Anatomy of a Large European IXP

Bernhard Ager*
ETH Zurich
bernhard.ager@tik.ee.ethz.ch

Nikolaos Chatzis
TU Berlin / T-Labs
nikos@net.t-labs.tu-berlin.de

Anja Feldmann
TU Berlin / T-Labs
anja@net.t-labs.tu-berlin.de

Nadi Sarrar
TU Berlin / T-Labs
nadi@net.t-labs.tu-berlin.de

Steve Uhlig*
Queen Mary, University of London
steve@eecs.qmul.ac.uk

Walter Willinger
AT&T Labs–Research
walter@research.att.com

ABSTRACT

The largest IXPs carry on a daily basis traffic volumes in the petabyte range, similar to what some of the largest global ISPs reportedly handle. This little-known fact is due to a few hundreds of member ASes exchanging traffic with one another over the IXP's infrastructure. This paper reports on a first-of-its-kind and in-depth analysis of one of the largest IXPs worldwide based on nine months' worth of sFlow records collected at that IXP in 2011.

A main finding of our study is that the number of actual peering links at this single IXP exceeds the number of total AS links of the peer-peer type in the entire Internet known as of 2010! To explain such a surprisingly rich peering fabric, we examine in detail this IXP's ecosystem and highlight the diversity of networks that are members at this IXP and connect there with other member ASes for reasons that are similarly diverse, but can be partially inferred from their business types and observed traffic patterns. In the process, we investigate this IXP's traffic matrix and illustrate what its temporal and structural properties can tell us about the member ASes that generated the traffic in the first place. While our results suggest that these large IXPs can be viewed as a microcosm of the Internet ecosystem itself, they also argue for a re-assessment of the mental picture that our community has about this ecosystem.

Categories and Subject Descriptors

C.2 [**Computer Systems Organization**]: Network Operations

Keywords

Internet Exchange Points, Internet topology, traffic characterization

1. INTRODUCTION

The basic role of Internet eXchange Points (IXPs) dates back to the establishment of Network Access Points (NAPs) as part of the decommissioning of the National Science Foundation Network (NSFNET) around 1994/95, a carefully orchestrated plan for transitioning the NSFNET backbone service to private industry. The vehicle that evolved in support of this transition was a set of four NAPs (i.e., MAE-East, Sprint NAP, PacBell NAP, and Ameritech NAP) that acted as connection points for the commercial carriers that were vying for offering backbone services (e.g., MCInet, Sprintlink, AGIS) and ensured that the network would remain connected at the top level once the NSFNET was retired.

Over the past 15 years, as the Internet grew by leaps and bounds by any imaginable metric, the original four NAPs were replaced by a steadily increasing number of modern IXPs. Originally providing largely just the bare necessities for supporting easy interconnection between their member ASes (e.g., physical space, caches, cabling, power, A/C, or secure access), IXPs themselves have evolved over time. Numbering now more than 300 worldwide [18], many of these IXPs are offering an array of different services that rely on advances in networking technology (e.g., VLANs or MPLS), exploit existing routing protocols in innovative ways (e.g., use of BGP for prefix-specific peering), or provide the economic incentives for an ever-increasing number of networks to join as paying members (e.g., remote peering offerings, support for IXP resellers).

In fact, large IXPs such as AMS-IX, situated in Amsterdam, and DE-CIX, in Frankfurt, offer high-end Service Level Agreements (SLAs) to their members that cover not only the initial provisioning and daily availability of a member's port(s) but also the level of performance of key service parameters. Such innovation on parts of the IXPs has enabled them to compete more directly with the traditional carriers and has led to today's environment where some of the largest IXPs worldwide (e.g., AMS-IX, DE-CIX, LINX, MSK-IX) reportedly carry on a daily basis similar amounts of traffic as some large ISPs (e.g., AT&T, Deutsche Telekom[1]). The traffic volumes at those IXPs are generated by some 300-500 networks that cover the whole spectrum of players in today's Internet marketplace. While there may be regional differences in how extensive in coverage or aggressive in the uptake of new members IXPs are, the critical role they have played in the Internet ecosystem has until recently gone largely unnoticed by the research community whose focus has traditionally been on large carriers and large content.

This paper reports on a first-of-its-kind measurement-based study of a large IXP and complements a body of existing literature that has focused squarely on large carriers or large content. To this end, we analyze a unique dataset consisting of nine months' worth of anonymized sFlow records that were collected at one of the largest IXPs in Europe, and worldwide, in 2011. We present our dataset, describe our data analysis and illustrate how our observations and findings contribute to an improved understanding of

*Part of this work was done while employed at TU Berlin / T-Labs.

[1]AT&T reports carrying 28.9 petabytes of data traffic on an average business day [2], Deutsche Telekom reports 422 petabytes per month corresponding to 14 petabytes per day on average [13].

- the AS-level Internet; that is, the structure and dynamics of the Internet as a network of networks or ASes;
- the Internet peering ecosystem; that is, the practices and economic incentives that drive the market for Internet interconnection and peering between ASes; and
- the Internet inter-domain traffic; that is, the quantity and quality of the traffic exchanged among ASes.

Specifically, the main contributions of our work fall into three categories. First, we show that this large IXP exhibits a surprisingly rich peering fabric in support of the many business objectives of its members. In particular, in terms of AS links of the peer-peer type that are typically established among member AS pairs we show that this IXP has close to 400 members which have established some 67 % (or more than 50,000) of all possible such peerings and use them for exchanging some 10 PB of IP traffic daily. *To put this number in perspective, note that as of 2010, the number of inferred AS links of the peer-peer type in the Internet was reported to be around 40,000 – less than what we observe at this particular IXP alone!*

To explain this startling difference between the number of peerings observed at this IXP (i.e., ground truth) and the number of known peer-peer AS links Internet-wide, we show which portion of the IXP's actual peering matrix is and is not visible when relying on the publicly available BGP data that has formed the basis for much of the past and recent work on inferring the Internet's AS-level connectivity. To further highlight this issue, we illustrate why a large portion of this IXP's actual peering matrix remains invisible even to measurement efforts that go beyond the current state-of-the-art, either with respect to BGP-derived data or traceroute-based measurements, or a combination of the two. By combining our finding of an enormously rich peering fabric among the members of this IXP with an accurate picture of their upstream connectivity (i.e., customer-provider relationships) that is reportedly quite accurate [11, 38], we are able to reconcile the traditionally-assumed hierarchical structure of the Internet with recent claims about a flattening of that structure. Indeed, while the traditional tier-structure of the Internet is still recognizable and can be largely recovered, the observed rich peering fabric at this IXP enables connectivity among networks of all different types and is essentially agnostic of any tier structure. Thus, at least as far as connectivity for the part of the Internet that involves this IXP is concerned, the observed IXP-related peerings provide a myriad of shortcuts and essentially complement any perceived or real hierarchical structure. Note that this realization of a much more elaborate interconnect structure than previously assumed says nothing about how these existing peering links are used to carry traffic.

To address the issue of how much and what type of traffic is traversing this IXP's infrastructure via public peering links, our second main contribution consists of an in-depth analysis of the available sFlow records. By examining the members of this large IXP and with which other members they peer and exchange traffic with, we highlight that large IXPs are a microcosm of the Internet as a whole in terms of types of networks, business relationships, or traffic. We observe various types of networks, from tier-1 to regional and local ISPs, large/medium/small content, host and service providers, content distribution networks (CDN), and a spectrum of academic and enterprise networks. With which other networks these networks establish peering connections at this IXP and exchange what type of traffic has nothing to do with their standing within the traditionally-assumed tier structure, but is largely dictated by economic considerations and business objectives and reflects a wealth of reasons and incentives for why the different types of networks make use of the various service offerings at this IXP.

Our third contribution is motivated by the existing large body of literature on traffic engineering for ISPs that relies critically on understanding how routing policies internal and external to the ISP affect the traffic flow over the ISP's infrastructure and ultimately result in what is commonly referred to as the ISP's intra-domain traffic matrix. In contrast to large ISPs, an IXP's infrastructure as well as the logical connectivity and routing at an IXP are significantly less complex and make it relatively easy to compute an IXP's traffic matrix and use it as a key input to IXP-specific methods in support of an efficient and effective management and operation of its infrastructure in a dynamic IXP marketplace. Despite the significant amount of traffic that the largest IXPs carry, especially when compared to the largest global ISPs, we are not aware of any research paper that provides an even remotely realistic picture of the traffic traversing an IXP. To fill this void, this paper is the first to obtain and characterize the traffic matrix of one of the largest IXPs, with a particular focus on traffic variability and dependency over time and in space (i.e., across ASes), and application mix. Knowledge of an IXP's traffic matrix and its properties in conjunction with emerging new routing strategies at IXPs is critical for exploring what-if scenarios that an IXP may want to run before announcing new services, encouraging members to send more traffic through the IXP, or for deciding when to upgrade its infrastructure and how.

The remaining part of the paper is structured as follows. In Section 2, we provide information about the large European IXP that we study and discuss the available sFlow records. Using this data, we examine in Section 3 the IXP's peering fabric and contrast our findings with results that rely exclusively on data that have been collected without the active participation of the IXP. Studying the members at this IXP as well as how and why they connect to other member ASes, we provide in Section 4 a detailed account of this IXP's ecosystem. In Section 5, we focus on the IXP's traffic matrix and report on some of its key characteristics. We elaborate on some of the implications of our findings and the new challenges they pose in Section 6, discuss related work in Section 7, and conclude in Section 8 with a summary of our main findings.

