Brazil's IXP requirements, limitations and security issues: a survey

Humberto Silva Galiza de Freitas[1], Christian Esteve Rothenberg[2], Jeronimo Aguiar Bezerra[3]

[1]Rede Nacional de Ensino e Pesquisa (RNP)
Campinas – SP – Brasil

[2]Universidade Estadual de Campinas (UNICAMP)
Campinas – SP – Brasil

[3]Florida International University (FIU)
Miami – FL – USA

humberto.galiza@rnp.br, christian@unicamp.br, jbezerra@fiu.edu

Abstract. *lorem ipsum lorem ipsum.*

1. Introduction

Internet eXchange Points (IXP) are the natural successors of the four Network Access Points (NAP) responsible for interconnecting different networks in the early stages of today's Internet. The benefits achieved regarding performance and security are the main reason for the key role that the IXPs plays in the present Internet ecosystem [Norton 2014].

Although the dynamics of operation, criticality, and security aspects of an IXP are well-known and quite often under discussion by the network operators community, only recently the scientific community began to delve into this subject. In this sense, the work of [Ager et al. 2012] pioneered with the flow analysis from a large European IXP. The authors concluded, among other things, that the traffic volume exchanged between the IXP members was analogous to levels observed in a Tier-1 Internet Service Provider (ISP) network.

In the same context in Brazil, the work of [Brito et al. 2015] innovated by bringing for the first time an analysis of the IXP ecosystem in operation in the country. The authors compiled valuable information about the characterization of the types of members present in these environments and their connectivity graphs at the Autonomous System (AS) level, being able to determine the peering potential in each IXP.

From an operational perspective, despite the fact that the Federal Interconnection eXchange Point (FIX) is considered the first public IXP in Brazil [FIX 2016], the

PTTMetro project launched in 2004 and coordinated by the Brazilian Regional Internet Registry Authority (NIC.BR) was, in fact, the responsible for the spreading of regional commercial IXPs in Brazil. Since then, the project has expanded the number of IXPs and the Points of Presence (PoPs) under its management across the country, and thus contributed to the development and improvement of national quality Internet. More recently, the project was renamed to IX.BR, and is currently operating in 24 cities in Brazil, exchanging near 2 Tbps of aggregated traffic [NIC.BR 2016], being considered the 4th biggest in the world based on the number of members [Putta Venkata and Ruan 2016].

Notwithstanding the clear benefits and advances brought about by large-scale deployment of IXPs in Brazil, the current model poses several problems, in the vast majority derived from its own architecture. These issues might end up affecting at different levels its security, scalability, and resiliency. The occurrence of network events or security incidents can compromise all or part of its operation, creating damages to its image and mission. In these cases, insecurity and instability set up in the routing environment contribute to the increase in distrust among the participants, as well as distancing possible new members.

The main contribution of the present work is to bring a Survey of the technical requirements and weaknesses of public IXPs in Brazil, presenting a non-exhaustive list of top security and limitations issues that affect these requirements. Despite the focus on Brazil, this study may contribute to a better understanding about the limitations and improvements that can be adopted in other IXPs using similar models. For each of the identified problems, possible solutions that follows the IXP model assumptions and operations guidelines and also that keeps IXP's stability, availability and scalability are provided.

Moreover, as some of the presented problems can not be addressed due to limitations in the current Ethernet/IP architecture, this paper also contributes in assessing the feasibility of adoption of some concepts from Software-Defined Networking (SDN) and OpenFlow protocol [ONF 2009] to address these deficiencies. The employment of such concepts can be the first step in the transformation of the current IXP model into a lightweight version of Software-Defined Internet eXchange (SDX) concept [Gupta et al. 2015].

The rest of this paper is structured as follows. Section 2 provides an overview about IXP's concepts, architecture and operations activities. Section 3 dissects the current architectural problems and limitations using a bottom-up approach. Following, Section 4, discusses some ongoing research efforts that address the issues raised, including the application of SDN and OpenFlow concepts. We conclude our work and discuss future directions in Section 5.

