Aplicando Redes Definidas por Software à gerência de um ponto de troca de tráfego (IX.br)

Luis Felipe Cunha Martins, Dorgival Guedes

[1] Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

{luisf,dorgival}@dcc.ufmg.br

Resumo. *Pontos de troca de tráfego Internet (IX) permitem que diferentes organizações se interliguem para trocar tráfego diretamente em uma certa localidade, visando a reduzir custos com provedores de mais alto nível. Sua operação é uma tarefa complexa e crítica, pelo volume de tráfego presente e pelos interesses comerciais envolvidos. Algumas propostas de utilização do paradigma de Redes Definidas por Software (SDN) no ambiente do IX já foram feitas, mas sem atentar para os desafios operacionais internos dessas instalações. Com base nessa observação, este trabalho tem por objetivo descrever como o princípio de SDN pode ser aplicado internamente a um IX e demonstrar as vantagens desse paradigma do ponto de vista do operador do IX. Tarefas, que em alguns casos podem envolver dezenas de comandos de configuração em diferentes dispositivos, podem ser simplificadas para algumas linhas de configuração de um controlador SDN. A solução proposta foi desenvolvida com base no IX Minas Gerais, do projeto PTTMetro (IX.br), e considerando características dos diversos pontos em operação no país.*

Abstract. *Internet traffic exchange points (IX) allow different organizations to interconnect and exchange traffic directly in a certain location, to reduce costs with upstream providers. Their operation is a complex and critical task, due to the volume of traffic present and the comercial interests in play. Some proposals for the use of Software Defined Networks (SDN) in this context have been made, but no attention has been given to the operational challenges in the IX themselves. Based on that observation, this work describes how the SDN paradigm can be applied in an IX and shows the advantages of that approach from the IX operator's point of view. Some tasks that may require tens of configuration commands in different devices to accomplish can be simplified to a few configuration lines in an SDN controller. The proposed solution was developed based on the operation of the Minas Gerais IX, part of the PTTMetro project (IX.br), considering the characteristics of the various points in operation in Brazil.*

1. Introdução

O grande crescimento e aceitação da Internet possibilitou inúmeros benefícios e inovações, incorporando diversas áreas do conhecimento. No entanto, não eram esperadas essa forte adesão e crescente utilização. Com isso, surgiram diversos problemas relativos à segurança, infraestrutura, topologia e comunicação. Para reduzir esses problemas, os pontos de troca de tráfego (IX) surgiram como parte da infraestrutura da Internet, representando um ponto neutro, central, interconectando redes de diversos Sistemas Autônomos (SA) que a constituem e permitindo uma melhor organização da rede,

redução de custos, maior confiabilidade e segurança para seus usuários. Dessa forma, ao invés de depender do roteamento de pacotes entre diversos provedores intermediários ou da contratação de enlaces dedicados para cada rede a qual um SA deseja se conectar, é possível conectar um SA a todos os outros participantes de um IX através de um único enlace, de capacidade adequada, para o IX.

Na evolução da Internet, os IX firmaram-se como ambientes de produção de suma importância, tendo um papel crítico no ecossistema da Internet. Os IX buscam incentivar a troca de tráfego entre os SA localmente nas cidades, objetivando otimizar o desempenho e a conectividade de seus participantes, mantendo a troca de tráfego o mais localizada possível, evitando que os pacotes percorram grandes distâncias da origem ao destino, diminuindo assim a latência da rede e melhorando a experiência dos usuários. Ao interconectar domínios administrativos diferentes, o IX cria rotas alternativas, anteriormente inexistentes [Akashi et al. 2006], provendo uma conexão redundante para os SA, o que interfere diretamente na conectividade da Internet como um todo [Ceron et al. 2009].

Alguns estudos já abordaram a possibilidade de que Redes Definidas por Software (SDN) poderiam simplificar a operação da rede no roteamento entre domínios. No entanto, dado a escala global da Internet, há certas dificuldades em se aplicar o paradigma de SDN ao roteamento entre domínios, sendo o custo um dos fatores mais limitantes.

Nesse contexto, os IX são ideais para se implementar SDN aplicado às redes de longa distância (WAN). A implantação de SDN, em um único IX, pode gerar benefícios para dezenas a centenas de SA conectados, necessitando apenas trocar a matriz de comutação do IX para afetarmos todos os participantes nele conectados, além de representar um ponto de neutralidade e inovação. Com uma implantação gradual e a capacidade de afetar diversos SA com apenas a troca do switch central, outro benefício direto é o custo significativamente menor para se utilizar a solução.

Em nosso estudo, propomos a construção de uma arquitetura de ponto de troca de tráfego baseado em redes definidas por software, abordando os desafios e apresentando diversos casos de usos passíveis de serem construídos em cima dessa arquitetura para demonstrar os potenciais ganhos com a adoção dessa abordagem. O foco do estudo realizado abrangeu as vantagens providas por essa estrutura para a administração do IX, como simplificação do gerenciamento, menor tempo de resposta, adoção de topologias mais robustas, entre outros, que podem diminuir a demanda na equipe de TI que administra o IX, além de prover um serviço de maior qualidade e mais ágil para os seus participantes. Essas tarefas requerem um grande tempo dos administradores que ficam impossibilitados de dar a vazão necessária às demais requisições ao IX. Ao simplificar e automatizar diversas operações, os operadores do IX conseguirão prover um serviço de maior qualidade aos participantes.

O restante do artigo está dividido da seguinte forma: na seção 2, discutimos os trabalhos relacionados, enquanto a seção 3 apresenta uma breve contextualização dos elementos principais de um IX; em seguida, a seção 4 ilustra como a solução proposta pode simplificar diversas operações do dia-a-dia desse ambiente; O protótipo, implementado nos moldes do IX Minas Gerais, é descrito na seção 5, e a seção 6 discute alguns dos resultados alcançados até o momento; por fim, observações finais são discutidas na seção 7.

