



Trabalho de Conclusão de Curso

Processos de Handover em Redes Móveis 5G com Redes Definidas por Software

Geymerson dos Santos Ramos
geymerson@ic.ufal.br

Orientador:
Prof. Dr. André Luiz Lins de Aquino

Maceió, 21 de maio de 2019

Geymerson dos Santos Ramos

Processos de Handover em Redes Móveis 5G com Redes Definidas por Software

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação do Instituto de Computação da Universidade Federal de Alagoas.

Orientador:

Prof. Dr. André Luiz Lins de Aquino

Maceió, 21 de maio de 2019

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação do Instituto de Computação da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina.

Prof. Dr. André Luiz Lins de Aquino - Orientador
Instituto de Computação
Universidade Federal de Alagoas

Rian Gabriel Santos Pinheiro - Examinador
Instituto de Computação
Universidade Federal de Alagoas

Ivan César Martins - Examinador
Instituto de Computação
Universidade Federal de Alagoas

Agradecimentos

Filho de pais simples, nascido em uma cidade do interior de Alagoas. Ao final dos meus 16 anos concluí meu ensino médio. Passei meus 17 anos ajudando minha mãe em uma lanchonete que ela mantinha. Aos 18 anos, comecei a trabalhar em uma usina de açúcar da cidade. Fiquei empregado por 8 meses até ser demitido e lembrar que aquilo não era algo que eu queria para minha vida, eu gostava de computadores, lembrei-me que queria ser um cientista da computação.

Ao pesquisar os cursos da Universidade Federal de Alagoas, percebi que havia um curso de engenharia de computação no qual eu poderia passar 1 ano a mais aprendendo, optei pela mudança. Com os estudos já defasados, passei a me preparar para ingressar na universidade estudando praticamente 10 horas por dia em casa, até perceber que estava sem disciplina para as matérias que eu não tinha interesse.

Felizmente, o dinheiro que ganhei trabalhando, guardei. Consegui me manter por 1 ano em Maceió, estudando manhã e tarde em um cursinho. No mesmo período, o Instituto Federal de Alagoas inaugurou um campus na cidade onde morava, fiz o processo seletivo para o curso técnico de segurança no trabalho, achei que ajudaria meus estudos. Fui aprovado, tive que estudar em três horários, manhã e tarde em Maceió, a noite em São Miguel dos Campos. No final do mesmo ano, consegui ingressar na Universidade.

O caminho foi longo, mas atingi meu objetivo, estou em vias de me graduar no curso de engenharia de computação. Transitei entre diferentes fases na universidade. Senti estresse, raiva, tristeza, privação de sono, felicidade, orgulho, satisfação. Tive a oportunidade de conhecer pessoas incríveis e extremamente esforçadas que me inspiraram. Lembro-me do meu primeiro dia de aula, quando fui perguntado o que eu esperava da universidade, a resposta foi instantânea: "Orientação".

Ao acaso um professor me encontrou nos corredores do instituto de matemática, um dos momentos cruciais de minha vida como pessoa e aluno. Fui apresentado ao Laboratório de Computação Científica e Análise Numérica, o LaCCAN, onde passei boa parte dos meus dias. Eu não mais estava restrito às salas de aula, agora tinha acesso à atividades de um laboratório constituído por pessoas extremamente competentes no que faziam. Eu tive o privilégio de participar de discussões e ver não professores, mas os profissionais da computação dispostos

a transmitir conhecimento.

Eu precisava disso, estava perdido. A universidade em primeira instância não se demonstrou diferente do ensino médio. Eu queria ver a engenharia, eu precisava ver as pessoas que me transmitissem o sentimento de que, se eu pudesse, algum dia, eu seria ao menos um pouquinho como elas.

O professor do corredor se tornou meu orientador. André Luiz Lins de Aquino. Para minha sorte, uma pessoa de excelência. Sempre entre os melhores colocados nos ranques de seleção de projetos. Recebi mais do que orientação, recebi conselhos, vi-o tratar as pessoas com humanidade. Coisas que assimilei para mim. Agora ao final do curso, obtive mais ganhos pessoais do que conhecimento técnico, e por isso também sou extremamente grato. Guardarei as experiências boas, para que me guiem no caminho da bondade e da responsabilidade. E guardarei da mesma forma as experiências ruins, para me lembrar do que não devo ser.

Quero encerrar este longo texto expressando minha gratidão. Sou grato aos meus pais, pela paciência e compreensão, por minha criação. Eu certamente não chegaria até aqui sem ajuda e apoio de vocês. Tentarei, mas dificilmente conseguirei retribuir tudo que me foi dado. Sinto-me feliz também por minhas irmãs, por serem alunas de curso superior e técnico de instituições federais. Agradeço aos meus avôs e avós, que infelizmente não presenciarão minha formação. Agradeço aos meus amigos e professores, em especial do LaCCAN, pelas conversas e por fazerem meus dias mais animados. Espero crescer como pessoa de valor para meu círculo social e familiar, que eu tenha forças para ajudá-los quando precisarem.

À minha mãe, pai, irmãs, avós, avôs, orientador, amigos e pessoas que me inspiram, admirá-los-ia por toda eternidade, não fosse minha breve existência.

If you live a pathological life, you pathologize your society, and if enough people do that, then it's hell.

– Peterson, J. B.

Resumo

Este trabalho apresenta uma modelagem matemática para alocação de usuários em torres de redes de telefonia móvel, levando em consideração sua implementação em aplicações que utilizam redes definidas por *software* no cenário 5G. O modelo utiliza menor distância, média de *handover*, largura de banda disponível por torre, quantidade mínima exigida pelos usuários e além disso, modelagens de minimização de *handover* entre áreas de cobertura e menor custo de comunicação entre torres e *data centers*. Através de dados coletados pelo GPS do dispositivo de um usuário, bases de dados com coordenadas de torres e um cenário simulado, foi verificado que em conformidade com o modelo, as alocações consideram menores distâncias, média de *handover* e disponibilidade de largura de banda.

Palavras-chave: Handover, Redes Definidas por Software, Redes 5G.

Abstract

This work presents a mathematical model for the allocation of users in cell towers of mobile networks, considering its implementation with 5G applications that use software-defined networks. The model uses the shortest distance, handover average, tower available bandwidth, user's minimal bandwidth requirements, and models of handover minimization between service areas and cost minimization between towers and data centers. With data collected from a user's GPS, tower coordinates from a database and a simulated scenario, it was verified in the results that the allocations considered the shortest distance, handover average, and network bandwidth availability, as established by our mathematical model.

Keywords: Handover, Software-Defined Networking, 5G networks.

Sumário

Lista de Figuras	vi
Lista de Tabelas	vii
1 Introdução	1
1.1 Motivação	1
1.2 Trabalhos Relacionados	3
1.3 Objetivo	4
1.4 Contribuições	4
1.5 Estrutura do texto	5
2 Tecnologias e Conceitos	6
2.1 Redes 5G	6
2.2 Redes Definidas Por Software	10
2.2.1 Controladores	13
2.3 Handover	14
3 Nossa proposta	16
3.1 Descrição do modelo	16
3.1.1 Modelo proposto	18
4 Resultados e Discussões	21
4.1 Resultados	21
4.1.1 Tempo de execução	21
4.1.2 Alocação de usuários	22
4.1.3 Gerenciamento de largura de banda	23
5 Considerações Finais	26
Referências bibliográficas	27

Lista de Figuras

1.1	<i>Handover</i> de um usuário conectado a torre 1 na área de cobertura 1 para a torre 2.	2
2.1	Projeções para número de dispositivos ou conexões no cenário global por tipo de rede e volume de dados em exabytes por mês.	7
2.2	Arquitetura para redes 5G composta de macrocélula com pequenas células fixas e móveis, comunicação D2D e direta.	8
2.3	Arquitetura de aplicações de redes definidas por <i>software</i>	11
2.4	Arquitetura simplificada de aplicação do OpenFlow.	12
2.5	<i>Handover</i> intra-macrocélula, inter-macrocélula e multi-RAT.	15
3.1	Arquitetura EPS para redes LTE.	17
3.2	Disposição das eNBs ao longo do eixo de posições.	18
4.1	Rota seguida pelo UE ao longo do tempo, partindo do ponto inicial e seguindo no sentido das eNBs de maior numeração.	22
4.2	Alocação do usuário nas eNBs ao longo do tempo.	23
4.3	Torre (eNB _{<i>i</i>}) cujo usuário (UE ₁) está alocado em função de sua posição (x). . . .	24
4.4	Quantidade de usuários alocados e largura de banda disponível em cada eNB. Um usuário está se deslocando e alterando a configuração da rede para cada instante (t).	24

