Roteamento e Alocação de Espectro Ciente da Aplicação em Redes Ópticas Elásticas

Léia S. de Sousa, Lucas R. Costa, Felipe R. de Oliveira, André C. Drummond, Eduardo Adilio Pelinson Alchieri

[1]Departamento de Ciência da Computação (CIC) – Universidade de Brasília (UnB)
Caixa Postal 15.064 – 70910-900 – Brasília – DF – Brasil

Abstract. *Based on the Cross-Layer Design (CLD) paradigm, this paper presents a application-aware routing and spectrum allocation (AA-RSA) solution of Multiple Bulk Data Transfer (MBDT), services which transports massive data amount in a geo-distributed network, taking into account their data traffic pattern to improve service to the demands of resynchronization in networks serving data centers (DC), or inter-data center network (IDC).*

Resumo. *Baseado no paradigma Cross-Layer Design (CLD), este trabalho apresenta uma solução de roteamento e atribuição de espectro ciente da aplicação (AA-RSA) de Múltiplas Transferências de Dados em Massa (MBDT), que são serviços de transporte de volumosas massas de dados em uma rede geo-distribuída. O algoritmo proposto leva em consideração o padrão de tráfego de dados para melhorar o atendimento às demandas de ressincronização em redes utilizadas por centros de dados (CD), ou redes inter-centro de dados (ICD).*

1. Introdução

Baseado no paradigma *Cross-Layer Design* (CLD), que rompe com o conceito de camadas isoladas promovendo a intercomunicação vertical entre camadas, o roteamento ciente da aplicação (*Application-Aware Routing*, AA-R) se destaca do roteamento convencional por explorar informações a respeito das características de cada tipo de aplicação, como *video streaming*, jogos *online* e de transferências de dados, e aplicá-las na melhoria do desempenho da rede e qualidade dessas aplicações [Gerber and Doverspike 2011], [Subramaniam et al. 2013].

As Múltiplas Transferências de Dados em Massa (MBDTs) são um exemplo de aplicação que, por definição, movimentam volumosas quantidades de dados ao longo de uma rede de centro de dados (CD) geo-distribuída, como forma de alcançar melhor desempenho e maior confiabilidade. São implementadas nos ambientes de computação em nuvem para replicar e sincronizar as bases de conteúdos de grandes provedores desse tipo de serviço [Laoutaris et al. 2009]. Hoje, a alocação de recursos para MBDT é estática com taxas previamente definidas, onde uma operação completa de transferência pode levar de dias a semanas, o que é feito tipicamente nos momentos de menor utilização da rede (padrão noturno) e leva os provedores a utilizarem armazenamento de dados em centros de dados alugados durante o dia, sendo que a rede não é ciente das aplicações que são executadas; além disso, não existem técnicas de engenharia de tráfego que garantam as transferências dentro de um específico período de tempo (*deadline*) [Zhang et al. 2015].

O tráfego elástico das MBDT pode ser paralisado e reiniciado conforme prioridade definida pelos *Internet Service Providers* (ISPs). Os canais utilizados são *backbones* de

redes ópticas de alta velocidade, que podem ser linhas proprietárias (ex: *Google*, *Yahoo!* e *Microsoft* [Ghosh et al. 2013]) ou públicas (ex: a *Internet* [Gerber and Doverspike 2011]).

Para atender às expectativas futuras, esses *backbones* poderão implementar o sistema de comunicação óptico elástico (EON). A EON utiliza menores espaçamentos de grades do espectro óptico e transporta mais *bits* por símbolos, o que permite transportar mais dados com menos recursos e melhor ajuste das demandas às larguras de banda [Christodoulopoulos et al. 2011], [Jinno et al. 2009]. Os *transponders* de largura de banda flexível (BVTs) da arquitetura EON, equipamentos que convertem o sinal vindo da rede cliente para um sinal que possa ser comutado no domínio óptico, implementam técnicas *Orthogonal Frequency-Division Multiplexing*(OFDM) capazes de alocar *slots*, ou fatias do espectro óptico, para um determinado caminho de acordo com a taxa de transmissão solicitada. Os *cross-connects* de largura de banda flexível (BV-WXCs) dessa arquitetura estabelecem os caminhos ópticos com os recursos solicitados e podem suportar uma infinidade de caminhos com as mais variadas capacidades [Christodoulopoulos et al. 2011].

Encontrar uma rota e atribuir recursos de espectro com o mais apropriado formato de modulação para acomodar uma demanda constitui o problema fundamental em EON, chamado de problema de roteamento, modulação e atribuição de espectro (RMSA). Tal problema está sujeito às restrições de continuidade do espectro (todos os enlaces do caminho devem utilizar os mesmos *slots*), de contiguidade do espectro (os *slots* que atendem uma demanda devem ser consecutivos), de não sobreposição (duas demandas não podem utilizar simultaneamente o mesmo *slot*), e de banda de guarda (separação entre os *slots* de demandas diferentes). A versão com modulação fixa do problema RMSA é o roteamento e atribuição de espectro (RSA). Ambos são NP-completos [Wan et al. 2012].

Este trabalho propõe um esquema de roteamento em EON, que, a partir de informações recebidas da camada de aplicação, atende um número maior de conexões, aumentando o desempenho do sistema. Para mostrar os ganhos advindos com um esquema de roteamento ciente da aplicação, é definido um cenário de ressincronização ICD, no qual um CD volta a fazer parte da rede após um período de indisponibilidade e precisa atualizar as informações armazenadas (seu estado). Para isso, seus pares encaminham grandes volumes de dados com as informações atualizadas para que ocorra esta ressincronização.

Este cenário reflete a execução de um sistema distribuído, cujas informações são replicadas para garantir tolerância a falhas [Castro and Liskov 2002]. A quantidade de réplicas necessárias varia de acordo com o protocolo de replicação e o modelo de sistema adotado. Neste trabalho, consideramos o modelo mais genérico de falhas onde são necessárias 4 réplicas para que uma delas possa falhar [Castro and Liskov 2002]. Embora este número possa ser aumentado para permitir que mais réplicas falhem, na prática utiliza-se poucas réplicas [Vukolic 2010]. Desta forma, durante a ressincronização, um lote de comunicação de muitos para um ($M \rightarrow 1, M \geq 3$) é estabelecido. Na solução proposta, emprega-se uma abordagem de transferência fim-a-fim, sem armazenamento intermediário.