2. Background

Prior to its worldwide deployment and coverage, the Internet routing was centralized at four NAPs located in the U.S. These NAPs were public exchange facilities where commercial ISPs got interconnected to exchange traffic. As part of the NSFNET transition strategy to the modern Internet, these centralized access points were replaced by distributed IXPs. Currently, they are present in a number of countries, playing a key role in the Internet ecosystem by carrying out a large amount of traffic [Chatzis et al. 2013].

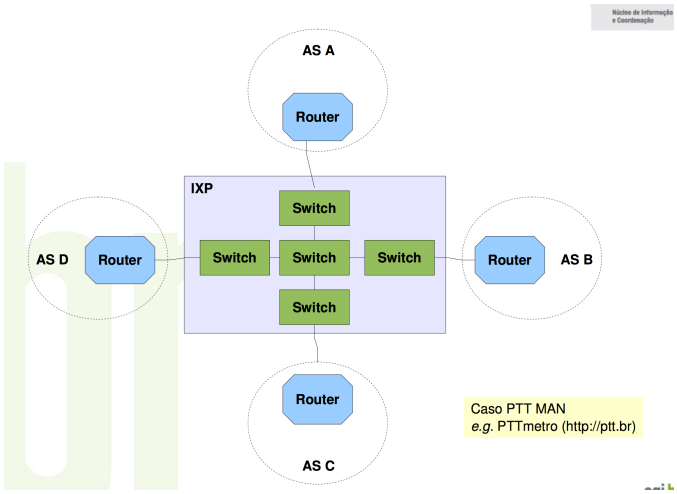

Figure 1. IXP traditional architecture with multiple PoPs. CHANGE PICTURE!!

From a design point of view, while the traditional IXPs can be classified into Layer-2 and Layer-3 IXPs, the prevalent model nowadays is the Layer-2 shared medium, like Ethernet, on top of which the IXP services are offered. This choice is due to the simplistic design requirements when providing a Layer-2 topology, as well as the easy and reduced costs for growing the switching fabric when compared to a Layer-3 solution.

As already mentioned, the simplest IXP design includes a Layer-2 pure switch, where the IXP connectors exchange traffic. However, as depicted in Figure 1, to scale this design and assure resiliency in the architecture, more switches may be added to the Layer-2 topology creating a switching fabric. In the same sense, the IXP can expand its own operations to cover a wide area (e.g., a metropolitan region). In this case, besides having multiple switches, the IXP can have multiple PoPs under the same Layer-2 shared medium, where all PoPs are interconnected, creating a full-mesh, or simply a distributed switching fabric.

In general, most IXPs in operation nowadays adopt a technical model where the switching fabric is essentially an enormous flat Ethernet domain, based on different technologies such as IEEE Std 802.1q Medium Access Control (MAC) Bridges, Virtual Bridge Local Area Networks (VLANs) and Virtual Private Lan Service (VPLS) [Kompella and Rekhter 2007, Lasserre and Kompella 2007]. On top of such (distributed) switching fabric, the members are free to set up peering agreements (BGP sessions) with other as they wish to exchange routes.

This method of peering at IXPs were very common in the early stages of IXPs deployment. However, it became a tough operational problem managing a large number of bilateral BGP sessions when the number of IXP members and peering agreements grew. In response to this challenge, the route-servers emerged as a method of interconnecting BGP speaking routers using a third-party system [Jasinska et al. 2016].

