UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

INTER-DOMAIN TRAFFIC MANAGEMENT IN AN EVOLVING INTERNET PEERING ECOSYSTEM

Autor:	Juan Camilo Cardona, IMDEA Networks Institute
Directores:	Pierre Francois, Cisco Systems
	Rade Stanojevic, Telefonica I+D
Tutor:	Ruben Cuevas Rumin, Universidad Carlos III de Madrid

DEPARTAMENTO DE INGENIERÍA TELEMÁTICA

Leganés (Madrid), Mayo de 2016

UNIVERSIDAD CARLOS III DE MADRID

Ph.D. THESIS

INTER-DOMAIN TRAFFIC MANAGEMENT IN AN EVOLVING INTERNET PEERING ECOSYSTEM

Author:	Juan Camilo Cardona, IMDEA Networks Institute
Directors:	Pierre Francois, Cisco Systems
	Rade Stanojevic, Telefonica I+D
Tutor:	Ruben Cuevas Rumin, Universidad Carlos III de Madrid

DEPARTMENT OF TELEMATIC ENGINEERING

Leganés (Madrid), May 2016

Inter-domain traffic management in an evolving Internet peering ecosystem

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Prepared by
Juan Camilo Cardona, IMDEA Networks Institute

Under the advice of
Pierre Francois, Cisco Systems
Rade Stanojevic, Telefonica I+D

Departamento de Ingeniería Telemática, Universidad Carlos III de Madrid

---

Date: May, 2016

Web/contact: [juancamilo.cardona@imdea.org]

This work has been supported by IMDEA Networks Institute.

TESIS DOCTORAL

INTER-DOMAIN TRAFFIC MANAGEMENT IN AN EVOLVING INTERNET PEERING ECOSYSTEM

Autor: Juan Camilo Cardona, IMDEA Networks Institute
Directores: Pierre Francois, Cisco Systems
Rade Stanojevic, Telefonica I+D
Tutor: Ruben Cuevas Rumin, Universidad Carlos III de Madrid

Firma del tribunal calificador:

Presidente: Jordi Domingo-Pascual

Vocal: Víctor López

Secretario: Francisco Valera Pintor

Calificación:

Leganés, 6 de Mayo de 2016

Acknowledgements

I learnt a lot on my own over these four years, but it is the lessons from Pierre and Rade what I appreciate the most. Pierre helped me go deeper in the networking world, by introducing me to people, anecdotes, knowledge, and experiences from the core of the industry. Rade opened my mind to the world of research, and showed me the thrill of data exploration, always in the most pragmatic way. It would have been impossible to have this professional and complementary training with a single person. I am a better engineer thanks to them.

Doing research requires a lot of work and resources. I wouldn't have been able to do much without the help of many. I would like to thank, Stefano, Nikos, Sergey, and all my other co-authors for their great help. Special thanks go to Ignacio de Castro, whose point of view helped me see the world and its problems in a very different way. I would also like to thank the operators of the anonymous Tier-2 network and of RedIRIS for supporting me through their network data. I also acknowledge Pradeep Bangera for making possible to obtain the data of the latter.

These years in Madrid would have been boring and lonely without the presence of many people. Thanks to Pablo and Isabel for their trust, their patience over my money saving habits, and the best pink birthday cake. To Fabio, for the always necessary football discussion, and to Chiara for her endless happiness. To Christian, for tolerating my daily complains. To Andrea, for always raising my self-esteem. To Jordi (and therefore to his complement, Mari Carmen), for being a great flat-mate and better friend. In general, thanks to my colleagues and staff of IMDEA for their great help and company.

I would like to thank my parents for their support at all possible levels. I wouldn't have been able to achieve anything abroad without them. My appreciation also goes to all my cousins, aunts, and uncles who remember me even when I am far, and are always happy to see me when I am back for a few days.

Finally, my gratitude to Monica for her unworldly patience. For keep fighting, although we knew it was going to be hard. She is a cornerstone of my life.

Abstract

The operators of the Autonomous Systems (ASes) composing the Internet must deal with a constant traffic growth, while striving to reduce the overall cost-per-bit and keep an acceptable quality of service. These challenges have motivated ASes to evolve their infrastructure from basic interconnectivity strategies, using a couple transit providers and a few settlement-free peerings, to employ geographically scoped transit services (e.g. partial transit) and multiplying their peering efforts. Internet Exchange Points (IXPs), facilities allowing the establishment of sessions to multiple networks using the same infrastructure. IXPs have hence become central entities of the Internet. Although the benefits of a diverse interconnection strategy are manifold, it also encumbers the inter-domain traffic engineering process, and potentially increases the effects of incompatible interests with neighboring ASes. To efficiently manage the inter-domain traffic under such challenges, operators should rely on monitoring systems and computer supported decisions.

This thesis explores the IXP-centric inter-domain environment, the managing obstacles arising from it, and proposes mechanisms for operators to tackle them. The thesis is divided in two parts. The *first part* examines and measures the global characteristics of the inter-domain ecosystem. We characterize several IXPs around the world, comparing them in terms of their number of members and the properties of the traffic they exchange. After highlighting the problems arising from the member overlapping among IXPs, we introduce remote peering, an interconnection service that facilitates the connection to multiple IXPs. We describe this service and measure its adoption in the Internet.

In the *second part* of the thesis, we take the position of the network operators. We detail the challenges surrounding the control of inter-domain traffic in the Internet environment previously described, and introduce an operational framework aimed at facilitating its management. Subsequently, we examine methods that peering coordinators and network engineers can use to plan their infrastructure investments, by quantifying the benefits of new interconnections. Finally, we delve into the effects of conflicting business objectives among ASes. These conflicts can result in traffic distributions which do not satisfy the (business) interests of one or more ASes. We describe these dissatisfactions, differentiating their impact on the ingress and egress traffic of a single AS. Furthermore, we develop a warning system that operators can use to detect and rank those conflicts. We test our warning system using data from two real networks, where we discover a large number of traffic flows that do not satisfy the interest of operators, thus stressing the need to identify the ones having a larger impact on their network.

Table of Contents

List of Abbreviations

API Application Program Interface.

AS Autonomous System.

BGP Border Gateway Protocol.

BMP BGP Monitoring Protocol.

CDN Content Delivery Network.

CLI Command Language Interpreter.

DSS Decision Support System.

FIB Forwarding Information Base.

I2RS Interface to the Routing System.

IETF Internet Engineering Task Force.

IRR Internet Routing Registries.

ISP Internet Service Provider.

IXP Internet eXchange Point.

KDS Knowledge Discovery System.

L2 Layer-2.

LG Looking Glass.

MPLS Multiprotocol Label Switching.

PoP Point of Presence.

QoS Quality of Service.

RIB Routing Information Base.

RPKI Resource Public Key Infrastructure.

RTT Round-Trip Time.

SDN Software Defined Networking.

SRLG Shared Risk Link Group.

VPLS Virtual Private LAN Service.

List of Tables

List of Figures

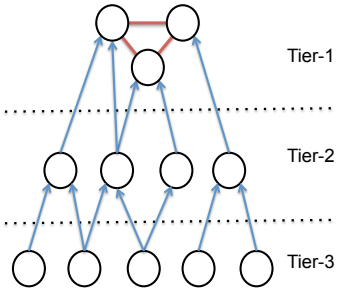
Chapter 1

Introduction

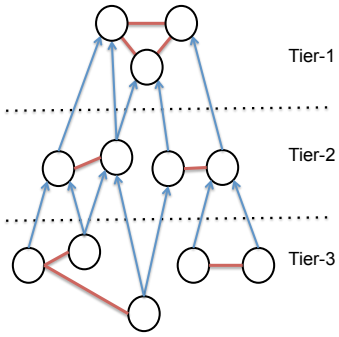
The Internet is formed by the interconnection of independently managed networks, referred to as Autonomous Systems (ASes). Operators of each AS define, following their own business interests, the interconnection infrastructure of their network, and the policy they apply to the reachability information they exchange with other ASes. At the routing level, the Internet structure and dynamics derives from the interaction among these policies.

While content linkage in the Internet looked like a web [19], the routing topology supporting it historically took the form of a *hierarchical structure* with various tiers [172] (Figure 1.1(a)). With time, this pyramid structure has been altered, driven by evolving interconnection policies and constantly changing business requirements. Namely, the appearance of new actors in the Internet ecosystem impacted the amount of traffic delivered and the manner in which it was done, leading to the establishment of many shortcuts in the original pyramid (i.e. direct connections among ASes in lower hierarchies, Figure 1.1(b)), thus leading to a *flatter structure* [57]. In practice, establishing new links to other ASes is easier said than done: Operators have to actually layout the interconnections with other networks, which normally demands significant investments in physical infrastructure and operations. These costs demotivate the extensive implementation of private peering in ASes, that is, the use of exclusive infrastructure to connect to neighboring networks.

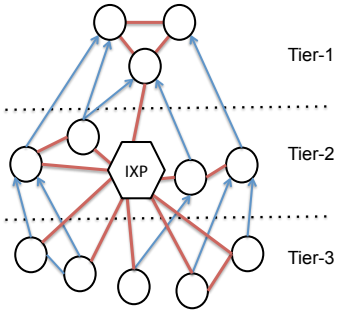
Internet eXchange Points (IXPs) are a key piece of the interconnection infrastructure, allowing for fast, flexible, and low-cost peering establishment. From a business point of view, when multiple ASes connect to an IXP, the investment in physically reaching the location of that IXP and the managing hardware at its facility opens the door to peering with any of its other members (Figure 1.1(c)). Technically, an IXP "is just a switch" providing its members the infrastructure needed to facilitate peering. It includes a Layer 2 backplane through which members exchange traffic, but also often provides complementary services such as IXP route servers, which facilitate the exchange of IP routes with its neighbors. Actually, IXPs transcend the mere technical aid, as they also become important social and governmental entities in the overall Internet ecosystem [168] [39] [186].



(a) Three tier Internet hierarchy.



(b) Peering links between ASes emerge. The structure becomes *flatter*.



(c) IXPs provide the infrastructure to connect to multiple ASes with a single link.

Figure 1.1: Evolution of the Internet structure.

The establishment of multiple peering sessions, facilitated by the proliferation of IXPs, offers operators various advantages, such as: (I) Reduction of the transit bill, as the traffic is delivered directly to settlement-free peers. (II) Decrease in latency and internal transport costs, since, by peering at multiple sites, traffic can be delivered closer to its destination point. (III) Increase the resiliency of the network, since the existence of multiple points of connectivity decreases the risk of a single point of failure.

The aforementioned benefits are certainly appealing to ASes; however, they come at a cost, as operators must deal with several challenges to fully enjoy them. In order to correctly manage their network in this environment, operators must be able to answer questions, such as:

- *Where to expand and with whom to peer?* IXPs open the possibility of peering with multiple ASes, but not all them are meaningful for the network. Many organizations have introduced to their operation the role of "peering coordinator". Their objective is to identify and negotiate potential interconnections with other networks and IXPs, in order to exchange traffic at lower cost compared to traditional transit or with better Quality of Service (QoS). To make decisions, peering coordinator and operators should be able to *quantify* the advantages of joining new IXPs. The decrease of transit traffic offers a direct way of justifying

expansion, however, due to the high number of overlapping members among IXPs, this advantage marginally decreases after the network joins a few IXPs. As described above, other valid advantages of peering with the same networks include an improved path diversity and less internal transport cost. As business-driven companies, network operators must be able to include all these other facts when planning any network expansion.

- *How to efficiently control the traffic exchanged through inter-domain links?* The connection with multiple peers increases the path diversity of the network. However, operators must actively tune their network to ensure that this diversity is utilized for the better. Operators should ensure that their inter-domain traffic is balanced across their network in the most cost-effective way (inter-domain traffic engineering). In addition, operators must leverage diversity to increase the resiliency of the network by, for instance, implementing the necessary mechanisms for back-up path propagation.

- *How to validate whether, and in which way, my neighbor's behavior is affecting the performance of my network?* The freedom offered by the Internet has facilitated its success, but complicates the overall operation of the network. Operators must be ready to detect and deal with anomalous situations such as prefix hijacking [12]), path hiding [181] [105], or route leaks [70]. Furthermore, policies of external ASes, even if considered correct, might be affecting the state of the network at different levels. There are multiple examples of the latter cases, such as "inconsistent advertisement" from neighbors connected at multiple IXPs [72], or the existence of traffic from directed connected peers through transit providers [149] [109]. Are operators ready to detect these problems and respond appropriately to avoid / tackle them?

Albeit the large efforts from both the research community and the industry, most operators do not have the necessary tools to answer these questions. Operators require the analysis of different sources of data, including traffic statistics, inter-domain paths, IXPs members, external AS policies, etc. The large size and nature of the data (which is incomplete and sometimes even erroneous), makes it hard for operators to process it. The typical heterogeneous composition of networks, which often consists of hardware from multiple vendors, and the lack of standard interfaces for data collection exacerbate this problem. A few companies certainly possess the internal developing capacity to deal with these issues and create inter-domain traffic management tools [104], however, they only represent a small percentage of the around 50000 (and counting) ASes forming the Internet.

The creation of tools for the management of inter-domain traffic can be boosted by introducing more powerful information analysis systems, as well as flexible interfaces for network data collection. The former can be made possible with the emergence of architectures and IT systems designed for big data analysis [3]. The latter could stem from the recent demand requirements of networks supporting the Software Defined Networking (SDN) model [36]. Indeed, in recent years, the demand for SDN has pressured manufacturers to implement into their systems more

flexible protocols and Application Program Interfaces (APIs), such as NETCONF/YANG [17]. Additionally, the introduction of SDN-like features in the network design cycle can create an environment prone for the analysis of network data. Therefore, we believe that network operators will have in the near future the necessary resources to analyze inter-domain traffic data.

The goal of this thesis is to define and implement a framework for the management of inter-domain traffic in the complex and evolving environment previously described. Inter-domain traffic management not only includes the techniques required to control the distribution of traffic, but also the procedures needed to handle the network infrastructure and the interaction with neighboring ASes. For instance, policies of external entities might actually be incompatible with those from the operator, i.e. no inter-domain traffic distribution will satisfy all parties involved. This problem is amplified in a network connecting with hundreds of peers (as would those joining various IXPs). Our research focuses on analytic procedures that use automated processes to aid operators and peering coordinators to make the best decisions for their network.

We divide the thesis in two main parts: **First**, in order to better understand the situation surrounding the ASes, we assess the current IXP-centric environment of the inter-domain ecosystem. We characterize and compare different IXPs around the world. Also, we examine the phenomena of remote peering, which helps ASes to reach various IXPs using the same physical infrastructure. **Second**, we focus on the specification of techniques that operators can use to improve their network management and provide the design and implementation of tools to sustain such techniques.

We describe the contents of each of these two parts hereafter.

1.1. Part one: Characterizing the IXP-centric Internet ecosystem

In this part of the thesis, we assess several characteristics of the support of the Internet ecosystem by IXP's, relying on publicly accessible resources, or measurements performed in the context of this Thesis. The objective is to understand the technical and behavioral aspects of Internet traffic exchanged at IXP, in order to set the stage for our research on the definition of efficient inter-domain traffic management procedures.

We start in **Chapter 3** by examining overall characteristics of several IXPs. We first characterize IXPs in terms of membership size and volume of exchanged traffic. We then examine the effects of member overlapping in terms of ASes and control-plane information. Finally, we look at the variation of traffic volume exchanged at IXPs at daily and monthly time granularities, notably focusing on the impact of weather on the traffic sustained by some national-scope IXPs.

Chapter 4 offers a detailed analysis of a regional IXP, the Slovakian IXP (SIX), from which we obtained an exhaustive amount of inter-temporal data. In the chapter, we study the temporal dynamics of SIX in terms of: membership size and composition (type of companies); the peering density among the members; the traffic distribution; port capacity/utilization; and the local AS-level traffic matrix. Our data revealed a number of invariant and dynamic properties of the studied system, such as the stagnation of member growth over time, counter-balanced by the explosion of

Content Providers traffic.

Chapter 5 focuses on remote peering. By using remote peering, an AS can peer at multiple IXPs without requiring the installation of infrastructure on each site. In the chapter, we describe this technique, and the benefits that it provides to both IXPs and companies. Also, the chapter includes the description of a measurement campaign we undertook to evaluate the adoption of remote peering. Concretely, we used looking glasses and active methods to measure the latency of the peering IP of members to the IXP switching back-plane. Our measurement campaigns reveal presence of remote peering at 20 IXPs worldwide, with remote peering members peaking at 20% in one IXP.

1.2. Part two: Inter-domain Traffic Management

In the second part of the thesis, we take the perspective of network operators solving the inter-domain traffic and peering engineering problem. We explain how different sources of data can be used to manage inter-domain traffic in the environment that we inspect in the first part of the thesis. Moreover, we describe two applications that operators can use to evaluate the expansion of their peering infrastructure, and detect conflicts with the policies of external ASes.

In **Chapter 6**, we introduce a general framework for inter-domain traffic management. We describe each of the procedures of the framework, including several mechanisms to gather the required inter-domain information. Furthermore, we highlight the difficulties that have hindered the implementation of inter-domain management procedures in many networks in the last decade.

Chapter 7 describes a peering expansion study that we perform for the Spanish Academic Network, RedIRIS. In this study, we leverage IXP information to enrich the peering management process. Using real network data from RedIRIS, we illustrate the decreasing benefits of traffic off-loading due to IXP member overlapping. We also discuss how Remote Peering can be used to reduce the cost of peering, and how operators could quantify other direct advantages of peering at multiple points.

In **Chapter 8**, we explain the effects of conflicting policies among different ASes. Conflicting policies lead to situations in which the (business) interests of two or more ASes are incompatible, thus driving to traffic states that eventually do not satisfy some of the networks involved. We denominate each of the cases in which an AS is not satisfied with how some of its inter-domain traffic is routed as *unsatisfied interests*. In this chapter, we meticulously define the concept of *unsatisfied interests*, differentiating those cases that affect outbound traffic to those affecting in-bound traffic. With the increasing number of peers and interconnection points, the possibility of unsatisfied interests appearing increases. We thus provide means for operators to detect and measure them in their network. We define algorithms that can be used by operators to detect these cases, and describe a prototype of a warning system that can signal to operators when critical *unsatisfied interests* occur. Furthermore, we use data from two real networks to demonstrate our tool and the frequency of *unsatisfied interests* in operative networks.

1.3. Thesis structure

Besides the two main parts described in the previous sections, the rest of this thesis is structured as follows. In Chapter 2, we provide an overview of the Internet ecosystem, including a background in BGP, IXPs, and typical traffic engineering practices performed by operators. We present ideas for future work, and conclude in Chapter 9. We provide relevant related work within each of the main chapters of this thesis.

1.4. Summary of thesis contributions and publications

Characterization of the evolution of regional IXPs. Using publicly available data, we analyzed several characteristics and practices of regional IXPs. Our study included traffic, peering, and capacity characteristics, under both a technical and economical perspective. In addition, we explored how the concentrated geographical footprint of regional IXPs could be leveraged to analyze effects of human behavior. The next are the publications related to this contribution:

- *Juan Camilo Cardona*, Rade Stanojevic. A History of an Internet eXchange Point. ACM Computer Communication Review, 42 (2). pp. 58-64. 2012.

- *Juan Camilo Cardona*, Rade Stanojevic, Ruben Cuevas. On Weather and Internet Traffic Demand (Poster). The Passive and Active Measurement Conference (PAM 2013), 18-19 March 2013, Hong Kong, China.

- *Juan Camilo Cardona*, Rade Stanojevic.IXP traffic: a macroscopic view (Paper). The 7th Latin American Networking Conference 2012, 4-5 October 2012, Medellin, Colombia.

Remote peering. We closely examine the concept of remote peering, and provide experimental measures of its adoption on world wide IXPs. Also, by using real data, we explored the economical and technical benefits of remote peering to Internet Service Providers (ISPs). The publication related to this contribution is:

- Ignacio Castro, *Juan Camilo Cardona*, Sergey Gorinsky, Pierre Francois. Remote Peering: More Peering without Internet Flattening. The 10th ACM International Conference on emerging Networking EXperiments and Technologies (ACM CoNEXT 2014), 2-5 December 2014, Sydney, Australia.

Analysis and detection of inter-domain traffic dissatisfactions due to incompatible polices. We identified scenarios in which incompatible interests lead to anomaly-free states which do not satisfy the (business) interests of one or more ASes. We defined methods for an operator to detect unsatisfied interests using network data and evaluate them using real network data. We analyzed one of these scenarios, in which the filtering of more-specific prefixes in one networks causes unexpected transit flows at remote ASes. The publications related to this contribution are:

- *Juan Camilo Cardona*, Pierre Francois, Paolo Lucente. Impact of BGP filtering on Inter-Domain Routing Policies. RFC 7789. 2016.

- *Juan Camilo Cardona*, Stefano Vissichio, Paolo Lucente, Pierre Francois. "I Can't Get No Satisfaction": Helping Autonomous Systems Identify Their Unsatisfied Inter-domain Interests. IEEE Transactions on Network and Service Management. 2016.

Inter-domain data collection and management. The difficulty of data collection is one of the problems blocking an efficient inter-domain traffic management. We provide ideas in how data could be more easily collected and processed it. The publications related to this contribution are:

- *Juan Camilo Cardona*, Pierre Francois, Paolo Lucente. Collection and Analysis of data for Inter-domain Traffic Engineering (Paper). I Workshop Pre-IETF, in conjunction with the 34th conference of the Brazilian Society of Computation (CSBC 2014), 28-31 July 2014, Brasilia, Brazil.

- Pierre Francois, *Juan Camilo Cardona*, Adam Simpson, Jeffrey Haas ADD-PATH for Route Servers. draft-francois-idr-rs-addpaths-01 (IETF Internet Draft). August 2014.

- Pierre Francois, *Juan Camilo Cardona*, Adam Simpson, Jeffrey Haas (February 2014) ADD-PATH limit capability. draft-francois-idr-addpath-limit-00 (IETF Internet Draft). February 2014.

- *Juan Camilo Cardona*, Pierre Francois, Saikat Ray, Keyur Patel, Paolo Lucente , Pradosh Mohapatra. BGP Path Marking. draft-bgp-path-marking-00. July 2013.

The next publications were also done and published during the development of this thesis:

- *Juan Camilo Cardona*, Rade Stanojevic, Nikolaos Laoutaris. Collaborative Consumption for Mobile Broadband: A Quantitative Study. The 10th ACM International Conference on emerging Networking EXperiments and Technologies (ACM CoNEXT 2014), 2-5 December 2014, Sydney, Australia.

- *Juan Camilo Cardona*, Pierre Francois, Bruno Decraene, John Scudder, Adam Simpson, Keyur Patel. Bringing High Availability to BGP: A Survey. Computer Networks 91, 788-803. 2015.

- Clarence Filsfils, Nagendra Kumar Nainar, Carlos Pignataro, *Juan Camilo Cardona*, Pierre Francois. The Segment Routing Architecture. IEEE GLOBECOMM: Next Generation Networking Symposium. 2015.

Chapter 2

Background

In this section, we provide technical and economic fundamentals required to understand the Internet ecosystem and the management of inter-domain traffic. We first describe the type of relationships that ASes establish to achieve global connectivity in Section 2.1. After that, in Section 2.2, we provide technical details of the Border Gateway Protocol (BGP), the protocol used between ASes to exchange reachability information. We review traffic engineering practices that operators use to control their inter-domain traffic in Section 2.3. We finish this chapter in Section 2.4 by explaining basic concepts of Internet Exchange Points.

2.1. Economic relationships between ASes

ASes on the Internet establish connectivity between them (i.e. Layer-2 / physical links) based on business-driven agreements. The relationships between ASes are characterized in economic terms (i.e. which AS pays for service and infrastructure costs), and in the transport services that each AS offers to the other (i.e. promise to exchange traffic to all the other networks, or just a subset of them). The basic types of agreements are *transit-customer* and *settlement-free peerings*, but other types of agreements are also possible. We extend the description of each of these below.

Transit-customer agreements are bilateral interconnections where the customer pays the provider for connectivity to the global Internet. In a common setting, transit traffic is metered at 5-minute intervals and billed on a monthly basis, with the charge computed by multiplying a per-Mbps price and the 95th percentile of the 5-minute traffic rates [59, 170]. In the early commercial Internet, traffic flowed mostly through a hierarchy of transit relationships, with a handful of tier-1 networks forming a clique at the top of the hierarchy.

Settlement-free Peering is an arrangement where two networks exchange traffic directly, rather than through a transit provider, and thereby reduce their transit costs. The exchange is commonly limited to the traffic belonging to the peering networks and their customer cones, i.e., their direct and indirect transit customers. Networks differ in their policies for recognizing another network as a potential peer. The peering policies are typically classified as open, selective, and

restrictive [120] [147]. An open policy allows the network to peer with every network. A network with a selective policy peers only if certain conditions are met. A restrictive policy has stringent terms that are difficult to satisfy. Costs of peering and transit have different structures. Peering involves a number of traffic-independent costs, e.g., collocation, IXP membership fees, etc. Peering might also have traffic-dependent costs, e.g., back-haul rates are more expensive. Over the years, peering relationships have proven themselves as cost-effective alternatives to transit.

Other peering types, such as paid-peering or partial-transit, are also possible between networks [67]. In paid-peering, a company pays the other to reach their IP prefixes and the one of its customers (similar to settlement-free peering with a cost). In partial-transit, a company provides transit services to a customer only to a subset of the Internet, normally for a lower rate than typical transit services.

Network managers decide how to control their inter-domain traffic based on technical and economic aspects. It is not economically convenient for operators to transport traffic from non-customer ASes (transit providers or settlement-free peers) to other non-customer ASes [84]. Operators selectively propagate reachability information to neighboring ASes in order to avoid these flows. The protocol used to exchange reachability information is BGP, which we explain in the next section. Afterwards, we describe the technical aspects of inter-domain traffic control in Section 2.3.

2.2. Introduction to BGP

BGP is used by an AS to inform neighboring ASes of the prefixes that can be reachable through it, and the characteristics of the path towards those prefixes. The characteristics of the path are denominated path attributes, and include information such as the sequence of ASes on the way to the origin, the preference of the local AS for each path (Local-preference), or the preference of the external AS to specific links (MED). BGP defines the best path algorithm, in which routers select the best routes from the ones available, based on their path attributes (Figure 2.3). The BGP version 4 is the base specification that an Internet router implementing BGP should support, which is specified in the RFC 4271 [155][1]. This section provides an overview of this version of the protocol. We divide the explanation in two different parts:

- **BGP Sessions.** BGP speakers connect with each other using BGP sessions, in order to exchange paths towards destinations, encoded as Network Layer Reachability Information (NLRI). BGP does not only convey routing information among Autonomous System (eBGP). The protocol is also used to exchange paths within Autonomous Systems (iBGP). We provide details on how BGP sessions are formed between peers, as well as the differences between eBGP and iBGP, in Section 2.2.1.

[1] RFCs 6286, 6608, and 6793 update the specifications of RFC 4271 and are thus also considered part of the base version of the protocol

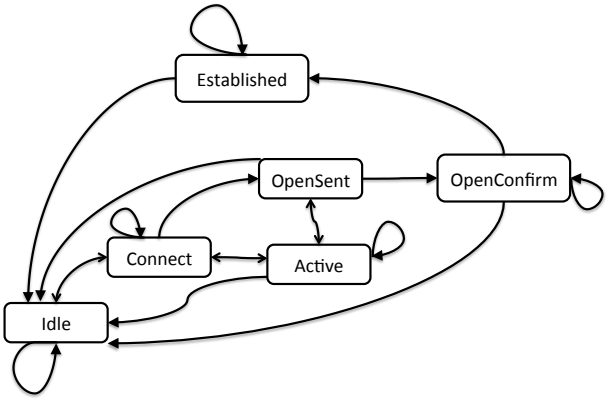


Figure 2.1: BGP Finite State Machine. [155] describes each state and the conditions for transition between states.

- **BGP Policy and Decision Process.** Each BGP speaker selects the best path for each destination, among the set of received paths through what is called the BGP decision process. This process uses the information included in BGP paths, referred to as BGP Path Attributes, to compare the known paths to the same NLRI. Since any BGP speaker can modify path attributes upon processing, ASes can influence the selection process to implement their policies, in order, for instance, to prefer settlement-free routes from those of transit providers. Such aspects of BGP are presented in Section 2.2.2.

2.2.1. BGP sessions

The BGP session between two peers is maintained over a TCP connection through which the peers exchange routing information, using different types of messages [15] [48]. The rules that govern the behavior of BGP speakers are described by the Finite State Machine (FSM) of the BGP standard (Figure 2.1) [155]. The BGP FSM defines six session states and the events that trigger state changes. The original BGP FSM was designed to be simple and did not differentiate disruptive events (e.g. node failure) from events that only partially interrupt the connection (e.g. a planned node restart). With the years, new features have been introduced into BGP sessions to preserve availability in cases where a less radical session recovery procedure can be applied. [25].

BGP defines different behaviors for external or internal BGP peers; these are refereed to as eBGP and iBGP respectively. While eBGP can be considered as the mechanism to exchange paths among ASes, iBGP is used to distribute external paths among the routers of an AS.

BGP defines several path propagation rules for iBGP sessions, such as the constraint of only announcing to iBGP peers routes received from eBGP peers. This rule would require the establishment of a full mesh of BGP sessions for path dissemination. As this approach can create scalability and operational issues in large networks, operators typically base the iBGP topology on route reflection [14] [144]. Route reflectors help controlling scalability in terms of number of BGP sessions to be maintained by each speaker. A route reflector is a BGP router that relays paths

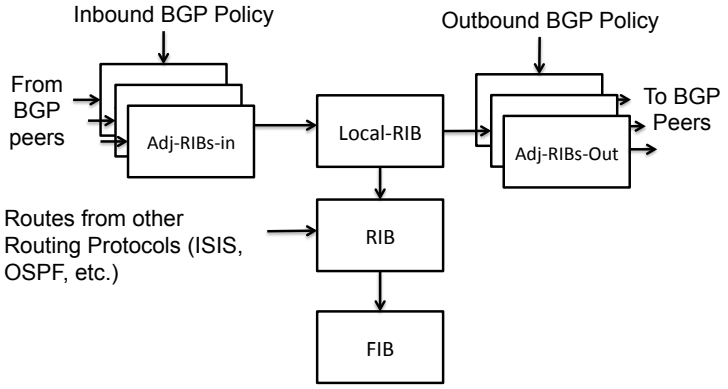


Figure 2.2: BGP RIBs.

received from iBGP peers to other iBGP peers. The set of routers to which the route reflector can announce iBGP routes are their clients. Operators can configure different route reflectors in their network and connect them using a full iBGP mesh or other route reflectors. [14] describes the rules governing the behavior of route reflectors.

Route reflectors are key components in today's iBGP networks. However, the inclusion of route reflectors in a network can reduce route diversity and lead to slow convergence in some situations [132] [96]. These problems arise from the fact that route reflectors must still obey the rule of only propagating a single best route for a given NLRI. The knowledge of multiple paths for each destination, at each node of the network, is critical for BGP high availability, as it can reduce convergence after failures for multiple seconds [179] [77].

2.2.2. BGP Policy and decision process

Each BGP router processes path updates received from neighboring BGP peers. Specifically, the router must decide whether paths should be filtered out on reception or its attributes modified; which paths should be selected as best and installed in the routing table; and which paths should be announced to other BGP peers. BGP defines three different conceptual types of Routing Information Base (RIB) that model this process inside a router. An illustration of the RIBs and their relationships is provided in Figure 2.2. Note that this description of the RIBs and their relationship are abstract concepts that model the management of routes in a BGP router. The actual implementation of the RIBs and their operation in a system depends on the router manufacturer.

1. **Adj-RIBs-In.** A BGP speaker maintains the routes received from its BGP peers in the Adj-RIBs-In and applies the locally configured policy to them. The policies applied to the received paths aim at rejecting incoming routes or modifying their attributes in order to tweak the selection of the best-path, according to the needs of the ISP. For example, paths received from customers tend to be preferred over paths received from transit providers [94] [84]. This policy is reflected by setting the local preference attribute of customer paths to a higher value than what is set for provider paths.

BGP best-path selection algorithm

1. Prefer path with highest Local preference.
2. Prefer path originated by local router.
3. Prefer path with shorter AS-path length.
4. Prefer path with lowest origin code.
5. Prefer path with lower MED (Only done if neighboring AS is the same)
6. Prefer EBGP to IBGP.
7. Prefer path with closest next-hop.
8. Prefer oldest path, if EBGP.
9. Prefer path in which the Router ID of NH is lowest.

Figure 2.3: BGP Algorithm [155].

2. **Loc-RIB.** After applying the policies, the BGP speaker selects the set of best paths using the best path selection algorithm (Figure 2.3) and stores them in the Loc-RIB. A router processes the Loc-RIB, together with other routes available to the router, to select its best path for each destination NLRI, and ultimately store them in the router's main RIB. The RIB is then further translated into a Forwarding Information Base (FIB) that is used by the router to forward packets. Note that as routes from other routing protocols might be preferred over BGP routes, not all routes in the Loc-RIB find their way into the FIB.

3. **Adj-RIBs-Out.** Finally, the router maintains RIBs aimed at tracking which paths were announced over which BGP session. These RIBs are denominated Adj-RIBs-Out and are populated after applying the policies to the routes present in the Loc-RIB. BGP outbound policies are necessary, as not all paths are to be propagated to neighboring ASes. For example, a path from a settlement-free peer should not be propagated to transit providers, since this will generate connectivity costs that would not provide any benefit for the ISP.

2.3. Inter-domain Traffic Engineering

One of the tasks of network operators is to control how inter-domain traffic flows through their network. This process is normally refereed to as *traffic engineering*. Networks are usually connected through a large number of links, each with different characteristics in terms of cost, capacity, or latency. In such environment, traffic engineering not only implies the management of link congestion, but also the reduction of expenses and the control of network performance. The distributed nature of the Internet makes it difficult for operators to perform traffic engineering on inter-domain traffic, since operators do not have complete control and knowledge over the policies of external networks. Moreover, the continuous stream of events affecting inter-domain traffic, such as network failures, traffic demand fluctuations, or new commercial agreements, forces operators to continually adjust their routing configurations to fit their policies.

Operators need to control outbound and inbound inter-domain traffic, both of them requiring different strategies. For outbound traffic, operators have certain control on the preference among the paths that they receive from external networks, even if these change in time. For inbound traffic, operators can only try to influence the decisions of external networks, thus driving operators to implement trial-and-error strategies: operators implement different techniques until finding one which fulfills their objective. We briefly expand on the techniques used for each type of traffic engineering.

Outbound Traffic engineering consists in selecting the paths that internal routers should use to forward traffic to external destinations. This is normally achieved by tweaking the attributes of the incoming paths to give priority to the ones they prefer. Operators can, for instance, change the local preference to achieve their goal [178] [91]. Some operators also use MED tweaking for this purpose, although it was initially designed for inbound TE. Other strategies rely on special communities to achieve more granular control [183].

With **Inbound Traffic engineering**, operators try to *influence* the routing decision of external ASes, in order to control where traffic for certain destination enters the network. Initially, operators should prevent incoming traffic that they are not willing to transport to other neighboring ASes, which normally means avoiding the existence of traffic flows between non-customer ASes. This is performed by not announcing (filtering) the prefixes of non-customer ASes to other non-customer ASes. The use of BGP communities to mark routes as non-customers and customer can be very useful to automatically implement this policy [60]. In addition, operator should control the amount of traffic arriving at ingress links, in order to avoid saturation. This objective is challenging. Since each AS selects its preferred path based on its private policy, it is difficult to estimate how each AS would react to changes. Therefore, ingress traffic engineering is a trial-and-error process [152]. Operators usually use AS-prepending or prefix deaggregation to influence the path selection of others AS [88] [71] [152]. Operators can also try to use MED tuning or append pre-arranged communities to influence the decision of adjacent ASes [153].

The control of the inter-domain traffic is only one of the tasks that operators must fulfill to manage their networks. The distribution of outbound and inbound inter-domain traffic depends not only on the policy of the operators, but also on the policy of external ASes. Networks operators must also be able to analyze external policies, detect when they are conflicting to their own, and make decisions on what to do if this occurs. In addition, operators must manage their network infrastructure and peering connections. In this thesis, we denominate as *traffic engineering* the process of controlling traffic in an existent network infrastructure. Complementary, we refer to as *traffic management* the overall process of controlling traffic, governing network infrastructure, and handling the policies of external peers.

2.4. Internet Exchange Points and Internet flattening

ASes benefit from establishing settlement-free relations because it reduces the traffic load of transit providers and can increase connectivity performance. Nevertheless, the cost of establishing direct peering links is non-negligible. Namely, each peering session can represent to the operator a high installation costs (including new router or switches) and considerable monthly costs (collocation in a common PoP, back-haul, operation, etc.). Normally, these high costs would prevent network to establish more than a handful of settlement-free peering sessions.

IXPs are facilities that provide ASes with the infrastructure required to establish interconnection agreements with multiple networks. IXPs offer ASes the opportunity to use the same back-haul, equipment, and operation personnel to support more than one interconnection, thus reducing the overall cost per individual peering.

Technically, an IXP offers a layer-2 domain that members use to establish BGP sessions and exchange traffic. An IXP can be implemented in different ways, from the installation of a few simple switches to more elaborated architectures that can include resiliency mechanism supported by recent technologies such as Multiprotocol Label Switching (MPLS), TRILL [63] or SDN [11]. By supporting communication and negotiation among ASes in many regions, IXPs have also become key players in the social and governmental aspects of the Internet ecosystem.

Partly due to the lower costs, peering has spread widely, with the IXPs growing into major hubs for Internet traffic. Since peering relationships bypass layer-3 transit providers, they have created a "flatter" Internet, at least on layer 3. Internet flattening refers to a reduction in the number of intermediary organizations on Internet paths [24, 57, 89]. For example, the Internet becomes flatter when a major content provider expands its own network to bypass transit providers and connect directly with eyeball networks, which primarily serve residential users.

There are more than 200 IXPs around the world [8]. IXPs can span one or multiple Point of Presence (PoP), all covered by the same layer-2 switching fabric. For instance, AMS-IX has collocation points in more than 10 sites in the Amsterdam area. In general, IXPs are different depending on their size, member composition. We examine various characteristics of IXPs in Chapters 3 and 4.

The large number of potential peers at some IXP can cause scalability problems with the respect to the number of BGP sessions on members' routers. Some IXPs offer **route servers** [105], which were designed to help network operators reduce the difficulties associated with maintaining a large number of sessions. Every route server client can receive paths from multiple ASes using a single eBGP session with the route server (similar to a route reflector for eBGP sessions). Note that in some cases, usually when there are many members in the IXP, multiple clients might announce a path to the same NLRI. Path diversity is an advantage for IXPs, as members can choose the path that better suits their policy. However, as a normal eBGP speaker, route servers can only advertise a single path per NLRI to each client. This limitation causes the route server to potentially hide paths that would be useful to their clients.

Part I

Characterizing the IXP-centric Internet ecosystem

Chapter 3

Overview of IXP characteristics

Establishing settlement-free peering sessions provides network managers with a way to lower transit cost and increase network resiliency [67]. IXPs membership is a cost-effective option to reach multiple ASes and, thus, IXPs have become key players in the modern Internet environment. Currently, there are more than 200 IXPs around the world. This chapter provides an overview of the IXP landscape, by comparing characteristics of different IXPs. We start in Section 3.1 by characterizing the size and evolution of several IXPs in terms of membership and traffic. Next, in Section 3.2, we measure and analyze the overlap across different IXPs based on their members and inter-domain routing data. For operators, the overlapping across different IXPs plays a role in technical and economic levels, which we will continue examining in Chapters 5 and 7. Finally, We discuss the common effects of human behavior on IXP traffic, and analyze the effect of weather on daily and yearly traffic trends.

3.1. Comparing number of members and maximum traffic

In this section, we make a basic comparison of the number of members and total traffic of several IXPs across the world. This study will provide an overview of the different types of IXPs and their evolution over the last years. To gather the data, we accessed the official websites of each of the IXPs. In order to obtain the data of previous years, we use the *WayBackMachine* [6], a project dedicated to storing the history of Internet websites. Table 3.1 lists the IXPs included in this analysis, together with some of their characteristics.

