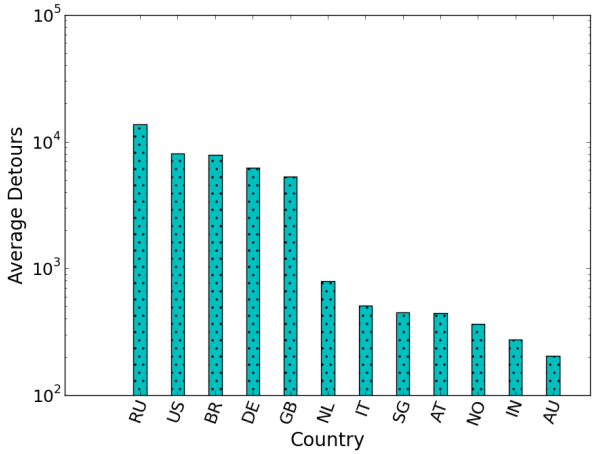


Figure 3.10: Average number of detours per country

Country-Wise Analysis

To provide an understanding on number of detours per peer in each country we normalize the data by dividing the number of detours by number of peers in the country. The reason to normalize data is simple, RouteViews and RIPE RIS peers are not evenly distributed among different countries. Therefore it is possible that more detours are seen in countries that have more peers due to more visibility. An average number of detours per peer per country provides better insight. Out of 30 countries, only 12 countries observed a detour. Figure 3.10 shows average number of detours per country. Russia showed most number of average detours. Understanding the total number of detours in different countries is important but it does not reflect if detours seen in different countries have different characteristics. In the next section we focus on characterizing these detours.

3.5.1 Characterizing Detours

To characterize detours we define two metrics:

1. **Detour Dynamics**

   (a) **Flap Rate**: Measure of *stability* of a detour; how many times a detour disappeared and reappeared.

   (b) **Duty Cycle**: Measure of *uptime* of a detour throughout the month measurement period.

2. **Persistence**: Total number of continuous hours a prefix was seen detoured.

Before using the above metrics to characterize the detours, we perform data pruning to avoid skewing of data towards ASes that have more peers that provide BGP feeds to RouteViews and RIPE RIS. Also, ASes with multiple peers and similar views can contribute duplicate detours to our dataset. We follow a simple approach to deal with this problem: if an AS contains more than one peer we select the peer that saw the most detours as the representative of that AS. This may potentially undercount detours since some peers in same AS may see different detours. After selecting a representative we are left with 36 (out of 79) peers. We now continue our characterization of detours by looking at **detour dynamics**. Specifically we focus on flap rate and duty cycle, defined as follows:

$$FlapRate = \frac{TotalTransitions}{TotalTime} \times 100$$

$$DutyCycle = \frac{TotalUptime}{TotalTime} \times 100$$

These metrics provide insights into the life cycle of detours by measuring route uptime and stability. BGP route flapping is a known problem and has been studied in [32] by looking at BGP updates and RFC 2439 provides methods to dampen these. However, in context of this work, duty cycle and flap rate are calculated from the RIBs. We extract detours from the RIBs and evaluate when they disappear and reappear.

To understand if country where detours occur plays a role in detour dynamics, next we drill into country specific detours. Figure 3.11 shows a scatter plot of flap rate vs. duty cycle for various detours in US, Brazil and Russia. We selected these three countries because they show the most detours in our dataset; they account for 93% of detours. We see a triangular pattern with some outliers. Large number of detours show high duty cycle and low flap rate. We divide each figure into 4 quadrants based on average flap rate and average duty cycle of all detours. We name quadrants anti-clockwise starting from top right. US detoured paths appear more stable (lower flap rate and higher duty cycle) in II^{nd} quadrant. On the other hand, Russian and Brazilian detoured paths fall mostly in the I^{st}, III^{rd} and IV^{th} quadrant. Russian detours in general showed lower duty cycle than US and Brazil. We also present a similar scatter plot for all the non US, BR and RU detours in Figure 3.12. In this case we observed detours mostly in extreme ends on II^{nd} and III^{rd} quadrant indicating two categories of detours, either long lasting or very rare events. A network operator can use information like this and decide which quadrant detours are more interesting to focus on. While all of detours may need attention, we believe detours with low duty cycle and low flap rate may need immediate attention. We talk more about this in Section 3.5.2.

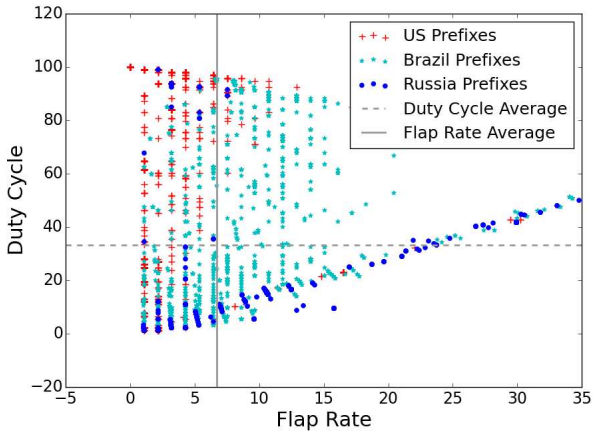


Figure 3.11: Flap Rate vs DC for US, RU and BR prefixes

Next, we examine the **persistence** of detours. Figure 3.13 shows the number of consecutive days a detour was visible by any peer. Note that persistence is measured in number of consecutive hours hence captures different characteristics than duty cycle which measures uptime throughout

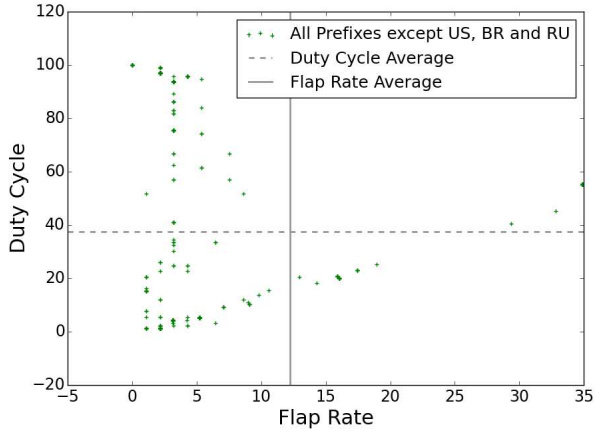


Figure 3.12: Flap Rate vs DC for Non US, RU and BR prefixes

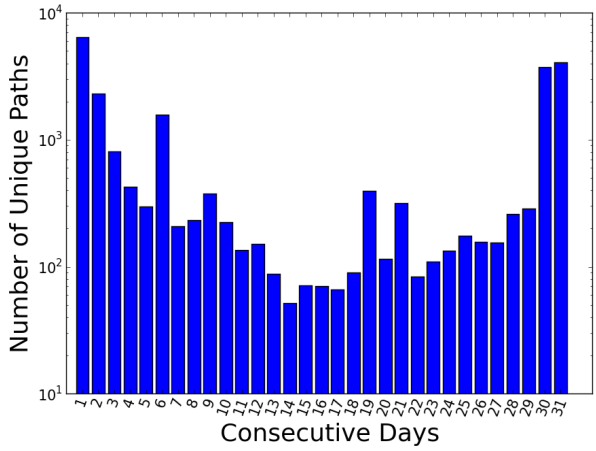


Figure 3.13: Persistence of definite detoured paths as seen by all peers

the dataset. We see a U-shaped pattern in Figure 3.13, meaning that many detours are either short lived (one day) or they persist for entire month. We take a different view at persistence in Figure 3.14 by plotting CDF of duration in hours. We see that most detours are short-lived, with about 92% lasting less than 72 hours, defined as *transient* detours. Finally, we examine a specific case of a transient detour, namely *flash detours* which appeared only once and never appeared again during the month.

In the following section we focus on transient and flash detours. Due to space limitations we do not characterize persistent detours further. We do note, however, that characterizing persistent detours is important for at least some of the reasons we enumerated earlier. We chose to focus on

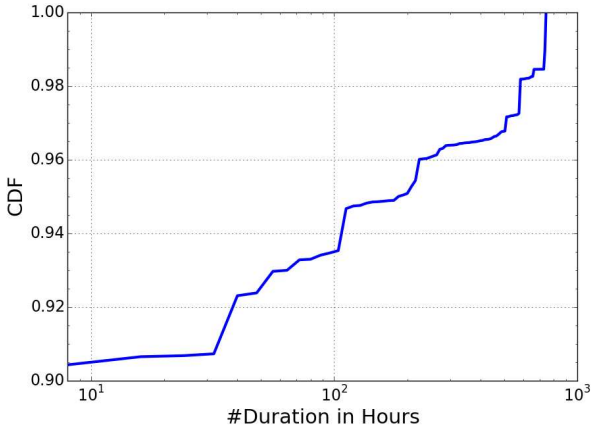


Figure 3.14: Distribution of detour duration

transient detours as they shed light on misconfigurations or even malicious activities, both aspects of routing we understand less.

3.5.2 Transient and Flash Detours

We first present an understanding of the transient detours on per-country basis. Since there are more than one peers in some countries and different peers see varying number of transient detours, we calculate an average number of transient detours per country by dividing total number of transient detours in a country by number of peers in the given country. This average value per country is presented in Figure 3.15. We detected transient detours in only 8 countries where Russia topped the list. In comparison to Figure 3.10 Italy and India showed more average number of transient detours than US. Figures 3.16 and 3.17 show a distribution of ASes that initiate detours and prefixes affected by detours. We observe that 4 ASes originate 50% of the transient detours and only 30 prefixes account for 50% of the transient detours.

Similar to Table 3.5, shown in Table 3.7 are the most common transient detour origins and Table 3.8 shows top impacted prefixes by transient detours. AS9002, RETN-AS, started the most number of transient detours in our dataset. We note that in *ASWatch* [29] authors gathered ground-truth data from security blogs which enlisted AS9002 as a malicious AS. Another previously know malicious AS that appeared in our findings was AS49934 as a detour destination for 7 Russian

Table 3.7: Top Transient Detour Origin ASNs

Transient Detour Origin AS	Total %	Frequent Detour Destination AS	% to frequent destination
9002 (RETN-AS RETN Limited,RU)	22.64%	2914 (NTT America)	99.07%
6939 (Hurricane Electric,IT)	10.94%	8551 (Bezeq International)	100%
1299 (TELIANET,IT)	10.87%	8708 (RCS-RDS)	100%

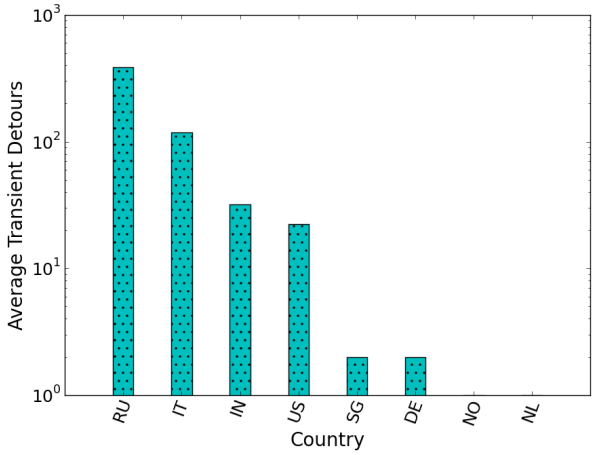


Figure 3.15: Average number of transient detours per country

prefixes. AS49934 is currently unassigned. It was assigned in Ukraine between 2009-10-14 and 2016-01-03 and was known to announce bogus prefixes and host bots.

Finally, we look at *flash* detours. These are detours that appeared only once and were observed in only one RIB of a peer. Flash detours account for 26% of the transient detours, 328 prefixes (6% of all prefixes that suffered detour) experienced at least one flash detour.

Owners of the prefix which suffered flash detours might be interested to know such findings. While 328 prefixes suffered flash detours in our dataset, due to space limitation we point out a few interesting ones in Table 3.9.

The list in Table 3.9 raises serious concerns. Data from government agencies, banks, insurance companies can easily be subject to wiretapping once it leaves national boundaries. Based on our control-plane only data, it is not possible to verify if these institutions were attacked or not. Nevertheless, we believe our findings will motivate network operators to look more closely into why their prefix detoured and if they intended it to happen.

Table 3.8: Prefixes affected the most by transient detoured BGP paths

Prefix Affected	Total %	Frequent Detour Destination AS	% to frequent destination
178.79.218.0/23 (Limelight Networks, Inc, IT)	5.5%	8551 (Bezeq International, IL)	100%
185.19.164.0/22 (Digi Italy S.R.L, IT)	5.5%	8708 (RCS-RDS, RO)	100%
46.21.30.0/24 (Tekka Digital, IT)	5.5%	8758 (Iway, CH)	67.08%

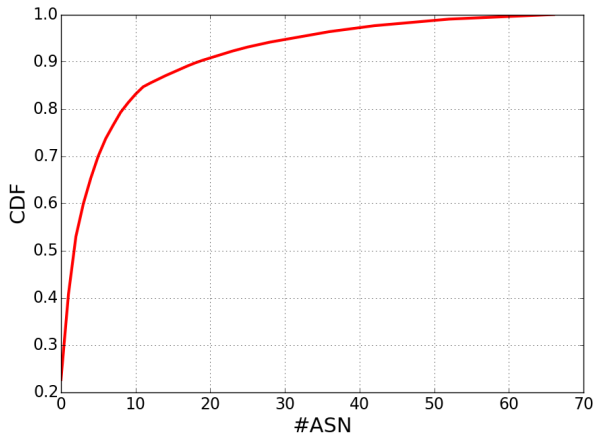


Figure 3.16: Distribution of ASes that originated a transient detour. The top 4 Detour Origin ASes account for 50% of all transient detours

3.6 Quantifying Geolocation Challenges

We presented AS Geolocation in Section 3.2 and used it to detect detours in Section 3.3. As with any measurement system, clean input data leads to better accuracy in detour detection as well. A key input to AS geolocation itself is infrastructure (routers , switches) geolocation. Here, we aim to quantify errors that inaccuracies in geolocation that might impact accurate detection of detours.

3.6.1 Evaluating Geolocation DB Accuracy

We first evaluate if the common belief, that infrastructure geolocation from public geolocation databases is not accurate, is correct or not. We compared geolocation routers with known locations (ground truth same as [33]) to the geolocation provided by Maxmind Free DB, Maxmind Paid DB, IP2Location (free) and NetAcuity (paid).

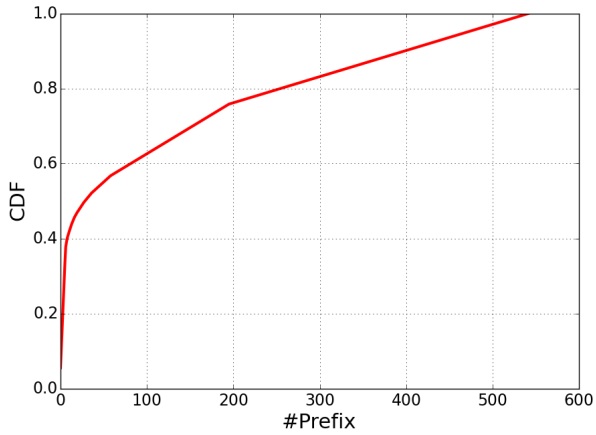


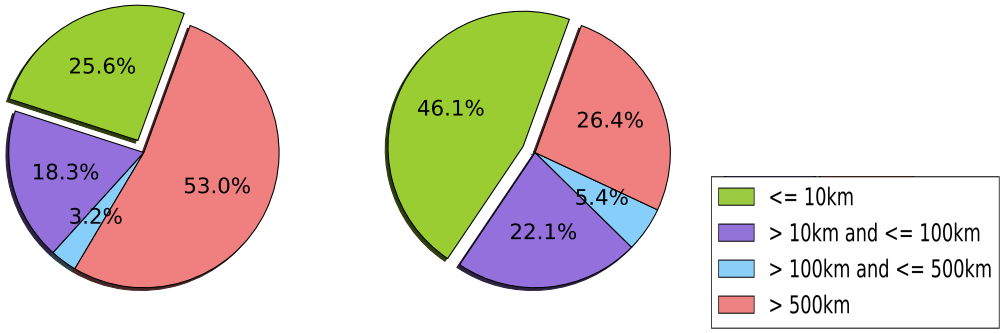
Figure 3.17: Distribution of prefixes that experienced a transient detour. About 30 prefixes account for 50% of all transient detours

Table 3.9: Some prefixes affected by flash detours

Prefix Affected	Owner	Detour Destination
170.61.199.0/24	Mellon Bank, US	28513 (Uninet, MX)
192.230.0.0/20	Washington State Department of Information Services, US	7660(Asia Pacific Advanced Network, JP)
212.11.152.0/21	Moscow Mayor Office, RU	2603(NORDUnet, NO)
208.79.7.0/24	Security Equipment Inc, US	53185(William Roberto Zago, BR)
161.151.72.0/21	The Prudential Insurance Company of America, US	2510(Infoweb Fujitsu, JP)

Once we obtain the latitude-longitude (lat-long) from the databases, we compute the distance between the ground truth lat-long and DB lat-long. If the database geolocation is correct, this distance should be close to zero.