In this context, each IXP member set up one or more BGP sessions with the route-server and it, in turn, forwards the route information to each route-server client connected. This method of peering is also known as Multilateral peering, and Figure 2 illustrates the dynamic of its operation. In fact, the route-server operation is very similar to the route-

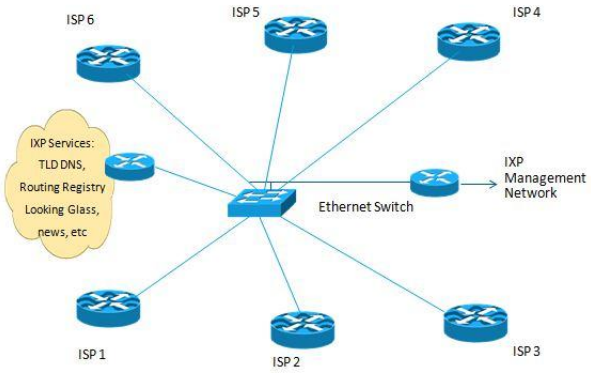

Figure 2. IXP route-server in operation. CHANGE PICTURE!!

reflector as specified in [Bates et al. 2006], except that it operates over an eBGP session instead of internal BGP (iBGP).

Besides the bilateral and multilateral peering arrangements, additional services may be offered, or even, sold using the peering fabric. A common ”package” often seen in different IXPs around the globe, includes, but do not limit to, DNS TLD service, NTP service, Internet Routing Registry (IRR) service, and Looking Glass. Usually these services are offered and maintened by the IXP operator.

In addition to this set of services, IXP members are allowed to buy/sell IP Transit from/to its partners. This can be achieved very straightforward by setting up a bilateral transit BGP session over the shared VLAN, or by asking the IXP operator to set up a dedicated VLAN between the two members, and then, setting up the transit BGP session. While this practice contrast with the IXP itself definition, in fact, it is a very common practice nowadays due to the easy of deployment, and the possibility of making business with a wide variety of members, in contrast with purchasing or selling IP transit from/to a dedicated partner.

Brazil's IX.BR project adopted the european IX model [EURO-IX 2012], i.e, the peering fabric is operated by a not-for-profit independent organization, usually NIC.BR and the Brazilian Research and Education Network (RNP), and the fabric may be spread across potentially many competing colocation facilities. Thus, the IX.BR members can choose the facility that best meets their needs. For instance, São Paulo's city IXP is the biggest one in operation, with around 500 connectors, exchanging near 1.5Tbps of traffic daily. Figure 3 illustrates São Paulo's city IXP PoPs interconnection in a live traffic map.

3. Issues and Limitations at IX.BR: A bottom-up Analysis

Despite of the success of the IXP model, and particularly in Brazil, the country-wide IXP deployment benefits, there are problems and limitations imposed by the architecture and technologies is use. In this section, we study the main issues and bottlenecks that affects this model, and in particular, the brazilian deployed model. For this analysis, we divided the issues and limitations into three main categories: control plane, data plane and management planes. Additionally, we have identified cross-layer issues, that will be studied in Subsection 3.4.

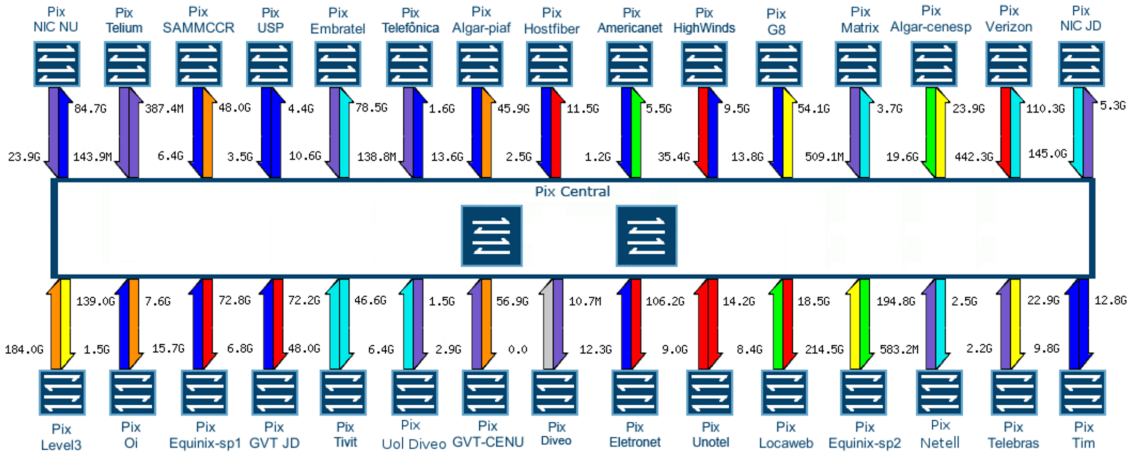

Figure 3. São Paulo's IXP links between colocation facilities (PoPs).

