Peering at Peerings: On the Role of IXP Route Servers

Philipp Richter
TU Berlin
prichter@inet.tu-berlin.de

Georgios Smaragdakis
MIT / TU Berlin
gsmaragd@csail.mit.edu

Anja Feldmann
TU Berlin
anja@inet.tu-berlin.de

Nikolaos Chatzis
TU Berlin
nikolaos@inet.tu-berlin.de

Jan Boettger
TU Berlin
jan@inet.tu-berlin.de

Walter Willinger
Niksun
wwillinger@niksun.com

ABSTRACT

During the last few years, more and more of the medium-to-large Internet eXchange Points (IXP) around the world have started to operate a *route server* and offer its use as a free value-added service to their members. This service has greatly simplified inter-domain routing for those members and has made it easy for them to peer with possibly hundreds of networks at those IXPs from the get-go.

In this paper, we report on an empirical analysis that is based on a unique collection of IXP-provided datasets from two different European IXPs that operate a route server and gave us access to a wealth of route server-specific BGP data. Both IXPs also made the traffic datasets that they routinely collect from their public switching infrastructures available to us. Using this information, we perform a first-of-its-kind study that correlates a detailed control plane view with a rich data plane view to reason about the different peering options available at these IXPs and how some of the major Internet players make use of them. In the process, we highlight the important role that the IXPs' route servers play for inter-domain routing in today's Internet and demonstrate the benefits of studying IXP peerings in a manner that is not agnostic but fully aware of traffic. We conclude with a discussion of some of the ramifications of our findings for both network researchers and operators.

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations—*Network Management*

Keywords

Internet Exchange Point (IXP); Peering; Routing; BGP.

1. INTRODUCTION

There are currently some 350+ Internet eXchange Points (IXP) worldwide, and some of the largest and most successful IXPs have more than 500-600 members and carry as much traffic as some of the global Tier-1 ISPs. With membership growth rates of 10-20% per year and annual traffic growth rates of 50-100%, these IXPs have emerged as key entities in the Internet infrastructure where a vast majority of today's peering connections are established. This growing importance of IXPs for the Internet peering ecosystem and the IXPs' increasing popularity with the full spectrum of Internet players have come in full view with recent studies such as [17, 21, 25, 36].

One of the main reasons for this emergence of IXPs as some of the most densely connected physical components in today's Internet has been the fact that many of them have started to offer the use of *route servers (RSes)*[1] as a free value-added service to their members. In essence, using an IXP's RS greatly simplifies routing for its members – a member AS has to set up just a single BGP session to connect to possibly hundreds of other member ASes. That is, most members use an IXP's RS to establish *multi-lateral* peerings (i.e., one-to-many) as compared to *bi-lateral* peerings (i.e., one-to-one). However, except for the fact that the peering fabrics of IXPs with RSes are significantly richer than of those that don't offer a RS, little is known in general about how IXPs operate RSes or how the members of such IXPs incorporate the use of RSes in their decision-making process that ultimately determines what sort of traffic is sent to/received from which of the other IXP members over what type of peering links.

The main contribution of this paper is an original and in-depth study of IXP RSes, from their architecture and deployment "in the wild" to their use by today's wide spectrum of Internet players. Our study combines a detailed control plane view of IXP-related peerings with a corresponding rich data plane view to examine how IXPs operate their RSes and, at the same time, reason about the different peering decisions that networks make when they connect with other networks at today's IXPs. To perform this analysis, we rely on our ongoing collaboration with an increasing number of European IXPs, which provided us access to a unique collection of datasets. Two of the IXPs operate a RS and gave us access to RS-specific BGP data and also made the traffic datasets that they routinely collect available to us.[2]

By correlating connectivity-related with traffic-specific aspects of peering at the level of individual links as well as routed prefixes, our analysis highlights the important role that the IXPs' RSes have started to play with respect to inter-domain routing in today's Internet. Among the key observations of our study are:

- RSes are key enablers for offering new and complex peering options to IXP members and also providing open access to a significant portion of Internet routes. At our IXPs, the

IMC '14, November 05–07 2014, Vancouver, BC, Canada.

http://dx.doi.org/10.1145/2663716.2663757.

[1]Note, while the term "route server" often refers to a "route collector/monitor" or "Routeviews server", we use it exclusively to refer to an IXP route server [29, 31] as defined in Section 2.2.

[2]We also have data from a third, small European IXP; however, since this small IXP does not operate a route server, its data is not a focus of this paper.

prefixes advertised via the RSes cover some 80-95% of the traffic and include many popular destinations.
- Due to the popularity of RSes, multi-lateral peering increasingly dominates classical bi-lateral peering in terms of number of peerings but not in terms of traffic; in fact, the majority of the traffic at our IXPs traverses bi-lateral peerings. Nevertheless, some of the top traffic contributors at each IXP use mainly multi-lateral peering.
- RS usage varies across IXPs and IXP members; while we observe certain patterns of RS usage among IXP members with similar business types, we illustrate the wide range of existing peering strategies with a number of case studies involving key Internet players as well as networks with interestingly complex peering setups.

These findings have ramifications for a number of recent developments that involve network operators and researchers alike. In particular, as many network operators are eagerly experimenting with and scrutinizing the wealth of peering alternatives that IXPs provide (e.g., to reduce transit costs, increase performance and robustness), a number of researchers are either exploring the evolution of the peering ecosystem as a whole or examine ways to innovate inter-domain routing. We contribute to each of these developments by providing an in-depth assessment of peering at IXPs with an emphasis on understanding and highlighting the role of IXP RSes in these different efforts.

For example, the recently launched Open-IX initiative [11] aims at establishing European-style IXPs in the USA. Its main objective is to increase the diversity in available peering options for networks in the US where the cost associated with the dominant form of interconnection (i.e., private peering or cross connects) is almost six times as much as in Europe [15], mainly because of limited competitive interconnection options. To understand the current Internet peering ecosystem and inform studies concerned with its evolution over time, our work provides a first in-depth look into a marketplace (i.e., Europe) where, thanks to the popularity of IXP RSes, networks have in general an abundance of interconnection options and have already acquired certain skills in using the available options strategically in support of their particular business.

At the same time, researchers have recently recognized the importance of the growing popularity of IXPs for trying to fundamentally improving inter-domain routing. In fact, the main premise of the work on Software Defined eXchanges (SDX) [27] is that IXPs are ideal venues for deploying new technologies such as Software Defined Networking (SDN) – only the mechanisms used at the IXP itself (e.g., RS operation) need to be changed while the rest of the Internet can remain unchanged. As a first account of practical RS usage "in the wild" that illustrates the innovative use of today's RSes for inter-domain routing, our work complements and informs these recent efforts. In particular, it contributes to the identification of new IXP service offerings that are beyond the capabilities of today's RSes but can be fully supported when relying on new technologies such as SDN.

Moreover, for the Internet measurement community, we demonstrate how our proprietary datasets can be used "for the good of science"; that is, calibrating the quality of the latest BGP measurements that are becoming publicly obtainable via Looking Glasses on some IXP RSes. By showing what parts of the ground truth of the peering fabrics of those IXPs are visible in these publicly available datasets and observing that the vast majority of the observed IXP traffic is destined to the RS prefixes, we significantly increase the value that these publicly available datasets will have for future studies by third parties.

The rest of the paper is structured as follows. Section 2 is an introduction to RS design and operation. In Section 3, we describe the different sets of control and data plane measurements that are available to us. We use this data in Section 4 to revisit connectivity-related issues and rely on it in Section 5 to study usage-related aspects of peerings. Moving beyond the traditional link perspective of peerings, we examine in Section 6 peering opportunities and traffic at the level of routed prefixes. We provide a longitudinal study of peering at IXPs in Section 7. We then study how different players implement peering in Section 8. We conclude in Section 9 and 10 with a discussion of trends in Internet peering and related open research problems.

2. TRENDS IN IXP OPERATIONS

IXPs offer a shared switching fabric where the members can exchange traffic with one another once they have established peering connections (public peering) between them. This generic service provided by IXPs is an example of a *(positive) network effect* (also known as *network externality* [34]) because the more members an IXP has, the more valuable that IXP is to each member. Especially in Europe where many IXPs operate on a not-for-profit basis [21], this observation has led to significant innovations at IXPs in the form of constantly expanding service offerings.

2.1 Background and motivation

An important example of a new IXP service offering has been the free use of IXP-operated RSes. With more IXPs operating RSes, the membership base of those IXPs keeps growing and more of their member ASes start using the RS service. The resulting proliferation of multi-lateral peerings creates new options for the member ASes to peer with one another which, in turn, causes them to reconsider their existing peering arrangements. At the same time, the ease with which networks can join IXPs and start using them has led to a trend whereby one and the same AS can be a member at multiple IXPs, often within the same city or geographic region.

In effect, offering the free use of their RSes has created a positive feedback loop for those IXPs. With more networks connecting to such an IXP, more member ASes use its RS and connect with most or all other members at the RS. This increases connectivity and typically also traffic, which in turn attracts yet more networks to join the IXP. Thus, while we are observing an Internet peering ecosystem in flux, a pre-requisite for studying the current system to ultimately reason about its evolutionary path is to have a basic understanding of "all things concerning RSes." Given the central role that IXP RSes play in this paper, we describe in this section the basic operations of an IXP RS and discuss the main reason for the observed bandwagon effect.

2.2 The IXP RS as enabler

The typical way to establish connectivity between two ASes is to establish a direct BGP session between two of their respective border routers. Initially, if two IXP member ASes wanted to exchange traffic via the IXP's switching fabric, they had to establish a *bi-lateral (BL)* BGP peering session at the IXP. However, as IXPs grew in size, to be able to exchange traffic with most or all of the other member ASes at an IXP and hence reap the benefits of its own membership, a member's border router had to maintain more and more individual BGP sessions. This started to create administrative overhead, operational burden, and the potential of pushing some router hardware to its limit.

