A survey on problems in generalized Internet eXchange Points architectures

Humberto Silva Galiza de Freitas
Academic Network at Sao Paulo (ANSP)
Americas Lightpaths Express and Protect (AmLight-ExP)
Sao Paulo, Brazil
Email: galiza@amlight.net

***Abstract*—Internet eXchange Points are the heart of today's Internet. However, a number of issues inherited from this architectural model have been identified and fixed, but others simply can not be remediated, due to the architecture conception. Regardless the IXP is commercial or for academic traffic exchange, these issues can compromise all or part of its operation, bringing risks, due to the insecurity and instability created in the traffic routing environment. Recent developments in Software-Defined Networks (SDN), which promises network management simplification as well as the addition of more capabilities and flexibility, have brought an interesting use case for IXPs called Software-Defined Internet Exchange (SDX). Nonetheless, while the current model presents limitations due to it is own nature, the SDX novel introduces many problems, which needs to be addressed to guarantee IXP's correctness, scalability, resiliency, management and evolution. This work presents a comprehensive study about the challenges affecting the current Internet eXchange Point (IXP) model in use, identifying and evaluating the impact of its principal weakness in the scope of security, scalability, resiliency and operations.**

I. Introduction

Though the critical importance the Internet eXchange Points (IXP) represent to the Internet nowadays, the architecture in place and *modus operandi* poses several challenges to its evolution. A number of issues inherited from this architectural model have been discussed and fixed in the past few years, but others simply can not be remediated due to architectural limitations.

Regardless the IXP is commercial or for academic traffic exchange, these issues can compromise all or part of its operation, damaging its image and mission, due to the insecurity and instability created in the traffic routing environment. This situation might generate mistrust to the current members, as well as distancing possible new associates.

Software-Defined Networking is a new networking approach that promises, among other things, turn the networks more smart by decoupling the control plane from the data plane, as well as adding programmability support to the network. OpenFlow is the most famous protocol in this novel, and was created as a joint effort of many different networking organizations, to provide all the SDN abstractions.

Recent developments in Software-Defined Networks (SDN), which promises network management simplification as well as the addition of more capabilities and flexibility, have brought an interesting use case for IXPs called Software-Defined Internet Exchange (SDX). [1] [2].

However, while the current model presents limitations due to it is own nature, still the SDX novel introduces many problems, which needs to be addressed to guarantee IXP's correctness, scalability, resiliency, management and evolution.

In this work we present a comprehensive analysis of the building blocks of a traditional traffic exchange infrastructure and their associated problems and limitations, using a functionality approach. The option for a functionality approach is ground on the fact that it is imperative to understand and catalog these issues affecting the resiliency of the current model as a previous step to move forward to a new architectural model such as an SDX.

This paper is organized as follows: Section II presents the current IXP architecture and terminology, and following, Section III dissects their limitations and issues.

II. Current IXP architecture and terminology

In the literature, the Internet Exchange Points are considered the natural successors of the old Network Access Points (NAP), place where the Autonomous Systems used to meet to exchange traffic and keep themselves reachable, what have turned to the current Internet. In the past few years a lot of work have been devoted to have a good comprehension of such complex ecossystem [3].

The IXP importance within the Internet ecosystem is clearly presented in the work of [4]. The authors sustain that although the academia does not like to see too many 'hot' topics involving IXP, there is a strong relationship between issues affecting network cloud services and data center services, mobility and even the trending SDN paradigm, with all the systematic operated by IXPs.

From a technical perspective, an IXP is basically an Ethernet based traffic matrix, as known as a single broadcast domain with the aim to facilitate the traffic exchange between Autonomous Systems (or simply participants) [5].