Os resultados obtidos indicam que o algoritmo ciente da aplicação estabelece cerca de 30% a mais de conexões para ressincronizações em comparação com o roteamento convencional, mantendo seu desempenho superior mesmo em condições de tráfego pesado na rede. Delegar mais responsabilidades para o roteamento aumenta o número de ressincronizações bem sucedidas, o que reduz as despesas com armazenamento em trânsito.

Neste sentido, as principais contribuições deste trabalho são: *(i)* proposta de um algoritmo de roteamento em EON ciente da aplicação de MBDT, chamado AA-RSA; e *(ii)* realização de uma série de simulações para avaliar o desempenho do algoritmo proposto face a um algoritmo RSA convencional, as quais demonstram claramente os ganhos advindos com a troca de informações entre as camadas de aplicação e de rede.

2. Trabalhos Relacionados

A partir do surgimento das pesquisas em EON [Jinno et al. 2009], soluções de roteamento e alocação de recursos vêm destacando o potencial elástico e dinâmico da tecnologia para melhorar a capacidade de transporte das redes, como [Wan et al. 2012] que propõe uma solução RSA com modulação fixa e outra que adapta modulações empregando níveis de modulação a partir do mais alto (com menor largura de banda e que transportam menos *bits* por símbolo) para satisfazer a mais demandas, sendo ambos roteamentos convencionais.

O paradigma CLD tem tradicionalmente relacionado as redes ópticas às pesquisas que tratam de *Physical Layer Impairments* (PLIs), mais especificamente *Impairment-Aware Routing and Wavelength Assignment* (IA-RWA) nas redes WDM, cujas soluções de roteamento apresentadas consideram as limitações da camada física da rede. Em [Zhao et al. 2015], por exemplo, limitações de não linearidade das fibras ópticas, que resultam em perdas nos sinais devido a mudanças de potência ao longo do caminho de propagação da luz em longas distâncias, bem como de defeitos de fabricação dessa fibra, são consideradas para melhorar o desempenho do roteamento. Já o roteamento ciente da aplicação foi explorado em [Song et al. 2012], onde propôs-se o mapeamento de máquinas virtuais aos seus respectivos *hosts* físicos para realizar migrações de acordo com a disponibilidade de recursos e com as distâncias entre os nós físicos, o que reduziu o tráfego de dados nos enlaces da rede e o consumo de energia.

Também motivado pela preocupação com o volume de tráfego, o problema MBDT foi destacado em [Li et al. 2012], com uma proposta para reduzir o tráfego inter-domínio utilizando os canais comerciais de ISPs associados a redes dedicadas, através das quais é imposto um esquema de escalonamento de dados em diferentes janelas de tempo. Em [Laoutaris et al. 2009], as MBDT são realizadas em horários noturnos. Também foi proposto o sistema *NetStitcher* [Laoutaris et al. 2011], capaz de identificar rapidamente as larguras de banda ociosas em qualquer enlace da rede geo-distribuída. Em comum, essas soluções propuseram variados modos de transferência aliados ao roteamento tradicional.

Em [Lu et al. 2015] e [Lu and Zhu 2015], são mostradas soluções para transferências de dados em massa (BDT) utilizando a tecnologia EON, as quais tentam aproveitar o espectro óptico fragmentado após o atendimento de requisições orientadas a fluxos para realizar as transferências orientadas a dados, embora não se tenha considerado aspectos dessas aplicações.

Este trabalho explora essa proposta de RSA com modulação fixa no âmbito do roteamento ciente das aplicações. Os resultados comparativos são apresentados para demonstrar o ganho de eficiência que pode ser alcançado quando a camada de rede possui mais informações para auxiliar a tomada de decisões.

3. Problema das Múltiplas Transferências de Massa de Dados na Ressincronização ICD

Um sistema distribuído capaz de tolerar falhas de *hardware* e até intrusões precisa realizar replicações geo-distribuídas dos seus dados. Cada nó de armazenamento desse sistema é responsável por um conjunto de informações, chamado de partição, que são replicados separadamente. Um grupo de replicação é um grupo de nós responsável pela mesma partição. Assim, cada CD pode participar simultaneamente de vários grupos diferentes. Os sistemas de gerenciamento de dados distribuídos (DMS) que lidam com esses CDs possuem um correto e consistente mapeamento dos grupos de replicação em seus membros e suas respectivas localizações [Sharov et al. 2015].

Várias rodadas de trocas de mensagens dentro de determinado grupo, que são definidas de acordo com cada protocolo específico, são realizadas quando CDs submetem requisições entre si. O pressuposto fundamental para que este mecanismo possa ocorrer é o estado comum das réplicas, alcançado mediante sincronizações e ressincronizações [Castro and Liskov 2002]. Um nó membro inativo que deseja voltar à rede após um período de indisponibilidade e cujo seu estado encontra-se desatualizado, pode submeter uma solicitação de integração a esse grupo e assim, acionar os serviços de ressincronização. O grupo de réplicas designado pelo DMS possui o mesmo estado e um mapeamento das partições a serem replicadas [Agrawal et al. 2013a]. Para atender a solicitação de ressincronização do nó que está retornando, o sistemas de gerenciamento de dados distribuídos define os respectivos nós candidatos capazes de atender a essa ressincronização e dispara as MBDTs dentro de um período de tempo suficiente para a completa atualização do nó solicitante.

Do ponto de vista da camada de rede, normalmente a aplicação MBDT é atendida juntamente com outras solicitações de roteamento, embora suas taxas de dados sejam muito variadas, sem distinção do tipo de tráfego ou do seu nível de prioridade, para as quais o roteamento estabelece um caminho com largura de banda disponível e move o tráfego fim-a-fim. Logo, embora as MBDTs sejam tolerantes ao atraso, algumas requisições podem não ser atendidas devido a indisponibilidade de recurso suficiente por um período estendido de tempo e, consequentemente, estrangulamento do seu *deadline* [Zhang et al. 2015].