To simplify routing for its members, IXPs introduced RSes and offered them as a free value-added service to their members. In short, an IXP RS is a process that collects routing information from

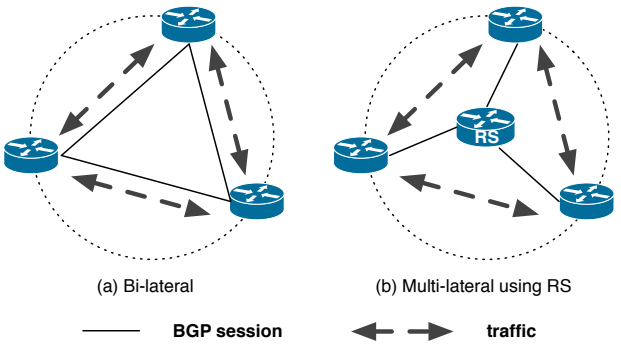


Figure 1: IXP peering options.

the RS's peers or participants (i.e., IXP members that connect to the RS), executes its own BGP decision process, and re-advertises the resulting information (i.e., best route selection) to all of the RS's peer routers. Figure 1 shows the flow of control plane information (BGP sessions) and data packets (data plane) for both traditional bi-lateral peering as well as peering via the RS at an IXP. The latter is referred to as *multi-lateral (ML)* peering because it typically involves more than two BGP partners.

A member AS just connects to the RS via a single BGP session to set up BGP peering with all other IXP members that peer with the RS.[3] Clearly, this lowers the maintenance overhead, in particular for small ASes. Note, however, that using the RS (i.e., ML peering) does not preclude BL peering by one and the same member AS. In particular, larger ASes can take advantage of the RS while still having the option to establish BL peerings with selectively-chosen IXP members. For example, if a large member AS finds the capabilities of the RS to be insufficient for its needs (e.g., with respect to traffic engineering or filtering) or prefers to have more control over the peerings with its most important peers, it can use BL peerings with the latter and ML peerings with those members that peer with the IXP's RS.

Note that the IXP RS is not involved in the data path. Moreover, it executes the BGP decision process based on the information it learns from its peers. Thus, there is no guarantee that the best route selected by the RS is the same as the one that the peer would have selected given a full set of BL peering sessions. In this sense, while using a RS solves a scalability problem, it has the potential of creating a new problem known as the *hidden path problem* [31]. For example, if different peers of the IXP's RS advertise routes for the same prefix, only one route will be selected and re-advertised to all other peers. If this route cannot be propagated to a certain peer (e.g., because of export filtering restrictions), that peer will not be able to receive any route for this prefix, even though an alternative path was advertised to the RS by a different peer. While this hidden path problem was an issue with earlier versions of RS software, recent advances have largely overcome the problem, and in Section 2.4, we describe a popular Internet routing daemon that addresses this problem by providing support for maintaining peer-specific routing tables. We recall that a similar problem also occurs in the context of large ISPs where *route reflectors* are used to redistribute routes to iBGP peers [20].

2.3 On the deployment history of RSes

The first RSes were designed and deployed in the US as part of the decommissioning of the NSFNET around 1995 [26]. However,

[3]For redundancy purposes, IXPs typically operate two RSes and encourage their members to connect to both RSes.

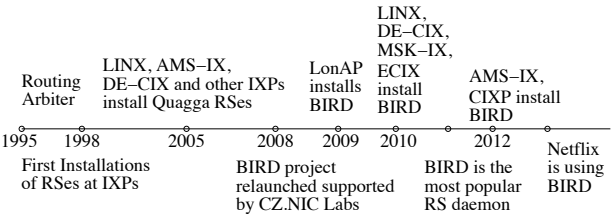


Figure 2: Route server deployment time line.

the service never took off due to the small size of the largest US IXPs at that time (i.e., MAE-East's membership approached some 20 ASes in mid-1990).[4]

Although some IXPs (e.g., DE-CIX in Frankfurt) offered the use of a route server as early as mid-1990, it took another 10 years and the increasing popularity of some European IXPs for the RS concept to re-emerge and gain momentum. In fact, around 2005 an increasing number of European IXPs already offered RS capabilities to their members. The de-facto RS at that time was Quagga [14]. However, early versions of Quagga suffered from the hidden path problem and had numerous operational problems (e.g., prefix filtering, resource consumption, performance limitations) [30, 32]. Later versions of Quagga and OpenBGPD [12] addressed most of these problems and started to be deployed around 2008. Difficulties with its maintenance prevented OpenBGPD from being widely adopted and deployed [38].

In 2008, the BIRD [16] project was relaunched (the initial project dates back to 1999), and it took just three years for BIRD to become the most popular IXP route server [4]. Figure 2 shows a time line when some of the larger IXPs migrated to BIRD. In fact, the success of BIRD goes beyond RSes. Since 2013, it is the core routing component of the Netflix Open Connect Appliance [10].

2.4 BIRD: A popular router daemon for RSes

BIRD is an open-source software routing daemon. The project was launched in 2008. The software was developed by CZ.NIC Labs and has been actively supported by the IXP community. In the following, we describe a BIRD configuration that has been abstracted from the Euro-IX RS example [4] and is the basis of the one in operational use by one of the IXPs with which we have an ongoing collaboration.

Like all routing daemons, BIRD maintains a Routing Information Base (RIB) which contains all BGP paths that it receives from its peers – the Master RIB. However, when using BIRD as RS, it can be configured to (i) maintain peer-specific RIBs and (ii) use them instead of the Master RIB for peer-specific BGP best path selection (see Figure 3). When configured this way, each member AS that peers with the RS maintains a BGP session with the RS, which results in a peer-specific RIB. When IXP member AS X advertises a prefix to the RS, it is first put into the AS X-specific RIB. Next, if this prefix passes the AS X-specific import filter, it is added to the RS' Master RIB. If the export filter of AS X allows it, then this prefix will also be added to each AS Y-specific RIB, where AS Y is any other IXP member that connects to the RS. Then the RS performs a peer-specific best path selection and exports the prefix by re-advertising it to each AS Y.

IXPs typically apply import filters to ensure that each member AS only advertises routes that it should advertise. To derive import filters, the IXPs usually rely on route registries such as IRR [7]. This policy limits the likelihood of unintended prefix hi-

[4]The RIPE RS [19] was a precursor of the RS proposed in [26].

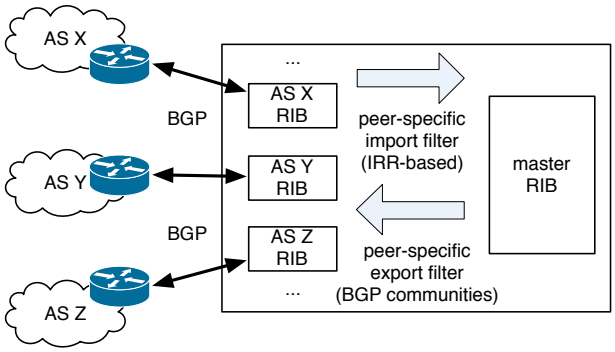


Figure 3: BIRD route server: Example setup with peer-specific RIBs and import/export filtering.

jacking and/or advertisements of bogon prefixes including private address space. With respect to export filters, they are typically triggered by the IXP members themselves to restrict the set of other IXP member ASes that receive their routes. The commonly used vehicle for achieving this objective is to tag route advertisements to the RS with RS-specific BGP community values [24]. These values are set on a per route basis and restrict to which members the route can be propagated. Thus, by using export filters, peers of the RS can express policies.

Using peer-specific RIBs overcomes the hidden path problem, because the BGP decision process is executed for each peer independently. If the best route for a certain prefix cannot be readvertised to some particular members (e.g., because of export filters), the RS can still select and advertise an alternative route to those particular members – presuming that another, non-blocked route is available.[5] Maintaining separate RIBs for each RS peer has a cost in terms of memory requirements and processing capabilities. However, recent stress tests and real deployment performance reports for BIRD at large IXPs [22, 33, 32] show that maintaining about 400 RIBs, each with up to 100K prefixes, consumes roughly 4 GB of memory and can be achieved with commodity servers.

2.5 Looking glasses and RSes

The fact that network operators typically cannot see the results of their route announcements within their own network makes debugging BGP routing difficult. To address this problem, the community has set up BGP looking glasses (LGes) and route monitors (RMes) in many different locations across the Internet. LGes are user-interfaces, co-located on routers. Many of them are publicly accessible and can be used by anyone to execute a limited number of commands. On conventional LGes the commands include ping, traceroute, as well as some BGP commands. The latter provides details about all BGP sessions of the router. In the past, LGes have also been used by researchers to gain better visibility into the AS-level topology and IXP-specific peering fabrics [28, 18].

LGes can also be co-located with RSes at IXPs. In this case, the LGes act as proxies for executing commands against the Master RIB of the RS and are equipped with additional capabilities that may include commands which list (a) all prefixes advertised by all peers and/or (b) the BGP attributes per prefix. Such LGes (referred to as LG-RS below) are already available at some IXPs, including DE-CIX and LonAP, and they have recently been used to explore AS connectivity at IXPs [25].

[5]Also for iBGP route reflectors, the advantage of using peer-specific RIBs and independent decision processes is discussed [40].

IXP	L-IXP	M-IXP	S-IXP	L-IXP & M-IXP
Member ASes	496	101	12	50
Tier-1 ISPs*	12	2	1	2
Large ISPs*	35	4	1	4
Major Content/ Cloud Providers*	17	5	0	5
RS (IPv4/IPv6)	BIRD Multi-RIB	BIRD Single-RIB	No	/
Public RS-LG (IPv4/IPv6)	Yes	Yes, limited commands	No	/
Member ASes using the RS	410	96	/	43

Table 1: IXP profiles: members and RS usage.

3. DATA SOURCES

In this section, we describe the European IXPs with which we have fruitful ongoing collaborations and that have been very forthcoming with sharing their IXP-internal datasets with us. We provide details about these proprietary datasets as well as the public IXP-external datasets that are at our disposal for this study.

