

Fatiamento Dinâmico de Redes em Computação em Névoa para Usuários Móveis

Diogo M. Gonçalves¹, Luiz F. Bittencourt¹, Edmundo R. M. Madeira¹

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Campinas – São Paulo – Brasil

diogomg@lrc.ic.unicamp.br, {bit, edmundo}@ic.unicamp.br

Abstract. *Fog Computing environments provide computing resources at the edge of the network. Based on the resource virtualization, Virtual Networks can be created on-demand over one shared physical infrastructure to serve the users. In such a context, a performance analysis of these networks on different scenarios is required to identify the strengths and weaknesses of that technology. This work fills that gap by analyzing different resource allocation approaches for network slicing, considering user' mobility support. Simulations made on MobFogSim show that static resource allocation may present performance issues due to variability of resource demand over time. Dynamic resource allocation is shown as a possible solution to that scenario. However, that approach is sensitive to a computing overhead.*

Resumo. *A Computação em Névoa é responsável por prover recursos computacionais na borda da rede a usuários com diferentes características e demandas. Por meio da virtualização de seus recursos, é possível criar múltiplas Redes Virtuais, ou Fatias de Rede, sobre uma mesma arquitetura física, cada uma atendendo um grupo de usuários. Nesse contexto, avaliar o desempenho de tais redes em diferentes cenários se torna primordial para identificar pontos fortes e fracos a serem considerados no desenvolvimento de mecanismos para gerenciar a rede. Este artigo apresenta uma análise de desempenho de diferentes abordagens de alocação de Fatias de Rede com o objetivo de otimizar o processo de migração de serviços na névoa. Resultados obtidos no simulador MobFogSim apontam que, devido a variações de demanda, o desempenho da alocação estática dessas redes pode se degradar ao longo do tempo. A alocação dinâmica de fatias de rede se apresentou como uma solução para esse cenário, a depender do custo computacional necessário para a reconfiguração dessas redes.*

1. Introdução

Com a crescente popularização de dispositivos conectados a infraestruturas de rede, além dos tradicionais celulares e computadores, novas demandas começaram a ser requisitadas por essas infraestruturas. Esses dispositivos, tais como veículos, aparelhos domésticos, máquinas agrícolas e da indústria, pertencentes a chamada Internet das Coisas, formam grupos de usuários com requisitos de rede e computação distintos entre si. Nesse contexto, as redes 5G propõem atender tais demandas não somente provendo uma infraestrutura para transmissão de dados como também um ambiente para interligar aplicações e serviços a serem utilizados pelos usuários.

Além dos recursos utilizados para a transmissão de dados, os recursos de armazenamento e processamento complementam a arquitetura das redes 5G para prover suporte

a execução desses serviços. O paradigma de Computação em Nuvem provê um gerenciamento flexível desses recursos físicos por meio da virtualização. Apesar dos benefícios introduzidos pela Computação em Nuvem, como flexibilidade e escalabilidade, essa arquitetura apresenta, em contrapartida, algumas limitações como alta latência e aumento de tráfego no núcleo da rede. A Computação em Névoa visa amenizar parte dessas limitações trazendo, de forma distribuída, parte desses recursos para a borda da rede. Nessa arquitetura em Névoa, *cloudlets* ou mini servidores, são utilizados como os nós da rede a fim oferecer recursos de armazenamento e processamento próximos aos usuários. Sendo arquiteturas complementares, Nuvem e Névoa atuam em domínios diferentes da rede.

Além disso, um dos principais diferenciais das redes modernas, como as redes 5G, tem sido a virtualização não somente dos recursos de armazenamento e processamento, mas também a virtualização dos recursos de rede. Nesse cenário, todos os recursos que permeiam a rede são virtualizados. Funções de Rede, que tradicionalmente eram fixadas ao *hardware*, agora podem ser virtualizadas e assim, instanciadas sob demanda em diferentes nós das redes. A virtualização desses recursos permitiu a criação das várias redes virtuais (RVs), também conhecidas como fatias de redes, sobre uma infraestrutura física em comum. Cada Fatia de Rede pode ser construída a fim de atender a demanda de um grupo de usuários em específico (priorizar latência, largura de banda ou processamento, instanciar funções de rede específicas, ou ainda selecionar uma topologia de rede personalizada). A Figura 1 ilustra esse conceito, apresentando duas Fatias de Rede, cada uma atendendo usuários distintos, sobre uma mesma infraestrutura física.

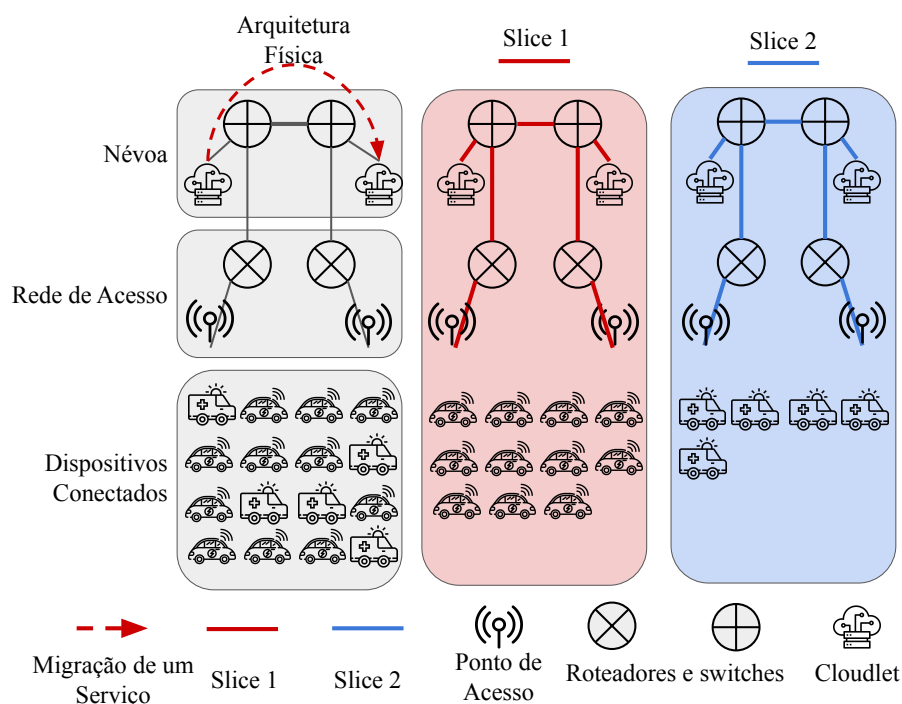


Figura 1. Criação de duas Fatias de Rede sobre uma arquitetura de Computação em Névoa com suporte a migração de Serviços. Adaptado de [Gonçalves et al. 2020].