Figure 3.1 compares the maximum traffic and number of members for the analyzed IXPs. In the top-right corner of this figure, we find the three largest European IXPs: LINX, AMS-IX, and DE-CIX. These three IXPs could be considered pioneers in many different aspects of IXP management. In fact, they are not only similar in terms of traffic and number of members, but also in governmental and marketing strategies [93]. The top left case is PTT, the largest IXP in South America. Although PTT is similar in membership size to the three top European IXPs, their level of traffic is barely over 1Tbps. At the time of this writing, PTT is still growing at a dramatic

IXP acronym	IXP name	Location		Number of members	Peak traffic (Gbps)
		Country	City		
AMS-IX	Amsterdam Internet Exchange	Netherlands	Amsterdam	782	4600
BCIX	Berlin Commercial Internet Exchange e.V.	Germany	Berlin	82	150
BIX	Budapest Internet eXchange	Hungary	Budapest	52	234
BIX.BG	Bulgarian Internet eXchange	Bulgaria	Sofia	66	120
BNIX	Belgian National Internet Exchange	Belgium	Brussels	57	100
DE-CIX	German Commercial Internet Exchange	Germany	Frankfurt	677	4800
DIX	Danish Internet eXchange point	Denmark	Lyngby	45	60
FRANCE-IX	France-IX	France	Paris	303	560
GR-IX	Greek Internet Exchange	Greece	Athens	25	35
INEX	Internet Neutral Exchange Association	Ireland	Dublin	82	120
INTERLAN	InterLAN - Internet Exchange	Romania	Bucharest	58	70
JPIX	Japan Internet Exchange	Japan	Tokyo	150	480
JPNAP	Japan Network Access Point	Japan	Tokyo	103	560
LINX	London Internet Exchange	UK	London	704	3000
LONAP	London Network Access Point	UK	London	174	135
LU-CIX	Luxembourg Commercial Internet Exchange	Luxembourg	Luxembourg	72	40
MIX-IT	Milan Internet Exchange	Italy	Milan	177	362
MSK-IX	Moscow Internet eXchange	Russia	Moscow	400	2000
NAMEX	Nautilus Mediterranean eXchange Point	Italy	Rome	69	30
NAPAFRICA	NAPAfrica	South Africa	Johannesburg	147	64
NETNOD	Netnod Internet Exchange	Sweden	Stockholm	177	1289
NIX	Norwegian Internet eXchange	Norway	Oslo	61	60
NIX.CZ	Neutral Internet eXchange of the Czech Republic	Czech Republic	Prague	139	415
NIXI	National Internet Exchange of India - Mumbai	India	Mumbai	38	28
NL-IX	Neutral Internet Exchange	Netherlands	The Hague	556	1300
PLIX	Polish Internet Exchange	Poland	Warsaw	283	550
PTT	PTTMetro Sao Paolo	Brazil	Sao Paolo	843	1120
RIX	Reykjavik Internet Exchange	Iceland	Reykjavik	50	13
SEATTLE-IX	Seattle Internet Exchange	US	Seattle	211	540
SIX	Slovenian Internet Exchange	Slovenia	Ljubljana	26	36
SIX-SK	The Slovak Internet eXchange	Slovakia	Bratislava	54	66
SWISSIX	Swiss Internet Exchange	Switzerland	Zurich	188	90
TOP-IX	Torino Piemonte Internet Exchange	Italy	Turin	85	57
TORIX	Toronto Internet Exchange	Canada	Toronto	197	220
VIX	Vienna Internet Exchange	Austria	Vienna	125	324

Table 3.1: Properties of several IXPs in April 2016

pace, as we will show next. MSK-IX, the largest Russian IXP, and NL-IX, another large IXP from Netherlands, populate the middle part of the figure. The rest of IXPs are mostly conformed by regional IXPs that serve individual countries, or larger domains.

Let us now look into the progression of member count and traffic for the IXPs. These are depicted, for some of the largest IXPs, in Figure 3.2. The growth between 2013 and 2016 for both characteristics is depicted for all IXPs in Figure 3.3. In the latter figure, we observe how the three top IXPs show a similar and consistent growth over this period both in terms of member count and traffic. Namely, they increased close to 40% and 100% for member count and traffic size, respectively. PTT highlights as the IXP with the largest growth over this period. The member

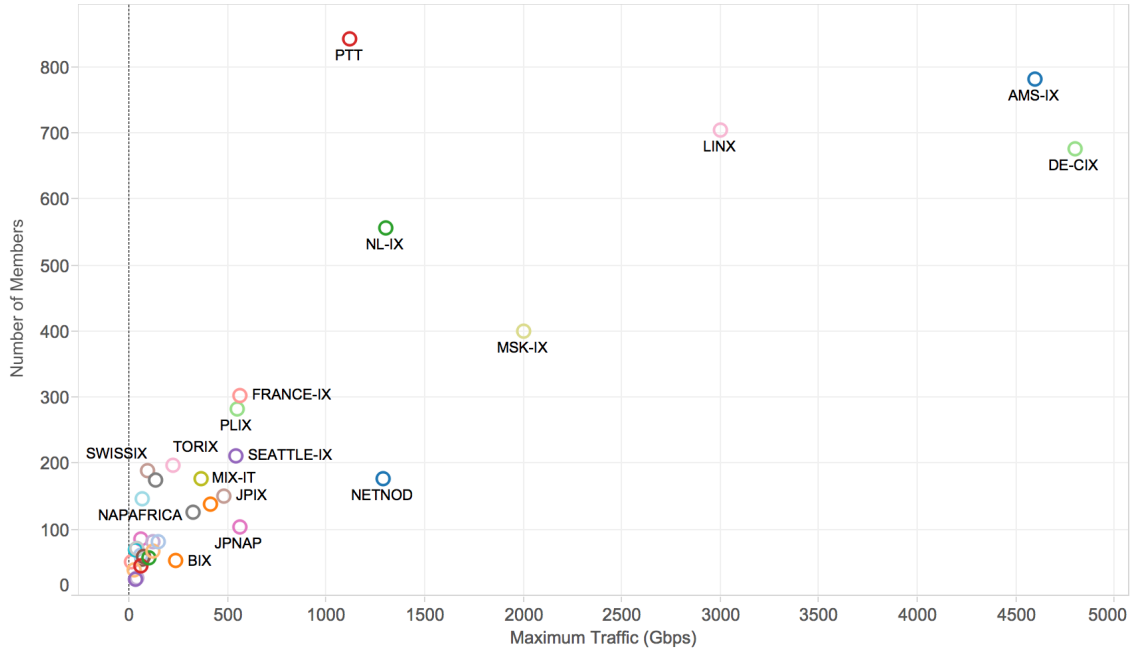


Figure 3.1: Maximum traffic versus number of members for the analyzed IXPs.

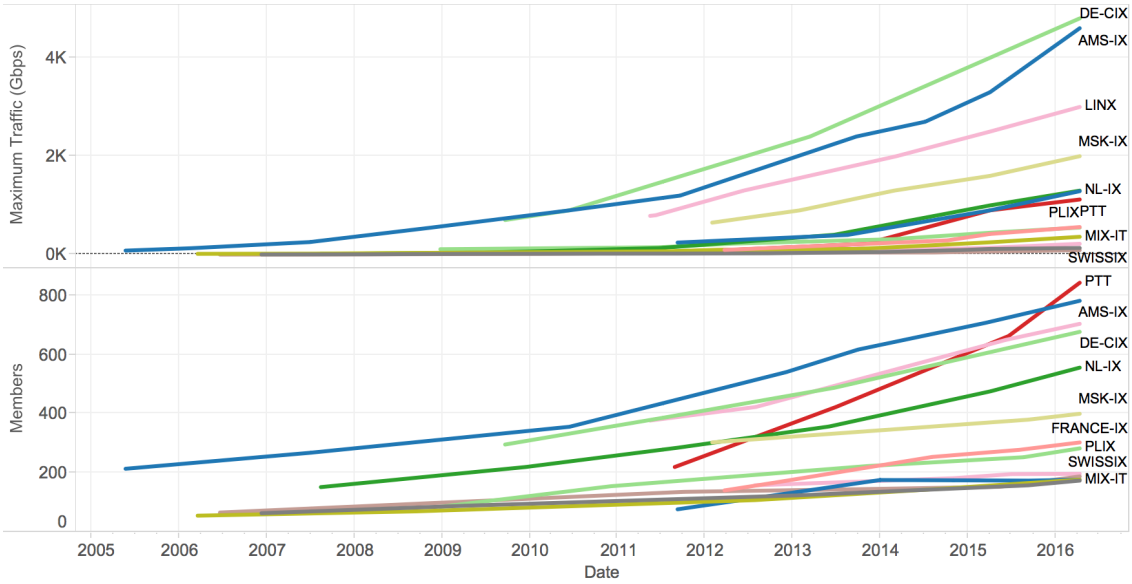


Figure 3.2: Evolution of maximum traffic and number of members for some of the large IXPs.

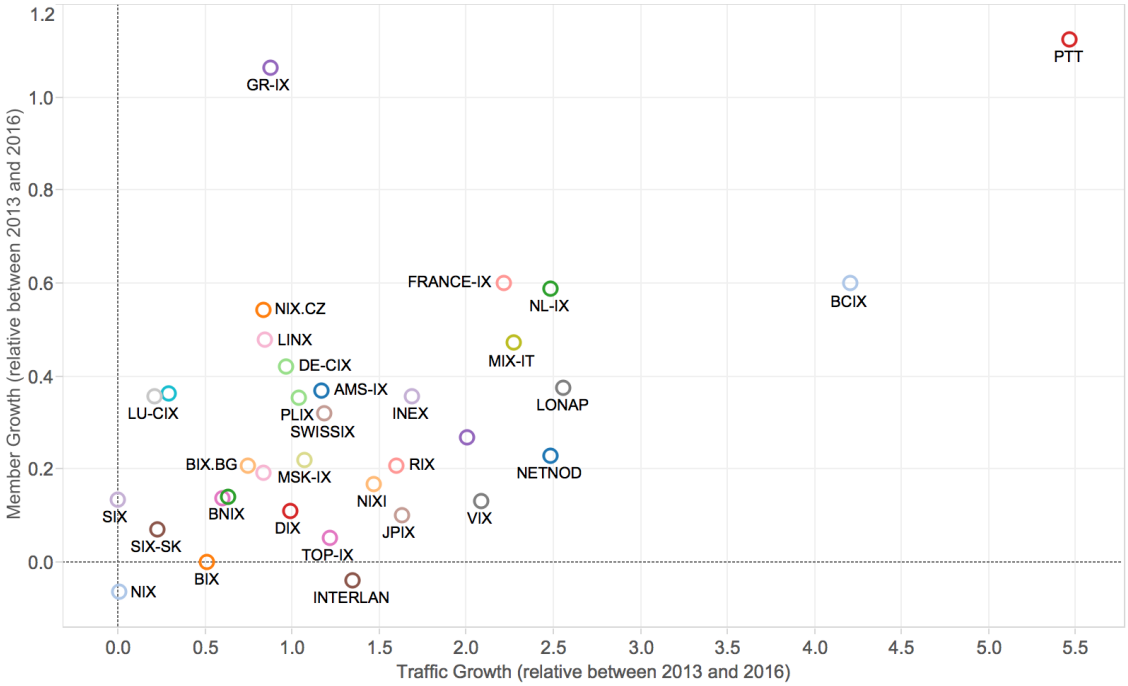


Figure 3.3: Maximum traffic growth versus number of members growth for the analyzed IXPs.

count growth of France-IX is also high among the studied IXPs. Several IXPs, including MS-IX, seem to be stagnating in terms of membership size, although their traffic growth has been steady.

The next two sections take a deeper look at common points across the IXPs for both member composition and traffic. First, in Section 3.2 we delve into the membership overlapping among IXPs, in terms of common members and IP prefix covering. Afterwards, in Section 3.3, we examine the common influence of human behavior in the traffic profile of IXPs, and, specifically, the effects of weather in daily and yearly time scales.

3.2. Overlapping among IXPs

The member composition of an IXP influences its potential to attract new members [138]. Since joining an IXP requires a non-negligible cost for a company, operators need to evaluate the benefits that this will provide to their network. Estimating the potential peers at the new IXP, and with them, the traffic the network would offload from the transit providers, is one direct way of quantifying the benefits. A more complicated situation arises when the interested company already peers with many of the ASes present at the IXP. Although there are technical advantages of peering with the same networks at multiple points, they are not as easily quantifiable and, therefore, networks might feel discourage to join a new IXP composed by many members with which they already peer.

The objective of this section is to provide a basic characterization of membership overlapping, and their effects, among our studied IXPs. We first look at this dimension using the number of

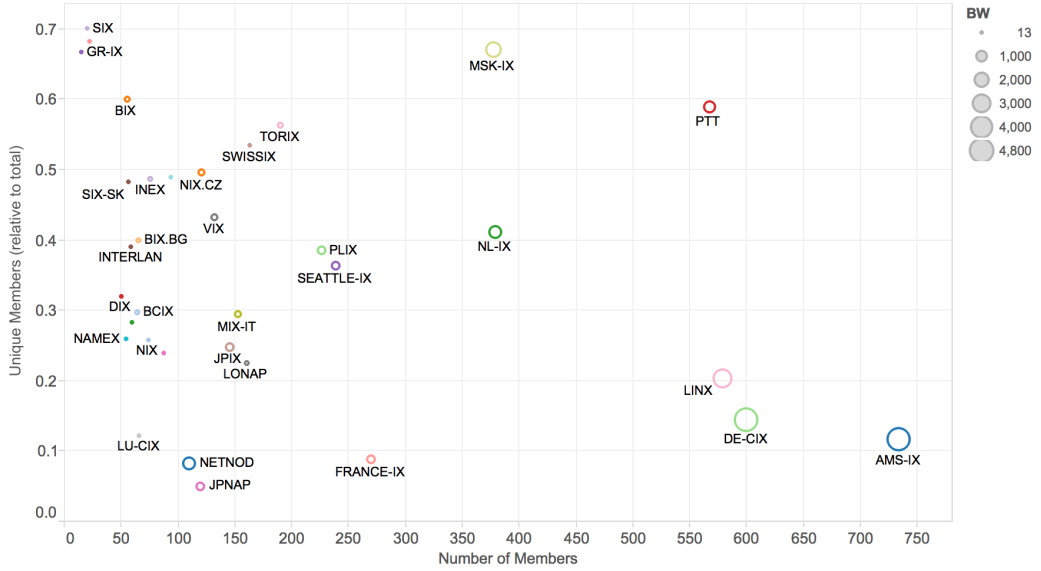


Figure 3.4: Unique members vs total number of members for the analyzed IXPs.

unique ASes of each IXP. We then perform a more detailed study of overlapping analysis using BGP information (control-plane).

3.2.1. Number of unique members

We start the analysis of member overlapping by looking at the number of ASes that are only present at each particular IXPs. To obtain the list of members for each IXP, we use the data from [111]. We calculated the unique number of members of each IXP also using this data.

Figure 3.4 compares the amount of unique ASes (relative to the total) against the total number of members. This graph helps us find characteristics of the IXPs, which where not feasible to obtain using the figures of the previous section. Hereafter, we analyze this figure, and provide some observations that might explain the location of each IXP therein.

Figure 3.4 locates LINX, DE-CIX and AMS-IX in a similar region of the figure since less than 20% of their members peer exclusively there. MSK-IK and PTT are the main IXPs of a large, and relatively disconnected region, which can explain their position at the upper-right part of the figure. NL-IX is an interesting case. Its size and unique member composition places it close to MSK-IX and PTT, yet, this IXP is in a well connected geographical area (Netherlands) and close to AMS-IX. The left side of the figure (less than 300 members) contains the rest of the IXPs. Torix (Toronto) and NapAfrica (South Africa) serve similar regions than MSK-IX and PTT, but they still show a relative small size compared to the latter. One wonders whether there is still growth potential for these IXPs, although more information on other IXPs of their regions, or even the social / political behavior of the companies operating on them is needed to understand the reason for this (for an example, please refer to [66]). The other IXPs in the upper-left side of the Figure (RIX, GR-IX, SwissIX, BIX, INEX, etc.) are IXPs serving smaller regions. We take a

detailed look at the characterization and evolution of one these IXPs, the Slovakian IX (SIX.SK), in Chapter 4. Finally, France-IX, NETNOD, LONAP, and LU-CIX stand out at the bottom left part of the figure. These IXPs could be located in regions in which companies can easily reach other IXPs, but still provide a valuable service. That is, they might not be the top of choice for companies only peering at single exchange, but they might be appropriate for companies that want to have back-up connectivity (e.g. LONAP at London), or want to exchange traffic locally (e.g. exchange the traffic in Paris, instead of Amsterdam).

3.2.2. Inter-domain routing reachability

Studying the unique ASes and common member base of IXPs provides information that can help us characterize IXPs in social or governmental dimensions. Nevertheless, the amount of unique networks does not offer an exact estimate of the technical benefits of joining a new IXP, since they do not directly reflect the traffic offload potential. In this Section, we study the member overlapping from the control-plane perspective, that is, based on the IP prefix reachability of each of the members of the IXPs.

Analyzing the control plane overlapping among IXPs is not a trivial task due to the large number of variables involved. First, it is unrealistic to only perform an analysis based on all members of each IXP. Many members have restrictive peering policies (e.g. Tier-1s) and, would never peer with small networks (in other words, for many companies, joining IXPs with many restrictive members is not beneficial). Therefore, a complete analysis would require the inclusion of the peering policy of the members. In addition, we would like to not restrict the analysis to pairs of IXPs (AMS-IX with LINX; SIX-SK with NETNOD, etc.), but a methodology that covers all IXPs.

In order to provide a short, but insightful analysis of control-plane overlapping comparison, we follow an approach similar to those of a company making decisions on IXP expansion: (I) We calculate the set of new hosts (/32 addresses) reachable through each individual IXP. (II) We then choose the IXP that provides the maximum number of new hosts. (III) We repeat the process until all IXPs are selected. At iteration N of the algorithm, we will get an approximation of the control plane covering obtained by joining N IXPs. We note that this greedy algorithm might yield sub-optimal results, however, it is much simpler to implement than an optimal one that would look for the N IXPs that maximize the covering IP space [121].

We repeat the experiment assuming that the company can only peer with members with four different types of peering policy, which we describe next. (1) all IXP members excluding Tier-1s. (2) members with polices publicly declared at the site Peeringdb [147] as open or selective. (3) Top ten members (in terms of number of hosts) with open and selective policies. (4) Members with open policy. We include for this study all IXPs described in the Euro-IX website on June 2014 [7]. We use the Prefix to AS mappings [22] and customer-cone data [23] from CAIDA to perform the analysis.

The results of this experiment are illustrated in Figure 3.5. The X-axis of the graph represents

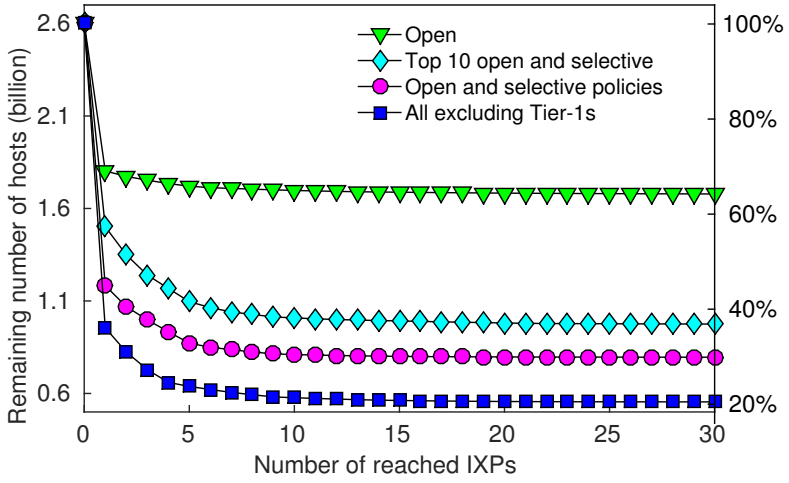


Figure 3.5: Generalized additional value of reaching an extra IXP.

Policy	First iteration	Second iteration	Third iteration
Open	AMS-IX	Terremark	DEC-IX
Top 10 open and selective	AMS-IX	Terremark	CoreSite Any2
Open and selective	LINX	JNAP	Terremark
All except Tier-1	Terremark	JNAP	CoreSite Any2

Table 3.2: IXPs corresponding to the first three iterations of control-plane offloading for different peering policies.

each iteration of the process. The Y-axis represents the amount of remaining reachable hosts after each iteration. Note that the IXP order depends on the type of policy selected. Table 3.2 includes the first 3 IXPs for all types of policies.

We observe that after the first iteration, the amount of new hosts quickly becomes marginal (around 5% additional covered hosts in the best case). In summary, we can see how member overlapping reduces the benefits in terms of new covered hosts after peering at one large IXP.

If joining a large IXP already covers most of the hosts that can be reached, why should a company join new ones? Peering at various IXPs delivers other types of benefit such as better resiliency in case the first one fails, or reducing of the internal network transport (bit-mile cost [150]). The problem is that joining each new IXP can be expensive. In Chapter 5 we analyze remote peering, a simple technique that can be used by operators to peer in various IXPs and reduce infrastructure / operational costs.

Remark 1. Note that the number of hosts is an inaccurate way of measuring the benefits of enrolling at an IXP. To calculate the real advantages (in terms of offload traffic), the egress traffic demand of the company is needed. The control-plane study, however, is globally valid, as it does not require the use of data from any company. In Chapter 7, we perform a specific peering analysis using data from the Spanish Network RedIRIS.

Remark 2. This study assumes that ASes announce all prefixes at each IXP where they peer.

This is not always the case. In Chapter 8, we examine the methods that allow an IXP to assess whether other companies are announcing all prefixes that they are expected to announce at each BGP session.

3.3. Short and long term trends in IXP traffic: A weather impact study

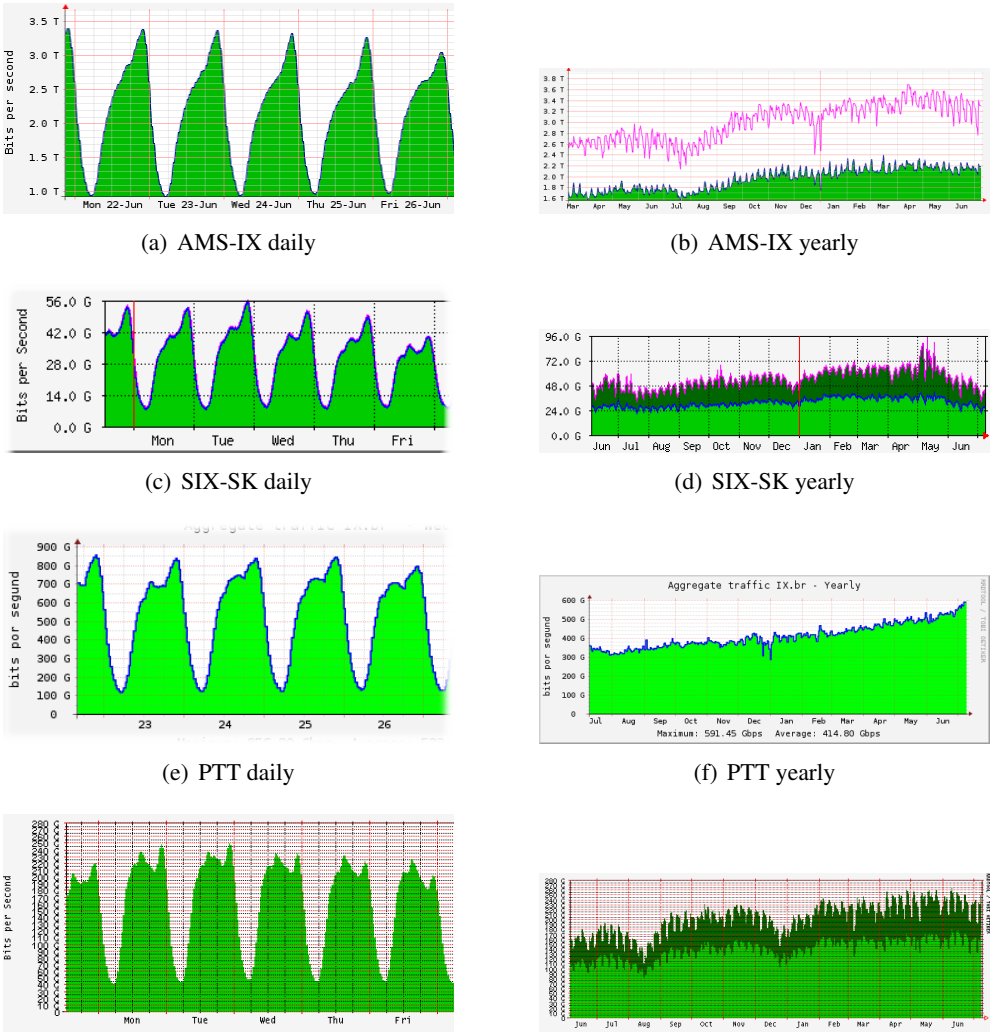
In Section 3.1, we compared the different IXPs in terms of the maximum total traffic, and their growth over various years. In this section, we examine IXP traffic at more granular time levels, by looking at the dynamics of this traffic over months or days.

Figure 3.6 contains the daily and yearly traffic for four different IXPs. These graphs allow us to observe shared characteristics of traffic among them. Concerning **daily traffic** (left column), we can see the periodic behavior related to a residential demand cycle, resembling those of the same type for utilities such as energy [145] [122] or mobile data consumption [112] [32]. The daily profiles for MIX, PTT, and SIX-SK also show the peaks related to the demand profile of commercial entities (i.e. two peaks around 11am and 5pm) [112]. In the **yearly traffic** (right column), we also observe common effects. All four IXPs show a drop of traffic around the end of the year. In addition, AMS-IX, MIX-IT and, in less effect, SIX-SK, show a smoother drop of traffic in the middle part of the year. One could argue that these drops correspond to new years holidays and the summer period on those countries, respectively. These daily and yearly patterns are common among many IXPs, which arguments that Internet traffic (at least on IXPs) is human driven[1].

As shown in the above figures, external factors affecting human behavior can have an impact on IXP traffic, yet, their effect has been overlooked. Previous studies have extensively leveraged the periodic behavior of Internet traffic for the analysis, modeling and forecasting of Internet traffic [143] [30, 41], but have not modeled the dependence of traffic with external factors, which are typically accounted as noise [45, 136, 143]. Curious about this effect, in this section we measure the direct impact in IXP traffic of one of the most important external factors affecting human behavior: *the weather*. While it has been known that the weather has a significant impact on the demand of utilities [74] or TV ratings [158], their relationship with the Internet traffic demand is not well understood. Next, we show that, in monthly periods, the traffic of IXPs has variations correlated to the temperature of the regions they are located. Also, we show the short-term effects caused by precipitation in three different regional IXPs. In these three IXPs, the effects of precipitation were similar, peaking at summer months and in the period between 16h and 18h. Our work complements other empirical approaches that have analyzed the effects of other external events, such as sport events [1], to Internet traffic.

We take a empirical approach to study the effects of weather on IXP traffic, and compare its

[1]Some reports, like [194], claim that a large percentage of Internet traffic comes from bots. If this were the case, one would expect a much lower effect of human related activities on the overall traffic of IXPs.



(a) AMS-IX daily

(b) AMS-IX yearly

(c) SIX-SK daily

(d) SIX-SK yearly

(e) PTT daily

(f) PTT yearly

(g) MIX-IT daily

(h) MIX-IT yearly

Figure 3.6: Weekly and yearly traffic for four IXPs.

influence among the different IXPs [33]. We obtain granular data for IXP traffic and the weather at the geographical domains they cover, and study short-term and long-term effects. Short-term weather events, like thunderstorms, snow or heavy wind, impact the accessibility/importance of the Internet access, and have direct, short-term, effect on the traffic demand. Longer-term effects reflected through seasonal changes in temperature and daylight duration, have slower and not so direct and immediate influence on the way the Internet traffic is generated.

Short-term effects. An example of a short-time effect is rainfall or other form of precipitation, typically lasting less than a few hours. We find that the precipitation periods have a tendency of increased traffic demand. However, this tendency is dependent on the time of the day and is most noticeable in the late afternoon, when precipitation tends to increase the traffic demand for around 6%.

Long-term effects. The periodicity due to end-user temporal cycles is a widely known prop-

erty of the Internet traffic demand. While the daily and weekly periodicity have been studied extensively [45, 136, 143], and applied in various domains from bulk-traffic transfer scheduling [115] to energy management [45], the seasonal variability over 12-month periods is not well understood. This is partly because the Internet traffic demand has been dominated by exponential growth, on top of which seasonal changes may be hard to observe and quantify. Using six IXPs as vantage points, geographically spread across the globe, allows us to study the season-dependent traffic variability in various climates. Our data suggests that seasonal traffic variability is strongly tied with temperature variability over the year: the regions far from the equator exhibiting strong seasonal traffic variation while the regions close to the equator show no such seasonal traffic changes.

3.3.1. Datasets description

To understand the interaction between the traffic demand and weather conditions we collect and use a number of datasets, which we describe below.

3.3.1.1. Traffic data

We use the traffic from several IXPs to perform our analysis. IXP traffic could be influenced by the weather, due to their tendency of transporting regional traffic [171] [8]. Many IXPs publish the traffic data aggregated across all members. We use this information for our analysis in short and long terms.

To analyze the *short-term* weather impact on IXP traffic, we have to focus on IXPs with a limited geographical area. A large IXP can carry traffic from a larger regional footprint and should, therefore, be avoided because it is difficult for the weather to be consistent over large geographies. Instead, for this purpose, we use the data from three, relatively small, regional IXPs: the Slovak-IX, FICIX and INEX. We obtained 5-minute granular traffic from each IXP by storing and processing their publicly available *mrtg* images textsfmrtg/rrdtool. Our Internet traffic dataset includes 8 months of data from INEX and 18 months of data from Slovak-IX and FICIX. FICIX was not included in the analysis of previous sections, due to the lack of historical data for more than 3 years. FICIX, however, provided the APIs to easily the traffic data automatically at low granularity levels, thus making it suitable for the weather study.

Concerning long-term effects, we need traffic demand trends over larger time intervals (12 months or more). We argue that the traffic over time scale is correlated with seasonal changes. We use the yearly graphs of *mrtg* output from several IXPs around the world: the largest European (AMS-IX), Northern American (TORIX), Southern American (PTT), Australian (WAIX), and Indian (NIXI) IXP that report their traffic information publicly. In addition to those 5 IXPs, we also utilize one medium sized IXP from central Europe, Slovak-IX (SIX). We did not include WAIX in our previous example due to its lack of historical data, but, in this section, it allows us to show the influence of the long-term trend in that part of the world.

IXP name	City	Duration (months)	Granularity	Peak traffic	# of ISPs
AMS-IX	Amsterdam	16	$24h$	$1.4Tbps$	463
TORIX	Toronto	13	$24h$	$70Gbps$	144
PTT	Sao Paolo	13	$24h$	$60Gbps$	243
WAIX	Perth	13	$24h$	$2Gbps$	64
NIXI	Mumbai	13	$24h$	$8Gbps$	31
SIX-SK	Bratislava	24	$24h$	$40Gbps$	52
SIX-SK	Bratislava	18	$5min$	$54Gbps$	52
FICIX	Helsinki	18	$5min$	$35Gbps$	28
INEX	Dublin	8	$5min$	$31Gbps$	56

Table 3.3: The traffic datasets and details of corresponding IXPs. The data was collected in 2012, corresponding traffic levels of that year.

Table 3.3 contains the details about the entire traffic dataset (duration, granularity) and the IXPs: the peak traffic, the number of ISPs participating at the IXP and the annual growth rate (AGR)[2].

3.3.1.2. Weather data

There are several online sources of free historical weather datasets. Since they report quantities that are easily measurable, they are very similar between each other. Some inconsistencies may exist between different datasets, but they are typically very small (e.g. the measurement point at the city center and the airport are likely to have slightly different weather conditions.). In this section, we use the weather data provided by the Weather Underground, easily accessible database available at `http://www.wunderground.com/`.

We queried the database for the cities that host the IXPs listed in Table 3.3, for the period that covers our traffic data. The arguments for a single query are the `location` and `day`. The output is a table that specifies the weather conditions in the given `location` and on the given `day` with the granularity of 30 minutes; i.e. every 30 minutes weather parameters are reported. The `wunderground.com` publishes a number of weather parameters including temperature, dew point, humidity, pressure, and precipitation. Granularity of 30 minutes allows fine analysis of the relationship between the weather and traffic, though some very short events such as hail or short storms, may be missed in the 30-minute sampling.

3.3.2. Short-term effects: Traffic vs. precipitation

We start the analysis by examining the dependence between the traffic demand and the precipitation. For that purpose, we use the fine grain traffic data described in Section 3.3.1. Such data allows us to notice changes that happen on short-time scales (hourly) and compare them against the weather conditions. As we explain above, we focus on precipitation as the parameter that is most likely to affect the instantaneous human behavior towards the Internet usage.

[2]AGR is defined as the ratio between the traffic in the same month(s) in 2011 and 2010; see Eq (3.1).

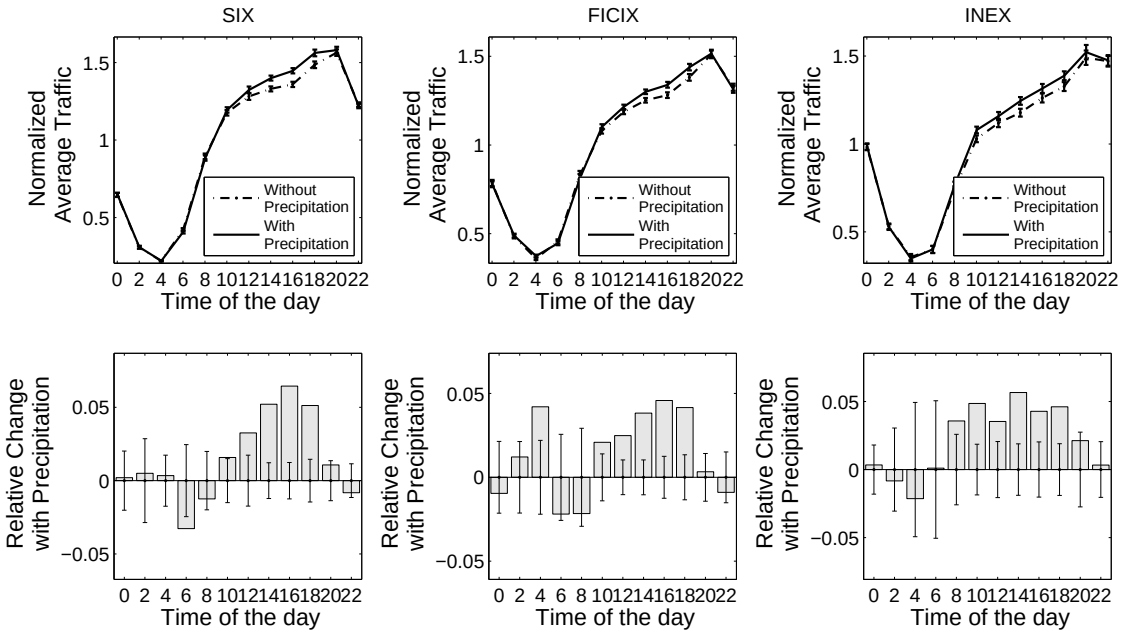


Figure 3.7: Normalized daily SIX demand with and without precipitation.

For that purpose, we split the time into 2-hour time-slots. For the ISP A, $u_A(t)$ and $d_A(t)$ denote the average upstream and downstream traffic (in $Mbps$) at time slot t. With

$$u(t) = d(t) = \sum_{\text{for all members } A \text{ of SIX}} u_A(t)$$

We will denote the aggregate traffic exchanged at the IXP with $u(t)$. Finally, in order to smooth-out the seasonal effects (e.g. Yearly graphs can show that March traffic is higher than the July traffic; see Section 3.3.3 for more details) we normalize $u(t)$ with the average traffic over a two week period centered at t:

$$\bar{u}(t) = \frac{u(t)}{average(u(t-84), \ldots, u(t+84))}.$$

For each time-slot t there are 4 weather reports (sometimes more than 4, depending on reporting system configuration) in our weather dataset, and we set a binary variable $wet_\alpha(t)$ to be 1 if the fraction of the weather reports from time slot t that report precipitation (snow, shower, rain, storm...) is not smaller than α, otherwise $wet_\alpha(t) = 0$, where $\alpha \in (0, 1]$ is a parameter. In other words if the fraction of the time it precipitates during the time slot is greater than or equal to α, we declare that the time slot to be wet, otherwise we declare it non-wet. We use $\alpha = 0.15$ unless we specify a different value, thus announcing a slot 'wet' if it receives precipitation for at least 15% of the time. Furthermore, we also declare a slot to be wet if any of the two preceding time slots presented an alpha of at least 15%.

Our goal is to observe whether time series $wet_\alpha(t)$ and $\bar{u}(t)$ are correlated. To that end, we

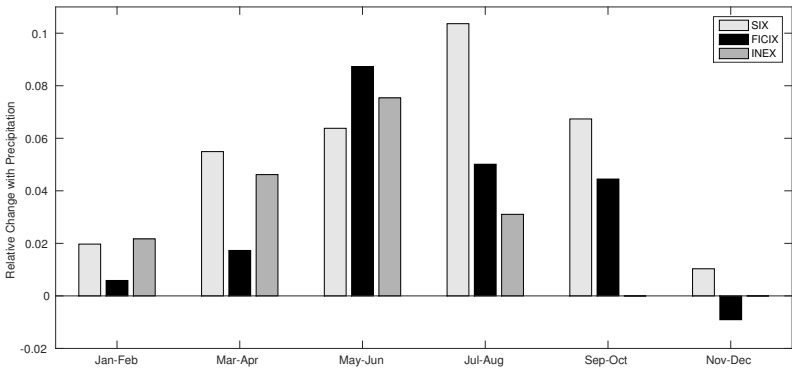


Figure 3.8: The relative change with precipitation during the $16h - 18h$ slot over the year.

split the day in twelve 2-hour intervals, and calculate average normalized traffic with and without precipitation for each of the twelve intervals:

$$A(i) = \frac{\sum_{\text{mod }(s,12)=i} \bar{u}(s) wet_\alpha(s)}{\sum_{\text{mod }(s,12)=i} wet_\alpha(s)} \quad i = 0..11$$

$$B(i) = \frac{\sum_{\text{mod }(s,12)=i} \bar{u}(s)(1 - wet_\alpha(s))}{\sum_{\text{mod }(s,12)=i} (1 - wet_\alpha(s))} \quad i = 0..11$$

thus for the twelve time intervals $0h - 2h, 2h - 4h, \ldots, 22h - 24h$, $A(i)$ and $B(i)$ represent the average normalized load in the interval $[2ih, (2i + 2)h]$ with and without precipitation, respectively.

In Figure 3.7(top) we depict the values of $A(i)$ and $B(i)$ together with the relative difference between $B(i)$ and $A(i)$ for the three IXPs. To determine if the difference between $A(i)$ and $B(i)$ is statistically significant to claim that the means of the samples with and without precipitation are different, we use Welch's t-test [190], which is well-suited for this case as the number of samples for each random variable is different and relatively large[3]. Figure 3.7(bottom) also includes the interval outside of which Welch's t-test rejects the null-hypothesis for a significance level of 0.05. Thus from early afternoon to early evening, with 95% of confidence we can affirm for all IXPs that the mean normalized traffic is larger in timeslots with precipitation than in timeslots without precipitation. For the other periods of the day (except for a couple of time period in the early morning), the difference between the means is not statistically significant to support that precipitation impacts the traffic.

Finally, we observe that the impact of precipitation is not uniform across the year. Namely, in Figure 3.8 we depict the relative increment of precipitation during the $16h - 18h$ interval for the 6 two-month periods and observe that the impact of precipitation is most pronounced in the

[3]In all cases the number of samples obtained is larger than 40.

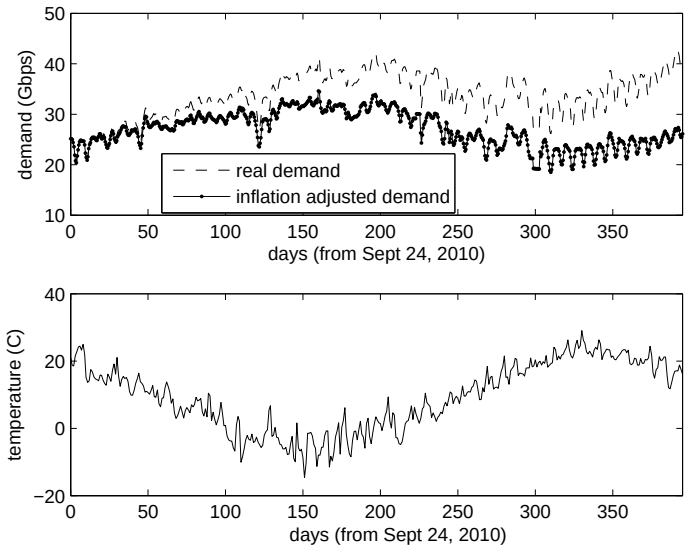


Figure 3.9: The real and inflation adjusted demand at TORIX (top). Average daily temperature at Toronto (bottom).

summer months, while it is insignificant over the winter.