3.1 Profiles of our IXPs

We focus in the following on the two IXPs (denoted L-IXP and M-IXP) that operate a RS. These IXPs differ greatly in size (e.g., number of members and traffic) and service offerings that they provide for their members. For comparison, Table 1 that summarizes some key features of our IXPs and their members also includes some basic information about our third IXP (denoted S-IXP). This IXP is representative of the large number of smaller IXPs (i.e., less than 20 members) that do not operate a RS. However, for the remainder of this paper, we will only consider L-IXP and M-IXP and will not use the S-IXP.

Large IXP (L-IXP): With close to 500 members and peak traffic that exceeded 3 Tbps in late 2013, this IXP is one of the largest IXPs in Europe and worldwide. It operates a layer-2 switching fabric that is distributed over a number of colocations/datacenters within a metropolitan area. Members can connect on ports ranging from 1 to 100 Gbps. Its membership includes the full spectrum of networks, including CDNs, content providers, hosters and the whole range of ISPs (from Tier-1 to regional to local providers). This IXP provides arguably one of the richest and most advanced service offerings, including SLAs, remote peering,[6] and black-holing (mitigation of distributed denial of service attacks e.g., [8]). One of its key service offerings is the free use of its RS. It operates two (for redundancy) IPv4 and IPv6 RSes using a BIRD configuration similar to the one discussed in Section 2.4 with advanced LG support.

Medium IXP (M-IXP): This medium-sized IXP operates a layer-2 switching fabric, and is present in several locations. Half of its members are also members at the L-IXP. As of late 2013, this IXP had 101 members and its peak traffic exceeded 250 Gbps. It offers ports ranging from 1 to 100 Gbps. Among service offerings such as remote peering, it also offers the free use of its RS (again, two for each IPv4 and IPv6). However, this RS does not support peer-specific RIBs, and its LG allows the execution of only a limited set of commands.

[6]The router of the "remote peer" can be located anywhere in the world and is connected to the IXP typically via an Ethernet-over-MPLS connection e.g., [2].

3.2 IXP-internal data: Route server

For both IXPs that operate a RS (i.e., L-IXP and M-IXP), we have access to the data from their BIRD deployment. In the case of L-IXP, we have weekly snapshots of the peer-specific RIBs, starting in June 2013. For M-IXP, we have several snapshots of the Master-RIB, starting in December 2013. In addition, for L-IXP, we have all BGP traffic to and from its RS that is captured on the network interface via tcpdump.

The unique advantage of having access to these IXP-provided control plane measurements is that they are rich enough to accurately and completely reconstruct the "ground truth" in terms of connectivity at these IXPs that has been established with the help of the RS, i.e., the IXPs' multi-lateral peering fabrics.

3.3 IXP-internal data: Traffic

For each of our IXPs, we have access to data plane measurements in the form of traffic that is routinely collected from the IXPs' public switching infrastructures. More precisely, for each IXP, the available datasets consist of massive amounts of sFlow records [42], sampled from their public switching infrastructure. The measured sFlow records contain Ethernet frame samples that have been collected using random sampling (1 out of 16K). sFlow captures the first 128 bytes of each sampled frame. Thus, they contain full Ethernet, network- and transport-layer headers, as well as some bytes of payload for each sampled packet. For further details about relevant aspects of these collected sFlow records (e.g., absence of sampling bias, removal of irrelevant traffic), see [17].

We rely on 4 continuous weeks of collected sFlow for each IXP. For L-IXP we cover a 4-week period in August/September 2013, and for M-IXP we cover a 4-week period in December 2013.[7] In addition, we also use snapshots collected at the L-IXP that date back to 2011 and cover periods of two weeks.

Having access to these IXP-provided data plane measurements makes it possible to examine the connectivity that has been established without the use of the IXPs' RSes (i.e., bi-lateral peerings). However, more importantly, these measurements provide valuable but hard-to-obtain information about how the two parties of an IXP peering use that link and for what purpose.

3.4 Public IXP-specific data

In our work, we also rely on a number of widely-used public datasets, including BGP data from the route collectors from RIPE RIS, Routeviews, PCH [13], and Abilene [6]. We refer to them as RM BGP data. Moreover, we use the data from the RS-LG at the L-IXP in accordance with the method described in [25]. In addition, we rely on data obtained from several publicly accessible LGes which query the routing tables of routers that belong to members peering at the L-IXP.

4. CONNECTIVITY: BI-/MULTI-LATERAL

Relying on our IXP-provided measurements, we show in this section how we can get close to recovering the actual peering fabrics at the IXPs. We also illustrate what portions of those actual peering fabrics can and cannot be recovered when using the different public BGP data.

[*]The numbers reflect only the well-known and easy-to-identify networks by business type to illustrate the existence of common members. No attempt at a complete classification has been made.

[7]To assess a possible bias due to the holiday season, we performed our analysis also using only the first week in December and got consistent results.

Multi-lateral Peerings: RS RIBs				
	IPv4		IPv6	
	symmetric	asym.	symmetric	asym.
L-IXP	65,599	14,153	34,596	5,086
M-IXP	3,140	594	1,173	434
Bi-lateral Peerings: Inferred from Data-Plane				
	bi-/multi	bi-only	bi-/multi	bi-only
L-IXP	14,673	5,705	4,256	3,727
M-IXP	399	61	223	75
Total Peerings				
L-IXP	85,457 (70%)		43,409 (35%)	
M-IXP	3,795 (75%)		1,682 (33%)	
Visibility in the RS Looking Glass				
L-IXP	all multi-lateral		all multi-lateral	
M-IXP	none		none	

Table 2: Multi-lateral and bi-lateral peering links.

4.1 Connectivity: IXP-provided data

To determine if IXP members AS X and AS Y are using a ML peering at the IXP, we rely on the IXP-provided RS data. More specifically, for the L-IXP, we first check if AS X and AS Y peer with the RS. If so, we next check in the peer-specific RIB of AS Y for a prefix with AS X as next hop. If we find such a prefix, we say that AS X uses a ML peering with AS Y. If we also find AS Y in the peer-specific RIB of AS X as next hop, we say that the ML peering between AS X and AS Y is *symmetric or bi-directional*; otherwise, we say that the ML peering between AS X and AS Y is *asymmetric*.

Given that the RS at the M-IXP only uses the Master RIB but no peer-specific RIBs, we re-implement the per-peer export policies based upon the Master RIB entries to determine peerings via the RS. More specifically, if there is a route for a prefix in the Master RIB with AS X as next hop, we postulate a ML peering with all member ASes that peer with the RS, including AS Y, unless the community values associated with the route explicitly filter the route via the peer-specific export filter to AS Y.

To determine if IXP members AS X and AS Y are using a BL peering at the IXP, we rely on the IXP-provided traffic measurements. In particular, to conclude that AS X and AS Y established a BL peering at the IXP, we require that there are sFlow records in the IXP-provided traffic data that show that BGP data was exchanged between the routers of AS X and AS Y over the IXP's public switching infrastructure.[8] We cannot however differentiate between asymmetric and symmetric BL peerings with these data plane measurements.

Note that our methodologies yield a lower bound for BL peerings and an upper bound for ML peerings, but there is evidence that these bounds are in general very tight. For example, with respect to BL peerings, our method is not significantly biased by the sFlow sampling rate because the numbers are very stable once we use data from more than two weeks. Indeed, Figure 4 shows that for the L-IXP, the additional BL peerings seen in the third (fourth) week are less than 1% (0.5%). As far as ML peerings are concerned, our method does not account for the fact that some RS peers might reject the advertisements of the RS, which can result in some overcounting by our method. At the same time, we find member ASes that use one and the same link for ML as well as for BL peering.

Our best efforts to reconstruct the actual ML and BL peering fabrics of our IXPs is summarized in Table 2. We further break

[8]The routers' IP addresses have to be within the publicly known subnets of the respective IXP.

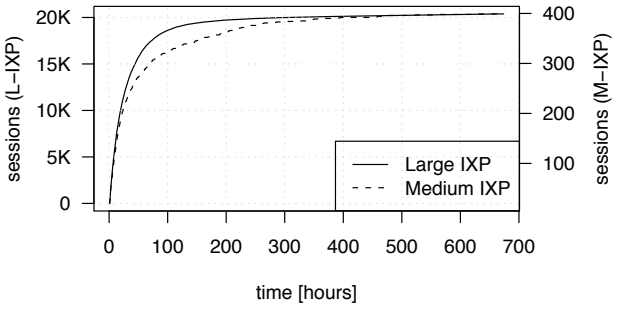


Figure 4: Inferred bi-lateral BGP sessions over time.

down (where possible) each of the ML and BL peering fabrics into links that are used for either IPv4 or IPv6 and in a symmetric or asymmetric manner. For each IXP, we also tally the total number of peerings along with the peering degree (percentage of established peering links compared to the number of possible peering links).

4.2 Connectivity: Public BGP data

To assess the ability of the various public BGP datasets to provide visibility into the peering fabrics of our IXPs and compare with our findings in Section 4.1 above, we first mine the RS-LG BGP data that has recently been considered and used in [25] and is publicly available. The results are shown in Table 2. There are three main take-aways. First, when restricting the study to the L-IXP that has an RS-LG with the capabilities listed in Section 2.5, the new methodology proposed in [25] is indeed successful in recovering the full-fledged ML peering fabric of that IXP. Second, as the example of the M-IXP shows, having a RS-LG that lacks the necessary querying capabilities is of no help for inferring the IXP's ML peering fabric.[9] Third, irrespective of the capabilities of the RS-LG, the RS-LG BGP data cannot be used to reveal any peerings of the L-IXP's and M-IXP's BL peering fabrics.