2. A LARGE EUROPEAN IXP

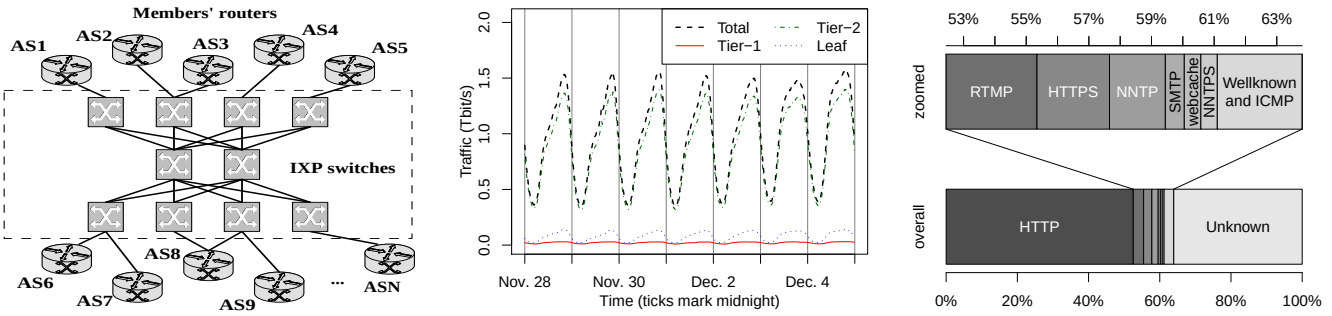
In this section, we provide a high-level overview of the infrastructure and operations of a large European IXP. We discuss the data that we obtained from this IXP, and present some basic facts about the member ASes of this IXP and about its overall traffic.

2.1 IXP overview

The main business model of an IXP is to operate and manage a physical infrastructure in support of public and private Internet interconnection. In this paper, we focus on the public part of an IXP's infrastructure where the IXP's revenues derive mainly from selling network interfaces or ports to customer networks (i.e., ASes) and supporting different types of interconnection arrangements. Such customer networks are referred to as member ASes. A member AS has the advantage to gain network connectivity to all other members of the IXP. However, interconnection arrangements reflect bi-lateral agreements[2] between a pair of member ASes, and these networks may want to impose certain conditions to ensure that they connect only to certain other networks or connect with them in ways that reflect their business model and support their market strategies.

What makes the IXP substrate of the Internet (i.e., the IXPs, their member ASes, and the peerings among these member ASes at those IXPs) such a vibrant marketplace is that the incentives for

[2]For the purposes of this paper, we view multi-lateral peering agreements as a collection of bi-lateral peering agreements.



(a) A typical IXP architecture. (b) Traffic volume distribution across time. (c) Application mix.

Figure 1: IXP architecture and traffic statistics for the Nov/Dec week.

networks to become members at such public peering platforms are as diverse as the growing number of increasingly diverse ASes. For example, a CDN interested in optimizing its performance while keeping its cost low might want to choose an open peering policy to encourage direct and settlement-free traffic exchange at an IXP with as many networks as possible. On the other hand, large ISPs are likely to be interested in establishing peering relationships with other ISPs of about the same size. To achieve this objective, they may want to base their peering decision on a selective peering policy that allows them to deny peering with small ISPs, thus retaining them as paying customers in customer-provider type interconnection arrangements that are more lucrative. Transit networks have yet different objectives for using an IXP – they look at an IXP as a point of sale of their upstream connectivity offerings. In general, the larger the number of member ASes at an IXP, the more attractive that IXP is as a peering platform. This explains to a large degree the high level of innovation that the IXP marketplace has experienced in the process of becoming a vital component of the Internet ecosystem.

2.2 IXP infrastructure and data

Figure 1(a) illustrates a high-level overview of the architecture of our IXP. Although complex to maintain and scale, the infrastructure of this large IXP is typical of large IXPs in general, and the IXP's operation can be described in simple terms. The IXP provides a layer-2 switching fabric and each of the member ASes connects its access router to that switching fabric. When a pair of member ASes decides to peer at the IXP, they establish a BGP session between their access routers which, in turn, enables the exchange of IP traffic over this peering link across the IXP's infrastructure.

The volume and properties of the traffic exchanged at an IXP depend on the number of member ASes, the location and scope of the activities of the IXP, the IXP's service offerings, and if the IXP operates for profit or as a non-profit organization [18]. In this paper, we consider the traffic that is exchanged over the public peering fabric supported by the switching infrastructure of the IXP. In particular, for this study, we rely on nine months' worth of continuous sFlow [47] records that were collected in 2011 at the IXP's infrastructure using a random sampling of 1 out of 16k packets. Our sFlow records capture the first 128 bytes of each sampled packet, thus giving us access to the IP and TCP headers. The sFlow capturing process includes an anonymization step in which IP addresses are scrambled while maintaining prefix consistency [19].

The efforts we made to assess the quality of the available sFlow records included checking for sampling bias and identifying and filtering out less than 1 % of the total traffic that was immaterial for our study. For example, since sFlow sampling is performed simultaneously and independently by multiple switches within the IXP's infrastructure, there may exist a bias toward such flows that traverse multiple sampling points. When counting the number of different sFlow probes that capture packets exchanged between the same pair of member router interfaces (MAC addresses), we found that more than 99 % of these flows were only sampled by a single probe, providing hard evidence that our data is not corrupted by this sampling bias. As for immaterial traffic, we filtered out all traffic contributed by the IXP's management machines (e.g., route servers) as well as broadcast and multicast traffic, except for ARP packets. Finally, we also eliminated all IPv6 traffic as it constitutes less than 1 % of the overall traffic (in bytes or packets) at this IXP.

Table 1: Overview of IXPs sFlow dataset.

	Apr 25 May 1	Aug 22 Aug 28	Oct 10 Oct 16	Nov 28 Dec 4
Identified member ASes	358	375	383	396
Router IPs	426	445	455	474
MAC addresses	428	448	458	474
Tier-1	13	13	13	13
Tier-2	281	292	297	306
Leaf	64	70	73	77
Countries of member ASes	43	44	45	47
Continents of member ASes	3	3	3	3
Average packet rate (Mpps)	142	150	166	174
Average bandwidth (Gbps)	838	863	954	992
Daily avg volume (PB)	9.0	9.3	10.3	10.7

2.3 IXPs: A moving target

Studying one of the largest IXPs means chasing a moving target. Large IXPs present a changing environment, with a number of different dynamic factors acting on different time scales. Over large time scales (i.e., annual or monthly), there are changes due to new IXP policies. On more medium time scales (i.e., weekly), there is churn in IXP membership (e.g., new members join, but there are also potential departures from the IXP associated with mergers and acquisitions), number of switch ports, and peerings (e.g., new peerings are established, de-peerings, or peering changes such as switching from a public peering arrangement to a private peering). On small time scales (e.g., daily or hourly and below), traffic variations are the main cause for changing IXP conditions.

To address this aspect, instead of analyzing the entire nine months of essentially uninterrupted sFlow measurements from our IXP, we selected four one week-long periods during late April, late August, mid-October, and late November/early December of 2011. We selected weekly periods based on the fact that the AS membership at our IXP was by and large stable during the course of a week. At the same time, choosing four one week-long periods from the nine

months long sFlow measurements results in four snapshots that – as seen from Table 1 – capture some of the churn that our IXP faces on the medium to large time scales. In particular, we note a steady increase in the number of members of our IXP and in the traffic volume they generated during the nine months long measurement period. How these and other changes manifest themselves in the IXP's peering fabric and its use is the theme of the next sections.

In the rest of this paper, we use the Nov/Dec data to illustrate our main findings. Where appropriate, we also include the results for the data of the other three weeks. Overall, we find that their analysis is consistent with the results we report here for the Nov/Dec data. In addition, we also spot-checked our results against a number of additional one week-long data and found no inconsistencies.

2.4 Membership and traffic statistics

To identify the active member ASes at our IXP during a given time period, we had to determine between which member ASes an observed IP packet is being forwarded. To this end, we relied on layer-2 information (i.e., MAC addresses) since the IP addresses in the header of the observed packets were those of the communication endpoints, not the routers on the path. We mapped MAC addresses to router IP addresses and their respective AS numbers by combining link-layer information from sampled ARP packets with routing data obtained from a publicly available looking glass at the IXP. This allowed us to identify 98 % of the members' routers. In the end, we succeeded in determining for more than 99 % of all observed sFlow packet samples their respective originating and receiving member ASes and as a result, we were able to identify for each week more than 350 AS members, each using between one and three logical router interfaces (see the first three rows in Table 1). The remaining less than 1 % of the exchanged traffic volume consists of IPv6 traffic and traffic that we could not associate with any member AS.