3.3.3. Long-term effects: Demand vs. temperature

Yearly seasonality of the Internet traffic has been poorly understood phenomenon for a number of reasons. Firstly, the data on the traffic over long-time periods is rarely available. Secondly, since the early days of the commercial Internet, the traffic dynamics have been dominated by an exponential growth, which makes the characterization of the seasonal effects challenging. Finally, seasonal effects may not be present at all in some geographical regions.

In this section, we examine the (yearly) seasonality trends of IXP traffic for several geographically diverse set of IXPs. The data used here covers 6 regions, from 5 continents and is described in detail in Section 3.3.1.1. The traffic demand datasets we collected cover over 1 year time span, which is a minimal duration for inferring basic characteristics of the yearly seasonality.

To extract the seasonal effects from the exponential growth of the traffic demand, we utilize the following procedure. For the traffic time series covering a time period of 365+Δ days[4], we calculate the *annual growth rate* (AGR) as the ratio between the traffic in the last Δ and the first Δ days:

$$AGR = \frac{\text{total traffic in days 366 to } 365 + \Delta}{\text{total traffic in days 1 to } \Delta}. \tag{3.1}$$

. Then, we calculate the inflation adjusted demand (IAD) at day t as:

[4]In different datasets, the number of days on top of a full year, Δ, varies from 30 to 180 days, see Table 3.3.

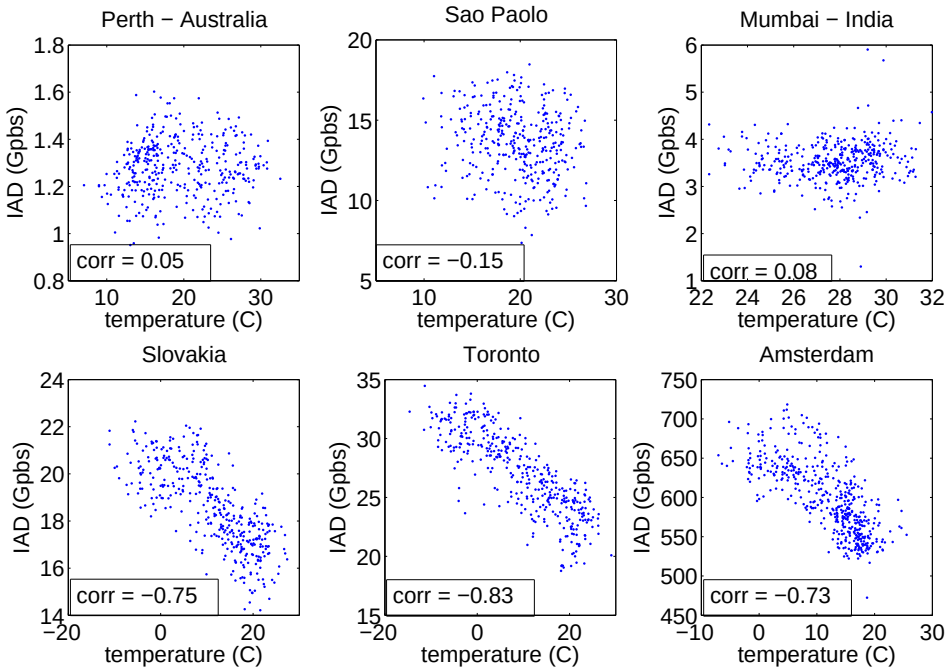


Figure 3.10: Correlation between the average daily temperature and the inflation adjusted demand (IAD).

$$IAD(t) = \frac{D(t)}{AGR^{\frac{t}{365}}},$$

where $D(t)$ is the average demand (in $Mbps$) at day t. Intuitively, the inflation adjusted demand, filters out the exponential growth of the Internet demand and allows us to focus on seasonal effects.

To illustrate the inflation adjustment operation we depict both the real and the inflation adjusted demand for the TORIX in Figure 3.9. We can see that the traffic demand, after inflation adjustment, appears to follow a sin-wave pattern that is out-of-phase with the average daily temperature observed in Toronto.

To visualize the correlation between the time series of average daily inflation adjusted demand $IAD(t)$ and the average daily temperature $Y(t)$, we use the scatter graphs in Figure 3.10. There we depict the scatter graphs for six IXPs listed in Table 3.3. One can observe that the IXPs located close to the equator (WAIX, PTT.br, NIXI) show no visually obvious dependence between the IAD and the temperature, while the dependence between the IAD and the temperature is visually observable from the Figure 3.10 for those IXPs that lie in the regions with large temperature variation. In order to quantify the correlation between $IAD(\cdot)$ and $Y(\cdot)$ we also calculate the

Pearson correlation coefficient:

$$corr = \frac{\sum_{t=1}^{N}(Y(t) - \bar{Y})(IAD(t) - \overline{IAD})}{\sqrt{\sum_{t=1}^{N}(Y(t) - \bar{Y})^2}\sqrt{\sum_{t=1}^{N}(IAD(t) - \overline{IAD})^2}},$$

where we used the notation $\bar{X}$ for the sample mean of the time series X. The Pearson coefficient is a metric that ranges between [-1,1]. The closer the value is to 0 the less correlated are the variables, whereas positive and negative values closer to 1 and -1 indicate strong positive and negative correlation, respectively.

The Pearson correlation coefficient is reported in the same figure, for all of the six studied IXPs. For the three IXPs close to the equator, the $corr$ value is very close to zero, indicating nonexistence of any significant correlation. For the other three IXPs, the $corr$ value is relatively high indicating strong negative correlation between the temperature and the demand.

From Figure 3.10 we can also observe relatively high variability of the daily traffic averages. While the correlation with the temperature may explain some of this variability, there are still many external factors (eg. social events) with unknown influence on the Internet traffic load.

3.4. Related Work

IXP characterization and analysis. [8] presents a measurement study of the peering relationships at IXPs worldwide, in which they report around 200 operational IXPs, and rich topological data on the peering relationships happening at these IXPs. [93] studies the community formation across the member composition of different IXPs. [168] performs a bottom-up analysis of network participation across IXPs from governance perspective.

Effects of human behavior and weather in the Internet. Aben [1] has analyzed the fluctuations of the traffic at different IXPs during major football events. Concerning the weather, Shulman and Spring [162] observe a strong correlation between the IP network failures and the weather conditions. Such failures are likely to be the consequence of direct impact of the weather on the ISP infrastructure (such as equipment damage by lightning strikes or degradation of satellite link quality due to high humidity) and are independent of the human behavior. More recently, Kondor et al. collected and provide visualization of the mobile traffic consumption for granular regions of different cities [112].

3.5. Summary

In this chapter, we provide an overview of several characteristics of IXPs. We first compare the IXPs in terms of their member number and maximum total traffic. We then investigated the membership overlapping across different IXPs. Finally, we examined the daily and monthly variations of traffic of IXPs, and their relationship to the weather and temperature at the regions

they serve. We describe hereafter the main conclusions of this section, and their relations with other parts of this thesis.

Different types of IXPs. The three large European IXPs (LINX, AMS-IX and DE-CIX) are similar in terms of size, traffic, and member compositions. Besides these cases, and other few large exceptions, most other IXPs serve medium to small regions, providing operators with sites in which local traffic can be exchanged. [2] contains a detailed description of one of the large IXPs. Complementary, in Chapter 4, we provide a complete characterization and evolution of SIX-SK, one of the regional IXPs.

High overlapping across IXPs. Member overlapping reduces the potential traffic offload, after a network establishes presence in one or two (large) IXPs. Peering at multiple IXPs is also valuable for resiliency and diversity purposes [25], but it is harder for operators to (economically) quantify the value of these benefits. Operators thus require methods to reduce the IXP entry costs, allowing them to scale their presence to multiple IXPs. Remote peering is one technique that can be used for that purpose. We examine remote peering with detail in Chapter 5.

Relationships between IXP traffic and weather. For IXPs located in regions with calendar seasons, we discovered an inverse relation between temperature and traffic. Regarding short-scale variations, we found significant influence of weather in the daily traffic profile of three regional IXPs, peaking in the period between 16h and 18h. In networks, capacity acquisition and network planning is performed using peak traffic values. SDN technologies could drive to flexible models, similar to those of energy [134], in which traffic is billed based on shorter time periods [46]. The relationship between traffic and weather can be useful in those scenarios to enhance the prediction models used for optimal capacity acquisition.

Chapter 4

History of an IXP

The previous chapter examines and compares overall characteristics of IXPs around the world. We illustrated some of the differences between large European IXPs to the regional ones. In this chapter, we study with detail the evolution of one regional IXP: the Slovak IX (SIX)[1]. Like other regional IXPs, SIX aggregates most ISPs operating in the country/region, thus allowing us to accurate characterize the ecosystem of the local ISPs. The detailed data on the SIX operation has been published on the SIX website since its inception. It offers an unique opportunity to understand the dynamics over a long (on the Internet timescale) time horizon of several structures including: the low-tier ISP peering, inter-AS link utilization, port capacity upgrade practices and the local AS-level traffic matrix. In particular, we observe high stability of IXP peering sessions; a pronounced imbalance of inbound/outbound traffic; and skewness of the amount of traffic exchanged among peering pairs.

We start this chapter by describing the data-set we use to perform our study in Section 4.1. We then illustrate the evolution of various characteristics of SIX in Section 4.2. We discuss several of our observations and their potential impact in Section 4.3. We describe related work in Section 4.4 and conclude in Section 4.5.

4.1. The dataset

IXPs sometimes publish data on their operation. Among IXPs that publish (or published) the peering matrix[2] and per-member traffic data, we choose to study the one that has done so from its beginning till today: the Slovak IX[3]. SIX hosted in 2011 52 ISPs, which exchange around $50Gbps$ of traffic in the peak hour. At each instance of time from 1997 onwards, SIX has published two sources of data that we use in our work: (1) the peering matrix and (2) Multi Router

[1] We refer to the Slovak IX as SIX in this chapter. We used SIX-SK in the previous chapter to differentiate it from the Slovenian IX.

[2] Boolean matrix indicating which ASes peer with each other.

[3] We study SIX in depth because it offers complete data in terms of peering and traffic since its inception to the present day. Studying incomplete data from few other IXPs leads to very similar qualitative findings, omitted here for brevity.

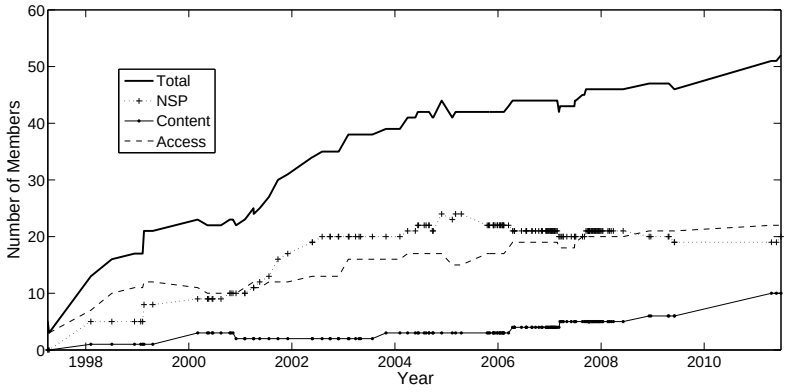


Figure 4.1: Number of members at SIX since 1997.

Traffic Grapher (mrtg) [142] data, described in detail below. While the snapshots of SIX webpage `http://www.six.sk/`, does provide some data on the traffic-stats over the previous 12 months, it does not store historic data for public use. In order to study the evolution of SIX, we take advantage of the *Internet archive* project [6], which stores 155 snapshots of the SIX website since 1997. In what follows, we will give more details on the data format and the data collection. Additionally, we continually monitor SIX since 03/2011 until the beginning of 2012, by taking a snapshot of the whole SIX website every day. In our analysis, we use 3 of these snapshots from the months of March, June and September of 2011. With 155 snapshots obtained from the `waybackmachine.org`, we have in total 158 snapshots covering years from 1997 to 2011. The data used in this chapter can be conveniently found at [31].

Peering matrix is a matrix that indicates whether two members (ISPs) of the IXP peer between each other or not. There could be many reasons for peering (or not) between two ISPs, and typically two ISPs would peer if and only if such peering provides (financial) benefits to *both* ISPs [7, 8, 52, 92]. At each snapshot[4] of the SIX website, we have the peering matrix, indicating who peers with whom at that time. Each participating ISP is identified by its name and AS number. In several cases, either name or AS number of an ISP change at some point, and for the purpose of our study we consider that ISP to be the same as the one before the change. However, some ISPs stopped peering at SIX. We found 32 ISPs since 1997 that used to peer at SIX that do not peer any more. With 52 ISPs that peer at SIX today, it brings the total number of ISPs that have participated in SIX to 84.

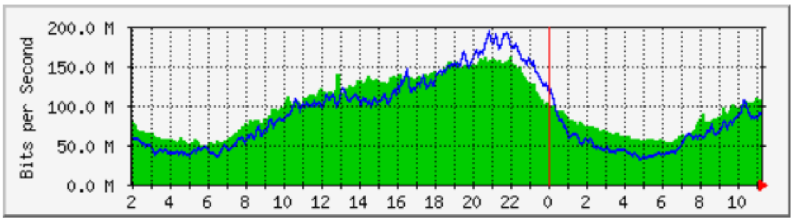
In addition to mapping SIX members to the ISPs, we classified each ISP based on the type of business they are involved with: access, content and network service providers (NSP). We used *peeringDB* [147] to map ISPs to their type, and in few cases in which no type info was found in *peeringDB*, we manually inspected the ISP type by checking its business offering. In Figure 4.1 we depict the evolution of the number and type of SIX participants.

Per-member traffic demands and port capacity are extracted from the mrtg data [142]

[4] 155 historic snapshots from the `waybackmachine.org` and daily snapshots from 03/2011.

System: SIX-UPC-gw
Interface: GigabitEthernet
IP: (192.108.148.205)
Max Speed: 1 Gbit/s (GigabitEthernet)
The statistics were last updated **Wednesday, 12 April 2006 at 5:20**,
at which time **'SIX-GigaSwitch'** had been up for **332 days, 9:38:14**.

`Daily' Graph (5 Minute Average)



	Max	Average	Current
In	170.0 Mb/s (17.0%)	100.9 Mb/s (10.1%)	54.0 Mb/s (5.4%)
Out	191.6 Mb/s (19.2%)	98.8 Mb/s (9.9%)	35.6 Mb/s (3.6%)

Figure 4.2: Snapshot of mrtg data.

available for *each* member at the SIX webpage. Figure 4.2 shows a (partial) snapshot of a webpage generated using the mrtg tool. It contains the data from UPC (large access provider), time-stamped on 12/4/2006, with the port capacity[5] of $1Gbps$ (value *Max Speed* in the graph), listing the average, max and current demand for both the inbound and outbound traffic, on daily (shown), weekly, monthly and yearly (not shown in Figure 4.2) basis. The mrtg tool also produces visual images depicting the daily/weekly/monthly/yearly traffic trends, and we designed a script that transforms these visual images into numeric data. However, for the purpose of this thesis, we exclusively work with the numeric data provided directly by mrtg: max and average.

Compared to peering matrix data that is stored each of the 155 times, the `waybackmachine.org` crawler hit `http://www.six.sk/`, the mrtg data is not archived every time, we believe because of the limitations of the `waybackmachine.org` in terms of bandwidth/storage and perhaps other implementation reasons. However, most ISPs do have at least one mrtg sample data point per year, which is enough for (relatively accurate) capturing dynamics on the yearly timescale. In the intervals between the available information, we use simple linear interpolation, to estimate the traffic at any point in time.

4.2. Evolution of SIX

In this section, we examine various dynamic and invariant properties of the Slovak IXP. We want to stress that several properties observed here (the exponential traffic growth, the rise of the

[5]Some members lease more than one port, in which case the sum of all port capacities is shown as the port capacity.

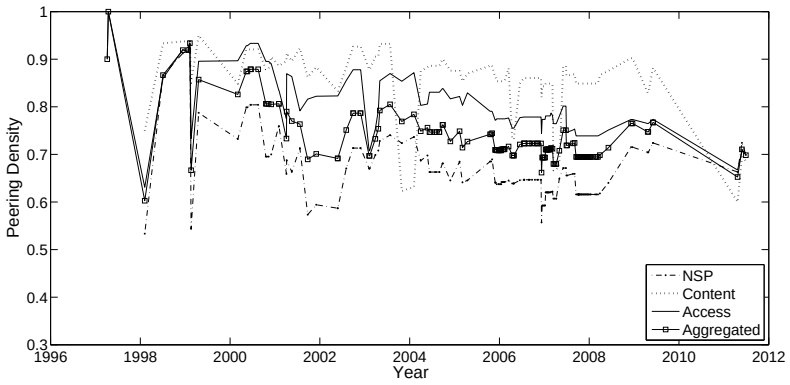


Figure 4.3: The evolution of peering density of SIX (aggregate) and per ISP type.

content providers and log-normality of the traffic matrix entry distribution) have been observed in other vantage points before, using confidential data [49, 114, 141]. However, we establish them in the IXP context using public-domain data, and report them here for the sake of completeness.

4.2.1. Peering matrix

Due to its importance, the AS-level topology of the Internet has been one of the most comprehensively studied topic in the networks research community. In spite of a tremendous amount of work on this topic, the existing measurement tools have very low accuracy in measuring (inferring) the AS-level topology in the lower tiers of the AS ecosystem [8, 92, 100]. For example, the most complete IXP-peering dataset [8] infers only 30% of the SIX peering links. The IXP data we use here, offer unique opportunity to accurately examine not only the state AS-level topology among a subset of low-tier ISPs, but also to evaluate its dynamics.

Peering density dynamics. A major difference between the AS-level topology at the IXP level and the global AS-level graph is in the density of interconnections. Namely an 'average' ISP typically peers with a large fraction of other ISPs from the IXP. To quantify how likely the two IXP members are to peer, we use the *peering density*, a common metric defined as the ratio between the number of peering links and the number of all possible pairs of ISPs participating at the IXP [8]. In Figure 4.3 we plot the density of SIX across the 14 years of operation and observe that this quantity has been fairly stable over the time and is in the range around 70% which is 'normal' for European standards. The peering densities in the US-based IXPs are reportedly lower than of the European IXPs, while the IXPs in Australia, New Zealand and far-east are slightly denser in terms of peering [8]. In the same figure we also plot, per-type peering density of content, access and network service providers, and observe no significant dependance between the type of an ISP and the peering density.

Peering link creation. Throughout the history of SIX until 2011, there have been 1711 pairs of ISPs peering between each other. For those pairs of ISPs that eventually start peering, how long does it take from the moment the newer of them appears at SIX until they engage into the peering

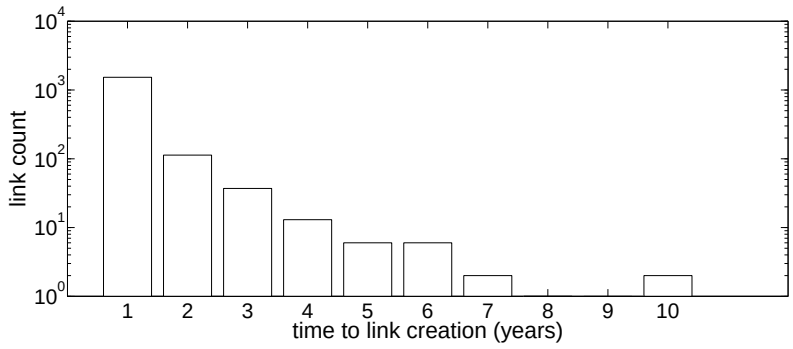


Figure 4.4: The histogram of link creation times for 1711 peering pairs in the history of SIX.

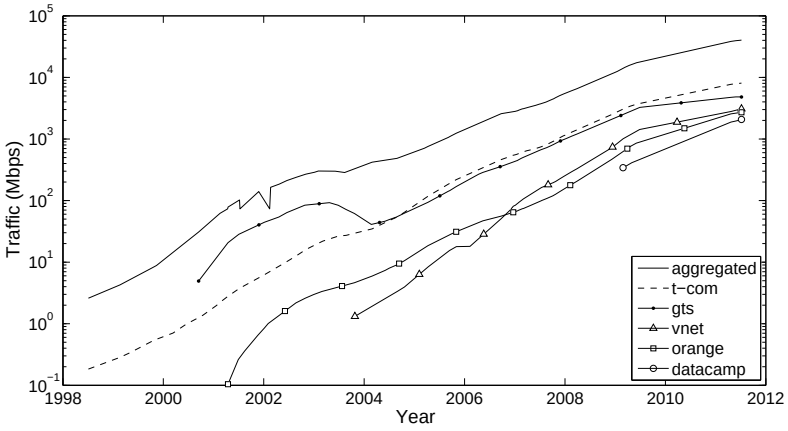


Figure 4.5: SIX aggregated traffic and the traffic of top-5 members in 2012.

(we call this value the *link creation time* - LCT)? In Figure 4.4 we depict the distribution of LCT for the 1711 peering pairs. We can observe that when the peering happens, it is usually created within a year from the appearance of the newer ISP, yet there is a dozen of pairs, that required more than 5 years of simultaneous existence before peering was established.

Peering link removal. Dhamdhere and Dovrolis [56], analyzed AS-level customer-provider (CP) links over a 10-year period, and showed that these links appear and disappear frequently, citing as a major reason the cost optimization process customers perform during the search for the most cost-effective provider. The peering links, on the other hand, have a different business objective and, thus, one may guess that peering links are less likely to be broken once they are created. Our data shows that this is indeed the case. Out of 1711 links, existing in SIX, only 20 link pairs de-peered. The other reason for the peering link removal is the disappearance of one of the ISPs from the IXP. As we mentioned earlier, 32 ISPs[6] that participated in SIX do not participate anymore and indeed peering links they were engaged with are not longer present.

ISP Name (ASN)	Type	AGR
aggregated	N/A	103.2%
t-com (AS6855)	Access	137.2%
gts (AS5578)	NSP	77.88%
vnet (AS29405)	Content	189.8%
orange (AS15962)	Access	143.4%
datacamp (AS39392)	Content	99.26%

Table 4.1: Type and Annualized Growth Rate (AGR) for aggregated traffic and Top-5 ISPs.

4.2.2. Traffic dynamics

Traffic growth. The growth of the Internet traffic has been shown to follow an exponential pattern at many vantage points: the residential broadband networks [49], the transpacific traffic [18], the global inter-domain traffic [114], etc. We observe a similar pattern in SIX as can be seen in Figure 4.5, which depicts the traffic growth for SIX and top-5 (in terms of traffic volume) ISPs of our data-set. In contrast to residential and global inter-domain traffic that grow with *annual growth rate* (AGR) of around 40-50% [49, 114, 174], the AGR of the SIX traffic both in aggregate and individual ISP terms is much higher: around 100% or even more for some ISPs. This hints at the growth of the relative fraction of the inter-domain traffic that is exchanged via peering at the IXP, and therefore a decay of the relative fraction of the inter-domain traffic that reaches end-customers via transit, a phenomenon consistent with the widely observed flattening-of-the-Internet trend [114].

Another interesting property of the growth is that it is not uniform among involved ISPs: the traffic of some ISPs grows quicker than for the others. Table 4.1 contains the AGR for top-5 active ISPs and the total SIX aggregate. Relatively wide range of AGR reveals significant differences in the growth of different ISPs.

Remark: The annualized growth rates are obtained using the linear least-square fitting of the traffic growth curves in log scale. This process is described in Section 3.3.3.

Traffic per ISP type. As the Internet ecosystem matures, we are likely to expect the emergence of the specialized ISPs that target specific customer groups. By looking in our data, we can see the emergence of one such class: content ISPs. Namely, until 2006, the content ISPs carried a very small fraction of SIX's traffic: under 10% in both directions. Since mid-2006 until 2011, the number of specialized content ISPs doubled from 5 to today's 10; see Figure 4.1. The relative outbound traffic of those members, however, grew for an order of magnitude, from under 5% to over 40%.

Traffic imbalance. From Figure 4.6 one can notice that inbound to outbound traffic ratios vary significantly for different ISP types. For example, a fully residential ISP traffic is likely to be heavily inbound, while the traffic of an ISP serving only content is likely to be heavily outbound. Additionally, balanced traffic is explicitly required in many peering contracts between tier-1 or tier-2 ISPs and the traffic imbalance has been cited as the main reason for de-peering in a number of recent de-peering incidents [20]. Here we look at the evolution over time of the traffic

[6] Most of these ISPs either do not exist anymore, or were merged/purchased with some of the existing SIX members.

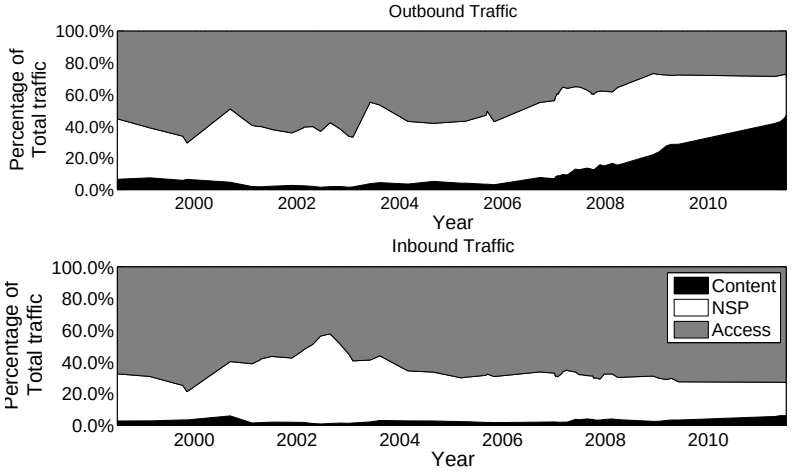


Figure 4.6: The percentage of total inbound/outbound traffic per ISP type.

imbalance, which we define for ISP X at time t as:

$$IB(X,t) = \max\left(\frac{T_{inbound}(X,t)}{T_{outbound}(X,t)}, \frac{T_{outbound}(X,t)}{T_{inbound}(X,t)}\right)$$

Where $T_{inbound}(X,t)$ and $T_{outbound}(X,t)$ is the inbound and outbound traffic of the ISP X at time t, respectively. In Figure 4.7 we depict the evolution of the median ISP IB, as well as the 10th and the 90th percentile. We also plot the (traffic) weighted average of the imbalance across all ISPs present in SIX at time t:

$$WA\_IB(t) = \frac{\sum_{X \in Active_t} T(X,t) \cdot IB(X,t)}{\sum_{X \in Active_t} T(X,t)}, \tag{4.1}$$

where $T(X,t) = T_{inbound}(X,t) + T_{outbound}(X,t)$ is the total traffic of ISP X at time t. From Figure 4.7 we can see a growing trend in the traffic imbalance, indicating the increasing focus of the ISPs in particular end-customer groups. The fact that the weighted average is close to the 90th-percentile indicates that the large ISPs are more pronounced in such traffic imbalance, compared to the small and medium ISPs.

4.2.3. Capacity and utilization dynamics

In order to exchange traffic at SIX and most other IXPs, each member needs to pay a monthly fee for each port it uses for traffic exchange. The price of a port increases with the port capacity, but are sub-additive, with price per $Mbps$ decreasing as a higher port capacity is purchased. The reasons for upgrading from lower port capacity to a higher one can be different, but roughly speaking most of the upgrades happen because the member's traffic reaches a port-utilization that is above some threshold.

Information on network utilization and upgrades by commercial ISPs are notoriously hard

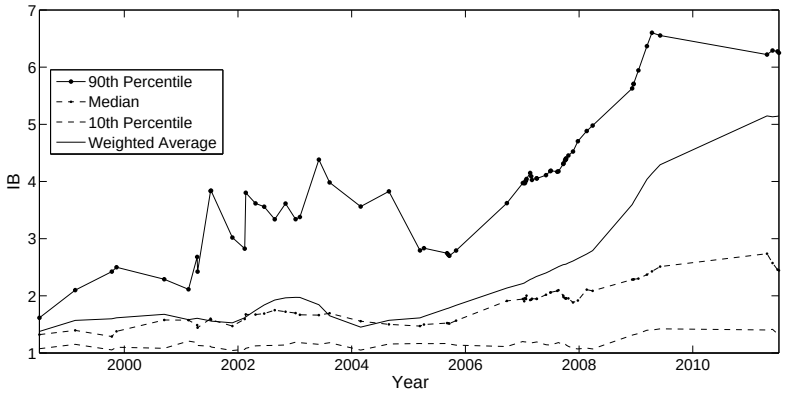


Figure 4.7: The evolution of ISP traffic imbalance (IB), median, 10th-, 90-th percentile and the weighted average.

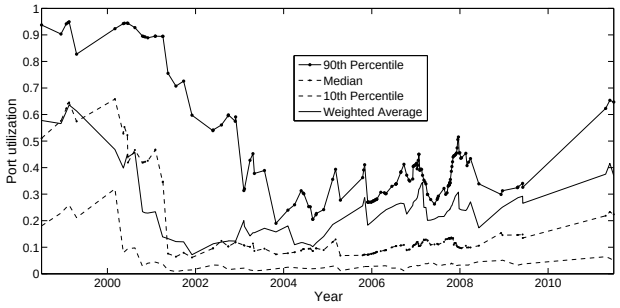


Figure 4.8: Per member port(s) utilization of SIX since 1998: median, 10th, 90th-percentile and weighted average.

to obtain. Literature often cites 50% rule-of-thumb for net upgrades [131, 176], but we are not aware on any empiric study across a set of diverse ISPs that evaluates such statements. The data we study here offers an unique opportunity to shed light on the upgrade practice and the port utilization in dozens of operational ISPs.

Port utilization. In Figure 4.8 we depict the median, the 10th and the 90th-percentile of peak utilization among all members of SIX at each time instance for which we have a SIX snapshot. We used the peak utilization as the maximum monthly utilization averaged in 2-hour slots, available in mrtg data. In the first couple of years, the SIX ports were very highly utilized with median members' peak utilization being greater than 50%. In the early 2000's, a significant increase in the port capacities occurred and brought the utilization of many members down. From mid-2000's until the end of our dataset we observe the increase in the utilization levels, yet still 90% of the members have the peak utilization of their port(s) under 65%. We also depict the evolution of the weighted average of the utilization, weighted with the traffic of the ISP, similarly to Eq. (4.1). We can see that, since 2004, most of the heavy ISPs are running their ports at a higher utilization than the median.

Port upgrades. In any ISP, increasing capacity of its networks is the most important mechanism for ensuring the high level of quality of service. Typically, the capacity upgrades in op-

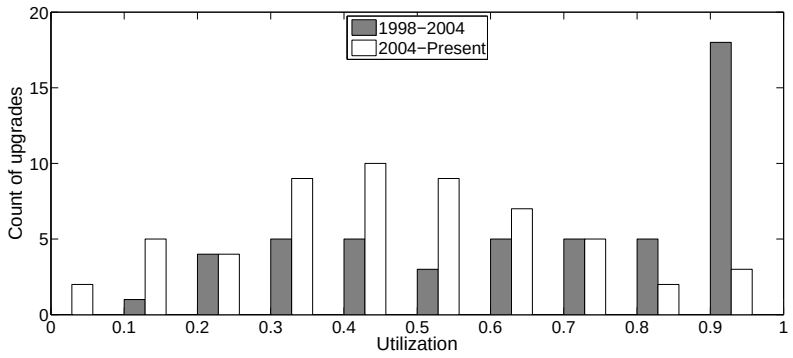


Figure 4.9: The estimated utilization at the time of port upgrade.

erational networks require a lot of planning and serious implementation efforts. In contrast, upgrading the port capacity at the IXP level is rather straightforward and can be done almost instantaneously. Hence, the data from SIX can offer insights on the operational practices of the involved ISPs. Figure 4.9 contains two histograms on the peak utilization at the moments of capacity-upgrades, one for the period from 1997-2004, and another for the period since 2004. In the earlier period, majority of upgrades happened because of the port overload. However since 2004, most ISPs upgrade their capacity early to avoid congestion of their SIX port. While most of the upgrades happen when the utilization reaches the range [30%-60%], there is still a non-negligible fraction of upgrades that happen outside of this range.

Comment. We analyzed the impact that port upgrades have on the traffic growth, and observed no statistically significant difference on the growth of the traffic before and after the upgrades.

4.2.4. Traffic matrix dynamics

Even though we do not have the exact *traffic matrix* (TM) between the pairs of ISPs peering at any instance of time, we can utilize the standard tools to estimate TM entries with reasonable accuracy. Here we choose to use the *tomogravity* method [195], that has been extensively deployed in operational ISPs and is simple to implement. The expected errors of any TM estimation tool can be relatively large for a single origin-destination (O-D) pair. However, we do not analyze specific O-D pairs, but rather focus on aggregate statistics of the TM entry distribution, to draw the relevant conclusions.

Per-peering traffic distribution is skewed. An invariant property of the SIX traffic matrix is the variability of its entries. Namely, the distribution of per-peering traffic appears to be log-normal with the exponentially growing mean (and variance), which is consistent with the previously observed property of intra-domain traffic matrix snapshots [141]. In Figure 4.10 we depict the histogram of per-peering traffic of peering pairs present in the years 1999, 2003, 2007 and 2011. Consequently, several heaviest pairs carry most of the traffic (e.g. top-1% and top-10% of the pairs carry 40% and 85% of the traffic, respectively) and majority of the peering pairs carry very little traffic.

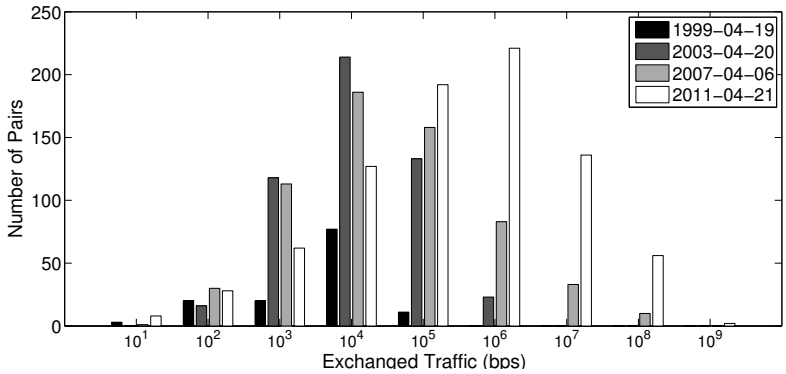


Figure 4.10: Histograms of traffic per peering pair in 4 points in 1999, 2003, 2007 and 2011.

Peering: cost-reduction or performance? The most widely cited reason for creating a peering relationship between two ISPs is cost reduction: if the traffic between ISP A and ISP B is exchanged directly via peering, there is no need to be delivered via transit provider(s), hence the transit cost for both ISPs is expected to reduce. While such reasoning is indeed valid when the traffic volume between ISPs A and B is large, it becomes questionable when the traffic is small, since engaging into a peering relationship has a non-zero monetary cost (for legal agreements, staff, maintenance, etc.) associated with it. In Figure 4.11 (top) we depict the (estimated) median traffic volume among all peering pairs at SIX present at any instance of time, as well as the corresponding wholesale price of IP transit per $Mbps$ per month[7] [137]. In spite of two-order-of-magnitude change in these two quantities, the value of the median peering (calculated as the product of the median peering volume and the wholesale IP transit price) remains stable and under 10 USD per month; see Figure 4.11 (bottom). Such low median peering value suggests that a majority of the peering links carry very low monetary savings and would not be economically viable outside Internet eXchange Points.

4.3. Discussion

Diurnal (daily) traffic dynamics. Most of the analysis we performed in this chapter treats the (traffic) dynamics on the yearly time-scale. The diurnal dynamic properties (such as the peak-to-valley ratio, peak-hour, etc.) is critical for the success of several recent proposals [45, 116, 130], and can be derived from the visual mrtg images. For example we observe that the peak-hour has shifted from early afternoon (1-2pm) in the late 1990's and early 2000's to late evening (9pm) nowadays. This probably is the result of residential traffic dominating that of commercial companies, as described in Section 3.3. Furthermore, the ratio between the peak hour and off-peak hour from 10:1 in the late 1990's, to 3:1 in the mid 2000's (coinciding with the p2p revolution), to 10:1 nowadays. These changes could be related to human-behavior effects, similar to the ones

[7]The wholesale price used here is for the ports in EU/USA hubs. In other continents, the price per $Mbps$ can be 5 to 10 times greater.

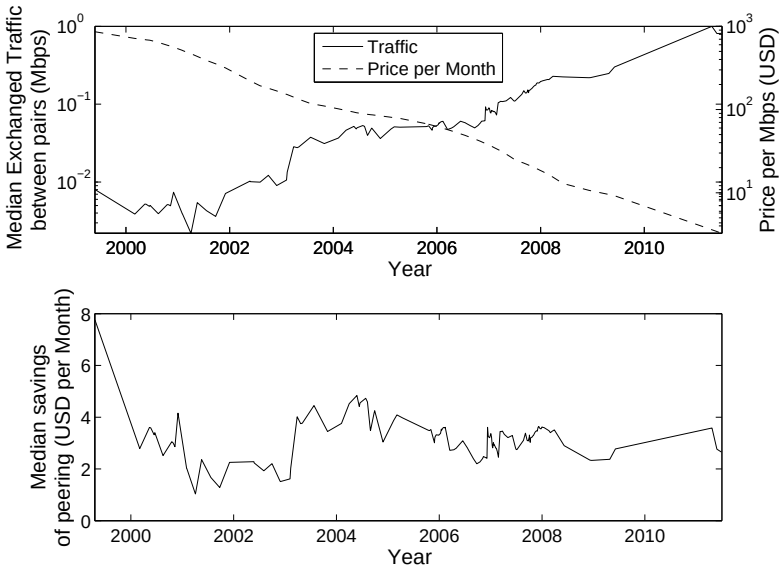


Figure 4.11: Median per-peering traffic and wholesale IP transit price (top) and peering value (bottom)

we described in Chapter 3. We however omit the detailed discussion of these properties due to space limitation.

Matching large shifts in traffic with external events. The historical traffic evolution of companies can be used to observe important external events (not necessarily human-driven), such as large-scale DDOS attacks, capacity upgrades or port blocking/throttling in an ISP. For example in 2003 a significant drop in SANET-AS2607[8] traffic happened (60% reduction in the traffic without any change in the peering) for which we speculate that is caused by sudden p2p port blocking by the SANET.

Effects of private peering. Public peering, via IXPs, offers many benefits to the ISPs interested in peering, by allowing inexpensive and convenient peering with many peers. However, some ISPs may choose to peer privately, in which case such those relationships may be invisible to the IXP. In context of the Slovak ISPs, it appears that only a small fraction of local peerings are private. Namely, the existing AS-level topology datasets [177] reports circa of 34 private peering links among Slovak ISPs, which is under 4% of the number peering relationships that we found at SIX.

4.4. Related work

Obtaining high-fidelity data for studying the Internet properties is notoriously hard for various reasons ranging from confidentiality concerns to the lack of measurement infrastructure. In addition, analyzing the evolution of the Internet properties requires not only capturing a single snap-

[8] Slovak academic network, acting as the access ISP to all Slovak universities.

shot of the data but also the system for continuous data collection and archiving. Consequently, the data that allows such longitudinal studies of the Internet properties is extremely scarce.

AS-level graph evolution. Dhamdhere and Dovrolis [56] use available historic data collected by RouteViews and RIPE to study the evolution of the customer-provider links in the AS-level graph, and the stability of such links. As they point out, these datasets have very poor accuracy in inferring the AS-peering links, and our work complements [56] by offering new insights on the evolution of the AS-level connectivity between the lower tier ISPs.

Longitudinal Internet traffic evolution. The data that involves IP traffic measurements is often kept confidential, for obvious business concerns. Several studies have appeared in the literature reporting the properties of traffic evolution in certain vantage points [18, 49, 83], with the caveat that such longitudinal data is typically collected at a single ISP and not available for public use. The analyzed traffic (and link capacity) data represent the 14-year evolution of the inter-domain traffic of *dozens* of ISPs, and is fully available for *public* use, which allows the analysis of the properties beyond those examined here.

Peering links over the Internet. The economics factors are believed to be the dominant force in the link creation (both the fee-based or settlement-free links) process at the AS-level [8, 57]. In this study we show that in fact a large fraction of peering links is (and has been) virtually valueless. In other words, majority of AS-level links[9] would not be economically viable without IXPs and the link bundling in which high-value peerings are bundled with low-value peering links over the same physical port.