Apesar da flexibilidade introduzida pelas Fatias de Rede, a dinamicidade apresentada em diversos cenários de uso, em especial, introduzidas pela mobilidade dos usuários e/ou mudança repentina na demanda por recursos, faz do gerenciamento dos recursos da

rede um desafio a ser estudado. Instanciar e/ou modificar uma Fatia de Rede adiciona custos a rede, tanto em termos de tempo quanto em poder computacional. A presença de tais custos demanda do processo de gerenciamento da rede, decisões mais assertivas para a escolha dos recursos a serem alocados para cada Fatia de Rede.

Nesse cenário, o desenvolvimento de mecanismos para um melhor gerenciamento dos recursos da rede por meio do seu fatiamento se faz necessário. Sendo uma tecnologia introduzida recentemente, seu potencial ainda não foi totalmente explorado pelos trabalhos presentes na literatura. Além disso, tão pouco houve tempo hábil para o desenvolvimento de diferentes ferramentas para desenvolver e validar novos mecanismos para essa tecnologia em variados cenários de uso, em especial, considerando a mobilidade dos usuários. Nesse contexto e considerando o custo para a construção de uma infraestrutura física para validação desses mecanismos, a criação de ferramentas como simuladores de rede tem se mostrado uma alternativa.

Em [Puliafito et al. 2020], apresentamos o MobFogSim, um simulador de rede em código aberto para Computação em Névoa com suporte a mobilidade de usuários e migração de Máquinas Virtuais. Em [Gonçalves et al. 2020], uma extensão ao simulador é apresentada para adicionar suporte a Fatiamento de Rede. Neste atual trabalho, propomos o fatiamento dinâmico de redes como uma solução para melhorar as migrações de serviços entre nós da Névoa requisitadas por usuários móveis. Neste trabalho, padrões realísticos para a mobilidade dos usuários guiam as variações de demanda em diferentes locais do mapa. Nesse contexto, o fatiamento de rede é utilizado para gerenciar recursos de rede (largura de banda e topologia da rede) a fim de distribuir recursos entre diferentes grupos de usuários. Variadas configurações e distribuições desses recursos de rede são consideradas a fim de analisar o seu impacto no processo de migração.

Simulações foram realizadas baseadas em parâmetros reais de oferta e demanda dos recursos de rede, bem como dados realísticos foram utilizados para definir a mobilidade dos usuários. Os resultados sugerem que o fatiamento de rede pode ser uma solução para introduzir prioridade a certos grupos de usuários que demandam recursos especiais, além de permitir uma melhor distribuição dos recursos da rede entre os demais usuários. Porém, este processo, dada a grande dinamicidade deste cenário causada pela mobilidade dos usuários, resulta em gargalos de demanda, o que torna não trivial uma distribuição justa dos recursos ao longo do tempo.

O restante deste artigo está organizado da seguinte forma: A Seção 2 faz uma revisão bibliográfica acerca dos trabalhos relacionados mais relevantes da literatura. A Seção 3 descreve a abordagem de fatiamento de redes utilizada neste trabalho, bem como o cenário proposto para avaliação. A Seção 4 apresenta a metodologia utilizada para a avaliação do cenário proposto, bem como as configurações e parâmetros utilizados nas simulações. Por fim, a Seção 5 apresenta os resultados obtidos por meio das simulações. As conclusões, trabalhos futuros e comentários finais são descritos na Seção 6.

2. Trabalhos Relacionados

Sendo uma tecnologia inovadora e promissora, o fatiamento de redes tem sido estudado em diversos contextos nos últimos anos. Cada trabalho propõe e avalia o uso dessa tecnologia com finalidades diferentes. A pesquisa apresentada neste artigo visa complementar esses trabalhos já produzidos ao apresentar o fatiamento dinâmico como uma abordagem para melhorar o processo de migração de serviços em Névoa para usuários móveis. Alguns dos trabalhos relacionados ao tema deste artigo são apresentados a seguir.

Tabela 1. Escopo dos trabalhos presentes na literatura

Trabalho	Suporte a mobilidade	Migração de Serviços	Objetivo
[Raza et al. 2018]	-	-	Melhorar a alocação de Fatias de Rede
[Zhang et al. 2017]	✓	-	Maximizar a capacidade de transmissão dos links
[Xiong et al. 2019]	✓	-	Minimizar tempo de migração e melhorar latência
[Addad et al. 2020]	✓	✓*	Minimizar tempo de migração
Este trabalho	✓	✓	Minimizar tempo de migração e melhorar latência

Os autores em [Raza et al. 2018] apresentam uma comparação entre as políticas de alocação estática e dinâmica, propondo duas soluções para a alocação dos recursos, sendo uma baseada em Programação Linear Inteira e outra em uma heurística. O objetivo das soluções propostas é aumentar o número de requisições atendidas. Apesar do artigo avaliar as políticas de alocação estática e dinâmica, a demanda por recursos é definida de forma aleatória. Tão pouco, a mobilidade dos usuários foi considerada na construção das requisições avaliadas.

Visando desenvolver um novo mecanismo de alocação de recursos para usuários móveis, o trabalho apresentado em [Zhang et al. 2017] propõe uma arquitetura lógica para 5G utilizando fatiamento de rede. Apesar de também considerar a mobilidade dos usuários nos cenários avaliados, os autores tinham como objetivo otimizar a alocação de recursos na transmissão *wireless* dos sistemas 5G. O trabalho apresentado neste artigo, apesar de utilizar uma interface *wireless* na conexão entre o usuário e arquitetura, foca na gerencia dos recursos posicionados na chamada rede de transporte, responsável por conectar os diferentes nós presentes na rede. Apesar do processo de *handover* (troca de Ponto de Acesso) também contribuir para fatores ligados a qualidade de serviço, como latência, confiabilidade e disponibilidade, o fatiamento de rede possui potencial para melhorar o desempenho das transmissões entre as *cloudlets*, fator esse não explorado no trabalho citado.