Lista de Tabelas

2.1	Comparativo das diferentes gerações de tecnologia de comunicação móvel sem fio.	7
2.2	Raio de alcance e quantidade de usuários atendidos por tipo de célula.	9
2.3	Lista de controladores <i>open source</i> com suporte ao OpenFlow.	14
3.1	Notação de constantes e variáveis utilizadas na modelagem do problema.	20
4.1	Tempo médio de execução para diferentes combinações de pesos. Número de diferentes instâncias computadas = 30, $N = 60$, $DC = 10$, $U = 10$	22

1

Introdução

1.1 Motivação

A evolução das tecnologias de comunicação impactaram de forma significativa protocolos e soluções de transferência de dados. Com o advento da Internet das Coisas (Al-Fuqaha et al., 2015), os dispositivos estão cada vez mais integrados e as redes sem fio (Akyildiz et al., 2002) se apresentam como recursos extremamente importantes. Usuários com equipamentos como *smartphones* podem se comunicar de praticamente qualquer lugar, usufruir de serviços de localização e trocar informações utilizando tecnologias como Wi-Fi, *Bluetooth*, 3G, 4G e a mais recente proposta 5G (Panwar et al., 2016). A quinta geração promete maiores velocidades, conexão ubíqua, suporte a redes mais densas, heterogêneas e escaláveis. Entre os pilares fundamentais das redes 5G estão:

- Conexão contínua e ininterrupta: usuários poderão permanecer conectados a todo momento e em qualquer lugar;
- Latência zero: será oferecido suporte à aplicações de tempo real, sistemas críticos e serviços com baixa tolerância a atraso, as restrições de latência da 5G são da ordem de 1 milissegundo;
- Conexão de alta velocidade: para suportar aplicações que dependem de latências virtualmente nulas, a velocidade de conexão será da ordem de Gigabit por segundo.

O novo cenário oferece oportunidade de inovações, melhorias de soluções existentes e a utilização de novos paradigmas computacionais, a exemplo das redes definidas por *software*, ou SDN (*Software-Defined Networking*) (Nunes et al., 2014). Empresas de impacto do setor tecnológico como Microsoft (Hong et al., 2013) e Google (Jain et al., 2013) já estudam e utilizam

SDNs em seus *data centers*. No contexto de redes definidas por *software*, existe uma separação entre o plano de dados, que contém dispositivos responsáveis unicamente por encaminhar mensagens, e o plano de controle, concentrando a inteligência de *software* da rede. A separação simplifica e alivia infraestruturas convencionais, que embutem o controle em dispositivos como roteadores, comutadores e *gateways*.

SDNs permitem programabilidade flexível, gerência de recursos, escalabilidade, privacidade, segurança, controle de tráfego, virtualização e realocação de serviços. Desacoplando implementações de *software* do *hardware*, a rede pode ser tratada como uma grande infraestrutura virtual que dispõe virtualização de suas funcionalidades (*Network Functions Virtualization* — NFV) (Han et al., 2015). Uma vez que a tecnologia 5G exigirá arquiteturas e aplicações mais rápidas, eficientes, heterogêneas e escaláveis, as SDNs atuarão como meio viabilizador essencial.

Soluções existentes precisarão ser melhoradas ou repensadas para atuação em latência zero. Os processos de *handover* são alguns dos problemas que podem ser abordados e que estão na lista de recebimento de melhoras. *Handovers* (Yan et al., 2010) ocorrem mediante a transferência de usuários entre pontos de acesso ou áreas de coberturas, como mostra a figura 1.1. Os impactos do processo abrangem experiência de mobilidade, percepção de descontinuidade e comutação entre diferentes tecnologias de comunicação, isto é, *Bluetooth*, 4G, Wi-Fi, etc. Portanto, é interessante que *handovers* ocorram imperceptivelmente, em curtos intervalos de tempo e para regiões que não influenciem negativamente a experiência de mobilidade e conectividade.

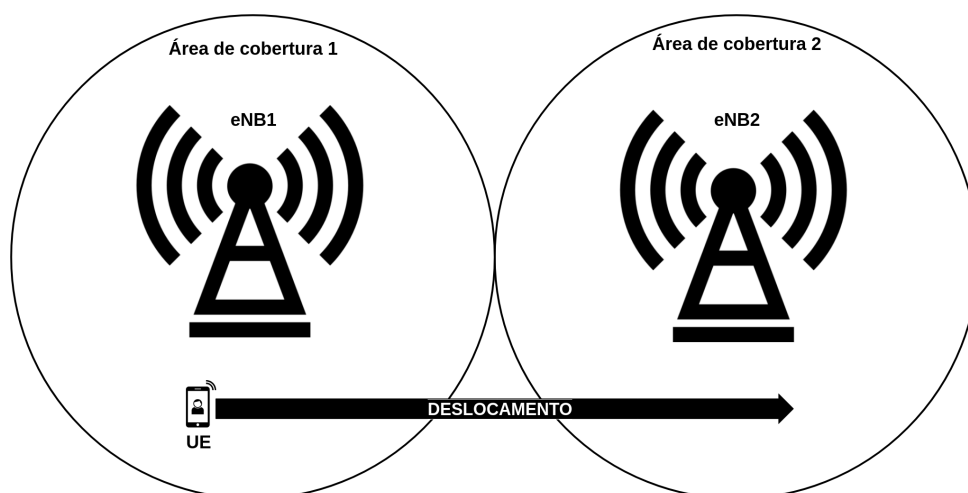


Figura 1.1: *Handover* de um usuário conectado a torre 1 na área de cobertura 1 para a torre 2.

Neste trabalho, apresentaremos uma modelagem matemática para alocação de usuários (*User Equipment* - UE) em torres (*Evolved Node B* - eNB). Será considerada menor distância entre UEs e eNBs, média de *handover*, largura de banda disponível por torre e a quantidade mínima exigida pelos usuários. Um conjunto de dados de GPS de um UE em deslocamento foi utilizado para realocações conforme sua proximidade e distanciamento de diferentes eNBs. O

modelo tem base na proposta em [Taleb et al. \(2015\)](#), que apresenta soluções para minimização de *handovers* entre áreas de cobertura e minimização do custo de comunicação entre eNBs e *data centers* (sem considerar usuários e largura de banda).

1.2 Trabalhos Relacionados

Os autores em [Kuklinski et al. \(2015\)](#) apresentam uma discussão de como aplicar tecnologias de SDN para aumentar a eficiência de redes móveis, levando em consideração o processo de *handover* em três tipos de arquitetura: i) *centralizada*: utiliza apenas um controlador; ii) *semi-centralizada*: múltiplos controladores, atuando em diferentes domínios ou regiões geográficas; iii) *hierárquica*: contém múltiplas camadas de modo que um controlador mestre no topo da hierarquia se conecta a controladores distribuídos em níveis inferiores.

O trabalho de [Prados-Garzon et al. \(2016\)](#) propõe uma implementação de *handover* em uma rede LTE (*Long Term Evolution*) parcialmente virtualizada e baseada em SDNs. Os autores implementaram a troca de mensagens envolvida no processo e simularam os atrasos de transmissão, propagação e processamento da rede, assim como o tempo de preparação e finalização do *handover*.

No contexto de Veículos Aéreos Não Tripulados (VANT), [Sharma et al. \(2017\)](#) disponibilizam uma arquitetura melhorada de SDNs, de modo que VANTs são utilizados sob demanda como comutadores, realizando encaminhamento de mensagens. Com a arquitetura, os autores diminuíram latência de *handover*, atraso de comunicação ponto-a-ponto e *overhead* de mensagens de sinalização.