4.5. Summary

With the trend of flattening-the-Internet [57,114], the relative importance of IXPs, and peering in general, is likely to increase. In this chapter, we study one regional IXP, Slovak IX, for which we acquired detailed peering and traffic data for 14 years of its existence. Such data allows us to characterize not only the state of the IXP at one particular moment, but also its dynamic and invariant properties.

Examining the evolution of SIX revealed a number of interesting facts in regards to the peering and traffic trends at the IXP. We summarize them next.

Stagnation of SIX's membership size. After almost a decade of steady growth, the number of members of SIX has been stable in recent years. As described in Chapter 3, this effect occurs on other regional IXPs an might be explained by the difficulties for IXPs to attract networks from other regions, after the local ones join, due to the relatively high cost of transport cost and the average low potential exchanged traffic. One potential way of reducing the cost of reaching multiple IXPs, thus favoring both IXPs and ASes, is the use of *remote peering*. Using remote peering, a company can share the connection infrastructure (e.g. interconnection, router, etc.)

[9] The number of peering links in the Internet dominates the number of customer-provider links in the Internet [8].

to connect to various IXPs. We explain remote peering, and describe our assessment of use at several IXPs, in Chapter 5.

IXP peerings are stable. We discovered that once it has been created, a peering between two ISPs is very unlikely to disappear unless one of the members leaves the IXP. This happens in contrast to customer-provider links that have been shown to change relatively frequently [56].

Evolution of port capacity and link utilization. We characterized the utilization of links among the members of SIX. While most of the upgrades happen when the utilization are within the rule of thumb range of [30%-60%], there is still a non-negligible fraction of upgrades that happen outside of this range.

Low monetary value for most peering sessions. By estimating SIX's traffic matrix, we observe that the distribution of traffic per peering-link has been skewed since the beginning of SIX. This property indicates that the majority of peering links carry very low amounts of traffic (and consequently have very low monetary value), which challenges the presumption that promotes financial gain as the main cause for peering. Our observations on SIX match those of [2] , which also observes a high skewed distribution of traffic on peering links, and traffic ratios that would not normally be allowed in private peering connections.

Chapter 5

Remote Peering

In Chapter 3, we showed how member overlapping among IXPs can quickly decrease the benefits of transit off-loading, even after peering only at one IXPs. Due to the non-negligible operational and infrastructural costs, this effect limits the number of IXPs that ASes can cost-effectively reach, which poses a problem for both operators and IXPs: On the one hand, by restricting to a few IXPs, operators lose the opportunity of increasing path diversity, and connecting directly to smaller networks. On the other hand, attracting only a smaller set of companies (usually the regional companies) restricts the growth potential of IXPs. This is a potentially reason for the halt, in terms of member size, experienced by SIX and other regional IXPs, as seen in Chapters 3 and 4.

Remote peering is a service in which an AS reaches an IXP via a Layer-2 (L2) intermediary (or remote-peering provider) and becomes a member of it without extending its own infrastructure or operational force. Remote-peering providers include companies specialized in this type of business and traditional transit providers that leverage their traffic-delivery expertise to act as remote-peering intermediaries [37]. A remote-peering provider can offer its customers connectivity with multiple IXPs using the same infrastructure.

Remote peering provides a solution for operators to profitable reach multiple IXPs, and for IXPs to attract a larger set of networks. In this chapter, we describe remote peering, provide the results of measuring its adoption across multiple IXPs, and discuss their Internet structure implications. For our adoption study, we develop a ping-based method that conservatively estimates the spread of remote peering. We apply the method in 22 IXPs worldwide and detect remote peers in more than 90% of the studied IXPs, with remote peering used by up to 20% of the members at an IXP.

Remote peering separates the trends of increasing peering and Internet flattening. At layer-3, remote peering is not distinguishable from direct peering. When a network buys a remote-peering service to establish new paths around a transit provider, the Internet becomes flatter at layer-3 because the new paths bypass the layer-3 transit provider. However, the layer-3 perspective is misleading because the new paths replace the transit provider with the layer-2 remote-peering

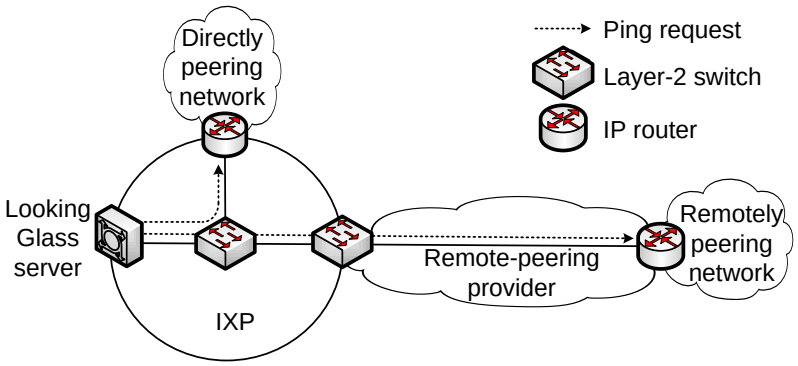


Figure 5.1: Directly and remotely peering networks, and probing of their IP interfaces from an LG server.

provider. When one broadens the layer-3 perspective to include the organizations that provide layer-2 services, remote peering does not necessarily reduce the number of intermediary organizations on Internet paths in spite of the enabled additional peering relationships. Hence, remote peering means more peering without Internet flattening. The wide spread of remote peering also has broader implications for Internet research. When a provider offers transit and remote peering, buying both might not yield reliable multihoming. The presence of intermediaries that are invisible to layer-3 protocols adds to existing security concerns, e.g., the invisible layer-2 intermediaries can monitor traffic or deliver it through undesired geographies. For Internet accountability, it is a challenge to associate an action with the responsible invisible entity. As a new economic option, remote peering opens a whole new ballgame for connectivity, routing, and traffic distribution, e.g., via newly enabled IXPs [140]. We further discuss the structural effects of remote peering in Section 5.3.

The rest of the chapter is organized as follows. Section 5.1 provides background on the technical details of remote peering. Section 5.2 empirically studies the spread of remote peering. Section 5.3 discusses broader implications of remote peering on the Internet infrastructure. Section 5.4 presents related work. Finally, section 5.5 presents a summary of the chapter.

5.1. Introduction to remote peering

5.1.1. Technical aspects

Remote peering constitutes an emerging type of interconnection where an AS reaches and peers at a distant IXP via a layer-2 provider [16]. The remote-peering provider delivers traffic between the switching infrastructure of the IXP and the remote interface of the customer. On the customer's behalf, the remote-peering provider also maintains networking equipment at the IXP to enable the remote network to peer with other IXP members, saving the customer from acquiring any new devices and hiring any operational support at the remote IXP (e.g. remote hands). Figure 5.1 depicts a typical setting for the remote-peering relationship.

Technologically, from the provider point of view, remote peering can be implemented with standard methods, such as those used in L2 MPLS (MultiProtocol Label Switching) VPNs (Virtual Private Networks). The remote peering provider can install devices with a large number of ports at the IXP or connect to the IXP infrastructure using VLANs. For ASes, a remote peering service is seen as a L2 service, differentiating the connectivity to different IXPs either logically via various VLANs or physical links. Operators connect the L2 service to a BGP router and use an IP of the IXP to establish peerings to other IXP members.

The L2 segment of the remote peering provider becomes an extension of the IXP backplane. This has deep operational repercussions, since the loss of connectivity of a member might be due to failures in the customer, remote peering provider, or IXP domain. Operators of the three parties should be aware of this aspect, in order to provide efficient troubleshooting procedures.

5.1.2. Economical aspects

A remote peering providing connectivity to the same IXP for multiple customers can save transport costs by avoiding traffic tromboning. That is, it can switch directly the traffic between two customers (in case they peer) without moving the traffic to the IXPs. This can be done, for instance, by levering multipoint-to-multipoint technologies such as Virtual Private LAN Service (VPLS)[1].

Remote peering has both traffic-dependent and traffic-independent costs. In comparison to direct peering, the traffic-independent cost is lower, and the traffic-dependent cost is higher: the remote-peering provider has multiple customers and reduces its per-unit costs due to traffic aggregation and acquisition of IXP resources in bulk. Compared to transit, remote peering has lower traffic-dependent costs. Thus, from the cost perspective, remote peering represents a trade-off between direct peering and transit.

IXPs and remote peering are highly symbiotic. IXPs benefit from remote peering because the latter brings extra traffic to IXPs, enriches geographical diversity of IXP memberships, and strengthens the position of IXPs in the Internet economic structure. To promote remote peering, AMS-IX (Amsterdam Internet Exchange), DE-CIX (German Commercial Internet Exchange), LINX (London Internet Exchange), and many other IXPs establish partnership programs that incentivize distant networks to peer remotely at the IXP. For example, some IXPs reduce membership fees for remotely peering networks. AMS-IX started its partnership program around year 2003. According to our personal communications with AMS-IX staff, about one fifth of the AMS-IX members were remote peers at the time of our study.

Implications of remote peering for transit providers are mixed. On the one hand, remote peering gives transit customers alternative means for reaching distant networks. On the other hand, remote peering is a new business niche where transit providers can leverage their traffic-delivery expertise.

[1] Note that this traffic would not appear at the traffic statistics of the IXP.

IXP acronym	IXP name	Location		Peak traffic (Tbps)	Number of members	Number of analyzed interfaces
		City	Country			
AMS-IX	Amsterdam Internet Exchange	Amsterdam	Netherlands	2.72	638	665
DE-CIX	German Commercial Internet Exchange	Frankfurt	Germany	3.21	463	535
LINX	London Internet Exchange	London	UK	2.60	497	521
HKIX	Hong Kong Internet Exchange	Hong Kong	China	0.48	213	278
NYIIX	New York International Internet Exchange	New York	USA	0.46	132	239
MSK-IX	Moscow Internet eXchange	Moscow	Russia	1.32	367	218
PLIX	Polish Internet Exchange	Warsaw	Poland	0.63	235	207
France-IX	France-IX	Paris	France	0.23	230	201
PTT	PTTMetro São Paolo	São Paolo	Brazil	0.30	482	180
SIX	Seattle Internet Exchange	Seattle	USA	0.53	177	175
LoNAP	London Network Access Point	London	UK	0.10	142	166
JPIX	Japan Internet Exchange	Tokyo	Japan	0.43	131	163
TorIX	Toronto Internet Exchange	Toronto	Canada	0.28	177	161
VIX	Vienna Internet Exchange	Vienna	Austria	0.19	121	134
MIX	Milan Internet Exchange	Milan	Italy	0.16	133	131
TOP-IX	Torino Piemonte Internet Exchange	Turin	Italy	0.05	80	91
Netnod	Netnod Internet Exchange	Stockholm	Sweden	1.34	89	71
KINX	Korea Internet Neutral Exchange	Seoul	South Korea	0.15	46	71
CABASE	Argentine Chamber of Internet	Buenos Aires	Argentina	0.02	101	68
INEX	Internet Neutral Exchange	Dublin	Ireland	0.13	63	66
DIX-IE	Distributed Internet Exchange in Edo	Tokyo	Japan	N/A	36	56
TIE	Telx Internet Exchange	New York	USA	0.02	149	54

Table 5.1: Properties of the 22 IXPs in our measurement study on the spread of remote peering (values for 2014)

According to anecdotal evidence, remote peering successfully gains ground and satisfies diverse needs in the Internet ecosystem. We focus on usages where remote peering at IXPs is purchased by distant networks or other IXPs. For example, AMS-IX Hong Kong and AMS-IX interconnect their infrastructures via remote peering to create additional peering opportunities for their members [168]. We do not consider an alternative usage where remote peering at an IXP is bought by a local network to benefit from cost reductions that remote peering provides even over short distances [44].

5.2. Spread of remote peering

In this section, we report measurements that conservatively estimate the spread of remote peering in the Internet.

5.2.1. Measurement methodology

Because remote peering is provided on layer 2, conventional layer-3 methods for Internet topology inference are unsuitable for the detection of remote peering. For instance, traceroute and BGP data do not reveal IP addresses or ASNs (AS Numbers) of remote-peering providers,

since the underlay L2 service is transparent to them.

The idea of our methodology for detecting a remotely peering network at an IXP is to measure propagation delay between the network and IXP. Specifically, we use the ping utility to estimate the minimum Round-Trip Time (RTT) between the IXP location and the IP interface of the network in the IXP subnet. If the minimum RTT estimate exceeds a threshold, we classify the network as remotely peering at the IXP.

While our ping-based method is intuitive, the main challenges lie in its careful implementation and include: identification of probed interfaces, selection of vantage points, adherence to straight routes, sensitivity to traffic conditions, identification of networks, choice of IXPs, threshold for remoteness, IXPs with multiple locations, impact of blackholing, and measurement overhead. We discuss these challenges below.

Identification of probed interfaces: The targets of our ping probes are the IP interfaces of the IXP members in the IXP subnet. IXP members do not typically announce the IP addresses of these interfaces via BGP. To determine the IP addresses of the targeted interfaces, we look up the addresses on the websites of PeeringDB [147], PCH (Packet Clearing House) [146], and the IXP itself.

Selection of vantage points: The ping requests need be launched into the IXP subnet from within the IXP location so that the requests take the direct route from the IXP location to the probed interface. We send the ping requests from Looking Glass (LG) servers that PCH and RIPE NCC (Réseaux IP Européens Network Coordination Centre) [135] maintain at IXP locations. Figure 5.1 depicts our probing of IP network interfaces from an LG server at an IXP.

Adherence to straight routes: With our choice of the vantage points, the ping requests and ping replies are expected to stay within the IXP subnet. It is important to keep the probe routes straight because otherwise the RTT measurements might be high even for a directly peering network. Potential dangers include an unexpected situation where the device of a probed IP interface replies from one of its other IP interfaces and thereby sends the ping reply through an indirect route with multiple IP hops. A more realistic risk is that some of our targeted IP addresses are actually not in the IXP subnet because the respective website information is incorrect. To protect our method from such dangers, we examine the TTL (Time To Live) field in the received ping replies. When ping replies stay within the layer-2 subnet, their TTL values stay at the maximum set by the replying interface [187]. When the path of a ping reply includes an extra IP hop, the TTL value in the reply decreases. Therefore, we discard the ping replies with different TTL values than an expected maximum. We refer to this discard rule as a *TTL-match filter*. For the expected maximum TTL, our experiments accept two typical values of 64 and 255 hops. Although ping software might set the maximum TTL to other values (e.g., 32 or 128 hops), these alternative settings are relatively infrequent, and ignoring them does not significantly increase the number of discarded ping replies in our experiments. Also, different ping replies from the same interface might arrive with different TTL values, e.g., because the replying interface changes its maximum TTL. Whereas we are interested in a conservative estimate for the extent of remote peering, we

discard all replies from an IP interface if their TTL value changes during the measurement period. We call this rule a *TTL-switch filter*.

Sensitivity to traffic conditions: Even if a probe stays within the IXP subnet, RTT might be high due to congestion. To deal with transient congestion, we repeat the measurements at different times of the day and different days of the week for each probed IP interface, and record the minimum RTT observed for the interface during the measurement period. This minimum RTT serves as a basis for deciding whether the interface is remote. Again to be on the conservative side, we exclude an IP interface from further consideration if we do not get at least 8 TTL-accepted ping replies from this interface for each probing LG server. We call this rule a *sample-size filter*. The limit of 8 replies and other parameter values in our study are empirically chosen to obtain reliable results while keeping the measurement overhead low. If less than 4 of the collected ping replies have RTT values within the maximum of 5 ms and 10% of the minimum RTT, i.e., below $\text{RTT}_{min} + \max\{5 \text{ ms}, 0.1 \cdot \text{RTT}_{min}\}$, we apply an *RTT-consistent filter* to disregard the interface. For an IXP that has both PCH and RIPE NCC servers, we probe each IP interface from both LG servers and exclude the interface from further consideration if the larger of the two respective minimum RTTs is not within the maximum of 5 ms and 10% of the smaller one. We refer to this rule as an *LG-consistent filter*.

Identification of networks: To identify the network that owns a probed IP interface, we use the network's ASN. We map the IP addresses to ASNs through a combination of looking up PeeringDB, using the IXPs' websites and LG servers, and issuing reverse DNS (Domain Name System) queries. If the ASN of an IP interface changes during the measurement period, we exclude the IP interface from further consideration. This exclusion rule is called an *ASN-change filter*.

Choice of IXPs: In choosing IXPs, we strive for a global scope surpassing the regional focuses of prior IXP studies. Our choice is constrained to those IXPs that have at least one LG server. Under the above constraints, we select and experiment at 22 IXPs in the following 4 continents: Asia, Europe, North America, and South America. After manually crawling the websites of the IXPs in January 2014, we collect data on their location, peak traffic, and number of members. Table 5.1 sums up these data. While information at IXP websites is often incomplete, out of date, or inconsistent in presenting a property (e.g., peak traffic), our measurement method does not rely on these data. We report this information just to give the reader a rough idea about the geography and size of the studied IXPs. For each studied IXP, table 5.1 also includes the *number of analyzed interfaces*, i.e., interfaces that stay in our analyzed dataset after applying all 6 aforementioned filters. Across all the 22 IXPs, we apply the filters in the following order: sample-size, TTL-switch, TTL-match, RTT-consistent, LG-consistent, and ASN-change. After the filters discard 20, 82, 20, 100, 28, and 5 interfaces respectively, we have a total of 4,451 analyzed interfaces. The high count of TTL-switch discards is likely due to operating system changes during our measurements.

Threshold for remoteness: We classify a network as remotely peering at an IXP if the mini-

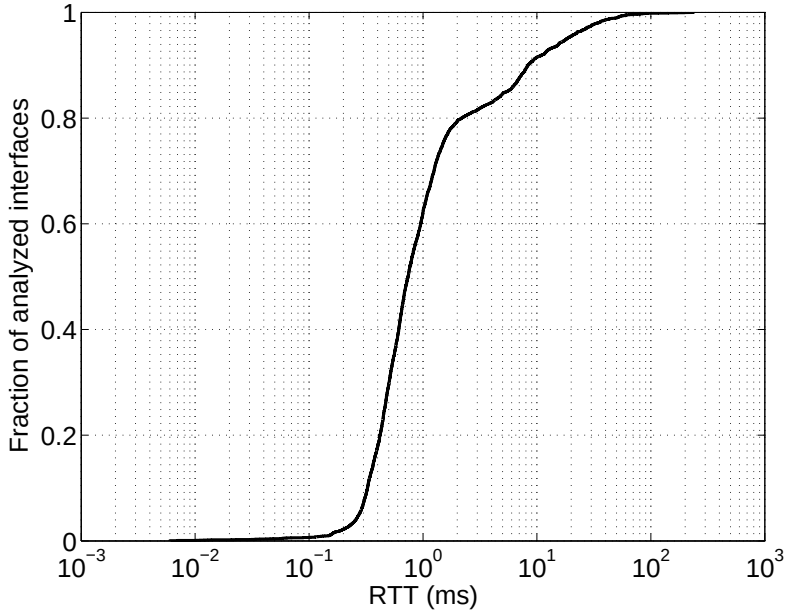


Figure 5.2: Cumulative distribution of the minimum RTTs for all the analyzed interfaces

mum RTT observed for its IP interface at the IXP exceeds a threshold. Despite the redundancy of our RTT measurements, the minimum RTT might still include non-propagation delays, e.g., due to persistent congestion of the IXP subnet or probe processing in the network devices. To minimize the possibility that such extra delays trigger an erroneous classification of a directly peering network as remote, the threshold should be sufficiently high. Figure 5.2 plots the cumulative distribution of the minimum RTTs for all the 4,451 analyzed interfaces. A majority of the analyzed interfaces have minimum RTTs distributed almost uniformly between 0.3 and 2 ms. This is a pattern expected for directly peering networks. The likelihood of a network being a direct peer declines as the minimum RTT increases. Our manual checks do not detect any directly peering network with the minimum RTT exceeding 10 ms. Thus, we set the remoteness threshold in our study to 10 ms. While this relatively high threshold value comes with a failure to recognize some remotely peering networks as remote peers, the false negatives do not constitute a significant concern because we mostly strive to avoid false positives in estimating the spread of remote peering conservatively.

IXPs with multiple locations: If an IXP operates interconnected switches in multiple locations, probes from an LG server at one location to an IP interface at another location might have a large RTT. The chosen remoteness threshold of 10 ms is sufficiently high to avoid false positives in cases where all locations of the IXP are in the same metropolitan area. False positives are possible if the geographic footprint of the IXP is significantly larger, e.g., spans multiple countries. We do not observe such situations in our experiments. In a more common scenario, two partner IXPs from different regions, e.g., AMS-IX Hong Kong and AMS-IX, interconnect by buying layer-2 connectivity from a third party. Our methodology correctly classifies such scenarios as remote

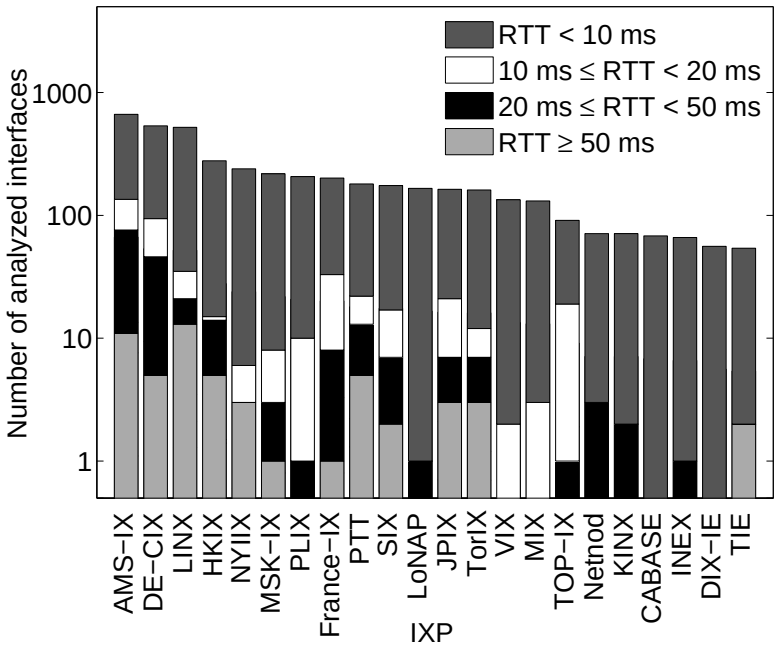


Figure 5.3: Classification of the analyzed interfaces with respect to 4 ranges of minimum RTTs

peering.

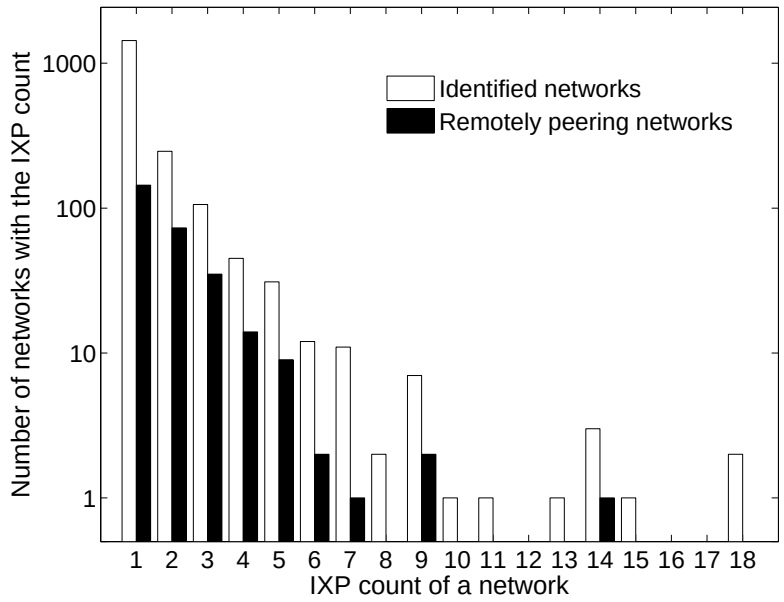
Impact of blackholing: If a probed interface intentionally blackholes or accidentally fails to respond to ping requests, the IP interface might be excluded from our analyzed data due to a low number of ping replies for the interface, as discussed above. In a hypothetical (not observed in our experiments) scenario where the probed interface forwards the probe to another machine that sends a ping reply on the interface's behalf, the ping reply is discarded by our TTL-match filter and does not affect accuracy of our RTT measurements.

Measurement overhead: While our method relies on probing from public LG servers, it is important to keep the measurement overhead low. The probes are launched through HTML (HyperText Markup Language) queries to the servers. The LG servers belonging to RIPE NCC and PCH react to an HTML query by issuing respectively 3 and 5 ping requests. For any LG server, we submit at most one HTML query per minute and spread the measurements over 4 months. The maximum number of ping replies received from any probed IP interface is 21 and 54 for respectively RIPE NCC and PCH servers.

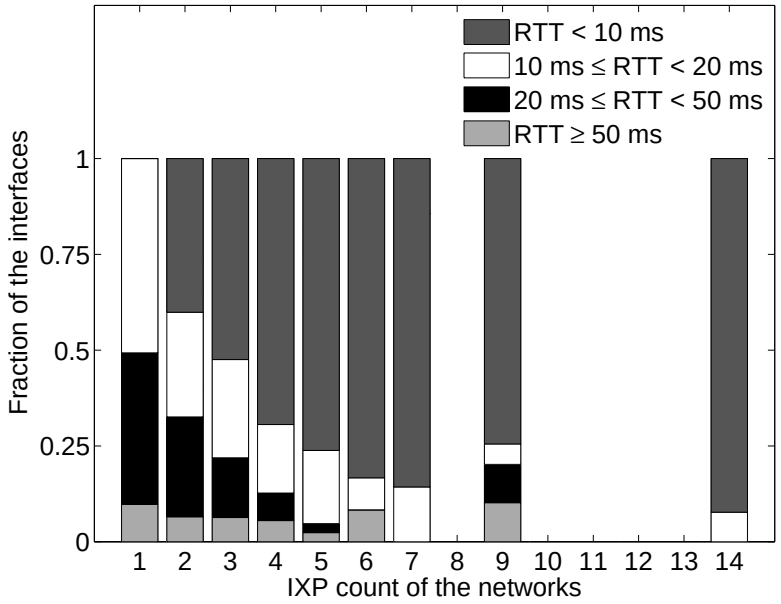
We conducted the measurements during the 4 months from October 2013 to January 2014. The measurement data are available at [27].

5.2.2. Experimental results

Figure 5.3 classifies all the 4,451 analyzed interfaces across the 22 IXPs according to the minimum RTT measured for each interface. Our conservative estimate finds remote peering in



(a) Distributions of the IXP counts



(b) Interfaces of all the 285 remotely peering networks

Figure 5.4: IXP-count distributions and interface classifications for identified networks

91% of the studied IXPs. While the numbers of remote interfaces are large in the 3 biggest IXPs (AMS-IX, DE-CIX, and LINX), these numbers are also large at smaller IXPs such as France-IX in France, PTT in Brazil, JPIX in Japan, and TOP-IX in Italy. Despite using the high remoteness threshold of 10 ms, the classification does not reveal remote interfaces in only two IXPs (DIX-IE and CABASE). Hence, our method independently confirms wide presence of remote peering in

the Internet economic structure.

The classification in figure 5.3 looks at the remote interfaces in greater detail by considering the following 3 ranges for the minimum RTT: [10 ms; 20 ms), [20 ms; 50 ms), and [50 ms; ∞) which roughly correspond to intercity, intercountry, and intercontinental distances respectively. We detect the intercontinental-range peering at 12 IXPs, i.e., a majority of the studied IXPs. For example, the Italian AS E4A remotely peers at both TIE and TorIX, based in the USA and Canada respectively. Brazilian ASes comprise most of the remote peers at PTT, the largest among the 21 IXPs of the PTTMetro project in Brazil. The high fraction of remote interfaces at the Turin-based TOP-IX likely results from the IXP's interconnections with VSIX and LyonIX, two other Southern European IXPs located in Padua and Lyon respectively.

Switching the perspective from the interfaces to the networks that own them, we apply our network identification method (described in section 5.2.1) to determine ASNs for 3,242 out of the 4,451 analyzed interfaces. While a network might have interfaces at multiple IXPs, we identify a total of 1,904 networks. We refer to the number of the studied IXPs where a network peers as an *IXP count* of the network. Figure 5.4a presents the distribution of the IXP counts for all the 1,904 identified networks. While a majority of the networks connect to only one IXP, some networks peer at as many as eighteen IXPs.

285 of the identified networks have a remote interface at a studied IXP. Business services offered by the remotely peering networks are diverse and include transit (e.g., Türk Telecom), access (e.g., E4A and Invitel), and hosting (e.g., Trunk Networks). Figure 5.4a also plots the distribution of the IXP counts for all the 285 remotely peering networks. Both distributions in figure 5.4a are qualitatively similar, suggesting that the choice of IXPs for a network to peer is relatively independent of whether the network peers directly or remotely.

We also examine the remotely peering networks with respect to the minimum RTTs of their analyzed interfaces. For each IXP count, we consider all the analyzed interfaces of the remotely peering networks with this IXP count and classify the interfaces in regard to the following 4 ranges of minimum RTTs: [0 ms; 10 ms), [10 ms; 20 ms), [20 ms; 50 ms), and [50 ms; ∞). Figure 5.4b depicts the fractions of these 4 categories. While our study sets the remoteness threshold to 10 ms, the remotely peering networks with the IXP count of 1 have no interfaces with the minimum RTT below 10 ms. As the IXP count increases, the fraction of the remote interfaces tends to decline because some interfaces of the remotely peering networks are used for direct peering. E4A exemplifies networks with a large number of remote interfaces: 6 of its 9 analyzed interfaces are classified as remote.

5.2.3. Method validation

While our methodology employs a series of filters and high remoteness threshold of 10 ms to avoid false positives, this section reports how we validate the method and its conservative estimates of remote peering.

First, we use ground truth from TorIX, an IXP located in Toronto. TorIX staff confirmed that

their members classified as remotely peering networks in our study are indeed remote peers. In one case, the TorIX staff initially thought that a network identified as a remote peer by our method was rather a local member with a direct peering connection. Nevertheless, a closer examination showed that throughout our measurement period this local member conducted maintenance of its Toronto PoP (Point of Presence) and connected to TorIX from its remote PoP via a contracted layer-2 facility.

Then, we take a network-centric perspective and focus on E4A and Invitel. Both networks specialize in providing Internet access. Based on the measurements, our method classifies the E4A interfaces at DE-CIX, France-IX, LoNAP, TorIX, and TlE as remote. Using public information on IXP websites [5,118] and insights from private conversations, we confirm that E4A indeed peers remotely at these 6 IXPs. Our method also identifies Invitel as a remote peer at AMS-IX and DE-CIX, with the minimum RTTs of 22 and 18 ms respectively. Our private inquiries indicate that Invitel uses remote-peering services of Atrato IP Networks to reach and peer at AMS-IX and DE-CIX.

Finally, we receive an independent confirmation that our RTT measurement methodology is accurate. On our request, the TorIX staff measured minimum RTTs between the TorIX route server and member interfaces. Their results for our analyzed interfaces closely match our RTT measurements from the local PCH LG server. The mean and variance of the differences are respectively 0.3 and 1.6 ms.

5.3. Discussion: Remote peering repercussions in the Internet infrastructure

The Internet economic structure is important for reliability, security, and other aspects of Internet design and operation. In spite of its importance, the economic structure remains poorly understood. It is typically modeled on layer 3 of Internet protocols because economic relationships can be inferred from BGP [155] and IP measurements. In particular, BGP identifies ASes on announced paths, enabling inference of layer-3 structures where ASes act as economic entities interconnected by transit or peering relationships [85]. ASes are imperfect proxies of organizations, e.g., multiple ASes can be owned by a single organization and act as a single unit. Nevertheless, AS-level topologies [43, 172] have proved themselves useful for reasoning about Internet connectivity, routing, and traffic delivery. While being useful, layer-3 models struggle to detect and correctly classify a significant portion of all economic relationships in the dynamic Internet.

As discussed in previous chapters, validated evolutionary changes in the economic structure of the Internet include a trend toward more peering. The proliferation of peering relationships is caused partly by their cost advantages over transit. Internet flattening refers to a reduction in the number of intermediary organizations on Internet paths [24, 57, 89]. For example, the Internet becomes flatter when a major content provider expands its own network to bypass transit providers and connect directly with eyeball networks, which primarily serve residential users.

Internet flattening is routinely conflated with the trend towards more peering. Indeed, peering relationships are commonly established to bypass transit providers and thus reduce the number of organizations on end-to-end paths.

The trends toward more peering and Internet flattening are typically conflated because direct peering interconnections bypass transit providers and thereby reduce the number of intermediary organizations on Internet paths. Complementing direct peering, remote peering enables additional peering as well. However, this increase in peering involves a remote-peering provider that acts as an intermediary. Furthermore, the intermediary that sells the remote-peering service can be the same company that provided the bypassed transit services. Hence, remote peering increases peering without necessarily flattening the Internet economic structure.

The observed separation of the two trends questions the usage of AS-level topologies for representing the Internet economic structure. With remote-peering services provided on layer 2, layer-3 modeling of the Internet structure fails to distinguish remote peering from direct peering and ignores the intermediary presence of remote-peering providers. Below, we elaborate on various risks posed by this omission of the intermediary economic entities.

Layer-3 topologies can make the Internet structure look more reliable than it is. When a company employs the same physical infrastructure to provide transit and remote-peering services, buying both might not translate the redundancy into higher reliability for the multihomed customers.

The emergence of remote peering makes AS-level paths even less representative of the underlying physical paths. The hidden presence of layer-2 remote-peering infrastructures in layer-3 paths creates an additional reason why a path with the smallest number of ASes does not necessary provide the shortest delay of data delivery.

While it is common to use AS-level models for reasoning about Internet security, the hidden presence of layer-2 intermediaries adds to existing security concerns. The invisible intermediaries might be unwanted entities, e.g., those associated with problematic governments. The risks include monitoring or modification of traffic by the intermediaries and exposure of traffic to other parties, e.g., by delivering it through undesired geographies.

The reliance on layer-3 models also compromises accountability. Whereas a layer-2 intermediary might delay or discard traffic, attribution of responsibility for such performance disruptions is complicated because the middleman is invisible on layer 3.

Because remote peering has different economics than transit and direct peering, the omission of the layer-2 intermediaries from layer-3 models weakens economic understanding of the Internet. In developing markets such as Africa, remote peering becomes a cost-effective alternative for reaching well-connected areas in Europe and North America. Since remote peering has a smaller connectivity scope than transit, adoption of remote peering necessitates new strategies for traffic distribution. IXPs greatly benefit from remote peering: existing IXPs gain members, and new IXPs are enabled by bringing together a critical mass of traffic [140]. Ignoring the remote-peering providers distorts substantially the Internet economic landscape.

Thus, the increase use of remote peering demonstrated by our study calls for alternative models of the Internet structure that explicitly represent layer-2 entities. The relevant additions include not only remote-peering providers but also other layer-2 economic entities such as IXPs. With the growing prominence of IXPs and remote-peering connectivity to them, integrated modeling of the Internet structure on layers 2 and 3 becomes increasingly important for understanding the Internet. The refined mapping of the Internet economic structure will likely require novel methods for inference of economic entities and their relationships [62].

5.4. Related work

Internet structure. The Internet structure is highly important for multihoming [4], routing security [90], and various problems in content delivery via overlay systems [106,110]. By clarifying the Internet structure, our study of remote peering enables further advances in these and other significant practical domains. Because network operators do not publicly disclose connectivity of their networks, the research community relies on measurements and inference to characterize the Internet structure [97]. A prominent means for the topology discovery is the traceroute tool that exposes routers on IP delivery paths [47]. For example, Hubble [108] use traceroute to generate and maintain annotated Internet maps. Paris traceroute enhances traceroute with the ability to discover multiple paths [9]. A complementary approach is to utilize BGP traces [85, 169]. Our work employs active probing in the data plane to understand the role of remote peering in the Internet structure. Delay measurements are common in Internet studies, e.g., to understand evolution of Internet delay properties [117] or Internet penetration into developing regions. We use delay measurements to investigate how geography affects peering of networks.

Remote peering and interconnection arrangements. While previous research studies only mention remote peering [2, 44, 168], the measurement study of this chapter is the first to closely examine this emerging type of network interconnection. Our results show wide spread, significant traffic offload potential, and conditions for economic viability of remote peering. The Internet structural evolution [128] changes the dominant sources of traffic [49] and diversifies network types and their interconnection arrangements [38]. For example, partial transit [180] and paid peering [57] complement the dominant relationships of transit and peering. Our study of remote peering confirms the trend toward diversification of interconnection types.

Flattening of the Internet. The arguments that the Internet structure becomes flatter are multifaceted [57,114], with the continued growth of IXPs [2,35,168] cited in support of this trend. Our work reveals separation between the trends of increasing peering and Internet flattening. Also, while analyses of interconnection options are typically restricted to networks that share a location [165], our work exhibits remote peering as a cost-effective solution that enables distant networks to peer over a layer-2 intermediary. With our results showing the increasing opaqueness of the Internet structure from layer-3 perspectives, the opaqueness is likely to increase further with adoption of software-defined networking [11]. The large reach of remote peering illustrated

in this chapter calls for new approaches to mapping the Internet economic structure on both layers 2 and 3.

5.5. Summary

This chapter described remote peering, a technique in which ASes connect to various IXP remotely using the same infrastructure. The remote peering provider provides layer-2 connectivity to the various IXPs and manages the required equipment to connect them to the IXP back-plane. Remote peering helps IXPs to expand their geographical footprint and maintain growth. We describe the technical details of remote peering, show the results of measuring remote peering in various IXPs across the world, and discuss its implications in the Internet structure. The main conclusions of this chapter are:

Remote peering adoption. Using careful measurements of RTTs at 22 IXPs worldwide, our ping-based method exposed wide spread of remote peering, with remote peering in more than 90% of the examined IXPs and peering on the intercontinental scale in a majority of them. Up to 20% of the members of AMS-IX were found to be using remote peering.

Remote peering implications for network operators. Remote peering represents a valid option for operators to reach multiple IXPs at a moderate cost. In Chapter 7, we leverage this opportunity and provide an example of how operators can study the benefits for their networks of joining a large number of IXPs, either simultaneously or incrementally. Operators should consider that remote-peering providers represent a potential single point of failure, and should perform their network planning accordingly.

Impact of remote peering in the Internet infrastructure. By demonstrating the wide spread and significant traffic offload potential of remote peering, our results challenge the research community's reliance on layer-3 topologies in representing the Internet economic structure. Although the layer-3 Internet topology sees a more flattening structure, remote peering is reflected into a hierarchical layer-2 structure, with remote-peering providers playing a central role. Due to the failure to include the organizations that provide layer-2 remote-peering services, layer-3 models substantially distort the economic structure and can lead to incorrect conclusions about its properties, e.g., by conflating the trends of increasing peering and Internet flattening. Thus, the omission of remote-peering providers from traditional layer-3 representations of the Internet topology compromises research on Internet reliability, security, accountability, and economics. Our findings call for new topological models to represent the prominent role of layer-2 organizations in the Internet economic structure.

Part II

Inter-domain Traffic Management

Chapter 6

Framework for Inter-domain traffic Management

In the first part of the thesis, we explored and characterized the inter-domain peering environment. In this second part, we take the perspective of network operators, introducing models and tools that they can use to manage the inter-domain aspects of their networks under this environment. We start, in this chapter, by describing a general framework for the management of inter-domain traffic, and the difficulties that hinder their implementation across ASes. In the two consequent chapters, we provide details of two applications for inter-domain traffic management. Concretely, we describe methodology for IXP infrastructure extension in Chapter 7, and provide a system for warnings on policy conflicts leading to unsatisfied interests in Chapter 8.

6.1. Introduction

For operators, inter-domain traffic management consists on imposing their *policies* into their inter-domain traffic distribution by monitoring, characterizing, optimizing, and controlling it [10]. The *policy* of each AS is the result of the business requirements given by the enterprise in charge of the network (e.g. SLAs fulfillment, cost constrains, etc.). Unfortunately, the imposition of policy into the traffic distribution is not simple. There are various reasons for this difficulty. **First**, operators control inter-domain traffic using exclusively the reachability information exchanged to other ASes; therefore, policies must be translated to inter-domain route management using device configurations and the BGP protocol, as illustrated in Figure 6.1 [70]. This task is not trivial and prone to errors. **Second**, in a distributed system such as the Internet, the policy of an AS can conflict with those of others. Operators must thus not only know how to steer the traffic, but also know *how to assess* the efficiency of their strategies and the impact of the policies of external ASes on them. **Third**, the process and techniques required to control ingress traffic vastly differ from those used to control the egress traffic. For outbound, operators are able to modify and select among all paths received from neighbors. For inbound, operators are limited to influence

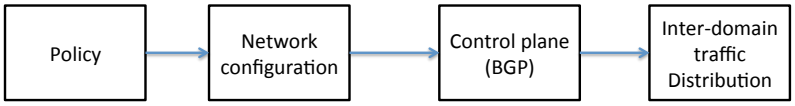


Figure 6.1: Process of imposing policy onto inter-domain traffic.

the decision process of others.