O trabalho apresentado em [Xiong et al. 2019] propõe um novo mecanismo para fatiamento de rede em uma arquitetura de Névoa veicular. Nessa arquitetura, não somente *cloudlets* podem compor os nós da rede, mas também outros dispositivos como os próprios veículos dos usuários. O trabalho visa, a partir da solução proposta, contornar as limitações presentes na arquitetura, como a variação na demanda e a indisponibilidade dos nós. Apesar da mobilidade dos usuários também ser fator importante no trabalho citado, os autores utilizaram outra arquitetura e diferentes objetivos no escopo do trabalho. A migração de Serviços também não é avaliada.

Os autores do trabalho apresentado em [Addad et al. 2020] propõem, o que eles denominam, Mobilidade de Fatias de Rede. Assim como neste trabalho, os autores ava-

liam o impacto da mobilidade de usuários no gerenciamento de serviços executados na Névoa sobre uma Fatia de Rede. Entretanto, a proposta dos autores é realizar migrações, não somente dos serviços presentes nas *cloudlets*, mas também de todo os agentes que compõe de uma Fatia de Rede. *Diferentemente da proposta deste artigo em que uma Fatia de Rede é utilizada como mecanismo para otimizar a migração de serviços, no trabalho citado, a Fatia de Rede em si é o objeto a ser migrado. No trabalho citado, todo o *layout* de uma Fatia de Rede (Topologia lógica, nós e links) é replicado em outra rede física e desfeito na infraestrutura anterior. Tal abordagem foge do escopo deste trabalho.

Como visto nos artigos citados acima, não há trabalhos presentes na literatura que proponham o uso do fatiamento dinâmico de redes para melhorar o processo de migração de Serviços considerando uma mobilidade realística dos usuários. Neste trabalho, o impacto das variações de demandas causadas pela mobilidade dos usuários foi analisado ao se utilizar diferentes políticas de alocação de recursos para as Fatias de Rede. Uma descrição dessas políticas é apresentada na Seção 3.

3. Políticas de alocação de recursos avaliadas

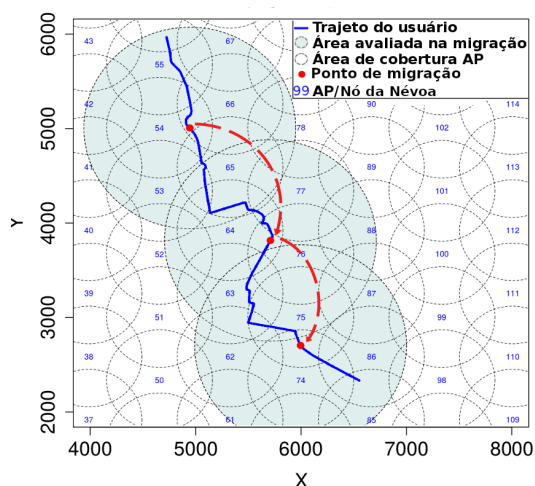
O Fatiamento de Rede tem se apresentado como uma solução para gerenciar grupos heterogêneos de usuários. Nesse cenário, cada Fatia poderá receber uma fração diferente dos recursos da infraestrutura de acordo com suas necessidades. Em muitos casos, a variação ao longo do tempo na demanda dos usuários pode comprometer a capacidade da Fatia em atender seus usuários utilizando os mesmos recursos pré-alocados. Nesse contexto, diferentes políticas de alocação e realocação de recursos podem ser adotadas. Dentre essas políticas, pode-se adotar uma estratégia de alocação estática ou dinâmica dos recursos.

Na política de alocação estática dos recursos da rede, no momento da criação de uma Fatia de Rede, tais recursos são reservados para ela até o final do seu ciclo de vida. Uma vez que a Fatia de Rede finaliza seu ciclo de vida, esses recursos são devolvidos para a infraestrutura, podendo ser disponibilizados para a criação de outras Fatias de Rede. Nesse cenário, os recursos de rede, armazenamento e processamento alocados para uma Fatia de Rede ficam garantidamente disponíveis para seu uso imediato.

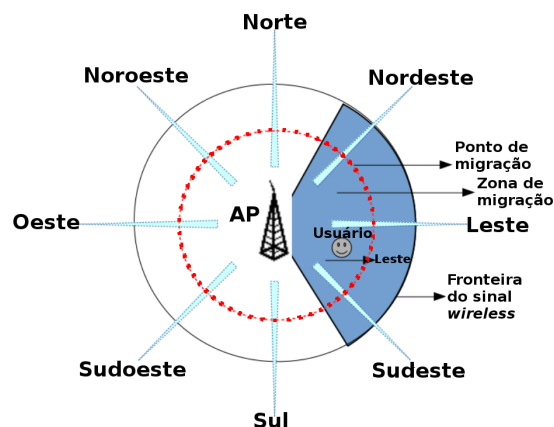
Em contrapartida, uma infraestrutura que suporte uma política de alocação dinâmica de recursos permite que alterações sejam feitas ao longo do tempo a fim de melhor atender a demanda dinâmica da rede. Dentro da alocação dinâmica, diversos critérios podem ser adotados para uma realocação (taxa de utilização máxima e mínima, presença de recursos ociosos, *ranking* de prioridades).

Considerando a demanda da rede entre as *cloudlets* sendo formada por requisições de migrações de serviços, sempre que esse processo de migração se iniciar, a rede poderá reavaliar a necessidade de requisitar mais recursos. O Algoritmo 1 apresenta, em alto nível, os passos realizados no processo de migração de serviços nessa infraestrutura de Névoa, bem como o momento da realocação de recursos. Neste trabalho, o escopo analisado para a realocação de recursos foi restrito a recursos de rede, como a largura de banda. No entanto, outros recursos, como o armazenamento e processamento das *cloudlets*, podem ser avaliados no contexto de Fatiamento de Redes. A Figura 1 ilustra a arquitetura em Névoa considerada neste trabalho. A Figura 2(a) ilustra as migrações de um serviço ao longo do trajeto de um usuário enquanto a Figura 2(b) ilustra a relação entre a posição do usuário e os mecanismos implementados no simulador para iniciar uma migração.