Being able to identify the IXP's members for a given time period, we examined next the member ASes of the IXP in each of the four weeks in more detail and report in rows 4 to 8 in Table 1 overall information about their tier level, country and continent. We considered the networks listed in Renesys "baker's dozen" [41] to be tier-1 ISPs and used the AS rank data provided by CAIDA [5] to classify the remaining members as tier-2 or leaf networks[3]. To this end, we classified a member AS as a tier-2 network iff it has both provider as well as customer ASes and as a leaf network iff it has only provider ASes. Based on this straightforward classification scheme that is largely agnostic to network specifics such as business, size, or traffic, we observed that irrespective of the considered time period, the vast majority of members of the IXP are tier-2 networks. At the same time, all the tier-1 ISPs in Renesys "baker's dozen" list are members of our IXP. To determine the country and continent of each member AS, we relied on the country code field that can be found in the AS's whois data. While the IXP members are from more than 40 countries in three continents, most of the member ASes are part of the European Internet scene. To highlight the geographic concentration even more, a majority of those European member ASes offer services in the same country in which our IXP is situated.

Figure 1(b) shows that for the Nov/Dec week, the total daily traffic volume generated by the member ASes and exchanged over the IXP's public switching infrastructure was in the petabyte range and followed a pronounced time-of-day pattern that is well-synchronized with the daily business or user activities in the country where our IXP is situated. In agreement with the observed tier-membership of the IXP's member ASes, the tier-2 networks were responsible for most of the total traffic volume. Rows 9, 10, and 11 in Table 1 provide information about the total traffic volume for each of the four weeks considered, namely the average packets per second, average bits per second, and average daily volume.

Not surprisingly, when breaking down the total traffic volume that traverses the IXP by member AS, we observe a skewed distribution, irrespective of whether we consider sent or received bytes. Indeed, less than 3 % of the member ASes were responsible for about 30 % of the total traffic and less than 30 % of the member ASes were responsible for close to 90 % of the traffic in the Nov/Dec week. Especially noteworthy is that while all tier-1 ISPs are members at this IXP and, with one exception, do exchange traffic over the IXP's public peering infrastructure, contrary to other parts of the Internet, they contribute relatively little to the total traffic volume, presumably because their high volume traffic travels over private peering links supported by the non-public part of this IXP's infrastructure. At the same time, the observed link load of public peerings involving tier-1 ISPs at this IXP was typically higher than average and causes some of those tier-1 ISPs to be included in the 30 % of members that generated close to 90 % of the total traffic volume.

Lastly, an added benefit of working with sFlow records is that it enabled us to examine the IXP traffic by applications. To this end, we separated the ICMP from the TCP and UDP packet samples by looking at their protocol field, and (when possible) associated the TCP and UDP packet samples with an application by looking at their source and destination port numbers and relying on the publicly available lists of port numbers used by the most popular applications. Figure 1(c) shows the application mix for the Nov/Dec week. While this straightforward approach cannot account for roughly 35 % of the bytes, we clearly see that HTTP is the most dominant application, accounting for more than 50 % of the bytes. This observation is consistent with recent reports from inter-AS traffic studies [28] and measurements of residential networks [30]. The next most popular applications are RTMP, HTTPS, and NNTP, the last of which has been reported to be used as a file sharing alternative [26].

3. AN IXP-CENTRIC VIEW OF AS-LEVEL CONNECTIVITY

The logical construct known as the AS-level Internet where nodes represent ASes and links denote AS relationships has no room for directly accounting for physical and geographically well-defined components of the Internet's infrastructure such as IXPs. As a result, an IXP's public peering fabric; i.e., the set of (bi-directional) AS relationships of the peer-peer type (P-P links[4], for short) that exist among pairs of member ASes of the IXP and express routing policies in support of settlement-free traffic exchange among pairs of members over the IXP's public infrastructure is not directly discernible and has received little attention in the past [3]. In this section, we rely on sFlow records from our IXP to obtain the ground truth of this IXP's public peering fabric. We call the compact description that summarizes which member AS is publicly peering with which other member ASes at this IXP the IXP's peering matrix. We then contrast this actual peering matrix with its counterparts derived from analyzing various BGP and traceroute datasets that have been used in the past for inferring AS-level connectivity and generating inferred AS-level maps of the Internet.

[3]In this paper, we decided against using the terms tier-3 and stub because of the different possible interpretations of their meaning.

[4]We are aware that the two parties of a bi-lateral peering agreement are free to (mis)use it as they see fit, e.g., as a customer-provider link. However, it is commonly assumed that most links at IXPs are of the peer-peer type [36, 50]. This has been confirmed by the IXP operator and is supported by our findings in Section 4.2 and 4.4.

3.1 Peering fabric seen from within the IXP

According to a commonly-used definition, two ASes are connected (at a particular time) in the logical AS graph if they can exchange routing information directly, i.e., without the help of an intermediary AS that provides transit, presumably for the purpose of exchanging IP traffic. In the case of our IXP where we know its topology (mapping of MAC and IP addresses to member ASes) and have access to its sFlow records, we use a more pragmatic definition and say that there exists a P-P link between a pair of member ASes if – during a given period of time – we see IP traffic being exchanged between these two member ASes over the IXP's public infrastructure. This pragmatic definition expresses our intention to focus on those P-P links of the IXP's peering fabric that matter; that is, carry actual IP traffic, e.g., BGP packets only in the case of backup links or IP packets generated by genuine application-level traffic. We call the thus-defined peering matrix the "ground truth" for our IXP as it provides the most useful and complete information about the actual status of the peerings between its member ASes.

After filtering the Nov/Dec sFlow records as described in Section 2 and analyzing the resulting traffic, we found that out of a total of 396 $\times$ 395 / 2 = 78,210 (bi-directional) P-P links that the 396 IXP member ASes could potentially establish at the IXP in that time period, more than 50,000 P-P links were actually established and were used to exchange IP traffic. This corresponds to a "peering rate" at our IXP or a "fill degree" of this IXP's (symmetric) peering matrix of about 67 %, meaning that on average, each member AS exchanges IP traffic over the IXP's public infrastructure with some 270 other member ASes. In total, the observed ground truth of this IXP's peering fabric with its more than 50,000 active P-P links is responsible for about 10 PB of traffic that traverses this IXP's public infrastructure daily. Next, we examine how well this IXP's actual peering matrix can be replicated when instead of relying on IXP-provided sFlow records, we are limited by measurements that do not involve the IXP and are obtained from outside the IXP.

3.2 Peering fabric seen from outside the IXP

In the past, BGP routing information (i.e., control-plane data) as well as traceroute measurements (i.e., data-plane information) have been widely used to analyze the structure and evolution of the AS-level Internet. Access to our IXP's actual peering fabric gives us a unique opportunity to evaluate how the various inferred peering matrices for this IXP that result from relying on these different IXP-external datasets compare to the IXP's ground truth.

In terms of BGP routing information, we relied on two well-known sources, i.e., Route-Views (RV) [45] and RIPE NCC (RIPE) [42], and on a non-public dataset (NP). For RV and RIPE, we relied on all their available route collectors, and used both BGP table dumps and updates from the same period when the Nov/Dec sFlow records were collected. NP consists of BGP dumps collected from about 70 routers worldwide which receive BGP information from 724 different ASes also covering the full week. Table 2 provides details about the total number of ASes from which the various datasets obtained BGP data and shows that despite varying significantly in magnitude, the three datasets are by and large complementary and contain routing information from almost 1,000 different ASes.

With respect to traceroute measurements, we used a dataset that resulted from a re-run of the targeted traceroute experiment described in [3]. This experiment was especially designed with the goal of discovering P-P links at IXPs and relied critically on the availability of publicly available traceroute-enabled looking glass (LG) servers throughout the Internet. The re-run was performed during Nov/Dec of 2011 using an updated list of available LG servers. The dataset we considered is derived from all traceroute probes launched as part of this recent campaign and consists of all inferred P-P links that involve our IXP and have an associated high confidence level of representing actual P-P links at our IXP (see [3] for details).

Table 2: Overview of routing and looking glass datasets for November. The numbers show P-P links.

Dataset	Unique LGs / ASN	Visible links	only in this dataset
RV	78	5,336	1,084
RIPE	319	10,913	5,460
NP	723	3,419	684
RV+RIPE+NP	997	13,051	10,472
LG	821 / 148	4,892	2,313
RV+RIPE+NP+LG	1,070	15,364	15,364

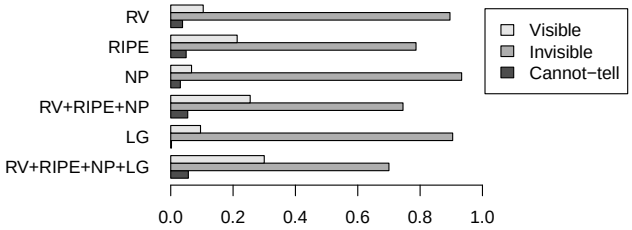


Figure 2: Peering links and visibility in control/data plane (normalized by number of detected P-P links).

To systematically examine which P-P links at our IXP can and which cannot be discovered with the help of which IXP-external datasets, we classify these links into three categories. A **visible P-P link** is a P-P link that is observed both in the IXP-provided sFlow records and the IXP-external datasets (e.g., BGP or traceroute data). A P-P link is called an **invisible P-P link** if it is visible from the IXP-provided traffic data (i.e., IP packets traverse the link), but not visible from the IXP-external datasets. Lastly, a **cannot-tell P-P link** is a P-P link that is visible in BGP data but no traffic exchange is observed between the two member ASes in question from our IXP-provided data. This scenario is typical for private peering arrangements supported by the IXP's non-public infrastructure, but could also arise in those rare situations where a peering is not established at the IXP, or simply not visible in the traffic due to packet sampling. Note that the visible and invisible P-P links add up to the more than 50,000 P-P links that constitute the ground truth of our IXP's peering fabric. Furthermore, since the cannot-tell P-P links cannot be seen from the IXP-provided data, they are not a subset of either the visible or invisible peerings.