A solução de roteamento em EON ciente da aplicação MBDT para ressincronização de CDs, AA-RSA, impõe maior complexidade por incorporar tomadas de decisões relacionadas a busca de combinações de um subconjunto dos nós aptos à transferência. Assim, o problema RSA tem seu espaço de busca ampliado de tal maneira que, além de tentar estabelecer caminhos ópticos com fatias espectrais suficientes, essas tentativas são experimentadas para todas os subconjuntos de 3 CDs transmissores que integram o grupo de replicação. Neste trabalho, consideramos grupos de transferências de 3 CDs porque o número de nós que compõem um grupo de ressincronização são geralmente 4 ou 5, sendo que as combinações de requisições consideram apenas os nós que farão transferências, que são 3 ou 4 [Vukolić 2010].

Esta solução é possível a partir da comunicação vertical com o DMS, que é encarregado de informar o tipo de tráfego a ser movido e de fornecer uma abstração da localização dos membros participantes do grupo de replicação, normalmente inacessível pelo roteamento. Essa troca de informações é garantida via CLD. A decisão por algum

subconjunto de replicação atende aos requisitos de tolerância a falhas da aplicação, garante melhor desempenho na ressincronização (mais requisições são aceitas) e reduz a utilização de recursos atendendo somente ao mínimo de transferências necessárias.

Para resolver o problema AA-RSA, a topologia física da EON foi modelada como um grafo direcionado $G(V, E)$, onde V e E denotam o conjunto de nós e enlaces da fibra, respectivamente. Assume-se que V é constituído por BV-WXCs e CDs geograficamente distribuídos e interconectados, representados como V^h. O conjunto V^h representa as localizações de CD e comutadores ópticos simultaneamente, e portanto, esse tipo de nó é capaz de solicitar e receber uma ressincronização. Cada enlace pode acomodar no máximo, BW *slots* de frequência do espectro.

Sobre essa rede é realizado o roteamento e alocação desses *slots* para determinar um caminho k com largura de banda B disponível a fim de transportar o tráfego ICD do tipo *Bulk Data Transfer* (BDT), que são chamadas orientadas a dados [Lu et al. 2015]. Uma requisição BDT é um *bulk*, modelado como $r_u = \{s_u, d_u, C_u, Dl_u\}$, onde $s_u \in V^h$ e $d_u \in V^h - \{s_u\}$ são a origem e o destino da chamada, C_u é a quantidade de dados a ser transferida e Dl_u é o *deadline* da transferência.

A requisição de ressincronização MBDT encaminhada do DMS e recebida no roteamento, é um conjunto de *bulks* r_u de tamanho n representado como $R = \{r_{u_1}, r_{u_2}, ..., r_{u_n}\}$, onde existe um conjunto S contendo todas as origens das transmissões e d e Dl são iguais para todas as requisições. Como principal característica deste tipo de aplicação, sua redundância inerente requer que a rede ICD seja altamente tolerante a falhas multi-nós. Por isso, um conjunto R_u deve ter uma cardinalidade mínima de 3 *bulks*. Com este conhecimento, as demandas de conexão terão semânticas que consideram a camada de aplicação, especificando o grupo de redundância participante, o que restringe o espaço de buscas por nós candidatos por parte da camada de rede. Devido a esta restrição, isto é, alocar recursos para as três transferências, o roteamento não tenta realizar alocações se a requisição viola esta condição.

Para exemplificar, a Figura 1 mostra um processo de ressincronização para o CD_5. Um lote MBDT com $S = \{CD_1, CD_2, CD_3, CD_4\}$ é enviado para o $d = \{CD_5\}$, movimentando um volume $C = \{1TB\}$, que deve ser recebido no destino dentro de um limite de tempo $Dl = \{30min\}$. Um algoritmo de roteamento convencional, como é o caso do RSA [Wan et al. 2012], adaptado para tratar de requisições orientadas a dados, considera esse lote como um conjunto de quatro chamadas independentes para as quais a busca por recursos é feita de maneira sequencial conforme a chegada dessas requisições. Nesse cenário, pode ocorrer de menos do que três chamadas serem atendidas e as demais serem bloqueadas devido ao esgotamento do tempo enquanto se aguardava por recursos de largura de banda ou mesmo devido a sua indisponibilidade nos enlaces compartilhados. Quando isso acontece, mesmo que algumas chamadas individuais sejam atendidas, não ocorre uma ressincronização com garantias de tolerância a falhas.

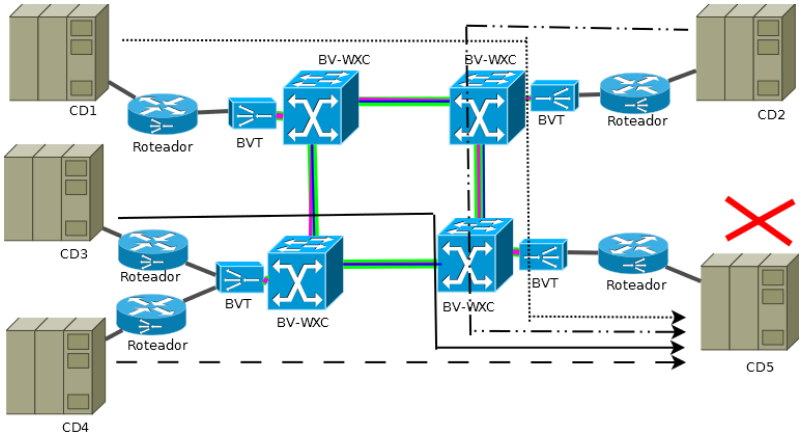

Figura 1. MBDT na ressincronização cross-DC

Ciente da importância dessa restrição para a aplicação, o roteamento pode esforçar-se em atender ao menos 3 *bulks* desse lote, limite inferior denominado fator de replicação, evitando bloqueá-las antes de ter certeza sobre a inexistência de recursos. Para isso, um algoritmo de combinações de lote pode acessar um lote e buscar por um subconjunto mínimo com essas características. Na Figura 1, as possíveis combinações seriam $\{\{CD1, CD2, CD3\}, \{CD1, CD2, CD4\}, \{CD2, CD3, CD4\}, \{CD1, CD3, CD4\}\}$. O AA-RSA propõe a busca por uma combinação mínima dos elementos de S. O atendimento de menos do que três demandas quebra a restrição de tolerância a falhas da aplicação de ressincronização de CDs. Por outro lado, atender mais do que 3 chamadas levaria ao desperdício de recursos.