Um esquema de *handover* em SDNs para 5G considerando informações de mobilidade dos usuários é apresentado em [Lee and Yoo \(2017\)](#). O controlador recebe dados de mobilidade e informações advindas de estações base para selecionar onde os usuários serão alocados (células). Através de simulação, foi verificado que a proposta consegue realizar escolha das células com maior força de sinal, menor carga e de acordo com a direção do movimento do usuário.

Uma solução para um problema multiobjetivo de *handover* em ambientes híbridos 5G é dada em [Qiang et al. \(2016\)](#). Os autores consideraram a maximização da taxa de dados recebidos e minimização da probabilidade de bloqueio (falha) do processo. As chances de bloqueio são calculadas com base em informações limitadas (restrições de privacidade) e no comportamento de outros usuários da rede. No trabalho de [Qiang et al. \(2014\)](#) para SDNs, um controlador é encarregado de realizar o processo utilizando informações privadas.

A proposta em [Duan and Wang \(2015\)](#) preza pela privacidade e mecanismos de autenticação de processos de *handover* em redes heterogêneas 5G com SDN. Para autenticação, são utilizados atributos como localização, direção e características da camada física para gerar identificadores únicos, sem a necessidade de soluções criptográficas e portanto simplificando o

procedimento de autenticação. Para proteção à privacidade, transmissões parciais de dados são realizadas por diferentes caminhos de uma SDN.

Como diferencial às abordagens citadas, a proposta desse trabalho considerará uma modelagem matemática para alocação de usuários em torres da arquitetura de redes móveis *Evolved Packet System*, levando em consideração sua implementação em aplicações que utilizam redes definidas por *software* no cenário 5G. O modelo utiliza menor distância, média de *handover*, largura de banda disponível por torre, quantidade mínima exigida pelos usuários e além disso, modelagens de minimização de *handover* entre áreas de cobertura e menor custo de comunicação entre torres e *data centers*.

1.3 Objetivo

A finalidade deste trabalho é a melhora de experiência de mobilidade em redes 5G através da utilização de redes definidas por *software*. Como proposta, uma modelagem matemática é apresentada para alocação de usuários em torres de telefonia móvel, de modo que transferências (*handovers*) ocorram sempre considerando menor distância, média de *handover*, largura de banda disponível em cada torre e a mínima exigida pelos usuários, que devem ter liberdade para se deslocar e transitar entre áreas de cobertura sem experienciar descontinuidades, perdas abruptas de qualidade de serviço ou deixarem de ser atendidos.

1.4 Contribuições

A contribuição deste trabalho é uma modelagem matemática para alocação de usuários em torres de telefonia móvel. Na modelagem também serão considerados os modelos de minimização de *handovers* entre áreas de cobertura e de minimização de custo de comunicação entre torres e *data centers*. Espera-se portanto que experiência de mobilidade apresente melhorias em termos de:

- Ganho de qualidade de comunicação entre torres de telefonia móvel e usuários, que são alocados sempre considerando menor distância, média de *handover*, e largura de banda para garantia de atendimento;
- Ganho de qualidade de comunicação entre torres e *data centers* pela minimização do custo de comunicação entre estas duas entidades, o que garante respostas em intervalos de tempo menores;
- Redução da quantidade de *handovers* que ocorrem na rede, através da organização das torres em áreas de cobertura maiores.

1.5 Estrutura do texto

A estrutura do texto segue com o Capítulo 2 discutindo sobre tecnologias e conceitos envolvidos no trabalho apresentado. No Capítulo 3 temos a apresentação da nossa proposta e subsequentemente se apresentam os resultados no Capítulo 4. Por fim, o Capítulo 5 contém a conclusão e considerações finais.

2

Tecnologias e Conceitos

2.1 Redes 5G

Até o ano de 2020, é previsto que o número de dispositivos conectados chegue a 50 bilhões (Ericsson, 2011). A quantidade já vem gradativamente aumentando e em função do agressivo crescimento no volume de dados, novas vertentes computacionais também surgiram, a exemplo da Computação na Nuvem (Mell and Grance, 2010) e *Big Data* (Chen et al., 2014), reinventando formas de armazenamento e obtenção de informação. As tecnologias de comunicação sem fio, que deveriam suportar aplicações no novo panorama, se encontraram em um cenário inóspito frente à redes demasiadamente densas, heterogêneas e com altas exigências de conectividade.

É mediante este cenário que a quinta geração se apresenta para telecomunicação móvel. As redes 5G abrirão caminho para implementação de paradigmas computacionais que viabilizam expansão das tecnologias de informação e comunicação (TICS), fundamentalmente importantes para Cidades Inteligentes (Mohanty et al., 2016). Se comparada à quarta geração, algumas melhorias a serem elencadas são: até 100 vezes mais dispositivos conectados; 1000 vezes mais volume em dados de dispositivos móveis; 100 vezes mais veloz; latência de 1 milissegundo; disponibilidade e cobertura de aproximadamente 100%; processamento de informações em tempo real e menor consumo energético (GSMA Intelligence, 2014). Portanto, a quinta geração vem oferecer conexão ubíqua, escalabilidade, heterogeneidade, e além disso, fatores que impactam significativamente na redução de custos operacionais de infraestrutura.

A tabela 2.1 contém um comparativo das gerações de tecnologia móvel, da 1ª à 5ª. Os primeiros serviços de telefonia ofereceram majoritariamente vias de comunicação para voz com sinais analógicos. As soluções de *handover* eram precárias e horizontais, isto é, transferências entre pontos de acesso não ocorriam entre tecnologias de comunicação sem fio diferentes (por exemplo, 1G para 2G). Além da baixa velocidade, a primeira geração apresentava graves problemas de segurança, visto que as conversas eram reproduzidas nas torres, o que consequentemente as tornavam suscetíveis ao acesso por terceiros (Vora, 2015). A tecnologia 2G

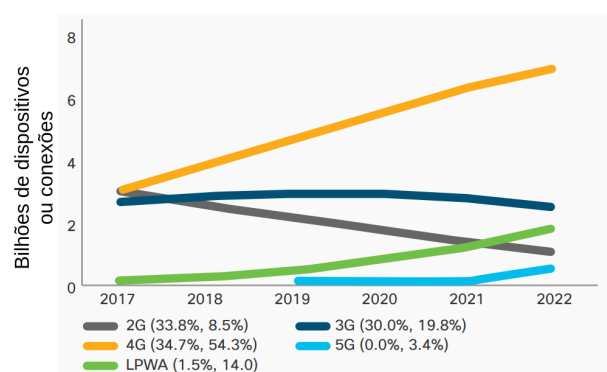
superou a geração anterior em velocidade, adoção de sinais digitais, envio de mensagens de texto e imagens. Sua obsolescência se tornou clara conforme a Internet e a disseminação de dados multimídia ganhavam popularidade, iniciando a era 3G.

Geração	Características	Velocidade	Latência (ms)	Handover	Limitações
1G	Sinais analógicos, mensagens de voz	< 2.4 Kbps	-	Horizontal	Baixa segurança
2G	Sinais digitais, mensagens de voz, texto e imagens	< 64 Kbps	< 1000	Horizontal	Pouco suporte a serviços de internet
3G	Mensagens de voz, acesso à Internet por telefonia fixa e móvel, vídeo chamadas	< 3.1 Mbps	< 500	Horizontal e Vertical	Baixa velocidade
4G	Maiores velocidades de conexão	< 300 Mbps	< 100	Horizontal e Vertical	Baixa escalabilidade e conectividade, suporte a redes densas
5G	Conexão ubíqua e de alta velocidade, escalabilidade, heterogeneidade	> 1 Gbps	< 1	Horizontal e Vertical	-

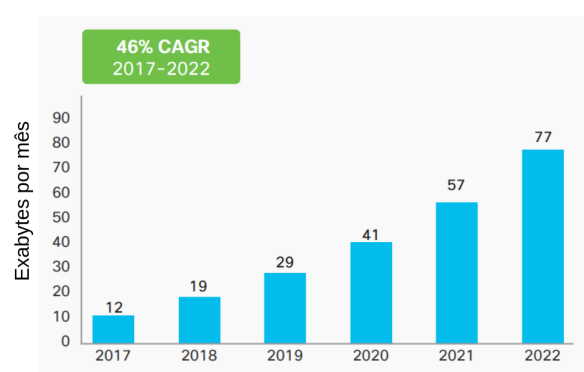
Tabela 2.1: Comparativo das diferentes gerações de tecnologia de comunicação móvel sem fio.