Based on the distance metric, we create four categories. Less than 10km, 10km to 100km, 100km to 500km and more than 500km to quantify accuracy. If the distance is less than 10km, then DB geolocation is correct. In Figure 3.18a and 3.18b we present these results for Maxmind free and NetAcuity. We observe that commonly used Maxmind free DB is correct at city level only for 25% of the IPs. Commonly used commercial DB, NetAcuity, has much higher accuracy with 46% of the IPs in the less than 10km category. Shown in Figure 3.19, we also compared OpenIPMap vs other DBs and ground truth. The distribution of distance shows that for router geolocation, OpenIPMap is almost always correct at the city level. About 90% of the routers are geolocation within 10km of known location.



(a) Maxmind Free vs Ground Truth **(b)** NetAcuity vs Ground Truth

Figure 3.18: Comparison of Maxmind Free DB and NetAcuity with ground truth.

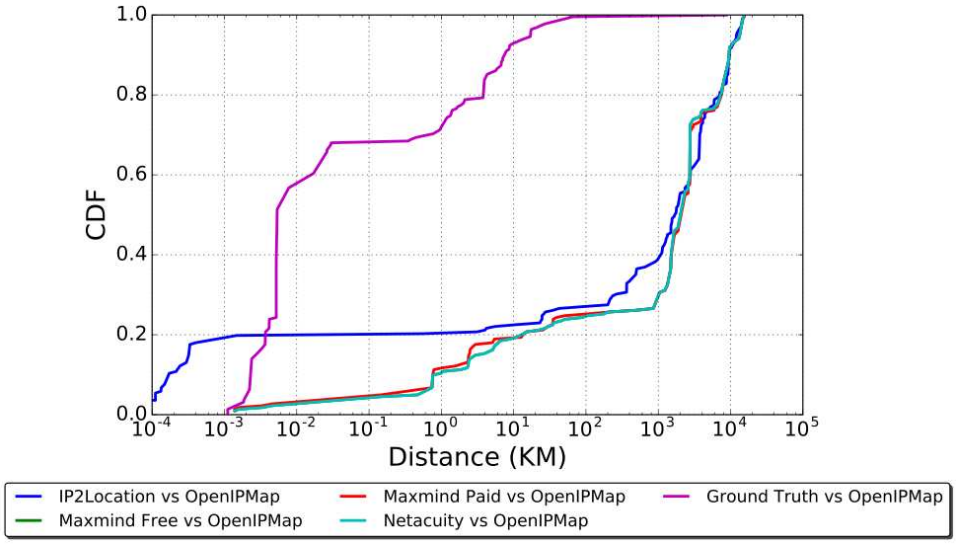


Figure 3.19: CDF showing performance of `OpenIPMap` in comparison to ground truth and other DBs.

This analysis shows that databases are fairly accurate at larger distance radius, for example, greater than 500km, but they do not capture infrastructure geolocation very accurately. In [33] authors provide a detailed evaluation of accuracy of databases across different regions and make recommendations to keep in mind while working with geolocation data. We recommend using OpenIPMap and NetAcuity for a productional and/or commercial system. Another possible direction one might pursue is running measurements to target IPs and geolocating using triangulation. While developing a robust active measurement geolocation system is out of scope of this dissertation, we present some preliminary results using RIPE Atlas to evaluate the efficacy of this approach.

3.6.2 Active Measurements

We begin by using an existing active measurement technique, Constraint Based Geolocation (CBG) [34] on the Atlas infrastructure.

CBG relies on the speed of light in fiber to estimate the location of a target IP from a given vantage point (probe). For example, if the RTT of a target from a probe is 1ms then assuming fastest possible fiber link between them, the target has to be within a 100km radius from the probe (2/3c speed in fiber, where c is the speed of light in vacuum). Overlapping such constraints from multiple probes, CBG obtains an intersection region which is the estimated geolocation of the said target.

To apply this method on Atlas infrastructure we develop `AtlasCBG`. This is the first implementation of CBG that seamlessly works with RIPE Atlas API calls, making it capable to run and process thousands of measurements in parallel[3].

For each target router IP, we first need to select Atlas probes that will provide best measurements. There are currently 10K probes, we select 25 probes for pinging each router. Selection of probes for each target involves using geolocation of the router and fetching nearby probes. We gather closest 25 probes using information from NetAcuity, Maxmind Paid, Maxmind Free, and IP2Location. Note that this information is just used a starting point to start active measurements. In an ideal case, measurements from these probes should overlap giving us an intersection region where the target is located. In reality, however, some constraints don't overlap. In such cases, we pick the intersection region where most probes agree.

Finally, to test the success of `AtlasCBG`, we geolocated about 500 router IPs from our ground truth dataset. We compute the distance between the lat-long obtained from CBG and the 4 databases vs the ground truth lat-long. In Figure 3.20 we show a CDF of the number of IP with corresponding distance. We observe that `AtlasCBG` outperforms not only the free databases Maxmind Free and IP2Location but also performs better than Maxmind Paid. NetAcuity, however, still performed better up to the distance of less than 100km.

[3]RIPE NCC graciously allowed this project to run 4000 concurrent measurements

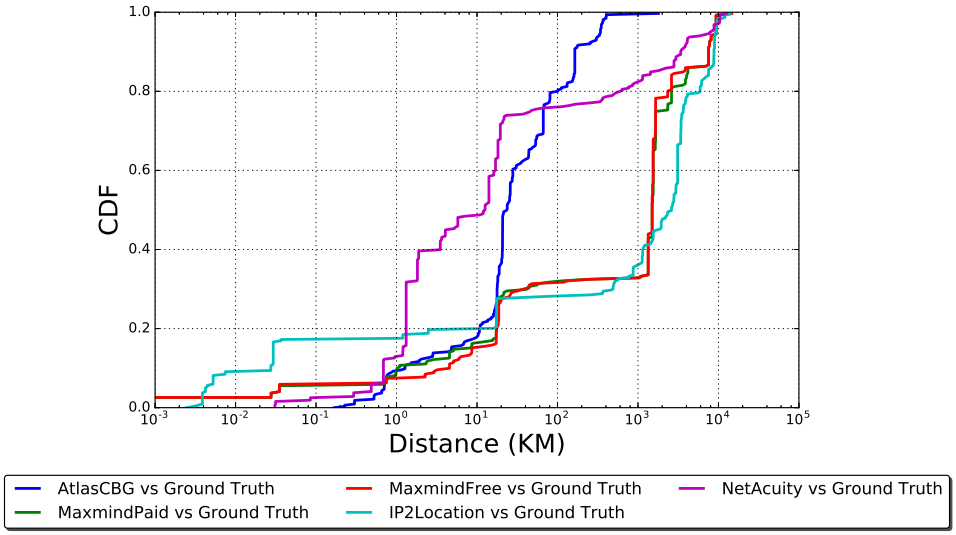


Figure 3.20: CDF showing performance of `AtlasCBG` in comparison to other databases.

We recommend complementing the databases with CBG geolocation thus providing better geolocation for at least some of the router IPs where Atlas probes are located nearby and performing CBG is possible.

Dataset Contributions

We make the geolocation and detours detection data available to the community via a public RESTful API interface. The motivation to do so is as follows. 1) Network operators can easily query our database and check if their prefix suffered a detour. 2) Internet measurement researchers can use this information to study various BGP anomalies such as route leaks, detecting malicious ASes, etc. Our results on AS and prefix geolocation are available at `http://geoinfo.bgpmon.io` and detours results can be accessed at `http://detours.bgpmon.io`.

3.7 Discussion

We present a first attempt to characterize detours in the Internet. We sampled BGP routing tables from 416 peers around the world over the entire month of January 2016 to investigate international detours. We see about 18.9K distinct entries in RIBs that show a detour. More than 90% of the detours last less than 72 hours. We also discover that a few ASes cause most of the

detours and detours affect a small fraction of prefixes. Some detours appear only once. Our work is the first to present different types of detours, namely, persistent and transient. We also present novel insights on their characteristics such as detour dynamics in different countries, top impacted prefixes and detour origins.

Based on our results, we believe that it will be hard to solve this problem without substantial data plane monitor deployment to corroborate control plane measurements. ISPs and IXPs may be required to install sophisticated data plane probe infrastructures and geolocation databases may have to become far more accurate for infrastructure IP addresses in order to detect international detours with good accuracy. Control plane monitoring is still very important as it provides efficient global monitoring and can immediately flag potential anomalies where data plane monitoring should be directed. Our work shows that it is effective and should be expanded.

3.8 Summary

This work sparks the conversation about the challenges new regulatory frameworks will pose to researchers, industry, and network operators. We investigate only a small part of the problem, namely finding the subset of paths where we can detect international detours with some confidence. We provide some answers, but also bring attention to the problem and will hopefully stimulate more work in this new direction. The gauntlet is thrown and we expect a lot more research in this area.

Within its scope, we believe this work is executed carefully by taking into account measurements from both control and data planes. We show that for the cases were able to study there is agreement between the two planes. This is a significant result. Equally significant, we also illuminated the difficulties in expanding the scope within the existing measurement infrastructures. One of the main difficulties we encountered for example, is finding measurement points with both control (BGP peers) and data (RIPE probes) monitors to correlate results. This problem cannot be easily solved, it would take substantial effort to scale the existing infrastructures by an order of magnitude or more. Another important obstacle is lack of knowledge about peering relationships

between ASes. This is also a hard problem to solve, since such relationships are not readily disclosed. It is interesting, however, to contemplate the issue if regulatory requirements require such disclosures.

Routing detours are one specific case of routing anomalies where users still have connectivity but suffer degraded performance. In the next chapter, we focus on loss of connectivity i.e., outages. We design and build a system to quickly detect outages and characterize its impact on routing.

Chapter 4

Internet Outage Detection

Outages pose a challenge for day-to-day operations of billions of users and businesses, unsurprisingly, outage detection has been studied from many angles, each providing a partial view of what is going on. We approach outage detection from a new perspective, `Disco`, a detection technique monitoring existing long-running TCP connections to existing infrastructure and identifying bursts of disconnections. The benefits are considerable as we can monitor Internet wide swaths of infrastructure that are not responsive to ICMP probes, are behind NATs, etc. without adding a single measurement packet to the traffic. We use RIPE Atlas probes' connections to their management infrastructure that are logged to report probe availability. The small footprint of this data and their availability in a publicly accessible live stream make light-weight near-real time outage detection possible. As disconnected probes continue to traceroute to DNS root servers and Atlas anchors we obtain a no cost advantage of viewing the outage inside out as the probes experienced it, characterizing the outage after the fact. We studied historical probe disconnections from 2011 to 2016 and report on the 443 most prominent outages. To validate our results we inspected traceroute results from affected probes and compared our detection to that of Trinocular [36].

4.1 Data Sources

We use separate data sources for detection, validation, and classification of outages as shown in Table 4.1. All data we use is public and pre-exists, meaning that it is gathered for other prior purposes to our work.

Detection Datasets

To detect outages we use the RIPE Atlas infrastructure [12], which is a global deployment of "probes" (tiny Linux machines) that continuously perform measurements such as pings and

Table 4.1: Datasets: Outage Detection

Usage	Name	Sources
Outage Detection	Probe Metadata	RIPE Atlas
Outage Validation	a) Pings b) Traceroutes	a) Trinocular (USC/ISI) b) RIPE Atlas
Outage Characterization	Traceroutes	RIPE Atlas

traceroutes. While there were more than 9,300 probes active in 178 countries on October 14, 2016, many more probes came and went during the study.

RIPE Atlas probes receive measurement commands and send measurement results back via SSH connections over TCP port 443 to a set of servers called "controllers". At probe boot, a connection is established to a single controller, which records the connect event. SSH keep-alives are used to check if both sides of the connection are up. If a controller finds a probe unresponsive for more than a minute, it tears down the TCP connection and records a disconnect event for that probe. The set of dis/connect events is available through the RIPE Atlas API both as an historic dataset [37] and a real-time stream [38], and is the sole input of our outage detector. In this work, we use data from 2011 to 2016 to detect bursts of disconnects and diagnose outages.

Validation and Characterization

For characterization we use traceroutes from RIPE Atlas probes. For validation we use the same traceroute dataset as well as pings from Trinocular [36].

RIPE Atlas probes frequently run traceroute measurements. When probes are not connected to controllers, they buffer the traceroute results for up to six hours and send it to a controller on the next successful connection. We take advantage of this to compare traceroutes before and during the detected outages. If the probe indeed experienced an outage, the buffered traceroute would not be complete i.e., would not reach the targeted destination during that time.

The periodic traceroutes we rely on are the "built-in" and "anchoring" measurements. These are available as public data to anyone.

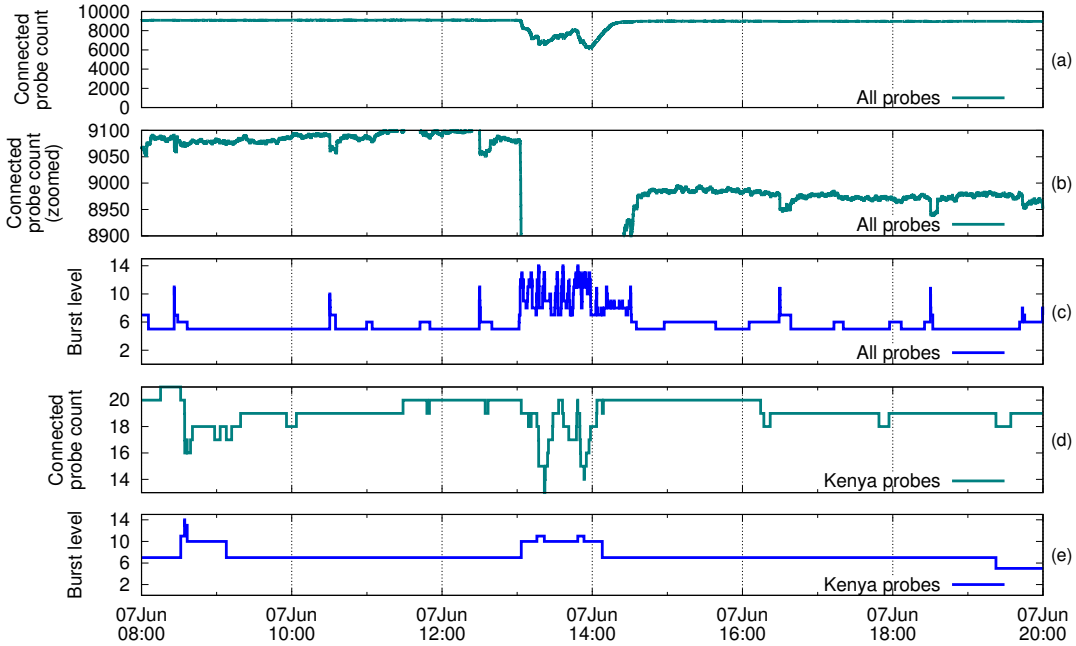


Figure 4.1: Example showing raw connected probe counts and using country sub-stream of results obtained from the burst model with disconnect events reported on June 7th, 2016.

Built-ins are traceroutes to DNS root servers (mostly anycast) while anchoring measurements are traceroutes to Atlas anchors located in various parts of the world. We use these traceroutes to validate that detected bursts of disconnects correspond effectively to outages (Section 4.3) and to characterize the detected outages (Section 4.4).

The second dataset we use for validation is that of the Trinocular project. From four geographically diverse vantage points, Trinocular pings four million /24s to track responsive blocks. Based on the responses Trinocular infers the state of a /24 prefix as up, down or uncertain using the methods explained in [36]. In particular, we used the processed 'outage adaptive datasets' [39] from April 2015 to December 2015. This gives us an external source of information to validate the unavailability of prefixes. However, out of the 10.5K /24s where RIPE Atlas probes are located, only 7.9K /24s are monitored by Trinocular and 2.6K /24s are not. Using RIPE Atlas probes for outage detection, we monitor this previously unseen address space without generating any probing traffic.

4.2 Outage Detection

We process disconnection data in three main steps: (1) model the arrival rate of probe disconnects (2) perform burst detection on sub-streams of data and (3) report significant events. The next subsections explain each in detail.

4.2.1 Burst Modeling

We define an outage as bursts of disconnects from a certain geography or topology. The significance of a burst is characterized by the number of disconnects within a very short time frame. The temporal characteristics of bursts are poorly modeled by simple time series of the number of disconnects. Indeed counting disconnects in time bins conceals the exact temporal relations between consecutive disconnects. Disconnects that are uniformly spread out through a time bin should be considered differently than synchronous disconnects. To account for the temporal structures of bursts we employ an infinite-state automaton [40] that models bursty activities in a stream of inter-arrival times $x = (x_1, x_2, ..., x_n)$, which represents, in our case, disconnect events in a period of time T. Each state q_j is associated with an exponential density function $f_j(x_t) = \alpha_j e^{-\alpha_j x_t}$ with rate $\alpha_j = \frac{n}{T} 2^j$. Therefore, with $0 \leq i < j$, the state q_j represents higher intensity bursts than the ones modeled by q_i. Kleinberg [40] proposes to find the optimal state sequence corresponding to a given stream by adapting the Viterbi algorithm [41] to the following model and cost function. $C_j(t)$ is the minimum cost of a state sequence ending with q_j for the input $x_1, x_2, x_3, ..., x_t$ and is defined by the recurrence relation:

$$C_j(t) = -\ln f_j(x_t) + \min_l (C_l(t-1) + \tau(l, j))$$

with initial conditions $C_0(0) = 0$ and $C_j(0) = \infty$ for $j > 0$, and τ the cost of transitions between states which is positive for transitions to higher states but null on return to lower states (see [40] for more details).