Using each of the IXP-external datasets, separately and in different combinations, Table 2 gives *(i)* the total number of visible P-P links that can be seen from the different IXP-external data and *(ii)* the number of unique visible P-P links; that is, those P-P links that can only be seen from exactly one of the IXP-external datasets. When compared to the ground truth, we see that each of the IXP-external datasets misses the vast majority of the observed links, and even when pooling all this available control- and data-plane information, we can still only account for a limited fraction of this IXP's actual peering fabric. A more detailed account of our findings is provided in Figure 2 and illustrates the breakdown of the P-P links into the three different categories of P-P links introduced above. We observe that even when relying on all the available datasets, about 70 % of the P-P links at this IXP remain invisible.

3.3 Some food for thought

A survey of the recent literature on measuring the AS-level Internet shows that as of late 2009, the total number of P-P links in

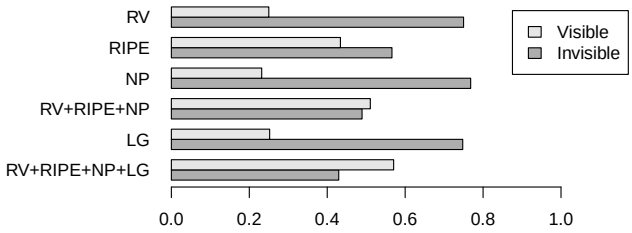


Figure 3: Peering traffic and visibility in control/data plane (normalized by total traffic volume).

the entire Internet was estimated to be in the 35,000-45,000 range. The low end of this range results from adding to the 15,000-20,000 P-P links reported in [16] the roughly 20,000 new P-P links that were discovered in [3] and passed very strict validation criteria. The high end of this range is reported by Chen et al. [11] who used less stringent criteria for validating newly detected P-P links. In stark contrast to these recent estimates, our above analysis shows that the more than 50,000 P-P links that we encountered in this single large European IXP exceed the total number of P-P links assumed to exist Internet-wide. In view of arguments that suggest that many of these P-P links at IXPs are not critical in topology inference [53] or for understanding the evolution of the Internet, for example due to their possible role as backup links [16], we use again our IXP as an example. We show in Figure 3 the fraction of the total traffic traversing the IXP infrastructure that would not be accounted for if we only knew about the visible links; that is, the P-P links whose existence at this IXP can be inferred from the various BGP or traceroute data. Figure 3 shows that when using these IXP-external datasets individually to infer the visible links, each of them misses between 56–78 % of the total traffic (in bytes or packets) handled by this IXP. Even when pooling all the IXP-external datasets, close to half of the total traffic would be missed, due to the large number of P-P links that are not seen. Therefore, trying to gain insight into the economic incentives and business reasons of the various member ASes of this IXP for establishing the encountered peering fabric would be very hard knowing the visible peerings only.

Table 2 and Figure 2 show why efforts to unveil the peering fabric at this IXP (or others) by using publicly available or even privately collected BGP data or relying on measurements obtained from carefully designed traceroute experiments are essentially doomed. The main reason is the well-known problem of vantage points [44]. On the one hand, as shown in [38], the locations within the AS-level Internet of the monitors traditionally used to collect the widely-used BGP data provide a relatively accurate picture of the Internet AS-level connectivity as far as AS links of the customer-provider type are concerned, reportedly missing less than 11,000 out of a total of about 94,000 of such links Internet-wide [11]. On the other hand, these monitors have hardly any visibility into the Internet's IXP substrate consisting of the various IXPs, their member ASes and the P-P links among them at those IXPs [37] and thus miss the majority of those links. At the same time, even when trying to use traceroute measurements and launch probes from LG servers close to an IXP to a target in an AS "on the other side" of the IXP, due to AS-specific routing policies, there is no guarantee that the probes traverse the IXP and improve the discovery of P-P links at the IXP.

4. DIVERSITY OF THE IXP ECOSYSTEM

In this section, we take a closer look at our IXP's ecosystem; that is, its member ASes, the rich peering fabric we described in Section 3, and various aspects of the traffic that is exchanged among the IXP's member ASes over this peering fabric.

4.1 Member ASes

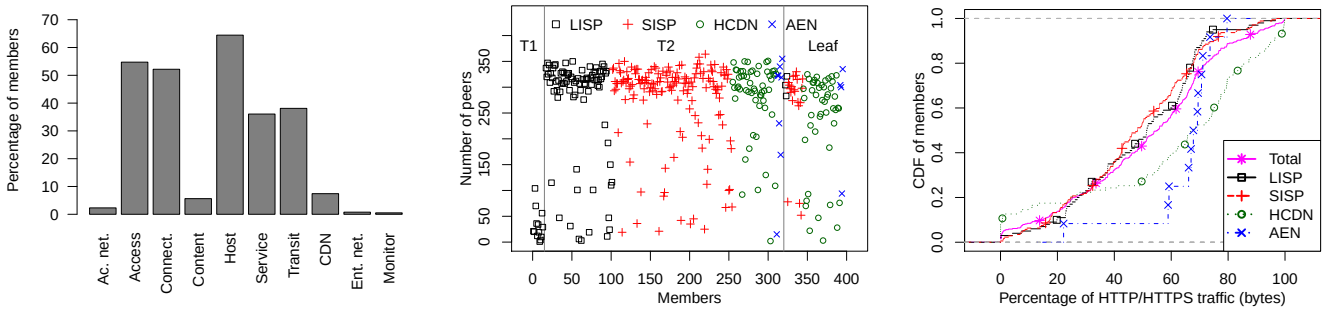
We have already noted that the traditional classification of networks into tiers says little about the nature and nothing about the business types of the networks that are members at this IXP (see Section 2). Unfortunately, there is no readily available dataset which lets us determine the business type(s) of each member AS. Therefore, we manually examined the information available on each of the member ASes' web sites and present in Figure 4(a) their business type(s). Clearly, the business model of the member ASes differ significantly, and it is not uncommon to encounter member ASes that are in multiple business types. Focusing on their main business type, we further classified each of the member ASes as a *large ISP* (LISP), *small ISP* (SISP), *hosting/service and content distribution network* (HCDN), and an *academic and enterprise network* (AEN). A large ISP is providing transit, connectivity, eyeball access and additional services such as hosting or even content distribution. A small ISP is an access provider and may also provide transit services. Hosting and service providers are hosting content, either indirectly through providing web-space or rack-space to actual creators of content, or as content owners. Some of them also provide special services such as DNS. The AEN category comprises all networks that are solely used to connect enterprises and universities.

4.2 Peering

For each of our IXP's member ASes, Figure 4(b) shows the number of its P-P links; that is, the number of other members with which it peers at this IXP. The member ASes are ordered (x-axis) according to our classification as introduced in Section 2 (i.e., tier-1, tier-2, and leaf networks), with no particular order within the resulting groups. Figure 4(b) reveals an enormous diversity with respect to the number of peers – some member ASes peer only with a few other members, while others peer with almost all of them. In particular, we see that the tier-1 ISPs have a small number of P-P links, typically peering with less than 25 % of all member ASes. This observation is consistent with their stated intention of offering only restrictive peering (e.g., on peeringDB [39], on the IXPs web site, or on the companies' web sites). Tier-1 ISPs apparently use the IXP, among other reasons, to augment their existing peerings, but need to do this with care because most other member ASes are either transit customers or potential transit customer for them.

Non-tier-1 members typically peer with a large number of other members, with about 71 % of the tier-2 and leaf member ASes peering with at least 70 % of the other members. Among the tier-2 and leaf networks (more than 96 % of the members), there are less than 10 % which peer with less than 25 % of all members. Tier-2 and leaf member ASes usually have an open peering policy, meaning that when asked what they look for in a peering, their answers are mainly performance and reducing transit costs. Some prefer selective over open peering policies, especially if setting some standards for a potential peer's network in terms of criteria such as traffic exchange ratio, geographic scope, backbone capacity, or traffic volume is in their interest. We find that most member ASes at this IXP use open peering policies and peer massively with other members. However, we also encounter ASes with open peering policies that do not peer with that many other members. Possible reasons for their low peering rate are *when* they joined the IXP, *if* they offer desirable content or *if* they provide Internet access to a significant number of eyeballs.

Figure 4(b) also shows a classification of the member ASes in the four business categories defined above: LISP, SISP, HCDN, and



(a) AS classification by business types. (b) Scatter-plot of num. of peers per member. (c) Fractions of web-traffic across members.

Figure 4: Diversity in members: business type, number of peerings, and application mix exemplified by web-traffic.

AEN. Based on this classification, we find that in the LISP group, the member ASes with a small number of peerings are the tier-1 ISPs and those ISPs with a selective peering policy. In the HCDN group, the networks with a few peerings include some of the large players, but also small hosting providers (e.g., for banks or online games). The picture is less clear for the SISP group. In general, the observed large number of member ASes that have a large number of peers at this IXP is testimony for the ease with which member ASes can peer at this (and other) IXP. In fact, the findings of a recent survey [50] provide compelling reasons – some 99 % of the surveyed peerings were a result of "handshake" agreements (with symmetric terms) rather than formal contracts, and an apparent prevalence of multi-lateral peering agreements; that is, the exchange of customer routes within groups of more than two parties.