4. Algoritmo AA-RSA

O algoritmo AA-RSA (Algoritmo 1) baseia-se no algoritmo RSA com modulação fixa [Wan et al. 2012]. Nas linhas 1 e 2 são inicializados o fator de replicação *b*, que representa a quantidade mínima de réplicas de CD para a ressincronização, e K, a quantidade de menores caminhos entre quaisquer dois nós. A função $COMBINATIONS(R_u, b)$, na linha 3, calcula todas as combinações possíveis de b CDs dentro do universo do lote R_u, que contém n chamadas para efetuar uma ressincronização. O seu processamento retorna uma matriz m de $\frac{n!}{b!(n-b)!}$ linhas e b colunas. Cada linha de m é um subconjunto de combinações *comb* de tamanho b. Em cada $comb$ executa-se para cada um de seus *bulks* i a função $KSP(i, K)$, linha 5, que retorna os K menores caminhos entre $s(i)$ e $d(i)$, origem e destino da chamada i, respectivamente, em ordem crescente de tamanho.

Na busca do *k*-ésimo caminho viável, nas linhas 6-13, para cada candidato $k \in K$ calcula-se a distância de k, linha 7, que é comparada com a fórmula $\mathscr{D} \times \tau$, onde $\mathscr{D}$ é o diâmetro da rede e τ é um parâmetro que restringe esse diâmetro a um limite de tamanho que permita utilizar a largura de banda máxima $MaxRate$ do caminho, equivalente à capacidade máxima de um BVT. Essa fórmula impede que requisições de conexões para os caminhos menores do que um determinado limite do maior caminho sejam penalizadas, com a alocação da mínima largura de banda disponível. A ideia é que recursos de caminhos relativamente pequenos sejam liberados mais rapidamente, desocupando recursos para as chamadas seguintes. Às requisições dos caminhos que não atendem a essa condição, na linha 10, é atribuída uma largura de banda B mínima, equivalente ao quociente da quantidade de dados transferida dentro do *deadline* total da sua chamada, $C_i \div Dl_i$.

Algoritmo 1 AA-RSA

1: $b \leftarrow 3$
2: $K \leftarrow 3$
3: $\mathbf{COMBINATIONS}(R_u, b)$
4: **for** $i = 1 \rightarrow b$ **do**
5: $\quad \mathbf{KSP}(i, K)$
6: $\quad$ **for** $k = 1 \rightarrow K$ **do**
7: $\quad\quad dist\,(k) = \sum_{l=0}^{|k|-1} v_l, v_l \in k$
8: $\quad\quad$ **if** $dist(k) \leqslant \mathscr{D} \times \tau$ **then**
9: $\quad\quad\quad B \leftarrow MaxRate$
10: $\quad\quad$ **else** $B \leftarrow C_i \div Dl_i$
11: $\quad\quad$ **end if**
12: $\quad\quad$ **Test RSA restrictions**
13: $\quad$ **end for**
14: $\quad$ **if** $\exists k \in K \mid k$ **is viable then**
15: $\quad\quad \mathscr{P} \leftarrow k$
16: $\quad\quad \mathscr{B} \leftarrow B$
17: $\quad$ **end if**
18: **end for**
19: $getNext(comb)$
20: **if** $comb$ can be attended **then**
21: $\quad$ accept $(R, \mathscr{P}, \mathscr{B})$
22: **else** block (R)
23: **end if**

Definidos o caminho e a largura de banda, na linha 12 é verificada a restrição de continuidade, de contiguidade e de não-sobreposição do espectro óptico. As linhas $14 - 17$ analisam se um caminho e sua respectiva taxa foi obtida para cada requisição do subconjunto de combinações, que em caso positivo, são previamente alocadas e guardadas em $\mathscr{P}$ e $\mathscr{B}$, conjunto dos caminhos da combinação e conjunto das larguras de banda das requisições dessa mesma combinação. A linha 19 faz tentativas com todos so subconjuntos de combinações até exaurir a matriz de combinações m ou até que encontre uma combinação viável. A aceitação das MBDTs para a ressincronização é feita na linha 20, caso um subconjunto possa ser aceito. Se não for o caso, a requisição é bloqueada.

O algoritmo recebe um lote R como entrada, mas a aceitação do serviço se dá com o atendimento de apenas um subconjunto de R, o que significa que chamadas remanescentes são marcadas como bloqueadas pelo plano de controle da rede. Isso é possível porque na comunicação *cross-layer*, o roteamento, em contato com a aplicação, sabe que se R é proveniente de um grupo de replicação e portanto, os dados das transferências são similares, então algumas das requisições podem ser bloqueadas sem prejuízo do serviço.

Complexidade do Algoritmo

O RSA dinâmico com caminho fixo proposto por [Wan et al. 2012], que é dividido em dois passos, computa os *k* menores caminhos sem *loops* usando o algoritmo KSP no primeiro desses passos com tempo $O(K|V|^3)$, e em seguida, utiliza operações de detecção e intersecção de espectro no segundo passo, levando $O(K^2)$ para a verificação em todos os enlaces que compõem o caminho. Assim, seu tempo total é $O(K^3|V|^3)$.

O AA-RSA dinâmico, baseado no RSA acima, para obter as combinações de *bulks* do lote na função $COMBINATIONS(R, b)$, com um lote R de tamanho n e um fator de replicação b, tem complexidade de tempo $O\left(\frac{n!}{b!(n-b)!}\right)$ que é exponencial. Para cada elemento de uma combinação é invocada a função $KSP\,(i, K)$, com complexidade de $O(V^3)$. A atribuição de espectro é feita pela política *First-Fit* de complexidade linear. O diâmetro da rede $\mathscr{D}$ é calculado uma única vez. A partir do algoritmo de *Dijkstra*, que leva

tempo $O(V^2)$, é encontrada a distância de cada nó s do grafo para todos os demais nós desse mesmo grafo. A maior distância dentre todas as menores distâncias representa $\mathscr{D}$. Assim, a complexidade de tempo é $O\left(\left(\frac{n!}{b!(n-b)!}\right) * (V^3)\right)$. No entanto, a literatura mostra que o tamanho de R geralmente é de 3 requisições de conexões de ressincronizações, visto que no mundo real é desvantajoso, do ponto de vista do custo capital e operacional, possuir um conglomerado muito grande de recursos que são poucos solicitados [Vukolić 2010].