One difficulty with Kleinberg's original formulation is comparing results obtained from different streams. This is because, as explained in the next section, we separate the global stream of

disconnects by country, AS and geolocation into multiple sub-streams and expect burst level j to represent the same disconnect arrival rate for every stream. However, as arrival rates α are function of the number of events and the time duration of the stream (respectively n and T), comparing burst levels from different datasets is meaningless. We modify the original burst model in order to obtain uniform states regardless of the input data by fixing constant values for n and T. The variable T is set to one day and the variable n is set to the average number of daily disconnects observed in RIPE Atlas. As we found about one daily disconnect per probe, n is equal to P, the number of probes active during the analyzed time period. Thereby, burst levels are stationary even if the number of active probes increases.

Given a sequence of disconnect timestamps, the above algorithm produces a sequence of burst levels such as the ones in Figure 4.1(c) and (e). Each level represents the disconnect arrival rate normalized by the number of probes in the analyzed stream. High burst levels indicate points in time where the rate of disconnects has greatly increased, hence helping to identify outages.

As we rely on the status of TCP connections to detect network outages and there are a myriad of causes of end-point failures, the analyzed streams are particularly noisy. Our burst model is, however, inherently robust to such noise as the cost of transitions to upper states increases exponentially.

Figure 4.1 illustrates the number of connected probes along with the corresponding burst levels for June 7$^{\text{th}}$ 2016. Two known events have happened on this day, a large Kenyan power outage occurred at about 08:30 UTC, and most of the RIPE Atlas controllers have been rebooted from 13:00 UTC. The burst levels computed with all probes (Figure 4.1(c)) exhibits clearly the controllers reboot, but not the Kenyan outage. Instead we found bursts that appear every two hours throughout the entire day. Our manual inspection of these periodic bursts revealed times of intense measurements on v1 probes causing the reboot of the highly loaded probes. These bursts correspond to meaningful events affecting the RIPE Atlas platform itself that we have reported and was acknowledged by RIPE NCC. Since the number of impacted probes is relatively small, these

synchronized disconnections are particularly hard to identify from the count of connected probes. This example illustrates the benefits of the temporal analysis of our burst model.

Due to the low number of probes in Kenya, neither a drop in probe count (Figure 4.1(a) and (b)) nor a burst is visible for the Kenyan outage (Figure 4.1(c)). Note that the first significant burst around 08:30 UTC in Figure 4.1(c) is from the v1 probes rebooting not the Kenyan power outage.

By looking at data only for probes in Kenya (raw data shown in Figure 4.1(d)), the power outage is easily identifiable through a burst of level 14 (Figure 4.1(e)). In addition, by using this *'sub-stream'* the controller reboots are less emphasized because Kenyan probes are connected to different controllers hence have been disconnected asynchronously. This example illustrates again the benefits of the burst model. Although there was more Kenyan probes disconnected during controllers reboot a higher burst level is reached during the power outage as probes have been synchronously powered down. Thanks to our modification of Kleinberg's burst model, burst levels with the entire stream and the Kenyan one are very comparable thus making detecting relevant events very easy.

4.2.2 Sub Streams

It is evident from the previous example that splitting the main stream of disconnections into *sub-streams* helps to mitigate the disparate probe deployment. A sub-stream is a collection of probes that share one of the following geographical or topological characteristic: the same country, the same AS or the same geographical area (within 50km radius). Therefore, a single probe appears in three sub-streams, a country, AS, and GeoProximate sub-stream.

We justify the choice of these sub-streams as follows:

a) *country sub-streams*: The Atlas probes are distributed among different countries disproportionately. By aggregating probes that belong to the same country and analyzing them separately allows us to detect impairments in a country even if it has deployed a small number of probes.

b) *AS sub-streams*: An AS sub-stream enables us to focus on a certain part of the network topology. This is particularly interesting to network operators aiming at monitoring connectivity of their ASes even across large geographical area.

c) *geoProximate sub-streams*: The goal of this sub-stream is to detect bursts of disconnections that may belong to different ASes or even countries but are geographically close to each other. This enables us to pinpoint the geographical effect of an outage. We believe outages due to natural calamities or power outages where entire cities loose connectivity stand out here. We chose to cluster probes that are within a 50km radius. Past research has suggested using 40km to 50km radius as a city-level assumption [42].

Depending on which type of sub-stream an outage is seen, we can identify additional information about the outage. For example, in a recent power outage in Amsterdam [43], many probes belonging to different ASes suffered disconnection. This was caught by geoProximate sub-streams. Similarly, we found cases where probes of same AS, but far away geographically, suffered an outage. These were caught by country and AS sub-streams. We revisit these examples, that emphasize advantages of sub-streams, in Section 4.5.

4.2.3 Outage Reporting

Once burst model is applied on multiple sub-streams we detect large events by applying a threshold and then aggregating events that appeared in different sub-streams but can be part of the same event. Next, we elaborate on each.

Burst Threshold

To systematically distinguish significant bursts from arbitrary disconnects, we consider only bursts that reach a certain level. Manual inspection of results obtained with the modified burst model shows the burst level usually fluctuates between 4 and 8, and goes beyond 10 during outages. In Figure 4.2 we show the number of detected events using different threshold values. We observe a significant drop in the number of detected events from threshold 8 to 10. Higher threshold values permit to detect the most significant events at the price of missing interesting outages. We recommend to use threshold values between 10 and 14. In the rest of this work we set this threshold to 12 as a mid value to keep a justifiable trade off. The precision and recall using this threshold is discussed in Section 4.3.

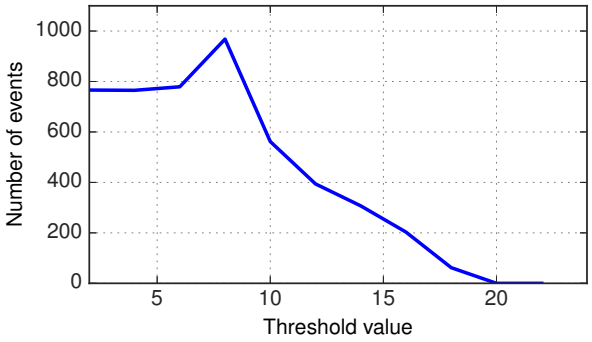


Figure 4.2: Number of reported events for different threshold values.

A burst of disconnects reveals the start of an outage but its time duration is unknown. To estimate the end of an outage, we select all probes disconnected during the detected burst and retrieve their corresponding re-connect events. Intuitively, the first reconnect signals the end of the outage, but some probes might be wrongly accounted for as they inadvertently get disconnected during the outage. We assume that at least 50% of the probes involved in the burst are affected by the outage, consequently, we define the end of the outage as the median re-connect timestamp of the disconnected probes.

Aggregation

The final step of outage reporting is aggregation. As described previously we split disconnections by country, AS or geolocation. Therefore, a burst might appear in multiple sub-streams. For example, during an outage in *AS1* which is located in country *C*, probe disconnect bursts will appear in both sub-streams, *AS1* and *C*. We group detected outages that appear in different sub-streams and start within the same 30 minute window and end within the same 30 minute window. We discard events that start and end within the same 30 minute window, hence deliberately discarding transient events such as a controller reboot. According to information obtained from RIPE NCC the reboots are usually in the order of few minutes.

4.2.4 Detection Results

After detecting burst events in the raw metadata of disconnect and connect events in the AS, country, and geoProximate sub-streams we perform aggregation as described above. This gave us

443 large outages between 2011 to 2016. We detected outages due to faulty maintenance issues that gained press attention: The Time Warner Cable outage on August 27th 2014 [44], the AMS-IX outage [9] and Kenyan power failure [45]. We also detected recurring outages in Benin and Andorra which were not in the limelight. We validate these results in the next section and further characterize these 443 outages in Section 4.4.

Performance

The proposed burst model has a time complexity of $\mathcal{O}(ns^2)$ where n is the number of disconnect events and s is the number of states in the implemented automaton. Our implementation (python2.7) takes 103 minutes to analyze all types of sub-stream for 6 years of historical disconnection data from RIPE Atlas. This execution was done on Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz with 24 processes in parallel. Each process takes less than 500MB of RAM. It can easily run on the live feed of disconnection data and raise alarms for events within a few minutes of it's occurrence.

Applicability

We study TCP connections between the RIPE infrastructure and all Atlas probes deployed in 3.3K ASes. The scope of our study is hence limited to the Atlas platform deployment, but `Disco` could be deployed in larger infrastructures with long-running TCP connections, for instance large scale video platforms.

In addition, Atlas controllers are updated and rebooted once in a while, breaking numerous TCP connections at once. For our study we empirically chose a minimum outage duration (30 minutes) much larger than the time observed for controller reboots, thus avoiding to report these irrelevant events. This would not be necessary if the planned events were announced in the stream of data.

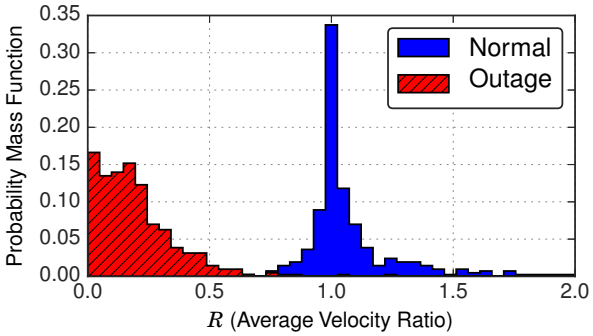


Figure 4.3: Distribution of Average Velocity Ratio for normal and outage durations.

4.3 Validation

We validate the results of our outage detector using two distinct data sources: traceroutes from the probes involved in the disconnect bursts and Trinocular [36] ping data. The first dataset is used to check whether traceroutes sourced by the affected probes are also affected. The second dataset has the advantage of being completely independent from our experiment's infrastructure.

4.3.1 Comparison to Traceroutes

Table 4.2: Outages reported by Disco vs. Trinocular

#Outages (out of 53)	Percentage of /24s also reported by Trinocular	R_outage (Average velocity ratio during outage)	Average outage duration (in hours)
23	100%	0.16	1.23
10	71-99%	0.22	1.39
4	51-70%	0.48	1.15
7	1-50%	0.31	1.13
9	0%	0.25	0.8

To measure the impact of outages with traceroutes, we define the velocity of an Atlas probe as the number of complete traceroutes (i.e. traceroutes that reached the intended destinations), x, per unit of time, namely, $\bar{v} = \frac{\Delta x}{\Delta t}$. Correspondingly, $\bar{V}$, is the average velocity for the n probes in a certain sub-stream and is defined as, $\bar{V} = \frac{1}{n}\sum_{i=1}^{n} \bar{v_i}$. Thereby, the ratio $R = \frac{\bar{V_T}}{\bar{V_R}}$ of the velocity for a testing ($\bar{V_T}$) and a reference ($\bar{V_R}$) period of time is the relative success of probes completing

traceroutes. We assume similar velocity for two periods of time representing normal operation; thus the value of R is expected to be around 1. During an outage, however, the average velocity is drastically decreased, hence, R is expected to be close to 0. This enables us to determine whether the events singled out by the detector correspond to outages.

We calculate *R* for the 443 intervals where `Disco` detected an outage versus normal operation for the same set of probes. Figure 4.3 shows that under normal conditions R is usually 1 while during an outage far fewer traceroutes complete, giving R values closer to 0. In Figure 4.3 we observe that all selected normal periods of times have velocity higher than the midpoint value, $R = 0.5$. Using this midpoint we determine if a reported event is considered as true positive ($R \leq 0.5$) or false positive ($R > 0.5$), and we obtain a precision of 95% for `Disco`.

4.3.2 Comparison to Trinocular

For a radically different view, we compare outages detected by our method with the Trinocular outage survey dataset. `Disco` detects 53 outages from April to December 2015, the period covered by both Trinocular and the RIPE Atlas metadata of interest. We compare the affected network prefixes revealed by `Disco` to the Trinocular observations for these prefixes.

First, we extract prefixes of the probes that belong to bursts of disconnect events. Then we query the Trinocular data for the same /24 prefixes to check if an outage was detected at the same time window.

We note that our methodology also detects prefixes that are part of an outage that Trinocular could not probe as they are unresponsive to ICMP traffic. Out of 851 /24s (from the 53 outages in consideration) detected by `Disco`, only 365 (43%) were pinged by Trinocular. This shows the potential of our method to detect outages in places where state-of-the-art active probing cannot reach. Moreover, due to use of *aggregations*, unlike Trinocular, `Disco` not only provides the prefixes that suffered an outage but also points out set of prefixes that are part of the same outage. Only the common 365 /24s from 53 outages could be considered for further validations.

We observe in Table 4.2 that for the /24s reported by Disco and also probed by Trinocular, both detectors agree on the down status of all the prefixes for 23 out of 53 (43.4%) outages (top row). There are some cases where Disco detected some prefixes affected by an outage but Trinocular did not. In general, about 62% (33 out of 53) of outages agree on the status for more than 70% of the prefixes. Surprisingly, there were 9 outages (16.98%) reported by Disco, with prefixes probed by Trinocular, that are not reported by Trinocular (see Table 4.2 bottom row).

As explained above, we look at pre-existing traceroutes from probes to check if they indeed lost connectivity during these outages. As shown in the third column of Table 4.2 the drop of average velocity (R) during these periods of time indicates that these are indeed outages missed by Trinocular. We verified that low R values are obtained for the 9 events. In addition, the last column of Table 4.2 shows that the average outage duration for these cases is significantly lower than other cases, suggesting that Trinocular missed short-lived outages.

We also investigated 32 events reported by Trinocular but not detected by Disco. We obtained these 32 events by extracting outages for all /24s visible to both detectors from Trinocular dataset. Then we aggregated outages of several prefixes into events if they started and ended in the same 30 minute window (same aggregation metric we used for Disco). The average velocity R for these events suggests that Trinocular reports 9 false positives. As Disco has reported 47 true positives (and 6 false positives) but missed 23 events found by Trinocular the recall in 2015 (April to December) is 67%. We note that the 23 events missed were due to their non-bursty nature, a use-case that Disco is not designed for.

4.4 Outage Characterization

In this section we analyze the traceroutes initiated by probes to gain more insight into the outages we detected. Unlike previous work, with a view from outside the affected networks, we present insights from the inside, i.e., the way probes saw the outages. We use the same traceroutes as in Section 4.3. The traceroutes are sourced by probes identified in the outages. They are buffered by probes waiting for communication to a controller to be reestablished.

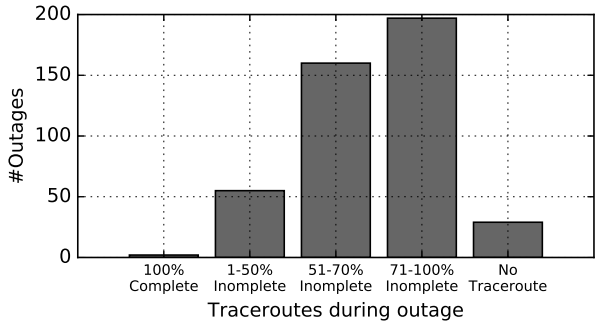


Figure 4.4: Number of outages with complete, incomplete, or no traceroute.

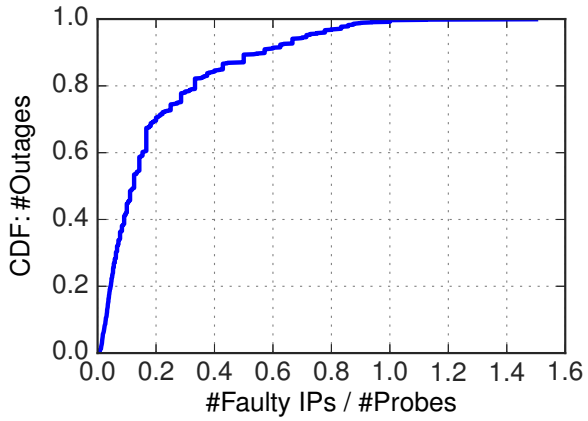


Figure 4.5: CDF: Ratio of unique faulty hops to the number of probes, low ratio indicates incomplete traceroutes ending at the same hop during an outage.

The first symptom of interest is a complete lack of traceroutes reported on reconnection. We believe this is an indicator of a power outage. This is verified in the case of the power outage in Kenya (Figure 4.1) and for a recent power outage in Amsterdam, NL on January 18th 2017 [43]. The probes stopped operating due to the lack of power and did not perform any of the regular scheduled traceroutes.

Second, during outages traceroutes may stop at earlier hops than during normal operation. Figure 4.4 shows the number of outages with percentage of traceroutes during the outage that were either complete, incomplete or had no traceroutes during the outage. Except for 2 out of 443 outages, there is usually a large fraction of incomplete traceroutes during the outage or no traceroute at all. In most cases, 71-100% of the traceroutes conducted during the outage were incomplete. This is a convincing sign that probes lost complete connectivity at the detected time.

For cases with 1-50% and 51-70% incomplete traceroutes, we found that some probes kept a limited connectivity and are still able to reach local targets, such as, anycasted root servers located near the probes.

We also investigated the 2 events where all traceroutes were complete. On closer inspection we notice these events had only 2 traceroutes each conducted during the outages and these 4 traceroutes were within 3 minutes of the outage end estimated by `Disco`. As stated in Section 4.2.3, to estimate the end of the outage we assume that at least half of the probes involved in the burst should reconnect. In this particular case, a few probes got connected and conducted 4 successful traceroutes within 3 minutes before our estimated outage end.

Next, we characterize where incomplete traceroutes end (i.e. inside or outside probes' local ASes), and study incomplete traceroutes due to forwarding loops.