2. Trabalhos Relacionados

Alguns estudos na literatura já abordaram se SDN poderia simplificar a operação da rede no roteamento entre domínios, trazendo diversos benefícios, como um controle mais direto sobre o encaminhamento, e solucionando alguns dos problemas relativos ao protocolo BGP. O estudo de [Kotronis et al. 2012] abordou maneiras de como melhorar o roteamento dentro de um único AS, possibilitando uma engenharia de tráfego mais eficiente e o controle remoto do caminho fim-a-fim, utilizando-se um controlador SDN. No entanto, todos esses trabalhos exigem modificar a infraestrutura existente, com a troca de equipamentos, por exemplo.

Outros estudos trataram da utilização de SDN no IX, porém sempre focando o problema no ponto de vista dos participantes, sem considerar o IX como uma entidade no processo. Em [Gupta et al. 2014] abordam o problema de expressão de rotas entre SA que desejam se conectar e como SDN pode ser usado para garantir a segurança e facilitar a expressão dessas rotas. Por sua vez, [Stringer et al. 2013] propõem um roteador distribuído que pode ser usado para implementar a troca de tráfego entre duas ou mais organizações autônomas. [Mambretti et al. 2014] apresentam considerações para a implementação de um IX focando nas questões de chaveamento multiprotocolar entre as organizações.

[Lin et al. 2013] propõem uma solução para a implantação gradual de redes SDN, coexistindo de forma transparente com a rede IP tradicional, lidando com desafios como manter a comunicação entre redes SDN e IP. De forma análoga, [Salsano et al. 2014] discutem a utilização de SDN em backbones IP e a coexistência do encaminhamento em redes SDN e IP tradicional. Os autores argumentam que a rede de um Provedor de Internet necessita de operações mais sofisticadas do que as providas por soluções SDN em camada 2, propondo uma arquitetura capaz de lidar com o roteamento IP e manter a inter-operabilidade entre switches OpenFlow e tradicionais. Assim como esses autores, buscamos estabelecer em nosso trabalho a implantação de SDN no ambiente do IX de forma transparente para os participantes, mantendo a inter-operabilidade entre as redes tradicionais e a rede SDN, de forma que os SA não precisem ajustar a sua infraestrutura às nossas mudanças.

Apesar de compartilharmos do mesmo ideal que os estudos citados — utilizar SDN para inovar e prover novas funcionalidades no IX — o projeto proposto aqui difere dos anteriores, uma vez que eles lidaram com os problemas de infraestrutura, como o roteamento entre múltiplos domínios SDN, ou com um foco restrito nos participantes do IX, apresentando os benefícios, desafios e novas funcionalidades, que podem ser utilizadas pelos SA conectados. No entanto, até onde temos conhecimento, nenhum estudo focou na visão da administração do IX em si, como uma entidade independente, visando levantar os desafios e benefícios da utilização de SDN nas tarefas internas do IX, como a centralização e automatização da configuração dos switches, o gerenciamento dos participantes, englobando desde as ativações de novos participantes até a filtragem de tráfego indesejado e verificação de abusos, a engenharia de tráfego e a adoção de topologias mais robustas do que as geralmente utilizadas em redes de camada 2.

3. Principais elementos de um Ponto de Troca de Tráfego

Usualmente, o IX oferece uma arquitetura simples aos SA conectados. Pode-se visualizar a estrutura interna como sendo uma topologia do tipo estrela, com cada SA se conec-

tando através de um enlace a um conjunto de switches de nível 2 que formam a matriz de comutação [Ceptro.br]. A figura 1 ilustra os diversos elementos essenciais ao funcionamento do IX, discutidos a seguir.

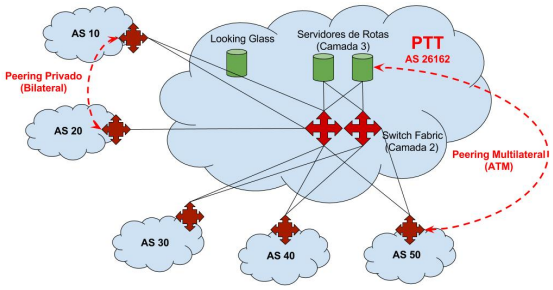

Figura 1. Estrutura interna de um IX, mostrando servidores de rotas, *looking glass* e o escopo de acordos de troca de tráfego (*peering*).

Um único IX é composto por vários PIXes, pontos de interconexão de redes comerciais e acadêmicas, interligados entre si, para formar o IX. Cada PIX representa um ou mais comutadores (switches) em uma posição geográfica, cuja função é prover o meio físico de acesso naquele local.

O anúncio das rotas de cada participante do IX é realizada pelo protocolo BGP através de pontos centrais denominados Servidores de Rotas (SR), de forma que um SA possa realizar a conexão BGP apenas com esses servidores, evitando topologias totalmente conectadas (*full-mesh*) e onerosos custos de interconexão dos roteadores de todos os participantes. A principal função de um SR é comunicar-se com todos os SA e propagar as rotas aprendidas de cada um deles, provendo alcançabilidade entre todos os SA, a partir do estabelecimento de uma única sessão BGP entre o SA e o SR.

Além da troca de tráfego, o IX oferece ainda um serviço conhecido como *looking glass* [LG - Wikipedia], que permite aos administradores dos SA visualizarem todas as rotas divulgadas dentro do IX, sem terem acesso físico à infraestrutura. Caso o participante deseje divulgar suas rotas, ele precisa apenas estabelecer uma sessão BGP com o servidor LG. A grosso modo, o LG nada mais é do que um servidor BGP que aprende rotas mas não repassa o que aprendeu, sendo muito utilizado pelos participantes para solucionar problemas de roteamento e consultar a tabela completa de roteamento do IX.

Uma vez fisicamente conectados no IX, seus membros podem acordar em realizar a troca de tráfego (*peering*) no acordo de troca multilateral (ATM), aberto, com todos os demais membros, ou através de acordos bilaterais (ATB), privado, de natureza seletiva ou restritiva. No primeiro caso, a administração do IX deve garantir a todos os participantes acesso às informações compartilhadas; no segundo, deve garantir uma forma de isolamento do tráfego dentro de sua rede para que o tráfego entre as duas partes flua sem ser acessível aos demais (por exemplo, alocando uma VLAN diferente para cada ATB).