Operators can boost the cost-effectiveness and performance of their networks by increasing their number of peering links, employing caches, or using special transport services such as remote peering or partial-transit. The Internet environment depicted in the first part of this thesis shows that ASes are using these strategies in practice. These benefits, however, are not free. Operators must ensure that they are able to manage the interaction of their networks with the policy of multiple peers, and that they receive all the benefits from the services they acquire. Operators should manage their inter-domain traffic using procedures adapted to an environment in which their networks interact with external ASes who possess dynamic policies and networks that can fail at multiple levels.

A general framework for inter-domain network management is depicted in Figure 6.2. which is based on the model found in [176] [78]. The framework consists of a cycle of four procedures: data-collection; validation and verification; optimization; and operation. These blocks are supported by a simulation procedure. In the **data collection** phase, operators use various tools to obtain the data of the network. The **validation and verification** procedure warns operators against situations (in deployment or in planning) that should be tackled by operators. These situations include: (1) performance degradation (link overload, dangerous failure cases, etc.); (2) potential configuration errors (e.g. Route-leaks); (3) divergence of the simulation model of the network (model verification); (4) or, prominently, the functions that ensure that internal policies are not violated by external ones. Using information coming from the validation and data collection blocks, operators decide on the actions required to **optimize** the network. Finally, the **operation** block contains all functions necessary to implement the decisions formulated in the optimization process. In the next sections, we extend the description of each of these procedures.

6.2. Data Collection

Data-collection functions gather the data required for a proper inter-domain management. This includes not only data from the analyzed network, but also data (control-plane and policies) from external ASes. The collection of network data for inter-domain management is usually considered complex for network operators due to its large and frequently incomplete nature, which limits the efficiency that traffic management methods can achieve. The introduction of SDN and big data management technologies may change this rigid context, as it is pushing operators to request flexible routing system architectures. In the future, the network operation team might have the resources to obtain, maintain, and analyze a rich variety of data from within and outside their

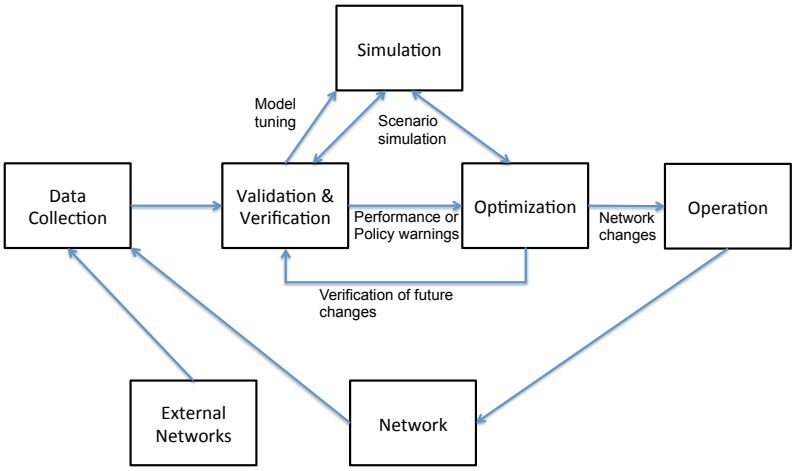


Figure 6.2: Inter-domain management framework.

networks. Without the need to change many of their routing devices, such data can be leveraged to implement more complete inter-domain management applications. In this section, we overview different methods to obtain this data, their limitations, and briefly provide recent techniques and protocols, proposed in the last few years, to facilitate data collection.

6.2.1. Collecting received BGP paths and paths installed in FIB

BGP feeds describing inter-domain paths reaching the border of an ISP are necessary for analyzing the different route alternatives available at each router in the network. In order to perform a complete network analysis, ISPs need to collect every path received from external neighbors. The simplest architecture for BGP data collection consists of a centralized collector, which communicates with all edge BGP devices. The completeness of the data collection depends on the approach / protocol followed to convey the data to the collector. We describe some of the options next.

- **CLI Scraping**. The usage of custom scripts, e.g., based on router Command Language Interpreter (CLI) commands and screen scraping.

- **iBGP.** The configuration of iBGP sessions with add-path [181] (or similar features to propagate all BGP routes) between edge routers and a route collector (such as [123]) [50];

- **BMP.** BGP Monitoring Protocol (BMP) is a simple protocol that provides access to the information stored in the Adj-RIB-in tables of BGP devices [163]. The BMP specifications support the signaling of paths before and after the application of inbound policies. Nevertheless, some manufacturers might opt to only support one of these modes.

- **I2RS (Future).** The Internet Engineering Task Force (IETF) is working on a standardized interface to the routing system of network devices denominated Interface to the

Routing System (I2RS) [99]. I2RS should provide the necessary interface for a centralized controller to fetch all paths received by edge devices from external ASes, including information of the state of the path in the Routing Information Base (RIB) and Forwarding Information Base (FIB) of the routers. The time of availability of such technology is however not yet known.

Some applications must be aware of the paths installed in the FIB of the routers. A BGP collector can learn them by connecting to all routers in the network (since each router would announce the best path). This is not always feasible, since some providers might opt to collect BGP information from route reflectors. Moreover, this technique would not function under cases where multi-path is required. Applications could simulate the extended BGP decision process under all paths received to estimate the paths installed in the FIB by the routers, although this process might be slightly different depending on the BGP implementation of each device. Another alternative is to transmit all paths available by each router and signal the paths installed in the FIB using a community, this idea is described in [29].

6.2.2. Collecting traffic data

Typically, ISPs collect traffic statistics to perform a variety of business-critical activities, ranging from accounting and billing to intra-domain traffic engineering. For inter-domain management purposes, an ISP should be able to monitor the traffic per prefix at every interface connecting the network to external ASes. More granular information, such as traffic per application or transport protocol, can also be very useful in specific cases. *Netflow* and *sFlow* are the two most popular technologies for collecting this data [55, 148]. The traffic data generated by ISP networks can be quite large. To cope with this overhead, operators can decide to focus only on the traffic observed in peak networking hour times. Likewise, operators could simplify many traffic management techniques by considering the prefixes that drive most of the network traffic [71].

6.2.3. Collecting external AS policies

Every AS in the Internet has the freedom of filtering, redistributing, and altering the paths it receives from neighboring ASes. ISPs might need to evaluate the impact of the policies of external ASes in their networks. The estimation of external polices is extremely complex, but, some data can be used to obtain a starting approximation of them. We summarize some of them in the following:

- **AS relationships.** The knowledge of the relationships among the AS conforming the close AS neighborhood of an ISP can be useful for different inter-domain traffic management applications (as the ones we will describe in Chapters 7 and 8). The estimation of the relationship is difficult, and requires a high level of analysis of BGP data [125]. An ISP can obtain this information through some commercial companies, through public data [125], or

thanks to social interaction, since operators usually know the relationships of many of their peers.

- **External BGP data.** Some public Internet projects, such as route views or RIPE, release the BGP data of devices located in different points of the Internet [133] [157]. This information can be used to obtain a partial view of the policies of external AS. For instance, it can help determine if some ASes filter prefixes or modify their path attributes [126] [60]. Also, as described in [193], the absence of events provides valuable of information for the analysis of distributed systems. For egress inter-domain traffic, using external BGP data to find missing paths that the network should be receiving, but are not, provides guidelines on how the policy of external ASes affects the network. We provide details on how to find missing paths from external ASes in Section 8.3.2.4.

The described data is a good starting point for the analysis of policies of external ASes. Nonetheless, ISPs must understand that it is unfeasible to obtain a complete view of the policies of the ASes in the Internet. Hence, this data can include some inaccurate information, and operators must consider this fact at the time of the analysis. In general, network operators still require tools for the conversion of large amounts of data (routing, traffic, etc.) into exploitable data that can be easily understood by their management, design, and architecture teams.

6.2.4. Collecting network infrastructure and Shared Risk Link Groups

An inventory of the physical devices of the network is useful for troubleshooting and network management. For network design, a summary of the physical location and the groups of devices that can fail simultaneously, also referred to as Shared Risk Link Group (SRLG), can provide engineers with information necessary for network failure analysis. The automatic discovery of SRLGs has been extensively explored in optical networks [164]. For the case of IP networks, operators must still rely on proprietary applications or protocols to perform network inventory and SRLG knowledge construction. For the specific case of inter-domain routing, SRLGs could group elements performing various inter-domain network functions such as route reflection or heavy BGP sessions nodes, which could fail simultaneously. IXP connections using the same remote peering provider should also be included in a single SRLG, even when the service to different IXPs is delivered via distinct physical ports (as explained in Chapter 5).

6.2.5. Collecting path performance details

Content provider networks or other ASes that offer real-time applications might need to obtain data on the performance characteristics of the paths received by their neighboring ASes. Delay or packet loss can be used by these companies to select the path that their packets should take to reach an end user. This type of companies might prefer good performance paths over paths traversing cheaper links. This can push them, for example, to send packets through a transit

provider, instead of a peering link, when the latter suffers from high delay or packet loss due to link congestion [161] [61]. The collection of performance data requires the measurement of packet statistics, which can be obtained using probes, or taken directly from user applications. Operators still face the challenge of correlating this data with control-plane information (BGP paths). These can be facilitated if the respective data is stored in information systems with flexible and standard interfaces [124].

6.2.6. Collecting internal Policies

Each autonomous system decides how to manage the routing information received from external ASes. The *routing policy* of each AS defines what to do with the paths it receives from their neighboring ASes: to which other ASes it can propagate the routes, how to modify the paths, whether to filter them or not, etc. This policy is reflected in the configuration of edge devices.

Maintaining the policy information not only at the edge devices, but also in a centralized location can help operators to match which network states and events conflict with the local policy. A collector could extract policy information from the configuration of edge devices, but this would require it to parse and simulate the policy descriptions for a large diversity of vendors and network operative systems (e.g. route maps / routing policy language in Cisco, Policy Framework for Junos). New network applications and devices are increasingly supporting protocols such as NETCONF/YANG [17], which provide a flexible interface for network configuration. Also, external policies can be obtained via the Internet Routing Registries (IRR), when available, although this has been proven to be inaccurate. These protocols could be leveraged to build centralized network management solutions, which would allow operators to maintain policy related information in a single database.

6.3. Simulation

Simulation tools estimate the full or partial state of the network under different scenarios. Optimization and validation applications rely on these tools to evaluate network performance under specific circumstances. Simulation tools designed for estimation of outbound traffic differ from the ones used for inbound traffic. In this section, we provide an overview of the each of them, and describe the difficulties that arise when developing tools for these purposes.

6.3.1. Simulation for inter-domain outbound traffic

For outbound traffic, an AS uses the reachability information (BGP paths) received from external networks to decide, from all available paths, the ones that the network can use for sending their traffic. Any tool used to estimate the egress inter-domain traffic should be able to simulate (i) the transformations performed on incoming paths (policy), (ii) the BGP decision process of the

network to define the paths that are selected, and (iii) the behavior of nodes with multiple paths installed in their FIB. Each of these steps carries significant technical challenges.

Concerning the simulation of inbound policies (i), difficulties arise when trying to express and simulate all possible policy logic options offered by router vendors, which can even involve algorithmic type of behaviors (e.g. defining policies with linked conditionals and priorities).

The network-wide BGP decision process (ii) is intrinsically complex to simulate. The BGP decision process is not always deterministic. It might converge differently depending on the order of the updates, or might even not converge at all [95]. This is exacerbated if internal iBGP policies are employed [184]. Also, simulation tools should also support different features of BGP such as ADD-PATH. Event-driven tools are powerful for simulating the BGP decision process under these characteristics, but they are more complex to build and maintain. Network operators can design their networks to be deterministic in the BGP sense, thus facilitating the simulation process [73]. Although the Internet routing table consists of more than 500k prefixes, simulation tools can employ techniques such as grouping the prefixes with the same policy [69] or analyzing only the prefixes with more traffic demand in order to speed up the simulation process. These simulation tools should also account for any architectural design, such as the use of different route reflector hierarchies, or the use of technologies like segment routing [76] to select a specific path for some prefixes.

Finally, simulation tools should include the load balancing behavior of routing nodes (iii). Vendors apply their own mechanism for traffic balancing when multiple paths are installed in the FIBs of routers. That is, not always the outbound traffic when N paths are available is equally shared among the N paths. Simulation tools should be able to provide acceptable estimations of the load balancing for the conditions of the network.

6.3.2. Simulation for inter-domain Inbound traffic

For inbound traffic, since each AS is free to choose the paths it uses for its outbound traffic, an AS can only try to influence the path selection of external ASes. A proper simulation of inbound traffic should estimate how external ASes would react to announcements made by the AS (i.e. which paths the other ASes would select). A perfect simulation of the decisions and policies of external ASes is intractable, due to difficulty of obtaining reliable or complete data on their policies. Processes relying on simulations of inter-domain inbound traffic should account for these difficulties, and provide methods to tune the models and deal with possible errors.

6.4. Validation and verification

The validation and verification block encompass tests for different levels of the network design phase, such as evaluating network performance, verifying the simulation models used for planning, or checking the influence of policies from external ASes. These methods not only involve checks at the technical level, but also at the economical level.

Validation and verification are functions used in system or model testing [160]. *Verification* functions check the internal operation of the model or system. Unit-testing or static analysis are typical verification tests. *Validation* functions check that the system fulfills the requirements of their users.

We classify the validation and verification functions according to the component / object they test. We explain each of them next.

- **Data collection.** Operators should verify that the data collected by their network is reliable, since this data supports the whole inter-domain management process. Controlled tests environments or comparison of different measurement systems can be useful for this. Also, since some of the data required might be prone to errors (such as the policies or relationships of external ASes), operators would need to validate them when they are important for the overall process.

- **Network state.** The network state should be verified and validated against the expectations and requirements of operators. Verification processes should ensure that the routers operate correctly and according to the protocols ruling them[1]. Validation processes for this component cover all tests required to ensure that the network offers the performance meeting the policy of the operators. For instance, making sure that links are not saturated, or that the company is not paying too much to its providers. Validation tests are not only limited to the current state of the network, but also the network state in hypothetical situations, such as the one created when traffic grows, or after some disruptive event (e.g. link failures).

- **External networks.** The performance and policies of external ASes influence the operation of the network. In terms of performance, operators often agree with their neighboring networks to SLAs, which can be verified using several metrics, which include latency, jitter, or packet loss. Operators must ensure that these values are within the agreed ones. The policy of external ASes are reflected in the reachability information they exchange and the traffic they sent to the network. Verification and validation of these policies encompass different tests. For instance, operators can use prefix filtering or Resource Public Key Infrastructure (RPKI) systems to detect routes that should not be announced by neighboring network. Operators can also analyze how external ASes propagate and modify the prefixes originated by them. Validation functions can also evaluate the state of the network if the policy of the external ASes were different, thus motivating network managers to try to persuade changes in policy to their counter-parts. In Chapter 8, we delve into the problem of policy verification and validation, and describe with detail several validation tests that operators can use to assess their impact in their network.

- **Simulation and optimization.** The simulation and optimization block supports the

[1] We refer here to the inter-domain related infrastructure of the network. Nevertheless, the correct operation of the inter-domain traffic depends on its intra-domain counterpart and should also be tested.

framework by evaluating the inter-domain network state under different scenarios and selecting the changes that would lead to a better state. Operators might in some cases be able to run labs [166], or configuration tests [82] to verify the simulation models and potential configuration before they are applied (this is symbolized by the feedback arrow between the optimization and validation process in Figure 6.2). Nevertheless, the nature of the inter-domain environment difficulties the implementation of these tests, forcing the use of the operative network data to check and tune simulation and provisioning components. Simulation models can fail, for instance, by not correctly inferring the internal propagation of paths, or the traffic distribution in a multi-path scenario. The optimization block can fail by not providing solutions that are sufficient for the requirements of the operators, or by not reflecting correctly the policies of the operators into router configuration and BGP changes (Figure 6.1). We stress that these processes are not necessarily implemented in automatic systems, but can be performed in the organization by the operators themselves.

The aforementioned tests could, in theory, be separated in the framework. We joined them under the same block due to different reasons. **First,** labs or testing facilities for the inter-domain environment are seldom available or reliable. Thus, operators can only use the data of the network itself to validate and verify all components. **Second**, many of these tests are entangled, in the sense that all of them should be run in order to test the overall system. For instance, a divergence of the inter-domain traffic to the calculations of a what-if scenario from the simulation process might be caused by a malfunctioning model, by incorrect data from the data-acquisition system, or by a conflicting policy from a neighboring ASes. **Third**, some tests can apply to many of these functions. For instance, testing whether traffic from a settlement-free neighbor appears at a transit link can be a test for the correctness of the simulation/optimization model, to validate the policy of external ASes, or to validate network state. *In practice, operators run different tests and correlate the information provided by them to find problems, root causes, and opportunities for improvement.*

6.5. Optimization

The optimization process encompasses all changes in traffic control or infrastructure to improve the performance and profitability of the network, either in the current state or in potential future scenarios.

Traffic control mechanisms include all techniques focused at steering the inter-domain traffic under the same infrastructure. This implies, in outbound traffic, for operators to decide the priority of the paths received [178]. For inbound traffic, operators can use different techniques for this end, such as selective advertisement, path prepending, or MED tuning to influence the path selection of external ASes [71].

Infrastructure optimizations extend or modify the structure of the network. For inter-domain routing, this can apply both to physical (e.g. expanding capacity, or connecting to a new IXP), or

logical resources (e.g. new peering sessions). Expanding the network to reach new neighboring ASes requires operators to evaluate the benefits that they would obtain. One could, for instance, use external data to estimate which prefixes can be exchanged with other ASes.

The optimization of traffic on ISP network is a multi-objective optimization process, in which operators should make decisions based on various techno and economical aspects, often selecting solutions that sacrifice some aspects. For instance, in case of saturation with a settlement-free peer, should the network steer some traffic exchanged with this peer to a transit provider? Is the operator willing to have a less optimal solution by limiting the number of configuration changes? Due to these conditions, optimization procedures might better be implemented using a hybrid solution between calculation systems, to explore potential changes, and human input accepting the best changes.

The granularity of the traffic flows that operators can steer affects the optimization model for both traffic directions. For outbound traffic, operators are limited to only modifying the attributes of received paths at the edge routers. In those cases, changes in the path attributes are propagated through the whole network, thus making it complex for optimization systems to distribute the traffic for a single prefix across possible exit points. This level of granularity might not be sufficient to achieve traffic balance in certain topologies. The use of iBGP policies (changing path attributes over iBGP sessions) can help increase the granularity of traffic changes, but they are discouraged as they can create control-plane instabilities [184]. A BGP controller could also inject artificial more specifics for this objective (i.e. prefix deaggregation), although operators must be careful on not propagate them into other neighbors (this was the source of one route-leak problem in 2015 [119]). Other option is to use a controller to signal the router to steer a sub-set of traffic through one of the exit points. Segment Routing , for instance, offers an Egress Peering Engineering option for this purpose [76]. Inbound traffic also suffers from problems related to the capacity of operators of steering granular amounts of traffic. Prefix deaggregation is the more the straightforward solution for this problem, but has caused the increase of the routing table [127]. Selective scoping communities [60] offer another option to deaggregate the traffic, but they are harder to manage.

6.6. Operation

The operational component includes all actions aimed at applying the changes outputted from the optimization process. This component contains elements like provisioning systems (either automatic or human based), acquisition processes, or new peering negotiation. The peering manager is an important figure in this component (and in the whole framework), as this one is in charge of negotiating with other companies any peering relationship.

6.7. Related Work

Network planning frameworks. The framework described in this section is based on the ones discussed in [176] and [78]. We extended this framework to emphasize the validation and verification process, which is particularly important for the inter-domain environment. Other network planning frameworks exists, such as the one portrayed in [182]. Different from individual networks, the inter-domain planning requires knowledge of the policies of external ASes. Some authors provide methods to estimate these policies, but their complex essence forces operators to use manual and automatic procedures, and probably many tests, to figure them out. In this regard, the discovery of external policies can resemble Knowledge Discovery Systems (KDSs) [51] or Decision Support Systems (DSSs) [151], which are cited more frequently in the Business Intelligence (BI) literature.

Inter-domain traffic engineering. Several authors have proposed optimization techniques for outbound [178] [73] [152], and inbound traffic [153] [88]. These proposals differ on the employed optimization techniques; their metrics; or whether they prepare the network for different traffic states and disruptive events (e.g. link failures). Some authors have also looked at the global optimization of inter- and intra-domain traffic [13] [188]; however, these can become very complex. Operators can isolate the inter-domain traffic from the intra-domain traffic by using technologies such as MPLS.

Validation of the Inter-domain operation. The RPKI system was developed with the purpose of avoiding forged or invalid routes in BGP [21]. Static analysis techniques have been proposed for validation of device configuration, thus preventing, for instance, the creation of route leaks [70]. Other tools examine policies of external ASes such as their compliance of consistent advertisement [72], or disruptive prefix filtering [126]. In Chapter 8, we discuss techniques to detect and classify policy conflicts with external ASes. We provide a further description of the literature related to this topic in that chapter.

6.8. Summary

This chapter introduces a framework for the management of inter-domain traffic. We explained the steps forming the framework, including data collection, validation procedures, and optimization techniques. In addition, we provided examples of procedures and tools fitting each one of these steps. We stress that not all inter-domain management processes might fit perfectly into the framework. However, the framework highlights the needs of validation, what-if scenario capabilities, and optimization functions that are important in practice.

In the next chapters, we provide details of inter-domain management applications dealing with peering infrastructure extension (Chapter 7) and external policy validation (Chapter 8). We expose next the main conclusions from this chapter.

Difficulties for data collection. Inter-domain traffic management requires the collection and

correlation of data from different systems. Also, operators might only be able to acquire partial or estimated data, such as the policy of neighboring ASes. Operators should thus include processes that can function with large, incomplete, or partially wrong data.

On the importance of validation. The infeasibility of simulating the Internet environment forces operators to use the running network data to validate the results of their inter-domain management procedures. These functions should provide operators with the feedback require to trigger tune the simulation models, launch new optimizations, perform internal configuration checks or deal with conflicting policies of external ASes. In Chapter 8, we illustrate how incompatible interest can result in unsatisfied interest for an AS. We present in that chapter a validation system that operators can use to detect these unsatisfied interest and highlight those that have a larger impact in the network.

Chapter 7

Peering expansion using remote peering

Recognizing the benefits of enriching the interconnections of their networks, many operators constantly evaluate the expansion of their peering infrastructure. The IXP based peering environment, described in Part I of this thesis, facilitates this operation. Although peering management is a social-centric process[1], operators and peering coordinators can use tools to help them find the best peering strategies, and the best IXPs to expand. These tools should consider scenarios in which the network potentially joins different IXPs, or peers with companies with different peering policies.

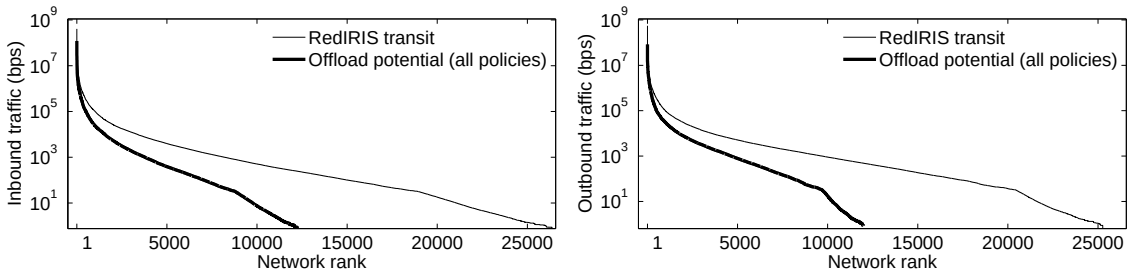
In this chapter, we provide an example of a peering expansion study leveraging the large IXP offer and remote peering services. Using the data of a real network, we first perform a typical peering study, analyzing the top potential ASes, to which the company should directly peer. Subsequently, we evaluate the potential for the company to go to single IXPs under different peering policy scenarios. We show how the overlapping of IP space diminishes the marginal gain of transit off-loading when peering at more than one IXP. However, thanks to remote peering, ASes can reduce the operative costs of joining more than one IXP and we calculate the potential off-loading traffic for extensively using this service. Finally, we discuss other advantages of improving peering and partially illustrate them using RedIRIS data.

The rest of the chapter is structured as follows. We describe the traffic data we use for our study in Section 7.1. Section 7.2 describes and evaluates the peering study based on transit off-load traffic. We discuss about other potential gains of peering at multiple IXPs in Section 7.3. We present related work in 7.4, and conclude in Section 7.3.

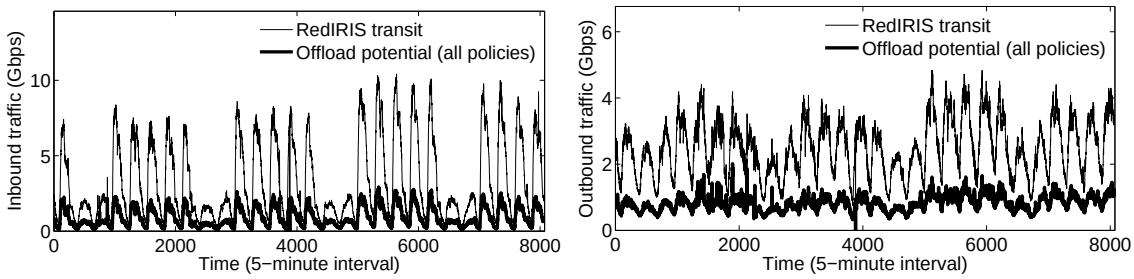
7.1. Traffic data

We collect and use traffic data from RedIRIS, the NREN (National Research and Education Network) in Spain. This network interconnects with GÉANT (backbone for European NRENs), buys transit from two tier-1 providers, peers with major Content Delivery Networks (CDNs), and

[1] As reported in [192] a large percentages of settlement-free peering arrangements are performed informally.



(a) Distributions



(b) Time series

Figure 7.1: Network contributions to the transit-provider traffic and offload potential with peer group 4

has memberships in two IXPs: CATNIX in Barcelona and ESpanix in Madrid. In February 2013, we used NetFlow to collect one month of traffic data at the 5-minute granularity in the ASBRs (Autonomous System Border Routers) of RedIRIS.

Among all the inter-domain traffic, only the traffic between RedIRIS and its transit providers might contribute to the offload potential. Depending on whether RedIRIS receives the traffic from its transit providers or sends the traffic to them, we respectively classify the transit-provider traffic as *inbound* or *outbound*. The collected dataset identifies networks by their ASNs and contains records for 29,570 networks that are origins of the inbound traffic or destinations of the outbound traffic.

To illustrate the contributions of the 29,570 networks to the transit-provider traffic of RedIRIS, we report how much traffic each individual network contributes as an origin of the inbound traffic and destination of the outbound traffic. Figure 7.1a plots the average traffic rates for the respective inbound and outbound contributions by the individual networks during the measurement period. The figure ranks the networks in the decreasing order of the contributions. While a few networks make huge contributions close to the Gbps mark, most networks contribute little. In the range where the networks are ranked about 20,000 and contribute average traffic rates around 100 bps, the distributions of the inbound and outbound traffic exhibit a similar change in the qualitative profile of the decreasing individual contributions: a bend toward a faster decline. While the raw data exhibit the bend as well, reasons for the bend constitute an interesting topic for future work. Figure 7.1b reveals daily and weekly fluctuations in the transit-provider traffic of RedIRIS, with

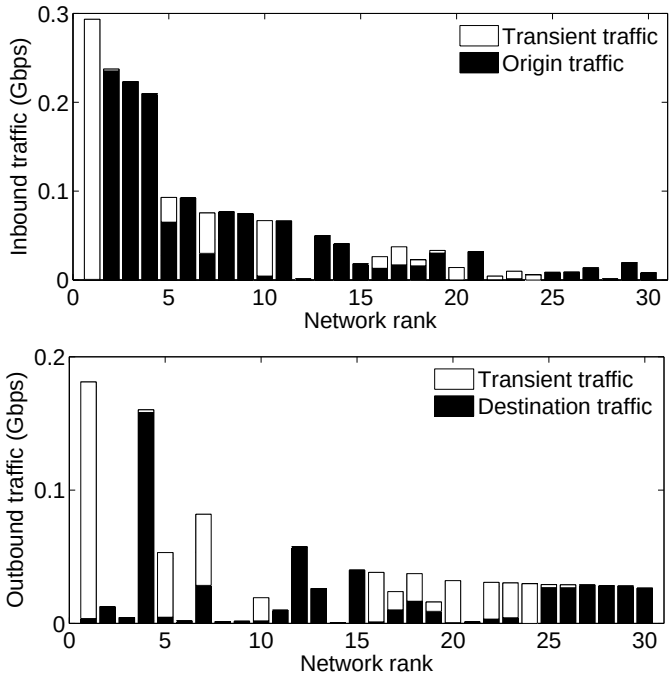


Figure 7.2: Origin and destination traffic vs. transient traffic for top contributors to the offload potential

periodic fluctuations being clearly pronounced for the inbound traffic.

7.2. Offload scenarios and evaluation

RedIRIS cannot offload all of its transit-provider traffic. The offload potential depends on the set of IXPs that the network is able to reach via remote peering. Also, the set of members of the reached IXPs do not include all the networks that contribute to the transit-provider traffic of RedIRIS. Finally, not all the members of the reached IXPs are likely to peer with RedIRIS.

For the set of IXPs that RedIRIS might be able to reach, we consider the Euro-IX association formed, as of February 2013, by 65 IXPs from all the continents [7]. Based on Euro-IX data from February 2013, we limit potential peers of RedIRIS to the members of these 65 IXPs.

We further trim the group of potential peers by excluding the networks that are highly unlikely to peer with RedIRIS. First, we do not consider the transit providers of RedIRIS as its potential peers because transit providers typically do not peer with their customers. It is worth noting that no network sells transit to these two tier-1 providers, and thus no such network needs to be excluded due to its transitive transit relation with RedIRIS. Second, since RedIRIS already has memberships in CATNIX and ESpanix, the other members of these two IXPs are disregarded as candidates for remote peering with RedIRIS. In particular, we exclude all the other tier-1 networks because they have memberships in ESpanix. Third, due to the cost-effective interconnectivity that comes with the GÉANT membership, we do not consider the other GÉANT members as

potential peers of RedIRIS. After applying the above three rules, the group of potential remote peers of RedIRIS reduces to 2,192 networks. Even after eliminating the highly unlikely peers, there remains a significant uncertainty as to which of the 2,192 networks might actually peer with RedIRIS.

To deal with the remaining uncertainty about potential peers, we examine a range of *peer groups*, i.e., groups of networks that might peer with RedIRIS. Using PeeringDB which reports peering policies of IXP members [120, 147], we compose the following 4 peer groups so that the peering policies of their members comprise:

[peer group 1] *all open policies*;

[peer group 2] *all open and top 10 selective policies*,
which adds to peer group 1 the 10 networks that have the largest offload potentials among the networks with selective policies;

[peer group 3] *all open and selective policies*;

[peer group 4] *all policies*, i.e., all open, selective, and restrictive policies.

Peer group 4 constitutes our upper bound on the likely peers of RedIRIS. When RedIRIS reaches all the 65 IXPs, this peer group 4 includes all the aforementioned 2,192 networks. Peer group 1 represents a lower bound on the networks that might actually peer with RedIRIS. It is common for such open-policy networks to automatically peer with any interested IXP member via the IXP route server [156].

For each peer group, we determine the offload potential of RedIRIS by fully shifting to remote peering the traffic that the networks of this peer group, and their customer cones, contribute to the transit-provider traffic of RedIRIS. The customer cone corresponds to all ASes that an AS can reach using customer links.

In addition to studying sensitivity of the offload potential to the peer groups, we also evaluate its sensitivity to the choice of reached IXPs. Specifically, our evaluation explores the set of reached IXPs from a single IXP to all the 65 IXPs in the Euro-IX data.

7.2.1. Offload evaluation results

We start by estimating the **maximal offload potential** with peer group 4 (all policies) when RedIRIS reaches all the 65 IXPs. In this scenario, RedIRIS offloads traffic of 12,238 networks which represent the joint customer cone of the 2,192 companies included in the peer group 4. Figure 7.1a shows how much traffic these 12,238 networks contribute to the offload potential in the inbound and outbound directions. The plot ranks the networks in the decreasing order of their traffic contributions. The results suggest that the maximal offload potential is substantial: RedIRIS offloads around 27% and 33% of its transit-provider traffic in the inbound and outbound directions respectively. While the inbound traffic dominates the outbound traffic, figure 7.1b reveals that the peaks of the transit-provider traffic and offload potential of RedIRIS consistently

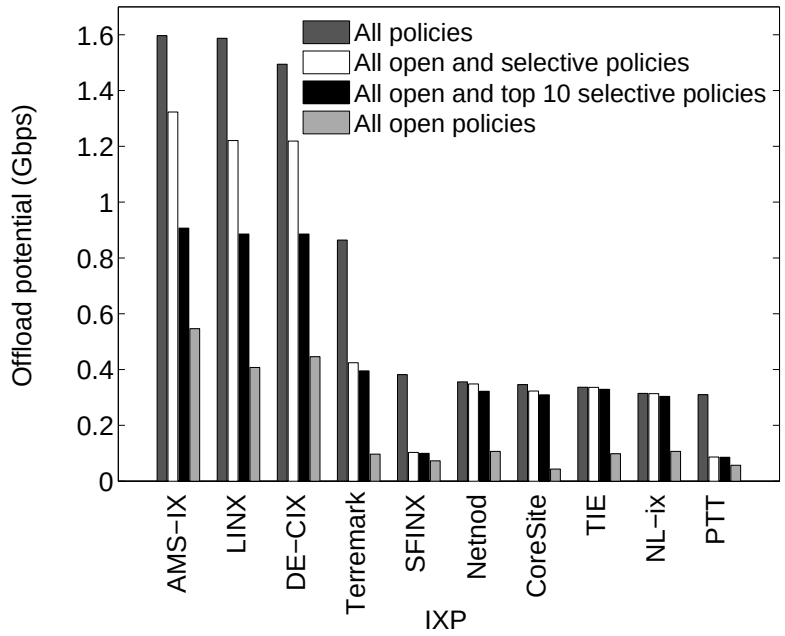


Figure 7.3: Offload potential at a single IXP

coincide, implying that the traffic offload can reduce transit bills, which are typically determined by traffic peaks.

Figure 7.2 zooms in on the top 30 contributors to the maximal offload potential. These 30 networks make the largest traffic contributions to the combined inbound and outbound offload potential and are the main candidates for **private peering**, in case the traffic exchanged fits their peering policies. The top contributors include Microsoft, Yahoo, and CDNs, suggesting that content-eyeball traffic features heavily in the offload potential. For a majority of the top contributors, the origin and destination traffic dominates the transient traffic.

Switching to the sensitivity analyses, we first evaluate the offload potential for all peer groups when RedIRIS **reaches a single IXP**. This single IXP is chosen among the 10 IXPs where RedIRIS has the largest offload potential. Figure 7.3 reports the offload potential of RedIRIS at each of the 10 IXPs. The top 4 of the IXPs include the big European trio (AMS-IX, LINX, and DE-CIX) and Terremark from Miami, USA. For any of the peer groups, the offload potential is similar across the 3 largest European IXPs because these IXPs have many common members (see Chapter 3). On the other hand, the offload potential at Terremark is significantly different due to its different membership: numerous members of Terremark from South and Central America [129] contribute significantly to the transit-provider traffic of RedIRIS and are not present in Europe.

We now assess the additional value of reaching a **second IXP** after RedIRIS fully realizes its offload potential at a single IXP. When the two reached IXPs have common members that contribute to the transit-provider traffic of RedIRIS, realizing the offload potential at the first IXP reduces the amount of traffic that RedIRIS can offload at the second IXP. For peer group 4 (all

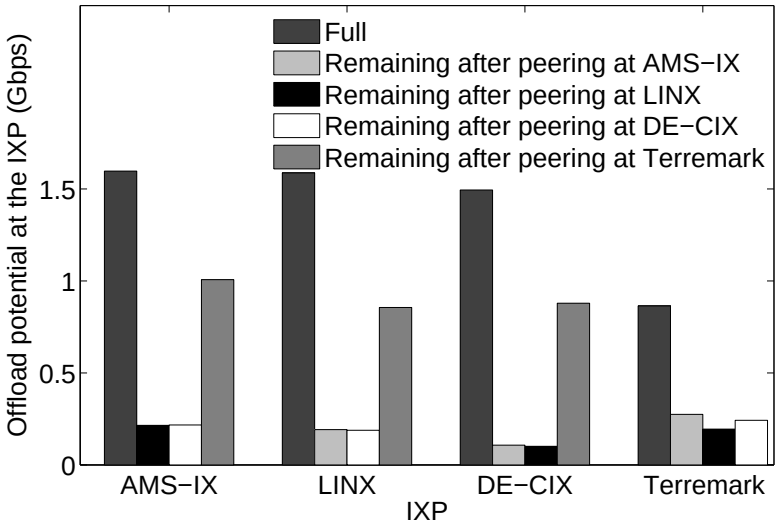


Figure 7.4: Additional value of reaching a second IXP after realizing the offload potential at a single IXP

policies), figure 7.4 illustrates this effect when AMS-IX, LINX, DE-CIX, and Terremark act as either first or second IXP. When LINX and AMS-IX act as the first and second IXPs respectively, the offload potential remaining at AMS-IX after fully realizing the offload potential at LINX is 0.2 Gbps, which is much lower than the full potential of 1.6 Gbps at AMS-IX. When Terremark acts as the second IXP, the decrease in its offload potential is less pronounced because Terremark shares only about 50 of its 267 members with either of the 3 largest European IXPs.

Generalizing the above, we examine the additional value for RedIRIS to reach **extra IXPs using remote peering**. We iteratively expand the set of reached IXPs by adding the IXP with the largest remaining offload potential. For peer group 4, the first 4 reached IXPs are added in the following order: AMS-IX, Terremark, DE-CIX, and CoreSite. For all the 4 peer groups, figure 7.5 plots the remaining transit-provider traffic of RedIRIS as the number of reached IXPs increases. The overall reduction in transit-provider traffic of RedIRIS varies from 8% for peer group 1 (all open policies) to 25% for peer group 4 (all policies). Figure 7.5 shows that the marginal utility of reaching an extra IXP diminishes exponentially and that reaching only 5 IXPs enables RedIRIS to realize most of its overall offload potential.

In Chapter 3, we calculated the number of hosts that could be reached by any company after joining different IXPs. Figure 7.5 reflects the specific case of RedIRIS using a similar procedure, but in terms of traffic. We re-plot the control-plane results in Figure 7.6, in order to be able to compare it with the results of RedIRIS. The two figures resemble a similar exponential diminishing shape (one in the control-plane, the other on the data-plane). In [37], we formulate an economic model evaluating the benefits of remote peering using this generalized pattern.

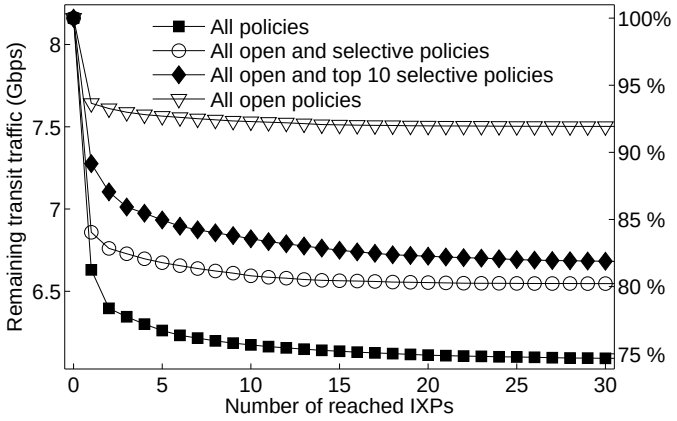


Figure 7.5: Additional value for RedIRIS to reach an extra IXP

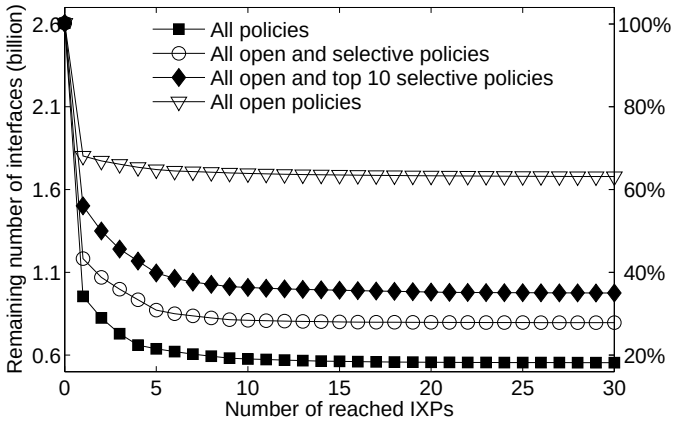


Figure 7.6: Generalized additional value of reaching an extra IXP

7.3. Quantifying other benefits

Transit traffic offloading provides a direct estimation of the savings that a company can obtain by peering at external IXPs, thus facilitating network managers to prepare a business case for peering expansion. Other advantages of establishing multiple peering links are the reduction of transport, or increasing resiliency bounding economic cost. The former is related to hot-potato routing, as by adding more peering links companies can potentially deliver traffic closer to the source. The latter advantage relates to the opportunity of still relying on peering links, even when other peering links fail.