Narrowing down the location of the outage

To identify faulty hops in incomplete traceroutes, we employ the probabilistic model proposed in [46]. In a nutshell, traceroutes before the outage are used to learn the visited IP addresses at every hop and to construct a probabilistic graph of the IP paths for each destination. Given an IP address in this graph we can then estimate what would be the next visited IP address. Thus, by looking at the last hops in incomplete traceroutes we estimate the addresses that are expected on the path. This technique is, however, not able to find unresponsive addresses if the last hop was not discovered during the learning phase, for example, a new route taken during an outage. For 80% of the outages we could estimate faulty next hops for up to 60% of the traceroutes in those outages.

For a given outage, we found that the incomplete traceroutes for each destination (DNS root server or anchor) usually end at the same estimated next hop regardless of the originating probe. For each destination, if traceroutes from n probes were incomplete, we estimate all the next hops and compute the ratio of the number of unique faulty hops to n. If all probes see the same faulty hop then this ratio is low. A CDF of this ratio is shown in Figure 4.5. In more than 80% of the cases, this ratio is lower than 0.35, indicating traceroutes from different probes to a given destination usually failed at the same hop. This, and the fact that the identification of expected hops works

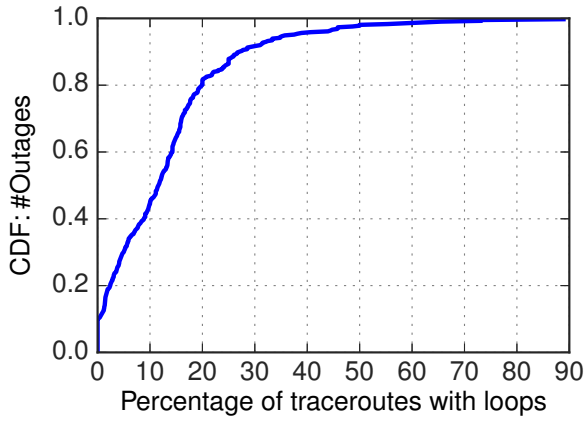


Figure 4.6: Distribution of the percentage of traceroute with forwarding loops per outage.

most of the time, means that in most cases we are able to discover precisely beyond which IP the outage occurred. Using the faulty hops we can also distinguish between outages occurring in the probe's AS from those occurring outside of that AS. Out of all the incomplete traceroutes, 73.5% failed outside the probe's AS and 26.5% within the probe's AS.

Outages with forwarding loops

We were surprised to find numerous incomplete traceroutes caught in *forwarding loops*. These are easily identifiable as IP addresses re-appearing in the same traceroute at different hops.

In Figure 4.6 we show the CDF of the number of outages we detected versus percentage of traceroutes during that outage which showed a forwarding loop. We observe that for 80% of the outages up to 20% of the incomplete traceroutes are caused by forwarding loops.

We keep track of IP addresses involved in loops, namely, pairs of adjacent IPs in the loop. For example, in traceroute {ip1-ip2-ip3-ip2}, {ip3,ip2} characterizes the observed loop. We observe that for 80% of the outages, when a forwarding loop is seen, up to 70% of the traceroutes during that outage see the exact same forwarding loop.

The characterizations above indicate the types of outages our method can detect. ISPs often have means to detect outages in their network but they have limited visibility into what happens in neighboring networks. The ability to detect large outages that occur in another AS, locate

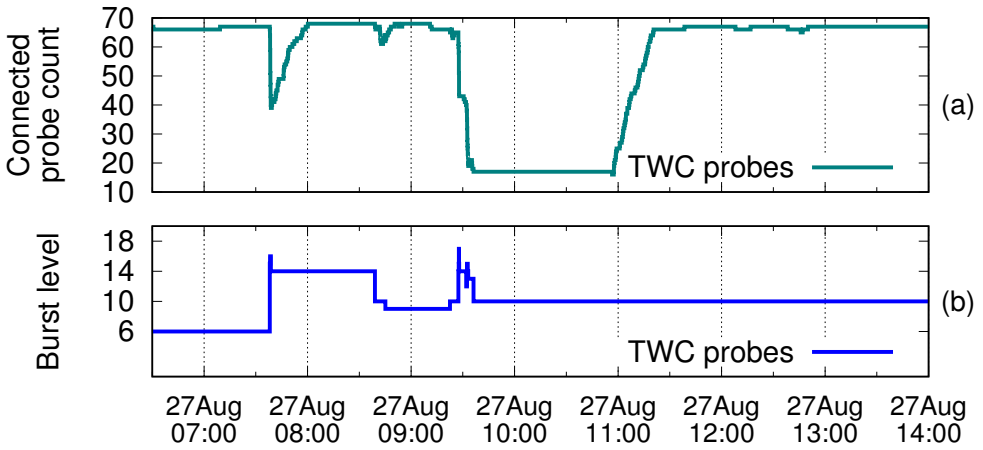


Figure 4.7: TWC outage: Probe counts and corresponding burst levels.

unreachable IPs, and detect forwarding loops motivates the use of `Disco` for ISPs willing to better diagnose reachability issues.

4.5 Case Studies

We focus on two outages in the Time Warner Cable (TWC) network and a power outage in Amsterdam. These use cases serve as examples of what network operators and researchers can learn about an outage using `Disco` on RIPE Atlas data stream.

4.5.1 Outages in TWC

The first outage is identified on August 27th 2014 [44]. During this outage the burst level of 17 was reached, indicating a very sudden drop in number of connected probes (Figure 4.7), this outage particularly stood out in the 6 years of data. It appeared in the AS-level sub-streams of AS10796, AS11351, AS11426, and AS20001, all belonging to TWC. Before the large outage of about 2 hours starting at 09:30 UTC, `Disco` also detected a burst of disconnection at 07:30 UTC. This outage was much shorter than the following one, but, we believe the alert at 07:30 could have been used by network operators as an early warning before the larger outage at 09:30. The outage characterization with traceroutes buffered during the outage reveals that 73% of the traceroutes that failed, suffered a forwarding loop. We also could pinpoint areas of fault by locating the common

failure points of the traceroutes. Probes from Honolulu, Hawaii could reach up to Pittsburgh and probes from LA, San Diego could reach up to San Jose.

Disco also reported an outage on December 27th 2015 for AS11426 sub-stream and a geoProximate sub-stream in North Carolina. TWC announced that a router issue has been identified in this area [47], affecting users' connectivity. This example illustrates the ability of Disco to precisely identify local network connectivity issues.

4.5.2 Outages in Amsterdam

On January 17th 2017 Amsterdam suffered a large power outage. As this area has one of the highest probe densities it is another interesting case-study that shows the benefits of the geoProximate sub-streams. Figure 4.8, shows in red the 56 disconnected probes we detected in geoProximate sub-streams and in green all other connected probes in the area. A large proportion of reported probes is concentrated within the boundaries of the cities affected by the power outage. Interestingly 19 of the probes in that disconnect burst are outside of the city boundaries. All these probes are hosted in a single network. Traceroute data and contact with the network operators revealed that, while these probes stayed physically powered, their Internet connectivity was disrupted between two network elements in the Amsterdam area, coinciding with the Amsterdam power outage. The operators of the affected network speculate that either a network element that terminates user sessions got overloaded by having to disconnect users in the power-outage affected area, or the network between these two network elements, which is opaque to the network operator in this case, got disrupted. The fact that Disco's geoProximate sub-streams emphasized this, shows that we capture real events and interesting side-effects of outages in confined geographic areas.

4.6 Outages in Control Plane

We have looked at outages that appear in data plane, i.e., TCP disconnections or active probing by Trinocular. Outages can also be viewed from the control plane. In such cases the routers (that

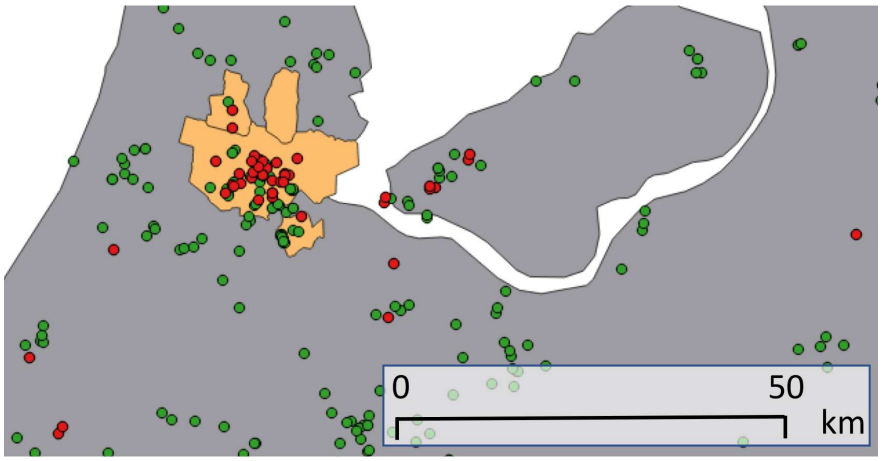


Figure 4.8: Amsterdam power outage. Probes detected using geoProximate sub-streams (red) and connected probes (green).

talk BGP) become aware that a certain prefix which they could reach earlier is now not reachable. In such case, the router makes a *Withdraw(W)* announcement to its neighbors. Similarly, when the prefix again becomes reachable, either via the same AS path as earlier or a new one, the router announces a *Network layer reachability information (NLRI) a.k.a Addition(A)* announcement.

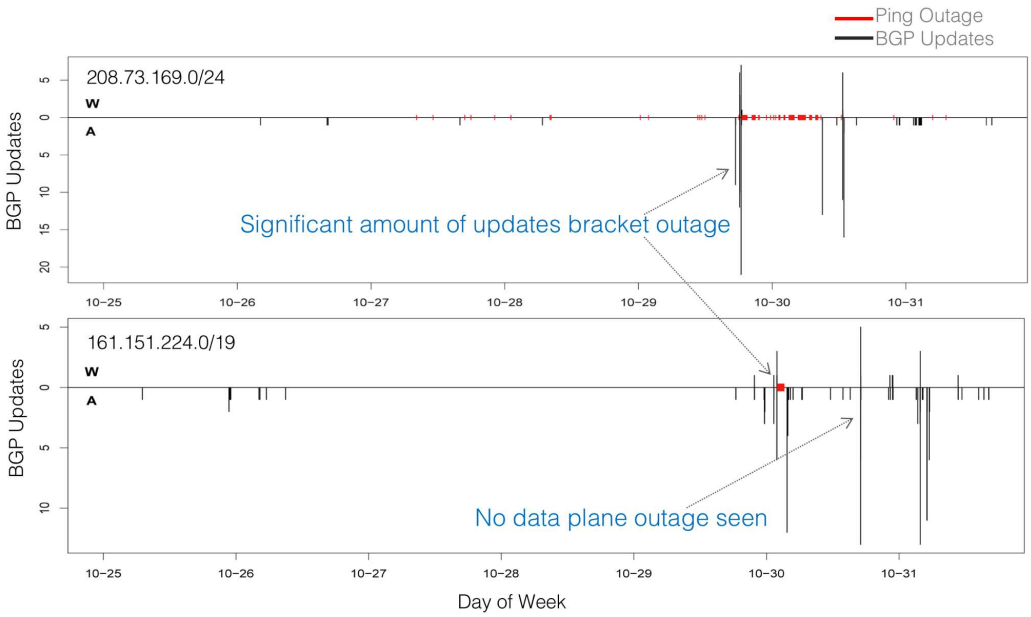


Figure 4.9: A timeline of BGP updates and data plane outage during hurricane Sandy

We study patterns of BGP updates during outages, we studied the view of control plane during Hurricane Sandy (2013). In a few cases, large number of BGP announcements bracket the outages detected in data plane. In Figure 4.9 we show observations from both planes for sample prefixes where we observed this behavior. The time duration in consideration is during hurricane Sandy. The BGP number of updates (W and A) from more than 100 peers on Y axis and markers during ping outage as seen by Trinocular are shown over the course of 7 days on X axis. We see a large churn of withdraws and additions of the prefix from various RouteViews vantage points.

It is not necessary that all outages will exhibit this behavior in both planes. Some outages are only observed in data plane and the contrary is true as well. Previous work [5] has observed correlation between control and data plane outages. Our work is the first to attempt a time series evaluation of them with outages detected using burst modeling on RIPE Atlas data. To achieve this we developed a separate version of `Disco` that can work with previous established work in this domain from CAIDA: `ioda.caida.org`. While more work needs to be done in this direction, the systems developed by CAIDA, and our continued feed of RIPE Atlas based outages can help researchers to investigate further.

4.7 Discussion

One obvious limitation of our application of `Disco` to RIPE Atlas is that RIPE Atlas probes are not ubiquitously deployed in the Internet. While RIPE Atlas is only deployed in 6% of IPv4 ASes, `Disco` could be deployed in larger infrastructures with long-running TCP connections, for instance large scale video platforms.

4.8 Summary

`Disco` is a light-weight outage diagnostic system based on detecting bursts of TCP disconnects and generating no new measurement packets. It allows network operators to monitor over the full extent of their network, from beyond customer premise equipment up to, and including, upstream networks, regardless of whether ICMP probing is allowed in the relevant network. Using `Disco`

we monitored the long-running connections between the RIPE infrastructure and Atlas probes. We found that 25% of the studied /24s are unresponsive to active probing techniques, thus proposed method contributes significantly to the current community outage detection systems.

We go beyond detecting outages. Multi-resolution analysis, i.e. geographical and topological sub-streams, provides information about affected AS, country or city-level radius. This extra knowledge can be a huge asset in localizing areas of impact and potentially reduce response time.

We use existing traceroute data from Atlas probes to validate our results and to characterize detected outages. `Disco` achieves a precision of 95% and detects both outages that happen inside and outside the Atlas probe ASes. Post-mortem analysis of existing Atlas traceroutes not only helps determining common network elements where failure was concentrated but also reveals interesting characteristics such as forwarding loops. Understanding these cases helps to identify routing configurations that may go amok.

This work also opens other interesting research questions such as studying partial connectivity during outages. Visibility into failures that occur outside the probe (customer) AS can reveal where ISPs need to focus on infrastructure development.

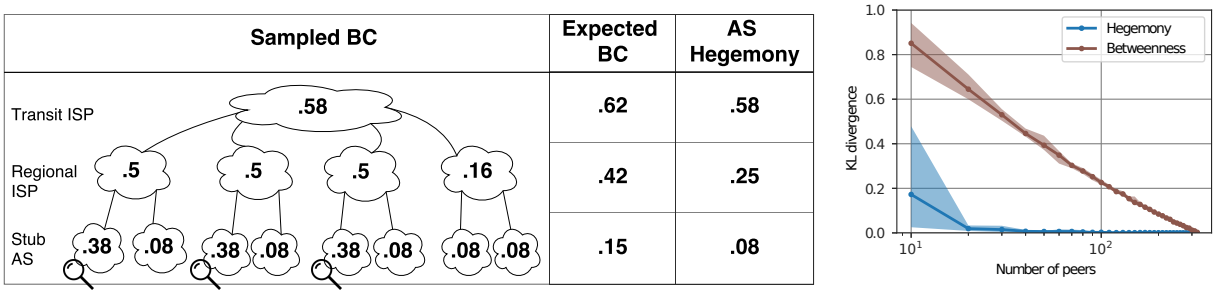
So far we have detected and studied routing anomalies that hamper connectivity. However, both detours and outages are an artifact of larger events that de-stabilize Internet routing. To achieve our goal of near-real time routing events detection there is a need for a system that monitors connectivity at an AS-level. In the next chapter, we develop and evaluate methods to monitor the AS graph while addressing challenges of sparse public data.

Chapter 5

Mapping the AS-Level Connectivity

Inter-domain routing is a crucial part of the Internet designed for arbitrary policies, economical models, and topologies. This versatility translates into a substantially complex system that is hard to monitor. Monitoring the inter-domain routing infrastructure is however essential for understanding the current state of the Internet, tracking unexpected changes, and even detecting potential areas for improving routing. We design a methodology to answer two simple questions: Which are the common transit networks used to reach a certain AS? How much does this AS depend on these transit networks? To answer these questions we digest AS paths advertised with the Border Gateway Protocol (BGP) into AS graphs and measure node centrality (i.e. the likelihood of an AS to lie on paths between two other ASes). Our proposal relies solely on the `AS Hegemony` metric, a new way to quantify node centrality while taking into account the bias towards the partial view offered by BGP. Our analysis using 14 years of BGP data refines our knowledge on Internet flattening but also exhibits the consolidated position of tier-1 networks in today's IPv4 and IPv6 Internet. We also study the connectivity to two content providers (Google and Akamai) and investigate the AS dependency of networks hosting DNS root servers. These case studies emphasize the benefits of our method to assist ISPs in planning and assessing infrastructure deployment. Sudden changes to this metric is often an indication of events such as prefix hijacks or route leaks which could further lead to detours or outages.

We demonstrate the value of the proposed method with 14 years of BGP data (Section 5.4). Overall we found that AS interdependencies in IPv4 are decreasing over time which corroborate with previous observations of the Internet flattening [49]. However, we also found that the important role played by tier-1 ISP is reinforced in today's Internet. The Internet flattening for IPv6 is happening at a faster rate, but we found that Hurricane Electric network is utterly central since the past 9 years. We also investigated the AS dependency of two popular networks, Akamai and Google, showing that their dependency to other networks is minimal although their peering poli-



(a) Simple graph with three viewpoints (illustrated by looking glasses). The sampled BC and AS hegemony are computed only with best paths from the three viewpoints, the expected BC is computed with all best paths.

(b) Sampling error forBC and AS hegemony in function of the number of viewpoints.