Para entender a importância de IX regionais, pode-se imaginar uma região onde estão presentes alguns SA. Se não há um IX, esses SA estarão trocando o tráfego local, via Internet. Isso pode significar longas distâncias percorridas pelos pacotes. Com a criação de um IX regional, os SA podem conectar-se diretamente, levando à redução da latência, o aumento na resiliência da rede e, provavelmente, economia de custos.

4. Detalhamento da aplicação de SDN à operação de um IX

Com base na experiência diária do IX.br-MG, apresentamos diversos casos de uso construídos para explorar os benefícios do uso de SDN na operação do IX.

Centralizar e automatizar a configuração dos switches: Em um IX tradicional toda a configuração dos switches é feita manualmente. A cada novo participante que deseja conectar-se ao IX, os administradores precisam, entre muitas outras tarefas, habilitar a porta do switch onde ele se ligará, configurá-la para aceitar apenas pacotes oriundos do endereço MAC específico do roteador do participante e adicionar a porta a todas as VLANs necessárias para atender os acordos estabelecidos.

Com o uso de SDN, a adição do membro é enormemente simplificada: após a habilitação da porta, todas as demais configurações seriam adicionadas ao controlador que as aplicaria quando e onde fosse necessário.

Mapear o IX em uma base de dados centralizada: Hoje as informações dos participantes estão cadastradas em um arquivo, sendo boa parte delas utilizadas nas configurações dos equipamentos da rede, como o endereço IP, endereços MAC e se fazem parte do ATM ou não. Além de permitir que haja inconsistências entre as configurações e as informações cadastradas, o sistema atual carece de informações como VLANs criadas, número de VLANs disponíveis e outras.

Na nossa proposta com SDN, todos os dados necessários à configuração da rede estão armazenados em um arquivo, sendo a única fonte de dados para que o controlador configure o plano de dados, evitando inconsistências, além de facilitar a consulta e atualização dos dados. Ao integrar a documentação à configuração da rede, evita-se o armazenamento de dados duplicados e passíveis de erros.

Facilitar a ativação de participantes: Além da configuração do switch, o processo de ativação de um novo participante envolve diversos testes para avaliar a conformidade da conexão física, da conexão BGP e das rotas divulgadas pelo mesmo. Para isolar os testes de ativação, o novo participante é inicialmente colocado em uma área de quarentena na qual existem servidores SR e LG falsos. Apesar de não serem os servidores de produção, esses servidores possuem exatamente as mesmas configurações dos reais. Assim, o participante pode estabelecer as conexões BGP e divulgar as rotas exatamente como o fará quando estiver na produção, e o IX pode avaliar se isso é feito exatamente como esperado, ou se há algum problema. Como no teste de abusos, além de analisar o BGP, outros parâmetros da configuração do participante são avaliados, como o envio de tráfego indesejado e se a conexão do participante suporta aprender os MACs de todos outros participantes. Quando toda a configuração do novo participante é validada, basta migrar sua porta para a VLAN do ATM em que as conexões BGP reais se estabelecerão. Como o processo de ativação é lento, e para evitar que um participante interfira em outro também em ativação, o IX possui diversas áreas de quarentena isoladas entre si. Dessa forma, cada participante é alocado individualmente a uma das áreas durante o processo de ativação.

Por ser uma tarefa complexa, envolvendo analisar o tráfego originado do participante e simular as interações com outros, a ativação pode tirar grande proveito de uma SDN. Esta pode ajudar tanto na geração do tráfego fictício quando na captura dos pacotes enviados pelo participante. Além disso, o isolamento entre os SA durante a ativação se torna trivial, não sendo necessária as migrações do ambiente de quarentena para produção.

Facilitar a detecção e filtragem de tráfego indesejado: Além das configurações de adição de participante mencionadas, são aplicados diversos filtros de pacotes nas portas dos switches para evitar o envio de tráfego indesejado. No IX, o único tráfego oriundo dos participantes aceito são pacotes IP destinados aos demais participantes e as conexões BGP com os SR, além do tráfego ARP. A detecção de outros tipos de tráfego é feita pela análise de dados SFLOW [P. Phaal and Mckee 2001] enviados pelos switches. Em muitos IX, como o de IX.br-MG, o ambiente ainda não está pronto para o SFLOW, sendo a detecção de tráfego indesejado não realizada, deixando o IX mais vulnerável.

Com o uso de uma SDN, todas as regras OpenFlow criadas pela aplicação nos switches realizam o *match* apenas para os tráfegos permitidos em cada tipo de comunicação no IX, rejeitando os demais tipos de tráfego.

Evitar abusos na estrutura: Os participantes necessitam seguir um conjunto de regras para se conectarem à estrutura do IX. Entre elas, não é permitido utilizar a estrutura para a troca de tráfego de um SA consigo mesmo (em duas localidades diferentes, por exemplo), através de duas conexões independentes ao IX e também não é permitido que um SA encaminhe tráfego de redes que ele não anuncia ou direcione a rota *default* para o ambiente do IX.

Em IX menores não há mecanismos implementados para tais checagens, sendo o ambiente vulnerável a tais abusos. Em uma rede SDN, pelo fato de o controlador ter conhecimento de todas as rotas, ele pode rejeitar tráfego de origens cujas rotas não estão anunciadas no IX. Além disso, basta não criar as regras de OpenFlow para a comunicação entre as portas de switches que estejam conectadas a um mesmo participante, impedindo qualquer forma de comunicação entre elementos de um mesmo SA.

Evitar *broadcasts*: O único *broadcast* permitido na rede do IX é do tipo ARP, para a resolução dos endereços MAC na camada 2. Em relação ao ARP, a quantidade de pacotes que trafegam na rede do IX está diretamente relacionada ao número de participantes e de VLANs existentes. Apesar da sua necessidade para o funcionamento da rede, *broadcasts* consomem recursos desnecessários, já que todos os participantes e os respectivos endereços MAC são de conhecimento prévio da administração do IX.

Nesse caso, a SDN impede o trânsito de tráfego *broadcast* não ARP e elimina completamente a propagação dos *broadcasts* ARP. Como o controlador sabe quais participantes estão ativos na rede, ele pode interceptar consultas ARP e gerar as respostas para elas diretamente.