No período 3G, as conexões chegaram a velocidades da ordem de Megabit por segundo, e já era possível realizar acesso à Internet por telefonia fixa ou móvel. Os serviços de *streaming* multimídia tiveram grande aumento em demanda. Com relação aos processos de *handover*, além de horizontal, também passaram a ser utilizadas arquiteturas verticais, possibilitando transferências entre pontos de acesso com tecnologia de comunicação sem fio diferentes (3G para 2G, por exemplo). As demandas subsequentes incidiram sobre a velocidade da rede, que como resultado impulsionou o desenvolvimento da 4G. Desde então, a quarta geração vem se tornando o principal meio de acesso, com limite teórico de até 300 Mbps.

De acordo com as previsões globais de tráfego de dados de dispositivos móveis da CISCO (CISCO, 2019), figura 2.1(a), desde 2017 com uma porção de 34.7%, o número de usuários da rede 4G já supera a geração anterior. As projeções aumentam este número para 54.3% até o ano de 2022. O gráfico também inclui conexões LPWA (*Low-Power Wide-Area*), que representam aplicações de baixa largura de banda e podem cobrir áreas extensas, com uso essencial em comunicação M2M (*Machine-to-Machine*) não encobertas por redes de telefonia móvel, a exemplo de postes de iluminação inteligentes com comunicação sem fio.



(a) Conexões globais por tipo de rede.



(b) Projeção de crescimento de dados de dispositivos móveis.

Figura 2.1: Projeções para número de dispositivos ou conexões no cenário global por tipo de rede e volume de dados em exabytes por mês.

A figura 2.1(b) exibe a taxa de crescimento anual composta (CAGR — *Compound Annual Growth Rate*) do tráfego global de dados de dispositivos móveis em exabytes por mês, entre os anos de 2017 e 2018, e as projeções até 2022. O crescimento será de 46% nessa janela de tempo, e atualmente o volume já é grande o suficiente para pressionar a infraestrutura das redes 4G, já em vias de substituição pela tecnologia sucessora.

A quinta geração não é simplesmente um aumento de velocidade em relação a sua antecessora. Trata-se de uma redefinição em termos de aplicações, utilização eficiente de recursos, modelos e arquiteturas para dar suporte a carga de dados e dispositivos diversos com diferentes tecnologias de comunicação à rádio. Deverão ser contemplados arcabouços heterogêneos com redes definidas por *software*, virtualização de funções de rede, comunicação à rádio *full duplex* e D2D (*Device-to-Device*) (Panwar et al., 2016). A figura 2.2 ilustra um exemplo da pluralidade que arquiteturas da rede 5G podem atingir. Esta mesma figura representa também uma arquitetura multi-camadas.

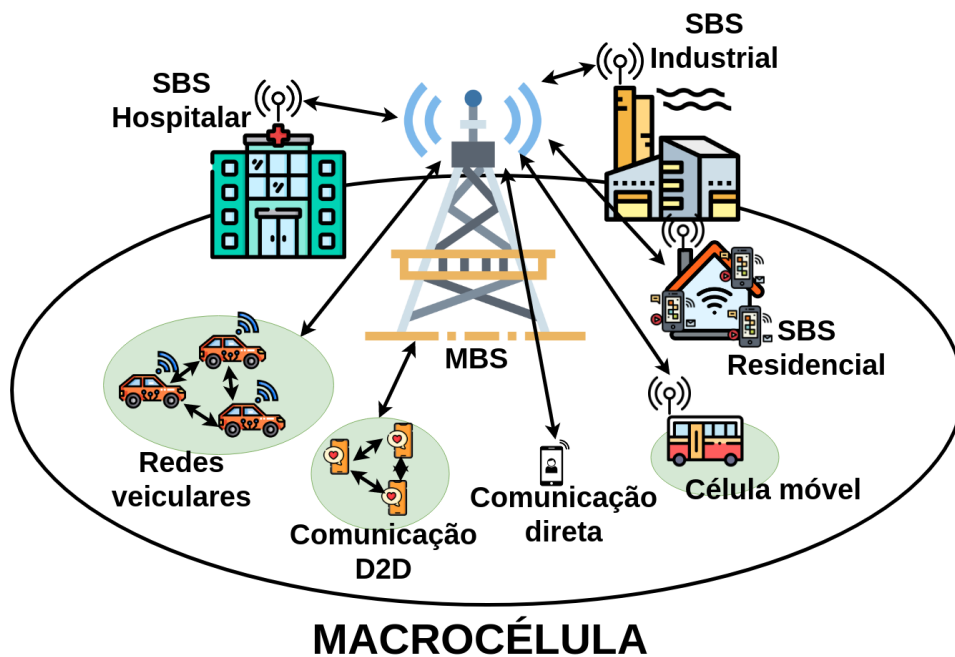


Figura 2.2: Arquitetura para redes 5G composta de macrocélula com pequenas células fixas e móveis, comunicação D2D e direta.

Nesta arquitetura, uma macrocélula com uma macro estação base (*Macro Base Station* — MBS) na camada superior recebe requisições de serviços das camadas inferiores, que estão divididas em pequenas células com pequenas estações bases (*Small Base Station* — SBS) para diferentes perfis de aplicação. No caso de uma pequena célula móvel, por exemplo um ônibus, os ocupantes do mesmo têm acesso à rede externa através de uma SBS acoplada no ônibus, que por fim se comunica com a MBS da macrocélula. Dispositivos móveis como *smartphones* podem realizar comunicação direta com a MBS ou constituírem pequenas células dinâmicas de comunicação D2D, nas quais apenas alguns destes dispositivos fazem conexão com a MBS

enquanto os outros fazem retransmissão de requisições e respostas.

SBSs podem ser instaladas em ambientes diversos como residências, hospitais e indústrias, constituindo naturezas diferentes de aplicações e pequenas células. A depender do alcance e número de usuários, células podem ser distinguidas como pode ser observado na tabela 2.2 adaptada de [Panwar et al. \(2016\)](#).

	Femtocélula	Pico-célula	Microcélula	Macrocélula
Alcance	10 - 20 m	200 m	2 km	30 - 35 km
Usuários	< 20	20 - 40	> 100	Muitos

Tabela 2.2: Raio de alcance e quantidade de usuários atendidos por tipo de célula.

A distinção e uso de pequenas células contribuem com aumento do alcance da macrocélula e também com:

- Maiores taxas de transferência: Dispositivos *indoor* (em ambientes internos) podem ser melhor atendidos por SBS locais que realizam a comunicação externa com a MBS, pois isso diminui efeitos de atenuação que existiriam na comunicação direta entre dispositivos e MBS;
- Melhor uso do espectro de comunicação a rádio: Pois há uma quantidade menor de dispositivos em comunicação direta com a MBS;
- Economia energética: A comunicação intermediada por uma SBS diminui o raio de alcance necessário para transmissão de dados, o que impacta significativamente no uso de bateria dos dispositivos;
- Redução de custo: Uma SBS tem menores custos operacionais e de instalação se comparada a uma MBS.

Entre os problemas a serem pontuados pela quinta geração, é explicitável o uso otimizado de recursos computacionais e energéticos. Nas atuais redes de telefonia móvel, estações bases são projetadas para funcionar em horários de pico, e seu processamento é destinado apenas aos usuários conectados. O consumo energético para horários de pico ou de quase inatividade são semelhantes ([Correia et al., 2010](#)), e isso aumenta os custos de operação da rede.

Durante o dia, estações bases em áreas comerciais ficam sobrecarregadas, enquanto áreas residenciais apresentam pouca atividade ([Wu et al., 2015](#)). Os cenários se invertem para diferentes momentos do dia. Um maior aproveitamento e balanceamento de carga poderia ser alcançado se o poder computacional ocioso fosse distribuído, ao invés de ter uso restrito aos usuários conectados.

Como demanda de uso eficiente de frequência de banda, também se busca solução para o uso redundante de dois canais (operando em frequências diferentes) para comunicação com as

estações base. Os dispositivos possuem um canal levando dados (*uplink*) até a estação, e outro no sentido contrário (*downlink*). Esta prática é considerada ineficiente, e portanto, espera-se que redes 5G apresentem comunicação *full-duplex* com único canal para envio e recebimento de dados, sem que haja co-interferência entre os processos.