No Algoritmo 1, dadas as Fatias de Rede já construídas, as métricas consideradas para a migração, e alguns dados sobre o usuário, tais como localidade, requisitos e



(a) Processo de migração entre os nós da Névoa a partir da mobilidade do usuário. Fonte: Adaptado de [Gonçalves et al. 2018]



(b) Relação entre posição do usuário sob a cobertura do ponto de acesso e o processo de migração. Adaptado de [Lopes et al. 2017]

Figura 2. Processo de migração de Serviços considerado neste estudo

cloudlet na qual ele está conectado, a infraestrutura passa a avaliar a necessidade de se realizar uma migração (linha 2). Dado o conjunto de *cloudlets* disponíveis e o critério de migração adotado (menor latência, maior disponibilidade, maior confiabilidade, dentre outros), o procedimento *seleciona_proxima_cloudlet* seleciona a *cloudlet* a receber a aplicação do usuário. Caso nenhuma *cloudlet* apresente um desempenho melhor que a atual, a aplicação do usuário se mantém no mesmo servidor, caso contrário a migração será realizada. Neste trabalho, define-se a zona de migração, ilustrada na Figura 2(b), como uma subárea dentro da área de cobertura do ponto de acesso na qual o usuário está conectado. Enquanto o usuário estiver sob a zona de migração (linha 1), o algoritmo verifica a disponibilidade de uma nova *cloudlet* oferecer melhores condições de hospedagem ao Serviço acessado pelo usuários (linha 2). Tais condições, definidas como *critério_migracao* podem considerar latência, proximidade física com o usuário ou quaisquer outros critérios. Uma vez definida a migração, esta acontecerá (linha 5) quando o usuário cruzar pelo, como definimos, ponto de migração (linhas 3 e 4), ilustrado na Figura 2(b). Este ponto delimita a fronteira da zona de migração. Esta fronteira é posicionada há alguns metros da interseção entre a área de cobertura entre dois pontos de acesso, no caso, o ponto de acesso atual e o seu vizinho. A implementação das políticas de realocação das Fatias de Rede ocorre nas linhas 6 e 7. Na alocação estática, as operações das linhas 6 e 7 são ausentes, uma vez que não há modificações na distribuição dos recursos uma vez que eles forem alocados. A análise e requisição de uma realocação, presente na linha 7, é realizada durante todo o processo de migração (linha 6). Dadas as configurações atuais das Fatias de Rede criadas, bem como as duas *cloudlets* envolvidas na migração, o procedimento *solicita_realocacao_banda* consulta a disponibilidade de recursos ociosos nas Fatias de Rede e, em caso afirmativo, os realoca para a Fatia de Rede em que o usuário está associado. A conexão entre as duas *cloudlets* presentes na Fatia de Rede passa então a usufruir desses recursos adicionais. Vale ressaltar, que o cenário adotado neste trabalho realiza apenas o empréstimo de recursos ociosos e não os devolve proativamente após o uso. Essa devolução ocorre apenas quando solicitado e apenas se este estiver novamente ocioso.

Algoritmo 1: Processo de migração de serviços e realocação dinâmica de recursos das fatias de rede.

```
input : redes_virtuais.cloudlets, redes_virtuais.links, usuario,
        usuario.cloudlet_atual, criterio_migracao
1 while usuário estiver na zona de migração do
2   usuario.cloudlet_destino =
   seleciona_proxima_cloudlet(redes_virtuais.cloudlets, usuario,
   criterio_migracao);
3   while usuario nao estiver no ponto de migracao do
4     └ espere;
5   inicia_migracao(usuario, usuario.cloudlet_atual, cloudlet_destino);
6   while usuario realiza migracao do
7     └ solicita_realocacao_banda(redes_virtuais.links, usuario,
   usuario.cloudlet_atual, cloudlet_destino);
```

Trabalhos anteriores [Gonçalves et al. 2020] avaliaram o desempenho dessas duas políticas em um ambiente de Computação em Névoa considerando diferentes demandas por migrações de serviços. Nesse estudo, pode-se observar a ocorrência de recursos ociosos ao se adotar a política de alocação estática. A política de alocação dinâmica, por sua vez, permitiu uma redistribuição desses recursos ociosos, permitindo um melhor desempenho geral da rede. No entanto, não se avaliou os efeitos dessas políticas considerando variações de demanda ao longo do tempo causadas pela mobilidade realística dos usuários. O trabalho citado utilizava uma mobilidade restrita (direção, velocidade e distância constantes), além de uma infraestrutura com escala reduzida (dois pontos de acesso e duas *cloudlets*). A proposta desse trabalho visa complementar a avaliação dessas políticas ao considerar cenários mais complexos na avaliação. Detalhes da metodologia adotada neste trabalho são apresentados na Seção 4.

4. Metodologia

Os cenários avaliados neste trabalho foram construídos com base em parâmetros realistas compatíveis com casos de uso de uma infraestrutura em Névoa. A validação dos cenários foi realizada no simulador MobFogSim. Uma descrição aprofundada sobre as características do simulador pode ser encontrada na Subseção 4.1. Detalhes sobre os cenários avaliados bem como os parâmetros utilizados neste trabalho são descritos na Subseção 4.2.

4.1. MobFogSim

MobFogSim [Puliafito et al. 2020] é um simulador para gerencia de recursos em Névoa com suporte a mobilidade dos usuários. Assim como as versões em que ele se baseia, iFogSim [Gupta et al. 2017] e MyIFogSim [Lopes et al. 2017], o MobFogSim também é disponibilizado como *software* livre¹. O IFogSim permite modelar dispositivos IoT sem mobilidade, Pontos de Acesso, e servidores de Nuvem e Névoa. MobFogSim estende o iFogSim ao suportar a mobilidade dos usuários, implementando mecanismos de gerência de recursos.

Dentre os mecanismos suportados, encontra-se a migração de serviços entre as *cloudlets*. Dada a mobilidade do usuário, sua aplicação que está sendo executada na

¹See <https://github.com/diogomg/MobFogSim>. Acessado em 07/04/2021

Névoa pode ser movida estrategicamente a fim de melhorar algum parâmetro de Qualidade de Serviço. O simulador simula duas técnicas de migração, completa e ao vivo, e dois tipos de virtualização, Máquinas Virtuais e Contêineres. Os critérios para migração podem ser baseados em redução da latência ou distância. Porém, outros critérios podem ser implementados.

Em [Gonçalves et al. 2020] foi apresentada uma extensão ao MobFogSim a fim de prover suporte ao fatiamento dinâmico de redes. Os recursos de processamento e armazenamento no simulador são modelados por meio de máquinas virtuais. Os recursos de rede são expressos por meio da largura de banda disponível nas conexões. Nesse primeiro momento, a implementação e avaliação do fatiamento de rede se restringiu aos recursos de rede. As Fatias de Rede modeladas no simulador podem compartilhar a mesma infraestrutura física. Entretanto, cada Fatia de Rede tem acesso a apenas uma fração desses recursos. O simulador possui suporte para uma realocação de recursos entre as Fatias de Rede. No entanto, o critério avaliado até o momento se restringe a realocação de recursos ociosos. Outras estratégias de realocação podem ser implementadas.