Isolamento do tráfego em acordos bilaterais: São muito comuns os pedidos de VLANs para isolar o tráfego de acordos bilaterais. Em IX com muitos participantes, a quantidade de VLANs começa a atingir os limites dos equipamentos e do próprio cabeçalho dos pacotes. Para eliminar a limitação da quantidade de VLANs, um dos recursos adotados pelos IX é substituí-lo pelo uso de MPLS no encapsulamento do tráfego dos acordos bilaterais, o que implica em custos operacionais e do uso de equipamento especializado.

Na nossa solução com SDN, o isolamento de tráfego é simples de ser alcançado já que o controlador, por ter conhecimento dos acordos bilaterais e o controle de encaminhamento da rede, pode facilmente rotear os pacotes entre os participantes, não sendo necessário o uso do MPLS ou outro recurso avançado.

Engenharia de tráfego entre PIXes: A inexistência de protocolos L2 que aproveitem toda a capacidade da rede em topologias complexas faz com que a maioria dos IX adote uma topologia em estrela, onerosa. Na existência de anéis na rede, utiliza-se do protocolo EAPS [Shah and Yip 2003] para deixar as conexões redundantes em *hot-standby*.

Por conhecer a topologia da rede, as características das conexões e a sua utilização, o controlador SDN permite a adoção topologias mais complexas e baratas, roteando os pacotes entre os PIXes da melhor forma possível e maximizando a utilização das conexões, inclusive utilizando técnicas de balanceamento através dos múltiplos caminhos do anel.

Realizar a tradução e contabilização de VLANs: Os participantes do IX costumam utilizar apenas a quantidade de portas de switch necessárias ao escoamento de todo o seu tráfego, mas sem a saturação das interfaces. Essa prática evita o desperdício de recursos, mas dificulta para a administração do IX contabilizar e tipificar o tráfego em cada porta de switch. O grande problema está na mistura de múltiplos tipos de tráfego em uma mesma porta. Como os participantes podem estar em inúmeros acordos de tráfego diferentes, em uma mesma porta passa tráfego de diversas VLANs, dificultando a contabilização do tráfego de cada VLAN individualmente, pois essa informação demanda recursos que poucas vezes estão disponíveis nos switches.

A solução comumente adotada nos IX, para contabilizar o tráfego de múltiplas VLANs, é separá-las fisicamente utilizando um switch de tradução, onde múltiplas VLANs entram por uma única porta e o tráfego de cada VLAN é encaminhado separadamente para outras portas físicas. Dessa forma, a contabilização pode ser realizada para cada VLAN através da coleta de estatísticas da porta para a qual a VLAN é encaminhada.

Diferentemente da contabilização por VLAN, que depende de implementação específica do fabricante, na rede SDN, este tipo de informação é facilmente obtida. Switches SDN nativos possuem contadores para regras arbitrárias e, para esse caso, bastaria criar regras independentes para cada tipo de fluxo, o que geraria a contabilização apropriada.

5. O protótipo do IX Minas Gerais

Para a validação dos casos de uso e funcionalidades propostas no ambiente SDN, desenvolvemos um sistema para a prototipação rápida e automatizada de topologias semelhantes às adotadas pelos IX no Brasil, mais especificamente o IX.br-MG. Para tanto, optamos por utilizar a ferramenta de criação de redes virtuais Mininet [1], que possibilita instanciar hosts, enlaces e switches.

No protótipo, cada SA foi mapeado em um roteador de borda, capaz de se comunicar via BGP com os SR e LG, sendo representado por um host no Mininet executando o software de roteamento Quagga [2] para divulgar as rotas daquele SA, e, pelo menos, um host para cada rede anunciada para testar a conectividade e o correto anúncio das rotas. Por exemplo, se o ASN1916 anuncia a rota para o bloco 200.131.0.0/24, cria-se um host com o IP 200.131.0.x/24 conectado ao roteador do ASN1916.

Para o protótipo, o ambiente do IX.br-MG foi replicado. Os switches que compõem os quatro PIXes e a rede física do IX são descritos através de switches vir-

[1] `http://mininet.org`
[2] `http://nongnu.org/quagga/`

tuais, conhecidos como Open vSwitch [3]. Cada SR é mapeado da mesma forma que um roteador de borda do SA, ou seja, um host no Mininet, executando o Quagga. O SR reside dentro da rede SDN e sua função é trocar informações, via eBGP, com os roteadores externos e repassá-las à aplicação que gerencia as rotas no controlador. Por sua vez, o LG é instanciado da mesma forma que os SR, alterando-se apenas a configuração do BGP no módulo do Quagga. Os detalhes do mapeamento dos elementos discutidos podem ser visualizados na figura 2. Observe os seis SA, com seus roteadores de borda e "redes internas" conectados ao switches dos cinco PIXes.