Figure 5.1: Comparison of Betweenness Centrality (BC) and AS hegemony with a toy example and BGP data.

cies are completely different. Finally, we look at two networks hosting DNS root servers and show how recent structural changes to these root servers have affected their AS dependencies.

We make our tools and updated results publicly available [50] to enable network operators quickly understand their networks' AS dependency.

5.1 Betweenness Centrality

Following past research, we initially conducted our experiments using Betweenness Centrality (BC) but faced fundamental shortcomings due to the incomplete view provided by BGP data. BC is a fundamental metric that represents the fraction of paths that goes through a node. Intuitively one expects high BC scores for transit ASes as they occur on numerous AS paths, and low BC scores for stub ASes. Formally, for a graph $G = (V, E)$ composed of a set of nodes V and edges E, the betweenness centrality is defined as: $BC(v) = \frac{1}{S} \sum_{u,w \in V} \sigma_{uw}(v)$ where $\sigma_{uw}(v)$ is the number of paths from u to w passing through v, and S is the total number of paths. BC ranges in $[0, 1]$, but the relative magnitudes of the scores are usually more significant than the absolute values.

Challenges

In theory, to compute BC one needs the set of all paths in the graph. With BGP data, however, we are restricted to paths bounded to a small number of viewpoints. We found that this singular type of path sampling greatly impairs BC results. To illustrate this, Figure 5.1a presents a simple

example with 13 ASes and three viewpoints. If we had viewpoints in all ASes, thus access to all paths in the graph, we would obtain the highest BC score for the transit ISP (.62) and lowest scores for the stub ASes (.15). But, using only paths bound to the three viewpoints, the computed BC scores are substantially different (Sampled BC in Fig.5.1a). Because a third of the paths converge to each viewpoint, BC values for ASes close to the viewpoints are undesirably high. This bias is so pronounced that the BC for stub ASes accommodating viewpoints (.38) is twice higher than the BC of one of the regional ISP (.16). Theoretical studies have already reported the shortcomings of BC with sampled data [51], but this issue has been rarely acknowledged in the networking literature. Mahadevan et al. [52] reported that BC is not a measure of centrality when computed with network data, but we stress that this issue comes from the non-random and opportunistic sampling method used to collect BGP data rather than the metric itself.

In our experiments we construct a global AS graph using all data from the RouteViews, RIS, and BGPmon project on June 1st 2016. This corresponds to an AS graph of more than 50k nodes with 326 viewpoints (we consider only full-feed BGP peers), and only 0.6% of all the AS paths on the Internet (16M paths out of the 2.5B). As collected paths all converge to the 326 viewpoints, ASes accommodating viewpoints and their neighboring ASes are seemingly more central than other ASes. To measure the bias obtained with real BGP data we conduct the following experiment. First, we compute the BC for all ASes with data from all 326 viewpoints, then we compare this distribution of BC values to BC values obtained with a smaller set of randomly selected viewpoints. The distance between two distributions is measured with the Kullback-Leibler divergence. Figure 5.1b shows that changing the number of viewpoints invariably reshapes the BC distribution, meaning that the obtained BC values are conditioned by the number of viewpoints. From these results, we hypothesize that having more viewpoints would yield different BC values, thus the BC values obtained with the 326 viewpoints might not be representative of AS centrality.

5.2 Methodology

To address the above BC shortcomings, we devise a monitoring method based on a robust centrality metric called AS hegemony. The proposed method consists of two basic steps. First we generate graphs from AS paths advertised via BGP. Then, using AS hegemony, we estimate the centrality of each AS in the graphs. We consider two types of graphs, global and local graphs.

Global AS Graph

A global graph is made from all AS paths reported by the BGP viewpoints regardless of the origin AS and announced prefix. Consequently, these graphs represent the global Internet and central nodes stand for transit networks that are commonly crossed to reach arbitrary IP addresses. In 2017, IPv4 global graphs typically contains about 58k nodes and 188k edges (14k nodes and 43k edges for IPv6). The structure of these graphs is complex, yet they are valuable to monitor the Internet altogether and reveal Internet-wide routing changes.

Local AS Graph

A local graph is made only from AS paths with the same origin AS. Thereby, we compute a local graph for each AS announcing IP space globally. Each local graph represents the different ways to reach its origin AS and dominant nodes highlight the main transit networks towards only this AS. These graphs are particularly useful to monitor the dependence of an AS to other networks. In addition, structural changes in local graphs can expose important routing changes that may be detrimental to the origin AS connectivity.

5.3 AS Hegemony

The core of the proposed method is to quantify the centrality of ASes in the generated graphs. To circumvent BC sampling problems we propose AS hegemony metric [53]. This metric measures the fraction of paths passing through a node while correcting for sampling bias.

Computing the hegemony of AS v from AS paths collected from several viewpoints consists of the two following steps. First, AS paths from viewpoints that are biased towards or against AS v

are discarded. A viewpoint bias towards AS v means that the viewpoint is located within AS v, or topologically very close to it, and reports numerous AS paths passing through AS v. In contrast, a viewpoint bias against AS v is topologically far from v and is reporting an usually low number of AS paths containing v. Therefore, viewpoints with an abnormally high, or low, number of paths passing through v are discarded and only other viewpoints are selected to compute the hegemony score. Second, the centrality of v is computed independently for each selected viewpoint and these scores are aggregated to give the final AS hegemony value. That is, for each selected viewpoint j the BC of v (hereafter referred as $BC_{(j)}(v)$) is computed only from AS paths reported by j. And the average BC value across all selected viewpoints is the AS hegemony score of v.

These steps can be formally summarized into one equation. Let n be the total number of viewpoints, $\lfloor . \rfloor$ be the floor function and $0 \leq 2\alpha < 1$ be the ratio of disregarded viewpoints. That is, the top $\lfloor \alpha n \rfloor$ viewpoints with the highest number of paths passing through the AS and do the same for viewpoints with the lowest number of paths. Then the AS hegemony is defined as:

$$\mathcal{H}(v, \alpha) = \frac{1}{n - (2\lfloor \alpha n \rfloor)} \sum_{j=\lfloor \alpha n \rfloor + 1}^{n - \lfloor \alpha n \rfloor} BC_{(j)}(v),$$

where $BC_{(j)}$ is the BC value computed with paths from only one viewpoint j (i.e. $BC_{(j)}(v) = 1/S \sum_{w \in V} \sigma_{jw}(v)$) and these values are arranged in ascending order such that $BC_{(1)}(v) \leq BC_{(2)}(v) \leq \cdots \leq BC_{(n)}(v)$.

Figure 5.1a depicts the AS hegemony obtained for the simple graph with three viewpoints ($\alpha = .34$). Unlike the sampled BC, the AS hegemony is consistent for each type of node: transit ($\mathcal{H} = 0.58$), regional ISP ($\mathcal{H} = 0.25$) and stub AS ($\mathcal{H} = 0.08$). AS hegemony scores are intuitively interpreted as the average fraction of paths crossing a node. For example, on average a viewpoint has one fourth of its paths crossing a regional ISP ($\mathcal{H} = 0.25$).

In order to compare the robustness of AS hegemony and BC with real data, we reproduce the experiment of Section 5.1 with AS hegemony. Figure 5.1b shows that the hegemony values with 20 or more viewpoints are very similar to the ones obtained with the 326 peers. Meaning that

path sampling has significantly less impact on AS hegemony than on BC. Note that we randomly select peers across different projects (e.g. RouteViews, RIS) to obtain a diverse set of viewpoints. Selecting viewpoints from the same collector may yield poor results [53].

Accounting for Prefix Size

We extend AS hegemony to account for path disparities. In a nutshell, we weight paths according to the amount of IP space they are bound to. For example, a path to a /24 IP prefix represents a route to a smaller network than a path to a /16 IP prefix, thus we want to emphasize the path to the /16. The network prefix length alone is however not sufficient to resolve the IP space bound to a path. IP space deaggregation [54, 55] should also be taken into account. For example, a viewpoint reports the path 'X Y Z' for the prefix $a.b.c.0/17$ and the path 'X W Z' for the prefix $a.b.0.0/16$. Meaning that BGP favors path 'X Y Z' for half of the advertised /16. Here there is no need to give more emphasis to the path bound to the /16 as each path represents a route to 2^{15} IP addresses.

Consequently, we modify our definition of BC to account for the size of the IP space reachable through a path. Formally, $\sigma_{uw}(v)$ is the number of IP addresses bound to the paths from u to w and passing through v. That is the number of IP addresses corresponding to the advertised IP prefixes minus the number of IP addresses from covered prefixes (i.e. deaggregated and delegated prefixes [54]) that are not passing through v. In the rest of the work, this weighted version of BC is applied for the calculation of AS hegemony in IPv4, but as the relation between number of addresses and prefix size in IPv6 is more ambiguous we keep the classical BC definition for the calculation of AS hegemony in IPv6.

5.4 Results

Our Python implementation of the above method fetches BGP data using the BGPStream framework from CAIDA [56] and computes AS hegemony of all ASes in the global graph as well as AS hegemony for ASes in all local graphs. Our tool and updated results are made publicly available [50].

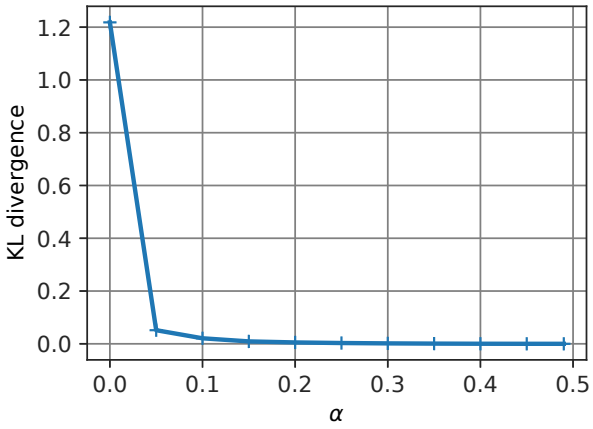


Figure 5.2: KL divergence between AS hegemony scores obtained with $\alpha = 0.49$ and different values of α (global graph on 2017/12/15 with rv2, LINX, rrc00, and rrc10 collectors).

Parameter Tuning

Setting the value of α is a trade-off between sampling robustness ($\alpha \approx 0.5$) and sensitivity to local routing changes ($\alpha \approx 0$). For example, setting $\alpha \approx 0.5$ achieves the most robust results to path sampling but conceals routing events affecting less than half of the monitored BGP peers. To monitor routing changes we seek for a small value of α that is still robust to path sampling. In Figure 5.2 we compare robust AS hegemony scores ($\alpha = 0.49$) to scores obtained with different values of α. For the following experiments we set the parameter $\alpha = 0.1$, as it provides results similar to those obtained with $\alpha = 0.49$ but is more sensitive to local changes.

BGP Datasets

The following results are all obtained using BGP data from four BGP collectors, two from the RouteViews project (route-views2 and LINX) and two from the RIS project (rrc00 and rrc10). These four collectors are selected from the collectors sensitivity results presented in [53]. For IPv4 they represent from 51 to 95 BGP peers respectively in 2004 and 2017. For IPv6, however, as the number of BGP peers is rather small before 2007 (i.e. less than 10 peers) and AS hegemony values might be irrelevant with such low number of peers (see Fig. 5.1b), we report only results obtained

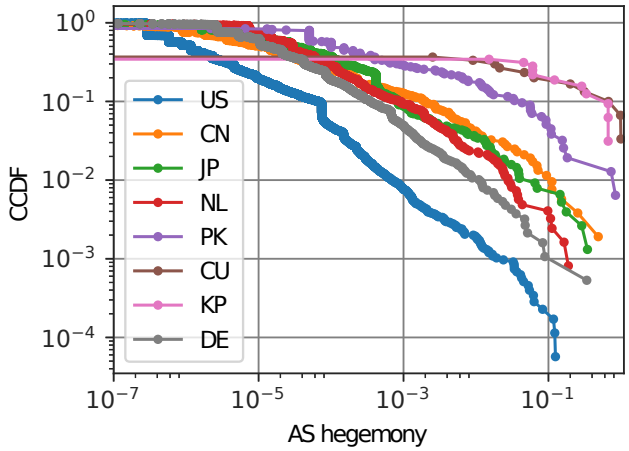


Figure 5.3: AS hegemony for paths toward countries

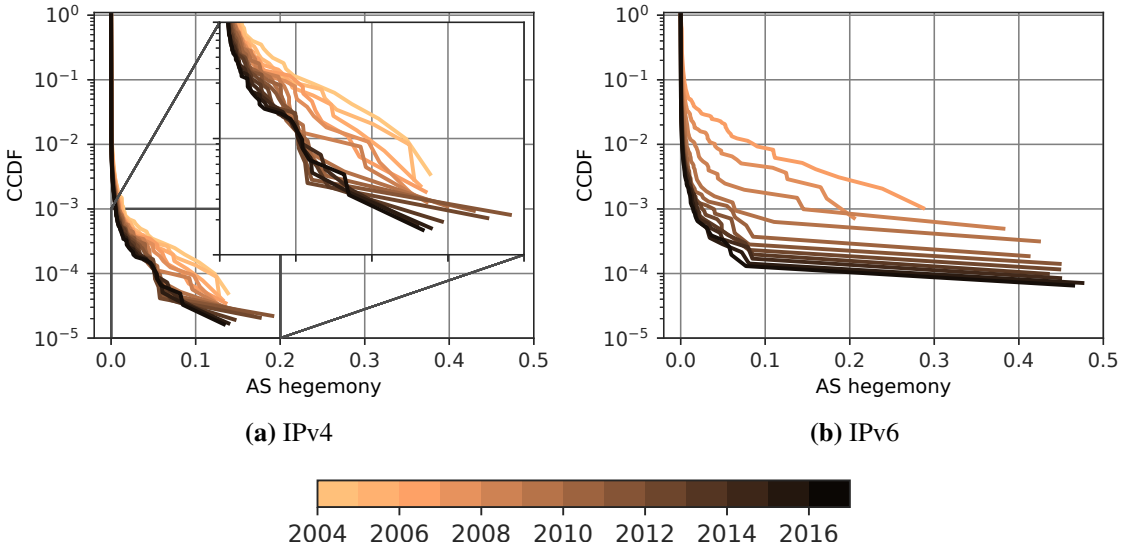
from 2007 onward using from 11 to 44 peers. The results presented below are obtained with RIB data of all peers for the 15th of each month from January 2004 to September 2017.

5.4.1 Country Level AS Graphs

Figure 5.3 illustrates the hegemony distribution for different AS graphs constructed with paths to prefixes mapped to distinct countries. Data points on the right hand side of the figure depict most central ASes for these countries. For Cuba, North Korea and Pakistan we observe a few ASes with an hegemony close to 1 meaning that all paths to these countries cross central ASes. The U.S. appeared to be the country where hegemony values are the most balanced. We found that the distribution of hegemony values is usually stable over time, and significant changes are a good indication of fundamental route changes usually attributed to BGP leaks or hijacks. Furthermore, the precision and robustness of AS hegemony enable us to accurately monitor very local changes in the AS graph.

5.4.2 IPv4 and IPv6 Global AS Graphs

As the starting point of our analysis, we investigate the AS interdependency for the entire IP space. We monitor the evolution of AS hegemony scores in the global AS graph from 2004 to 2017. Here large AS hegemony scores represent transit networks that are commonly crossed to



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

Figure 5.4: Distribution of AS hegemony for all ASes in the global graph.

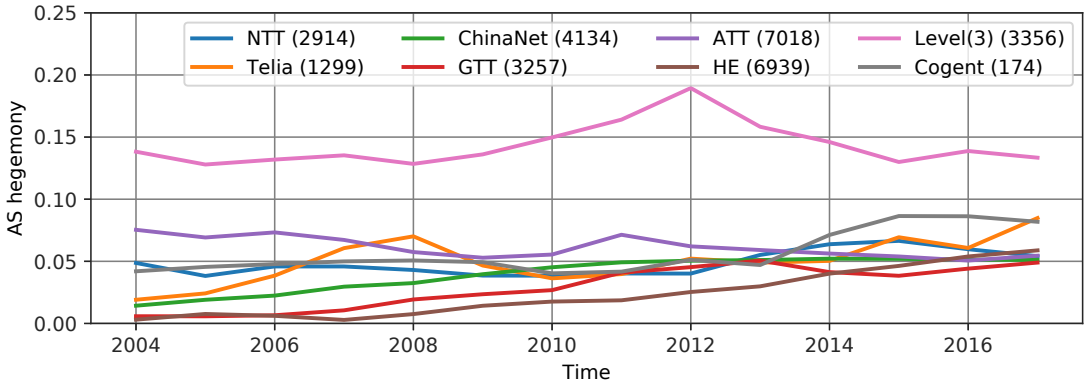


Figure 5.5: AS hegemony for Tier-1 ISPs from 2004 to 2017 (global graph, IPv4).

reach arbitrary IP addresses. Figure 5.4 depicts the distribution of the yearly average AS hegemony for all ASes in the IPv4 and IPv6 global AS graphs. In these figures each point represent an AS, and those on the right hand side of the figures stand for nodes with the highest hegemony values.

As the distribution of AS hegemony values for IPv4 is overall shifting to the left over time (Fig. 5.4a), we observe a global and steady decrease of AS hegemony values. This is another evidence of Internet's flattening [49], as networks are peering with more networks we observe less dominant ASes. Nonetheless, Figure 5.4a suggests that the AS hegemony for the most dominant networks (i.e. points on the right hand side) is quite stable.