Até então, as arquiteturas das redes de telefonia móvel não haviam sido projetadas para fazer distinção entre usuário *indoor* e *outdoor* (em ambientes externos). A comunicação é realizada diretamente com as estações base, a despeito de localização, e isso implica em graves atenuações e perda de qualidade de sinal se o usuário estiver cercado por paredes ou algum outro obstáculo. Como a quinta geração introduz uma SBS como intermédio entre usuários e MBS, a distinção torna-se prática, e conseqüentemente espera-se ganho de qualidade.

Em termos de redes sem fio heterogêneas (Damjanovic et al., 2011), compostas por tecnologias como *Bluetooth*, 3G, Wi-Fi, etc., o arcabouço de soluções da 5G deve implementar contribuições significativas. Já existe certo suporte na 4G, mas sua arquitetura não foi desenvolvida inicialmente para tal. Atualmente, é permitido apenas que um dispositivo tenha os canais de *uplink* e *downlink* associados a mesma estação base.

A quinta geração deve implementar um desacoplamento no qual usuários podem ter os canais associados a estações bases diferentes. Segundo Boccardi et al. (2016), o desacoplamento pode trazer melhorias de redução de potência de transmissão, redução de interferência, maiores taxas de transmissão, e além disso, redução de custos e tratamento distinto de balanceamento de carga para *uplink* e *downlink*.

A experiência de mobilidade e processos de *handover* também exigirão atenção, uma vez que transferências entre diferentes tecnologias de comunicação sem fio devem se tornar mais comuns. No deslocamento entre áreas comerciais e residenciais por exemplo, as sessões devem ser mantidas, sem que haja percepção de descontinuidade ou queda de taxas de transferência.

O conjunto de soluções da 5G precisa garantir os serviços dos usuários mesmo em cenários de alta mobilidade, com velocidades de até 500 km/h (Zhang et al., 2017). Para uma quantidade crescente de dispositivos e redes cada vez mais densas, heterogêneas e com exigências de escalabilidade e baixa latência, as redes definidas por *software* estão se demonstrando de uso fundamental pela quinta geração.

2.2 Redes Definidas Por Software

Como forma de inovação, SDNs vêm para mudar uma atual integração vertical das redes tradicionais de computadores. Decisões de como gerir e lidar com o tráfego e o encaminhamento dos dados conforme estas decisões eram ações aglutinadas em elementos como roteadores, comutadores e *gateways*, mas no contexto de SDNs, tomadas de decisão e encaminhamento passam a ser divididas entre plano de dados e de controle.

A divisão visa principalmente a flexibilização da programabilidade e reconfiguração frente à falhas diversas e variações abruptas de tráfego, tendo em vista os mecanismos de resposta e auto-reconfiguração praticamente inexistentes das redes IP (Kreutz et al., 2015). Nas SDNs, o plano de controle é detentor da inteligência computacional, logicamente centralizada em um controlador.

As decisões de encaminhamento são realizadas no controlador e transmitidas para comutadores (*switches*) programáveis, que simplesmente executam o encaminhamento. Os dispositivos de encaminhamento constituem o plano de dados e não estão sujeitos aos encargos computacionais das tomadas de decisões.

Uma vez que há uma centralização lógica em um elemento controlador, visão e controle global da rede (da perspectiva do controlador) são alguns dos benefícios que podem ser mencionados. É importante enfatizar que a centralização é lógica e não física. Isso permite que controladores robustos sejam desenvolvidos em sistemas distribuídos (Jain et al., 2013), garantindo confiança em termos de escalabilidade, desempenho, tolerância à falhas, etc. A figura 2.3 exibe a arquitetura de aplicações SDN envolvendo plano de dados, plano de controle, camada de aplicações e as interfaces norte e sul.

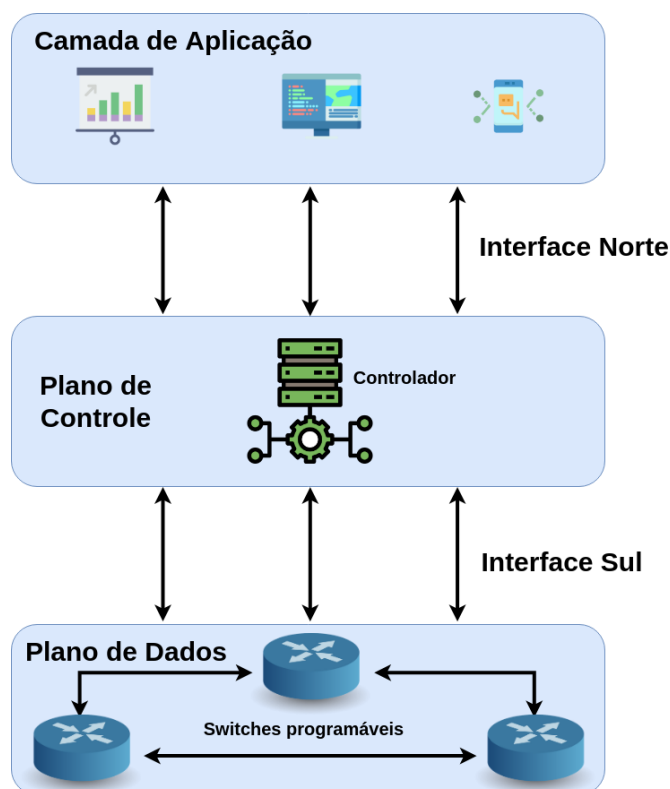


Figura 2.3: Arquitetura de aplicações de redes definidas por *software*.

A interface norte, ou *northbound interface*, oferece um arcabouço de desenvolvimento para aplicações que consomem e monitoram serviços da rede, abstraindo complexidades de comunicação entre plano de controle e a camada de aplicação. Uma vez que esta interface é inteiramente definida por *software* e as aplicações são extremamente diversas, não existem

padronizações ou protocolos amplamente adotados para interações entre plano de controle e camada de aplicação (Nunes et al., 2014). Os desenvolvedores possuem liberdade de implementação e tipicamente apresentam soluções ad hoc.

A interface sul, também conhecida como *southbound interface*, define o conjunto de instruções entre plano de controle e elementos no plano de dados, responsáveis pelos encaminhamentos. Nesta interface, determinadas funções do plano de dados são definidas de acordo com o *hardware* dos elementos no plano, e conseqüentemente é possível encontrar protocolos como OpenFlow, Open vSwitch Database (OVSDB), OpFlex da Cisco e muitos outros.

O OpenFlow faz papel da interface sul, e portanto liga os dispositivos do plano de dados ao controlador. Uma arquitetura simplificada da interação entre controlador, o protocolo e o *switch* é exibida na figura 2.4. Para encaminhamento de dados, os *switches* possuem tabelas que armazenam conjuntos de entradas de fluxo e cada entrada contém campos de compatibilidade, contadores, e um conjunto definido de ações a serem aplicadas aos pacotes de dados que apresentarem compatibilidade com um dado fluxo (ONF, 2015).