A fim de oferecer um suporte a mobilidades realística, o MobFogSim possui integração com base de dados interpretáveis por outros simuladores de mobilidade, como Simulation of Urban MObility (SUMO) [Behrisch et al. 2011]. Utilizando saídas do SUMO como entrada para o MobFogSim, é possível avaliar a relação entre a mobilidade desses usuários com os impactos da rede.

4.2. Cenários avaliados

Neste trabalho, a avaliação do fatiamento de rede é realizado principalmente sobre a óptica de duas políticas de alocação diferentes, o Fatiamento Estático e Fatiamento Dinâmico. No estático, uma vez alocado o recurso para uma Fatia de Rede, este permanece até o fim da simulação. O fatiamento dinâmico, por sua vez, permite a realocação de recursos ociosos entre as Fatias de Rede. Nesse trabalho, duas Fatias de Rede foram criadas, denominadas Slice 1 e Slice 2, cada uma com diferentes distribuições de oferta e demanda de recursos. Três diferentes combinações foram avaliadas para a distribuição dos recursos físicos entre as Fatias de Rede (10%/90%, 30%/70%, 50%/50%). Além disso, para cada um dos cenários de distribuição de recursos, três diferentes níveis de demanda foram considerados. O Slice 1 foi avaliado servindo 10%, 20% e 40% dos usuários totais da simulação. Para o Slice 2, 90%, 80% e 60% dos usuários foram alocados nas simulações.

Os valores utilizados como parâmetros para oferta e demanda de recursos foram baseados em infraestruturas reais compatíveis com o cenário avaliado [Puliafito et al. 2020]. Cada máquina virtual, responsável por executar a aplicação do usuário, possui 128MB de RAM a ser migrada. A infraestrutura utilizada nessa avaliação é composta por 16 *cloudlets* uniformemente distribuídas em uma mapa 5kmx5km. Outros 64 Pontos de Acesso realizam a conexão sem fio entre os usuários e a infraestrutura. Uma conexão de 74 Mbps com 3.47 ms de latência realiza a conexão entre as *cloudlets*. O critério utilizado para realizar as migrações se baseia na escolha da *cloudlet* que ofereça a menor latência. A Tabela 2 resume os valores utilizados como parâmetros para definir a oferta e demanda de recursos utilizados neste trabalho. As métricas de processamento são apresentadas em MIPS (Milhões de Instruções Por Segundo).

Para a avaliação, 2400 usuários com mobilidades diferentes foram utilizados nas simulações. Esses usuários foram divididos aleatoriamente em 30 grupos de 80 usuários. Cada simulação utiliza um grupo de 80 usuários sob os cenários descritos anteriormente.

Tabela 2. Parâmetros e seus valores utilizados nas simulações. Valores baseados na infraestrutura real apresentada em [Puliafito et al. 2020]

Parâmetro	Valor
Capacidade de processamento do cliente	2901MIPS
Tamanho dos pacotes de requisição	87B
Tamanho dos pacotes de resposta	54B
Uso de armazenamento do cliente	4MB
Uso de armazenamento do servidor	412MB
Capacidade de processamento do cliente	46534MIPS
Capacidade de processamento do servidor	3234MIPS
Latência entre cliente e servidor via Ponto de Acesso	4.78ms
Latência média entre os servidores	3.47ms
Demanda de RAM no cliente	49MB
Demanda de RAM no servidor	128MB
Largura de banda entre cliente e servidor	13640kbps
Largura de banda entre os servidores	74148kbps

A duração de cada simulação é restrita a 60 minutos. A mobilidade de cada um dos 2400 usuários foi baseado nos dados de veículos de Luxemburgo disponibilizados pela base de dados LustScenario [Codeca et al. 2015]. O desempenho das Fatias de Rede foi avaliado em termos de latência, número de migrações realizadas e duração de cada migração. Os resultados são apresentados na Seção 5 utilizando a média entre as 30 simulações com um intervalo de confiança de 95%.

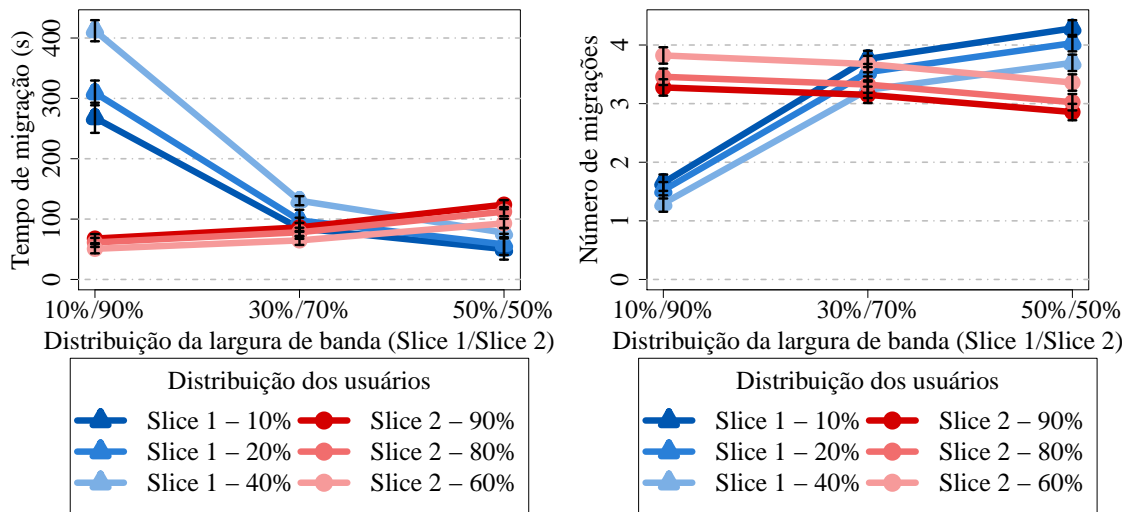
5. Resultados

A partir dos cenários descritos anteriormente e das simulações realizadas, pode-se identificar o impacto de diferentes políticas de alocação de recursos no desempenho da rede. No primeiro cenário, a Subseção 5.1 apresenta os resultados da política de alocação estática de recursos para as fatias de rede. A Subseção 5.2 complementa o cenário anterior avaliando o comportamento de uma política de alocação dinâmica de recursos, ou seja, adaptável ao longo do tempo à medida que os parâmetros para alocação mudem, como por exemplo, demanda dos usuários. Os resultados são apresentados a seguir.