A. Next Generation IXPs: Software-Defined Internet Exchange (SDX)

Introduzir SDN e SDX

To the best of this author knows, there are no consesus on Software-Defined Exchange Points definition in the literature,

but on [6] works a SDX classification is presented based in three main approachs:

1) SDX de camada 3
2) SDX de camada 2
3) SDX para interconexo entre ilhas SDN

Os requisitos fundamentais dentro da abordagem SDX de camada 3 so a necessidade do utilizao do protolo BGP [7] para o envio e recebimento dos prefixos IP dos participantes, tal como feito no modelo atual de PTT, alm da incluso de um controlador SDN ao *switching fabric* com a responsabilidade de instalar os fluxos entre os participantes. Nesse escopo, destaca-se a contribuio feita por [8], na qual so demonstrados os desafios de construo e implantao de um simples roteador distribudo baseado no protocolo OpenFlow.

III. Current IXP architectures: Problems and Limitations

In this section, we categorize and describe the problems and limitations affecting the current IXP architectures based purely on packet forwarding. These issues were fit into three different functional perspectives: Control, Data, and Management planes.

This approach is suitable because each perspective represent, in the end, one aspect of the IXP: the peering fabric, the connectors or participants, and finally, the IXP operator. Thus, this approach can provide a clear view about the architectural limitations, and point future directions in terms of platform evolution.

A. Control Plane issues and limitations

1) Broadcast Storming in the Switching Fabric: As described in Section II, most of the IXPs in operation nowadays utilize a model where the switching fabric is essentially an enormous flat Ethernet domain. Also, a number of big IXPs have migrated their fabrics from VLANs to the Multi-Protocol Label Switching (MPLS) and its Virtual Private Lan Service (VPLS) [9] [10], mainly due to the protocol's high scalability and resiliency, when compared to the restrictions imposed in the utilization of VLANs. Nonetheless, regardless the Layer-2 technology chosen, all IXPs members set up BGP sessions to exchange routes on top of this fabric.

The common sense when designing data-center networks is to split large broadcast domains into multiple network segments. However, 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.

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) in the fabric 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. 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 [11].

Although this interaction does not represent a huge amount of bandwidth, the efficiency of the network can be improved by reducing the allocated power for broadcast traffic [12]. Sources of broadcast storms in the switching fabric include, but does not limit to: (1) poor implementations of loop detection and prevention protocols; (2) IXP members misconfiguration or equipment errors; (3) large number of IXP participants.

Some IXP operators around the world have developed processes to validate the presence of not authorized traffic in the switching fabric, including .... (To Do - Survey com IXP operators?) To address (1)... To address (2)... To address (3)...

Besides that, the IXP operator should be able to distinguish ARP between legitimate ARP requests and genuine broadcast storms, because both 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.

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 DHCP and IPv6 Neighbour Discovery / Router Advertisements, 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) 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 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 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 III-A1.

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. Although 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.

4) Attacks Targeted to Control Plane: In early IXP deployments, each member had to set up a BGP session with each other, leading to full mesh routing connectivity scheme. Route Servers emerged as a response to this scalability challenge, because it reduces all overhead turning the IXP management task easier [13].

As Route Servers plays an important 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, all Layer-2 shared medium as well as Layer-3 attacks (e.g. mac flooding, arp spoofing, IP spoofing, and so forth), targeted (D)DDoS to the Route Server IP, and man-in-the-middle attacks.

Some counter-measures already exists in todays 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.

5) 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 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 loadbalancing is achieved pretty straight forward: based on the destination prefix the IXP participant can install multiple routes using different paths. However, most of a ISP traffic, for example is inbound, since they have the eyeballs for the content. Based on that, inbound Traffic Engineering for Multihomed ASes is achievied today basically using AS Path Prepending or prefix deagregation. Both mecanisms are easy to achieve but generate many lateral problems, such as BGP table polution and undesired effects.

In terms of scalability, current IXP limit the tenants to grow because the fabric doesn't have a mechanism to support multiple speed ports expansion. For instance, if an IXP participant wants to have a 10G port in an edge PoP of the IXP and a 100G in another edge 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), the IXP usually offer them using LACP as a mechanism to improve their troughput to the fabric. However, 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 IXP, LACP will not work, since protocols particularities with timers and so forth. So, it's not a feasible option.

6) Networking Programmability Support: Current IXP have a very limited support or even does not support 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.