For completeness, we also obtained the number of peerings that can be discovered from mining the traditional and widely-used RM BGP data. The results confirm past findings that a majority of the peerings (i.e., some 70-80%) cannot be seen in those data [17, 18, 24, 41]. In addition, we also notice a significant bias in this data towards BL peerings. Interestingly, this data does produce peerings between IXP member ASes that we do not see even in our most complete peering fabrics discussed in Section 4.1. One possible explanation for seeing such peerings in the public but not in the IXP-provided BGP data is that the member ASes in question engage in *private peering* at the IXP[10] or peer in a different location.

Summary

We observe a dense peering mesh at our IXPs, with ML peerings outnumbering BL peerings by a ratio of 4:1 and 8:1 at the L-IXP and the M-IXP respectively. Thus, connectivity at these IXPs is clearly driven by their RSes and the resulting ML peerings. For IPv6 connectivity we note that at both IXPs, the number of IPv6 peerings is roughly half of the number of IPv4 peerings. LGes co-located with IXP RSes are able to uncover the full ML peering fabric, assuming they support an advanced set of commands. BL peerings remain hidden and can only partly be inferred using public RM BGP data.

[9]Some portion of the ML peering fabric may be recovered but only with substantial additional efforts. For example, by using prefixes extracted from the RIPE RIS or Routeviews data, a portion of the IPv4 prefixes on the Master RIB of the M-IXP can be recovered and used to submit queries to this RS-LG.

[10]Private peerings at IXPs are handled completely separate from the IXP's public switching infrastructure and are not part of our study.

5. FROM CONNECTIVITY TO TRAFFIC

In this section, we are mainly interested in how many of the established peerings reported in Table 2 are actually "used"; that is, see traffic. Furthermore, we are interested in the actual distribution of traffic when taking the type of the peering link into account.

5.1 Identifying traffic-carrying peerings

To identify a traffic-carrying peering between AS X and AS Y, we look for sFlow records that (i) contain MAC addresses which belong to AS X and AS Y, respectively, and (ii) have IP addresses that are not part of the IP address space assigned to the IXP. Thus, we only count the exchange of non-local IP traffic, which allows us to clearly separate control traffic (i.e., BGP sessions) and actual data traffic, thus we can distinguish between BL peerings with and without traffic. Once we identified a traffic-carrying peering link, we assign it to be either BL or ML, depending on our earlier introduced inference. For a small portion of the traffic (less than 0.5% for both IXPs) we did not find a corresponding BL or ML peering link. We discard this traffic from our analysis.[11]

In this context, for IXP member ASes that peer with other member ASes at the IXP both bi-laterally and multi-laterally, we are faced with the problem of determining whether the observed traffic between two such ASes is traversing the BL or ML peering link between them; that is, identifying the traffic-carrying peering(s). Taking a pragmatic approach, when two IXP member ASes peer with one another at the IXP both bi-laterally and multi-laterally, we tag the BL peering between them as the traffic-carrying peering and associate any observed traffic with it. Intuitively, our argument for this approach is based on the observation that compared to ML peering, establishing a BL peering requires work (e.g., manually setting up BGP sessions) and is an indication of joint incentives and needs between the involved parties. On the other hand, peering multi-laterally at the IXP (i.e., using the RS) is designed to be easy and informal, making it in general possible to exchange traffic with all the RS's peers from the get-go.

To provide empirical support for our argument, we manually searched for LGes that query the routing tables of member routers that peer both bi-laterally and multi-laterally with other members. We found six such LGes with sufficient capabilities to reason about the best path selected. In all cases, advertisements via BL sessions were selected as best path over advertisements from the RS.[12]

5.2 Traffic over ML and BL peerings

The results of our analysis of the traffic-carrying links are summarized in Table 3. When compared to Table 2, the first column in Table 3 shows that most peering links (i.e., more than 80% at both IXPs) are actually "used" in the sense of the binary attribute "carry traffic/no traffic". Moreover, we note that the ratio of traffic-carrying peerings is largest for BL peerings, followed by symmetric ML peering, followed by asymmetric ML peering.

Moving beyond this binary classification of peering links, this usage picture sharpens when examining the second column of Table 3. This column shows for each IXP the number of peerings that are responsible for 99.9% of the IXP's total traffic; that is, all

[11]Possible explanations for this traffic are either non-detected BGP sessions or peerings using protocols other than BGP (e.g., static routing).

[12]Selection of BL over ML was typically done by setting the local preference to a higher value for routes received via BL sessions. However, we point out that is not necessarily true for all peerings.

IPv4	L-IXP		M-IXP	
	all	99.9p	all	99.9p
% BL	92.4	55.6	93.5	47.7
% ML sym.	85.9	31.3	83.7	24.0
% ML asym.	23.8	5.43	38.5	7.89
links total	67,915	28,849	2,968	918
IPv6	**L-IXP**		**M-IXP**	
	all	99.9p	all	99.9p
% BL	76.2	4.92	74.9	7.17
% ML sym.	54.0	0.52	52.2	0.48
% ML asym.	30.4	0.04	25.3	0.07
links total	24,159	556	819	24

Table 3: Percentage of links that carry traffic (all traffic vs. top 99.9% of all traffic) and their corresponding type.

peerings that collectively see less than 0.1% of the overall traffic are discarded.[13] When imposing such thresholds to eliminate peerings that carry only comparably little amounts of traffic, we observe a drastic reduction of the number of active peerings. Indeed, the main take-away from this thresholding exercise is that it puts the connectivity-related findings reported in Section 4.2 into proper perspective. Specifically, it demonstrates that while RSes increase connectivity and are responsible for the larger part of peerings, a majority of those ML peerings typically does not carry much traffic. At the same time, the smaller number of BL peerings that are established at IXPs carry in general the bulk of the traffic.

As a by-product of this thresholding, we also notice that essentially none of the large number of IPv6 peerings that have been established at our IXPs carry any significant traffic. Thus, while connectivity for IPv6 is there, traffic still ranges in the order of less than 1% at each of our IXPs. As a result, for the remainder of this paper, we will focus exclusively on IPv4 traffic that traverses the IXPs' switching infrastructures. Nevertheless, it is worthwhile to mention that some links are only present for IPv6 but not for IPv4.

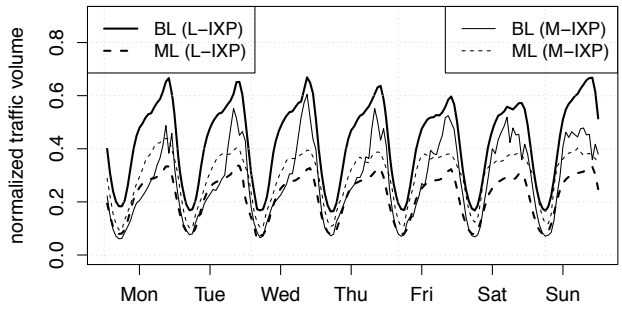
To assess the overall contribution of BL and ML peerings, we map the total traffic of the L-IXP and M-IXP onto the BL and ML peerings that have been established at each of these IXPs. Figure 5(a) shows a time series for the traffic on BL and ML peerings at both IXPs over a one-week period. We observe that even though there are much fewer BL peerings than ML peerings at L-IXP, the total amount of BL traffic is more than twice the total amount of ML traffic. To contrast, for M-IXP, this BL:ML traffic ratio is more like 1:1.

Even though this traffic-centric picture of IXP peerings emphasizes the significance of BL peerings, ML peerings do play an important role. Figure 5(b) shows a CCDF of the distribution of traffic over BL/ML peerings. As far as the top traffic-contributing peerings are concerned, BL and ML peerings contribute similar amounts of traffic. In fact, for both the L-IXP and the M-IXP, the top traffic-contributing peering is a ML peering which highlights the critical role that the use of IXP RSes play in today's Internet.

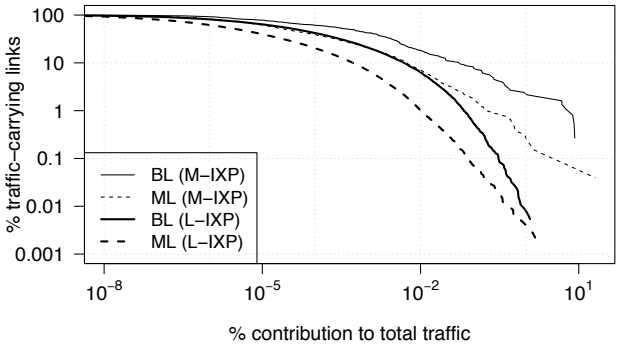
Summary

While most peerings at our IXPs (e.g., some 80%) do carry traffic, BL peerings are more likely to carry traffic when compared to ML peerings. Taking traffic volumes into account, the small number of BL peerings clearly dominate the large number of ML peerings, with observed traffic ratios of 2:1 (L-IXP) and 1:1 (M-IXP). Thus, the majority of ML peerings only carry small amounts of traffic,

[13]Note that the corresponding thresholds still require a monthly traffic volume of some tens resp. hundreds of GBs at each IXP.



(a) Timeseries of traffic over BL/ML links.



(b) CCDF: Traffic distribution over BL/ML links.

Figure 5: Traffic: bi-lateral (BL) and multi-lateral (ML).

but some of the top traffic-contributing links at both IXPs are ML peerings.

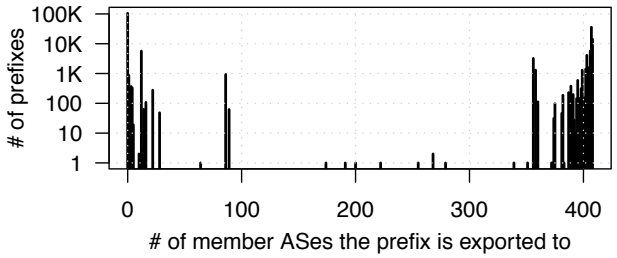
6. FROM TRAFFIC TO ROUTES: PREFIXES

In this section, we move beyond the link perspective of peerings and peering traffic and instead examine peerings at the level of routed prefixes. We show how this new perspective allows us to reason about peering opportunities at IXPs and the different peering options chosen by member ASes.