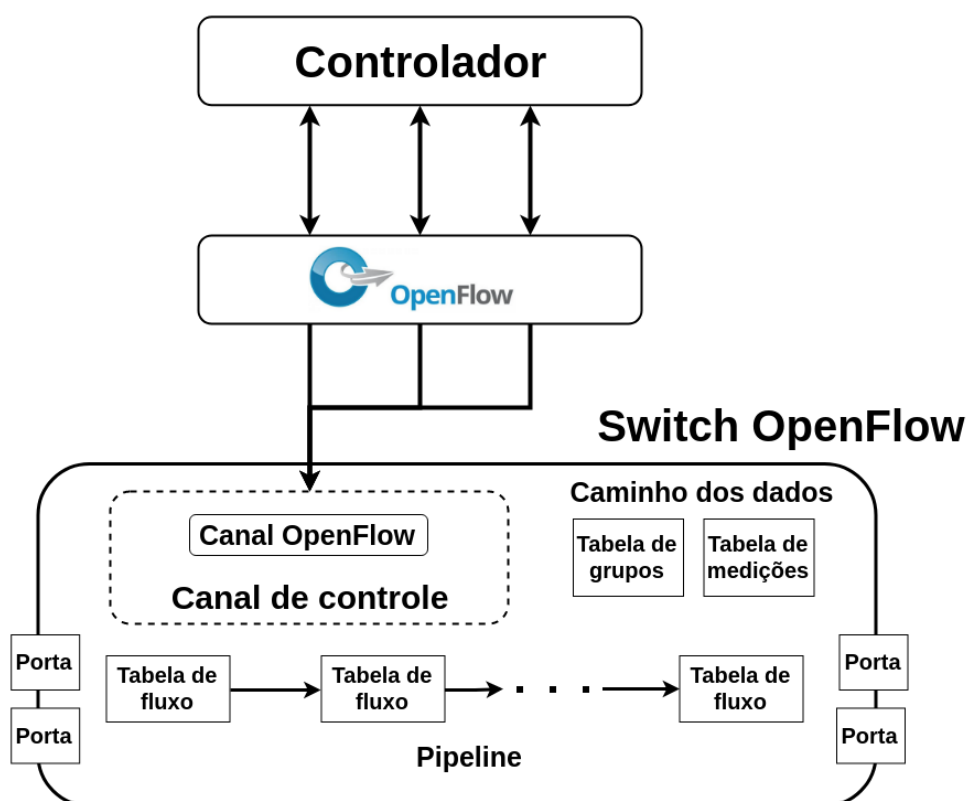


Figura 2.4: Arquitetura simplificada de aplicação do OpenFlow.

Descrição dos componentes do Switch OpenFlow:

- Grupo: Lista de ações a serem aplicadas em um pacote de dados qualquer;
- Tabela de grupos: Armazena conjuntos de grupos;
- Tabela de fluxo: Elemento parte do *pipeline* que armazena entradas de fluxo;

- Tabela de medições: Contém entradas de medição que permitem a implementação de limitadores de largura de banda como métrica de qualidade de serviço;
- Porta: Ponto de entrada ou saída de pacotes de dados, seja esta física, lógica, ou reservada pelo protocolo;
- Caminho dos dados: Conjunto dos componentes envolvidos no encaminhamento e processamento, isto é, tabelas de fluxo, tabela de grupos e as portas;
- Canal de controle: Via de comunicação e gerenciamento entre o *switch* e o controlador;
- *Pipeline*: Cadeia de tabelas de fluxo para verificação de compatibilidade, encaminhamento e modificação de pacotes.

Discutindo brevemente sobre o OVSDB e o OpFlex, observa-se que é possível encontrar protocolos com características muito semelhantes ou distintas do OpenFlow. O OVSDB por exemplo é um complemento do OpenFlow, de modo que seus elementos de controle podem criar múltiplas instâncias de *switches* virtuais, ajustar políticas de qualidade de serviço, gerenciar filas e coletar estatísticas.

Por outro lado, o OpFlex mantém as funções de controle nos dispositivos de encaminhamento pela premissa de que o controlador pode se tornar o gargalo das aplicações SDN. O que se tem centralizado em seus controladores são as políticas de controle da rede. As políticas são formas de controle para assegurar que os usuários não utilizem mais largura de banda que a permitida, e acessem apenas endereços e serviços permitidos.

2.2.1 Controladores

Até o presente momento, as redes tradicionais de computadores eram programadas com instruções de baixo nível ou sistemas operacionais proprietários para redes (Kreutz et al., 2015). Muitos dos problemas resolvidos recebiam soluções ad hoc que tratavam das eventuais complicações de baixo nível que pudessem surgir.

Como abstração às complicações de programação de baixo nível, as SDNs introduzem os controladores, também considerados sistemas operacionais para redes em virtude do desengargalo gerado para os desenvolvedores com relação ao tratamento de atividades dos subníveis das aplicações de redes. O número de controladores vem crescendo desde então, visando oferecer mais desempenho, segurança e escalabilidade. A tabela 2.3 contém uma lista de controladores *open source* com suporte ao OpenFlow.

Muitos desses controladores se encontram sob escrutínio da academia e da indústria de tecnologia. No trabalho de Shalimov et al. (2013) uma avaliação é feita para NOX, POX, Beacon, Floodlight, MuL, Maestro e Ryu. Nem todo controlador consegue atender exigências mínimas para atuar em um ambiente de produção, e os autores mostram que dentre os avaliados, o Beacon apresentou maior escalabilidade com taxas de transferência de aproximadamente 7 bilhões

Controlador	Linguagem de Programação
POX	Python
NOX	Python/C++
MUL	C
Maestro	Java
Trema	Ruby/C
Beacon	Java
Jaxon	Java
Floodlight	Java
Ryu	Python
NodeFlow	JavaScript
ovs-controller	C
Flowvisor	C
RouteFlow	C++

Tabela 2.3: Lista de controladores *open source* com suporte ao OpenFlow.

de fluxos por segundo, com possibilidade de crescimento proporcional ao número de núcleos disponíveis ao controlador. POX e Ryu não suportam *multi-threading*, e portanto não são alternativas escaláveis. Os resultados para métricas como latência, segurança e confiabilidade podem ser verificados no trabalho dos autores.

Recorrentemente surgem discussões sobre controle distribuído ou centralizado, mas é fácil perceber que a centralização em um único elemento da rede aumenta as chances de colapso total do sistema, além de também não ser uma solução escalável. Logo, é admissível uma centralização lógica, mas com a instanciação física do controlador em diversas máquinas, como por exemplo em um *cluster* centralizado, o que aumentaria consideravelmente as taxas de transferência de fluxos, ou em conjunto de máquinas geograficamente espaçadas, aumentando a tolerância do sistema à falhas.

Os controladores podem também ter natureza reativa ou proativa. No primeiro caso, os dispositivos de encaminhamento de dados precisam obrigatoriamente fazer consultas ao controle para cada tomada de decisão. No modo proativo, o controlador passa para os dispositivos apenas as políticas de controle da rede.

2.3 Handover

Como já explicado, o *handover* se dá pela transferência de um usuário conectado a um ponto de acesso para outro. De modo geral, o processo pode ser dividido em 3 fases: preparação, execução e finalização. A preparação abrange a análise de uma série de informações do usuário, como posição, serviços consumidos, provável rota futura, pontos de acesso vizinhos e a qualidade dos serviços disponíveis nestes.

As informações permitem que os melhores pontos de acesso sejam selecionados, e para tal,

a sofisticação das técnicas pode abranger até mesmo algoritmos de aprendizado de máquina (Al et al., 2016). Na fase de execução são trocadas mensagens de sincronização e reconhecimento entre os elementos envolvidos no processo, isto é, o usuário, os pontos de acesso e as entidades superiores que coordenam o processo de maneira geral.

No contexto de uma rede de telefonia móvel sem fio com SDN, em um cenário hipotético o controlador poderia solicitar e receber informações de usuários e, quando exigido, tomar a decisão de *handover*, cabendo aos dispositivos apenas encaminhar as mensagens do processo para as respectivas entidades envolvidas. Quando o usuário já está devidamente sincronizado em seu novo ponto de acesso, temos a efetivação da transferência e fase de finalização.

Os tipos de *handover* também podem variar. Nas redes 5G é possível ter transferências entre duas SBS pertencentes à mesma macrocélula, entre macrocélulas diferentes e até mesmo para um outro tipo de tecnologia de comunicação à rádio, como exhibe a figura 2.5. Se as estações bases estão na mesma macrocélula, o *handover* é intra-macrocélula, como acontece para a transferência entre SBS1 e SBS2. Na transferência de SBS2 para SBS3, o dispositivo está migrando de célula e portanto sofrendo um *handover* inter-macrocélula. Quando o dispositivo desconecta das torres da rede 5G e é realocado em um outro tipo de tecnologia de comunicação, como por exemplo a 4G, temos o *handover* multi-RAT (*Radio Access Technology*).