We further investigate this by selecting the eight most dominant ASes found in our dataset and monitor their yearly AS hegemony (Fig. 5.5). The AS hegemony for these networks is indeed either steady, or increasing, which is contradictory with the global Internet flattening observed earlier. These two observations provide evidences of dense connectivity at the edge of the Internet but the role of large transit ISP is still very central to connect remote places in the Internet. This can be explained by the growth of public peering facilities (IXP) that allows regional networks to keep traffic locally and peer directly with content providers. Yet transiting to remote locations requires the international networks of tier-1 ISPs. In recent years this distinction between tier-1 ISP and other networks is event more visible, as we observe in Fig. 5.4a a clear gap between most networks ($\mathcal{H} < 0.03$) and tier-1 ISPs ($\mathcal{H} > 0.05$).

Figure 5.5 also depicts the dominance of Level(3) through the entire study period. After Level(3) acquisition of Global Crossing (AS3549) in 2011, it reached in 2012 the highest AS hegemony score monitored for the IPv4 global graph ($\mathcal{H} = 0.19$). We also found that from 2008 to 2010 Global Crossing was the most dominant AS in Level(3) local graph, meaning that it was the most common transit network to reach Level(3). These results thus assert that Global Crossing acquisition was the most effective way for Level(3) to attain new customers. It also illustrates the benefits of our tools for deployment and business decisions.

For IPv6 (Fig. 5.4b) we observe a faster Internet flattening than for IPv4. We hypothesize that this is mainly because the Internet topology for IPv6 in 2007 was quite archaic. But IPv6 has drastically gained in maturity, the AS hegemony distribution for IPv6 in 2017 is then very close to the one for IPv4 in 2009. The most striking difference with IPv4 is the central role played by Hurricane Electric (HE) in the IPv6 topology. After doubling its number of peers in 2009 [57], HE has been clearly dominating the IPv6 space from 2009 onward. It reaches an impressive AS hegemony $\mathcal{H} = 0.46$ in 2017, largely above the second and third highest scores (0.07 and 0.05), respectively, for Level(3) and Telia. Consequently, our tools confirm the dominant position of HE observed previously [58] and permit to systematically quantify the overall IPv6 dependency to HE.

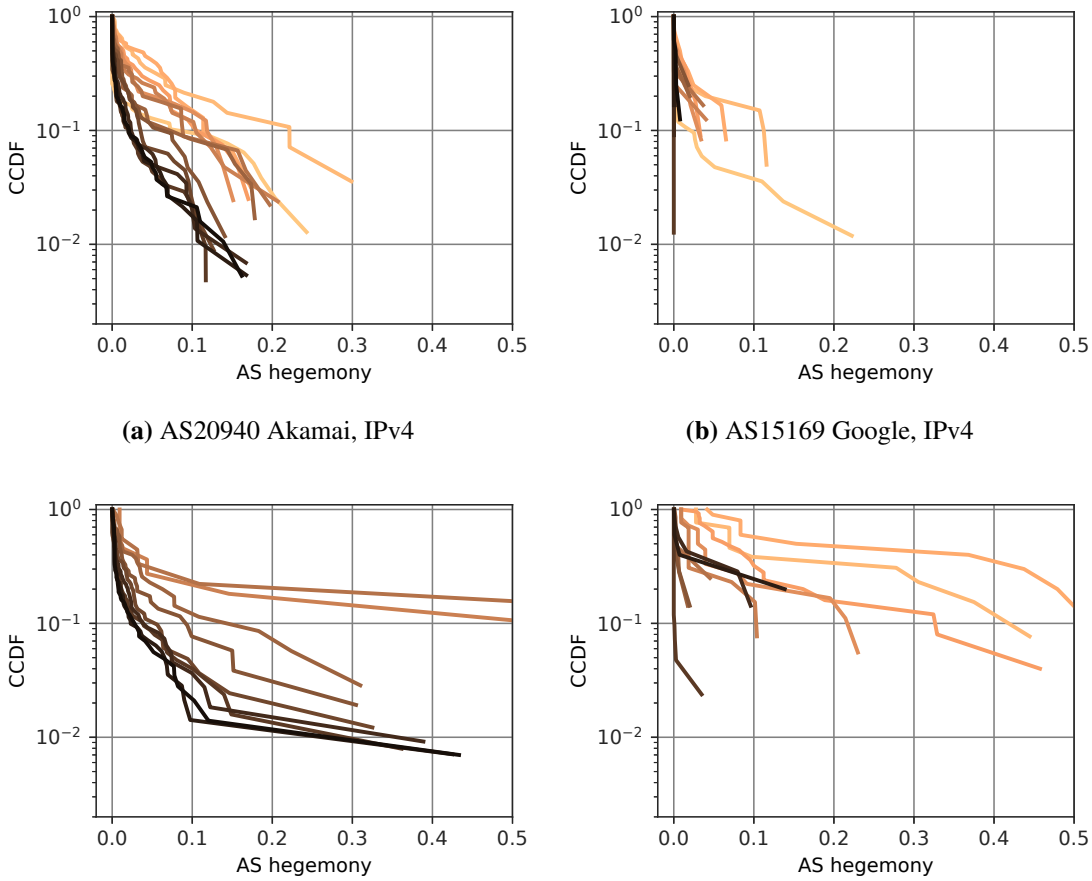
5.4.3 Case Studies

Our analysis now focuses on results obtained with local graphs. Unlike the global ones, local graphs shed light to AS dependency only for a specific origin AS. We found that the structure of local graphs is very different depending on the size and peering policies of the origin AS. On average in 2017, an IPv4 local graph contains 98 nodes but 93% of these nodes have an hegemony null ($\mathcal{H} = 0$). Typically ASes hosting BGP peers have an hegemony null and AS hegemony scores increases as the paths converge towards the origin AS. Thereby, the upstream provider of a single-homed origin AS gets the maximum hegemony score, $\mathcal{H} = 1$. By definition the origin AS of each local graph also features $\mathcal{H} = 1$, therefore, we are not reporting the AS hegemony of the origin AS in the following results.

In 2017, local graphs have on average 5 ASes with $\mathcal{H} > 0.01$, which usually corresponds to a set of upstream providers and tier-1 ASes. We also noticed interesting graphs containing no dominant AS, and other graphs containing numerous nodes with non-negligible AS hegemony scores. To illustrate this we pick a local graph from both end of the spectrum, namely, AS20940 from Akamai and AS15169 from Google.

Akamai and Google

The IPv4 graph for Akamai's main network, AS20940, is the local graph with the largest number of nodes in our results. In 2017, it contains on average 30 nodes with an AS hegemony greater than 0.01 (see Fig.5.6a). Meaning that accessing Akamai IP space relies on a large set of transit networks. This is true for our entire analysis period as shown in Figure 5.6a. Our manual inspection of Akamai BGP announcements reveals that Akamai is heavily fragmenting its IP space and advertising small prefixes at various Points of Presence (PoPs). Consequently, each prefix is accessible only through a very limited number of upstream providers and all BGP peers report AS paths going through these providers. In summary, Akamai local graph contains a lot of nodes with weak but non-negligible AS hegemony scores implying that Akamai has numerous weak AS-dependencies.



(a) AS20940 Akamai, IPv4

(b) AS15169 Google, IPv4

(c) AS20940 Akamai, IPv6

(d) AS15169 Google, IPv6

Figure 5.6: Distribution of AS hegemony for Google and Akamai local graphs. Same color scale as Fig. 5.4.

On the other hand, the IPv4 graph for Google (AS15169) in 2017 contains no node with an hegemony greater than 0.01 (see Fig. 5.6b). Our manual inspection of Google BGP advertisements reveals that, unlike Akamai, Google announces all their prefixes at each PoP. Because Google is peering at numerous places, all BGP peers report very short and different AS paths with almost no AS in common hence no relevant hegemony score. Nonetheless, Google's local graphs before 2012 feature a different AS hegemony distribution with a few high AS hegemony scores (Fig. 5.6b). Level(3) is the most dominant AS observed until 2012. But then Google has clearly succeeded to bypass Level(3) and alleviate its dependency to this AS (usually $\mathcal{H} < 0.00005$ from 2014). Now Level(3) is rarely seen in paths towards Google. In summary, we observe that Google used to depend on a few ASes but it is now mostly independent from all ASes. This is not an isolated

case, we measured no AS dependency for a few other ASes, notably, Microsoft (AS8075), Level(3) (AS3356), HE (AS6939), and Verisign (AS7342).

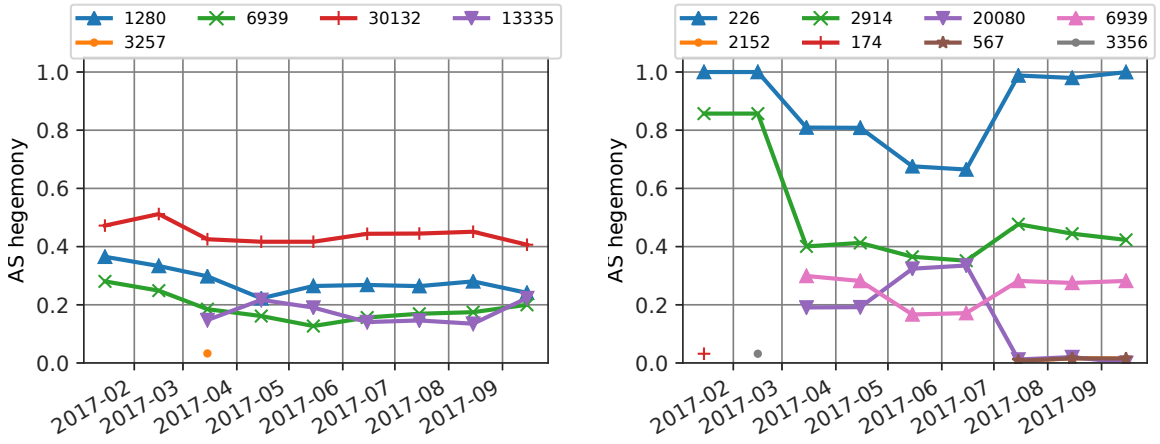
For IPv6, the situations for Akamai and Google are a bit different. The local graph for Akamai contains a lot of nodes with a high AS hegemony (Fig.5.6c). But HE is quite outstanding and features an AS hegemony ($\mathcal{H} = 0.43$) very close to the one observed for HE in the IPv6 global graph (Fig. 5.4b). HE is also the dominant node in Google's IPv6 local graph (Fig. 5.6d) but at a much lower magnitude ($\mathcal{H} = 0.12$). Thereby, our results show that Google's aggressive peering policy has partially succeeded to bypass HE ubiquitous IPv6 network.

DNS Root Servers

Monitoring an AS with our tools provides valuable insights into its AS dependency. This is particularly useful for networks hosting critical infrastructure, as operators of these ASes try to minimize their dependencies to third-party networks. To illustrate the benefits of our tools, we present results for the local graphs of ASes hosting DNS root servers. Notice that understanding AS dependency of root servers is usually a complicated task as most root servers are using anycast and more than 500 instances are deployed worldwide. We detail IPv4 networks hosting the F-root (AS3557) and B-root (AS394353) servers as they had significant structural changes in 2017.

In early 2017, we observe three dominant transit ASes for the network hosting the F-root server (Fig. 5.7a). AS30132 and AS1280 are direct upstream networks managed by ISC, the administrator of the F-root server. AS6939 is HE, the main provider for AS1280, and is found in about a third of the AS paths toward the F-root server. From March, Cloudflare (AS13335) starts providing connectivity to new F-root instances [59]. This new infrastructure is clearly visible in our results. Starting from March 2017, Cloudflare hegemony is fluctuating around 0.2 and seems to divert traffic from other instances as the three other transit networks have their hegemony proportionally decreased. From these results we deduce that the addition of Cloudflare has successfully reduced F-root dependencies on other ASes.

For the B-root server (Fig. 5.7b), we observe two dominant ASes in January and February 2017, Los Nettos (AS226) and NTT America (AS2914). Los Nettos reaches $\mathcal{H} = 1$ because



(a) F-root (AS3557) **(b)** B-root (AS394353)

Figure 5.7: AS hegemony for nodes in F-root (AS3557) and B-root (AS394353) local graphs from 15th January to 15th September 2017.

at that time the B-root server was unicasted and Los Nettos was the sole provider. NTT also has a very high AS hegemony score, in fact more than 80% of analyzed AS paths also cross NTT's network. From March 2017, we observe two other transit nodes AMPATH (AS20080) and HE (AS6939). Our manual inspection of the advertised paths reveals that a single /24 prefix is advertised with AMPATH as the first hop and usually HE as the second hop. This prefix is one of the two /24 prefixes advertised by the network hosting the B-root server (AS394353) but is not the one containing the server IP address. We believe that B-root operators were testing anycast in preparation for the deployment of the second instance of B-root at Miami that happened in May [60]. In May we acknowledge the deployment of the second instance hosted at AMPATH as the hegemony of that AS is raising again and the one for Los Nettos had significantly decreased. From July onward, however, we observe a sudden decrease of AMPATH hegemony while hegemony for Los Nettos is getting back close to 1. A manual comparison of AS paths in June and July reveals that Los Nettos is trying to fix this by prepending its ASN to paths through HE. Despite these efforts, most of the paths that were transiting through HE and AMPATH in June are replaced by paths going through HE and Los Nettos in July. The addition of the second instance in Miami had uncertain benefits, first, it considerably mitigated the dependence on NTT and Los Nettos networks in May and June, but then, from July Los Nettos is once again totally dominating the B-

root connectivity. Results for IPv6 are quite different, after the deployment in Miami we observe higher hegemony values for AMPATH. Both the IPv4 and IPv6 observations have been confirmed by the B-root operators.

5.5 Discussion

The structural changes observed for the F and B root servers illustrate the value of AS hegemony to monitor significant routing events. Our work enables possibility of designing an automated detection process to identify significant changes in AS hegemony scores. This detector should be able to report sudden routing changes such as the recent BGP leak from Google [61]. During this event Google became a transit provider for NTT OCN, which exhibits a sudden and significant increase in Google's AS hegemony for NTT's local graph. Using AS hegemony detecting this type of event is fairly easy, while state of the art tools employed by network operators (e.g. BGPmon provided by OpenDNS) have usually missed this significant event. The implementation and evaluation of this detector go beyond the scope of this work, but we believe AS Hegemony lays a strong foundation to build an effective system to do so. In Chapter 6 we show how the near-real time reporting of Hegemony values can be queried to produced such a alerting tool.

With more availability of data, this mechanism can benefit a lot. It could be possible to evaluate different weighting schemes. For example, by assigning paths' weight based on traffic volume an ISP can emphasize destinations that are favored by its customers.

5.6 Summary

We presented a methodology to quantify the AS interdependency in the Internet. It deals with the various AS paths reported via BGP and produce AS hegemony scores, that are robust estimates of the ASes centrality. Using 14 years of BGP data we proved that this method permits to monitor structural changes in the Internet and identify most important ASes to reach a certain part of the IP space. We also demonstrated with case studies the benefits of our tools to help ISPs to plan and assess infrastructure deployment.

Chapter 6

Internet Health Report

This dissertation aimed to make resources available to larger research and network operations community. Romain Fontugne at the Internet Initiative Japan Research Lab is leading the development of `Internet Health Report`: `ihr.iijlab.net`. The work presented in this dissertation is one of the primary input sources to this portal. Here, routing data from various ASes and countries is monitored and processed using algorithms developed in this dissertation (`Disco`, `ASMap`, `AS Hegemony`) as well as other methodologies from IIJ Research Lab.

In this chapter we provide a glance into this portal (Figure 6.1) and point out key functionalities that researchers could use.

6.1 Visualization

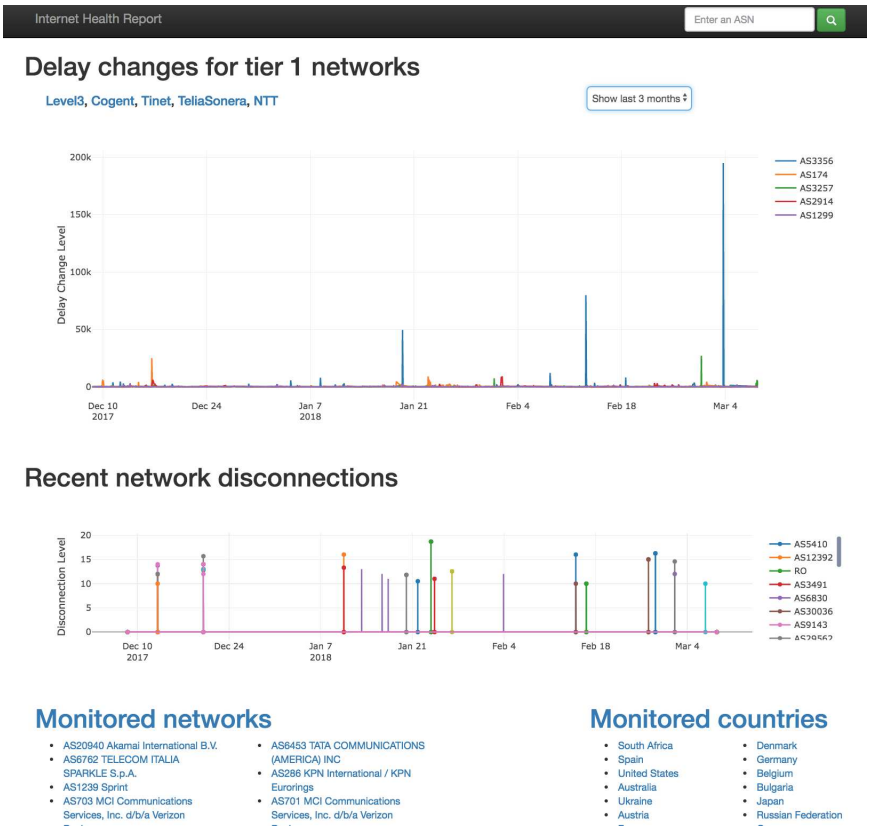


Figure 6.1: Internet Health Report

`Disco`, presented in Chapter 4, is running live at IIJ Research Lab and feeds detected outages into this portal. In Figure 6.2 we show an example outage and associated probes with prefixes that were unreachable in South Africa on February 16^{th} 2018. Using this live visualization we were able to detect and report on many outages at are of interest to the community. We not only visualize the detected event but also report it live using a twitter bot (`https://twitter.com/ihr_alerts`).