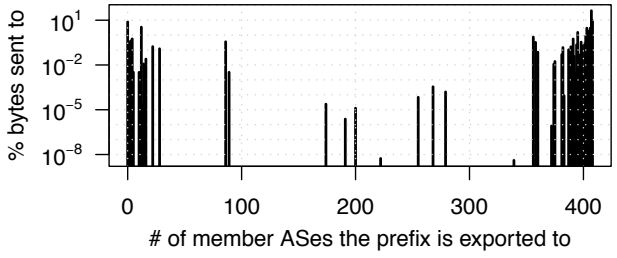
6.1 A prefix view of peering

An outward sign of the popularity of the RS service offering with the member ASes of the L-IXP and M-IXP is the large number of ML peerings that have been established at those IXPs. To further understand what set of routes RSes at popular IXPs offer – e.g., to understand the instant benefits that a new member AS can reap when connecting to the IXP's RS – we check each prefix that is advertised via the RS and count to how many of the RS's peers this prefix is exported. For the L-IXP, Figure 6(a) shows the resulting histogram (y-axis in log-scale) and reveals a striking bi-modality – either a prefix is exported to almost all members or to only a few selected ones. Focusing on the mode to the right, the sizable fraction of prefixes that is exported to almost all member ASes is a reflection of the very peering-friendly policies at the IXP. Summing up all the prefixes that are exported by the RS to more than 90% of its peers shows that a new IXP member will receive more than 65K routes from some 11K different origin ASes at the L-IXP and more than 12.5K routes from about 3K different origin ASes at the M-IXP (see Table 4) once connected to the RS.

As far as the mode to the left in Figure 6(a) is concerned, the sizable fraction of prefixes that are exported to fewer than 10% of the RS's peers is an indication that the strategy offered by the IXPs to restrict (block) the propagation of certain routes to certain members is effective and used. Upon closer inspection, we also observe that



(a) Number of prefixes.

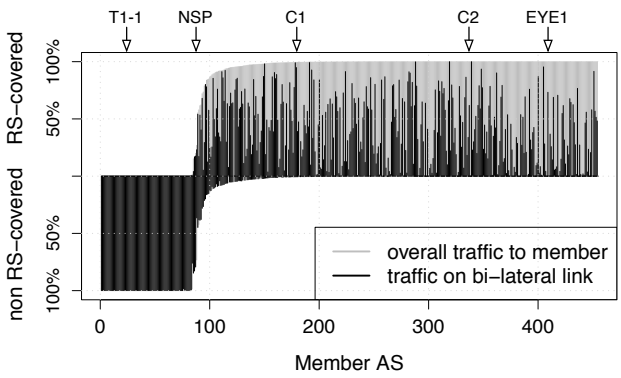


(b) Traffic share.

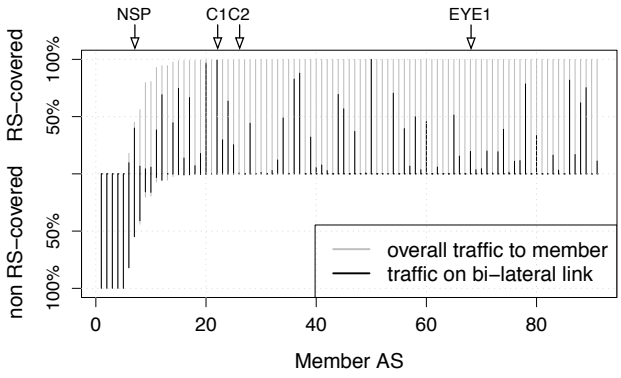
Figure 6: Prefixes advertised via the RS in relation to the number of member ASes they are exported to (L-IXP).

	L-IXP		M-IXP	
Export to % of peers	$< 10\%$	$> 90\%$	$< 10\%$	$> 90\%$
Prefixes	112.5K	68.0K	171	12.6K
/24 Equivalent	1.97M	819K	7.4K	337K
Origin ASes	13.06K	11.1K	44	3.0K

Table 4: Breakdown of advertised IPv4 address space.



(a) L-IXP.



(b) M-IXP.

Figure 7: Traffic to member: to RS prefixes (upper half in each plot) / non-RS prefixes (lower half in each plot) with ML/BL traffic grey/black.

the sets of reachable origin ASes for the mode to the left and mode to the right are largely disjoint. For M-IXP (not shown), we note that while the number of advertised prefixes and the corresponding reachable address space is much smaller, the vast majority of prefixes are exported to almost all peers.

To illustrate, of the 408 member ASes that connect to the L-IXP RS, we find that only some 24 limit the export of some or all of their prefixes. However, 371 members export their prefixes to more than 90% of the members participating in the RS. At M-IXP, we see an even more peering-friendly environment with only very few members applying strict export filtering for a small number of prefixes.

6.2 A prefix view of traffic

To complement the connectivity-centric prefix-level, we are now interested in how this corresponds to actual traffic exchanged. By matching all destination IP addresses of traffic exchanged (irrespective of the link type) on the aggregate of RS prefixes, we see that more than 80% of the overall traffic at L-IXP (95% at M-IXP) is sent towards RS prefixes. Hence, the prefixes advertised at RSes give significant insight into spatial aspects of its actual traffic components.

In view of Figure 6(a), we are especially curious about how much traffic is related to those prefixes that are exported to almost every peer of the RS and to those that are exported to only a few selected peers of the RS. We compute the percentage of traffic that each prefix at the RS is responsible for and plot in Figure 6(b) their sum as a function of the number of the RS's peers to which the RS exported the prefix. While we observe a similar bi-modality as in Figure 6(a), the openly-advertised prefixes are responsible for the largest traffic share. While the percentage of traffic covered by the very selectively-advertised prefixes (exported to less than 10% of the member ASes using the RS) is about 9% of all traffic, the more openly-advertised prefixes (exported to more than 90% of the member ASes using the RS) cover the destination addresses of almost 70% of all exchanged traffic in terms of bytes.

6.3 A closer look at RS usage

To this point, we have mainly focused on the overall traffic at the IXPs and how it relates to the prefixes that are advertised via the IXP's RS. In particular, we have paid little attention to the per-member AS policies that determine which ASes advertise their prefixes to which other ASes and over what kind of IXP peering. On the one hand, a majority of members openly advertise their prefixes via the RS, and these prefixes cover a majority of the IXP traffic. At the same time, we see the bulk of traffic traversing BL links. Thus, we would like to know if members advertise different prefixes to the IXP's RS and over their BL peering session, i.e., what is the overlap in terms of routes advertised via both sessions.

The problem we face is that while we know the route set advertised over ML peerings, we only have binary information about the existence of BL peerings (i.e., we sampled BGP packets indicating an active BL session). To deal with this issue, we rely on properties of the actual traffic exchanged between members. In particular, we check for each IXP member if the traffic it receives is fully covered by the prefixes advertised via the RS or if there is traffic to a superset of RS prefixes. More precisely, to obtain Figure 7(a), we compute for each member AS of the L-IXP (x-axis) which fraction of the traffic sent to this member is (i) covered by the prefixes that

this specific member advertises via the IXP's RS (shown in upper half of the plot) and (ii) not covered by the prefixes this member advertises to the RS (see lower part of the plot). Before plotting these values, we sort them in increasing order, starting with those members for which none of their traffic is covered by the RS prefixes. To minimize the impact of churn (new route advertisements, route withdrawals), we use one week of traffic, starting on the same day for which we have the corresponding RS dump.

Figure 7(a) shows that for the majority of members (to the right of $x \approx 120$), all traffic they receive is covered by the prefixes they advertise via the RS; that is, there is no traffic at the IXP destined to any addresses outside these prefixes – for these members, the RS is all there is! For another set of members (to the left of $x \approx 85$), we see that none of the traffic they receive is covered by the RS prefixes advertised by them. These members either do not connect to the RS at all or do not advertise any prefixes via the RS. Finally, there is a small group of member ASes in the middle (around $x = 100$) that use prefixes that they advertise via the RS as well as prefixes that they do not advertise via the RS. We discuss these cases in more detail in Section 8.2. In terms of numbers, about 67% of all traffic is sent to members on the right part of this figure and roughly 26% of the traffic is destined to the members in the cluster on the left. The members in the middle-section receive about 7% of all traffic – a significant share given the small number of members in the middle-section. For M-IXP, the members in the right part account for more than 95% of the traffic.

Figure 7 uses grey-shading to indicate which portion of a member's traffic is associated with BL or ML peerings. For example, the members that only peer bi-laterally are all black (the few exceptions are due to churn in route advertisements). Member ASes that only peer multi-laterally are all grey. In addition, some members are highlighted in terms of their main business type, and we will return to this cast of players in Section 8 below. Together, Figures 7(a) and 7(b) show that while most member ASes use BL and ML peerings, the majority of members only receive traffic for those prefixes that they also advertise via the RS. Thus, the need to advertise a different route set bi-laterally and multi-laterally can — in most cases — be ruled out as the reason for two member ASes to peer both bi-laterally and multi-laterally with each other at an IXP (following up on the discussion in Section 5.1). Instead, the reasons are more likely a question of control and trust (see discussion in Section 9.3), but more mundane reasons such as the ease with which ML peerings can be established and the lack of incentives to remove/block prefix advertisements via the RS once a BL peering has been established might also play a role.

Summary

Our IXP RSes provide access to a significant portion of the Internet's routes and networks and most member ASes provide open access to their RS-prefixes to all IXP RS participants. The routes advertised openly via the RS are indeed routes to popular destinations and they receive a majority of the traffic at both IXPs. In turn, the route set of an IXP RS is a good approximation of the popular destinations as seen from this IXP. Individual member ASes typically exhibit a binary behavior: either all routes for which traffic is received are advertised via the RS or none at all. Many member ASes that openly advertise their prefixes via the RS also establish BL sessions to selected other members ASes - without advertising a superset of prefixes on these sessions. This observation agrees with publicly-stated peering policies from large content providers (see, e.g., Google [5]), which state that BL peerings can be established once a significant amount of traffic is exchanged. We also see cases in which member ASes advertise a different route set on BL peerings than on ML peerings and discuss some of them in Section 8 below.