3.1. Control Plane Issues and Limitations

Issues affecting the network control plane can cause tremendous damages to the network operation. While the current discussion today is to decouple or not it from, especially with the novel SDN approach, the legacy networks still have to deal with classical problems affecting its control plane functionality, and will continue having to handle for a couple of years during a possible transition to SDN.

3.1.1. Broadcast Storms in the Switching fabric

Flat Layer-2 networks have long been known to have scaling problems. As the size of a broadcast domain increases, the level of broadcast traffic from protocols like Address Resolution Protocol (ARP) increases. Significant amounts of broadcast traffic pose a particular burden on the network because every device in such domain must process and possibly act on such traffic.

Sources of broadcast storms in the IXP switching fabric include, but does not limit to: (1) poor implementations of loop detection and prevention protocols; (2) IXP members misconfiguration errors. In extreme cases, these storms can occur where the quantity of broadcast traffic reaches a level that actually brings down part or all of a network [Narten et al. 2013].

Although the common sense when designing datacenter networks is to split large broadcast domains into multiple network segments, such solution is not applicable to an IXP fabric, because offering a single shared environment to the members is one of the IXP fundamental premises. Besides that, the IXP operator should be able to distinguish ARP between legitimate ARP requests and genuine broadcast storms. Either a restrict ARP blocking policy or excessive high ARP timeouts may result in the fabric aging out the MAC address of the receiving party from its MAC and CAM tables, and consequently may cause traffic blackholing.

3.1.2. Prohibited and Unknown Ethernet Frames in the Switching Fabric

Usually, under normal operations, the only Layer-2 traffic allowed in most current IXPs are: (1) ARP (ethertype: 0x0806), (2) IPv4 (ethertype: 0x0800) and (3) IPv6 (ethertype: 0x86dd). Despite this, it is not unusual to see in the switching fabric frames from protocols varying since Bridge Protocol Data Units (BPDU) and its variants, topology discovery protocols such as IEEE Std 802.1AB - Link Layer Discovery Protocol (LLDP) and the proprietary Cisco Discovery Protocol (CDP).

The occurrence of such ethertypes in the Exchange platform as well as frames generated by upper layers protocols such as Dynamic Host Control Protocol (DHCP) [Droms 1997] and IPv6 Neighbour Discovery / Router Advertisements [Narten et al. 2007], is strictly linked with device member misconfiguration. As long as the number of IXP members and the IXP Points of Presence (PoP) grows, more hard is to track down and address these problems.

Each of these already mentioned ethertypes could potentially impact the normal IXP operation leading to a total or partial interruption of the service. Even though there are simple mechanisms to easily address each of these issues, the process require some networking monitoring tools as well as network operator attention to network logs and flows. Based on that information gathered, the operator can trigger an action to address the issue.

3.1.3. Building and Maintaining a Loop-Free Topology

Designing and maintaining a loop free topology to an IXP is not something tough nowadays, considering the vast variety of resiliency protocols options in the market. Nevertheless, problems to this approach could arise in a very simple manner. As as a member can extend the IXP shared medium to his own backbone, he can cause (intentionally or not) a switching loop by creating more than one Layer-2 path between the IXP edge switch and an internal network switch, or simply by having two ports on the same switch connected to each other. The loop creates a broadcast storm and its effects already were described in section 3.1.1.

In response to this issue, IXP network operators have been using mainly two features: (1) storm broadcast control, to limit packet per seconds (pps) volume incoming from customer edge port; (2) setting a MAC learning limit to 1 in the member assigned interface.