5. Avaliação de Desempenho

Através do simulador ONS[Costa et al.][1] desenvolvido com base no simulador WDM-Sim [Drummond 2015], simulações foram realizadas para verificar o ganho de desempenho do algoritmo proposto em relação ao algoritmo RSA convencional [Wan et al. 2012]. Eventos dinâmicos de chegadas e partidas de requisições foram simuladas na topologia NSFNET (Figura 2) com 14 nós, dos quais cinco deles $(0, 7, 11, 12, 13)$ foram definidos como CDs, e nas topologias USA e Pan-Euro reunidas por cabo de fibra óptica submarino de capacidade suficiente, com 24 e 28 nós, respectivamente (Figura 3), onde os nós $\{1, 10, 12, 20, 21, 22, 28, 30, 46\}$ são as representações de CDs. As distâncias físicas são destacadas nos enlaces. As definições dos CDs foram feitas baseadas nas localizações de centros de dados do Google [Google 2015], nas quais foram consideradas ressincronizações de uma única partição.

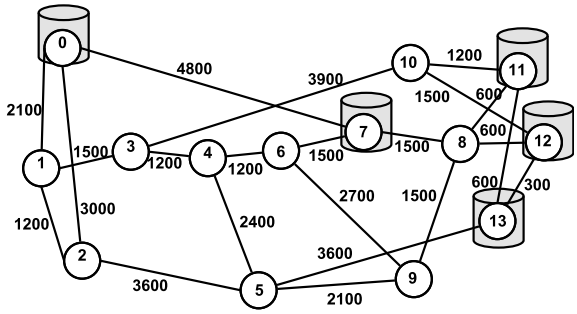

Figura 2. Topologia da rede NSFNET

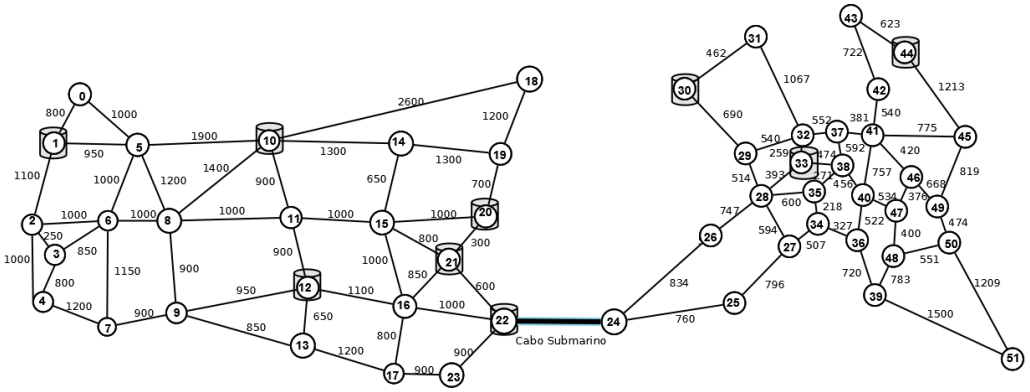

Figura 3. Topologia das redes USA e Pan-Euro, reunidas

Na implementação foram considerados 15 *transponders* por nó, cada um com capacidade de 8 *slots*. Cada *slot* possui largura de banda de $12.5GHz$ e cada enlace possui

[1]O simulador ONS é um simulador híbrido para redes WDM e EON, disponibilizado em http://comnet.unb.br/br/grupos/get/ons/download

120 *slots* de frequência. Assumiu-se ainda que são empregados dois *slots* como banda de guarda. Nos dois cenários, os dois algoritmos foram testados com as modulações *Binary Phase Shift Keying* (BPSK), que transmite um *bit* por símbolo e por isso fornece baixa taxa de dados, mostrando-se ideal para as distâncias mais longas, e *Quadrature Phase Shift Keyed* (QPSK), com dois *bits* por símbolo e uma taxa de dados mais elevada, que permite transmitir duas vezes mais quantidade de dados no mesmo canal com a mesma taxa oferecida à BPSK, e que portanto, é mais adequada para distâncias menores. Além disso, por se tratar do problema RSA, a solução não considera distâncias para definição de qualquer modulação, no entanto, o ajuste da modulação implica na definição da quantidade de *bits* por símbolo empregados no transporte dos dados. Nos testes, essas modulações tiveram resultados próximos quanto a taxa de bloqueio, embora QPSK tenha se sobressaído.

Cada simulação foi realizada 5 vezes utilizando o método de replicações independentes. Para os resultados apresentados foram calculados intervalos de confiança com 95% de confiabilidade. Foram realizadas 100.000 chamadas com origens e destinos distribuídos uniformemente dentro do subconjunto de localizações dos CDs. Os *deadlines* das chamadas foram de 10, 15 e 20 unidades de tempo. Foram empregados dois cenários de carga, um deles considerado leve, com taxas de chegadas de 2 a 10 chamadas por unidade de tempo com incrementos de 2 chegadas, e um outro cenário com tráfego pesado, cujos números de chegada de chamadas variam de 20 a 100 por unidade de tempo com incrementos de 20 chegadas. Quanto aos volumes dos *bulks*, cada terça parte das chamadas transferiram $100GB$, $500GB$ e $1TB$, respectivamente. Para essas chamadas foram configurados lotes com 3 e 4 requisições, que são parâmetros largamente encontrados na literatura [Agrawal et al. 2013b]. A simulação foi executada em uma máquina Intel Core 2Quad de 2.66GHz com 4GB de RAM, onde verificou-se que o tempo médio de execução do algoritmo AA-RSA, para atendimento de um lote, é da ordem de $1 : 27ms$ com a menor topologia e $90 : 31ms$ com a topologia reunida.

A atribuição de espectro é realizada utilizando a política *First-Fit*, onde para cada enlace pertencente a rota estabelecida, são considerados os seus respectivos BW *slots* de frequência em ordem crescente dos índices, e a demanda é acomodada iniciado-se pelos *slots* de índices menores que estão com a capacidade livre. O parâmetro τ, de restrição do diâmetro da rede utilizado no algoritmo, foi assumido como $0,5$ de modo a comparar os tamanhos dos caminhos candidatos à metade do maior caminho que pode ser encontrado.