5.1. Alocação estática de recursos

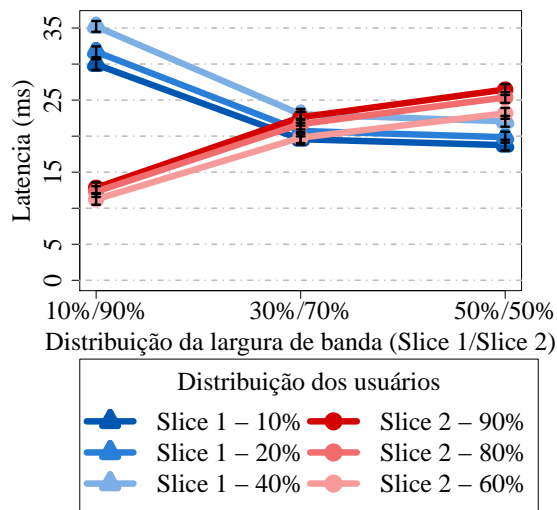
Uma das principais métricas dependentes da disponibilidade de recursos diz respeito ao tempo necessário para a realização da migração de um serviço entre os servidores. A Figura 3(a) apresenta a duração média da realização desse processo de migração. Neste cenário, como esperado, ambos os fatores, recursos disponíveis e demanda dos usuários, interferiram no desempenho do processo de migração. No entanto, a disponibilidade de recursos apresentou um maior impacto.

Em um cenário com uma distribuição igualitária da largura de banda entre as Fatias de Rede (50% para cada), as Fatias com mais usuários apresentaram um pior desempenho. Uma Fatia de Rede atendendo 90% dos usuários provê a eles migrações cerca de 240% mais lentas do que o desempenho observado ao servir 10% dos usuários. Para cenários com desequilíbrio na distribuição dos recursos avaliados, a queda de desempenho na duração das migrações pode chegar a 400% ao comparar as Fatia de Rede com 10% e 90% dos recursos totais disponíveis.



(a) Tempo de duração de cada migração, em média, por usuário

(b) Número de migrações por usuário



(c) Latência média obtida por cada usuário

Figura 3. Desempenho da rede utilizando fatiamento estático

Outro ponto observado nas simulações foi a não linearidade no ganho de desempenho esperado pelo acréscimo de recursos. Este comportamento pode ser observado ao comparar o desempenho de uma Fatia de Rede com 90% dos recursos servindo 90% dos usuários e uma Fatia de Rede com 10% dos recursos servindo 10% dos usuários. Neste cenário, apesar da demanda total ser proporcional a quantidade de recursos disponíveis, cada área do mapa pode receber demandas diferentes. Nesse contexto, usuários em zonas menos congestionadas se beneficiam mais de Fatia de Rede com mais recursos disponíveis. Um usuário alocado a uma Fatia com acesso a 10% dos recursos, mesmo em casos em que ele não encontra concorrência na sua região, ele estará limitado a acessar somente 10% da capacidade da rede. Outro usuário nas mesmas condições, mas podendo acessar 90% dos recursos da rede, irá obter um desempenho muito superior.

O tempo necessário para a realização de uma migração afeta diretamente outras

métricas da rede. Considerando a mobilidade dos usuários sendo restrita por um trajeto finito, o número total de migrações realizadas ao longo da viagem do usuário é limitado pela soma do tempo necessário para a execução de cada uma delas. A Figura 3(b) apresenta o número de migrações realizadas, em média, por cada veículo durante as simulações. Nota-se uma grande correlação inversamente proporcional entre as duas métricas, tempo necessário para realizar uma migração e número de migrações realizadas. Assim como apresentado na Figura 3(a), as Fatias de Rede com mais recursos apresentaram um número maior de migrações, assim como nos cenários com menos usuários demandando recursos.

O resultado da duração das migrações associado ao número de migrações que a infraestrutura consegue prover ao usuário ao longo do seu trajeto resultam na proximidade que a aplicação consegue ser reposicionada à medida que o usuário se move pelo mapa. Como apresentado na Figura 3(c), as Fatias de Rede que realizaram mais migrações, proveram uma latência reduzida para seus usuários. Quanto mais migrações a infraestrutura é capaz de realizar, mais rápida a aplicação pode ser reposicionada, impedindo que a latência aumente à medida que o usuário se distancia da aplicação. Como visto nas figuras anteriores, essa capacidade de migração está fortemente associada ao acesso a maiores larguras de banda disponíveis para a realização desse procedimento. Na Figura, as Fatias de Rede com menos recursos apresentaram uma latência, em média, até 350% maior. Em um cenário de distribuição igualitária de recursos (50% para cada Fatia de Rede), a variação na demanda dos usuários resultou em uma diferença de até 40%.