From To	04-2011 12-2011	12-2011 06-2012	06-2012 12-2012	12-2012 06-2013
# (ML ⇒ BL)	577	440	546	435
Δ Traffic	+86%	+230%	+82%	+204%
# (BL ⇒ ML)	172	175	189	242
Δ Traffic	+20%	-77%	-65%	-42%

Table 5: Number of peerings changed from BL type to ML type and vice versa and corresponding changes in traffic volumes carried over these peerings (L-IXP).

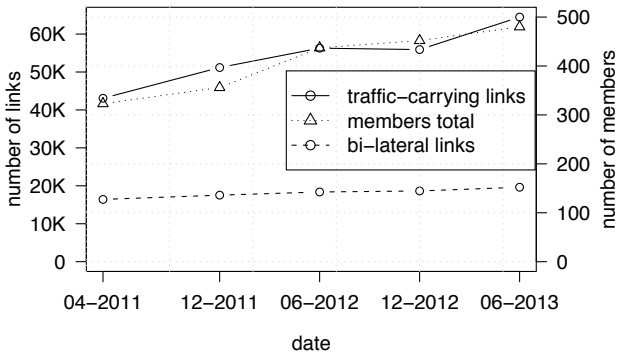


Figure 8: Number of peerings over time (L-IXP).

7. CHANGE CHANGES THINGS

With our IXP datasets, we are observing an Internet peering ecosystem in flux. With more IXPs operating RSes, more members peer multi-laterally and connect to more than one IXP. In this section, we examine in more detail how IXP peerings at the L-IXP change over time and how the common members at the L-IXP and the M-IXP use those IXPs.

7.1 Same IXP: IXP peerings over time

Taking advantage of our historical collection of sFlow records from L-IXP, we perform a longitudinal study of the IXP peerings at this IXP. To examine the relationship between BL and ML peerings during the last 2+ years, we focus on the traffic-carrying peerings and rely on the methodology described in the latter part of Section 4.1 to infer BL peerings. The results are depicted in Figure 8 and illustrate that while the total number of observed traffic-carrying links has increased noticeably during this time period, the number of BL peerings has only increased slightly. Thus, the observed increase in the number of traffic-carrying peerings is mainly due to the proliferation of ML peerings. Taking into account our earlier observation that only about 85% (24%) of symmetric (asymmetric) ML peerings do carry actual traffic, we can conclude that the number of established ML peerings exceeds the number of traffic-carrying ML peerings by some 5-10k links. At the same time, we find that the ratio of total BL peering traffic to total ML peering traffic has stayed constant at around 65–67% (not shown), a reminder that from a traffic perspective one has to consider both types of peerings.

To examine in more detail the dynamics of peerings over time, we consider next five different two-week snapshots of sFlow data, dating back to 2011. For each snapshot, we identify the traffic-carrying BL peerings and tag the remaining traffic-carrying links as

M-IXP \ L-IXP	yes	no
yes	67.9%	11.4%
no	12.1%	8.6%

(a) Connectivity across IXPs.

M-IXP \ L-IXP	yes	no
yes	50.9%	13.6%
no	22.8%	12.7%

(b) Traffic across IXPs.

M-IXP \ L-IXP	BL	ML
BL	27.8%	3.2%
ML	22.6%	46.4%

(c) Peering type across IXPs.

Figure 9: Peerings between common members across L-IXP and M-IXP. Consistency regarding connectivity, traffic exchange and peering type.

ML peerings. Then we check if the same traffic-carrying peerings that are present in two consecutive snapshots have changed their peering type or not (see Table 5). Although some small amount of churn could result from our inability to detect all BL sessions in a given snapshot (see Section 4.1), we observe that the number of changes from ML-to-BL and BL-to-ML, respectively, are relatively stable. The churn caused by ML peerings being replaced by BL peerings is consistently higher than the churn that is due to the peering types changing from BL to ML. We also notice that the traffic volume on these peerings increases substantially if the type changes from ML to BL. In contrast, a peering type change from BL to ML typically goes along with a significant decrease in the amount of traffic that gets exchanged. These observations support the argument that BL peerings are typically established and used if there is significant traffic volume that may or may not need additional BGP capabilities (see e.g., [5]). Another reason for encountering more ML-to-BL than BL-to-ML changes is that when new members join an IXP, they typically start with ML peerings and may switch later on to BL peerings with a subset of IXP member ASes.

7.2 Different IXPs: Common members

We next perform a comparative analysis and check how networks make use of the two IXPs that cover by and large the same geographic region. The two IXPs have 50 members in common (see Table 1). Figure 9 summarizes the results of our comparison with respect to connectivity (i.e., peerings and peering type) and traffic. Of these 50 common ASes, more than 75% use consistent peering setups in the sense that if they are peering/not peering with one another at one IXP, they also do so at the other IXP. At the same time, more than 20% do peer at one IXP but not at the other IXP. However, not all of these peerings see traffic, and the percentages of traffic-carrying peerings are reported in the second sub-table of Figure 9. As far a the types of these traffic-carrying peerings are concerned, we observe that in 46% of the cases ML peering is used at both locations and BL peering in 28% of the cases. This observation is consistent with the overall picture (not presented) and shows

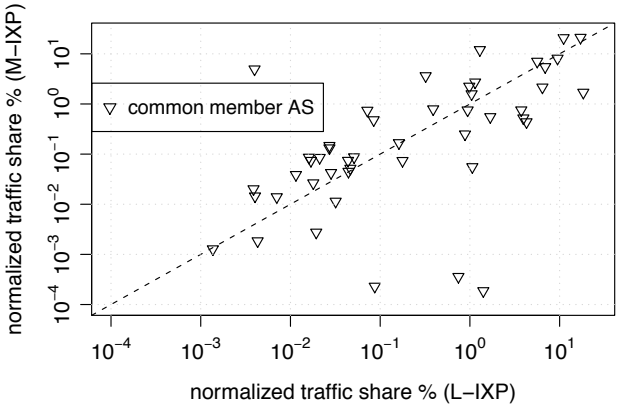


Figure 10: Common ASes: Normalized traffic share.

that while the member ASes at M-IXP use predominantly ML peerings, it is more often the case that the member ASes use BL peering at the L-IXP. In fact, the BL peerings common to both IXPs make up almost half of all traffic-carrying BL peerings at M-IXP.

To illustrate the common members' contributions to the traffic volume at the two IXPs, Figure 10 shows for all common member ASes that exchange traffic at the two IXPs how their normalized traffic shares compare. The scatter-plot shows for each common member AS the fraction of its traffic at the L-IXP (x-axis) vs. the fraction of its traffic at the M-IXP, where we normalize by all traffic over the common peerings. The observed clustering around the diagonal is an indication that the common member ASes' (relative) contributions to the respective overall peering traffic at the two IXPs are similar. A similar strong clustering can also be seen when examining the common members' contributions at the level of common peering links (not shown).

The overall picture that emerges from studying the common member ASes at the two IXPs and how they use the different IXPs is consistent with their profiles (see Section 3). For example, the ASes in the upper right corner of Figure 10 are the well-known big content providers/CDNs and make extensive use of both IXPs – they use the L-IXP because of its global role as a critical hub in today's Internet, and they are at M-IXP because of its mainly regional role as a place for small-medium eyeball networks to connect. On the other hand, the few outliers in Figure 10 (e.g., point in upper left and points in lower right) are typically smaller regional networks that prefer one IXP over the other for reasons such as location, reach, and needs. In addition, for the latter to establish connectivity at the two IXPs via the IXPs' RSes is very low overhead and justifies membership at the IXPs, irrespective of any imminent need to exchange traffic.

Summary

Our longitudinal analysis of the L-IXP shows that while its RS is clearly the main driver for increasing connectivity among member ASes at this IXP, the traffic ratio between BL and ML peering stays largely constant, with BL peerings accounting for twice as much traffic. Switch-overs from ML to BL peering typically occur when there is an increase of traffic carried over that link. For ASes that are members at two IXPs, we observe that peering links with other networks are largely consistent across the different IXPs. We also notice that the relative traffic contributions of members peering at our two IXPs exhibit strong correlations. This is also the case for common traffic-carrying peerings between the same ASes across IXPs.

AS	RS usage L/M IXP	Notes	# traffic links	# BL links	%BL traffic
C1	yes / yes	open peering	417 / 82	329 / 41	91 / 99
C2	yes / yes	open peering	392 / 73	138 / 2	35 / 0.5
OSN1	no /-	only BL	248 / -	256 / -	100 / -
OSN2	yes / -	open peering	337 / -	0 / -	0 / -
T1-1	no / no	very selective	3 / 0	22 / 0	100 / -
T1-2	yes / -	no-export	18 / -	19 / -	100 / -
EYE1	yes / yes	open peering	405 / 77	134 / 11	74 / 20
EYE2	yes / yes	open peering	401 / 77	198 / 41	84 / 72

Table 6: Case Studies: RS usage, number of traffic-carrying links, number of BL links, fraction of traffic seen on BL links at the L-IXP/M-IXP.

8. CASE STUDIES

In this section, we take a closer look at some of the well-known major players in today's Internet that send significant traffic over the public switching fabrics of either L-IXP or M-IXP. We are interested in understanding how they make use of these IXPs so that we can reason about their choices of peering options, and assess whether there are typical patterns for ASes of the same business type.

8.1 Our cast of big players

Our selection of "big players" was driven by the volume of traffic they exchange at L-IXP or M-IXP, by the type of businesses they represent, and by their geographic reach. To this end, we chose two ASes that are responsible for a significant share of content in today's Internet and are among the top traffic contributors at both IXPs where each of their traffic share is over 10% (C1 and C2). Moreover, we picked two popular Online Social Networks (OSN) that use the IXPs extensively (OSN1 and OSN2), two regional eyeball providers (EYE2 and EYE1), and two Tier-1s (T1-1 and T1-2).