Taxa de Bloqueio de Banda Passante (BBR)

A taxa de bloqueio de largura de banda (BBR) de *bulks* equivale a taxa do uso de largura de banda correspondente aos *bulks* bloqueados dividida pelo total de largura de banda em uso por todas as chamadas. A Figura 4 mostra que o algoritmo RSA bloqueia mais *bulks* do que o algoritmo AA-RSA no cenário das topologias reunidas, onde os caminhos ópticos estabelecidos são maiores. No cenário de tráfego leve, os algoritmos mantêm o seu comportamento até o limite de 10 chegadas. Para as duas modulações testadas, o AA-RSA apresenta melhor desempenho do que o RSA. Para ambos os algoritmos, a modulação QPSK resultou em menores taxas de bloqueio de largura de banda do que a modulação BPSK. O resultado de BBR do AA-RSA_QPSK inicia com uma taxa de cerca de 16% e chega a atingir 55% de bloqueio, ao passo que o RSA com essa mesma modulação inicializa com 28% de bloqueio, chegando a 80%.

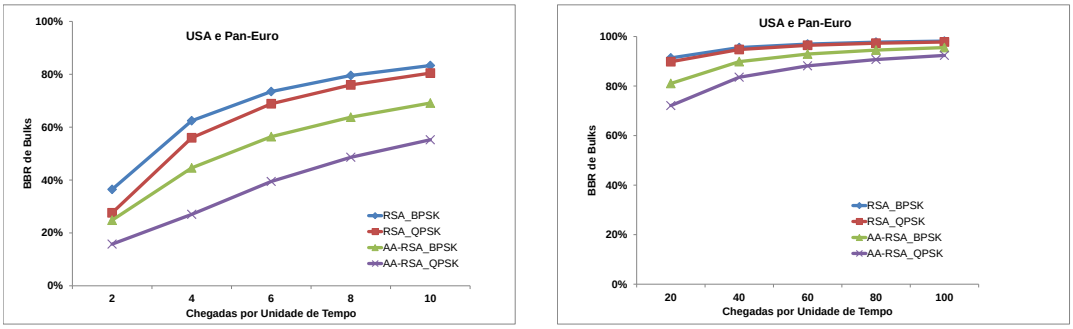

Figura 4. BBR de *bulks* nas topologias USA e Pan-Euro reunidas com tráfego leve (esquerda) e pesado (direita).

Cabe ressaltar que parte dos bloqueios do AA-RSA são na verdade chamadas descartadas dos lotes com 4 *bulks* e, portanto, não significa prejuízo no atendimento a aplicações. Quando ocorrem duas chegadas por unidade de tempo, cerca de 90% dos bloqueios dizem respeito a descartes, mas esta taxa cai para 17% quando o número de chegadas aumenta para 10. O AA-RSA_BPSK descarta menos chamadas do que o AA-RSA_QPSK. A explicação para o bom desempenho da modulação QPSK é a alta capacidade de transporte em relação a BPSK.

Para o tráfego pesado, os dois RSAs tiveram taxa de desempenho próximas e seus percentuais de utilização de BVTs e de espectro óptico foram praticamente os mesmos, embora o RSA_BPSK tenha bloqueado mais recursos mesmo após atingir o ponto de saturação a partir de 60 chegadas por unidade de tempo. As diferenças em torno de 1% entre eles, mostram que em condições pesadas de tráfego, BPSK e QPSK exibem o mesmo comportamento. Já no caso do AA-RSA, o uso da modulação BPSK levou a mais bloqueio com uma desvantagem de cerca de 4% em comparação com QPSK. O mecanismo de busca por recurso disponível desse algoritmo eleva a taxa de utilização de BVTs e de espectro óptico em pelo menos 3%, levando a uma menor BBR. Os AA-RSAs, comparados com os RSAs, aumentam a taxa de utilização desses recursos em quase 50% a medida que o número de chegadas também cresce.

Já no cenário da rede NSFNET (Figura 5), percebeu-se no gráfico à esquerda que a utilização de BVTs manteve-se em torno de 22 e 24% pelo RSA_BPSK e RSA_QPSK, respectivamente, sendo que os resultados das taxas de bloqueio de largura de banda tiveram mínimas diferenças de 1% em condições leves de tráfego. A justificativa está na quantidade e nas localizações dos nós CDs. Em um subconjunto de CDs relativamente pequeno, a possibilidade de sorteio de um nó cujos recursos nos caminhos adjacentes a ele ainda não tenha sido desocupados é muito alta. Além disso, 80% deles são concentrados no lado leste da topologia enquanto que existe um único CD situado no lado oposto, resultando em mais transferências de dados entre os membros a leste do que entre qualquer um desses membros e o CD na região oposta da rede. Os algoritmos AA-RSAs bloquearam menos recursos e novamente a modulação QPSK resultou em menor taxa de bloqueio e menor taxa de utilização de BVTs e espectro.

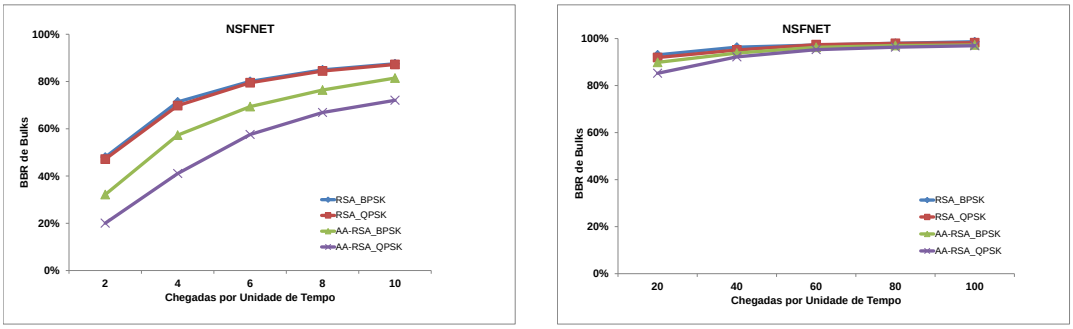

Figura 5. BBR de *bulks* na topologia NSFNET com tráfego leve (esquerda) e pesada (direita).