Peering at multiple IXPs would not decrease the overall transport distance of packets on our case, since RedIRIS already peers at the two IXPs that are close to their networks and customers. In fact, in case RedIRIS does peer at other distant IXPs, its network managers should be careful to not send traffic to these if closer exit points exist.

To study the resiliency advantages of extending to new IXP, we first look at the peering traffic from the network. Figure 7.7 depicts the inbound and outbound traffic exchanges with the largest IXP the network uses (ESpanix). The traffic over this link is comparable to the transit traffic, and

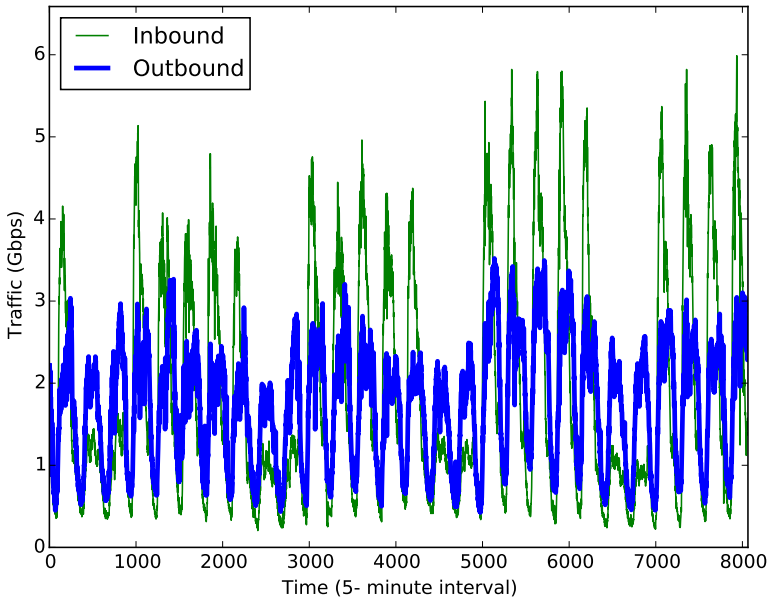


Figure 7.7: Traffic over link to main IXP.

would cause a potential bill increase in case there is a failure at the IXP or on the link connecting with it, driving this traffic to the transit provider links.

Figure 7.8 show the top 30 ASes contributing to this traffic (values based on the 95-percentile). The top peer is causing around 50% of the inbound traffic, and 14% of the outbound. The top 5 ASes contribute 65% and 80% of the inbound and outbound traffic, respectively. The top 8 peering ASes are connecting directly at ESpanix. The reminding ASes either connect at ESpanix or are reached via one AS connecting there.

Operators can evaluate the benefits of peering at other IXP in terms of redundant peering links, by simulating the traffic that would go to them in case of failure of the main IXP, and comparing the costs of this to the costs of exchanging this traffic with transit providers. Figure 7.9 illustrate for the top 30 ASes whether they peer or not at 10 large IXPs. This provides a guideline to the IXPs where top contributing ASes are present. From the figure, we observe that the top contributor AS (top row) is present at many other IXPs, and the third contributor is present at both LINX and AMS-IX. There should be no reason for one of these ASes to not peer with RedIRIS at other IXPs, as a peering relationship already exists. Finally, Figure 7.10 shows the potential traffic exchanged to individual IXPs in case the connection with the ESpanix fails. The comparable traffic levels with most of these IXPs are explained by the presence of the top contributor AS in many of them.

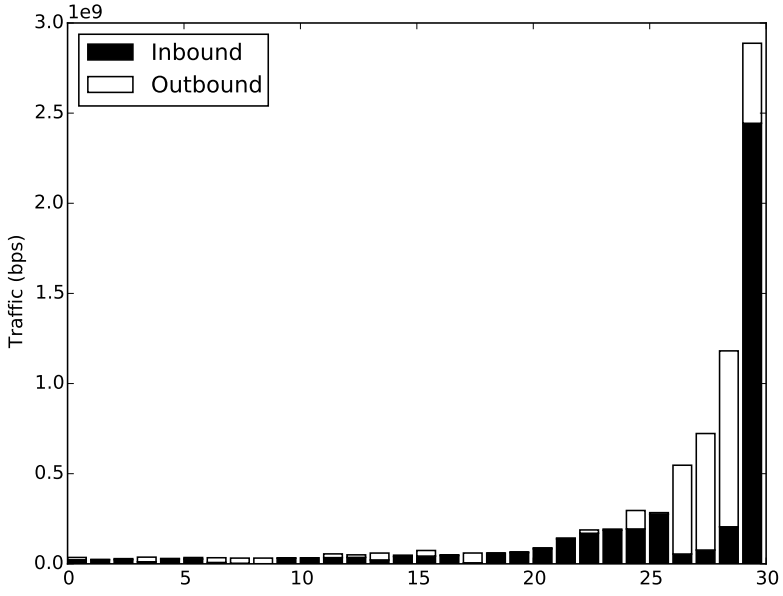


Figure 7.8: Top ASes contributing to IXP peering traffic.

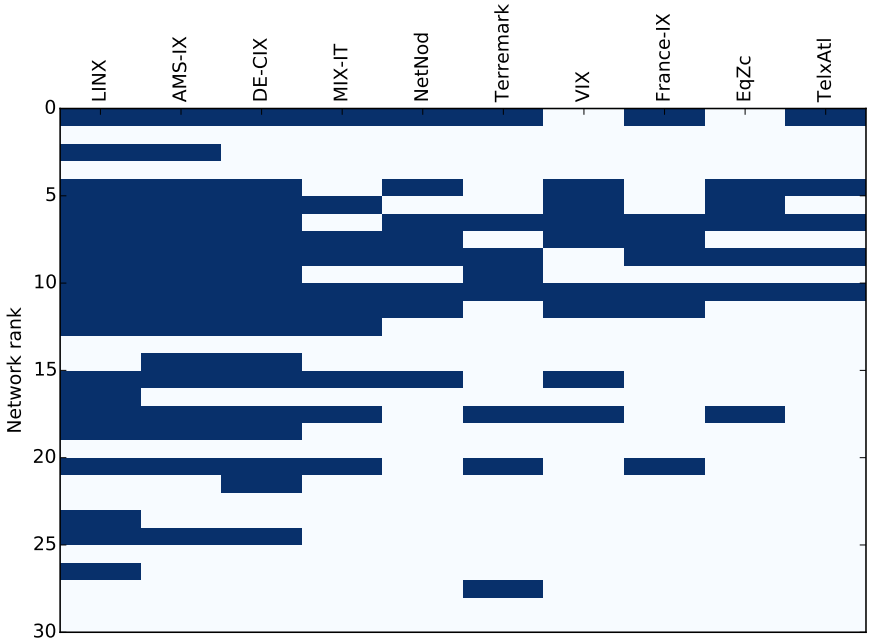


Figure 7.9: IXP membership for top contributing peering ASes.

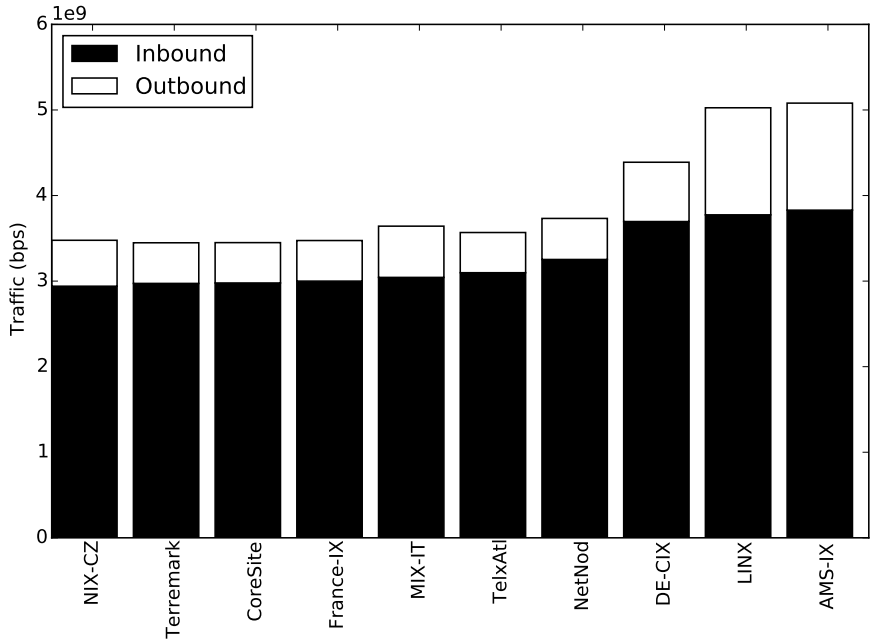


Figure 7.10: Potential back-up traffic after failure of main IXP.

7.4. Related Work

Peering selection and optimization. The data required to perform peering studies is difficult to obtain, therefore, previous work has focus on creating analytic frameworks that offer generic conclusions. [87] formulates the peering study as an optimization problem determining a limited set of ASes to which the network should peer and the peering agreement type used to do it. [121] studies the complications involving peering selection due to the inaccurate traffic prediction, involved pricing structure, and optimization complexity. The work in this chapter relaxes some of these assumptions to provide guidelines to operators and peering coordinators.

Peering modeling. Several authors have worked toward modeling the peering strategies of ASes in the Internet [42] [128]. Paid-peering agreements have received special attention from researchers, which have built frameworks to quantify the value involved in them [58], or analyze models for revenue distribution [128].

7.5. Summary

This chapter presents a peering study using real network data. This analysis can serve as a guideline for tools that operators can use to estimate the benefits of direct connections with other ASes, or peering at multiple IXPs. The main study uses transit off-load potential to measure the advantages of peering. We also discuss other benefits of peering, and provide a discussion on how operators can do to quantify them.

Using data to assist Peering management. A large part of the peering management process is social-centric. Operator and peering coordinators engage frequently with their peers at other companies to informally set up settlement-free peering agreements [192]. Nonetheless, business-intelligence and planning type of tools reporting information similar to the one exposed in this chapter, can aid network managements to focus their efforts to the IXPs and companies that provide larger returning benefits. The potential transit traffic off-loading offers a direct way of quantifying the savings. Network resiliency or internal transport offer additional benefits. Managers can use business impact analysis, or detailed bit-mile calculations to quantity these advantages and justify network expansions.

External policies should be assumed, validated later. Peering study tools must assume the policies of external ASes, reflected in the prefixes that they advertise, and the traffic that can be attracted from them and their customers. As seen in Chapter 6, the policies of ASes are dynamic and hard to estimate. After the actual peering sessions are established, the exchanged traffic might not be exactly as network managers planned. Operators and peering coordinators could use more insight information to tune the estimation process in order to avoid bad surprises. They can, for instance, obtain a priory the prefixes advertised in IXP route server, or use social contact to inquire the conditions of potential peering agreements. In Chapter 8, we describe several techniques that operators can use to validate the policy of external networks after the peering links have been provisioned.

Chapter 8

Detecting inter-domain policy conflicts

The inter-domain traffic of a network depends not only on the policy imposed by network operators, but also on its interaction with the policies of external ASes. In some cases, the interplay of policies of various ASes can lead to traffic distributions that do not satisfy the business interest of a network. In this chapter, we explain how these unsatisfied interests can occur, even when external ASes do not intend to cause them. Also, we describe a set of techniques that operators can employ in the validation steps of the management process (Chapter 6) to detect these cases, and use data from two European networks to test them.

Due to the nature of inter-domain routing and the lack of global coordination between ASes, the interest (or policies) of ASes may be incompatible. Consider, for instance, Fig. 8.1. In this example, $AS4$ has an interest to receive incoming traffic destined to its prefix $1/8$ from $AS2$. However, $AS3$ prefers to send traffic directly to $AS4$, and $AS1$ favors the link to $AS3$ to forward this traffic. Unfortunately, those interests are actually *incompatible*, that is, no valid distribution of traffic will realize the interests of all the ASes involved [26]. In this scenario, depending on the specific policies configured by $AS4$, $AS3$ and $AS1$, we have three possible cases, which we detail in Figure 8.2. In the first case (shown in Fig. 8.2(a)), $AS1$ forwards traffic to $AS2$. This configuration realizes the interests of $AS4$ but neither those of $AS1$ nor of $AS3$. In the second case (illustrated in Fig. 8.2(b)), $AS1$ forwards traffic to $AS3$, which subsequently transmits it to $AS2$, thus sacrificing its own interests. In the last case (see Fig. 8.2(c)), both $AS1$ and $AS3$ realize their respective interests, at the expense of $AS4$.

The interests and policy interactions of the example from Figure 8.1 can realistically occur in the Internet, if (I) AS4 is a customer of AS3 and AS2, (II) AS1, AS3, and AS4 are all customers of AS2, and (III) AS1 and AS3 are settlement-free peers. Also, the scenarios depicted in Figure 8.2 reflect policy decisions and router configurations that can be found in operational networks [28] [113]. BGP does not provide any guarantee in the presence of incompatible interests. Actually, incompatible interests may result in so-called policy disputes, which can trigger routing (i.e., control-plane instabilities) and forwarding anomalies (i.e., inter-domain forwarding loops) [95]. Policy disputes have been the target of numerous research efforts, covering the full range between

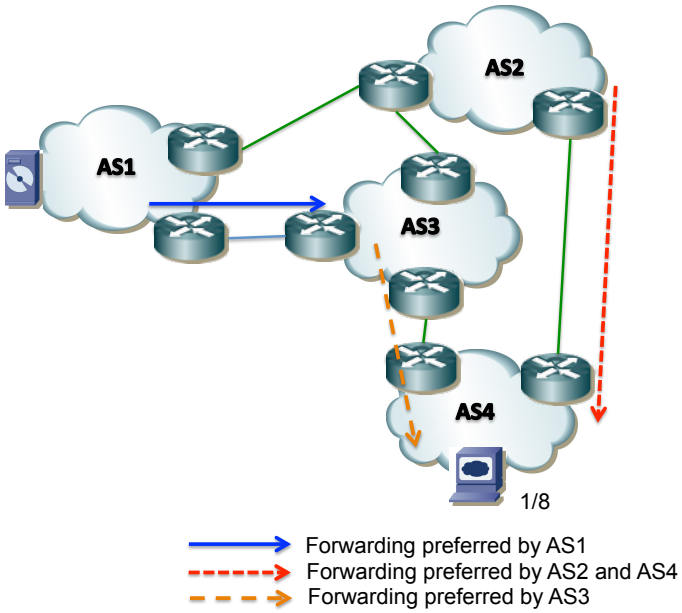


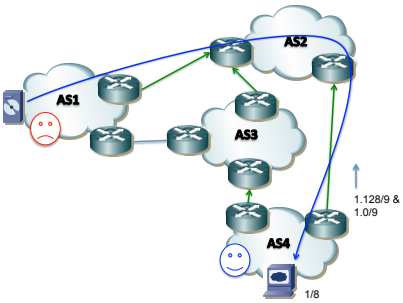
Figure 8.1: Network topology and inter-domain traffic interests for four different ASes.

theoretical (e.g., [53, 159, 173]) and practical (e.g., [54, 81]) contributions. Nevertheless, much less work has been done on the class of incompatible policies that are anomaly-free (as illustrated in Fig. 8.2).

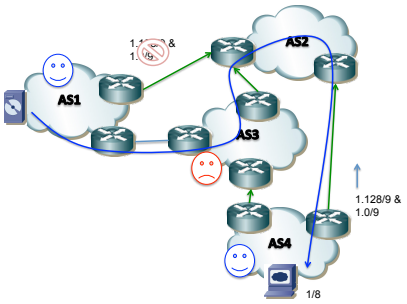
In this work, we complement previous research by focusing on incompatible policies that do not result in routing or forwarding anomalies. Incompatible policies can trigger traffic flows not respecting the original interest (preference) of one or more ASes. We show how these *unsatisfied interests*, as we denominate them, can theoretically and practically lead to *significant economic losses* for individual ASes.

For example, in Fig. 8.2(a), $AS1$ may be forced to pay for the traffic forwarded to $AS2$, while its (violated) interest to send traffic to $AS3$ could have led to no expenses. Similar economic losses can occur for $AS3$ and $AS4$, as explained Fig. 8.2(b) and 8.2(c), respectively. Such an economic impact, along with the impossibility to resolve incompatible interests in an automatic way (by definition), highlights the operational need for ASes to timely detect, understand, and assess unsatisfied interests. Unfortunately, current commercial and research tools do not address this need. In order to fill this gap, we make several contributions:

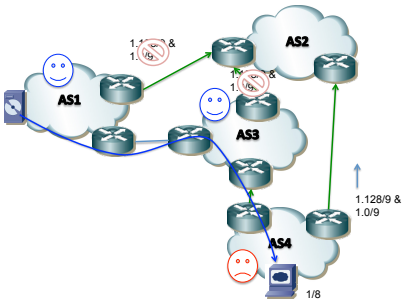
First, we provide a characterization of interest inconsistencies not leading to BGP anomalies (Section 8.1). In particular, taking the perspective of a single AS X, we distinguish between unsatisfied interests affecting outbound and inbound traffic. The former category is related to the BGP routes that neighboring ASes send to X. For example, if a single neighbor advertises a deaggregated prefix p on a specific inter-domain link with X, this will attract all traffic from X to p to that link, independently of the policies (and the interests) of X. Conversely, inbound unsatisfied interests are typically reflected in the policies applied by X's neighbors on routes



(a) $AS4$ performs selective advertisement with (more-specific) prefixes $1.0/9$ and $1.128/9$, in order to lead traffic heading to prefix $1/8$ to its link with $AS2$. This conflicts with the interest of $AS1$.



(b) $AS1$ filters the incoming prefixes $1.0/9$ and $1.128/9$ from its transit provider ($AS2$), and sends traffic to its peer $AS3$ using covering prefix $1/8$. $AS3$, however, is affected, as it still has routes based on the more-specific prefixes $1.0/9$ and $1.128/9$.



(c) $AS3$ could decide to filter the more-specific prefixes $1.0/9$ and $1.128/9$. This policy is incompatible with the one of its customer, $AS4$.

Figure 8.2: Traffic state and interest fulfillment for 3 different policy configurations.

propagated by X. For instance, filters and partial modifications applied by neighbors to X's routes can change the ingress points of traffic traversing X.

Second, we develop algorithms to (I) detect outbound and inbound unsatisfied interests; and (II) measure their impact, especially in terms of affected traffic volumes (Section 8.2). Our algorithms exploit the variety of data sources on the Internet traffic, typically available to network operators. In particular, we show ways to detect unsatisfied interests by combining multiple BGP views (internal and external to the given AS) and traffic data.

Third, we build upon our algorithms to design and prototype a warning system, which raises alarms for the most critical unsatisfied interests (Section 8.3). Our warning system is intended to support network managers in their strategic business decisions, and to contact their counterparts in other ASes (e.g., to modify business agreements). Since unsatisfied interests may coincide with unfulfilled peering contracts (possibly due to misconfigurations), our system also offers technical support for verification of commercial agreements.

Fourth, we leverage our warning system to measure unsatisfied interests and their impact

on traffic in a European Tier-2 and a national academic network (Section 8.4). This original measurement campaign shows (I) the effectiveness of our algorithms to detect business-affecting unsatisfied interests; (II) the practical feasibility of our system; and (III) the unexpectedly high frequency and relevance (e.g., in terms of impacted traffic) of unsatisfied interests in operative networks.

8.1. Classification of unsatisfied interests

In this section, we study unsatisfied interests of single ASes and their impact. We define AS interests in terms of profit and costs as established by commercial agreements. This definition reflects long-term policies set by the given AS and their effectiveness in the stable routing state. The analysis in this section can however be extended to shorter-term interests by including the consideration of transient network conditions in the definition of interests. For example, interests may encompass specific neighbor preferences (prefer A over B) - or no preferences at all - for given flows, in the presence of given failures (C is not available) or congestion.

We classify unsatisfied interest into outbound and inbound, depending on the traffic that they affect, and describe realistic examples for each class. Our examples also show that (i) our classification covers all types (inbound, outbound and transit) of traffic traversing the considered AS; and (ii) unsatisfied interests may be due to various economic reasons, and be realized via different technical means.

8.1.1. Outbound unsatisfied interests

We define outbound unsatisfied interests, or outbound dissatisfactions, as follows. An AS X suffers from an outbound dissatisfaction if X is prevented from sending some traffic flows through an intended inter-domain link. That is, BGP forces the traffic to a given destination to exit X via an inter-domain link l_1 while X has interest to send it to $l_2 \neq l_1$.

Figure 8.3 shows a simple scenario in which $AS1$ is affected by an outbound dissatisfaction, as a result of the incompatible interest between $AS1$ and $AS2$. The two ASes are connected in different locations, over multiple physical links. However, they disagree on which inter-domain link should be used for the traffic from the source S to the destination prefix $1/8$. The dashed (red) and the solid (blue) arrows respectively indicate that $AS1$ would like to forward such traffic through $R1b$, while $AS2$ prefers to receive this traffic at $R2a$.

Disagreements like the one in Figure 8.3 realistically happen in the Internet [72]. For instance, if S is closer to $R1b$ and the machines hosting prefix $1/8$ are closer to $R2a$. In this case, the incompatible interest is the result of the adoption of the *hot-potato* policy from both $AS1$ and $AS2$, according to which ASes tries to reduce the internal path followed by Internet traffic (e.g., for resource consumption minimization).

In the example, the inter-domain link preferred by $AS2$ is eventually used, forcing $AS1$ to send outbound traffic against its economic interests. This unsatisfied interest is due to $AS2$ se-

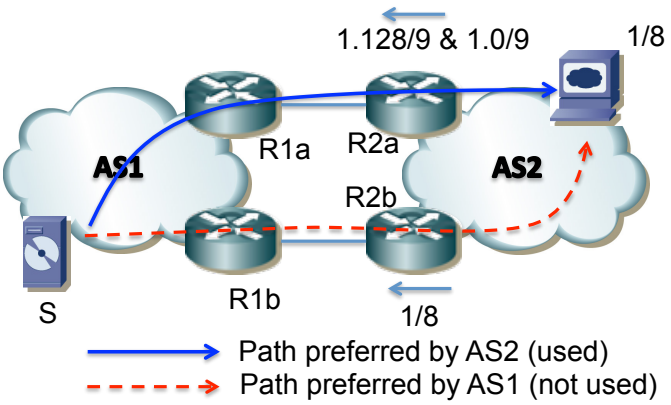


Figure 8.3: Incompatible inter-domain traffic interests resulting in an outbound dissatisfaction for $AS1$.

lectively announcing paths to more specific prefixes, on the $(R1a, R2a)$ link. Note that $AS2$ has plenty of ways to (try to) enforce its interest, e.g., it can also set different BGP attributes (e.g., AS-path or MED) in the announcements that it propagates on the two inter-domain links [152]. All those cases can be categorized as "inconsistent advertisements", and are traditionally considered a bad practice when applied to private peerings [139] [72], as they typically violate business agreements formalized by the contracts. With the proliferation of peerings using IXP route servers [156], the peering ecosystem has however become more informal, increasing the likelihood of occurrence of such situations.

Of course, $AS1$ can take actions aimed at preventing the outbound unsatisfied interest from occurring. For example, with respect to the example in Figure 8.3, $AS1$ can filter the more specific prefix announced by $AS2$ on the $(R1a, R2a)$ link. However, the actions that $AS1$ can take to avoid the outbound dissatisfaction depend on the specific techniques used by $AS2$ to attract the traffic on its preferred link. Moreover, the safety (from an inter-domain routing viewpoint) and the legality of $AS1$'s reaction may depend on specific circumstances, e.g., specifics of the business agreement between $AS1$ and $AS2$. For this reason, we consider the technical solutions to react to an unsatisfied interest outside the scope of this work.

Inconsistent advertisements between two neighboring ASes only represent a simple example of outbound dissatisfaction. $AS1$ in Figure 8.2(a), for instance, also suffers from an outbound dissatisfaction, since the more specific prefixes force it to send traffic to $AS2$ instead of $AS3$. In that case, the conflict is with $AS4$, a remote AS. Moreover, observe that outbound unsatisfied interests do not necessarily affect outbound traffic only. In particular, they can also affect transit traffic, that is, traffic which the considered AS did not originate but has to transfer from one of its neighbor to another. As an illustration, the unsatisfied interest to which $AS3$ is subject for the transit traffic from $AS1$ to $AS4$ in Figure 8.2(b) is an outbound one. In fact, in that case, $AS3$ is forced by the global inter-domain configuration to send such traffic to $AS2$ while it has an economical interest to send it directly to $AS4$.

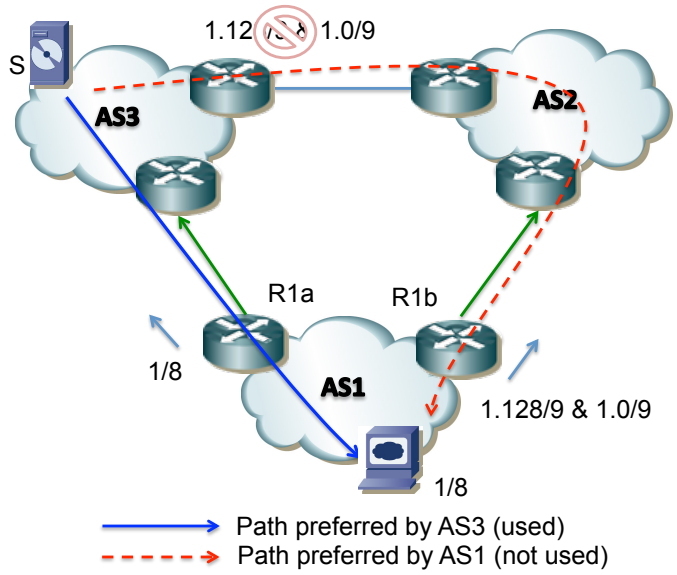


Figure 8.4: Incompatible interest resulting in an inbound dissatisfaction for $AS1$.

8.1.2. Inbound unsatisfied interests

We define inbound unsatisfied interests, or inbound dissatisfactions, as complementary to the outbound ones. In particular, we say that an AS X is subject to an *inbound dissatisfaction* if X is prevented from *receiving* certain traffic over a given inter-domain link. That is, BGP forces the traffic to a given destination to enter X from an inter-domain link l_1 while X has interest to receive it to $l_2 \neq l_1$.

An example of inbound dissatisfaction is displayed in Figure 8.4. As in the previous example, we take the perspective of $AS1$, and consider the traffic to prefix $1/8$ in $AS1$. For this traffic, $AS1$ has an economic advantage in receiving it at $R1b$ (solid blue path). However, this clashes with the interest of $AS3$ to send such traffic directly to $AS1$ (dashed red path in the figure). Note that, as the previous one, this example is also realistic. On one hand, the path preferences of $AS1$ can stem from an economic advantage for it to balance incoming inter-domain traffic load between its two border routers (i.e., depending on the destination prefix). On the other hand, the interests of $AS3$ are likely as illustrated in the figure, if $AS3$ is a service provider of $AS1$ (getting money for the traffic exchanged on their direct inter-domain link) and is a settlement-free peer of $AS2$ (with free of charge traffic exchange agreement).

In Figure 8.4, the incompatible interests eventually leads to selecting the direct path between $AS3$ and $AS1$. Indeed, after noticing that a less specific route is received on its preferred link, $AS3$ discards (e.g., filters) the more specific advertisement received from $AS2$. While implicitly ignoring the preferences of $AS1$, $AS3$ may indeed not be forced to consider all $AS1$'s announcements, e.g., by contractual obligations.

Observe that $AS1$ can still try to pursue its interest, by withdrawing the less specific route propagated to $AS3$ and announce only two disjoint prefixes. However, as a side effect, such a

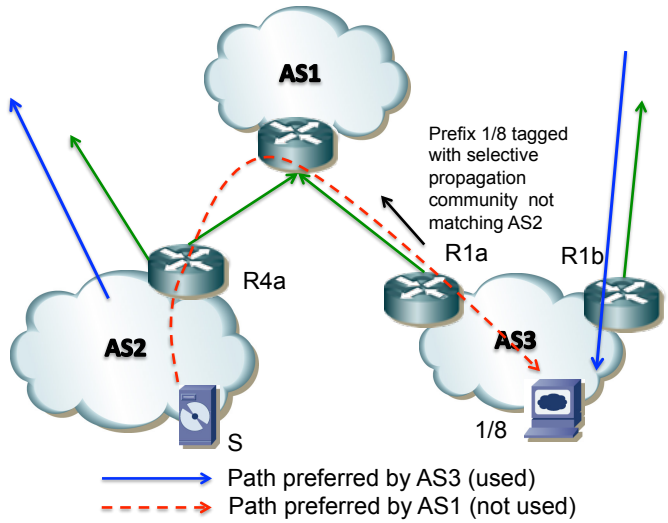


Figure 8.5: Incompatible interests between two customers of the same provider.

choice may lead to a less robust configuration, in which $AS3$ has no available routes to forward traffic towards $1/8$ if the inter-domain link between $AS1$ and $AS2$ fails. An alternative for $AS1$ would be to revise or renegotiate the contract with $AS3$.

Finally, note that transit traffic can also be subject to inbound dissatisfactions. Consider the example depicted in Figure 8.5. $AS3$ has interest to receive the traffic from source S in $AS2$ to a given destination $1/8$ using router $R1b$. Nevertheless, $AS1$ may have economic benefits in forwarding the traffic from $AS2$ to $AS3$, e.g., if both $AS2$ and $AS3$ are customers paying for transit through $AS1$. In the example, $AS3$ indicates into the BGP announcement to $AS1$ (e.g., with pre-agreed communities [60]) that $AS1$ should not propagate the announcement to $AS2$. Since $AS1$ may be forced by contractual agreements respect such an indication, it ends up not propagating to $AS2$ the announcement from $AS3$, which results into an incoming dissatisfaction at $AS1$.

8.2. Detection of unsatisfied interests

Unsatisfied interests can have significant detrimental effects on the inter-domain traffic of a network. Yet, operators typically do not check for them. One of the main reasons why unsatisfied interests are rarely monitored is that they are hard to detect by manual inspection. Indeed, discovering them requires mining and correlating huge amounts of data coming from different data sources, like BGP configurations, control-plane messages, and data-plane traffic statistics.

In this section, we propose algorithms to detect unsatisfied interests and assess their economical and technical impact. We designed those algorithms to have the following features.

We designed different algorithms to exploit peculiarities of inbound and outbound dissatisfactions. Unsatisfied interests affecting outbound and inbound traffic are intrinsically differ-

ent. We provide distinct detection algorithms that match each of the two cases. For the outbound case, we can rely on control-plane information only. Indeed, knowing the received route and decision process used by routers, we can compare the received routes with the missing ones to detect outbound dissatisfactions, infer their cause, and estimate their impact by analyzing what-if scenarios. For the inbound case, in contrast, we cannot assume that policies of external ASes are known. Hence, rather than relying on control-plane messages, we use data-plane information to detect unexpected ingress points for given flows.

Our algorithms estimate the impact of detected unsatisfied interests. The gravity of an unsatisfied interests generally depends on the traffic affected by it. Informal conversation with network operators confirmed that unsatisfied interests affecting certain traffic flows (e.g., to popular or critical destinations) would be more important than those affecting others. Moreover, unsatisfied interests taking place after link failures may be considered less problematic than those occurring during normal network operation. Our algorithms include metrics to estimate such an impact for each detected dissatisfaction. To this end, we correlate the prefixes affected by unsatisfied interests with the traffic destined to them. As a positive side effect, this allows us to pinpoint the most practically relevant unsatisfied interests. Moreover, we classify unsatisfied interests based on their qualitative impact (reduction of route diversity, possible occurrence in the case of failure, etc).

Our algorithms can be customized according to specific needs of operators. This includes both new unsatisfied interests types and impact estimation. Prominently, the algorithms can be easily extended to include more or different metrics on the impact of unsatisfied interests. In addition, while we focus on the standard BGP implementation [155], our algorithms can be slightly modified to take into account different inter-domain routing protocols or implementations. For example, in the following, we assume that routers run the standard BGP decision process [73], however our algorithms can be easily adapted to modified versions of such decision process, as those used to implement routing policies on iBGP sessions [184].

We now detail algorithms for outbound and inbound dissatisfactions in Section 8.2.1 and 8.2.2, respectively.

8.2.1. Detection of Outbound unsatisfied interests

For a given AS, the exit points of inter-domain traffic flows depend on the routes received from neighboring ASes; the selected routes among those ones; and the intra-domain routing (i.e., IGP or iBGP) configuration. Operators have control of the two latter parameters, and the neighboring ASes only influence the announced paths.

To detect outbound unsatisfied interests, we designed an algorithm that compares received routes with the set of missing ones, i.e., those expected but not received. In this comparison, we automatically assess *whether* and *how much* the traffic would differ if the missing paths were announced to the network. For instance, if a network detects that is not receiving a route from a settlement-free peer, the algorithm checks how the traffic of the prefix is currently being routed.

Algorithm 1: Detection of outbound unsatisfied interests.

input : **1.** Missing paths ($MissingPaths$)
2. Current BGP paths per prefix ($CurrentBestPath$ function),
3. Outbound Traffic per Prefix ($OutboundTrafficDemand$).
4. Preference of AS ($ASPreference$ function)

output: For each missing path returns the Impact type(s) (contained in $CurrentLabels$) and Impact level ($ImpactMetric$).

/* Go over each missing path and analyze its impact. Each Missing path is composed by an NLRI ($NLRI$) and a set of path attributes (MP). */
[1] **for** $(MP, NLRI) \in MissingPaths$ **do**
  /* Store the current best path (CBP) and backup path ($CBaP$) for $NLRI$. */
[2] $CBP = CurrentBestPath(NLRI)$;
[3] $CBaP = CurrentBackupPaths(NLRI)$;
  /* Calculate the best paths if the missing path (MP) were received for $NLRI$. Store this best path in NBP. */
[4] $NBP = BGPBestPathAlgorithm(CBP \cup MP)$;
  /* 1) Detection of unsatisfied interests: Perform the classification tests and apply labels. The next part can be modified to fit the requirements of each operator. */
[5] $CurrentLabels = \emptyset$;
  /* If the preference of any AS in NBP ($NewPreference$) is higher than the preference of the current path ($CurrentPreference$), apply label $NeighbourPreferenceImprovement$. */
[6] $NewPreference = Max(\{ASPreference(AS) \mid AS \in GetNeighboringASes(NBP)\})$;
[7] $CurrentPreference = Max(\{ASPreference(AS) \mid AS \in GetNeighboringASes(CBP)\})$;
[8] **if** $NewPreference > CurrentPreference$ **then**
[9] $CurrentLabels = CurrentLabels \cup \{NeighbourPreferenceImprovement\}$;
  /* If the current NHs ($CurrentNHs$) is a strict subset of the new NHs ($NewNHs$) under the missing paths, apply label $IncreaseNHDiversity$. */
[10] $CurrentNHs = GetNHs(CBP)$;
[11] $NewNHs = GetNHs(NBP)$;
[12] **if** $NewNHs \supsetneq CurrentNHs$ **then**
[13] $CurrentLabels = CurrentLabels \cup \{IncreaseNHDiversity\}$;
  /* If there is currently a single active path for the prefix, and the missing path improves the preference of the back-up AS, apply label $IncPrefofBKforSingleActivePath$. */
[14] **if** $|CBP| == 1$ **then**
[15] $NewBackupPaths = BGPBestPathAlgorithm(CBaP \cup MP)$;
[16] $CurrentBKPreference = Max(\{ASPreference(AS) \mid AS \in GetNeighboringASes(CBaP)\})$;
[17] $NewPreference = Max(\{ASPreference(AS) \mid AS \in GetNeighboringASes(NewBackupPaths)\})$;
[18] **if** $NewPreference > CurrentBKPreference$ **then**
[19] $CurrentLabels = CurrentLabels \cup \{IncPrefofBKforSingleActivePath\}$;
  /* If we find a path ($CoveringP$, $CoveringNLRI$), in which $CoveringNLRI$ covers the $NLRI$, and if ($CoveringP$, $CoveringNLRI$) is propagated to other non-customer ASes, apply label $UnexpectedTransit$. */
[20] **if** *($\exists$ ($CoveringP$, $CoveringNLRI$) which:*
[21] *$CoveringNLRI$ Covers $NLRI$ **and***
[22] $IsPropagatedToNonCustomerneighbors(CoveringP)$ **then**
[23] $CurrentLabels = CurrentLabels \cup \{UnexpectedTransit\}$;
  /*
  /* 2) Impact assessment: Assess impact of interest conflict for the paths that match at least one classification. */
[24] **if** *$CurrentLabels$ is not $\emptyset$* **then**
[25] $ImpactMetric = OutboundTrafficDemand(NLRI_P)$;
[26] **Register** (P, $NLRI$) with *labels $CurrentLabels$*
[27] and *Impact $ImpactMetric$*

If the traffic is currently being routed through a transit provider, and its volume is significant, the algorithm would detect this case and rank it high. In a case of inconsistent advertisement (Figure 8.3), the algorithm checks whether the inconsistency of the neighbor is reducing the next-hop diversity of the network, and rank it based on the outbound volume of these prefixes.

The algorithm is summarized in Algorithm 1. Its input consists of (1) the set of all current best routes; (2) the list of missing routes (expected but not received from neighboring ASes); and (3) statistics of outbound traffic per prefix. The last input is used to assess the impact on the network of the missing paths. We discuss several methods that operators can use to collect these inputs in Section 8.3.

The algorithm is based on two macro-steps, explained hereafter.

8.2.1.1. Detection of unsatisfied interests

The purpose of the first part of the algorithm is to compare how the traffic distribution of the network would improve, if the missing paths where actually received. In this part, we classify missing paths into different categories, each representing different types of network impacts (a missing path could be in none or even in more than one category). Based on private conversations with operators, we identified four main effects of unsatisfied interests on outbound traffic. We detail them in the following. Anyway, note that the additional categories can be added to our algorithm by applying additional comparisons between missing and received paths.

- **Neighbor preference dissatisfaction**. A missing path is added to this category if its announcement would lead to the selection of a more preferred neighbor. In Algorithm 1, we use function $ASPreference$ to compare the preference of the operators among neighboring ASes. In simple implementations, however, the comparison can be easily performed by checking the Local Preference that operators normally consistently apply to each neighbor [91]. If the operator decides, this same procedure could be extended to cover *links* instead of neighboring ASes.

- **Next-hop diversity dissatisfaction**. Missing paths are added into this category if the exit point of the missing path is equally preferred to the current best paths, but it increases the number of next-hops (NHs) available to outbound traffic. This is important for operators who want to bring traffic balance to their network, by providing a large diversity of exit points (this case is important when the operator connects to various IXPs). Note that inconsistencies advertisements are a subset of this case, however, an operator might prefer to classify them in their separate category.

- **Back-up path dissatisfaction**. Missing paths are added into this category if the missing path comes from a neighbor more preferred than the one of any of the current back-up paths, when there is a single active path. The idea with this category is to cover cases in which a single link failure would let traffic be sent to less preferred neighbors.

- **Unexpected transit dissatisfaction**. Missing paths are added into this category if they are generating transit flows between two non-customer ASes (unexpected transit flows). This is a special case of *Neighbor preference dissatisfaction* paths, and is the problem experienced by AS3 in Figure 8.2(b). Network operators must avoid transporting traffic between non-customer neighboring ASes, as this does not provide any economical benefit [86]. To do this, operators do not advertise routes coming from non-customer neighbors, to other non-customer neighbors. However, an operator might receive a route to a prefix p from a customer, which it propagates to neighboring ASes, while, it receives from non-customer neighbors a route for a prefix p', more specific than p. Since routers in the network forward packets based on the more specific prefix (p'), the network might start transiting traffic between non-customer neighbors [28]. The missing routes from the customer towards the more-specific prefixes (p') are the ones added to this category.