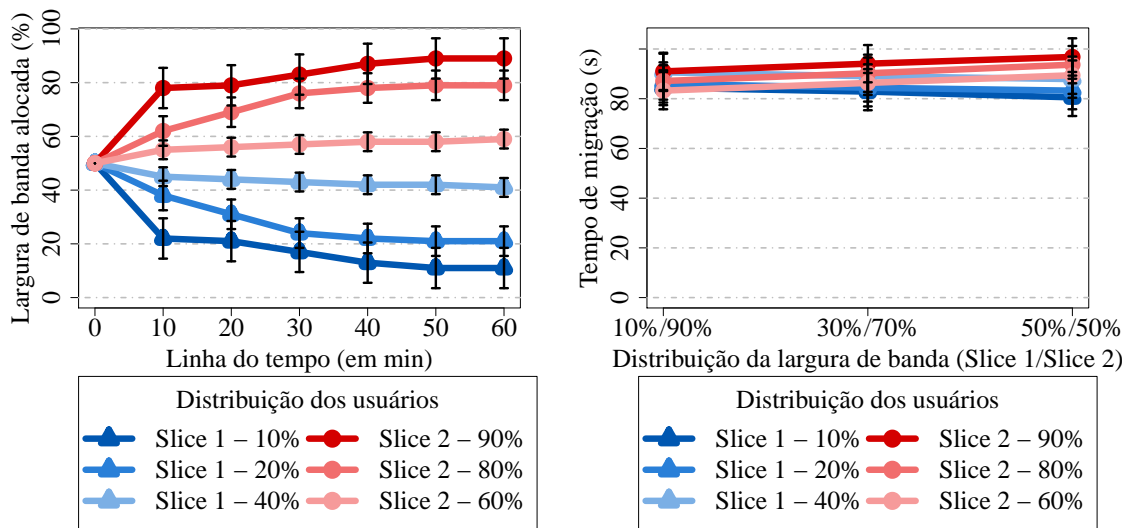
O fatiamento de rede permitiu a distribuição dos recursos da rede entre diferentes Fatias de Rede, cada uma podendo atender diferentes demandas. A possibilidade de alocar recursos de forma desproporcional entre as Fatias de Rede pode ser considerada como um comportamento desejado ou um efeito colateral, a depender do contexto. Como apresentado na motivação deste trabalho, certas aplicações necessitam de uma disponibilidade maior de recursos. Fatias de Rede que recebem mais recursos introduzem privilégios ou prioridades aos seus usuários se comparadas às demais Fatias de Rede, ao custo de prejudicar o desempenho dos demais usuários. Nesse contexto, fatores dinâmicos como a mobilidade dos usuários, podem agravar essa desproporção, aumentando ou diminuindo a demanda em determinadas regiões. A política de alocação estática de recursos, avaliada neste primeiro cenário, não permite rápidas correções dessas configurações da rede. Uma alternativa para esse cenário é a alocação dinâmica de recursos, na qual os recursos atribuídos às Fatias de Rede podem ser realocados ao longo do tempo a fim de melhor servir a demanda dos usuários.

5.2. Alocação Dinâmica de Recursos

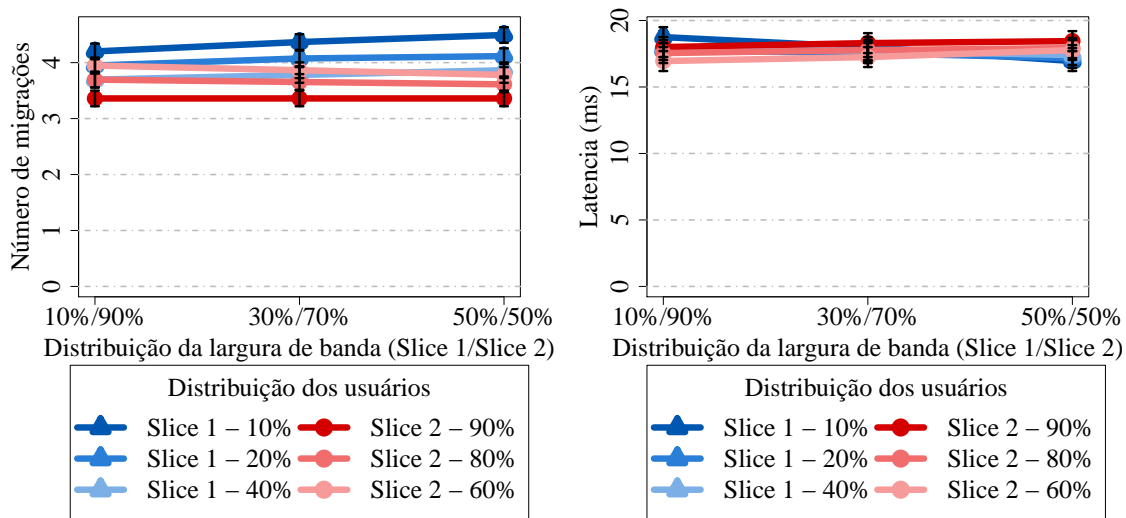
A partir dos cenários de avaliação do fatiamento estático da rede, novos experimentos foram realizados considerando a capacidade da rede em redistribuir seus recursos à medida em que ocorressem mudanças na demanda de cada Fatia de Rede. Nesse novo cenário de avaliação, recursos ociosos em uma Fatia de Rede podem ser realocados para a outra fatia de rede se esta assim demandar. O Algoritmo 1 apresenta a execução desse processo, em especial, nas linhas 6 e 7.

A fim de ilustrar o comportamento da rede a partir desse cenário de realocações, a Figura 4(a) apresenta o percentual da largura de banda disponível no canal alocado para cada uma das Fatias de Rede ao longo do tempo. No experimento ilustrado na Figura 4(a), cada Fatia de Rede recebeu metade da largura de banda disponível na infraestrutura física. Nesse cenário avaliado, a medida em que os usuários interagem com a infraestrutura e, conseqüentemente, requisitavam recursos para realizar as migrações de seus serviços

entre os nós da Névoa, a demanda das Fatias de Rede foi se alterando de acordo com a quantidade de usuários servidos por aquela Fatia. Tendo cada Fatia de Rede iniciado o experimento com a mesma disponibilidade de recursos (50%), as fatias que serviam mais usuários, conseqüentemente, apresentaram maior demanda. Diferentemente do cenário estático, essas Fatias de Rede requisitaram recursos das fatias mais ociosas. Na Figura, pode-se observar que em todos os cenários, essa política de realocação resultou em uma redistribuição dos recursos que tende a ser proporcional a demanda de cada Fatia de Rede.



(a) Largura de banda alocada para cada Fatia de Rede ao longo do tempo (b) Tempo de duração de cada migração, em média, por usuário



(c) Número de migrações por usuário (d) Latência média obtida por cada usuário

Figura 4. Desempenho da rede utilizando fatiamento dinâmico

Dada a política de realocação adotada neste trabalho na qual o empréstimo de recursos só ocorre quando estes estiverem ociosos, redes com maior demanda tendem a ser mais suscetíveis a requisitar empréstimos do que a devolvê-los. Esse comportamento, apesar de apresentar uma redistribuição mais igualitária dos recursos, remove a prioridade introduzida a determinadas Fatias de Rede. Esse fator prioritário de algumas Fatias pode

ser desejado em alguns cenários, como apresentado na motivação deste trabalho.