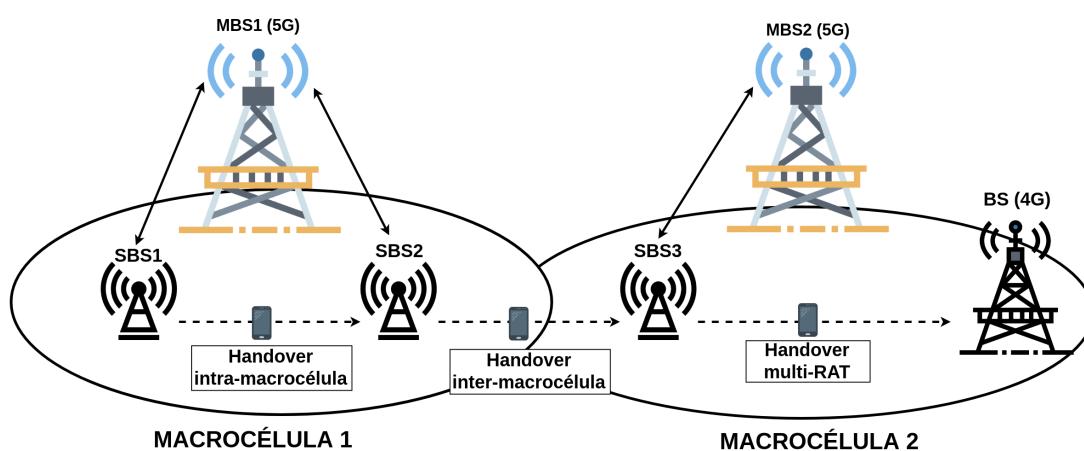


Figura 2.5: *Handover* intra-macrocélula, inter-macrocélula e multi-RAT.

Transferências intra-macrocélulas são relativamente menos custosas quando comparadas a inter-macrocélulas, uma vez que no primeiro caso apenas as SBS estão envolvidas no processo. A nível inter-macrocélula, não só as SBSs precisam ser analisadas, as MBSs também estão envolvidas e isso consequentemente aumenta a complexidade do *handover*, que pode se tornar ainda mais custoso para transferências multi-RAT.

3

Nossa proposta

Neste capítulo apresentaremos a modelagem matemática para a alocação de usuários em torres de telefonia móvel. A partir daqui, conforme a literatura, a terminologia utilizada será eNB (*Evolved Node B*) para torres e UE (*User Equipment*) para usuários ou dispositivos da rede.

3.1 Descrição do modelo

Conforme o modelo multiobjetivo de [Taleb et al. \(2015\)](#), a proposta visa aprimorar a experiência de conexão e mobilidade em redes móveis, descritas pela arquitetura EPS (*Evolved Packet System*), como ilustra a figura 3.1. O elemento S-GW (*Serving Gateway*) é responsável por rotear pacotes, gerenciar mobilidade e processos de *handover* no plano do usuário (entre eNBs ou outras redes). HSS (*Home Subscriber Server*) armazena informações de identificação e endereçamento dos usuários UEs, contém também funcionalidades de autenticação e autorização. O PDN-GW (*Packet Data Network Gateway*) atua como intermediário para redes externas, como por exemplo serviços de nuvem de terceiros ou a Internet propriamente dita, neste caso representada como uma nuvem constituída de diferentes *data centers*. Na MME (*Mobile Management Entity*), estão englobadas funcionalidades de gerência de mobilidade dos UEs e retenção de informação de localização, interagindo com a HSS, também realiza autenticação. As torres, ou eNBs, fazem ponte entre UEs e os demais componentes da rede.

No trabalho de [Taleb et al. \(2015\)](#), é realizada uma modelagem matemática para minimização do custo de comunicação entre eNBs e a fonte de seus dados, neste caso os *data centers*. Também é minimizada a quantidade de *handovers* que os usuários na rede são submetidos (problema multiobjetivo). No contexto de SDNs, considera-se aqui as infraestruturas virtuais de rede, nas quais componentes como S-GW e PDN-GW possuem conjuntos de funcionalidades virtuais que podem ser alocadas para entidades distintas, a depender da demanda e comportamento dos usuários. No problema de minimização de custo de comunicação, uma forma de

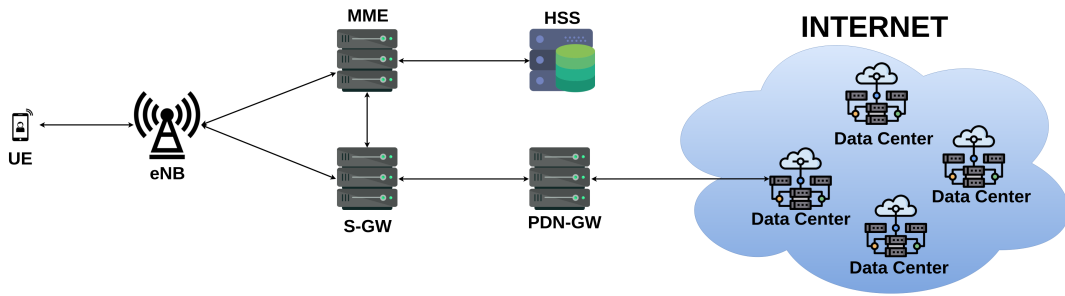


Figura 3.1: Arquitetura EPS para redes LTE.

solução é a alocação das funcionalidades virtuais do PDN-GW em *data centers* mais próximos das eNBs. Para a minimização de *handovers*, basta realizar alocação das funcionalidades virtuais do S-GW em regiões mais distantes, o que criará áreas de coberturas maiores.

Dado um conjunto de eNBs N e um conjunto de *data centers* DC , o modelo completo de Taleb et al. (2015) é formulado como:

$$\min (f(h_{ij}, x_{ij}), g(c_{is}, y_{is})) \quad (3.1)$$

$$\text{s. a } y_{is} + y_{js} \leq 1 + x_{ij} \quad \forall i, j \in N, \forall s \in DC \quad (3.2)$$

$$y_{is} - y_{js} \leq 1 - x_{ij} \quad \forall i, j \in N, \forall s \in DC \quad (3.3)$$

$$\sum_{s \in DC} y_{is} = 1 \quad \forall i \in N \quad (3.4)$$

$$y_{is}, x_{ij} \in \mathbb{B} \quad \forall i, j \in N, \forall s \in DC \quad (3.5)$$

em que, as funções objetivos f e g são definidas por:

$$f(h_{ij}, x_{ij}) = \min \sum_{i \in N} \sum_{j \in N} h_{ij}(1 - x_{ij}) \quad (3.6)$$

$$g(c_{is}, y_{is}) = \min \sum_{i \in N} \sum_{s \in DC} c_{is} y_{is} \quad (3.7)$$

neste modelo, a variável binária x_{ij} vale 1 se a eNB $_i$ e a eNB $_j$ estão conectadas ao mesmo *data center* e 0 caso contrário; já a variável y_{js} vale 1 se a eNB $_j$ está conectada ao *data center* s e 0 caso contrário. A função multiobjetivo (3.1) utiliza dois critérios f e g . A restrição (3.2) impõe que duas eNBs não conectadas não devem pertencer ao mesmo *data center*, isto é, $(x_{ij} = 0) \implies (y_{is} = 0) \vee (y_{js} = 0)$. A restrição (3.3) determina que duas eNBs conectadas não devem pertencer a *data centers* distintos, isto é, $(x_{ij} = 1) \implies (y_{is} = y_{js})$. A restrição (3.4) estabelece que toda eNB deve se conectar a apenas 1 *data center*. Por fim, (3.5) define o domínio binário das variáveis.

Com relação aos critérios da função multiobjetivo, na equação (3.6), temos a variável h_{ij} representando a frequência média de *handovers* entre as torres eNB $_i$ e eNB $_j$. Torres ligadas ao mesmo *data center* fazem parte da mesma área de cobertura, onde não há *handover* entre

eNBs. Logo, o valor ótimo para a equação (3.6) é obtido quando todas as torres pertencem a mesma área. Para a equação (3.7), o objetivo é selecionar apenas as torres mais próximas dos *data centers* no conjunto DC , garantindo sempre o menor caminho e consequentemente custo de recebimento e envio de mensagens. O valor de c_{is} é o custo de envio de mensagens entre eNB $_i$ e o *data center* s .

