Humberto Silva Galiza de Freitas

Research proposal: Adding capabilities to the Brazilian IXP model towards the application of SDN and OpenFlow protocol

Campinas

2016

Contents

List of Figures

List of Tables

1 Introduction

Introduction to the subject - last part to finish!!!

1.1 Research Topic

Evolution of the current Internet Exchange Point model towards a Software-Defined approach.

1.2 Problem Statement

Though the critical importance that IXPs represent to the Internet, as could be easily verified in the literature, its *modus operandi* carries several problems affecting at different levels its security, scalability, resiliency and management.

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, generating mistrust to the current members, as well as distancing possible new associates.

In this work, some of the principal issues affecting the current model of traffic exchange in Brazil are outlined.

1.3 Hypothesis/Research subtopics

In this work the following hypothesis will be verified:

1. Are the problems listed in this document a big concern to the growth of current IXPs in Brazil?
2. What are the benefits brought by using SDN/OpenFlow in such environment?
3. Is the proposed model more advantageous to the IX.BR project from a financial and operational view?

1.4 Justification

Internet eXchange Points are the heart of today's Internet. In Brazil, the roll out of PTTMetro project in 200x have contributed to develop Brazil's Internet as well as to improve

it's quality. On the other hand, many problems affect the current model, not only in Brazil but over the world, since most IXPs use pretty much the same Ethernet based model today.

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.

These are trend topics today and have a good relevance to the academia. Understanding how SDN/OpenFlow could bring benefits to a real world use case will bring new challenges, but the most important achievement of this research will be delivering a true scalable model, with support to more refined Traffic Engineering capabilities and improved security..

This work proposes to study the challenges affecting the current Internet eXchange Point (IXP) model used in Brazil, identifying and evaluating the impact of its principal weakness in the scope of security, scalability, resiliency and operations. As a response to these challenges, a new design, based on the novel Software-Defined Networking approach, will be proposed and is expected that this new model could be able to address the primary concerns affecting the current model, adding more capabilities to it, while keeping the original foundations that drives the IXP operation.

1.5 Objectives

1.5.1 General Objective

Propose and validate a SDN/OpenFlow model that respects the current IX.BR operation way.

1.5.2 Specific Objectives

1. Propose and validate a strategy to control broadcast storm in the switching fabric using SDN

2. Propose and validate the addition of OpenFlow programmability support to the current IXP model in Brazil

3. Propose and validate new TE use-cases based on the addition of OpenFlow support to the switching fabric

4. Propose and validate a strategy for prefix advertisment validation (to avoid prefix/next_hop hijacking)
5. Propose and validate a strategy to avoid routing leaks in the switching fabric
6. Propose and validate a strategy to avoid hot-potato routing in the switching fabric

2 Literature Review

2.1 IXP importance to the Internet in Brazil

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 yeards a lot of work have been devoted to have a good comprehension of such complex ecossystem (HADDADI; BONAVENTURE, 2013).

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) (EURO-IX, 2012).

The IXP important within the Internet ecossytem is clearly presented in the work of (CHATZIS *et al.*, 2013). According to the authors, although the academia doesn't like to see too many 'hot' topics involving IXP, there's a strong relationship between issues affecting network cloud services and datacenter services, mobility and even the trending SDN paradigm, with all the systematic operatated by IXPs.

No Brasil, o primeiro PTT foi criado em 1992(?), através de iniciativa da Rede Nacional de Ensino e Pesquisa (RNP) (?), e teve como objetivo interconectar a infraestrutura de Internet no Brasil. Em meados de 2003 o Núcleo e Coordenação do Ponto BR (NIC.BR) em parceria com a RNP, lançou o projeto PTTMetro[1] (NIC.BR, 2016), com o objetivo de melhorar a interconexão regional no Brasil. O projeto atualmente tem PTTs implantados em XX cidades, conectando um total de YYY ASs participantes, tendo um tráfego agregado de ZZZZGbps.