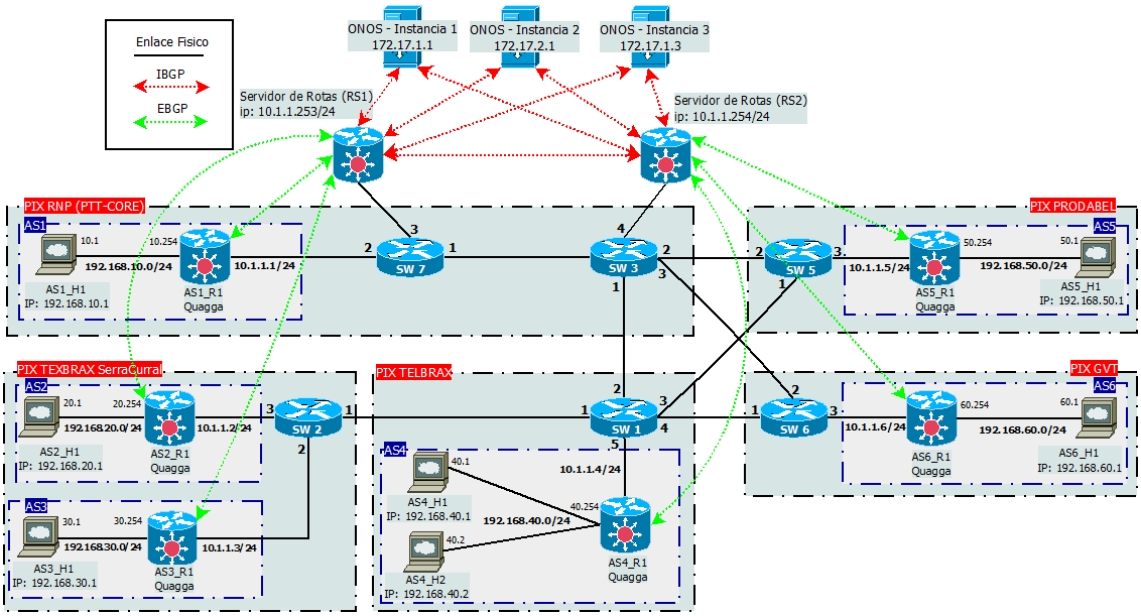

Figura 2. Protótipo do IX.br-MG

O ambiente foi construído pensando na resiliência, alto desempenho e disponibilidade em mente. Dessa forma, após um estudo sobre os controladores existentes, optou-se pelo uso do Open Source Network Operating System (ONOS), um controlador recente, de Dezembro/2014, com foco em operadoras de telecomunicação e redes de longa distância (WAN). Seu desenvolvimento foi norteado por requisitos de alto desempenho, disponibilidade, baixa latência e capacidade de manipulação de grandes redes [Berde et al. 2014].

Os SR atuam como refletores de rotas não se colocando no caminho do tráfego (AS-PATH), provendo a troca de rotas BGP entre os SA e também repassando as rotas ao controlador. Os controladores operam em forma de um cluster, portando-se como uma única entidade para o mundo exterior. Dessa forma, é possível obter tolerância de falhas, tanto no plano de controle, onde um SR pode assumir a demanda do outro em caso de falha, assim como os três controladores entre si, quanto no plano de dados, uma vez que o ONOS monitora os caminhos em utilização, recalculando-os em caso de falhas.

Um das grandes diferenças do ONOS para os demais controladores é a sua abstração chamada ***intents***. O framework de *Intents* é um subsistema que permite as aplicações a especificarem o comportamento da rede, através de diretivas baseadas em políticas - chamadas *intents*. Uma *intent* pode ser definida como **o que** se deseja reali-

[3] `http://openvswitch.org`

zar no lugar de **como** deve ser feito para se alcançar o propósito desejado. Em termos da rede, as *intents* fornecem uma abstração para os desenvolvedores de forma que a aplicação possa requisitar os recursos da rede, sem ter conhecimento de como eles serão executados e disponibilizados, permitindo que os operadores da rede possam 'programar' a rede em alto nível e não através de regras de baixo nível, abstraindo a complexidade da camada de rede. Com o uso de *intents*, acordos de troca de tráfego podem ser representados, usando essa abstração. Devido à sua visão global da rede, o framework de *intents* pode reagir aos eventos da rede. Quando um host deseja comunicar com outro host, cria-se uma *intent* host-to-host e o ONOS calcula o melhor caminho, instalando o fluxo. Na ocorrência de uma falha no caminho entre os dois hosts, o framework recalcula automaticamente um novo caminho e instala a regra correspondente.

Para aproveitar o potencial do framework de *Intents* e integrar o BGP com o ONOS, foi construída uma aplicação, SDX-APP, baseada em um caso de uso já existente no ONOS, para prover conectividade entre os membros, gerenciar a rede do IX e prover o *peering* transparente entre a rede SDN e as demais redes IP tradicionais. Dessa forma, a SDX-APP é responsável por configurar o plano de dados de acordo com as configurações do ambiente do IX, visando à comunicação entre os roteadores de borda dos SA e também deles para os SR e possibilitando a troca de rotas através do BGP.

A aplicação SDX-APP foi construída para integrar a estrutura do protótipo, interagindo com os SR, encaminhando as rotas aprendidas ao ONOS que, por sua vez, transforma-as em *intents*, responsáveis por programar os switches OpenFlow (OF). O objetivo final é prover funcionalidade L3 em uma rede de switches OF, transformando a rede SDN em uma rede de trânsito, capaz de trocar tráfego entre diferentes SA. Um ponto importante é que uma vez que a SDX-APP permite a comunicação transparente entre as redes tradicionais e o IX, através do BGP (e qualquer protocolo legado), os participantes do IX podem aproveitar dos benefícios do SDN em suas redes e realizar a migração de sua infraestrutura para SDN de forma gradual e de acordo com as suas próprias necessidades. A figura 3 descreve a arquitetura da aplicação em alto nível:

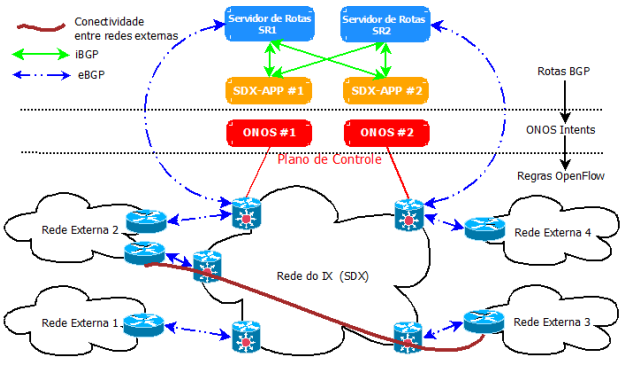

Figura 3. Arquitetura da aplicação SDX-APP

O plano de dados do IX, com as tabelas de fluxo dos switches do núcleo, é mostrado na figura 4. Nela podemos observar as regras para a troca de tráfego entre os participantes e as regras para o encaminhamento do tráfego BGP entre os SA e os SR, assim como a tradução de VLANs quando necessário. Como usamos o framework de *intents* no SDX-APP, é possível aproveitar de seus benefícios, de forma que, sempre que uma *intent*

é solicitada, o controlador checa quais regras já existem e podem ser agregadas com a *intent* solicitada, de forma a diminuir o número de regras no plano de dados.