However, even though both solutions are straightforward and easy to be deployed, network operators have to manually configure the settings, and the process requires also some networking monitoring tools and network operator needs to be heads up to strange network behavior. Additionally, it is not clear in the literature or in the best current practices how much is the limit of packets to avoid a storm broadcast but maintaining the IXP under normal operations. In the context, the IXP operator should be aware about the historic data from his network to understand the ”normal” broadcast traffic on a day-to-day operation.

Furthermore, limiting the number of MAC learning doesn’t guarantee or enforce

that the learnt MAC is the correct address from the IXP member, and also generates a limitation in case of remote peering from more than one member in one unique port. (EXPANDIR UM POUCO MAIS ESSE PARAGRAFO. FALAR DOS ATAQUES DE MANIPULACAO DA TABELA MAC, E TAMBEM DO CASO DOS PEERING REMOTO - CIX).

3.1.4. Attacks Targeted to Control Plane

As previously mentioned, in early IXP deployments, each member had to set up a BGP session with each other, leading to full mesh routing connectivity scheme. In this context, Route Servers emerged as a response to this scalability challenge, because it reduces all overhead turning the IXP management task easier [Lu and Zhao 2005].

Citar aqui RFC 7948 e 7947 - descrever mais problemas relacionados à operação dos route-servers

As Route Servers plays a crucial role in any current IXP today, attacks directed to them can cause serious operational problems to the infrastructure. Well-known attacks targeted to Route Servers include, but is not limited to, (1) all Layer-2 shared medium as well as Layer-3 attacks, for instance, mac flooding, cam overflow, arp spoofing, IP spoofing, and so forth. Besides that, targeted (D)DDoS to the Route Server IP and man-in-the-middle attacks are common forms of attacks that can take down an IXP.

Some counter-measures already exists in today's implementations to address specifically each of these points. Nevertheless, securing and maintaing the Route Server operation is one more task that requires both the network operator attention, as well as his careful in configuring network devices that connects to the Route Server.

3.1.5. Limited Traffic Engineering Support

Due to the nature of IXP model currently in use, support for Traffic Engineering both in inbound and outbound directions are performed through BGP attributes manipulation. Nevertheless, these techniques are restricted to a common place, the destination prefix. This is a current architectural limitation because jeopardize applying a most granular policy, for instance, source based routing, as well as doesn't provide an elegant solution to resiliency. For example, if an AS wants to have more that one port connected to distinct Points of Prosence (PoPs) of such an IXP, BGP doesn't provide TE capabilities to have a good traffic load balacing.

Outbound load balancing is achieved pretty straight forward: based on the destination prefix the IXP participant can install multiple routes using different paths. However, most of an ISP traffic, for example, is inbound, since they have the eyeballs for the content. Based on that, inbound Traffic Engineering for Multihomed ASes is achieved today basically using AS Path prepending or prefix deagregation. Both mecanisms are easy to implement, but may generate many side effects, such as BGP table polution, as well as undesired effects such as traffic black holing.

In terms of scalability, the current IXP model limit the tenants to grow because the

fabric doesn't have a mechanism to support multiple speed ports expansion. For instance, if a member wants to have a 10G port connected in a IXP PoP, and a 100G connected in another IXP PoP, they don't have a proper mechanism, using only BGP or even MPLS-TE, to achieve a good TE for the incoming traffic towards their backbone.

Even in the case that the participant wants to just bring another similar port (e.g the participant has a 10Gbps port and wants another 10Gbps port), it's usual that the IXP offer LACP as a mechanism to improve their throughput to the peering fabric. This approach carries several limitations. First, in the case that the participant is not locally connected (e.g it's a remote peering case), and it's using a leasing line or MPLS circuit to reach the peering fabric, LACP will not work, due to protocols particularities with timers and so forth. Second, link-aggregation configuration will require a proper maintenance window for it's configuration, because there is a risk of Layer-2 loops in case of misconfiguration in one side. In this case, to achieve a LACP scenario the member will need to drain their traffic, and possibly saturate other links in their network while the link-aggregation configuration is performed and validate. So, in summary, while this a feasible option currently, doesn't provide a non-stop upgrade mechanism, contributing to reduce member SLAs and possible IXP SLA.