No gráfico à direita, percebe-se que o RSA_QPSK tem uma taxa de bloqueio por volta de 92% com 20 chegadas registradas, que aumenta para 95% com 40 chegadas e em seguida, atinge 97% na ocorrência de 60 chegadas por unidade de tempo, quando ultrapassa em quase 0,5% o RSA_BPSK, que vinha mantendo a BBR mais alta. A medida que a saturação da rede vai sendo atingida, esses dois algoritmos vão mantendo diferenças muito pequenas até que, com o número máximo de chamadas registradas o RSA_BPSK acaba bloqueado mais. Novamente, as localizações dos nós sorteados e a ocupação de recursos dos canais adjacentes com tráfego pesado justificam esses resultados. O AA-RSA_QPSK segue com menor bloqueio e experimenta uma rápida elevação logo com o aumento de tráfego, onde ocorre uma ligeira diminuição de utilização de BVTs, o que leva a entender que a taxa de ocupação da rede não permitiu o estabelecimento de novos caminhos ópticos para as requisições que solicitavam conexões. Sua diferença percentual para o AA-RSA_BPSK foi de 0,6% com a máxima taxa de chegadas registradas.

Note que nas simulações, os RSAs cientes da aplicação resultaram em menores taxas de bloqueio de largura de banda e maior utilização de BVTs devido a sua política de busca de uma combinação de chamadas para a qual existem recursos. Os RSAs clássicos tiveram taxas mais elevadas de bloqueio embora os número de recursos disponíveis como BVTs e espectro de frequência estivesse entre 6% e 12% a mais que os outros algoritmos. No geral, a modulação QPSK, de maior capacidade, demostrou melhor eficiência.

Avaliação da Qualidade do Serviço

A taxa de sucesso do serviço diz respeito às ressincronizações que se efetuaram, ou seja, o percentual de lotes atendidos. Nas topologias reunidas (Figura 6) onde se registra o tráfego leve no gráfico à esquerda, verifica-se que, como esperado, o AA-RSA_QPSK leva a um maior estabelecimento de conexões de lotes e consegue manter sua taxa de aceitação em nível elevado. O AA-RSA_BPSK tem desempenho inferior mas ainda assim, apresenta vantagens quando comparado com os RSAs clássicos. No gráfico à direita, os algoritmos cientes da aplicação tendem a manter o mesmo comportamento, mesmo com uma queda um pouco acentuada para o QPSK quando ocorrem 40 chegadas por unidade de tempo. Como a taxa de aceitação vai diminuindo, a taxa de utilização de recursos vai se reduzindo no mesmo passo. Por fim, com número máximo de chegadas, a versão com QPSK aceita 9,94% das conexões enquanto que a versão com BPSK tem taxa de 6,44%.

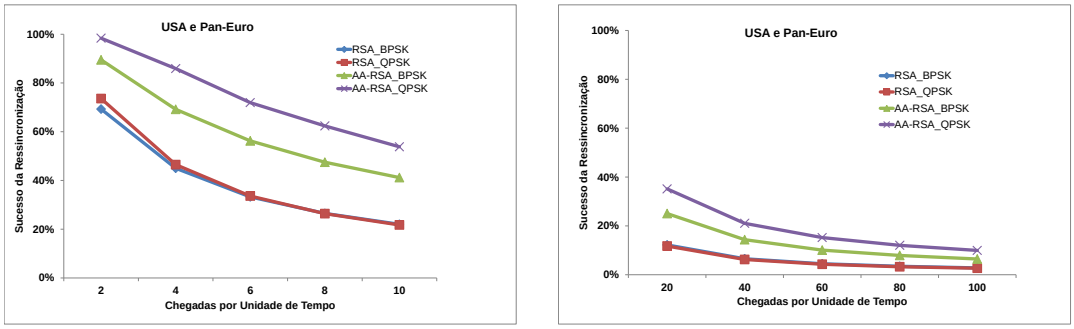

Figura 6. Taxa de sucesso de conexões atendidas nas topologias USA e Pan-Euro reunidas mediante tráfego leve (esquerda) e pesado (direita).

Quanto aos dois RSAs na Figura 6 à esquerda, como as taxas de bloqueio de largura de banda foram altas nos resultados anteriores, esse fator se reflete nas taxas de aceitação de conexões. O QPSK começa com aceitação de pouco mais de 73% enquanto que o BPSK tem 69%. Como esses dois algoritmos tratam as requisições na sequência em que são dadas, a probabilidade de atender a pelo menos 3 chamadas de um lote e resultar em uma ressincronização é pequena. O mesmo desempenho é percebido no gráfico à direita, quando seus resultados tendem a ficar muito próximos, com diferenças menores que 0,5%.

Já com relação a topologia NSFNET (Figura 7) foi registrado que a taxa de aceitação nessa rede é inferior se comparada com as topologias reunidas, mostradas anteriormente. O gráfico à direita mostra queda acentuada da taxa de aceitação do AA-RSA_QPSK em condições de tráfego pesado, em uma rede com concentração de nós. Os RSAs conservam a taxa de utilização de BVTs na faixa dos 43 a 44% e 11 a 13% de utilização do espectro, enquanto que os AA-RSAs tem variações de 28 a 32% de utilização de BVTs e até 25% de utilização do espectro. Com a diminuição de aceitações, o número de chamadas descartadas tende a diminuir enquanto que o número de chamadas efetivamente bloqueadas aumenta, e lotes acabam sendo inteiramente bloqueados devido ao alto congestionamento na rede por conexões ativas.

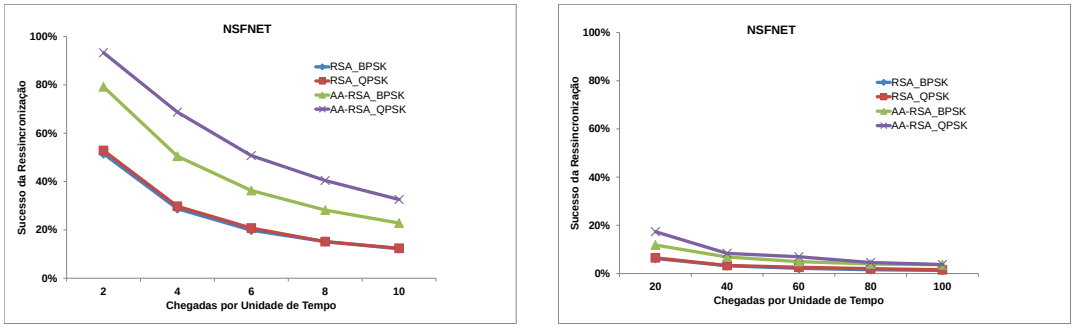

Figura 7. Taxa de sucesso de conexões atendidas na topologia NSFNET mediante tráfego leve (esquerda) e pesado (direita).