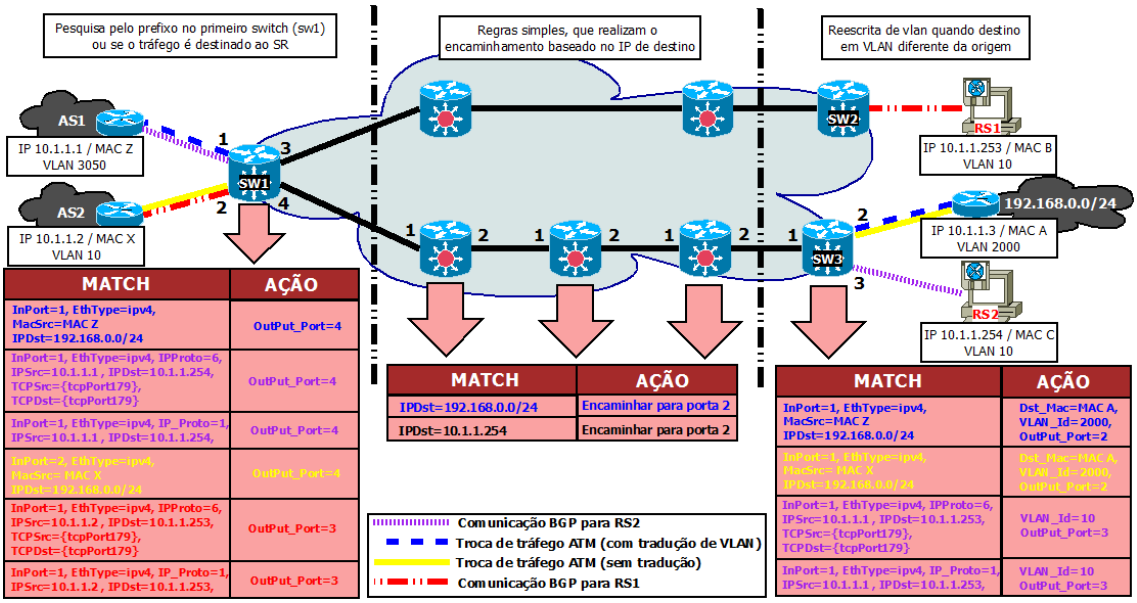

Figura 4. Regras de fluxos instaladas pela aplicação SDX-APP

6. Validação do protótipo

Para a validação do protótipo construído e a sua integração com o BGP e redes IP tradicionais, várias baterias de testes foram realizadas, com os critérios de alta disponibilidade e tolerância a falhas, premissas do ONOS, em mente. Para os testes de alta disponibilidade e tolerância a falhas foram criadas três instâncias do controlador ONOS, implementadas como containers individuais, através do uso da ferramenta de virtualização Docker [4].

6.1. 1ª Bateria - Conectividade Geral

A primeira bateria buscou verificar a conectividade geral entre as redes externas dos participantes, ou seja, se os anúncios de rotas estavam sendo repassados aos SR e se estes repassam as rotas aprendidas aos roteadores de borda de cada participantes, provendo conectividade e a troca de tráfego entre os SA do IX.

Na listagem 1, exibem-se as rotas aprendidas pelas instância do ONOS, mostrando que a SDX-APP recebeu as rotas corretamente, se comunicando via BGP com os dois SR.

```
# Rotas antes da falha do SR1
onos> bgp-routes
    Network            Next-Hop    Origin  LocalPref   MED   BGP-ID          AsPath
    192.168.10.0/24    10.1.1.1    IGP     100         0     10.10.10.31     65006
    192.168.20.0/24    10.1.1.2    IGP     100         0     10.10.10.31     65005
    192.168.30.0/24    10.1.1.3    IGP     100         0     10.10.10.31     65004
    192.168.40.0/24    10.1.1.4    IGP     100         0     10.10.10.31     65003
    192.168.50.0/24    10.1.1.5    IGP     100         0     10.10.10.31     65002
    192.168.60.0/24    10.1.1.6    IGP     100         0     10.10.10.31     65001
```

Listagem 1. SDX-APP BGP-ROUTES - ONOS #1

Quanto ao teste de conectividade, todos os hosts externos foram capazes de comunicarem entre si.

[4] https://www.docker.com

6.2. 2ª Bateria - Falha de um controlador

A SDX-APP, quando executada em um cluster, utiliza os algoritmos de consenso do ONOS para eleger um líder, responsável por gerenciar as rotas BGP e comunicar as informações com as outras instâncias. Dessa forma, no momento da falha, outra instancia deverá assumir o controle, como novo líder, sem prejuízos para o tráfego no IX. Indiretamente, as sessões iBGP entre a instância com problemas e os SR serão removidas.

O funcionamento esperado foi validado pelo teste, conforme mostrado na listagem 2. Toda a conectividade entre as redes externas foi mantida e outro controlador assumiu a liderança de todos os dispositivos, tornando-se líder da SDX-APP. Observe que as sessões estão *up* a vários minutos e que a queda do controlador não interferiu nas sessões iBGP estabelecidas com os demais controladores (172.17.1.1 e 172.17.1.3).

```
# Ping funcionando entre diversos SA diferentes
mininet> AS1h1 ping AS2h1 -c1
PING 192.168.20.1 (192.168.20.1) 56(84) bytes of data.
64 bytes from 192.168.2.1: icmp_seq=1 ttl=62 time=80ms

mininet> AS2h1 ping AS4h1 -c1
PING 192.168.40.1 (192.168.40.1) 56(84) bytes of data.
64 bytes from 192.168.40.1: icmp_seq=1 ttl=62 time=111ms

# Sumario BGP no Quagga do SR1
routerServer1> show ip bgp summary
Neighbor        V    AS      MsgRcvd MsgSent TblVer  InQ    QutQ    Up/Down     State/PFxRcd
10.1.1.1        4    65001   56      58      0       0      0       00:02:39    2
10.1.1.2        4    65001   56      58      0       0      0       00:02:38    2
10.1.1.3        4    65001   56      58      0       0      0       00:02:39    2
10.1.1.4        4    65001   56      58      0       0      0       00:02:38    2
10.1.1.5        4    65001   56      58      0       0      0       00:02:39    2
10.1.1.6        4    65001   56      58      0       0      0       00:02:38    2
172.17.1.1      4    65001   56      58      0       0      0       00:02:37    2
172.17.1.2      4    65001   56      58      0       0      0       00:01:07    Active
172.17.1.3      4    65001   56      58      0       0      0       00:02:37    2
```

Listagem 2. Rotas recebidas antes e após a falha do SR1

Após corrigir a falha na instância interrompida, verificou-se que todos os relacionamentos BGP foram restabelecidos com ela.