4.3 Traffic

The contributions to the IXP's overall traffic by the individual member ASes is highly skewed, with the top 30 % of member ASes contributing close to 90 % of the overall IXP traffic. Examining in more detail the traffic volume that each member AS contributes to the IXP's overall traffic, we first investigate what role the traffic exchange ratio plays in establishing P-P links. To this end, we consider the traffic asymmetry across all peerings between any two member ASes and show in Figure 5(a) the empirical cumulative probability distribution of this asymmetry. For improved readability we only show the part of the curve for ratios up to 100:1 (75 % of all peerings). The figure reveals a high variability in terms of exchanged traffic between the two member ASes of a peering. Indeed, only 27 % of the links have a traffic ratio of up to 3:1 (see support lines), where a 3:1 ratio is often stated as a typical requirement in common formal peering agreements [35]. Moreover, for 8 % of the peerings the ratio exceeds 100:1, and for another 17 % we observe traffic in only one direction. Figure 5(a) also depicts the empirical cumulative probability distribution for the P-P links at this IXP involving only tier-1 ISPs and shows that these peerings are less asymmetric, with more than 33 % of them having a ratio below 3:1.

Figure 5(b) shows the traffic asymmetry of the member ASes (i.e., the ratio of outgoing bytes vs. incoming bytes of a given member AS). The traffic of 52 % of the member ASes is more or less symmetric and within the range of 1:3 to 3:1. However, a significant number of member ASes fall in the 3:20 to 20:3 range[5]. In agreement with expectations, HCDNs have more outgoing than incoming traffic, while the opposite is true for LISPs and SISPs. However, there are various exceptions to this rule, and we find HCDNs with significantly more incoming than outgoing traffic and LISPs and SISPs where the opposite holds true. Note that despite the significant diversity in the ratio of incoming and outgoing traffic, more than half of the member ASes that send most of the traffic also receive most of the traffic. Indeed, there is a 50 % overlap among the top 50 member ASes according to bytes sent and the top 50 member ASes according to bytes received.

We can also examine how similar or dissimilar the overall application mix (see Section 2) is across all the IXP member ASes. For example, when computing for each member AS the fraction of HTTP/HTTPS traffic relative to the total number of bytes sent and received, we find in Figure 4(c) that this application mix differs significantly across the member ASes and follows almost a uniform distribution, indicating that without additional information, it would be difficult to predict which percentage of a member AS's traffic is HTTP. However, as soon as we include for example information about the member AS's business type, we observe that as expected, hosting providers and CDNs tend to send a larger fraction of HTTP traffic. However, rather unexpectedly, we also see more than 10 % of the hosting providers and CDNs with only marginal fractions of HTTP traffic. Closer inspection shows that these member ASes are primarily service providers that do not provide web content.

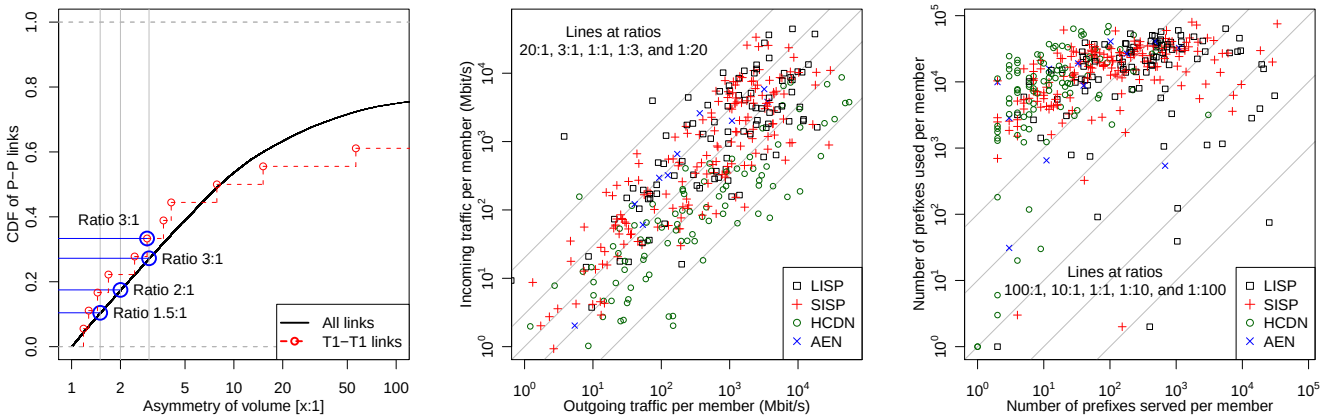
4.4 Prefixes

We next consider the prefix exchange ratio. For this purpose, we say that a prefix is *served* by a member AS if the member AS receives traffic for that prefix. Vice verse, we say that a prefix is *used* by a member AS if output traffic of its access router is destined toward that prefix. Figure 5(c) depicts a scatter-plot of the ratio of the number of prefixes used vs. the number of prefixes served by each member AS and provides clear evidence that the vast majority of the member ASes of our IXP use more than 10-times the number of prefixes they serve. Specifically, we see that hosting providers and CDNs have a tendency to serve a smaller number of prefixes but to use some two orders of magnitude more prefixes. Focusing on the ISPs, we can identify two groups. The first, larger group, serves a diverse but limited set of prefixes, from a few tens to a few thousands. The second, smaller group, serves and uses a large number of prefixes, some tens of thousands. Members that serve such large numbers of prefixes are likely acting as transit networks for other member ASes. However, we again observe exceptions to these general observations in almost all categories.

4.5 Geographical aspects

Conventional wisdom about IXPs states that ASes join regional

[5]To illustrate, if we had a member AS that would only deliver content using 1,500 byte-sized packets, the ratio could be as bad as 1:58, assuming on average one ACK of 52 bytes for every two data packets of 1,500 bytes and no overhead for the TCP connection establishment.



(a) Traffic asymmetry across P-P links. (b) Traffic asymmetry (out/in) per member. (c) Prefix ratio per member AS.

Figure 5: Diversity in traffic asymmetry and use of prefixes.

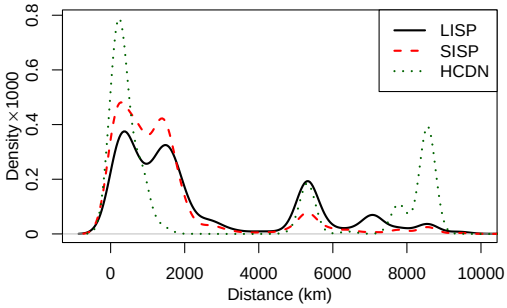


Figure 6: Geographic distances of IP endpoints to IXP.

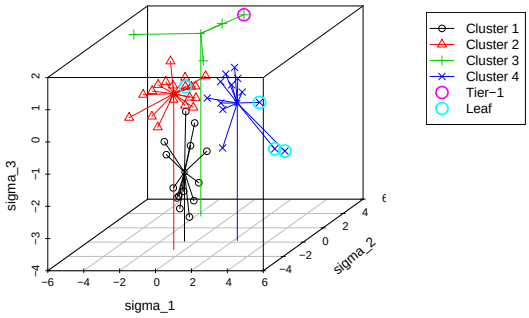


Figure 7: SVD-based 3D projection based on 12 features of the top 50 member ASes by bytes sent.

IXPs to exchange local traffic. To examine this general belief, we used the MaxMind GeoLite City database [31] to identify the geographic coordinates of both source and destination IP addresses for each sampled packet. Despite known inaccuracies of this geolocation database, using it for the needs of this study is appropriate, as we are only interested in approximate distances at the country level [40]. Contrary to our expectations, we found that only 10 % of the traffic is exchanged within the country in which the IXP is situated, while another 26 % originates from that country, and another 3 % is destined for that country. However, when relaxing the geographic constraint and considering a geographic region within a radius of 2,000 km from our IXP, we confirm the local nature of IXP traffic – almost 80 % and 72 % of the traffic terminates and originates, respectively, within this relatively close proximity.

To better understand the geographic reach of our IXP, Figure 6 depicts the density of the distances of traffic originated by member ASes in the LISP, SISP, and HCDN groups, weighted by byte volume. The distances shown in this figure are measured from the IP source address to the IXP, i.e., they represent the geographic range from which the IXP attracts traffic. We find that HCDNs have the largest fractions of very short distance traffic and at the same time the largest fraction of very long distance traffic–37 % of traffic volume with a distance larger than 5,000 km suggests the presence of significant intercontinental traffic. This indicates either mismatches in the IP address location [40] and/or that some of the traffic is indeed being served from remote locations. Likewise, member ASes in the LISP group show strong presence at around 5,000 km, which is mainly contributed by a small number of large international ISPs. SISPs and AENs (not shown) typically send traffic from closer to the IXP than members in the other business groups.

4.6 Tiers without tears

The above analysis highlights the diversity of the member ASes in terms of business types, the number of peerings, as well as their traffic characteristics. We have already seen indications that the traditional classification of networks by tiers cannot account for this observed diversity, mainly because it is agnostic to features of the member ASes such as their business type, traffic, peerings, prefixes, and geographic properties. Clearly, these and other features have the potential of painting a much more interesting and relevant picture of networks compared to what is possible knowing only the presence or absence of provider and customer networks.