De maneira geral, os resultados mostram que algoritmos de roteamento que conhecem a aplicação para a qual precisam fornecer recursos tendem a ter melhor desempenho do ponto de vista dos serviços se comparado com algoritmos convencionais.

6. Considerações Finais

Este trabalho propõe uma solução de roteamento ciente da aplicação que eficientemente atende ao serviço de ressincronização ICD e reduz as taxas de bloqueio de banda de chamadas orientadas a dados. Os resultados comparativos mostram que a aplicação MBDT alcança maior sucesso quando a abordagem CLD é aplicada, visto que as características da aplicação podem servir de entrada para as tomadas de decisões na camada de rede.

Como trabalhos futuros pretendemos estender a proposta AA-RSA para AA-RMSA, fornecendo espaço de busca para a modulação mais adequada à distância da transmissão, adaptando as taxas oferecidas de acordo com as demandas. Com isso, será possível que mais recursos de largura de banda estejam disponíveis e assim resultem em maior aceitação de chamadas. Além disso, outros tipos de tráfegos serão acrescidos a esse cenário para simular um ambiente mais próximo da realidade, com compartilhamento dos canais de comunicação. Técnicas de escalonamento de recursos serão consideradas no atendimento das demandas com variação de tipos de tráfego e prioridades.

Referências

Agrawal, D., El Abbadi, A., Mahmoud, H., Nawab, F., and Salem, K. (2013a). Managing geo-replicated data in multi-datacenters. In Madaan, A., Kikuchi, S., and Bhalla, S., editors, *Databases in Networked Information Systems*, volume 7813 of *Lecture Notes in Computer Science*, pages 23–43. Springer Berlin Heidelberg.

Agrawal, D., El Abbadi, A., Mahmoud, H. A., Nawab, F., and Salem, K. (2013b). Managing geo-replicated data in multi-datacenters. In *Databases in Networked Information Systems*, pages 23–43. Springer.

Castro, M. and Liskov, B. (2002). Practical Byzantine fault-tolerance and proactive recovery. *ACM Transactions on Computer Systems*, 20(4):398–461.

Christodoulopoulos, K., Tomkos, I., and Varvarigos, E. (2011). Elastic bandwidth allocation in flexible ofdm-based optical networks. *Journal of Lightwave Technology*, 29(9):1354–1366.

Costa, L. R., de Sousa, L. S., de Oliveira, F. R., da Silva, K. A., Júnior, P. J. S., and Drummond, A. C. ONS: Optical Network Simulator - WDM/EON. `http://comnet.unb.br/br/grupos/get/ons/download`.

Drummond, A. C. (2015). Wdmsim: Wdm optical network simulator. `http://www.lrc.ic.unicamp.br/wdmsim/`.

Gerber, A. and Doverspike, R. (2011). Traffic types and growth in backbone networks. In *Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2011 and the National Fiber Optic Engineers Conference*, pages 1–3.

Ghosh, A., Ha, S., Crabbe, E., and Rexford, J. (2013). Scalable multi-class traffic management in data center backbone networks. *Selected Areas in Communications, IEEE Journal on*, 31(12):2673–2684.

Google (2015). Google data centers. `https://www.google.com/maps/d/viewer?mid=zXkGqQo6GvyA.kRMGK2Zpmpbg&hl=en`. Acessed em 16/03/2015.

Jinno, M., Takara, H., Kozicki, B., Tsukishima, Y., Sone, Y., and Matsuoka, S. (2009). Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. *Communications Magazine, IEEE*, 47(11):66–73.

Laoutaris, N., Sirivianos, M., Yang, X., and Rodriguez, P. (2011). Inter-datacenter bulk transfers with netstitcher. In *ACM SIGCOMM Computer Communication Review*, volume 41, pages 74–85. ACM.

Laoutaris, N., Smaragdakis, G., Rodriguez, P., and Sundaram, R. (2009). Delay tolerant bulk data transfers on the internet. In *ACM SIGMETRICS Performance Evaluation Review*, volume 37, pages 229–238. ACM.

Li, Y., Wang, H., Zhang, P., Dong, J., and Cheng, S. (2012). D4d: Inter-datacenter bulk transfers with isp friendliness. In *Cluster Computing (CLUSTER), 2012 IEEE International Conference on*, pages 597–600. IEEE.

Lu, W. and Zhu, Z. (2015). Malleable reservation based bulk-data transfer to recycle spectrum fragments in elastic optical networks. *Lightwave Technology, Journal of*, 33(10):2078–2086.

Lu, W., Zhu, Z., and Mukherjee, B. (2015). Data-oriented malleable reservation to revitalize spectrum fragments in elastic optical networks. In *Optical Fiber Communications Conference and Exhibition (OFC), 2015*, pages 1–3.

Sharov, A., Shraer, A., Merchant, A., and Stokely, M. (2015). Automatic reconfiguration of distributed storage. In *Autonomic Computing (ICAC), 2015 IEEE International Conference on*, pages 133–134. IEEE.

Song, F., Huang, D., Zhou, H., and You, I. (2012). Application-aware virtual machine placement in data centers. In *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on*, pages 191–196. IEEE.

Subramaniam, S., Brandt-Pearce, M., Demeester, P., and Saradhi, C. V. (2013). *Cross-layer design in optical networks*. Springer.

Vukolić, M. (2010). The byzantine empire in the intercloud. *ACM SIGACT News*, 41(3):105–111.

Wan, X., Hua, N., and Zheng, X. (2012). Dynamic routing and spectrum assignment in spectrum-flexible transparent optical networks. *Journal of Optical Communications and Networking*, 4(8):603–613.

Zhang, H., Chen, K., Bai, W., Han, D., Tian, C., Wang, H., Guan, H., and Zhang, M. (2015). Guaranteeing deadlines for inter-datacenter transfers. In *Proceedings of the Tenth European Conference on Computer Systems*, EuroSys '15, pages 20:1–20:14, New York, NY, USA. ACM.

Zhao, J., Wymeersch, H., and Agrell, E. (2015). Nonlinear impairment aware resource allocation in elastic optical networks. In *Optical Fiber Communication Conference*, pages M2I–1. Optical Society of America.