6.3. 3ª Bateria - Falha de um enlace

Na 3ª bateria de testes, objetivou-se verificar se os controladores são capazes de ajustar o encaminhamento após a falha de um enlace específico. O enlace escolhido foi o que conectava a 3ª porta do switch sw1 do PIX-Telbrax com 1ª porta do switch sw5 do PIX-Prodabel (00000000000000a1/3 $\Leftrightarrow$ 00000000000000a5/1), conforme mostrado na figura 2. O comportamento esperado é que a aplicação irá recalcular as rotas para ajustar as *intents* correspondentes, mantendo a conectividade e o encaminhamento do tráfego.

Ao remover o enlace descrito, é possível ver que o controlador averiguou a falha e recalculou os caminhos corretamente, removendo as regras antigas e instalando novas regras de fluxo, de acordo com o novo caminho calculado. Pode-se observar que o caminho antigo (sw1 $\Leftrightarrow$ sw5) foi alterado para o próximo caminho mais curto disponível, no caso sw1 $\Leftrightarrow$ sw3 $\Leftrightarrow$ sw5. O comportamento acima descrito pode ser visto na listagem 3. Conforme a saída do comando *paths*, verifica-se que o caminho foi recalculado pelo ONOS corretamente após a falha do enlace. O caminho entre sw1 e sw5 era composto por um enlace direto, de custo igual a 1 e, após a queda do enlace, o caminho foi modificado para um novo de custo igual a 2, sendo o caminho mais curto disponível.

```
# Caminho utilizado pelo ONOS para comunicar o sw1 com o sw5 antes da falha do enlace
onos> paths of:00000000000000a1 of:00000000000000a5
of:00000000000000a1/3-of:00000000000000a5/1; cost=1.0


# Caminho utilizado pelo ONOS para comunicar o sw1 com o sw5 apos a falha do enlace
onos> paths of:00000000000000a1 of:00000000000000a5
of:00000000000000a1/2-of:00000000000000a3/1⇒00000000000000a3/2-00000000000000a5/2;
cost=2.0
```

Listagem 3. 3ª Bateria de testes - Mudança no caminho apos a queda do enlace

A análise das regras de fluxos e um ping em execução contínua durante o teste mostrou que a conectividade foi mantida entre os hosts, não havendo perdas de pacotes tanto no caminho antigo quanto no novo, validando a mudança no plano de dados.

Após o enlace ter sido restaurado, o ONOS recalculou o melhor caminho e alterou as regras de fluxo para o caminho direto entre sw1 e sw5, de custo igual a 1, que é a rota mais curta, atualizando a tabela de encaminhamento dos switches envolvidos.

6.4. 4ª Bateria - Falha de um Servidor de Rotas

Na última bateria de testes programada, verificou-se o comportamento do sistema após a falha de um dos SR, ocasionando a queda das sessões BGP estabelecidas com ele.

Para esse teste específico, os roteadores dos AS1, AS2 e AS3 foram alterados para estabelecerem sessão BGP apenas com o SR1 e não com todos os SR como é realizado normalmente. Desse modo, novos prefixos divulgados por esses SA não serão repassados para os demais, causando o isolamento de novas redes divulgadas por eles. É esperado que o ONOS invalide as *intents* e entradas de fluxos nas tabelas dos switches referentes às rotas já existentes, uma vez que a SDX-APP não receberá as informações das rotas repassadas pelo SR1, isolando os SA em questão. Os prefixos anunciados apenas ao SR1 também serão retirados pelo BGP da tabela de rotas dos roteadores dos demais participantes.

Como esperado, as rotas correspondentes foram retiradas em todos os roteadores de borda dos participantes e invalidada na aplicação SDX-APP, causando o isolamento dos SA afetados. Nos roteadores de borda desses participantes as sessões do BGP foram finalizadas e apenas a rota para rede local, o núcleo do IX, e para a rede externa do próprio participante continuaram em suas tabelas de rotas.

O comando *bgp-routes* exibe os dados dos anúncios recebidos pela SDX-APP através da sessão BGP estabelecida com os SR. Na listagem 4, executado antes e após a falha do SR1, é possível ver que o controlador retirou as *intents* para as redes 192.168.10.0/24, 192.168.20.0/24 e 192.168.30.0/24, uma vez que elas não estão sendo repassadas pelo SR1 (10.10.10.31) à aplicação SDX-APP e ao ONOS. Dessa forma, as rotas para essas redes, na tabela de rotas dos roteadores de borda dos demais participantes, foram retiradas, interrompendo a conectividade aos SA afetados. Por fim, a aplicação SDX-APP atualizou as rotas das redes 192.168.40.0/24, 192.168.50.0/24 e 192.168.60.0/24 como recebidas pelo segundo servidor de rotas (10.10.10.32).

```
# Rotas antes da falha do SR1
onos> bgp-routes
    Network            Next-Hop    Origin  LocalPref   MED   BGP-ID          AsPath
    192.168.60.0/24    10.1.1.6    IGP     100         0     10.10.10.31     65006
    192.168.50.0/24    10.1.1.5    IGP     100         0     10.10.10.31     65005
    192.168.40.0/24    10.1.1.4    IGP     100         0     10.10.10.31     65004
    192.168.30.0/24    10.1.1.3    IGP     100         0     10.10.10.31     65003
```

```
    192.168.20.0/24     10.1.1.2    IGP     100         0     10.10.10.31     65002
    192.168.10.0/24     10.1.1.1    IGP     100         0     10.10.10.31     65001

# Rotas apos a falha do SR1
onos> bgp-routes
    Network             Next-Hop    Origin  LocalPref   MED   BGP-ID          AsPath
    192.168.60.0/24     10.1.1.6    IGP     100         0     10.10.10.32     65006
    192.168.50.0/24     10.1.1.5    IGP     100         0     10.10.10.32     65005
    192.168.40.0/24     10.1.1.4    IGP     100         0     10.10.10.32     65004
```

Listagem 4. Rotas recebidas antes e após a falha do SR1

Ao restaurarmos a conectividade ao SR1, as sessões BGP com ele foram restabelecidas. As rotas aprendidas foram repassadas à aplicação do SDX-APP que gerou corretamente as *intents* para as redes 192.168.10.0/24, 192.168.20.0/24 e 192.168.30.0/24. O ONOS, por sua vez, reinstalou as regras de OpenFlow - criadas a partir das *intents* - na tabela de fluxos dos switches e os roteadores dos SA aprenderam as rotas para as referidas redes novamente, sendo a conectividade restaurada.