In the rest of this section, we explore the possibility of combining some of these features and identify meaningful clusters. To this end, we consider 12 features in an attempt to characterize the member ASes' peerings and traffic characteristics: number of bytes sent, number of bytes received, number of peers, number of ASes and prefixes they send traffic to and receive traffic from, percentage of HTTP traffic that they send and receive, and 25-percentile of the distances from the traffic source to the destination, as well as from the IXP location to the destination for outbound traffic, and from the traffic source to the IXP location for inbound traffic.

We consider the Singular Value Decomposition [27] (SVD) of the 396×12 matrix and look at its projection into the 3D space defined by the three eigenvectors corresponding to the three largest singular values. Intuitively, the SVD produces a set of combined features (i.e., linear combinations of the original variables) so that the variability of the values of the first few combined features is maximized. Figure 7 shows the resulting 3D figure as a scatter-plot.

To keep the number of points reasonable, we sub-selected the top 50 member ASes according to sent bytes. Similar plots result if we sub-select according to number of bytes received or if we increase the number of member ASes to the top-100. In generating Figure 7, we repeatedly clustered the selected member ASes using the k-means clustering algorithm with random starting points into four clusters. Increasing or decreasing the number of clusters considered (i.e., value k) had no major impact on the nature of the results.

The output of this combined SVD/clustering method typically consists of one small cluster (~5 member ASes), two medium clusters (~10 member ASes) and one large cluster (~20 member ASes). When examining the clusters in more detail, we find that (i) the small cluster contains only large content and service providers, (ii) one of the medium cluster contains mainly small to medium ISPs that provide access to residential and enterprise customers and are all located in a country different from the IXP, (iii) the second medium cluster has mainly large ISPs and backbone networks that provide transit, and (iv) the big cluster has mainly data centers, big hosting providers, CDNs and some ISPs that provide web and server hosting. To highlight one such example clustering, we use in Figure 7 different symbols and colors and connect the clusters with a spider and mark their centers by support lines. We view Figure 7 as evidence that it is possible to classify an IXP's member ASes in ways that are practical and meaningful and respect the real-world diversity among networks that are critical elements of an IXP's ecosystem. Importantly, annotating the individual points in Figure 7 with tier-information shows why conventional tier classification is uninformative and of little help when trying to understand the Internet ecosystem locally (i.e., at this IXP) or globally.

5. IXP TRAFFIC MATRIX

Many IXPs report up-to-date traffic statistics on their web sites, and some of the largest European IXPs show daily traffic volumes for their public switching infrastructure that have been consistently in the petabyte range for some time. In Section 2, we confirm this publicly available but little-known fact for our IXP. A breakdown of this overall traffic by member ASes can be compactly described by an IXP's traffic matrix that specifies for example the hourly or daily traffic volumes exchanged between all member ASes that have a P-P link at this IXP. Despite the many similarities with an ISP's intra-AS traffic matrix, we are not aware of any published research paper that has considered IXP-specific traffic matrices and their properties. This is largely a reflection of the attention that large ISPs (and big content) have received from researchers and an indication that IXPs have been viewed as relatively uninteresting in terms of their topology, traffic, and routing. As is the case with ISP-provided measurements for computing or inferring an ISP's intra-domain traffic matrix, access to IXP-provided data for studying IXP-specific traffic matrices is similarly critical, and we rely in the following on our IXP-provided sFlow measurements. We report below on a first-of-its-kind analysis of the actual traffic matrix of one of the largest IXPs worldwide, examine in detail properties such as the diurnal pattern, sparsity, and (approximately) low rank, and discuss possible applications of our findings in support of managing and operating a large IXP's infrastructure.

5.1 Temporal properties

We have seen in Section 2 (see Table 1) that during our nine months-long measurement period, the overall traffic seen by our IXP steadily increased, mainly due to a similarly steady increase in the number of its member ASes. In addition, Figure 1(b) shows that the total traffic volume over time is dominated by a pronounced time-of-day characteristic, where the observed diurnal cycle coincides with the daily business cycle in the country where this IXP is located. Such diurnal behavior has long been a trademark of the temporal nature of measured intra-domain ISP traffic matrices.

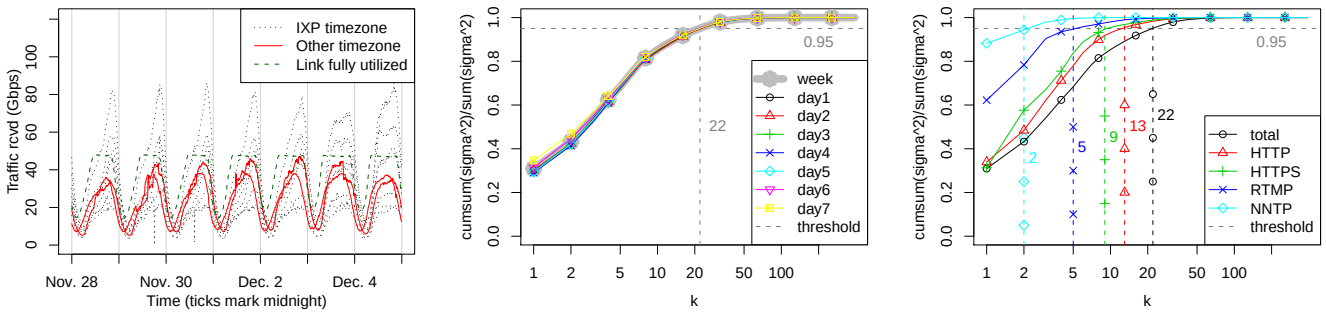
To explain this diurnal behavior for our IXP, note that *(i)* the tier-2 member ASes are responsible for a majority of the overall traffic (see Figure 1(b)) *(ii)* the top-10 of these tier-2 member ASes generate much of the IXP's traffic (more than 33 %), and *(iii)* despite the tier-2 member ASes being a very heterogeneous group of networks, a majority of them cover the city, region, or country where our IXP is situated. Therefore, when plotting the incoming traffic volume for the top-10 receivers among the tier-2 member ASes in Figure 8(a), we see that the temporal behavior of their traffic is well-correlated with the overall traffic and well-aligned with the daily business cycle of the area where our IXP is located. An exception to this rule is the traffic of the two member ASes shown with solid lines in Figure 8(a) that are shifted by some 1–5 hours as a result of representing the traffic of ISPs serving geographic areas in different time zones, one in Europe and one outside of Europe. The traffic of the member AS plotted with dashed lines in Figure 8(a) visually reaches the bandwidth limits of its IXP network interfaces. We were able to confirm this observation by closer inspecting the sFlow records, which revealed that this particular member AS connects with five physical 10 Gbps links to the switching fabric of the IXP, thus it is indeed reaching its 50 Gbps capacity limit.

Knowing such temporal properties of an IXP's traffic matrix takes on a new meaning in an environment where innovation in the IXP marketplace in the form of new types of service offerings or refined routing policies is likely to increase the volatility and decrease the predictability of the traffic seen by the IXP. For example, in the presence of IXP port re-sellers, the IXP has limited visibility into the networks that use the re-seller as an intermediary and only indirect control over the traffic that these networks send or receive at the IXP. Similarly, members that can select prefix-specific peering to send certain traffic at certain times through the IXP instead of handing it off to their upstream provider(s) are likely to be a significant source for increased traffic volatility at IXPs. These are instances where a detailed understanding of the temporal dynamics of an IXP's traffic matrix will be critical for accurate prediction and successful root cause analysis of observed infrastructure performance problems.

5.2 Structural properties

A readily observable structural property of measured ISP-specific traffic matrices is their sparsity. The fact that generally only a small number of entries in these traffic matrices are populated or non-zero is largely due to a carefully planned network and complex routing decisions. In stark contrast, as reported in Section 3, with a "fill degree" of more than 65 %, our IXP-specific traffic matrix cannot be called sparse, and the main reason for this observed non-sparsity is economics. More precisely, an IXP's infrastructure and routing requirements are purposefully kept simple to facilitate easy network interconnection among member ASes, resulting in a rich peering fabric in support of traffic exchanges at the IXP between as many member ASes as dictated by "good" business practices.

Next, we examine how our IXP-specific traffic matrix compares to measured ISP-specific traffic matrices with respect to their widely-reported property of having approximately low rank. Recall that in SVD terminology, a matrix X of (algebraic) rank k has approximately or nearly low rank if only a small number $k^* \ll k$ of the largest singular values are needed to well-approximate X by a $k^* \times k^*$ matrix under a certain norm of the approximation error (see for example [52] for details). In practice, to check if a matrix has approximately low rank, we consider its singular values $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_n$, compute the energy function defined



(a) Daily pattern of top-10 tier-2 members. (b) Traffic matrix energy. (c) Traffic matrix energy by application.

Figure 8: Traffic matrix properties.

by $f(k) = (\sum_{i=1}^{k} \sigma_i^2)/(\sum_{i=1}^{n} \sigma_i^2)$, and find the smallest k^* that captures, say, 95 % of the total energy.