There are two basic observations from our characterization work described in the previous sections. First, the big content providers and the regional eyeball networks are actively using the IXPs' RSes and peer openly, most likely to reduce transit cost and reduce latency. On the other hand, the Tier-1s which include some of the major eyeball networks with presence in Europe and overseas peer more restrictively and often do not use the RS. Our findings about the selected players are summarized in Table 6 and show some unexpected yet important differences.

In the case of the two major content providers, we note that C1 exchanges most of its traffic via BL peerings. However, this content provider actively markets the RS connectivity and maintains many ML peering sessions. On the other hand, C2 exchanges more than half of its traffic via the ML peerings and we notice an increasing tendency towards ML peerings for this player over the last two years. In both cases, their prefixes are openly advertised to all members and we see 100% resp. 99% of the traffic sent towards these members covered by the prefixes they advertise via the RS.

When examining the two OSNs, the most striking finding is that they approach peering completely different, choosing the two extremes of the spectrum of available peering options. On the one hand, OSN1 only offers BL peering and is not present at the IXPs' RS. A likely reason is that this OSN wants to have control over its traffic flows. On the other hand, OSN2 only does ML peering at IXPs like L-IXP that are outside of this OSN's home country. In fact, their publicly announced peering policy states that they try to avoid BL BGP sessions wherever peering using a RS is possible. A plausible reason to choose this option is to reduce the administrative overhead associated with BL peering.

As for the Tier-1s, in contrast to the US where Tier-1s typically do not use the IXPs' public switching fabrics, many Tier-1s connect at L-IXP, including our players T1-1 and T1-2. One of them also connects at M-IXP. Focusing on our two Tier-1s, they all use a selective peering policy which is often interpreted to mean that they do not use the RS. While this is true for most of them, we see some exceptions. T1-1, for example, does indeed not use the RS at all at either of the IXPs. While T1-2 connects to the L-IXP RS, it tags all of its routes with the NO-EXPORT community which implies that none of them are shared with any of the other ASes that are at the RS. Thus, this Tier-1 also solely relies on bi-lateral connections. The reason for its presence at the RS is unclear to us but might be a hint towards early exploration of RS usage.

The regional eyeball providers, EYE1 and EYE2 both peer openly utilizing the RS at both IXPs. We see all of the traffic sent to these members covered by their RS prefixes. While we notice a tendency towards more ML peerings at the M-IXP vs. the L-IXP, both eyeball networks also use BL peerings, with EYE2 relying mostly on BL sessions.

8.2 Players with hybrid peerings

Figure 7 (left lower part) clearly shows that some members only peer bi-laterally at L-IXP or M-IXP. The other large group of members consists of those that advertise all prefixes for which they receive traffic via the RS but receive varying amounts of traffic via BL peerings. However, the more intriguing cases are those members that occupy the small middle part in Figure 7 (e.g., around $x = 100$) and advertise some prefixes via the RS, but receive traffic for a superset of these prefixes. While this fraction of members is rather small compared to the overall number of members, we see some important players here, carrying significant traffic volumes. We select two cases, namely CDN, a mid-sized content delivery network and NSP, a large transit provider. We have seen that content providers typically advertise their prefixes openly and large transit providers advertise them more selectively. CDN and NSP differ from these typical cases by following a "hybrid" peering strategy by advertising some prefixes multi-laterally, some others only bi-laterally to selected member ASes.

In the case of CDN, we find that close to 90% of its incoming traffic is covered by the small number of openly advertised RS prefixes. However, this AS also has some 59 BL peering links, and on most of them we see traffic to a superset of the prefixes advertised via the RS.[14] We assume that this is as an example of complex traffic engineering performed by today's CDNs that, in this case, involves both using the vast connectivity provided by the RS, as well as the application of more fine-granular prefix advertisements for specific networks, e.g., to fulfill performance requirements for certain customers.

NSP advertises some 5K prefixes openly via the RS, and most of them contain subsequent ASNs in the AS path that are not other ASNs of the NSP. Nevertheless, most traffic is exchanged via BL peering sessions and only about 20% of NSP's incoming traffic is covered by its RS prefixes. While at L-IXP, NSP receives traffic from most members ASes, many of them peer bi-laterally with NSP and some of them send traffic to a large superset of the prefixes advertised via the RS. Interestingly, we see a very similar pattern for NSP also at M-IXP where it receives about 45% of its traffic for prefixes advertised via the RS and the other part for a super-

[14]We verified that a superset of prefixes is advertised on these sessions by using the LGes of IXP members. Moreover we note that the corresponding paths were prepended.

set. Again, we validated the advertisement of a superset of prefixes using LGes of IXP members having BL sessions with NSP. We assume that NSP's hybrid approach to peering is in support of its main business (e.g., transit provider), and hints towards growing diversification of models in the transit business (see e.g., [35], [23]). We note that NSP is present at many IXPs and thus might have more confidence and experience in making use of RS functionality, tailored to its business interests, than other transit providers.

Summary

Among the networks that have an open peering policy and utilize IXP RSes, we observe a diverse set of players ranging from content providers to eyeball networks. While many networks that utilize the RS also rely on additional BL peerings, some only use the RS (i.e., no BL peerings) and others (particularly larger transit networks) typically do not use RSes at all. We also observe networks with hybrid peering strategies in the sense that they advertise parts of their routes via the RS and others via BL peerings, hinting towards advanced usage of RSes.

9. DISCUSSION

Our empirical study of real-world IXP RS deployments and how IXP RSes are being used by a wide range of today's Internet players has implications for IXP and network operators as well as researchers.

9.1 On the benefits of connecting to an IXP

The success of the large European IXPs is an example of a (positive) network effect "in action." As these IXPs have been growing in size, their value for each of their members has increased, but managing and maintaining the ensuing bi-lateral peerings (i.e., large number of corresponding BGP sessions) has became a burden for many of these members. To alleviate this burden, these IXPs have started to offer their members the free use of their route servers, and this service has contributed to making those IXPs even more attractive to yet more networks.

To quantify this attractiveness, our empirical results suggest a concrete step that IXPs and network operators could take for effectively evaluating the instant benefits that a network can reap when connecting (either directly or remotely) to an IXP. In particular, we have seen that at our two IXPs, where a majority of the members connect to the RS (i.e., more than 80% of the L-IXP's members and more than 90% of the M-IXP's members), the prefixes advertised by each of the two IXPs' RSes cover some 80-95% of all traffic seen and include many of the most popular destinations. Thus, if IXPs provide the profile of routes that are advertised via their RSes (e.g., via adequately-supported LGes), network operators can immediately determine how much of their individual traffic would reach these destinations from "day one" (i.e., as soon as they start connecting to the IXP's RS). They can then compare the different available interconnection options in terms of costs, performance or other criteria to make an informed decision about which IXP(s) they want to connect to and how they want to make use of the IXP's RS (see Section 8).

Although we lack comparable data from the other large European IXPs that operate a RS, the fact that a majority of their members connect to the RS (e.g., some 70% at LINX, over 80% at AMS-IX) suggests that their RSes cover a similarly large portion of traffic seen at those IXPs as observed at the L-IXP and M-IXP. However, we caution against generalizing our RS-specific observations across the spectrum of IXPs worldwide. For example, even in Europe there are large regional IXPs (e.g., NL-IX) where the reported member participation at the RS is less than 50%. Looking beyond Europe, while many of the larger IXPs in North America operate a RS (e.g., Equinix, Any2), little is known about member participation in the RS[15] and typically no information is available about the portion of traffic covered by such an IXP's RS.

9.2 A peering ecosystem in flux

Our findings show that the popularity of the IXPs' RSes has led to a proliferation of ML peerings, but at this point, a majority of the traffic that is exchanged at a typical IXP is still exchanged via BL peerings. However, as more networks explore how to best use their many easy-to-establish and simple-to-maintain ML peerings (e.g., as back-up, as primary connection, only for certain traffic or particular peers) and try to incorporate them into their complex peering decision making processes, the current picture is bound to undergo constant changes, especially if the IXPs keep on innovating and improving RS functionalities and operations. Thus, new research efforts are required for developing and deploying appropriate monitoring infrastructures that yield reliable data for sensing the onset of possible "sea changes" in the future Internet peering ecosystem and also providing insight into the reasons for their occurrence.

In this context, although the IXP-provided data used in this study is proprietary, we show in Section 4 how it can be put to good use by increasing the value of certain publicly available datasets for future studies by other researchers. For example, proprietary data often enables a direct comparison between inferred connections (based on publicly available information) and actual connections (based on proprietary information), and we illustrate this capability in this paper by calibrating the quality of the BGP measurements that are becoming publicly available from RS-LGes in selective IXPs.

The use of such vetted public data would be of particular interest in the context of monitoring the impact (or lack thereof) of the recently launched *Open IX (OIX)* initiative [11]. This initiative's main objective is to bring more peering opportunities to the US interconnection marketplace that has historically been dominated by an inordinately small number of players whose main focus has been on private peering offerings (i.e., cross connects). As a result of this situation, prices for cross connects in the major metropolitan areas in the US are about six times as high as in the major European cities, where peering opportunities are abundant, especially in places with highly successful IXPs. Among the early developments related to the OIX initiative that was launched less than a year ago are the establishment of new European-style IXPs in the USA by the largest European IXPs (e.g., LINX in Northern Virginia [9], AMS-IX and DE-CIX in New York [1, 3]) and a formal certification process whereby the certified commercial data center and colocation provider companies explicitly agree to house certified IXPs in their facilities. Only time will tell if these early developments will bring about the "sea change" within the US peering ecosystem that the supporters of OIX are advocating, but our findings point to concrete data collection efforts that should be pursued to assess the actual impact of OIX in the US (e.g., encouraging the deployment of RS-LGes by the IXPs in the US and the continuous data collection from these emerging new vantage points).