3.1.6. Network Programmability Support Limitations

Current IXP model have no or even very limited support to Networking Programmability. This feature could bring interesting benefits to the tenants since could allow them to have more Traffic Engineering capabilities, with more options than just the destination prefix as it is today as mentioned in Subsection 3.1.5. DESCREVER MAIS SOBRE PROGRAMABILIDADE / APLICACOES / ETC

3.2. Data Plane Issues and Limitations

In this subsection, we will describe some issues and limitations affecting the way the packets are forwarded in the switching fabric. Similarly to the Control Plane, the Data Plane problems can lead to interruptions in the IXP operation, and even worse, can propagate the incident to other routing points in the Internet.

3.2.1. Unauthorized Hot-Potato Routing

Due to the current IXP's shared medium architecture, it's trivial to a member point out a static route to someone's routers to reach a 3rd party network. This is a form of Hot-Potato routing situation with no authorization. Despite this kind of conduct is denied by IXP operators, is hard to monitor the occurrence of such situation, and each IXP member usually have to take their own counter measures to avoid being maliciously used by another ASN in the fabric.

ESCREVER MAIS UM PARAGRAFO

3.2.2. Routing Leaks

To the best of our knowledge, the current IXP model doesn't have an efficient mechanism to prevent non-intentional routing leaking. There are network operators reports around the world that shows these routing leaks within the peering fabric caused a lot of problems, not only to the IXP environment, but towards the entire Internet. For instance, the case of... CITAR ESTUDOS DA NLNET

What has been done until today is just set a maximum prefix-limit per participant and in case the participant reach a set number of prefixes advertised the BGP session will be placed in a shutdown state. Although this action could prevent the spread of route leak, it is not focused on the main problem: a participant should advertise only what they are allowed to. Such validation mechanism doesn't exist in the current IXP model, despite there are reports that some IXP's have been running RPKI to validate the origin of the advertised prefixes.

3.2.3. Prefix Hijacking

Besides the routing leaking section mentioned in subsection 3.2.2, another routing problem is a huge concern in todays IXP: prefix hijacking. This is more difficult to be identified as long as there are no controls about what each participant shall advertise in the BGP sessions with the Route-Servers.

As a response to this challenge, all participants as well as the Route-Server operator should be able to verify all prefixes being adverstised and accept only the ones that pass in the validation process. Nevertheless, such mechanism doesn't exist today as a standard, and potential problems could happen in a very easy way.

ESCREVER MAIS

3.2.4. BGP *NEXT_HOP* Hijacking

Another concern in terms of hijacking is when a participant advertise a prefix with a wrong NEXT_HOP information, intentionally or not. As there is no verification to the NEXT_HOP information in the Route Server nor at the BGP implemenation in each participant, this attack is very easy to be deployed in an IXP fabric. To the best of our knowledge, there's no singular solution to this problem today due to the characteristics of the architecture deployed.

3.2.5. Member Sourced Attacks

As already said in other problems in this section, current IXP model doesn't validate what's is being advertised to the Route Servers of the Exchange. Another situation that raise from this characteristic is that (Distributed) Denial of Services attack can could easily be generated by participants in the fabric.

The simplest and straight solution to this is to apply anti-spoofing filters and val-

idate using an external mechanism such as RPKI to validate what's being advertised. Although applying filters is a easy task, manually maintaining the filters is a hard task to accomplish. Monitoring the current flows across the fabric is also a good starting point to identify such attacks, however is a very time consuming task to the network operators.

ESCREVER MAIS

3.2.6. Member Destined Attacks

The same limitations occur when the (D)DoS is originated outside of the fabric, and the target is a specific member of the IXP. This could destabilize all the switching fabric, as well as spread over the Internet due the degree of connectivity the IXP represents today.