Plotting the energy $f(k)$ as a function of k, Figure 8(b) shows the results of applying this method to a week-long traffic matrix (trace from Nov/Dec, thick line) and the corresponding seven daily traffic matrices (thin lines). In addition to confirming the approximately low rank of these traffic matrices (i.e., out of some 380 non-zero singular values, only the 22 largest are needed to capture 95 % of the energy), the plot also illustrates consistency among the week-long trace and the different daily traces. Similarly, recalling the application mix of the total traffic (see Figure 1(c)), Figure 8(c) shows the approximately low rank nature of the week-long application-specific traffic matrices. We observe that the low rank nature becomes more pronounced as we consider, in order, HTTP, HTTPS, RTMP, and NNTP, even though all these application-specific traffic matrices have almost full rank. In fact, in the case of NNTP, a simple 2×2 matrix can successfully capture most of the NNTP portion of the traffic in the entire original IXP-specific traffic matrix of size 396×396. Even for the case of RTMP, only a handful of singular values are needed to well approximate the portion of the overall traffic generated by this application. Another more specialized set of traffic matrices that also have approximately low rank can be constructed by considering only that portion of the overall IXP traffic that is produced by the top-50 member ASes (in terms of sent bytes) that were used as input to the clustering study described in Section 4 (see Figure 7) and ended up in one and the same cluster. A reason to examine such specialized traffic matrices is to look for any connection between the different business types that roughly specify these clusters and the low rank nature of the corresponding traffic matrices, and preliminary results (not shown here) suggest that the answer is affirmative.

While there are many other types of IXP-specific traffic matrices that can be considered, having approximately low rank is a common property among them and hints at the presence of enormous amounts of structure that is hidden in real-world IXP-specific traffic matrices and begs the question how this structure could possibly be exploited for practical purposes. One promising such application concerns efficient data acquisition. Specifically, as an IXP's infrastructure grows in terms of member ASes, peerings, and traffic, existing monitoring infrastructures that rely on increasingly lower sampling rate to keep up with the encountered growth in traffic may no longer be viable, especially in view of more stringent performance criteria that are promised by the IXP and typically require higher-resolution IXP measurements. The approximately low rank nature of IXP traffic matrices suggests a viable alternative whereby fewer but more intelligently collected (and hence more expensive) measurements can provide as much information as the current brute-force method of throwing more hardware at the problem to support the collection of maximal amounts of highly redundant data.

6. DISCUSSION

It is generally known and understood that BGP-based efforts for discovering P-P links in the AS-level Internet have somewhat limited success and provide at best a lower bound for the number of such links in the Internet [16]. However, that the number of P-P links at a single IXP exceeds even very recently reported such lower bounds [3, 11, 16] has come as a big surprise. This and the realization that even when laboriously combining the most up-to-date publicly available BGP data with hard-to-get non-public control-plane measurements and the latest available state-of-the-art data-plane measurements, our pooled data can only account for some 30 % of all known P-P links at our IXP begs the question how we can miss so many actual P-P links and why. Clearly, obscuring the existence of a "live" P-P link can occur with control-plane data (e.g., the route server from which BGP data is pulled not being close enough to the IXP) and with data-plane measurements (e.g., due to routing policies that prevent direct traffic exchange via an existing P-P link at the IXP and force traffic between two member ASes to take the upstream path). We plan to study the role of routing policies in hiding existing P-P links at IXPs as part of our future work and expect that an in-depth understanding of the root causes will suggest novel measurement experiments that have the potential of providing an accurate and near-complete peering matrix for each IXP and, in turn, an approximately valid snapshot of the AS-level Internet.

Irrespective of the reasons for the enormous number of encountered P-P links at our IXP, we have seen that the observed rich peering fabric by and large defies the well-known tiered structure of the Internet. As such, it is much more a reflection of the varied economic incentives and business benefits that drive the different member ASes to massively peer at this IXP and an indication of the ease with which such P-P links can be established. The resulting rich peering fabric supports massive direct interconnections or "shortcuts" as viable alternatives to sending traffic upstream and by and large agrees with recently reported findings of a "flattening" of the Internet. In fact, while some of our results are largely complementary to those of reported in [28], they do provide a different perspective. Relying exclusively on IXP data (as compared to the use of non-IXP-only data), we use an analysis of the IXP members' business types and a port-based application classification of their traffic to observe a similar consolidation of content and applications. However, our explanation for this consolidation centers more around the discovered massive peerings among all kinds of networks at this IXP and relies less on the presence and formation of "hyper-giants".

While this "flattening" of the Internet can fully co-exist with

the Internet's traditionally assumed hierarchical structure, our IXP-specific findings and the following cautious extrapolations to the Internet as a whole suggest an even more radical change in perspective of the AS-level Internet. Indeed, considering only the European IXP scene (see [18] for details) and being conservative in using our large IXP as a baseline (i.e., assuming only a 50 % peering rate at IXPs), counting up the P-P links we expect to encounter at the four largest IXPs (with, say, 400 unique members each) and at the 10 next-largest IXPs (with, say, 100 unique members each), we obtain a realistic lower bound for the estimated number of P-P links for just the European portion of the Internet of some 200,000. This number is more than 100 % larger than the number of *all* AS links (i.e., customer-provider and peer-peer) in the entire Internet in 2010 as reported in [16] or the number of *all AS links of the customer-provider type* Internet-wide in 2008 as reported in [11][6].

Despite being extremely conservative, this estimate by itself should give reasons to pause. First, it indicates that there are easily an order of magnitude more P-P links in today's Internet than previously assumed. Second, contrary to conventional wisdom, there are many more P-P links in today's Internet than customer-provider type peerings, with twice as many being a conservative estimate. Third, judging from what we have seen at our IXP, most of these massive amounts of P-P links are of critical importance as they carry significant traffic. Given that past studies of the Internet peering ecosystem assumed exactly the opposite of what our findings support, there is a need for a major overhaul of the mental picture that our community has about the AS-level Internet, not only in terms of local and overall structure, but also with respect to its evolution in response to often rapidly changing business conditions locally at an IXP or within the larger Internet. In particular, we argue that assessing the "standing" of a network within the Internet's ecosystem has to account for network-specific features (e.g., business type) and some notion of traffic that this network is responsible for, either as a source, sink, or transit entity. This renewed focus on traffic requires novel approaches for the measurement, modeling, analysis, and inference of the Internet's inter-AS traffic matrix. Despite some initial efforts dealing with this matrix and a large body of existing work on intra-AS traffic matrices, the Internet inter-AS traffic matrix has remained a big enigma, but the availability of IXP-specific traffic matrices promises to invigorate research activities in this area.

7. RELATED WORK

Over the past years, the AS-level Internet has been a much-studied graph structure and continues to fascinate networking and non-networking researchers alike, though typically for different reasons. Instead of attempting to provide a necessarily incomplete overview of the existing literature on this topic, we refer the reader to a number of recent studies that serve as useful surveys [14, 16, 22, 44]. A majority of published research in this area has focused on measuring, inferring, modeling, and characterizing the AS-level Internet, often for the purpose of applying inferred AS graphs or carefully-tuned models to specific problems; e.g., [15, 17, 21, 32, 46].

Given that as logical constructs, AS topologies have no explicit space for physical infrastructure components, IXPs have long been neglected and generally viewed as an unimportant "detail". This view has slowly started to change with the gradual realization of the existence of "hidden" links in the Internet AS topology [8, 12, 23, 24, 37, 51]. [3] is the first traceroute-based study that purposefully targeted the existing IXPs worldwide. In conjunction with improvements on the measurement side, there has also been an increasing awareness of the important role that IXPs play in the Internet ecosystem, e.g., see [6, 7, 48, 49]. Ironically, among network operators, this realization has been there pretty much from the beginning of the commercial Internet [25, 33, 34], and there are signs that the research community is starting to give the Internet's IXP substrate the attention it deserves.

As far as traffic matrix research is concerned, despite some recent efforts [4, 9, 28, 44], little (if anything) is known about the Internet's inter-domain or inter-AS traffic matrix. The pieces of this puzzle that have received most attention by researchers have been the tier-1 ISPs and their intra-domain traffic matrices (e.g., see [1, 10, 53] and references therein). In fact, the latter have been a key ingredient for many ISP-critical tasks such as traffic engineering, capacity planning, and anomaly detection [20, 29, 43]. In stark contrast, when it comes to IXP-specific traffic matrices, only very recent work [6] has considered peering and traffic trends through a longitudinal study of a small European IXP across a period of 14 years. However, no study has gone into the same level of details as for intra-domain ISP traffic matrices. As is the case with ISP-provided measurements for computing or inferring an ISP's intra-domain traffic matrix, access to IXP-provided data for studying IXP-specific traffic matrices is similarly critical and requires close collaborations with IXPs. For the work reported in this paper, we have been fortunate to be able to work closely with one of the largest IXPs worldwide.

8. CONCLUSION

Examining readily available public information about a number of large European IXPs shows that they are very similar, not only with respect to the make-up of their constituents (i.e., member ASes) and overall traffic volumes, but also in terms of offered services, underlying technologies, business models, and overall purpose. As such, access to detailed internal measurements from even just one such IXP can highlight the important role that these largely ignored entities play for the Internet as a whole and for the particular geographic regions where they are located.