Figure 6.2: Outage detection reporting example

In Figure 6.3 we show integration of our visualization with that of RIPE NCC's TraceMon [63] tool. On February 27^{th} 2018 probe in Zen Internet, UK lost connectivity. Using the visualization and TraceMon integration, one can easily click on the outage and see traceroutes failing during the detect period, investigate which ASes suffered and where did the traceroutes fail.

`AS Hegemony`, presented in Chapter 5, is also tracked and reported live on this portal. It is now possible to quickly check which ASes a content provider depends on. We could also

Visualisation of traceroutes from disconnected probes

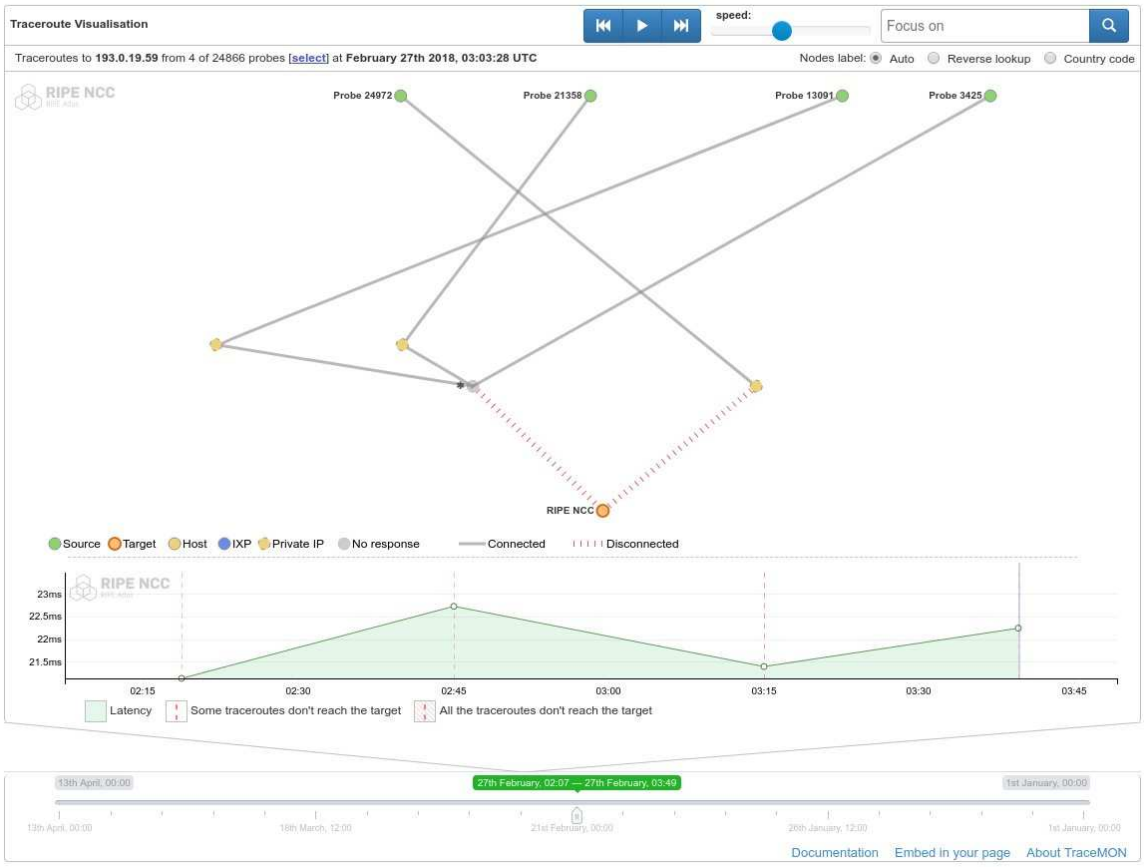


Figure 6.3: TraceMon Integration

quickly understand change in AS topology as a result of a DDoS attack. On February 28^{th} 2018, github.com was the subject of largest DDoS attack ever recorded (>1Tbps) [64]. As a response to this attack github switched to Akamai's Prolexic Cyber Security offering where paths to github now only went through Prolexic so they could be scrubbed out. This significant change in AS topology of github was easily captured by the Internet Health Report. We show this visualization in Figure 6.4.

Figure 6.4: AS Hegemony for Github

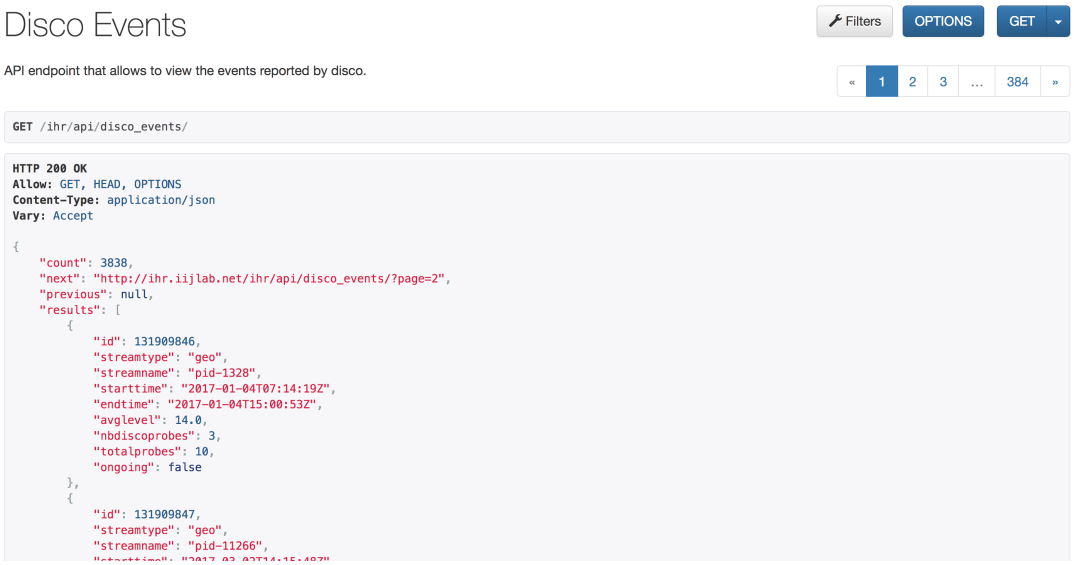


Figure 6.5: API example

6.2 API

Researchers might want to make use of data that powers Internet Health Report and produce more analysis or use inputs to open new directions to explore. We make all the data available via RESTful API. An example of the API is shown in Figure 6.5. A detailed documentation of the API can be found here: `ihr.iijlab.net/ihr/api/`

Chapter 7

Related Work

Detour Detection

In November 2013 Renesys reported a few suspicious paths [4]. One went from Guadalajara, Mexico to Washington, D.C. via Belarus; another went from Denver, CO through Reykjavik, Iceland, back to Denver. They used mostly data plane information from traceroute for their analysis. In [3] the authors focus on ISP inter-connectivity in the continent of Africa. They searched for paths that leave Africa only to return back. The goal, however, was to investigate large latencies in Africa and ways to reduce it. The premise was that if a route crosses international boundaries it would exhibit high latency. The work pointed to cases where local ISPs are not present at regional IXPs and IXP participants don't peer with each other. Similar to Renesys, they also use traceroute measurements, this time from the BISmark infrastructure (a deployment of home routers with custom firmware) in South Africa. Our study extends beyond Africa and investigates transient in addition to long-lasting detours. In *Boomerang* [2], the authors again use traceroute to identify routes from Canada to Canada that detour through the US. In this work the motivation was concerns about potential surveillance by the NSA. This work differs from ours in a number of ways: we characterize detours not just for one but 30 countries using control plane information rather than data plane. We use data plane measurement only for validation purposes.

To detect detours we only use only control plane data. This has a number of advantages: 1) Collecting data plane information at an Internet scale is hard. It needs infrastructure and visibility provided by Atlas probes or Ark monitors is limited. 2) Small footprint of our methodology makes it easily reproducible. Any network operator can pull a RIB dump from his/her border router and run `Netra` to detect detours for prefixes they own. Our goal is to not only detect detours but show characteristics about them which previous work does not present.

Data Plane vs Control Plane Incongruities

In [65] authors focus on routing policies and point out cases where routing decisions taken by ASes do not conform to expected behavior. There are complex AS relationships, such as, hybrid or partial transit which impact routing. Such relationships may lead to false positives in our results. However, the paper points out that most violations of expected routing behavior caused by complex AS relationships are very few and most violations were caused by major content providers. Our work identifies detours for variety of ASes, including both large content providers and small institutions. Moreover, in [66] authors argue that such incongruities are caused due to incorrect IP to AS mappings. About 60% of mismatches occur due to IP sharing between adjacent ASes. Authors here show that 63% to 88% of paths observed in control plane are valid in data plane as well. The work in [27] also analyzes the control plane (RIBs and AS paths) to construct a network topology and then uses traceroute to construct country-level paths. The goal of this work was to understand the role of different countries that act as hubs in cross-country Internet paths. Their results show that western countries are important players in country level internet connectivity.

Malicious AS Detection

In [29] authors present *ASwatch*, an AS reputation system to detect bulletproof hosting ASes. Similar to our work ASwatch relies on control plane information to detect malicious ASes (that may host botnet C&C servers, phishing sites, etc). The motivation of this work is different than ours. ASwatch attempts to detect malicious ASes by mining their link stability, IP space fragmentation and prefix reachability. ASwatch will not detect ASes that cause detours. The detour origin ASes that our work detects could complement features that ASwatch uses. As authors in [29] point out malicious ASes rewire their routes more frequently than legitimate ones, transient detours might be particularly useful to improve detection capability of ASwatch.

Geolocation Accuracy

In context of MaxMind geolocation accuracy, [67] and [68] have shown MaxMind country geolocation to be 99.8% in consensus with other geolocation DBs. In [69] authors use data from Routing Information Registries (RIRs), RIPE DB and Team Cymru to determine all IP blocks and ASes that geolocate to Germany. To validate their geolocation accuracy, authors query the MaxMind database which allows mapping IP addresses to their country of presence. We adopt a more exhaustive strategy than [69].

Control-plane-only for detection

One way to detect detours is to use *traceroute*, analyze reported hops and use latency as an indication of a detour. This approach was followed by [3] that studied peering relationships in Africa; we too use this approach to validate our results on live data.

Outage Detection

Outage detection has been studied from different angles. Operationally, important outages are likely to be discussed on network operations mailing lists which can be data-mined [70]. Censorship can be implemented as country-wide Internet outages, which have been studied from BGP, traceroute, and Internet background radiation data in [71].

For parts of the Internet that generate enough background radiation, network telescopes [72] can be used to detect outages. The alternative is sending probing packets and detecting changes in responses. To do this at Internet scale one needs to inject a massive number of packets in the network, and this works only for the part of the Internet where targets respond to active probing and packets are not blocked by upstream infrastructure. Dainotti et. al. [73] find that, out of 10.5M routed /24 prefixes equivalents, IP addresses in 3.1M /24 prefixes are seen in the UCSD network telescope, and 4.5M /24 prefixes are visible using active probing, but both techniques observe mostly the same /24 prefixes. Also, outage detection methods based on network telescope data are opportunistic, as they are restrained to monitor whatever networks generating background radiation. Our study relies on a pre-existing metadata stream which is deterministic thus more

reliable. Furthermore, we observed that out of the total Internet space visible to RIPE Atlas, about 25% is not monitored by active probing.

In [74], an approach is proposed solely based on flow data at a network border. To make this work across multiple networks, the problem of sharing potentially privacy sensitive flow data must be tackled, which is explored in privacy preserving distributed outage monitoring [75, 76], with non-trivial complexity.

In active probing, the Trinocular [36] and Hubble [77] projects are especially notable, as they do Internet-wide adaptive scanning. Trinocular uses adaptive ICMP probing, exploring the trade-off between probing rate and accuracy. Hubble uses BGP feeds and ICMP probing to find potential problems which are investigated and classified using traceroute-based approaches. PlanetSeer [78] detects anomalies in P2P traffic. Upon anomaly detection it performs active probing to locate and quantify the outage.

AS Centrality

In the literature AS centrality is commonly measured using Betweenness Centrality (BC). This is one of the basic metric used to characterize the topology of the Internet [52, 79]. It was also applied for similar motivation as ours. Karlin et al. [80] consider Internet routing at the country-level to investigate the interdependencies of countries and identify countries relying on other countries enforcing censorship or wiretapping. BC is also used to identify critical ASes for industrial and public sectors in Germany [10]. Similarly, Schuchard et al. [81] select targets for control plane attacks using a ranking based on BC. Finally, researchers have also applied BC to detect changes in the AS-topology. For example, Liu et al. [82] employ BC to monitor rerouting events caused by important disruptive events such as sea cable faults. Our work presented a more robust way to detect AS centrality.

Chapter 8

Future Work

The work presented in this dissertation paves the way for future work in several keys areas. In Sections 3.7, 4.8, and 5.5 we highlighted some immediate next steps for detour detection, outage detection and AS connectivity analysis respectively.

BGP plays a very important role in Internet connectivity. While data sources such as BGPmon and RouteViews have provided researchers with a rich BGP archive and tools to understand routing at the AS level, there is lack of systems that can sit on top of these archives and provide analytics. The methods presented in this thesis are only the beginning of such systems. Network operators and regulators play an important role in the growth of Internet's economy. A study, both from a technical and economic perspective is possible, to measure the impact of routing anomalies. One might build upon the systems presented here to measure the growth of the Internet in developing economies.

There are signs that routing, principally controlled by BGP or the control plane, is merging into a more software-defined, data-driven world. More and more content providers are building systems that measure data points such as capacity at peering and transit links, user QoE and application optimizations to drive route selection at the core routers. In such cases, the clear separation between data and control plane is hard. In the future, studies of routing impact during such optimizations should be done, especially for traffic egressing towards the public Internet.

Public measurement platforms such RIPE Atlas provide free access to traceroute measurements that traverse a large part of the Internet. It is possible to merge information from such sources and develop predictive models. By learning destinations that have a common routing segment, a rise in latency or an outage from one traceroute can be used as an indicator to predict routing anomalies for another set of ASes.

Internet protocol standardization can also leverage from this work. For example, RPKI (a public-private key-based BGP security mechanism) has seen growth much less than it should have,

given its advantages. Detecting prefix hijacks using AS Hegemony, or routing detours can provide feedback into discussions about such security mechanisms driving the growth forward.

As we covered in Section 3.6, geolocation enables a better understanding of routing. However, the accuracy of geolocation sources is not very reliable. Improvements in geolocation accuracy, especially router geolocation, will certainly propel routing research forward. Platforms such as RIPE's OpenIPMap and TraceMon are steps in right direction but more research needs to be done in carefully maintaining and improving them.

Another possible area of future work is IPv6. According to recent numbers from Google and Akamai, about 20% of the clients now reach their platform on IPv6. More than 90% of Verizon's customers reach Google on IPv6 rather than IPv4. There is no doubt that, while slow in adoption, IPv6 is the future of the Internet and for the foreseeable future we will live with both the versions of the IP protocol. There is a clear need for building systems to understand IPv6 routing better. With new protocols, there are new complexities. The methodology developed for IPv4 often fails when measuring IPv6 because of the latter's huge size. More work needs to be done on how to cover the large footprint of IPv6 before developing tools to geolocate, detect detours and outages.

Finally, there is future work in making the systems presented in this dissertation translate directly into actionable items. We have reached out to network operation centers (NOCs) at large CDNs and see a need for automated systems that work with existing monitoring tools and use them as a data source to refine what actions an operator can take to address outages, hot circuits, performance degradation, etc.

Chapter 9

Conclusions

The Internet constitutes a trillion dollar economy. We rely on the Internet for communication, business, entertainment, education, etc. Even though failure or degradation of Internet routing leads to large financial losses, these events commonly occur both as a result of attack or misconfiguration. There is an increasing need fueled by new national regulations in Europe and Australia, for ISPs to ensure that personal information belonging to their users does not leave the country. It is unclear whether such regulations cover data in transit as well as storage, but data can certainly be sniffed while in transit, violating the original intent. Such regulations may place a substantial burden on ISPs to prove that data remains within a country for its entire lifetime, even when it moves. It is still far from clear what the implications are for ISP operations. Current public routing monitoring systems do not have the tools to monitor data in transit and state with confidence that data has not left a country, even briefly. Similarly, there is a lack of public monitoring systems that report routing failure to public and present characteristics about them.