Ainda como evidência da sua importância, o recente trabalho de (BRITO *et al.*, 2015) trouxe uma contribuição inovadora para o assunto, oferecendo a primeira análise detalhada desse microcosmo dentro da Internet brasileira. Os autores caracterizaram um conjunto de informações relevantes compreendendo desde os tipos de membros conectados aos PTTs, até os respectivos grafos de conectividade em nível Sistema Autônomo.

[1] Mais recentemente o projeto teve seu nome alterado para IX.BR - http://www.ix.br

2.2 Problems Affecting the Current Model

2.2.1 Issues Affecting IXP Control Plane

2.2.1.1 Broadcast Storming in the Switching Fabric

In general, most of the IXPs in operation today adopt a 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 and Virtual Bridge Local Area Networks and Virtual Private Lan Service (VPLS). On top of such switching fabric, all members set up BGP sessions to exchange routes.

However, 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 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 data-center 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, 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.

2.2.1.2 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 DHCP and IPv6 Neighbour Discovery / Router Advertise-

ments, 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.

2.2.1.3 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 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 2.2.1.1.

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. 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.

2.2.1.4 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 (LU; ZHAO, 2005).

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.

2.2.1.5 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.

2.2.1.6 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.

2.2.2 Issues Affecting IXP Data Plane

2.2.2.1 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.

2.2.2.2 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.

2.2.2.3 Prefix Hijacking

Besides the routing leaking section mentioned in subsection 2.2.2.2, 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.

2.2.2.4 BGP NEXT_HOP Hijacking

Another concern in terms of hijacking is when a participant advertise a prefix with a wrong NEXT_HOP information. 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.

2.2.2.5 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.

2.2.3 Operational Issues Affecting IXP Today

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.

2.3 Software-Defined Networks

citar aqui artigos chave sobre SDN

2.4 Software-Defined Exchange Points

To the best of this author knows, there are no consesus on Software-Defined Exchange Points definition in the literature, but on (CHUNG *et al.*,) works a SDX classification is presented based in three main approachs:

1. SDX de camada 3
2. SDX de camada 2

3. SDX para interconexão entre ilhas SDN

Os requisitos fundamentais dentro da abordagem SDX de camada 3 são a necessidade do utilização do protolo BGP (REKHTER *et al.*, 2006) para o envio e recebimento dos prefixos IP dos participantes, tal como é feito no modelo atual de PTT, além da inclusão de um controlador SDN ao *switching fabric* com a responsabilidade de instalar os fluxos entre os participantes. Nesse escopo, destaca-se a contribuição feita por (STRINGER *et al.*, 2013), na qual são demonstrados os desafios de construção e implantação de um simples roteador distribuído baseado no protocolo OpenFlow.

3 Methodology

Descrever a metodologia: - Métodos e técnicas a serem usados - Caracterização do objeto de estudo - Definição do universo da pesquisa - Aspectos e procedimentos éticos no envolvimento com os sujeitos da pesquisa - Plano de amostragem (se houver) - Procedimentos previstos para coleta de dados e ou experimentos - Procedimentos previstos para análise de dados - Avaliar se os problemas que já tem solução não foram impactados pelas soluções propostas neste trabalho

4 Cronograma

Descrição do cronograma

4.1 Etapas do desenvolvimento

Deve-se descrever as atividades a serem desenvolvidas e os marcos indicativos (componentes, equipamentos, textos, resultados de pequisas, *software*, etc.) que permitiro perceber o progresso das atividades.

Etapa 1: Estudo bibliográfico

Esta primeira etapa destinada formao da equipe do projeto, atravs do estudo das especificaes relacionadas ao trabalho a ser desenvolvido. No decorrer das demais etapas do projeto, outros estudo bibliogrficos mais especficos e detalhados sero realizados.

Atividades a serem desenvolvidas:

1.1 Buscar por informaes em diversos lugares

1.2 Ler bastante

1.3 Ler ainda mais

Etapa 2: Desenvolvimento da parte inicial

Nesta etapa sero dados os primeiros passos em busca da concluso deste projeto de final de curso. O caminho poder ser difcil mas com certeza resultar em uma grande satisfao.