8.2.1.2. Impact assessment

The final step in the algorithm is to measure the level of impact of each missing path. This value depends on the amount of outbound traffic and on its classification. Each operator could select their own impact metrics based on their needs. We follow a basic approach and use the outbound traffic demand of the each prefix in the peak hour of the network. Operators could also employ more complex metrics such as bit-mile calculations, or metrics that estimate the potential revenue reduction due to the missing path. We decided to not implement such metric as setting its parameters is difficult to achieve from a researcher point of view.

The missing path, the categories, and impact value for each discovered case become descriptive features of the unsatisfied interests. These features can be used by the alarm system (Section 8.3) to highlight cases that should be analyzed individually by operators.
Note that, since the algorithm is based on re-simulating the BGP decision process, it always correctly provides outbound policy dissatisfactions, provided that the input is correct. In Section 8.3, we discuss different approaches to collect such input data in real networks.

8.2.2. Detection of inbound unsatisfied interests

The distribution of inbound inter-domain traffic into a network depends on the paths announced and the policies of the other ASes. Since each AS independently sets their own policies, an operator looking to balance their inbound traffic can only try to persuade the selection of others by tweaking the paths that they announce. The algorithm in this section aims at detecting neighboring ASes whose policies work against those tweakings. In Figure 8.2(c), for example, the algorithm, running at $AS4$, would identify the traffic for prefixes 1.128/9 and 1.0/9 coming from $AS3$, when this is not the intention of $AS4$.

Since operators rarely know the policies of external AS, and considering that they can be very complex, it is almost unfeasible to analyze the inbound unsatisfied interests by using control

plane data available from external looking glasses. Therefore, we rely only in local data plane information to detect these cases. This type of test is simpler than the one in Section 8.2.1, but it provides less information.

Algorithm 2: Detection of inbound unsatisfied interests.

input : **1.** Inbound flow ($InboundFlow$) , with attributes $InboundFlowAttributes$ (containing attributes such as $SourceIP$, $DestinationIP$, Bw over the peak hour, etc.) arriving over link L.
2. Inbound policy contained in a function $IsFlowUndesired$.
output: Returns the inbound flows that are conflicting with the policy of the operator, together with their impact level ($ImpactMetric$).

```
   /* For each inbound flows InF on each link L.                    */
[1]foreach Link L do
[2]   foreach InboundFlow, with attributes InboundFlowAttributes (SourceIP, DestinationIP, BW, etc.) do
[3]      if IsFlowUndesired(InboundFlowAttributes, L, Bw) returns True then
[4]         Register InboundFlow, L, Bw;
```

Algorithm 2 describes our method to detect inbound traffic dissatisfactions. Shortly explained, the algorithm searches for inbound traffic that, based on the policy of the operator, should not be received on each link. We provide in the rest of this section a extended description of the algorithm.

8.2.2.1. Input

To identify the undesired traffic, we need to know its characteristics in terms of origin AS, origin prefix, or destination prefix. Therefore, the traffic description must be grouped under these characteristics. In other words, traffic data must be divided in *traffic flows*. Based on this, the algorithm takes two basic inputs:

- **Inbound Inter-domain policy**. For our algorithm, the policy is defined as a function $IsFlowUndesired$, which based on the attributes of inbound flows, defines whether the flow should enter the network through the link L. Operators can build this policy using automatic or manual methods. In simple scenarios, the function looks for source prefixes of ASes for which inbound traffic should not be detected. For example, operators usually do not expect traffic from customers or peering ASes in transit providers links, or traffic from ASes to which propagation is being remotely filtered through special communities [60]. More involved examples could lead the function $IsFlowUndesired$ to calculate complex metrics, such as the bit-mile of the flow [150], in order to define whether the flow is undesired or not.

- **Inbound traffic statistics**. The algorithm requires disaggregated information of the inbound traffic of the network. Specifically, operators should provide statistics of inbound traffic flows per prefix for individual ingress links of the network.

8.2.2.2. Undesired flow detection

The algorithm cycles over each inbound traffic flow received over each external link. The function $IsFlowUndesired$ is then used to check whether the flow should be present in the link or not, based on the policy of the operator. In a simple implementation, the function would, for instance, check whether the origin AS of the flow is not connected through a more preferred peer. An example of this case is when traffic from a peering content provider is coming through the link of a transit provider. Operators implementing more complex traffic engineering would perform more checks. For instance, in cases in which they are doing prefix deaggregated, they would check if there are inbound flows to prefixes with more specific masks announced in links different from the one being analyzed.

8.2.2.3. Impact assessment

After an inbound unsatisfied interest is detected, the algorithm assesses the impact of the flow in the network. We follow a similar approach to the outbound case: we account for the actual traffic of the inbound flow in the peak hour of the network. We stress again that operators could use more complex metrics such as bit-mile calculations, for this end.

Finally, the event is stored using the tuple ($externallink$, $Flowattributes$, $amountofinboundtrafficoftheflow$). Operators can query this information to isolate the more important cases and study them in detail.

8.2.2.4. Final remarks

Similar to the outbound case, the inbound unsatisfied interest detection algorithm is correct, given that the input of the system is also correct. Since this algorithm only uses data-plane information, operators should probably gather more information before actually judging whether the problem lies outside the AS. For instance, a routing leak [107] or a router misconfiguration could generate an invalid flow. Although the algorithm could detect these problems, we stress that its philosophy is to relay in other mechanism, such as [54] [70] to isolate these cases and deal with them. The algorithm should just tackle the cases of real inbound unsatisfied interests in which external ASes are not following the actual announcements performed by the network.

8.3. System architecture

In this section, we describe our inter-domain unsatisfied interests warning system. This system is designed with the purpose of detecting, ranking, and creating alerts for inter-domain unsatisfied interests. We first describe the general architecture of the system, and then discuss each of its components.

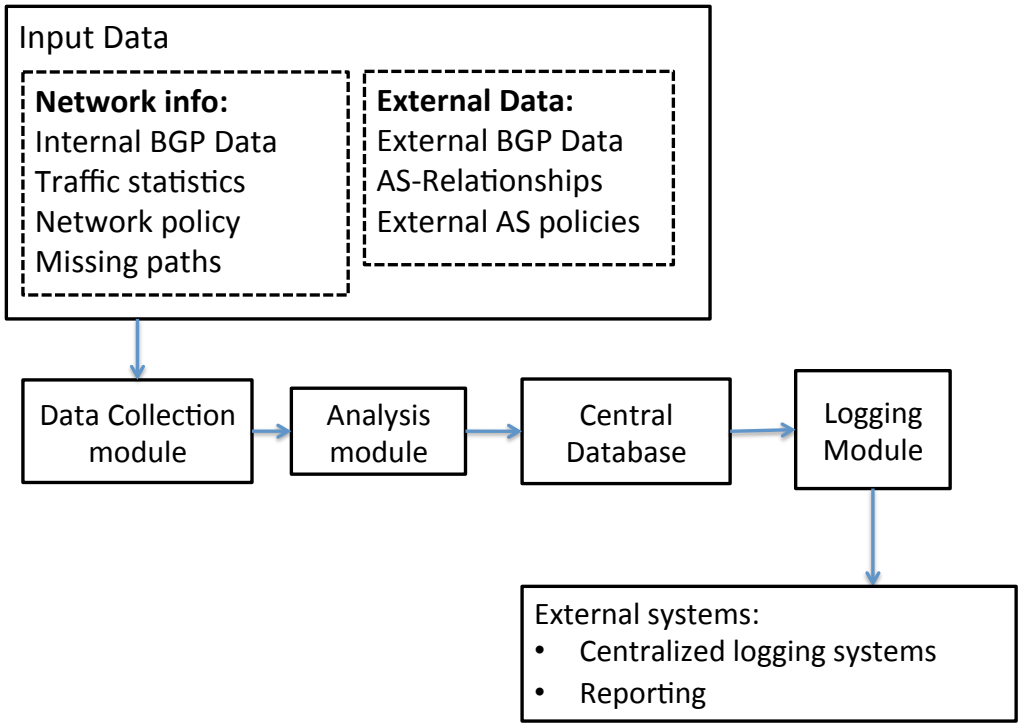


Figure 8.6: Architecture of our warning system.

8.3.1. General Architecture

Our warning system relies on four basic modules, as depicted in Figure 8.6. A *data collection module* gathers the required *input data* by interfacing with internal network devices or data collectors. The *logging module* communicates with other monitoring or management systems, e.g., triggering warnings or outputting data in a convenient format. Finally, the *analysis and central database modules* implement the logic to detect, rank, and store the unsatisfied interests warnings. Following the inter-domain framework presented in Chapter 6, the data collection module will be part of the *data collection* procedures of the framework, and the rest of the modules would fall into the *validation* and *simulation* procedures.

By relying on those four modules, this architecture decouples the implementation of unsatisfied interest detection algorithms (analysis module) from external interfaces, hence facilitating algorithm implementation. Moreover, the system is easy to adapt to different networks by changing the external modules (data collection and logging) to specific requirements.

We now provide more details on each module.

8.3.1.1. Data collection module

This module provides a standard interface for our system with respect to input data sources. Abstractly, it isolates the complexity of all those interfaces, thus simplifying the implementation

of the rest of the modules.

A key ability for detecting unsatisfied interests consists in correlating different types of data. These data can only be fetched by interacting with multiple monitoring systems (e.g. BGP collectors, network controllers, traffic monitoring, routers), and protocols (e.g. JSON, XML, CSV, etc.). We describe possible data extraction methods in Sec. 8.3.2.

8.3.1.2. Analysis module

The analysis module is the heart of the system. This module implements the algorithms described in Section 8.2 using the network data obtained via the data collection module. Operators can tune the analysis module's parameters to fit the behavior of the algorithms to their needs (e.g. incident classification, frequency of operation, etc.). Observe that our algorithms can be easily parallelized. For example, different missing routes in the outbound unsatisfied interest algorithm as well as traffic flows in the inbound unsatisfied interest algorithm can be processed in parallel, since their analysis does not required any share information.

8.3.1.3. Central Database

The central database stores the output of the unsatisfied interest detection algorithms. As discussed in Section 8.2, this output contains fine-grained attributes to generate detailed unsatisfied interest reports. In particular, for every detected unsatisfied interest, it includes its class (inbound or outbound), its impact according to the implemented metrics, the category to which it belongs (neighbor preference dissatisfaction, next-hop diversity dissatisfaction, etc.), and additional information (for example, attributes of the missing route in the case of outbound unsatisfied interest).

8.3.1.4. Logging Module

The main purpose of this module is to log any warnings obtained from the analysis. The logging information could be used directly by operators, or can be sent to other management systems (e.g., a general warning system or an SDN controller) deployed in the network. The purpose of this component is to isolate the other modules of our architecture to external systems, and to translate the corresponding information into specific formats. This module could include the necessary code to generate reports for different network management roles, or could serve as an interface to external log management systems to perform this task. Operators can configure the information included in the logs generated by the system, including thresholds to generate alarms.

8.3.2. Implementation

We developed a Proof of Concept of the system described in the previous section using a server with 16 cores and 32GB of RAM. Python was used to implement the algorithms and the

logic of the *data collection* and *analysis* modules. We employ MySQL to store the data required for the system and to implement the *Central Database module*. The logging module generates summary files on CSV format that are later used to generate reports. The reports contained multiple figures that we plotted using the MatplotLib [102]. We will look at these reports in the next section.

The main challenge for the implementation of the system is the ability to collect the necessary **input data** required to run the algorithms. In Chapter 6 (Section 6.2), We discussed different methods that operators can use to gather most of these input. We extend this description here, for the specific case of the interest unsatisfied interest detector.

8.3.2.1. Traffic data

Traffic statistics are needed for both inbound and outbound unsatisfied interest detection. ISPs normally gather traffic statistics to perform different applications. We can leverage this data for the warning system. The system requires granular statistics of the inter-domain traffic, thus forcing the use of traffic flow type of collectors, such as those supporting *Netflow* and *sflow*.

8.3.2.2. Received BGP routes

One of the inputs of the outbound unsatisfied interest detection algorithm is represented by the BGP routes received by edge routers for every destination prefix (see Section 8.2.1). In order to perform a complete network analysis, all paths received from external neighbors need to be collected. We describe Several methods can be used for this purpose in Section 6.2.1. (I) the usage of custom scripts, e.g., based on router CLI commands and screen scraping; (II) the configuration of iBGP sessions with add-path [181] (or similar features to propagate all BGP routes) between edge routers and a route collector (such as [123]) [50]; (III) the usage of monitoring protocols like BMP [163] and (IV) the configuration of selective port mirroring on edge routers, as proposed in [185].

8.3.2.3. Intended internal policies

The inbound unsatisfied interest detection algorithm requires knowledge of network policies (see Section 8.2.2).

For *outbound traffic*, the algorithm needs the preference of operators for the traffic leaving the network. For many companies this could be calculated automatically by checking the default local-preference given to neighboring ASes. For cases in which this does not work, the preference could be given manually.

For *inbound traffic*, the algorithm needs to know the attributes of the traffic that should not arrive over specific inter-domain links. Different sources of information can be used to obtain this data automatically. The peering relationships of the network can be used to build a starting policy for unexpected or unwanted inbound traffic. In a typical set-up, for instance, an operator

does not want traffic from settlement-free peers, or its customers, on transit links. The peering relationships to neighboring AS can be obtained using router configurations, BGP data, or by fetching information from Internet Routing Registries (IRR), when available. In cases in which ASes are allowed to steer inbound traffic over links with the same AS, using BGP communities or MED, operators would like to check if their neighbor is respecting their commands. This information is reflected in the configuration of the routers.

8.3.2.4. Missing paths

The outbound unsatisfied interest algorithm takes missing paths as input. We recall that missing paths are those, which are supposed to be received, but are actually not received due to policies of external ASes (see Section 8.2.1). Our system currently focuses on two general and practically relevant classes of missing paths, i.e., inconsistent advertisements and incomplete sets of paths.

Inconsistent advertisements identify BGP messages that are ranked differently from the receiving AS, because of different attributes, although they refer to the same destination prefixes, and are received from the same neighboring AS (on different inter-domain links). Inconsistent advertisements frequently correspond to missing routes, since not all of them have the same preference contrary to what typically expected [40, 72, 139]. Note that inconsistent advertisements do not necessarily correspond to a contract violation, especially in cases in which the AS peers often to open IXP route servers [34]. In any case, they have an effect of the selected egress point for inter-domain traffic. For this reason, we consider them in the outbound unsatisfied interest detection algorithm. In particular, we gather inconsistent advertisements comparing the routes announced by each peer on different physical location, as in [72].

Incomplete sets of routes represent cases in which a neighboring AS does not announce routes to some prefixes while it was supposed to. Of course, determining incomplete sets of routes depends on operators' expectations. While expectations on the route set received from BGP neighbors can be case-specific, our system currently focuses on two basic expectations that are commonly shared by the large majority of operators [40]. Namely, we check that (I) transit providers announce routes to all destination prefixes, and (II) peers propagate all routes originated by the peer itself, and its customers routes (Note that our tool supports partial peering, in the sense that we can define the subset of customer routes that the ISP is expecting to receive). To perform those checks, for each neighboring AS X, we compare the routes received from X with those that X announces to other ASes, leveraging public BGP collectors, like Routeviews [133] and RIPE RIS [157] ones, and AS relationship datasets, like the one provided by CAIDA [23]. ISPs can also rely on their own data sources or on services from commercial companies [64] to acquire this data or complement public data sources. The estimation of the relationships of external ASes can be complicated, and not entirely reliable. As we described in the previous section, the output of our algorithm is correct, given that the input is also correct. Operators should always check whether an unsatisfied interest warning triggered by the system can be the result of a incorrect estimation of the relationship of two external ASes.

The algorithm used to find incomplete sets of routes is detailed in Algorithm 3. For every neighboring AS X, we consider the list of prefixes in which X appears in the AS-PATH of some BGP route. We then compare the list of prefixes obtained from BGP routes received by the local ASes with the one extracted from external BGP sources (e.g., RIS and Routeviews). If X is an eBGP peer, we only need to analyze the routes where the successive AS in the AS-PATH is X or one of its customers.

Algorithm 3: Algorithm used to obtain the incomplete set of routes.

input : External BGP paths.
output: Incomplete paths.

```
   /* Only analyze those paths where a peering AS is seen:                    */
[1]foreach Path P, with ASPATH containing a neighboring AS Neigh do
[2]    if Neigh is a transit provider; or Neigh is a peer and the path arrives from one of the customers of Neigh then
[3]        PathAttributes ← BGP attributes from P;
[4]        BestPaths ← Current Best paths of the network towards GetNLRI(P);
[5]        BackupPaths ← Best paths of the current network towards GetNLRI(P) when BestPaths are removed;
[6]        if P is better than any path in BestPaths or BackupPaths then
[7]            Return P;
```

8.4. Evaluation

In this section, we present the results obtained after deploying a prototype of the warning system in the network of two service providers. We used our system in an offline mode, for a-posteriori analyses of unsatisfied interests. In the following, we first describe our datasets in Section 8.4.1. We then discuss unsatisfied interests detected by our system for outbound and inbound traffic in Sections 8.4.2 and 8.4.3, respectively.

8.4.1. Data-sets

Our datasets are provided by the following two networks.

1. **Tier-2**. This dataset consists of the BGP routing tables and traffic data from an European Tier-2 network for the month of June 2014. Its network spans several countries, and exchanging prefixes with around 900 neighboring ASes. The routing tables are taken directly from the border routers of the network using Command Line Interface (CLI) commands. Hence, we avoid the existence of hidden BGP paths in our monitoring system, which can trigger false alarms in some of our test (in a running system, a more automatic technique, such as BGP ADD-PATH or BMP, would have been employed to obtain the same information) [179]. Concerning traffic data, we have the aggregated traffic per destination and per source prefix flowing through the core routers of the network.

2. **Academic Network**. The academic network encompasses around 20 nodes, has connections to two transit providers, and peers with several neighboring ASes over private links and exchange points. We have the BGP routing tables and traffic data from the Spanish

academic network (RedIRIS) for the month of March 2013. The BGP tables are obtained directly from each of the border routers. The traffic data consists of Netflow dumps from the routers of the network. Netflows dumps allow us to obtain the total traffic aggregated over different characteristics, such as ingress/egress interface; source/destination prefix or transport protocol.

We employ the *Tier-2* data-set to illustrate the outbound traffic algorithm, and the academic network data-set to show results for the inbound traffic algorithms. We use the *Tier-2* data-set of the outbound part, due to large path diversity that this network possesses, which can provide a rich set of results for this type of traffic. We move to the data-set of the *academic network* for the second part, as the granularity of the traffic of the *Tier-2* data-set is too coarse to perform the algorithms of the inbound traffic (e.g. cannot be divided in ingress traffic, per origin prefix, per physical link).

8.4.2. Outbound traffic measurements

Starting from the *Tier-2* dataset, our system first searched for missing paths using the procedures described in Section 8.3.2.4. We found 645 peers with 654,779 missing paths affecting 232,193 prefixes. 78,876 of these prefixes were affected by inconsistent advertisement, and 192,545 were affected by incomplete paths. Not all of these missing paths are meaningful for our analysis. Our system indeed runs the algorithm described in Section 8.2.1 to identify which of these paths lead to unsatisfied interests, assess their impact, and classify them correctly. After feeding the missing paths to the algorithm, we found that 439,273 of them (i.e., about 67%) have some kind of impact to the network. Those paths globally affect 144,622 prefixes. The collection and processing of network data dominates the overall execution time of the system, as it takes almost 10 hours to be completed. After network data is indexed, the algorithm can run in around 1 hour.

Moreover, we queried the Central Database module storing the output of the previous algorithm, in order to find the cases with a larger operational impact. In the following, we analyze the results of our queries, by providing an overview across all unsatisfied interests and a case-by-case analysis of the most impacting unsatisfied interests.

8.4.2.1. Results overview

We aggregated results according to different dimensions.

First, we group unsatisfied interests by neighboring AS and the impact type. Results are shown in Figure 8.7. The X-axis ranks the different grouped unsatisfied interests based on their impact, while the Y-axis shows the amount of traffic impacted by each unsatisfied interest. Since we are not able to show exact traffic values for confidentiality reasons, we use a non-disclosed value, in the order of Mbps, to scale the graph. Figure 8.7 exhibits many interesting things. Primarily, it shows that unsatisfied interests do have impact on a significant amount of traffic:

Globally, the impact of unsatisfied interests summed up to more than 1.5Gb of traffic per second! Moreover, note that distribution of unsatisfied interests with respect to their impact is quite skewed. In total, 740 grouped unsatisfied interests were found, but only 84 (approximately 11%) had an impact larger than zero. The impact of unsatisfied interests on traffic is also skewed. Indeed, the top 10 dissatisfactions account for 85% of this traffic. Finally, both frequency and impact of different types of unsatisfied interests are uneven. Generally speaking, the unexpected transit dissatisfaction is the less common, with only 3 cases in the top 50 of most impacting unsatisfied interests. Nevertheless, this type of dissatisfaction has the overall greatest impact (the first and third most impacting dissatisfactions are of this type). Neighboring preference dissatisfaction, next-hop diversity dissatisfaction, and backup path dissatisfaction complete the top-50 of cases with 9, 18, and 20 times, respectively.

Observe that different unsatisfied interests in Figure 8.7 can belong to the same neighboring AS. We found that from the top 50 cases with more aggregated traffic from the figure belong to 41 different neighboring AS.

To show the impact of unsatisfied interests due to specific neighbors, we plot detected unsatisfied interests aggregated only on a per neighboring AS basis in Figure 8.8. The X axis ranks the neighboring ASes based on the impact of their unsatisfied interests respectively induced by them. The Y axis shows the impact of unsatisfied interests in terms of affected traffic, using the same scaling process as in Figure 8.7. We found outbound dissatisfactions for about 471 neighboring ASes, for which 66 had an impact larger than zero. Even the distribution of ASes responsible for impacted traffic is highly skewed. Indeed, the top 10 ASes account for 87% of the total impacted traffic.

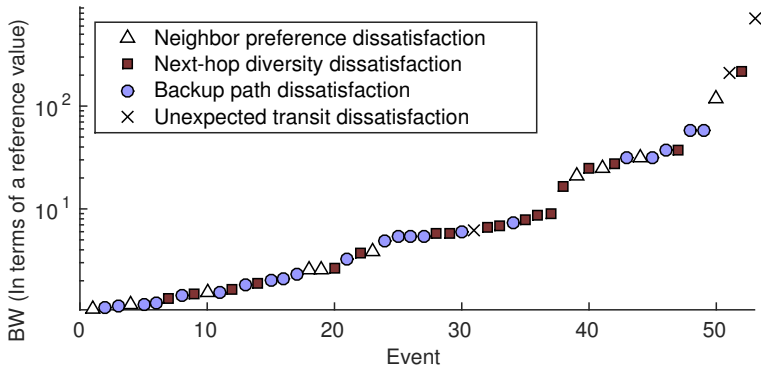


Figure 8.7: Outbound interest unsatisfied interests for the Tier-2 network, grouped by neighboring AS and type of impact.

8.4.2.2. Study

Figures 8.7 and 8.8 show that few dissatisfactions and few ASes are responsible for most of the unsatisfied interests traffic. We now take the perspective of an operator, and delve into

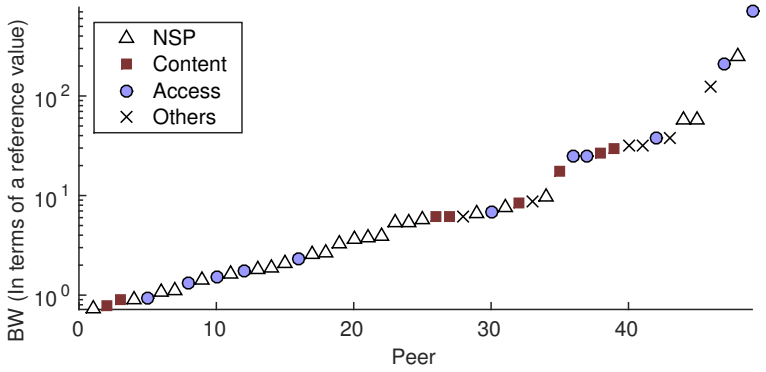


Figure 8.8: Outbound traffic affected by unsatisfied interests for every neighboring AS.

the dissatisfactions with the highest impact, in the hope of understanding causes and possible solutions to them. In particular, we focus on the unsatisfied interests caused by the top six AS (peers 45 to 50) in Figure 8.8.

The first and third AS of Figure 8.8 have a similar type of impact to the network. Their effects mostly correspond to the first and third unsatisfied interest in Figure 8.7, which are classified as a *Unexpected transit dissatisfaction*. Both of these ASes are multi-homed customers of the Tier-2. The detected unsatisfied interest are generated by the advertisement of more specific prefixes to another neighbor with respect to the prefixes sent to the Tier-2 network. More concretely, in those cases, at least one prefix p is advertised to the other transit provider but not to the Tier-2 network, which is left with a less specific prefix p' covering p. We assume that this is likely done for redundancy purposes. The Tier-2 receives those more specifics from non-customer neighbors. Note that, although the Tier-2 has a direct path towards these prefixes, the more specific prefixes force the traffic towards these non-customers. Unfortunately, there is no easy way to solve this situation [28]. Filtering the more specific prefixes could be, in some cases, considered as a contradiction to their policy from the point of view of the customers[1]. This information, however, could be very useful for network managers and planners. If the customers decide, for instance, to not change its behavior, the company could account the expenses that these unexpected transit traffic causes in the network, and add them to the business model of the customer.

The effect of the second AS of Figure 8.8 corresponds in most part to the second impact of Figure 8.7, a unsatisfied interest of type *Next-hop diversity dissatisfaction*. This particular AS is a service provider with multiple points of presence in different countries in Europe. By looking at the missing paths creating this dissatisfaction, we find that all of them are due to inconsistent advertisement. Figure 8.9 depicts the distribution of this type of missing paths for this peer. The top figure depicts the number of missing paths per ingress link in the network, while the lower figure weights each missing path with the amount of outbound traffic carried by the prefix. This peer is connected to the ISP via three different links. Each bar corresponds to

[1]The customers could use the inbound detection algorithm to detect if a provider is ignoring the more specifics.

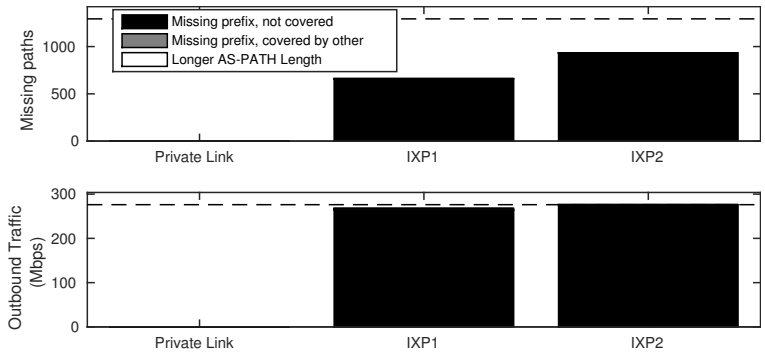


Figure 8.9: Inconsistencies from an individual peer. The dashed lines identify the total number of prefixes or total outbound traffic for this neighboring AS.

the amount of inconsistencies in terms of number of prefixes (up) or the outbound traffic (down). We see from the figure that the neighboring AS is attracting most of the traffic to the private link by doing selective advertisement. Note that half of the total prefixes are correctly announced through the IXP1 (top figure), but these prefixes do not attract much traffic (down figure). The two sessions suffering from the missing paths are through two different European IXPs. We also observe that the missing paths are "not covered", which means that the neighboring AS does not announce a more general prefix covering those routes. The neighboring AS could be incurring in inconsistency advertisement because it does not want to transport traffic from the two IXPs to these specific destinations. Depending on the contractual situation between the Tier-2 and this neighboring AS, the Tier-2 may contact the neighboring AS to enforce pre-agreed policies. In case the missing paths are due to route server path hiding [105], the two ASes may decide to establish a direct BGP section between each other.

The fourth AS of Figure 8.8 is the one causing the fourth unsatisfied interest of 8.7, which is of type *Neighbor preference dissatisfaction*. To understand the reasons of this case, we need to take a deeper look at the data from this neighboring AS. The neighboring AS is a settlement-free peer of the network, which hosts a top-30 Alexa site, and is a source and destination of a considerable amount of traffic for the network. The missing paths causing the dissatisfaction for this case are of type *incomplete*. In other words, these are paths that we detect that the neighboring AS possesses, yet they are not announced directly, but are received by the Tier-2 through its transit providers. These missing paths are originated and announced to the neighboring AS by another AS that, after a quick examination, indistinctly belongs to the same organization. The neighboring AS could be filtering the paths because it wants to avoid transporting traffic to these prefixes from the Tier-2. Independent of the root cause, it is evident that the neighboring AS has valid routes to them. By detecting this case, the Tier-2 could analyze whether it could demand the neighbor to announce this routes directly, in case when the peering agreement explicitly includes this on its terms [40] [139].

Finally, the fifth and sixth ASes from Figure are suffering from *Backup path dissatisfaction*. These two ASes face a similar type of unsatisfied interest, due to a common customer. The

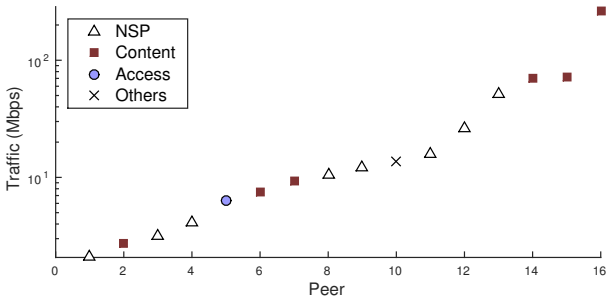


Figure 8.10: Peers traffic received over transit links.

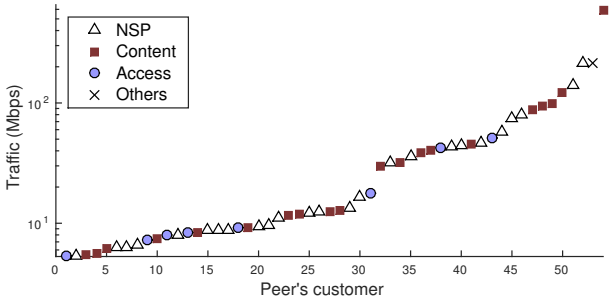


Figure 8.11: Peers customers traffic received over transit links.

problem, specifically, is that both ASes are not advertising paths of this common customer. The two neighboring ASes, and the customer AS, also happen to be settlement-free peers of the Tier-2, but connecting to the network through only one physical link (an IXP). In case of a failure of this IXP, there is the risk of sending the significant amount of traffic that the prefixes of this customer AS attracts, through transit providers. The reason why these two peers do not send these prefix is not yet determined. It can be that customer AS is performing selective advertisement, or that the two peers simply decided to not announce their prefixes. Our tool, however, still serves as a way of pointing network managers to this case, in order to obtain the attention needed, to prevent a big impact in the network in case of a failure.

8.4.3. Inbound traffic measurements

To evaluate the warning system for inbound traffic, we use the *Academic network* data-set. The inbound traffic analysis is conceptually simpler than the outbound traffic counterpart. The algorithm basically loops over the ingress flows per source prefix, and decides whether those flows fit into the *interest* of each operator. The core of the algorithm is to obtain, in a semi, or automatic matter, the interest of each network. For the case of the academic network, we reverse-engineered the applied BGP policy directly from the configuration of the routers, by correlating default Local-preference values to the type of neighboring AS. The result was a very basic policy targeting to enforce two key sub-interests 1. receive peer's traffic only over peer's links. 2. receive peer's customer's traffic only over peer's links.

8.4.3.1. Results overview

Figure 8.10 summarizes the cases in which peer's traffic is found over transit provider's links, hence violating sub-interest 1. We found cases for 33 peers, 16 with more than 1Mbps. The X axis ranks each settlement-free peering ASes based on the amount of traffic originating in those AS found in transit links. The Y axis measures directly that traffic in Mbps. The top 3 ASes account for about 75% of the total affected traffic.

Figure 8.11 summarizes the cases in which peer's customer's traffic is found over transit provider's links, that is, those violating sub-interest 2. We found traffic of more than 500 peer's customers over the transit links of the network (in the figure, we just show the cases with more than 5Mbps). The X axis ranks the customers of the settlement-free peers based on the amount of traffic originating in those AS found in transit links. The Y axis measures directly that traffic in Mbps.

8.4.3.2. Case-by-case analysis

Although the philosophy of the tool is not to provide root causes for the undesired traffic, we can use some additional information to provide assumptions on the reason of this traffic. Regarding dissatisfactions of sub-interest 1 (see Figure 8.10), the top three settlement-free peers with more traffic on the transit links are content providers. This type of companies performs Traffic Engineering practices that differ from those of access or transit networks. Indeed, content providers use different sources of information, such as the DNS system [149] or cache location [109], to select the source host and the path towards the end user. This behavior many times neglects BGP data, making it thus possible for them to send traffic ignoring the direct connections [109]. Given these circumstances, the Academic-network could decide to expand its infrastructure with these content providers [109], or explore other collaboration techniques, as described, for instance in [149] to reduce the amount of traffic of these companies over its transit links. The next three companies with most traffic are NSPs. These companies could operate under outbound TE practices in which they prefer to send some traffic to the academic network through indirect ASes, for instance due to back-haul transport costs, or other technical details. However, this are just speculations as there is no clear reason why there is traffic from these peers over the transit links. The Academic network would be required to examine each case, probably contacting each network individually.

In contrast to sub-policy 1, dissatisfactions to sub-policy 2 (Fig. 8.11) are harder to analyze because of the (routing) distance of ASes causing them to the academic network. Still, there might cases of undesired policies from the direct neighbors: for instance, they can perform intermediate filtering [126]. The undesired traffic can also be due to the outbound traffic policies of the origin networks. Note that many of the top contributors are content providers, for which the same analysis of the last case applies. Concretely, these companies might select the paths from the caches connected to the transit providers of the Academic network, instead to the ones available

through the settlement-free peers. Operators and peering managers can use this data to potentially look for new peering agreements that can reduce the inbound traffic through transit providers.

8.5. Related Work

Routing divergence. Several authors have studied the effects of routing conflicts in the Internet. These works have examined the conditions in which uncoordinated policies can cause the BGP algorithm to diverge [95] [159] [53]; and to propose systems that can detect these cases [81] [54], or prevent them [159]. The unsatisfied interests that we analyze in this work do not fit in this category, as they focus only in the policy interest of the operator under a stable state. In other words, we analyze cases in which BGP converges, but where one or more ASes do not obtain the state that they originally intended.

BGP security. Although our warning system could detect the effect of forged BGP routes, we encourage operators to have dedicated systems to detect these cases. Multiple proposals have emerged to secure the Internet [101] [103]. In recent years, the IETF standardized the Resource Public Key Infrastructure (RPKI) [21]. RPKI uses a public key system to validate some of the information included in BGP updates.

External peering auditing. The distributed nature of the Internet makes it prone to situations in which the behavior of a single system (either allowed or not) can affect many others. Many authors have analyzed how to detect the cause of specific situations affecting ASes, even under environments in which only partial information can be obtained. [72] describes the problem of inconsistency advertisement from neighboring peers and how they could be detected using local data. [79] uses data-plane data (traceroute) to find disruptions between what is announced in the control-plane and the actual path of each packet. [196] proposes a cryptography system that can test several properties of the received routes from neighboring ASes. [98] [175] [193] use network information to pinpoint any AS behaving in an unexpected manner, or causing specific route changes. [126] checks for prefix filtering that can limit the visibility of prefixes for other ASes. These systems complement the information provided by the warning application. Some of these systems (e.g. [72]) can feed our application with missing paths that can be used to run the outbound traffic algorithm. Other systems (e.g. [126]) could help operators find root causes for specific unsatisfied interests that with greatest impact on the network.

Internal configuration checking. Some unsatisfied interests can be explained, not by policies of external ASes, but by mistaken configurations of internal devices. Route leaks, for instance, can arise due to this problem, and can trigger the warning system (e.g. by detecting traffic to transit prefixes arriving in a settlement-free peer link) [107]. Operators can implement configuration checking systems that can avoid misalignment between internal policies and configurations [70]. If the operators actively use Internet Routing Registry (IRR) to publish their inter-domain policy, systems like [167] can be employed to check the policy against the BGP updates.

Different inter-domain routing protocols. Due to the limitations and problems experienced

by the current Internet, different authors have proposed improved inter-domain network architectures. These can either provide flexibility to the systems [154]; improve the behavior of BGP by extending some aspects of the protocol [189]; or working with overlay protocols that provide more features and control [68]. Our system operates abstractly in policy decisions from ASes, and not directly in information generated exclusively by BGP. Therefore, the system can be adapted to any of these inter-domain architectures.

Content providers and network neutrality. Content providers have established themselves as the origin of a large proportion of Internet traffic. These companies have particular operational practices and policies concerning inter-domain routing, which can clash with the interest of eyeballs or transit providers, triggering warnings in our system. For instance, some content providers have no backbone, and decide forwarding routes disregarding any policy reflected in BGP announcements [109] [75]. Access and transit networks should be aware of these policies, and use the warning systems as a notification system that can point operators to cases with a large impact on the network. The operators can then tackle each case in conjunction with content provider. Providers can use systems like the ones proposed in [149] [191] [80], which can be used to establish a collaboration between content and access providers, in order to reduce incompatible interests. A large part of the discussions around *network neutrality* have been generated by conflicting relations between access and content provider networks [65]. Although specific details of agreements fitting network neutrality clauses is beyond the scope of this work, we envision that our system could be use by *both sides* of this relationship. The warning system cold provide useful information to specific situations that do not fit points of any peering agreement, such as missing prefix announcement or disrespect for inbound policy.

8.6. Summary

In this chapter, we studied inter-domain routing configurations in which the (economic) interests of one or more ASes are not satisfied. Taking an AS-centric perspective, we (I) classified possible unsatisfied interests; (II) proposed algorithms to detect them and assess their impact; and (III) described a warning system providing operators with critical input on business-impacting unsatisfied interests. We used our system to perform real-world measurements. Our results show that unsatisfied interests do occur in practice and can affect a non-negligible amount of traffic (a couple of Gbps in one of our cases). The main conclusions from this chapter are:

Focus on the unsatisfied interests with larger impact. In our measurements, we found a large amount of unsatisfied interests of different types. We believe this situation might be common for many other ASes in the Internet. Due to the partial information that one can obtain from external policies, it might be hard to detect the root cause of all cases. This requires operators to focus on each case individually, analyzing data or making calls to find the reason behind them. As this is operative intensive operation, network managers should focus on the cases with larger network impact.

Unsatisfied interests are not always solvable. unsatisfied interests caused by routing leaks or forged advertisement can be detected and solved, however other unsatisfied interests might lead to one AS being discontent with the final traffic distribution. Network managers suffering from unsatisfied interests should understand each case, in order to evaluate the steps to deal with them. Some operators might, for instance, decide to continue with the dissatisfaction, but incorporate the missing revenue in their service's pricing model.

Chapter 9

Conclusion

Network operators must cope with continually increasing traffic volumes and more stringent service levels agreements, while at the same time striving for cost reduction. Increasing the number of inter-domain peering links offers operators a direct way of off-loading transit traffic, increasing resiliency and reducing the transport bill, while shortening the path to the content consumer. In the last years, the massive increase of settlement-free peering links has transformed the Internet infrastructure into a flatter network, of which IXPs become central entities. Indeed, our own observations, and those of other authors, have shown that this amount of peering links would not be feasible via dedicated private connections. IXPs provide the technical means to support these connections, and, complementary, provide a space to discover new peering opportunities. Remote peering has also emerged as a new interconnection service that allows ASes to efficiently join various IXPs.

The benefits of joining multiple IXPs are not for free, since they also come with management overhead. Concretely, the large number of peering relationships, frequently informally established or with no previous contact, forces operators to monitor the behavior of external ASes, which can be affecting their own policies. Operators should thus manage their inter-domain traffic using procedures adapted to an environment in which external ASes have dynamic policies, and networks that are prone to errors. This management procedure not only concerns the traffic control under a given infrastructure, but also those processes related to network expansion, or peering management.

The goal of this thesis was to provide methods for operators to manage their inter-domain traffic in such an environment. For this purpose, we divided the thesis in two first parts. We first explored the IXP-centric Internet ecosystem. We characterized different IXPs around the world, leveraging public information to analyze the evolution of several of their characteristics. In the second part of the thesis, we took the perspective of an individual AS operator. We defined a framework for the management of inter-domain traffic, highlighting the difficulties behind obtaining the data required for this purpose, and the need of policy validation. In addition, using real network data, we provided examples of applications of this framework.