Characterizing detours on the Internet is very useful. Customers gain more insight into how their providers route traffic. There is perhaps an expectation from users that if they send traffic to other users in the same country the packets will not step outside national borders; our work provides evidence to the contrary. Network operators can use our methodology and results for diagnostic purposes. A sudden change in RTT may be traced to a detour or keeping track of what the routing system does. The latter is important to assure customers that their traffic is not subject to monitoring by other governments.

Our work is useful to regulators and state officials responsible for network infrastructure since our work quantifies information about a practice that may run afoul of state policy. State officials can use such information to assure citizens that their traffic stays within national borders or direct ISPs to alter their practices. State agencies that transmit sensitive information may monitor detours to alert for potential MITM attacks. Finally, entrepreneurs may use our results when decid-

ing where to establish new Internet exchange points (IXP) or deploy infrastructure in developing countries.

To detect routing failures, we presented a light-weight outage diagnostic system based on detecting bursts of TCP disconnects. In combination with RIPE Atlas, our work allows network operators to have a fast outage detection system that tests the full extent of their network, from beyond customer premise equipment up to (and including) upstream networks, regardless of whether ICMP probing is allowed in the relevant network.

Large content providers can leverage the outage detection method presented in this dissertation. Providers keep a log of TCP dis/connections towards clients. Using such logs it is possible to apply burst detection and detect outages or performance degradation and act accordingly, such as direct users to another PoP and avoid the link that failed.

Understanding dependencies between networks is integral to detecting possible bottlenecks, attacks, or to aid in decision making for deployment of new infrastructure. In this dissertation, we presented a novel study of AS Hegemony, a way to measure which ASes are important for a network's connectivity. A near-real time study of this metric reveals interesting characteristics of AS topology and helps in detection of misconfigurations, route leaks and other significant changes in the Internet routing infrastructure.

Internet measurements is a thriving research domain and this dissertation makes a contribution towards improving the measurement of connectivity on the Internet. It is very apparent that we need to include multiple data sources from both the control and data planes, and study algorithms that provide actionable insights into the observed events. We provide ISPs and content providers with tools to alert them when a routing event has taken place, followed by information about it. This study enables emerging regulatory requirements to evaluate the impact of routing anomalies on security and privacy. With this dissertation, we hope to have drawn attention to the intricacies of Internet routing, different protocols that govern them, how we can measure and characterize events and make near-real time reporting possible which helps the growth of the Internet while the next billion people get plugged in.

Bibliography

[1] UK has world's most Internet-dependent economy. http://www.information-age.com/bt-wins-39m-nato-networking-deal-2093003/.

[2] Jonathan A. Obar and Andrew Clement. Internet Surveillance and Boomerang Routing: A Call for Canadian Network Sovereignty. *SSRN Electronic Journal*, 2013.

[3] Arpit Gupta, Matt Calder, Nick Feamster, Marshini Chetty, Enrico Calandro, and Ethan Katz-Bassett. Peering at the internetâĂŹs frontier: A first look at isp interconnectivity in africa. In Michalis Faloutsos and Aleksandar Kuzmanovic, editors, *Passive and Active Measurement*, volume 8362 of *Lecture Notes in Computer Science*, pages 204–213. Springer International Publishing, 2014.

[4] Jim Cowie. The new threat: Targeted internet traffic misdirection, Nov 2013. http://www.renesys.com/2013/11/mitm-internet-hijacking/.

[5] Huffington Post. What Libya Learned From Egypt. http://www.huffingtonpost.com/jim-cowie/libya-egypt-internet_b_831794.html, 2011.

[6] Huffington Post. Syria Internet Goes Down, Traffic Monitoring Sites Confirm. http://www.huffingtonpost.com/2013/05/07/syria-internet-down_n_3232433.html, 2013.

[7] Thousand Eyes Blog. Route Leak Causes Global Outage in Level 3 Network. https://blog.thousandeyes.com/route-leak-causes-global-outage-level-3-network/, 2015.

[8] Dyn Research Blog. Routing Leak briefly takes down Google. http://research.dyn.com/2015/03/routing-leak-briefly-takes-google/, 2015.

[9] Amsterdam internet exchange (amsix) outage, may 13, 2016.

[10] Matthias Wählisch, Thomas C Schmidt, Markus de Brün, and Thomas Häberlen. Exposing a nation-centric view on the german internet–a change in perspective on as-level. In *PAM*, pages 200–210. Springer, 2012.

[11] University of oregon route views project. http://www.routeviews.org/.

[12] RIPE NCC. RIPE Atlas. https://atlas.ripe.net.

[13] Thousandeyes: Route leak causes global outage in level 3 network. https://blog.thousandeyes.com/route-leak-causes-global-outage-level-3-network/.

[14] MaxMind LLC. Maxmind geoip country database. http://dev.maxmind.com/geoip/legacy/geolite/.

[15] Caida ark dataset. http://www.caida.org/projects/ark/.

[16] Ripe ncc openipmap. https://github.com/RIPE-Atlas-Community/openipmap.

[17] iplane datasets, university of washington. http://iplane.cs.washington.edu/data/data.html.

[18] The caida internet topology data kit. http://www.caida.org/data/internet-topology-data-kit/.

[19] Packet clearing house ixp datasets. https://prefix.pch.net/applications/ixpdir/menu_download.php.

[20] Peering db 2.0 api. https://prefix.pch.net/applications/ixpdir/menu_download.php.

[21] Caida as relationships. http://www.caida.org/data/as-relationships/.

[22] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan. The internet topology zoo. *IEEE Journal on Selected Areas in Communications*, 29(9):1765–1775, October 2011.

[23] Ethan Katz-Bassett, John P. John, Arvind Krishnamurthy, David Wetherall, Thomas Anderson, and Yatin Chawathe. Towards ip geolocation using delay and topology measurements. In *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, IMC '06, pages 71–84, New York, NY, USA, 2006. ACM.

[24] Bradley Huffaker, Amogh Dhamdhere, Marina Fomenkov, and Kc Claffy. Toward topology dualism: Improving the accuracy of as annotations for routers. In *Proceedings of the 11th International Conference on Passive and Active Measurement*, PAM'10, pages 101–110, Berlin, Heidelberg, 2010. Springer-Verlag.

[25] CAIDA AS Ranks. `http://as-rank.caida.org`.

[26] Rachee Singh, Hyungjoon Koo, Najmehalsadat Miramirkhani, Fahimeh Mirhaj, Phillipa Gill, and Leman Akoglu. The politics of routing: Investigating the relationship between as connectivity and internet freedom. In *6th USENIX Workshop on Free and Open Communications on the Internet (FOCI 16)*, Austin, TX, August 2016. USENIX Association.

[27] Josh Karlin, Stephanie Forrest, and Jennifer Rexford. Nation-state routing: Censorship, wiretapping, and BGP. *CoRR*, abs/0903.3218, 2009.

[28] Anne Edmundson, Roya Ensafi, Nick Feamster, and Jennifer Rexford. A first look into transnational routing detours. In *Proceedings of the 2016 Conference on ACM SIGCOMM 2016 Conference*, SIGCOMM '16, pages 567–568, New York, NY, USA, 2016. ACM.

[29] Maria Konte, Roberto Perdisci, and Nick Feamster. Aswatch: An as reputation system to expose bulletproof hosting ases. *SIGCOMM Comput. Commun. Rev.*, 45(5):625–638, August 2015.

[30] Matthew Luckie, Young Hyun, and Bradley Huffaker. Traceroute probe method and forward ip path inference. In *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, IMC '08, pages 311–324, New York, NY, USA, 2008. ACM.

[31] kc claffy Young Hyun, Andre Broido. Traceroute and bgp as path incongruities.

[32] Beichuan Zhang, Daniel Massey, and Lixia Zhang. Bgp dynamics during route flap damping. Technical report.

[33] Manaf Gharaibeh, Anant Shah, Bradley Huffaker, Han Zhang, Roya Ensafi, and Christos Papadopoulos. A look at router geolocation in public and commercial databases. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, pages 463–469, New York, NY, USA, 2017. ACM.

[34] Bamba Gueye, Artur Ziviani, Mark Crovella, and Serge Fdida. Constraint-based geolocation of Internet hosts. *IEEE/ACM Trans. Netw.*, 14(6):1219–1232, December 2006.

[35] Anant Shah, Romain Fontugne, and Christos Papadopoulos. Towards characterizing international routing detours. In *Proceedings of the 12th Asian Internet Engineering Conference*, AINTEC '16, pages 17–24, New York, NY, USA, 2016. ACM.

[36] Lin Quan, John Heidemann, and Yuri Pradkin. Trinocular: Understanding Internet reliability through adaptive probing. In *Proceedings of the ACM SIGCOMM Conference*, pages 255–266, Hong Kong, China, August 2013. ACM.

[37] RIPE NCC. RIPE Atlas Built-in Measurements. `https://atlas.ripe.net/docs/built-in/`.

[38] RIPE NCC. RIPE Atlas Result Streams. `https://atlas.ripe.net/docs/result-streaming/`.

[39] USC/LANDER Project. Internet Outage Dataset, PREDICT ID: "USC-LANDER/internet_outage_adaptive_a[20-22]all". http://www.isi.edu/ant/lander.

[40] Jon Kleinberg. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4):373–397, 2003.

[41] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[42] B. Huffaker, M. Fomenkov, and k. claffy. Geocompare: a comparison of public and commercial geolocation databases - Technical Report . Technical report, Cooperative Association for Internet Data Analysis (CAIDA), May 2011.

[43] Dutch Public Broadcasting. Power failure in Amsterdam, 2017. http://nos.nl/artikel/2153383-stroomstoring-regio-amsterdam-na-uren-opgelost.html.

[44] Time warner cable outage, august 27, 2014.

[45] BBC. Kenya nationwide blackout caused by rogue monkey. http://www.bbc.com/news/world-africa-36475667, 2016.

[46] Romain Fontugne, Emile Aben, Cristel Pelsser, and Randy Bush. Pinpointing delay and forwarding anomalies using large-scale traceroute measurements. *CoRR*, abs/1605.04784, 2016.

[47] Router Issue Caused Time Warner Cable Outage in the Carolinas. http://www.twcnews.com/nc/triangle-sandhills/news/2015/12/27/time-warner-cable-outages-reported-statewide.html.

[48] A. Shah, R. Fontugne, E. Aben, C. Pelsser, and R. Bush. Disco: Fast, good, and cheap outage detection. In *2017 Network Traffic Measurement and Analysis Conference (TMA)*, pages 1–9, June 2017.

[49] Giovanni Comarela, Evimaria Terzi, and Mark Crovella. Detecting unusually-routed ases: Methods and applications. In *IMC*, pages 445–459. ACM, 2016.

[50] AS Hegemony Results. `http://ihr.iijlab.net/ihr/hegemony/`, 2017.

[51] Sang Hoon Lee, Pan-Jun Kim, and Hawoong Jeong. Statistical properties of sampled networks. *Phys. Rev. E*, 73:016102, Jan 2006.

[52] Priya Mahadevan, Dmitri Krioukov, Marina Fomenkov, Xenofontas Dimitropoulos, k c claffy, and Amin Vahdat. The internet as-level topology: Three data sources and one definitive metric. *SIGCOMM CCR*, 36(1):17–26, January 2006.

[53] Romain Fontugne, Anant Shah, and Emile Aben. As hegemony: A robust metric for as centrality. In *Proceedings of the SIGCOMM Posters and Demos*, pages 48–50. ACM, 2017.

[54] Luca Cittadini, Wolfgang MuȊĹhlbauer, Steve Uhlig, Randy Bush, Pierre François, and Olaf Maennel. Evolution of internet address space deaggregation: Myths and reality. *IEEE JSAC*, 8(28):1238–1249, 2010.

[55] Julien Gamba, Romain Fontugne, Cristel Pelsser, Randy Bush, and Emile Aben. Bgp table fragmentation: what & who? In *CoRes*, 2017.

[56] Chiara Orsini, Alistair King, Danilo Giordano, Vasileios Giotsas, and Alberto Dainotti. Bgpstream: a software framework for live and historical bgp data analysis. In *IMC*, pages 429–444. ACM, 2016.

[57] E. Hui Pan. *Gigabit/ATM Monthly Newsletter November 2009*. Information Gatekeepers Inc.

[58] Amogh Dhamdhere, Matthew Luckie, Bradley Huffaker, Ahmed Elmokashfi, Emile Aben, et al. Measuring the deployment of ipv6: topology, routing and performance. In *IMC*, pages 537–550. ACM, 2012.

[59] Dani Grant. Delivering Dot . https://blog.cloudflare.com/f-root/, 2017.

[60] Root Operators. B-Root Begins Anycast in May. http://root-servers.org/news/b-root-begins-anycast-in-may.txt, 2017.

[61] Andree Toonk. BGP leak causing Internet outages in Japan and beyond. https://bgpmon.net/bgp-leak-causing-internet-outages-in-japan-and-beyond/, August 2017.

[62] R. Fontugne, A. Shah, and E. Aben. The (thin) bridges of as connectivity: Measuring dependency using AS hegemony. In *2018 Passive and Active Measurement Conference (PAM)*, March 2018.

[63] TraceMon, RIPE NCC. https://labs.ripe.net/Members/massimo_candela/tracemon-traceroute-visualisation-network-debugging-tool.

[64] Wired, github ddos, 2018. https://www.wired.com/story/github-ddos-memcached/.

[65] Ruwaifa Anwar, Haseeb Niaz, David Choffnes, Ítalo Cunha, Phillipa Gill, and Ethan Katz-Bassett. Investigating interdomain routing policies in the wild. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, IMC '15, pages 71–77, New York, NY, USA, 2015. ACM.

[66] Yu Zhang, Ricardo Oliveira, Hongli Zhang, and Lixia Zhang. Quantifying the pitfalls of traceroute in as connectivity inference. In *Proceedings of the 11th International Conference on Passive and Active Measurement*, PAM'10, pages 91–100, Berlin, Heidelberg, 2010. Springer-Verlag.

[67] B. Huffaker, M. Fomenkov, and k. claffy. Geocompare: a comparison of public and commercial geolocation databases - Technical Report . Technical report, Cooperative Association for Internet Data Analysis (CAIDA), May 2011.

[68] Yuval Shavitt and Noa Zilberman. A study of geolocation databases. *CoRR*, abs/1005.5674, 2010.

[69] Matthias Wählisch, Thomas C. Schmidt, Markus de Brün, and Thomas Häberlen. Exposing a nation-centric view on the german internet — a change in perspective on as-level. In *Proceedings of the 13th International Conference on Passive and Active Measurement*, PAM'12, pages 200–210, Berlin, Heidelberg, 2012. Springer-Verlag.

[70] Ritwik Banerjee, Abbas Razaghpanah, Luis Chiang, Akassh Mishra, Vyas Sekar, Yejin Choi, and Phillipa Gill. Internet outages, the eyewitness accounts: Analysis of the outages mailing list. In *International Conference on Passive and Active Network Measurement*, pages 206–219. Springer, 2015.

[71] Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C. Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, pages 1–18, New York, NY, USA, 2011. ACM.

[72] UCSD Network Telescope. `https://www.caida.org/projects/network_telescope/`.

[73] A. Dainotti, K. Benson, A. King, B. Huffaker, E. Glatz, X. Dimitropoulos, P. Richter, A. Finamore, and A. Snoeren. Lost in Space: Improving Inference of IPv4 Address Space Utilization. *IEEE Journal on Selected Areas in Communications (JSAC)*, 34(6):1862–1876, Jun 2016.

[74] Dominik Schatzmann, Simon Leinen, Jochen Kögel, and Wolfgang Mühlbauer. *FACT: Flow-Based Approach for Connectivity Tracking*, pages 214–223. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[75] Mentari Djatmiko, Dominik Schatzmann, Arik Friedman, Xenofontas Dimitropoulos, and Roksana Boreli. Privacy preserving distributed network outage monitoring. In *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on*, pages 69–70. IEEE, 2013.

[76] Mentari Djatmiko, Dominik Schatzmann, Xenofontas Dimitropoulos, Arik Friedman, and Roksana Boreli. Federated flow-based approach for privacy preserving connectivity tracking. In *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*, pages 429–440. ACM, 2013.

[77] Ethan Katz-Bassett, Harsha V. Madhyastha, John P. John, Arvind Krishnamurthy, David Wetherall, and Thomas Anderson. Studying black holes in the internet with hubble. In *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, NSDI'08, pages 247–262, Berkeley, CA, USA, 2008. USENIX Association.

[78] Ming Zhang, Chi Zhang, Vivek Pai, Larry Peterson, and Randy Wang. Planetseer: Internet path failure monitoring and characterization in wide-area services. In *Proceedings of the 6th Conference on Symposium on Opearting Systems Design & Implementation - Volume 6*, OSDI'04, pages 12–12, Berkeley, CA, USA, 2004. USENIX Association.

[79] Shi Zhou and Raúl J Mondragón. Accurately modeling the internet topology. *Physical Review E*, 70(6):066108, 2004.

[80] Josh Karlin, Stephanie Forrest, and Jennifer Rexford. Nation-state routing: Censorship, wiretapping, and BGP. *CoRR*, abs/0903.3218, 2009.

[81] Max Schuchard, Abedelaziz Mohaisen, Denis Foo Kune, Nicholas Hopper, Yongdae Kim, and Eugene Y Vasserman. Losing control of the internet: using the data plane to attack the control plane. In *CCS*, pages 726–728. ACM, 2010.

[82] Yujing Liu, Xiapu Luo, Rocky KC Chang, and Jinshu Su. Characterizing inter-domain rerouting by betweenness centrality after disruptive events. *IEEE JSAC*, 31(6):1147–1157, 2013.