O resultado dessa distribuição proporcional dos recursos utilizando a política adotada neste trabalho, apresentada na Figura 4(a), impacta outras métricas utilizadas para avaliar o desempenho da rede. Tendo em vista que cada Fatia de Rede passou a receber recursos proporcionais a sua demanda, o tempo de duração de cada migração deixou de ser destoante entre cada Fatia. A Figura 4(b) ilustra o desempenho de cada Fatia de Rede baseado nessa métrica. Nesse novo experimento, a diferença de desempenho entre as Fatias de Rede não ultrapassou a marca de 30%, diferentemente dos cerca de 400% observados no experimento anterior. Pode-se observar também a melhora significativa de desempenho das Fatias de Rede que previamente receberam menos recursos, se comparadas com o cenário de alocação estático. No cenário anterior, tais redes estavam limitadas ao baixo percentual alocado a elas. No cenário dinâmico, as redes menos favorecidas puderam acessar mais recursos em regiões com menos concorrências. Em regiões mais densas, a presença de recursos ociosos é mais rara. Como consequência da duração das migrações igualitária, o número de migrações realizadas, apresentada na Figura 4(c), bem como a latência média observada para cada usuário, apresentada na Figura 4(d), também mostraram desempenho semelhantes.

6. Conclusão

O fatiamento de redes permite criar redes virtuais sob demanda para servir usuários com diferentes requisitos. Dada a dinamicidade desse cenário, introduzida principalmente pela mobilidade dos usuários, a rede deve ser capaz de se reconfigurar de forma eficiente à medida que a demanda por recursos se modifica ao longo do tempo. Este trabalho apresentou o fatiamento dinâmico como uma tecnologia para auxiliar o processo de migração de Serviços em Névoa. Neste trabalho foram avaliadas duas políticas de alocação de recursos para Fatias de Rede considerando vários cenários de oferta e demanda.

Simulações apontaram que o fatiamento de rede pode introduzir prioridade de acesso a recursos para determinadas Fatias ao custo de precarizar o desempenho das demais. No entanto, considerando a política de alocação estática, uma porção dos recursos alocados se mostrou ociosa durante a simulação, podendo ter sido utilizada por Fatias de Rede menos favorecidas. Na política de alocação dinâmica, as Fatias podem ser reconfiguradas a fim de suprir variações de demanda ao longo do tempo. Nesse cenário, notou-se uma redistribuição proporcional à demanda dos usuários, melhorando o desempenho geral da rede. No entanto, considerando a metodologia adotada neste trabalho, esse cenário removeu a prioridade de acesso a recursos introduzida a determinadas Fatias de Rede.

Apesar da política de realocação se mostrar mais promissora para gerenciar os recursos da infraestrutura de rede nos cenários avaliados, cada realocação demanda custos, tanto em termos computacionais, para calcular e realizar tais modificações, quanto em tempo consumido para realizar tal operação. A depender do custo despendido para realizar tal procedimento, os benefícios podem não ser recompensados. Uma avaliação considerando tais fatores pode ser necessária para a adoção dessa abordagem.

A partir dos resultados obtidos deste trabalho, alguns tópicos de pesquisa poderão ser estudados. Dentre os trabalhos futuros, espera-se desenvolver novos mecanismos para o gerenciamento dos recursos virtuais da rede, o que inclui políticas de realocação de recursos não ociosos, e a avaliação desses cenários considerando outras métricas. Um estudo acerca do impacto do custo da realocação dos recursos no desempenho da rede também deverá ser desenvolvido.

Agradecimentos

Este trabalho é parte do INCT sobre Internet do Futuro para Cidades Inteligentes (CNPq 465446/2014-0, CAPES 88887.136422/2017-00 e FAPESP 2014/50937-1). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo nº 2021/05253-0.

Referências

- Addad, R. A., Taleb, T., Flinck, H., Bagaa, M., and Dutra, D. (2020). Network Slice Mobility in Next Generation Mobile Systems: Challenges and Potential Solutions. *IEEE Network*, 34(1):84–93.
- Behrisch, M., Bieker, L., Erdmann, J., and Krajzewicz, D. (2011). SUMO - Simulation of Urban Mobility: An Overview. In *3rd International Conference on Advances in System Simulation (SIMUL)*.
- Codeca, L., Frank, R., and Engel, T. (2015). Luxembourg SUMO Traffic (LuST) Scenario: 24 Hours of Mobility for Vehicular Networking Research. In *IEEE Conference on Vehicular Networking (VNC)*, pages 1–8.
- Gonçalves, D., Puliafito, C., Mingozi, E., Rana, O., Bittencourt, L., and Madeira, E. (2020). Dynamic network slicing in fog computing for mobile users in mobfogsim. In *2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC)*, pages 237–246. IEEE.
- Gonçalves, D. M., Bittencourt, L. F., and Madeira, E. R. M. (2018). Migração proativa de máquinas virtuais para aplicações móveis na computação em névoa. In *Simpósio Brasileiro de Redes de Computadores (SBRC)*, volume 36.
- Gupta, H., Vahid Dastjerdi, A., Ghosh, S. K., and Buyya, R. (2017). iFogSim: A Toolkit for Modeling and Simulation of Resource Management Techniques in the Internet of Things, Edge and Fog Computing Environments. *Software: Practice and Experience*, 47(9):1275–1296.
- Lopes, M. M., Higashino, W. A., Capretz, M. A., and Bittencourt, L. F. (2017). MyiFogSim: A Simulator for Virtual Machine Migration in Fog Computing. In *ACM 6th International Workshop on Clouds and (eScience) Applications Management (CloudAM). Companion Proceedings of the 10th International Conference on Utility and Cloud Computing*, pages 47–52.
- Puliafito, C., Goncalves, D. M., Lopes, M. M., Martins, L. L., Madeira, E., Mingozi, E., Rana, O., and Bittencourt, L. F. (2020). Mobfogsim: Simulation of mobility and migration for fog computing. *Simulation Modelling Practice and Theory*, 101:102062.
- Raza, M. R., Fiorani, M., Rostami, A., Öhlen, P., Wosinska, L., and Monti, P. (2018). Dynamic Slicing Approach for Multi-tenant 5G Transport Networks. *IEEE/OSA Journal of Optical Communications and Networking*, 10(1):A77–A90.
- Xiong, K., Leng, S., Hu, J., Chen, X., and Yang, K. (2019). Smart Network Slicing for Vehicular Fog-RANs. *IEEE Transactions on Vehicular Technology*, 68(4):3075–3085.
- Zhang, H., Liu, N., Chu, X., Long, K., Aghvami, A., and Leung, V. C. (2017). Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges. *IEEE Communications Magazine*, 55(8):138–145.