7. Conclusão

O presente trabalho abordou a construção de um ponto de troca de tráfego baseado em SDN, operado nos padrões dos IX brasileiros, criados pelo projeto PTTMetro, do NIC.br. O foco do estudo foi nos benefícios providos por essa estrutura, objetivando simplificar as tarefas de administração do IX, diminuindo a demanda na equipe de TI, tornando sua operação mais simples e eficaz e provendo um serviço de maior qualidade e ágil para os SA conectados, fomentando a entrada de novos participantes e a continuidade do projeto.

Com a experiência na operação do IX de Minas Gerais (IX.br-MG), foram analisadas diversas tarefas da rotina de um IX, levantando-se as demandas mais onerosas, seja em tempo ou complexidade, dificuldades, tarefas passíveis de falhas e outros problemas que poderiam ser simplificados e até automatizados através do uso de uma solução SDN nesse ecossistema. Desse estudo, chegou-se a diversos casos de uso e funcionalidades a serem implementadas, englobando tarefas administrativas, operacionais e de segurança, como a adoção de topologias mais robustas na infraestrutura do IX, simplificação do processo de ativação de novos participantes, gerenciamento dos acordos de troca de tráfego (privados e públicos), tradução de VLANs, verificação de tráfego indesejado, entre outros.

Além dos benefícios diretos como a maior automatização das tarefas, evitar erros devido a configuração erradas ou inconsistência e a maior vazão nas demandas dos participantes à administração do IX, outra grande vantagem da abordagem é não exigir qualquer recurso adicional dos equipamentos de encaminhamento no ambiente físico, desde que sejam compatíveis com o protocolo OpenFlow. Isso permite, inclusive, que recursos avançados como MPLS ou tradução de VLANs sejam implementados em software, reduzindo o custo dos equipamentos e aumentando a flexibilidade do ambiente.

Agradecimentos

Este trabalho foi parcialmente financiado por Fapemig, CAPES, CNPq, e pelos projetos MCT/CNPq-InWeb (573871/2008-6), FAPEMIG-PRONEX-MASWeb (APQ-01400-14), e H2020-EUB-2015 EUBra-BIGSEA (EU GA 690116, MCT/RNP/CETIC/Brazil 0650/04).

Referências

Akashi, O., Fukuda, K., Hirotsu, T., and Sugawara, T. (2006). Policy-based BGP control architecture for autonomous routing management. In *Proceedings of the 2006 SIGCOMM workshop on Internet network management*, pages 77–82. ACM.

Berde, P., Gerola, M., Hart, J., Higuchi, Y., Kobayashi, M., Koide, T., Lantz, B., O'Connor, B., Radoslavov, P., Snow, W., and Parulkar, G. (2014). ONOS: Towards an open, distributed SDN OS. In *Proceedings of the Third Workshop on Hot Topics in Software Defined Networking*, HotSDN '14, pages 1–6, New York, NY, USA. ACM.

Ceptro.br. PTTMetro - Pontos de Troca de Trafégo Metropolitanos. Disponível em `http://www.ceptro.br/CEPTRO/MenuCEPTROSPPTTMetro`, visitado em 05/12/2015.

Ceron, J., Lemes, L., Granville, L., Tarouco, L., and Bertholdo, L. (2009). Uma solução para gerenciamento de bgp em pontos de troca de tráfego internet. *XXVII Simpósio Brasileiro de Redes de Computadores (SBRC)*.

Gupta, A., Vanbever, L., Shahbaz, M., Donovan, S. P., Schlinker, B., Feamster, N., Rexford, J., Shenker, S., Clark, R., and Katz-Bassett, E. (2014). SDX: A software defined Internet exchange. In *Proceedings of the 2014 ACM conference on SIGCOMM*, pages 551–562. ACM.

Kotronis, V., Dimitropoulos, X., and Ager, B. (2012). Outsourcing the routing control logic: Better internet routing based on SDN principles. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pages 55–60. ACM.

LG - Wikipedia. Looking glass server. Disponível em `https://en.wikipedia.org/wiki/Looking_Glass_server`, visitado em 14/06/2015.

Lin, P., Hart, J., Krishnaswamy, U., Murakami, T., Kobayashi, M., Al-Shabibi, A., Wang, K.-C., and Bi, J. (2013). Seamless interworking of SDN and IP. *SIGCOMM Comput. Commun. Rev.*, 43(4):475–476.

Mambretti, J., Chen, J., and Yeh, F. (2014). Software-defined network exchanges (SDXs): Architecture, services, capabilities, and foundation technologies. In *Teletraffic Congress (ITC), 2014 26th International*, pages 1–6. IEEE.

P. Phaal, S. P. and Mckee, N. (2001). Inmon corporation's sflow: A method for monitoring traffic in switched and routed networks. RFC 3176, RFC Editor. Disponível em `https://tools.ietf.org/html/rfc3176`, visitado em 29/11/2015.

Salsano, S., Ventre, P. L., Prete, L., Siracusano, G., and Gerola, M. (2014). OSHI - open source hybrid IP/SDN networking (and its emulation on Mininet and on distributed SDN testbeds). *arXiv preprint arXiv:1404.4806*.

Shah, S. and Yip, M. (2003). Ethernet automatic protection switching (EAPS). RFC 3619, RFC Editor. Disponível em `https://tools.ietf.org/html/rfc3619`, visitado em 29/11/2015.

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