2.1 Buscar por ferramentas para o desenvolvimento

2.2 Concepo do prottipo

2.3 Testes e reviso do projeto inicial

Etapa 3: Desenvolvimento da parte final

Nesta etapa o trabalho j estar bem encaminhado e restar apenas aparar algumas arestas. Apesar do caminho ter sido difcil, os resultados obtidos nos deram uma grande satisfao.

3.1 Verificao dos resultados obtidos

3.2 Novos experimentos com base nas correes

3.3 Escrita sobre os novos experimentos

Etapa 4: Escrita do documento e defesa do projeto

Uma vez que o trabalho j foi finalizado, cabe a esta etapa a escrita da monografia, a preparao da apresentao e por fim a defesa do trabalho.

4.1 Preparao do texto

4.2 Preparao da apresentao

4.3 Defesa do projeto

4.2 Cronograma do desenvolvimento

Indicar quando cada etapa descrita na seo 4.1 ser executada. preciso indicar as semanas onde os marcos indicativos estaro finalizados. Esse cronograma ser de grande importncia para determinar se o projeto est caminhando bem.

As etapas apresentadas no item 4.1 sero executadas da seguinte forma:

Etapa	Semanas											
	01	02	03	04	05	06	07	08	09	10	11	12
1												
2												
3												
4												

Table 1 – Cronograma das atividades previstas

4.3 Cronograma financeiro

Indicar o valor e os momentos de desembolsos e a finalidade dos mesmos.

5 Proposta de Estrutura Capitular

1 Introdução

2 Capítulo 2

3 Capítulo 3

Bibliography

BRITO, S. H. B.; SANTOS, M. A. S.; FONTES, R. dos R.; PEREZ, D. A. L.; ROTHENBERG, C. E. Anatomia do ecossistema de pontos de troca de tráfego publicos na internet do brasil. *XXXIII Simpósio Brasileiro de Redes de Computadores (SBRC)*, 2015. Cited on page 5.

CHATZIS, N.; SMARAGDAKIS, G.; FELDMANN, A. On the importance of internet exchange points for today's internet ecosystem. *arXiv preprint arXiv:1307.5264*, 2013. Cited on page 5.

CHUNG, J.; COX, J.; IBARRA, J.; BEZERRA, J.; MORGAN, H.; CLARK, R.; OWEN, H. Atlanticwave-sdx: An international sdx to support science data applications. Cited on page 10.

EURO-IX. *European Internet Exchange Association 2012 Report on European IXPs*. [S.l.], 2012. Disponível em: <https://www.euro-ix.net/documents/1117-Euro-IX-IXP-Report-2012-pdf>. Cited on page 5.

HADDADI, H.; BONAVENTURE, O. Recent advances in networking. chapter 1: Internet topology research redux. *ACM SIGCOMM eBook*, 2013. Cited on page 5.

LU, X.; ZHAO, W. *Networking and Mobile Computing: 3rd International Conference, ICCNMC 2005, Zhangjiajie, China, August 2-4, 2005, Proceedings*. [S.l.]: Springer, 2005. v. 3619. Cited on page 7.

NARTEN, T.; KARIR, M.; FOO, I. *Address Resolution Problems in Large Data Center Networks*. IETF, 2013. RFC 6820 (Informational). (Request for Comments, 6820). Disponível em: <http://www.ietf.org/rfc/rfc6820.txt>. Cited on page 6.

NIC.BR. *Projeto PTTMetro*. 2016. Disponível online. Disponível em: <http://www.ptt.br>. Cited on page 5.

REKHTER, Y.; LI, T.; HARES, S. *A Border Gateway Protocol 4 (BGP-4)*. IETF, 2006. RFC 4271 (Draft Standard). (Request for Comments, 4271). Updated by RFCs 6286, 6608, 6793, 7606, 7607, 7705. Disponível em: <http://www.ietf.org/rfc/rfc4271.txt>. Cited on page 11.

STRINGER, J. P.; FU, Q.; LORIER, C.; NELSON, R.; ROTHENBERG, C. E. Cardigan: Deploying a distributed routing fabric. In: ACM. *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking*. [S.l.], 2013. p. 169–170. Cited on page 11.