IXP Characterization. By studying the characteristics of various IXPs, we observed the differences between various types of IXPs. On the one hand, most IXPs focus on covering particular geographical domains. The Brazilian and Russian IXPs (PTT and MSK) are examples of exchanges serving large areas; yet, PTT shows the largest growing rate in term of members, while MSK size has been stable in the last years. In general, other regional IXPs, such as the Slovakian-IX, which we examined with detail in Chapter 4, are also experiencing member stagnation, as they already attracted the companies that might be willing to peer in their particular domains. On the other hand, the three large European IXPs, AMS-IX, LINX, and DEC-IX share very similar characteristics of size and traffic level. We use the number of unique members and control plane information to show the member overlapping of most IXPs. In fact, a few isolated IXPs are composed by a large percentage of ASes peering at a single exchange. Member overlapping affects companies and IXPs alike. For the first, it reduces the opportunity for traffic off-loading after having already joined a couple of IXPs. For the second, it can discourage companies to join them, limiting their footprint to the companies that can easily reach them (normally the close ones).

Remote Peering. By using remote peering services, a company can join multiple IXPs without extending their infrastructure or operational overhead towards new locations. Remote peering thus helps operators reduce the entry cost to an IXP. Simultaneously, by embracing remote peering providers, IXPs broaden their geographical footprint. Using active measurements from multiple looking glasses, we evaluated the use of remote peering at several IXPs. We discovered members using remote peering in more than 90% of the studied IXPs. We also discussed the impact of remote peering for the Internet structure. Certainly, remote peering providers become central entities in the layer-1 or layer-2 view of the Internet, while they go unnoticed with measurements performed at layer-3. In a practical perspective, this means that a single remote peering provider might connect the same ASes at different IXPs. Operators must be aware that remote peering can hide single points of failure, since a problem in a remote peering provider could cause simultaneous failures in the links to different IXPs. ASes requiring high availability should ensure that this is not occurring in the networks of critical settlement-free peers.

Inter-domain management framework. We described the difficulties related to the management of inter-domain management in Chapter 6. First, we exposed the challenges of procuring and analyzing the data required for this process. The main challenges for this task is the large volume of data, heterogeneity of devices producing the data, and incompleteness or inaccuracy of data. The second challenge is the difficulty of simulating changes in labs, given the undetermined nature of external policies; the non-deterministic behavior of BGP in some network topologies; and the size of the overall dataset needed to process. Operators are hence obliged to trial and investigate network changes over their own live network, monitor its actual effect, and react based on the perceived impact. Third, the policies of external ASes are dynamic and might conflict with the interest of the operators, and establish procedures that specifically deal with those cases.

Peering study using remote peering. We performed a peering study, evaluating potential network expansions needed to improve the performance and profitability of the inter-domain as-

pect of a real network. We assessed the transit off-load potential after connecting privately to new peers, or publicly through IXPs. We showed how the off-loaded transit traffic decreases quickly after joining a couple of large IXP, hinting at the use of remote peering to reach additional exchanges. Further, we discussed different methods for operators to quantify other benefits of peering at multiple sites, such as the increase of resiliency, and the reduction of internal transport cost. Any type of verified information that operators can obtain from external peers, including their relationships to other ASes or the prefixes they announce on IXP route servers, reduces the risk of yielding incorrect results in a peering study.

Incompatible inter-domain policies. The distributed nature of the Internet makes it feasible for different ASes to have incompatible policies. In those cases, the resulting inter-domain traffic distribution might not satisfy the interest of one or more ASes. Assuming the position of one AS, we describe how these unsatisfied interest can affect both traffic directions. Outbound traffic is affected, for instance, when an AS does not receive the same paths from a neighboring AS at their sessions in different IXPs. Inbound traffic is affected when external ASes select a path that is less preferable for the network operation and business. We provide several algorithms that operators can use to find these cases in their networks, and describe the implementation of a system that warns operators when they occur. Using data from two real networks, we show that unsatisfied interest are frequent, highlighting the need for operators to use such tools. We envision our system to complement other validation tools, such as RPKI systems or route leaks detectors.

Despite the large efforts from researchers and manufacturers, network operators still lack the tools to perform proper inter-domain network management. A few companies enjoy the internal resources needed to create in-house solutions for this purpose, but they are usually specific for their networks. The commoditization of big data architectures; the support of open APIs by network devices; and the expansion of an analytic and development culture among network engineers provides an environment feasible for the procedures required for inter-domain management. In parallel, the development of technologies like Segment Routing allows the granular control of inter-domain traffic in a centralized way, facilitating the implementation of optimal traffic engineering strategies. IXPs could also leverage these technologies to increase their efficiency, avoid undesired effects such as Route Server path-hiding, and provide direct control on policy implementation to their members.

References

[1] Emile Aben. Internet Traffic During the World Cup 2014. https://labs.ripe.net/Members/emileaben/internet-traffic-during-the-world-cup-2014, 2014.

[2] Bernhard Ager, Nikolaos Chatzis, Anja Feldmann, Nadi Sarrar, Steve Uhlig, and Walter Willinger. Anatomy of a Large European IXP. In *Proceedings of ACM SIGCOMM*, 2012.

[3] Divyakant Agrawal, Sudipto Das, and Amr El Abbadi. Big data and cloud computing: current state and future opportunities. In *Proceedings of the 14th International Conference on Extending Database Technology*, pages 530–533. ACM, 2011.

[4] Aditya Akella, Bruce Maggs, Srinivasan Seshan, Anees Shaikh, and Ramesh Sitaraman. A measurement-based analysis of multihoming. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 353–364. ACM, 2003.

[5] AMS-IX. Amsterdam IXP (AMS-IX) webpage. https://www.ams-ix.net/.

[6] Internet Archive. The Wayback Machine. urlhttp://waybackmachine.org/.

[7] European Internet Exchange Association. Euro-IX webpage. www.euro-ix.net.

[8] B. Augustin, B. Krishnamurthy, and W. Willinger. IXPs: Mapped? In *Proceedings of ACM IMC*, 2009.

[9] Brice Augustin, Xavier Cuvellier, Benjamin Orgogozo, Fabien Viger, Timur Friedman, Matthieu Latapy, Clémence Magnien, and Renata Teixeira. Avoiding traceroute anomalies with Paris traceroute. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 153–158. ACM, 2006.

[10] Daniel Awduche, Angela Chiu, Anwar Elwalid, Indra Widjaja, and XiPeng Xiao. Overview and principles of Internet traffic engineering. *IETF RFC 3272*, 2002.

[11] Josh Bailey, Russ Clark, Nick Feamster, Dave Levin, Jennifer Rexford, and Scott Shenker. SDX: A Software Defined Internet Exchange. *ACM SIGCOMM Computer Communication Review*, 44(4):551–562, 2015.

[12] Hitesh Ballani, Paul Francis, and Xinyang Zhang. A study of prefix hijacking and interception in the Internet. In *ACM SIGCOMM Computer Communication Review*, volume 37, pages 265–276. ACM, 2007.

[13] Simon Balon and Guy Leduc. Combined intra-and inter-domain traffic engineering using hot-potato aware link weights optimization. *ACM SIGMETRICS Performance Evaluation Review*, 36(1):441–442, 2008.

[14] Tony Bates, Enke Chen, and Ravi Chandra. BGP Route Reflection-An Alternative to Full Mesh IBGP. *IETF RFC 4456*, 2006.

[15] Zied Ben Houidi, Mickael Meulle, and Renata Teixeira. Understanding slow BGP routing table transfers. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 350–355. ACM, 2009.

[16] S. Biernacki. Remote Peering at IXPs. *EURONOG 1 presentation*, 2011.

[17] Martin Bjorklund. YANG-A data modeling language for the Network Configuration Protocol (NETCONF). *RFC 6020*, 2010.

[18] P. Borgnat and et al. "Seven years and one day: Sketching the evolution of internet traffic". In *Proceedings of* IEEE INFOCOM, 2009.

[19] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer networks*, 33(1):309–320, 2000.

[20] M. Brown and et al. Peering Wars: Lessons Learned from the Cogent-Telia De-peering. *NANOG*, 43, 2008.

[21] Randy Bush and Rob Austein. The Resource Public Key Infrastructure (RPKI) to Router Protocol. *IETF RFC 6810*, 2013.

[22] CAIDA. Routeviews Prefix to AS mappings Dataset for IPv4 and IPv6. `http://www.caida.org/data/routing/routeviews-prefix2as.xml`, 2014.

[23] CAIDA. The CAIDA AS Relationships Dataset. `http://www.caida.org/data/active/as-relationships/`, 2014.

[24] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann, and R. Govindan. Mapping the Expansion of Google's Serving Infrastructure. *IMC*, 2013.

[25] J Camilo Cardona, Pierre Francois, Bruno Decraene, John Scudder, Adam Simpson, and Keyur Patel. Bringing high availability to BGP: A survey. *Computer Networks*, 91:788–803, 2015.

[26] Juan Cardona, Stefano Vissicchio, Paolo Lucente, and Pierre Francois. "I Can't Get No Satisfaction": Helping Autonomous Systems Identify Their Unsatisfied Inter-domain Interests. *Transactions on Network and Service Management*, 2016.

[27] Juan Camilo Cardona and Ignacio Castro. Remote Peering Data. `https://svnext.networks.imdea.org/repos/RemotePeering`.

[28] Juan Camilo Cardona, Pierre Francois, and Paolo Lucente. Impact of BGP filtering on Inter-Domain Routing Policies. *RFC 7789*, 2016.

[29] Juan Camilo Cardona, Pierre Francois, Saikat Ray, Keyur Patel, Paolo Lucente, and Pradosh Mohapatra. BGP Path Marking. *draft-bgp-path-marking-00. Work in Progress. IETF Draft.*, 2013.

[30] Juan Camilo Cardona, Claus G Gruber, and Carmen Mas Machuca. Energy profile aware routing. In *Communications Workshops, 2009. ICC Workshops 2009. IEEE International Conference on*, pages 1–5. IEEE, 2009.

[31] Juan Camilo Cardona and Rade Stanojevic. IXP history Dataset. `https://svnext.networks.imdea.org/repos/ixp_history/`, 2013.

[32] Juan Camilo Cardona, Rade Stanojevic, and Nikolaos Laoutaris. Collaborative Consumption for Mobile Broadband: A Quantitative Study. In *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies*, pages 307–318. ACM, 2014.

[33] J. C. Cardona Restrepo, R. Stanojevic, and R. Cuevas. On weather and Internet traffic demand. *PAM*, 2013.

[34] Juan Camilo Cardona Restrepo and Rade Stanojevic. A History of an Internet eXchange Point. *ACM SIGCOMM CCR*, 42(2):58–64, 2012.

[35] Juan Camilo Cardona Restrepo and Rade Stanojevic. IXP traffic: a macroscopic view. In *Proceedings of the 7th Latin American Networking Conference*, pages 1–8. ACM, 2012.

[36] Martin Casado, Teemu Koponen, Scott Shenker, and Amin Tootoonchian. Fabric: a retrospective on evolving SDN. In *Proceedings of the first workshop on Hot topics in software defined networks*, pages 85–90. ACM, 2012.

[37] Ignacio Castro, Juan Camilo Cardona, Sergey Gorinsky, and Pierre Francois. Remote Peering: More Peering without Internet Flattening. In *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies*, pages 185–198. ACM, 2014.

[38] Ignacio Castro, Rade Stanojevic, and Sergey Gorinsky. Using tuangou to reduce IP transit costs. *Networking, IEEE/ACM Transactions on*, 22(5):1415–1428, 2014.

[39] D Cavalcanti. The Role of Internet Exchange Points in Broadband Policy and Regulation. *Revista de Direito, Estado e Telecomunicações*, 3(1):75–88, 2011.

[40] CenturyLink. CenturyLink's North America IP Network Peering Policy. `http://www.centurylink.com/legal/peering_na.html`, 2011.

[41] Joseph Chabarek, Joel Sommers, Paul Barford, Cristian Estan, David Tsiang, and Steve Wright. Power awareness in network design and routing. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*. IEEE, 2008.

[42] Hyunseok Chang and Sugih Jamin. To Peer or Not to Peer: Modeling the Evolution of the Internet AS-level Topology. *INFOCOM*, 2006.

[43] Hyunseok Chang, Sugih Jamin, and Walter Willinger. Inferring AS-level Internet Topology from Router-Level Path Traces. In *Proceedings of SPIE ITCOM*, 2001.

[44] N. Chatzis, G. Smaragdakis, A. Feldmann, and W. Willinger. There Is More to IXPs than Meets the Ey. *2013*, CCR.

[45] G. Chen and et al. Energy-Aware Server Provisioning and Load Dispatching for Connection-Intensive Internet Services. In *Proceedings of NSDI*, 2008.

[46] Hao Chen, Michael C Caramanis, and Ayse K Coskun. Reducing the data center electricity costs through participation in smart grid programs. In *Green Computing Conference (IGCC), 2014 International*, pages 1–10. IEEE, 2014.

[47] Kai Chen, David R Choffnes, Rahul Potharaju, Yan Chen, Fabian E Bustamante, Dan Pei, and Yao Zhao. Where the sidewalk ends: Extending the Internet AS graph using traceroutes from P2P users. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pages 217–228. ACM, 2009.

[48] Pei-chun Cheng, Jong Han Park, Keyur Patel, Shane Amante, and Lixia Zhang. Explaining BGP Slow Table Transfers. In *Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on*, pages 657–666. IEEE, 2012.

[49] Kenjiro Cho, Kensuke Fukuda, Hiroshi Esaki, and Akira Kato. Observing slow crustal movement in residential user traffic. *CoNEXT*, 2008.

[50] Jaeyoung Choi, Jong Han Park, Pei-chun Cheng, Dorian Kim, and Lixia Zhang. Understanding BGP next-hop diversity. In *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, 2011.

[51] Krzysztof J Cios, Witold Pedrycz, and Roman W Swiniarski. *Data Mining and Knowledge Discovery*. Springer, 1998.

[52] Cisco. The value of peering. ISP/IXP Workshop. `http://www.pacnog.net/pacnog6/IXP/IXP-peering.pdf`.

[53] Luca Cittadini, Giuseppe Di Battista, Massimo Rimondini, and Stefano Vissicchio. Wheel+ ring= reel: The impact of route filtering on the stability of policy routing. *IEEE/ACM ToN*, 2011.

[54] Luca Cittadini, Massimo Rimondini, Stefano Vissicchio, Matteo Corea, and Giuseppe Di Battista. From theory to practice: Efficiently checking BGP configurations for guaranteed convergence. *IEEE Transactions on Network and Service Management*, 8(4):387–400, 2011.

[55] Benoit Claise. Cisco systems NetFlow services export version 9. *IETF RFC 3954*, 2004.

[56] A. Dhamdhere and C. Dovrolis. Ten years in the evolution of the internet ecosystem. In *Proceedings of ACM IMC*, 2008.

[57] A. Dhamdhere and C. Dovrolis. The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh. In *Proceedings of ACM CoNEXT*, 2010.

[58] A. Dhamdhere, C. Dovrolis, and P. Francois. A Value-based Framework for Internet Peering Agreements. In *Proceedings of ITC*, 2010.

[59] X. Dimitropoulos, P. Hurley, A. Kind, and M. Stoecklin. On the 95-percentile Billing Method. *PAM*, 2009.

[60] Benoit Donnet and Olivier Bonaventure. On BGP communities. *ACM SIGCOMM CCR*, 2008.

[61] Nick Duffield, Kartik Gopalan, Michael R Hines, Aman Shaikh, and Jacobus E Van Der Merwe. Measurement informed route selection. *Passive and Active Network Measurement*, pages 250–254, 2007.

[62] Ramakrishnan Durairajan, Joel Sommers, and Paul Barford. Layer 1-Informed Internet Topology Measurement. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 381–394. ACM, 2014.

[63] Marian Durkovic. TRILL Deployment in SIX. `http://www.six.sk/trill.pdf`, 2014.

[64] Dyn. Dyn. `http://dyn.com`, 2014.

[65] Nicholas Economides and Joacim Tåg. Network neutrality on the Internet: A two-sided market analysis. *Information Economics and Policy*, 2012.

[66] Rodérick Fanou, Pierre Francois, and Emile Aben. On the Diversity of Interdomain Routing in Africa. In *Passive and Active Measurement*, pages 41–54. Springer, 2015.

[67] Peyman Faratin, David D Clark, Steven Bauer, William Lehr, Patrick W Gilmore, and Arthur Berger. The growing complexity of Internet interconnection. *Communications & strategies*, 2008.

[68] Dino Farinacci, Darrel Lewis, David Meyer, and Vince Fuller. The locator/ID separation protocol (LISP). *IETF RFC 6830*, 2013.

[69] Nicholas Greer Feamster. *Proactive techniques for correct and predictable internet routing*. PhD thesis, Massachusetts Institute of Technology, 2005.

[70] Nick Feamster and Hari Balakrishnan. Detecting BGP configuration faults with static analysis. In *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2*, pages 43–56. USENIX Association, 2005.

[71] Nick Feamster, Jay Borkenhagen, and Jennifer Rexford. Guidelines for interdomain traffic engineering. *ACM SIGCOMM CCR*, 2003.

[72] Nick Feamster, Zhuoqing Morley Mao, and Jennifer Rexford. BorderGuard: Detecting cold potatoes from peers. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 213–218. ACM, 2004.

[73] Nick Feamster, Jared Winick, and Jennifer Rexford. A model of BGP routing for network engineering. In *ACM SIGMETRICS Performance Evaluation Review*, 2004.

[74] Eugene A Feinberg and Dora Genethliou. Load forecasting. In *Applied mathematics for restructured electric power systems*, pages 269–285. Springer, 2005.

[75] Pierdomenico Fiadino, Alessandro D'Alconzo, Arian Bar, Alessandro Finamore, and Pedro Casas. On the detection of network traffic anomalies in content delivery network services. In *Teletraffic Congress (ITC)*, 2014.

[76] C. Filsfils, N. Kumar, C. Pignataro, J. C. Cardona Restrepo, and P. Francois. The Segment Routing Architecture. *IEEE GLOBECOMM: Next Generation Networking Symposium*, 2015.

[77] Clarence Filsfils, Pradosh Mohapatra, John Bettink, Pranav Dharwadkar, Peter De Vriendt, Yuri Tsier, Virginie Van Den Schrieck, Olivier Bonaventure, and Pierre Francois. BGP prefix independent convergence (PIC) technical report. Technical report, Cisco, Tech. Rep, 2011.

[78] Clarence Filsfils, Thomas Telkamp, Paolo Lucente, and Arman Maghbouleh. Best Practices in Network Planning and Traffic Engineering. *NANOG 52*, 2011.

[79] Tobias Flach, Ethan Katz-Bassett, and Ramesh Govindan. Quantifying violations of destination-based forwarding on the internet. In *Proceedings of IMC*, 2012.

[80] Ashley Flavel, Pradeepkumar Mani, David A Maltz, Nick Holt, Jie Liu, Yingying Chen, and Oleg Surmachev. FastRoute: A Scalable Load-Aware Anycast Routing Architecture for Modern CDNs. *NSDI*, 2015.

[81] Ashley Flavel, Jeremy McMahon, Aman Shaikh, Matthew Roughan, and Nigel Bean. BGP route prediction within ISPs. *Computer Communications*, 33(10):1180–1190, 2010.

[82] Ari Fogel, Stanley Fung, Luis Pedrosa, Meg Walraed-Sullivan, Ramesh Govindan, Ratul Mahajan, and Todd Millstein. A general approach to network configuration analysis. In *Networked Systems Design and Implementation*, 2015.

[83] M. Fomenkov and et al. Longitudinal study of internet traffic in 1998-2003. In *Proceedings of the WISICT*, 2004.

[84] Lixin Gao. On inferring autonomous system relationships in the Internet. *IEEE/ACM Transactions on Networking (ToN)*, 9(6):733–745, 2001.

[85] Lixin Gao, Roch Guerin, Zhi-Li Zhang, and Yong Liao. Safe Inter-domain Routing under Diverse Commercial Agreements. *IEEE/ACM Transactions on Networking*, pages –, May 2010.

[86] Lixin Gao and Jennifer Rexford. Stable Internet routing without global coordination. *IEEE/ACM ToN*, 2001.

[87] Qixin Gao, Feng Wang, and Lixin Gao. Routing-Policy Aware Peering for Large Content Providers. *Computer Communications*, 2015.

[88] Ruomei Gao, Constantinos Dovrolis, and Ellen W Zegura. Interdomain ingress traffic engineering through optimized AS-path prepending. In *NETWORKING 2005*. Springer, 2005.

[89] Phillipa. Gill, M. Arlitt, Z. Li, and A. Mahanti. The Flattening Internet Topology: Natural Evolution, Unsightly Barnacles or Contrived Collapse? In *Proceedings of PAM*, 2008.

[90] Phillipa Gill, Michael Schapira, and Sharon Goldberg. Let the market drive deployment: A strategy for transitioning to BGP security. In *ACM SIGCOMM Computer Communication Review*, volume 41, pages 14–25. ACM, 2011.

[91] Phillipa Gill, Michael Schapira, and Sharon Goldberg. A survey of interdomain routing policies. *ACM SIGCOMM CCR*, 2013.

[92] E. Gregori, A. Improta, L. Lenzini, and C. Orsini. *The impact of IXPs on the AS-level topology structure of the Internet*. Computer Communications, 2010.

[93] Enrico Gregori, Luciano Lenzini, and Chiara Orsini. k-clique Communities in the Internet AS-level Topology Graph. In *Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference on*, pages 134–139. IEEE, 2011.

[94] Tim Griffin and Geoff Huston. BGP wedgies. *IETF RFC 4264*, 2005.

[95] Timothy G Griffin, F Bruce Shepherd, and Gordon Wilfong. The stable paths problem and interdomain routing. *IEEE/ACM ToN*, 2002.

[96] Timothy G Griffin and Gordon Wilfong. Analysis of the MED Oscillation Problem in BGP. In *Network Protocols, 2002. Proceedings. 10th IEEE International Conference on*, pages 90–99. IEEE, 2002.

[97] Hamed Haddadi, Miguel Rio, Gianluca Iannaccone, Andrew Moore, and Richard Mortier. Network topologies: inference, modeling, and generation. *Communications Surveys & Tutorials, IEEE*, 10(2):48–69, 2008.

[98] Andreas Haeberlen, Ioannis C Avramopoulos, Jennifer Rexford, and Peter Druschel. NetReview: Detecting When Interdomain Routing Goes Wrong. In *NSDI*, 2009.

[99] Susan Hares and Russ White. Software-Defined Networks and the Interface to the Routing System (I2RS). *IEEE Internet Computing*, 17(4), 2013.

[100] Y. He and et al. Lord of the links: a framework for discovering missing links in the internet topology. *IEEE/ACM Transactions on Networking*, 17:2, 2009.

[101] Yih-Chun Hu, Adrian Perrig, and Marvin Sirbu. SPV: Secure path vector routing for securing BGP. *ACM SIGCOMM CCR*, 2004.

[102] John D Hunter. Matplotlib: A 2D graphics environment. *Computing in science and engineering*, 2007.

[103] Geoff Huston, Mattia Rossi, and Grenville Armitage. Securing BGP-A literature survey. *Communications Surveys & Tutorials, IEEE*, 2011.

[104] Sushant Jain, Alok Kumar, Subhasree Mandal, Joon Ong, Leon Poutievski, Arjun Singh, Subbaiah Venkata, Jim Wanderer, Junlan Zhou, Min Zhu, et al. B4: Experience with a globally-deployed software defined WAN. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 3–14. ACM, 2013.

[105] Elisa Jasinska, Nick Hilliard, Robert Raszuk, and Niels Bakker. Internet Exchange Route Server. *draft-ietf-idr-ix-bgp-route-server-07. Work in Progress. IETF Draft.*, 2015.

[106] Wenjie Jiang, Stratis Ioannidis, Laurent Massoulié, and Fabio Picconi. Orchestrating massively distributed CDNs. In *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, pages 133–144. ACM, 2012.

[107] Sriram. K., D. Montgomery, D. McPherson, and E. Osterweil. Problem Definition and Classification of BGP Route Leaks. *draft-ietf-grow-route-leak-problem-definition-01. Work in Progress. IETF Draft.*, 2015.

[108] Ethan Katz-Bassett, Harsha V Madhyastha, John P John, Arvind Krishnamurthy, David Wetherall, and Thomas E Anderson. Studying Black Holes in the Internet with Hubble. In *NSDI*, pages 247–262, 2008.

[109] Christian Kaufmann. BGP and Traffic Engineering with Akamai. `http://www.menog.org/presentations/menog-14/282-20140331_MENOG_BGP_and_Traffic_Engineering_with_Akamai.pdf`, 2014.

[110] Ram Keralapura, Nina Taft, Chen-Nee Chuah, and Gianluca Iannaccone. Can ISPs take the heat from overlay networks. In *ACM SIGCOMM Workshop on Hot Topics in Networks (HotNets)*. Citeseer, 2004.

[111] Rowan Klöti, Bernhard Ager, Vasileios Kotronis, George Nomikos, and Xenofontas Dimitropoulos. A Comparative Look into Public IXP Datasets. *ACM SIGCOMM Computer Communication Review*, 46(1):21–29, 2016.

[112] Dániel Kondor, Pierrick Thebault, Sebastian Grauwin, István Gódor, Simon Moritz, Stanislav Sobolevsky, and Carlo Ratti. Visualizing signatures of human activity in cities across the globe. *arXiv preprint arXiv:1509.00459*, 2015.

[113] Fredy Künzler. How More Specifics increase your transit bill (and ways to avoid it). `https://ripe63.ripe.net/presentations/48-How-more-specifics-increase-your-transit-bill-v0.2.pdf`, 2011.

[114] Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, and Farnam Jahanian. Internet Inter-domain Traffic. In *Proceedings of ACM SIGCOMM*, 2010.

[115] Nikolaos Laoutaris, Michael Sirivianos, Xiaoyuan Yang, and Pablo Rodriguez. Inter-datacenter bulk transfers with netstitcher. In *ACM SIGCOMM Computer Communication Review*, volume 41, pages 74–85. ACM, 2011.

[116] Nikolaos Laoutaris, Georgios Smaragdakis, Pablo Rodriguez, and Ravi Sundaram. Delay Tolerant Bulk Data Transfers on the Internet. In *Proceedings of ACM SIGMETRICS*, 2009.

[117] DK Lee, Kenjiro Cho, Gianluca Iannaccone, and Sue Moon. Has internet delay gotten better or worse? In *Proceedings of the 5th International Conference on Future Internet Technologies*, pages 51–54. ACM, 2010.

[118] LINX. London IXP (LINX) webpage. `https://www.linx.net/`.

[119] NANOG Mailing list. More specifics from AS18978. `http://mailman.nanog.org/pipermail/nanog/2015-March/074348.html`, 2015.

[120] A. Lodhi, N. Larson, A. Dhamdhere, C. Dovrolis, and K. Claffy. Using PeeringDB to Understand the Peering Ecosystem. *CCR*, 2014.

[121] Aemen Lodhi, Nikolaos Laoutaris, Amogh Dhamdhere, and Constantine Dovrolis. Complexities in Internet Peering: Understanding the Black in the Black Art. *Infocom*, 2015.

[122] Woodbank Communications Ltd. Electricity Demand. `http://www.mpoweruk.com/electricity_demand.htm`.

[123] Paolo Lucente. PMACCT. `http://wiki.pmacct.net/`, 2015.

[124] Paolo Lucente and Elisa Jasinska. NetFlow and BGP multi-path: quo vadis? `https://www.nanog.org/sites/default/files/monday_general_lucente_netflow_32.pdf`, 2014.

[125] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, et al. AS relationships, customer cones, and validation. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 243–256. ACM, 2013.

[126] Andra Lutu, Marcelo Bagnulo, Jesus Cid-Sueiro, and Olaf Maennel. Separating wheat from chaff: Winnowing unintended prefixes using machine learning. *Proceedings of INFOCOM.*, 2014.

[127] Andra Lutu, Marcelo Bagnulo, and Rade Stanojevic. An economic side-effect for prefix deaggregation. In *Computer Communications Workshops (INFOCOM WKSHPS), 2012 IEEE Conference on*, pages 190–195. IEEE, 2012.

[128] Richard TB Ma, John Lui, and Vishal Misra. Evolution of the Internet Economic Ecosystem. *Networking, IEEE/ACM Transactions on*, 23(1):85–98, 2015.

[129] Doug Madory. Crecimiento in Latin America. `http://www.renesys.com/2013/05/crecimiento-in-latin-america/`.

[130] Massimiliano Marcon, Marcel Dischinger, Krishna P Gummadi, and Amin Vahdat. The Local and Global Effects of Traffic Shaping in the Internet. In *Proceedings of ACM SIGCOMM(poster*, 2008.

[131] J. Martin and A. Nilsson. On service level agreements for ip networks. In *Proceedings of IEEE INFOCOM*, 2002.

[132] Danny McPherson, Vijay Gill, Daniel Walton, and Alvaro Retana. Border gateway protocol (BGP) persistent route oscillation condition. *IETF RFC 3345*, 2002.

[133] David Meyer et al. University of oregon route views project, 2005.

[134] S Mirasgedis, Y Sarafidis, E Georgopoulou, DP Lalas, M Moschovits, F Karagiannis, and D Papakonstantinou. Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, 31(2):208–227, 2006.

[135] RIPE NCC. Réseaux IP Européens Network Coordination Centre (RIPE NCC) webpage. `http://www.ris.ripe.net/cgi-bin/lg/index.cgi`, 2013.

[136] Di Niu, Zimu Liu, Baochun Li, and Shuqiao Zhao. Demand forecast and performance prediction in peer-assisted on-demand streaming systems. In *INFOCOM, 2011 Proceedings IEEE*, pages 421–425. IEEE, 2011.

[137] W Norton. Internet Transit Prices - Historical and Projected, 2010.

[138] W Norton. The Value of an IXP. `http://drpeering.net/AskDrPeering/blog/articles/Ask_DrPeering/Entries/2011/8/18_The_Value_of_an_IXP.html`, 2011.

[139] W Norton. A Brief Study of 28 Peering Policies. `http://drpeering.net/AskDrPeering/blog/articles/Peering_Rules_of_the_Road_-_A_Brief_Study_of_28_Peering_Policies.htmls`, 2012.

[140] LINX NoVA. LINX NoVA. `https://www.linx.net/service/publicpeering/nova`, 2014.

[141] A. Nucci, A. Sridharan, and N. Taft. The problem of synthetically generating IP traffic matrices: initial recommendations. *ACM SIGCOMM Computer Communication Review*, 35:3, 2005.

[142] T. Oetiker. MRTG. urlhttp://oss.oetiker.ch/mrtg/.

[143] Konstantina Papagiannaki, Nina Taft, Zhi-Li Zhang, and Christophe Diot. Long-term forecasting of Internet backbone traffic. *Neural Networks, IEEE Transactions on*, 16(5):1110–1124, 2005.

[144] Jong Han Park, Ricardo Oliveira, Shane Amante, Danny McPherson, and Lixia Zhang. BGP route reflection revisited. *Communications Magazine, IEEE*, 50(7):70–75, 2012.

[145] Danny S Parker, Maria D Mazzara, and John R Sherwin. Monitored energy use patterns in low-income housing in a hot and humid climate. *Symposium on Improving Building Systems in Hot and Humid climates*, 1996.

[146] PCH. Packet Clearing Houswebpage webpage. `https://www.pch.net/`.

[147] PeeringDB. PeeringDB. `http://www.peeringdb.com/`.

[148] Peter Phaal and Marc Lavine. sflow version 5. *URL: http://www. sflow. org/sflow_version_5. txt, J uli*, 2004.

[149] Ingmar Poese, Benjamin Frank, Bernhard Ager, Georgios Smaragdakis, and Anja Feldmann. Improving content delivery using provider-aided distance information. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pages 22–34. ACM, 2010.

[150] R. Powell. XO Joins Level 3 For Bit-Mile Peering. `http://www.telecomramblings.com/2013/01/xo-joins-level-3-for-bit-mile-peering/`, 2013.

[151] Daniel J Power, Ramesh Sharda, and Frada Burstein. *Decision support systems*. Wiley Online Library, 2002.

[152] Bruno Quoitin, Cristel Pelsser, Louis Swinnen, Ouvier Bonaventure, and Steve Uhlig. Interdomain traffic engineering with BGP. *Communications Magazine, IEEE*, 41(5):122–128, 2003.

[153] Bruno Quoitin, Sébastien Tandel, Steve Uhlig, and Olivier Bonaventure. Interdomain traffic engineering with redistribution communities. *Computer Communications*, 27(4):355–363, 2004.

[154] Barath Raghavan, Martín Casado, Teemu Koponen, Sylvia Ratnasamy, Ali Ghodsi, and Scott Shenker. Software-defined Internet architecture: Decoupling architecture from infrastructure. In *HotNets*, pages 43–48. ACM, 2012.

[155] Y Rekhter, T Li, and S Hares. Border gateway protocol 4. *IETF RFC 4271*, 2006.

[156] P. Richter, G. Smaragdakis, A. Feldmann, N. Chatzis, J. Boettger, and W. Willinger. Peering at Peerings: On the Role of IXP Route Servers. *IMC*, 2014.

[157] RIPE. RIPE NCC RIS. `http://www.ripe.net/data-tools/stats/ris`, 2013.

[158] Keith Roe and Heidi Vandebosch. Weather to view or not: That is the question. *European Journal of Communication*, 11(2):201–216, 1996.

[159] Rahul Sami, Michael Schapira, and Aviv Zohar. Searching for stability in interdomain routing. In *Proc. INFOCOM*, 2009.

[160] Robert G Sargent. Validation and verification of simulation models. In *Simulation Conference, 2004. Proceedings of the 2004 Winter*, volume 1. IEEE, 2004.

[161] Stefan Savage, Andy Collins, Eric Hoffman, John Snell, and Thomas Anderson. The end-to-end effects of Internet path selection. In *ACM SIGCOMM Computer Communication Review*, volume 29, pages 289–299. ACM, 1999.

[162] Aaron Schulman and Neil Spring. Pingin'in the rain. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 19–28. ACM, 2011.

[163] John Scudder, Rex Fernando, and Stephen Stuart. BGP monitoring protocol. *draft-ietf-grow-bmp-07. Work in Progress. IETF Draft.*, 2012.

[164] Panagiotis Sebos, Jennifer Yates, Gisli Hjalmtysson, and Albert Greenberg. Auto-discovery of shared risk link groups. *Optical Fiber Communication Conference*, 4, 2001.

[165] Srinivas Shakkottai and Rayadurgam Srikant. Economics of network pricing with multiple ISPs. *IEEE/ACM Transactions on Networking (TON)*, 14(6):1233–1245, 2006.

[166] Rob Sherwood, Glen Gibb, Kok-Kiong Yap, Guido Appenzeller, Martin Casado, Nick McKeown, and Guru Parulkar. Flowvisor: A network virtualization layer. *OpenFlow Switch Consortium, Tech. Rep*, 2009.

[167] Georgos Siganos and Michalis Faloutsos. Analyzing BGP policies: Methodology and tool. In *INFOCOM 2004*, volume 3, pages 1640–1651. IEEE, 2004.

[168] Jesse H Sowell. Empirical Studies of Bottom-Up Internet Governance. *IETF*, 2012.

[169] Neil Spring, Ratul Mahajan, and David Wetherall. Measuring ISP topologies with Rocketfuel. In *ACM SIGCOMM Computer Communication Review*, volume 32, pages 133–145. ACM, 2002.

[170] R. Stanojevic, N. Laoutaris, and P. Rodriguez. On Economic Heavy Hitters: Shapley Value Analysis of 95th-percentile Pricing. In *Proceedings of ACM IMC*, 2010.

[171] Rade Stanojevic, Ignacio Castro, and Sergey Gorinsky. CIPT: using tuangou to reduce IP transit costs. In *Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies*, page 17. ACM, 2011.

[172] L. Subramanian, S. Agarwal, J. Rexford, and R.H. Katz. Characterizing the Internet Hierarchy from Multiple Vantage Points. In *Proceedings of IEEE INFOCOM*, 2002.

[173] Martin Suchara, Alex Fabrikant, and Jennifer Rexford. BGP safety with spurious updates. In *Proc. INFOCOM*, 2011.

[174] Cisco Systems. *Cisco Visual Networking Index: Forecast and Methodology*. Cisco White Paper, 2008.

[175] Renata Teixeira and Jennifer Rexford. A measurement framework for pin-pointing routing changes. In *Proceedings of the ACM SIGCOMM workshop on NetT*, 2004.

[176] T. Telkamp. Traffic characteristics and network planning. *NANOG 26*, 2002.

[177] UCLA. Internet topology collection. `http://irl.cs.ucla.edu/topology/`.

[178] Steve Uhlig and Olivier Bonaventure. Designing BGP-based outbound traffic engineering techniques for stub ASes. *ACM SIGCOMM CCR*, 34(5):89–106, 2004.

[179] Steve Uhlig and Sébastien Tandel. Quantifying the BGP routes diversity inside a tier-1 network. In *NETWORKING 2006. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems*, pages 1002–1013. Springer, 2006.

[180] Vytautas Valancius, Cristian Lumezanu, Nick Feamster, Ramesh Johari, and Vijay V Vazirani. How many tiers?: pricing in the internet transit market. In *ACM SIGCOMM Computer Communication Review*, volume 41, pages 194–205. ACM, 2011.

[181] Virginie Van den Schrieck, Pierre Francois, and Olivier Bonaventure. BGP add-paths: the scaling/performance tradeoffs. *Selected Areas in Communications, IEEE Journal on*, 28(8):1299–1307, 2010.

[182] Luis Velasco, Alberto Castro, Daniel King, Ori Gerstel, Ramon Casellas, and Victor Lopez. In-operation network planning. *Communications Magazine, IEEE*, 52(1):52–60, 2014.

[183] Patrick Verkaik, Dan Pei, Tom Scholl, Aman Shaikh, Alex C Snoeren, and Jacobus E Van Der Merwe. Wresting Control from BGP: Scalable Fine-Grained Route Control. In *USENIX Annual Technical Conference*, pages 295–308, 2007.

[184] Stefano Vissicchio, Luca Cittadini, and Giuseppe Di Battista. On iBGP routing policies. *IEEE/ACM ToN*, 2015.

[185] Stefano Vissicchio, Luca Cittadini, Maurizio Pizzonia, Luca Vergantini, Valerio Mezzapesa, and Maria Luisa Papagni. Beyond the best: Real-time non-invasive collection of BGP messages. *Proc. INM/WREN*, 2010.

[186] Ben Wagner and Patricia Mindus. Multistakeholder Governance and Nodal Authority - Understanding Internet Exchange Points. *NoC Internet Governance Case Studies Series*, 2015.

[187] H. Wang, C. Jin, and K. G. Shin. Defense Against Spoofed IP Traffic Using Hop-count Filtering. *ToN*, 2007.

[188] Hao Wang, Haiyong Xie, Lili Qiu, Yang Richard Yang, Yin Zhang, and Albert Greenberg. COPE: traffic engineering in dynamic networks. *ACM SIGCOMM Computer Communication Review*, 36(4):99–110, 2006.

[189] Yi Wang, Michael Schapira, and Jennifer Rexford. Neighbor-specific BGP: more flexible routing policies while improving global stability. In *Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems*, pages 217–228. ACM, 2009.

[190] Bernard L Welch. The generalization ofstudent's' problem when several different population variances are involved. *Biometrika*, pages 28–35, 1947.

[191] Matthias Wichtlhuber, Robert Reinecke, and David Hausheer. An SDN-Based CDN/ISP Collaboration Architecture for Managing High-Volume Flows. *Network and Service Management, IEEE Transactions on*, 12(1):48–60, 2015.

[192] Bill Woodcock and Vijay Adhikari. Survey of characteristics of Internet carrier interconnection agreements. *Packet Clearing House, May*, 2, 2011.

[193] Yang Wu, Mingchen Zhao, Andreas Haeberlen, Wenchao Zhou, and Boon Thau Loo. Diagnosing missing events in distributed systems with negative provenance. In *Proceedings of the 2014 ACM conference on SIGCOMM*, pages 383–394. ACM, 2014.

[194] Igal Zeifman. Report: Bot traffic is up to 61.5 `https://www.incapsula.com/blog/bot-traffic-report-2013.html`, 2014.

[195] Y. Zhang and et al. Fast accurate computation of large-scale IP traffic matrices from link loads. *Proceedings of ACM SIGMETRICS*, 2003.

[196] Mingchen Zhao, Wenchao Zhou, Alexander JT Gurney, Andreas Haeberlen, Micah Sherr, and Boon Thau Loo. Private and verifiable interdomain routing decisions. In *Proceedings of SIGCOMM*, pages 383–394. ACM, 2012.