To this end, we analyze in this paper a unique data set of nine months' worth of continuous sFlow measurements from one of the largest IXPs in Europe, and worldwide, and clarify in the process some common misconceptions that exist regarding IXPs and the AS-level Internet. These include, among others, that tier-1 ISPs do not peer at IXPs (they do), IXPs are not used for transit (they are), the number of peer-peer links in the Internet is small (it is at least an order of magnitude larger than what has been assumed), the number of customer-provider links in the Internet is much larger than the number of peer-peer links (there are easily twice as many peer-peer links than customer-provider links), and IXP peerings are mostly used for back-up (they are not). In particular, we examine in detail the peering fabric and traffic matrix of our IXP and show the existence of a very diverse ecosystem in terms of the member ASes' business types, peering strategies, traffic exchanges, and geographic coverage that mimics the Internet's AS ecosystem as a whole. We argue that these findings are a proof that the mental picture our community has about IXPs and the AS-level Internet is in much need for a major overhaul.

Acknowledgments

We want to express our gratitude towards the IXP for their generous cooperation and support. We thank Mario Sanchez (Northwestern University) for re-designing and re-running the traceroute experiment originally described in [3] and for providing an up-to-date dataset (i.e., LG in Table 2). We also thank our shepherd Sergey

[6]Note that the reported numbers involving customer-provider links are from the published literature and cannot be derived from our IXP data.

Gorinsky and the anonymous reviewers who provided useful feedback. This work was supported by a G-Lab grant from the Federal Ministry of Education and Research of Germany (BMBF) and a Leibniz Prize grant from the German Research Foundation (DFG).

9. REFERENCES

[1] V. K. Adhikari, S. Jain, and Z.-L. Zhang. YouTube Traffic Dynamics and Its Interplay with a Tier-1 ISP: An ISP Perspective. In *Proc. of ACM IMC*, 2010.

[2] AT&T Company Information, 2012. http://www.att.com/gen/investor-relations?pid=5711.

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

[4] V. Bharti, P. Kankar, L. Setia, G. Gürsun, A. Lakhina, and M. Crovella. Inferring Invisible Traffic. In *Proc. of ACM CoNEXT*, 2010.

[5] CAIDA AS Rank:AS Ranking. http://as-rank.caida.org/.

[6] J. C. Cardona Restrepo and R. Stanojevic. A history of an Internet eXchange Point. *ACM CCR*, Apr. 2012.

[7] I. Castro and S. Gorinsky. T4P: Hybrid Interconnection for Cost Reduction. In *Proc. of NetEcon*, 2012.

[8] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Towards Capturing Representative AS-Level Internet Topologies. *Computer Networks*, 2004.

[9] H. Chang, S. Jamin, Z. Mao, and W. Willinger. An Empirical Approach to Modeling Inter-AS Traffic Matrices. In *Proc. of ACM IMC*, 2005.

[10] H. Chang, M. Roughan, S. Uhlig, D. Alderson, and W. Willinger. The Many Facets of Internet Topology and Traffic. *Networks and Heterogeneous Media*, 2006.

[11] K. Chen, D. R. Choffnes, R. Potharaju, Y. Chen, F. E. Bustamante, D. Pei, and Y. Zhao. Where the Sidewalk Ends: Extending the Internet AS Graph Using Traceroutes From P2P Users. In *Proc. of ACM CoNEXT*, 2009.

[12] R. Cohen and D. Raz. The Internet Dark Matter—On the Missing Links in the AS Connectivity Map. In *Proc. of IEEE INFOCOM*, 2006.

[13] Deutsche Telekom ICSS, 2012. http://ghs-internet.telekom.de/dtag/cms/content/ICSS/en/1222498.

[14] A. Dhamdhere and C. Dovrolis. Ten Years in the Evolution of the Internet Ecosystem. In *Proc. of ACM IMC*, 2008.

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

[16] A. Dhamdhere and C. Dovrolis. Twelve Years in the Evolution of the Internet Ecosystem. *IEEE/ACM Trans. Netw.*, 2011.

[17] D. Dolev, S. Jamin, O. Mokryn, and Y. Shavitt. Internet Resiliency to Attacks and Failures under BGP Policy Routing. *Computer Networks*, 2006.

[18] Euro-IX—European Internet Exchange Association. https://www.euro-ix.net/resources.

[19] J. Fan, J. Xu, M. H. Ammar, and S. Moon. Prefix-Preserving IP Address Anonymization: Measurement-Based Security Evaluation and a New Cryptography-Based Scheme. *Computer Networks*, 2004.

[20] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, J. Rexford, and F. True. Deriving Traffic Demands for Operational IP networks: Methodology and Experience. *IEEE/ACM Trans. Networking*, 2001.

[21] P. Gill, M. Schapira, and S. Goldberg. Let the Market Drive Deployment: A Strategy for Transitioning to BGP Security. In *Proc. of ACM SIGCOMM*, 2011.

[22] H. Haddadi, G. Iannaccone, A. Moore, R. Mortier, and M. Rio. Network Topologies: Inference, Modelling and Generation. *IEEE Communications Surveys and Tutorials*, 2008.

[23] Y. He, G. Siganos, M. Faloutsos, and S. Krishnamurthy. A Systematic Framework for Unearthing the Missing Links: Measurements and Impact. In *Proc. of NSDI*, 2007.

[24] Y. He, G. Siganos, M. Faloutsos, and S. Krishnamurthy. Lord of the Links: A Framework for Discovering Missing Links in the Internet Topology. *IEEE/ACM Trans. Networking*, 2009.

[25] G. Huston. Interconnections, Peering and Financial Settlements. In *Proc. of INET*, 1999.

[26] J. Kim, F. Schneider, B. Ager, and A. Feldmann. Today's Usenet Usage: Characterizing NNTP Traffic. In *Proc. of IEEE Global Internet*, 2010.

[27] V. Klema and A. Laub. The singular value decomposition: Its computation and some applications. *IEEE Transactions on Automatic Control*, 1980.

[28] C. Labovitz, S. Lekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet Inter-Domain Traffic. In *Proc. of ACM SIGCOMM*, 2010.

[29] A. Lakhina, M. Crovella, and C. Diot. Diagnosing Network-Wide Traffic Anomalies. In *Proc. of ACM SIGCOMM*, 2004.

[30] G. Maier, A. Feldmann, V. Paxson, and M. Allman. On Dominant Characteristics of Residential Broadband Internet Traffic. In *Proc. of ACM IMC*, 2009.

[31] GeoLite City. http://www.maxmind.com/app/geolitecity/.

[32] W. Muhlbauer, A. Feldmann, O. Maennel, M. Roughan, and S. Uhlig. Building an AS-topology Model that Captures Route Diversity. In *Proc. of ACM SIGCOMM*, 2006.

[33] W. B. Norton. The Evolution of the U.S. Internet Peering Ecosystem. *Equinix White Papers*, 2004.

[34] W. B. Norton. A Business Case for Peering in 2010, 2010. http://drpeering.net/.

[35] W. B. Norton. A Study of 28 Peering Policies, 2010. http://drpeering.net/.

[36] W. B. Norton. Public vs Private Peering : The Great Debate, 2010. http://drpeering.net/.

[37] R. Oliveira, D. Pei, W. Willinger, B. Zhang, , and L. Zhang. In Search of the Elusive Ground Truth: The Internet's AS-level Connectivity Structure. In *Proc. of ACM SIGMETRICS*, 2008.

[38] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. The (In)completeness of the Observed Internet AS-Level Structure. *IEEE/ACM Trans. Networking*, 2010.

[39] PeeringDB. http://www.peeringdb.com/.

[40] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye. IP Geolocation Databases: Unreliable? *ACM CCR*, 2011.

[41] Renesys. http://renesys.com/.

[42] Routing Information Service (RIS)—RIPE Network Coordination Center. http://www.ris.ripe.net/.

[43] M. Roughan. Robust Network Planning. In *Guide to Reliable Internet Services and Applications*. Springer, 2009.

[44] M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush. 10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems. *IEEE J. on Selected Areas in Communications*, 2012.

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

[46] M. Schapira, Y. Zhu, and J. Rexford. Putting BGP on the Right Path: Better Performance via Next-Hop Routing. In *Proc. of SIGCOMM HotNets*, 2010.

[47] InMon—sFlow. http://sflow.org/.

[48] R. Stanojevic, I. Castro, and S. Gorinsky. CIPT: Using Tuangou to Reduce IP Transit Costs. In *Proc. CoNEXT*, 2011.

[49] V. Valancius, C. Lumezanu, N. Feamster, R. Johari, and V. Vazirani. How Many Tiers? Pricing in the Internet Transit Market. In *Proc. of ACM SIGCOMM*, 2011.

[50] B. Woodcock and V. Adhikari. Survey of Characteristics of Internet Carrier Interconnection Agreements, 2011. http://www.pch.net/docs/papers/peering-survey/.

[51] K. Xu, Z. Duan, Z.-L. Zhang, and J. Chandrashekar. On Properties of Internet Exchange Points and their Impact on AS Topology and Relationship. In *Networking*, 2004.

[52] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu. Spatio-Temporal Compressive Sensing and Internet Traffic Matrices. In *Proc. of ACM SIGCOMM*, 2009.

[53] Y. Zhang, Z. Zhang, Z. Mao, C. Hu, and B. Maggs. On the Impact of Route Monitor Selection. In *Proc. of IMC*, 2007.