ESCREVER MAIS

3.3. Management Plane

REVISAR ESSA SEÇÃO

There are serveral operational issues affecting IXPs today. In Brazil, despite the success of the deployed model, network operators have been claiming to the SLA offered by the IXP operator (NIC.BR). Most of time is spent with manual tasks, such as new participant validation process, filling forms and the interaction process itself over support tickets.

In this sense, there are tools today to automate such tasks, but the limitation is that most of them don't have the proper mechanisms to interact with all switching fabric devices using an uniform language or abstraction model.

3.4. Cross-Layer issues

DESCREVER PROBLEMAS CROSS-LAYER

4. Ongoing Research Efforts and Challenges

4.1. Anatomy and Evolution

4.2. Resiliency and Scalability

4.3. Performance e Security

4.4. Migration to SDN and Hybrid Models

5. Conclusion

In this work we presented a Survey about the current issues and limitations affecting the IXP model deployed in Brazil. We pointed out the major of these problems....TERMINAR...

References

Ager, B., Chatzis, N., Feldmann, A., Sarrar, N., Uhlig, S., and Willinger, W. (2012). Anatomy of a large european ixp. In *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*, pages 163–174. ACM.

Bates, T., Chen, E., and Chandra, R. (2006). BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP). RFC 4456 (Draft Standard). Updated by RFC 7606.

Brito, S. H. B., Santos, M. A. S., dos Reis Fontes, R., Perez, D. A. L., and Rothenberg, C. E. (2015). Anatomia do ecossistema de pontos de troca de tráfego publicos na internet do brasil. *XXXIII Simpósio Brasileiro de Redes de Computadores (SBRC)*.

Chatzis, N., Smaragdakis, G., and Feldmann, A. (2013). On the importance of internet exchange points for today's internet ecosystem. *arXiv preprint arXiv:1307.5264*.

Droms, R. (1997). Dynamic Host Configuration Protocol. RFC 2131 (Draft Standard). Updated by RFCs 3396, 4361, 5494, 6842.

EURO-IX (2012). European internet exchange association 2012 report on european ixps. Report, European Internet Exchange Association (EURO-IX).

FIX (2016). Ponto federal de interconexão. *online website: http://www.fix.org.br*.

Gupta, A., Vanbever, L., Shahbaz, M., Donovan, S. P., Schlinker, B., Feamster, N., Rexford, J., Shenker, S., Clark, R., and Katz-Bassett, E. (2015). Sdx: A software defined internet exchange. *ACM SIGCOMM Computer Communication Review*, 44(4):551–562.

Jasinska, E., Hilliard, N., Raszuk, R., and Bakker, N. (2016). Internet Exchange BGP Route Server. RFC 7947 (Proposed Standard).

Kompella, K. and Rekhter, Y. (2007). Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling. RFC 4761 (Proposed Standard). Updated by RFC 5462.

Lasserre, M. and Kompella, V. (2007). Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling. RFC 4762 (Proposed Standard).

Lu, X. and Zhao, W. (2005). *Networking and Mobile Computing: 3rd International Conference, ICCNMC 2005, Zhangjiajie, China, August 2-4, 2005, Proceedings*, volume 3619. Springer.

Narten, T., Karir, M., and Foo, I. (2013). Address Resolution Problems in Large Data Center Networks. RFC 6820 (Informational).

Narten, T., Nordmark, E., Simpson, W., and Soliman, H. (2007). Neighbor Discovery for IP version 6 (IPv6). RFC 4861 (Draft Standard). Updated by RFCs 5942, 6980, 7048, 7527, 7559.

NIC.BR (2016). Projeto ix.br - http://www.ix.br. Disponível online.

Norton, W. B. (2014). *The 2014 Internet Peering Playbook: Connecting to the Core of the Internet*. DrPeering Press.

ONF, O. N. F. (2009). Openflow switch specification 1.0.0 (wire protocol 0x01). *Online website: https://www.openflow.org/documents/openflow-spec-v1.0.0.pdf*.

Putta Venkata, R. and Ruan, L. (2016). A region-centric analysis of the internet peering ecosystem.