B. Issues Affecting IXP Data Plane

1) MAC and IP Spoofing: To Do

2) Hot-potato routing: Current IXP model doesn't have a solution when one participant points out a route towards another partipant to reach a 3rd party network. This is a 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 the participation have to take their own counter measures to avoid being mailiciously used by another ASN in the fabric.

3) Routing Leaks: Current IXP model doesnt have an efficient mechanism to prevent non-intentional routing leaking. There are network operators reports around the world that shows these routing leakings within a switching fabric caused a lot of problems, not only to the IXP environment, but towards the entire Internet. What's 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's 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're reports of some IXP's have been running RPKI to validate the origin of the advertised prefix.

4) Prefix Hijacking: Besides the routing leaking section mentioned in subsection III-B3, 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.

5) BGP NEXT_HOP Hijacking: Another concern in terms of hijacking is when a participant advertise a prefix with a wrong NEXT_HOP information. 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 this author knows, there's no singular solution to this problem today.

6) Source Member Attacks and Targeted Member 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 caracheristic 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 antispoofing filters and validate using an external mechanism such as RPKI to validate what's being advertised. Although applying filters is a easy task, manually maintaning 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.

The same limitations occur when the (D)DoS is originated outside of the fabric, and the target is a specfic 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.

C. Issues Affecting IXP Management Plane

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.

IV. Conclusion

The conclusion goes here.

Acknowledgment

The authors would like to thank...

References

[1] A. Gupta, L. Vanbever, M. Shahbaz, S. P. Donovan, B. Schlinker, N. Feamster, J. Rexford, S. Shenker, R. Clark, and E. Katz-Bassett, "Sdx: A software defined internet exchange," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 551–562, 2015.

[2] A. Gupta, R. MacDavid, R. Birkner, M. Canini, N. Feamster, J. Rexford, and L. Vanbever, "An industrial-scale software defined internet exchange point," *USENIX NSDI, Santa Clara, CA*, p. 1, 2016.

[3] H. Haddadi and O. Bonaventure, "Recent advances in networking. chapter 1: Internet topology research redux," *ACM SIGCOMM eBook*, 2013.

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

[5] EURO-IX, "European internet exchange association 2012 report on european ixps," European Internet Exchange Association (EURO-IX), Report, 2012. [Online]. Available: https://www.euro-ix.net/documents/1117-Euro-IX-IXP-Report-2012-pdf

[6] J. Chung, J. Cox, J. Ibarra, J. Bezerra, H. Morgan, R. Clark, and H. Owen, "Atlanticwave-sdx: An international sdx to support science data applications."

[7] Y. Rekhter, T. Li, and S. Hares, "A Border Gateway Protocol 4 (BGP-4)," RFC 4271 (Draft Standard), Internet Engineering Task Force, Jan. 2006, updated by RFCs 6286, 6608, 6793, 7606, 7607, 7705. [Online]. Available: http://www.ietf.org/rfc/rfc4271.txt

[8] J. P. Stringer, Q. Fu, C. Lorier, R. Nelson, and C. E. Rothenberg, "Cardigan: Deploying a distributed routing fabric," in *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking*. ACM, 2013, pp. 169–170.

[9] K. Kompella and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling," RFC 4761 (Proposed Standard), Internet Engineering Task Force, Jan. 2007, updated by RFC 5462. [Online]. Available: http://www.ietf.org/rfc/rfc4761.txt

[10] M. Lasserre and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling," RFC 4762 (Proposed Standard), Internet Engineering Task Force, Jan. 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4762.txt

[11] T. Narten, M. Karir, and I. Foo, "Address Resolution Problems in Large Data Center Networks," RFC 6820 (Informational), Internet Engineering Task Force, Jan. 2013. [Online]. Available: http://www.ietf.org/rfc/rfc6820.txt

[12] V. Boteanu and H. Bagheri, "Minimizing arp traffic in the ams-ix switching platform using openflow," University of Amsterdam, Tech. Rep., August 2013.

[13] X. Lu and W. Zhao, *Networking and Mobile Computing: 3rd International Conference, ICCNMC 2005, Zhangjiajie, China, August 2-4, 2005, Proceedings*. Springer, 2005, vol. 3619.