9.3 Innovation in inter-domain routing

Based on our description of today's RSes, their popularity is understandable because establishing connectivity in the control plane via a RS is simple (i.e., requiring basically one BGP session setup), efficient (in terms of receiving routes to a significant part of the Internet), and quick (i.e., an ability to start exchanging traffic from the

[15]A noteworthy exception is SIX in Seattle, which started offering RS services in 2014 and provides a Looking Glass showing all members connecting to the RS and their advertised routes.

get-go). In this sense, the relatively recent RS offerings by IXPs can be viewed as a concrete example of an innovation in inter-domain routing, an area that is notorious for resisting changes to or deploying improvements over the status quo (i.e., BGP).

More importantly, our discussions of the design and operations of today's IXP RSes show that they do not forward any data traffic and adhere to a strict separation of control plane and data plane. This observation makes RS design and operations a prime candidate for Software Defined Networking (SDN) which, in turn, suggests that IXPs are natural places for harnessing SDN's power to improve if not revolutionize inter-domain routing. In fact, the concept of a software defined Internet exchange (SDX) is the topic of a recent paper [27] in which the authors consider novel approaches to innovation at the IXP RSes, including software support for a range of policies for inter-domain traffic engineering with BGP that are either difficult to perform using current RS capabilities or simply out of reach for today's RS implementations.

In fact, by describing how RSes are used in practice, our work is complementary to [27] and raises a number of questions that are mainly concerned with issues such as flexibility, trust, and security. Addressing these and related issues as part of the SDX effort would have far-reaching implications, mainly because they have prevented some of the larger players to rely and depend on the IXPs' RSes. For example, member ASes that want to control which prefixes are advertised to which other member ASes typically have the BGP know-how to peer bi-laterally to achieve the desired control over the traffic they receive. While the use of the RS does not preclude such control, peering bi-laterally offers a member AS additional opportunities for BGP traffic engineering, e.g., AS-path prepending, MED, scoped advertisements [39] etc., that are currently not fully supported by RSes. More generally, how to expand the effort described in [27] to also include bi-lateral peering as part of the proposed SDX approach looms as an interesting open problem.

With respect to trust-related issues, IXP operators are aware of concerns that some of their member ASes have about RS availability, RS operations, RS monitoring/debugging capabilities, and security at the RS.[16] For example, there currently exist no SLAs for the RS itself, but most of the IXPs operate redundant RSes.[17] From an operational perspective, there are numerous opportunities for misconfigurations at the RS (e.g., filter setup, mis-shapes with routing registries), and the monitoring capabilities at today's RSes are still in their infancy. In terms of security, an RS adds an additional middleman that has to be trusted with the proper handling of all route advertisements. In this context, large IXPs may be opportune places for future deployment of BGP security mechanisms such as sBGP [37].

10. CONCLUSION

Using a collection of IXP-provided datasets that offer unprecedented visibility into the control plane (i.e., BGP data from the RS) and data plane (i.e., sFlow measurements) as seen at two European IXPs, we report in this paper on an in-depth study of IXP RSes for the purpose of understanding an IXP's full-fledged public peering fabric; that is, the existence of peerings (i.e., connectivity) as well as their usage (i.e., traffic) and how they are used by which player and for what reasons. In the process, we obtain an accurate picture of the relative importance of ML peerings (large in numbers, but responsible for only a minority of the traffic) vs. BL peerings (small in numbers, but carrying the bulk of the traffic). More importantly, when moving beyond the traditional link-level perspective of peering and examining the correlation between the control and data plane views at the prefix-level, reasoning about which ASes peer with which other networks (e.g., how and for what reasons) becomes feasible. This, in turn, leads to new insight into how the IXPs' RSes are used by their member ASes and demonstrates that these RSes are quickly becoming critical enablers for connectivity and, as a result, integral entities of today's Internet inter-domain routing system that deserve the full attention of networking researchers.

[16]Personal communication with IXP operators.

[17]The L-IXP reported zero downtime during 2013.

Acknowledgments

We want to express our gratitude towards the IXP operators for their generous support and feedback. We thank our shepherd Phillipa Gill and the anonymous reviewers for their helpful feedback.

This work was supported in part by the EU project BigFoot (FP7-ICT-317858). Georgios Smaragdakis was supported by the EU Marie Curie International Outgoing Fellowship "CDN-H" (PEOPLE-628441).

11. REFERENCES

[1] AMS-IX New York. https://nynj.ams-ix.net/.

[2] AMS-IX Partner Program. https://ams-ix.net/connect-to-ams-ix/partner-program.

[3] DE-CIX New York. http://nyc.de-cix.net/.

[4] Euro-IX Resources: Traffic, Reports, and Best Practices. https://www.euro-ix.net/resources.

[5] Google Peering Policy. https://peering.google.com/about/peering_policy.html.

[6] Internet2 Network Research Data. http://noc.net.internet2.edu/i2network/research-data.html.

[7] IRR - Internet Routing Registry. http://www.irr.net.

[8] LINX: Black Holing Support for DDoS Attack. HotLINX 34, https://www.linx.net/files/hotlinx/hotlinx-34.pdf.

[9] LINX NoVA. https://www.linx.net/service/publicpeering/nova.

[10] Netflix Open Connect. https://signup.netflix.com/openconnect.

[11] Open-IX Association. http://www.open-ix.org/.

[12] OpenBGPD Border Gateway Protocol. http://www.openbgpd.org/.

[13] Packet Clearing House routing archive. https://www.pch.net/resources/data.php.

[14] Quagga Routing Suite. http://www.nongnu.org/quagga/.

[15] TeleGeography (press release on September 24, 2013): U.S., Europe Colocation Pricing Models Vary Significantly. http://www.telegeography.com/press/marketing-emails/2013/09/24/u-s-europe-colocation-pricing-models-vary-significantly/index.html.

[16] The BIRD Internet Routing Daemon. http://bird.network.cz.

[17] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger. Anatomy of a Large European IXP. In *ACM SIGCOMM*, 2012.

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

[19] T. Bates. Implementation of a Route Server for Policy Based Routing across the GIX Project, 1993.

[20] M. O. Buob, S. Uhlig, and M. Meulle. Designing optimal iBGP route-reflection topologies. In *IFIP Networking*, 2008.

[21] N. Chatzis, G. Smaragdakis, A. Feldmann, and W. Willinger. There is More to IXPs than Meets the Eye. *ACM CCR*, 43(5), 2013.

[22] O. Filip. BIRD's flight from Lisbon to Prague. RIPE 60.

[23] P. Gill, M. Schapira, and S. Goldberg. A survey of interdomain routing policies. *ACM CCR*, 44(1), 2014.

[24] V. Giotsas and S. Zhou. Improving the Discovery of IXP Peering Links through Passive BGP Measurements. In *Glob. Internet*, 2013.

[25] V. Giotsas, S. Zhou, M. Luckie, and kc claffy. Inferring Multilateral Peering. In *ACM CoNEXT*, 2013.

[26] R. Govindan, C. Alaettinoglou, K. Varadhan, and D. Estrin. Route Servers for Inter-domain Routing. *Computer Networks*, 30, 1998.

[27] A. Gupta, Vanbever L, M. Shahbaz, S. P. Donovan, B. Schlinker, N. Feamster, J. Rexford, S. Shenker, R. Clark, and E. Katz-Bassett. SDX: A Software Defined Internet Exchange. In *ACM SIGCOMM*, 2014.

[28] Y. He, G. Siganos, M. Faloutsos, and S. V. Krishnamurthy. A systematic framework for unearthing the missing links: Measurements and Impact. In *NSDI*, 2007.

[29] N. Hilliard, E. Jasinska, R. Raszuk, and N. Bakker. Internet Exchange Route Server Operations. IETF draft, draft-ietf-grow-ix-bgp-route-srver-operations-01, 2013.

[30] M. Hughes. Route Servers at IXPs – Bugs and Scaling issues with Quagga. UKNOF 13.

[31] E. Jasinska, N. Hilliard, R. Raszuk, and N. Bakker. Internet Exchange Route Server. IETF draft, draft-ietf-idr-ix-bgp-route-server-03, 2013.

[32] E. Jasinska and C. Malayter. (Ab)Using Route Servers. NANOG 48.

[33] A. Labrinidis and E. Nguyenduy. Route Server Implementations Performance. 20th Euro-IX Forum, April 2012.

[34] S. J. Liebowitz and S. E. Margolis. Network Externality: An Uncommon Tragedy. *J. Econ. Perspectives*, 8(2), 1994.

[35] A. Lodhi, A. Dhamdhere, and C. Dovrolis. Open Peering by Internet Transit Providers: Peer Preference or Peer Pressure? In *IEEE INFOCOM*, 2014.

[36] A. Lodhi, N. Larson, A. Dhamdhere, C. Dovrolis, and kc claffy. Using PeeringDB to Understand the Peering Ecosystem. *ACM CCR*, 44(2), 2014.

[37] R. Lychev, S. Goldberg, and M. Schapira. Is the Juice Worth the Squeeze? BGP Security in Partial Deployment. In *ACM SIGCOMM*, 2013.

[38] C. Malayter. Route Servers, Mergers, Features, & More. NANOG 51.

[39] B. Quoitin, C. Pelsser, L. Swinnen, O. Bonaventure, and S. Uhlig. Interdomain traffic engineering with BGP. *IEEE Communications Magazine*, 2003.

[40] R. Raszuk, C. Cassar, E. Aman, B. Decraene, and S. Litkowski. BGP Optimal Route Reflection. IETF draft, draft-ietf-idr-bgp-optimal-route-reflection-06, 2013.

[41] M. Roughan, W. Willinger, O. Maennel, D. Pertouli, and R. Bush. 10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems. *IEEE J. on Sel. Areas in Comm.*, 29(9), 2011.

[42] InMon – sFlow. `http://sflow.org/`.