Como a alocação de usuários em torres não foi considerada no modelo multiobjetivo, a proposta deste trabalho incrementará a modelagem original. Será portanto adicionada a alocação de UEs em eNBs, considerando distância, quantidade média de *handovers* em cada eNB, e os requisitos de largura de banda de usuários e torres. Vale ressaltar que os critérios caminho mínimo e quantidade mínima de *handovers* apresentam características opostas, uma vez que para se ter o menor custo, a área de cobertura precisa ser a menor possível, indo em contradição a redução da quantidade de *handovers*, que demanda maior área. Dessa forma, é necessário encontrar soluções de meio termo para as duas situações e uma possível abordagem são as técnicas de otimização multiobjetivo (Marler and Arora, 2004).

3.1.1 Modelo proposto

Para contribuir com o modelo, será introduzida a alocação de UEs em eNBs, que pode ser feita de acordo com largura de banda disponível, distância e média de *handovers* em cada torre. A figura 3.2 ilustra um problema que não é tratado pelo modelo de Taleb et al. (2015). Este exemplo mostra a disposição de 5 eNBs, distanciadas horizontalmente e espaçadas por 50 unidades de distância. Considere que uma largura de banda $L_i = 20$ Mbps para cada eNB foi fixada. Cada UE tem um requisito mínimo de largura de banda $l_k = 3$ Mbps. Isso implica que cada torre não poderá oferecer serviços para mais do que 6 usuários simultaneamente. Note também que um UE se desloca linearmente com o passar do tempo, fazendo com que uma nova solução seja recalculada dinamicamente.

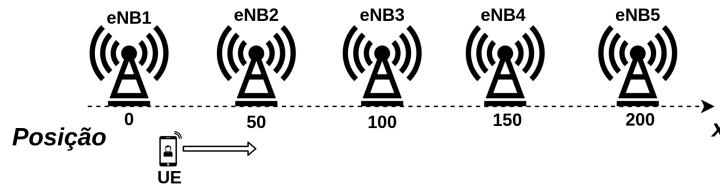


Figura 3.2: Disposição das eNBs ao longo do eixo de posições.

Neste caso, as seguintes restrições são propostas:

$$z(b_{ki}, d_{ki}, \bar{h}_i) = \sum_{k \in U} \sum_{i \in N} b_{ki} (d_{ki} + \bar{h}_i) \quad (3.8)$$

$$\bar{h}_i = \sum_{j \in N \setminus \{i\}} \frac{h_{ij}}{(N-1)} \quad \forall i \in N \quad (3.9)$$

$$\sum_{i \in N} b_{ki} = 1 \quad \forall k \in U \quad (3.10)$$

$$\sum_{k \in U} l_k b_{ki} \leq L_i \quad \forall i \in N \quad (3.11)$$

$$b_{ki} \in \mathbb{B} \quad \forall i \in N \quad (3.12)$$

No modelo, uma nova variável binária b_{ki} mapeia um usuário k de um conjunto U conectado a eNB_i , de modo que para $b_{ki} = 1$, k está conectado, e do contrário $b_{ki} = 0$. Na equação (3.8), um novo critério de otimização estabelece que os usuários devem se conectar às torres mais próximas e com menor média \bar{h}_i de *handovers*. O cálculo de \bar{h}_i é dado pela equação (3.9), que utiliza os valores de h_{ij} para cada eNB_j no conjunto N . A distância entre k e eNB_i é dada por d_{ki} . As restrições adicionais do modelo são dadas em (3.10–3.12). Na restrição (3.10), cada usuário deve se conectar a exatamente uma eNB . A restrição (3.11) determina que cada eNB_i pode disponibilizar no máximo uma largura de banda L_i , que não pode ser ultrapassada pela soma das larguras de banda mínimas l_k , exigidas por cada usuário conectado. Por fim, a restrição (3.12) define o domínio binário das variáveis.

O valor da distância d_{ki} pode ser obtido pela equação (3.13) ou (3.14), a depender do sistema de localização. Para posições Cartesianas, $d_{ki} = d_{cart}$ é a distância Euclidiana entre k , no ponto $p_k = (x_k, y_k, z_k)$, e eNB_i em $p_i = (x_i, y_i, z_i)$. Para coordenadas geográficas de latitude e longitude, a distância é dada pela Equação de Haversine para superfícies esféricas, isto é, $d_{ki} = d_{harv}$. Neste caso (ψ_i, β_i) representa respectivamente latitude e longitude de eNB_i . Analogamente (ψ_k, β_k) é referente ao usuário. O raio da esfera é denotado por R , adotaremos $R = 6371$ km (raio da Terra).

$$d_{cart} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2 + (z_i - z_k)^2} \quad (3.13)$$

$$d_{harv} = (2R) \arcsin \left(\sqrt{\sin^2 \left(\frac{\psi_i - \psi_k}{2} \right) + \cos(\psi_k) \cos(\psi_i) \sin^2 \left(\frac{\beta_i - \beta_k}{2} \right)} \right) \quad (3.14)$$

Considerando as equações (3.6), (3.7) e (3.8) como critérios de uma função multiobjetivo, o modelo proposto pode ser implementado no *framework* de programação linear CPLEX (IBM, 2017). Alguns ajustes precisam ser realizados, dado que o CPLEX não foi desenvolvido especificamente para otimizações multiobjetivo. Uma possível abordagem é o método da soma ponderada, descrito nas expressões (3.15) e (3.16). Uma vez que a modelagem possui três funções objetivo, o método da soma ponderada fornece uma conversão mono-objetivo pela equação (3.15). Dado um conjunto com m critérios de otimização, cada critério F_i será ponderado por um peso $w_i \in [0, 1]$. Variar os valores de w_i significa atribuir importâncias a problemas específicos, pois a equação (3.16) determina uma combinação convexa. Se $w_2 = 0.6$, então

$g(c_{is}, y_{is})$ terá maior contribuição de piora da solução ótima, conseqüentemente valores cada vez menores de c_{is} serão considerados (priorização de minimização de custo).

$$\sum_{i=1}^m w_i F_i(x) \quad (3.15)$$

$$\sum_{i=1}^m w_i = 1 \quad (3.16)$$

Com isso, será proposta a seguinte formulação linear:

$$\begin{aligned} \min \quad & obj = w_1 \times f(h_{ij}, x_{ij}) + w_2 \times g(c_{is}, y_{is}) + w_3 \times z(b_{ki}, d_{ki}, \bar{h}_i) \\ \text{s. a} \quad & (3.2 - 3.5) \text{ e } (3.10 - 3.12) \end{aligned} \quad (3.17)$$

Para simplificar o entendimento, a notação e descrição das variáveis utilizadas na modelagem pode ser observada na tabela 3.1.

Notação	Descrição
h_{ij}	Frequência média de <i>handovers</i> entre eNB _i e eNB _j
\bar{h}_i	Valor médio de <i>handovers</i> em eNB _i
x_{ij}	Variável que indica se eNB _i e eNB _j estão conectadas
y_{is}	Variável que indica se eNB _i está conectada ao <i>data center</i> s
c_{is}	Custo de envio de mensagem entre eNB _i e s
b_{ki}	Variável que indica se usuário k está alocado em eNB _i
d_{ki}	Distância entre k e eNB _i
d_{cart}	Distância Euclidiana
d_{harv}	Distância pela fórmula de Haversine
ψ	Latitude
β	Longitude
R	Raio da Terra
N	Conjunto de torres (eNBs)
DC	Conjunto de <i>data centers</i>
U	Conjunto de usuários
L_i	Largura de banda máxima oferecida por eNB _i
l_k	Largura de banda mínima exigida por k

Tabela 3.1: Notação de constantes e variáveis utilizadas na modelagem do problema.



Resultados e Discussões

4.1 Resultados

O modelo de Programação Linear Inteira foi implementado e todos os experimentos foram executados no *solver* IBM ILOG CPLEX Optimization Studio v12.5.0 em um computador Intel Core i5-6200U (2.30GHz e 4 cores) com 4GB de RAM em um Sistema operacional Ubuntu 16.04 LTS.

4.1.1 Tempo de execução

Na Tabela 4.1, estão dispostos os tempos médios de execução para diferentes combinações dos pesos w_1 , w_2 e w_3 na equação (3.17). Para este caso de teste, foram considerados 10 *data centers* e uma amostra de 60 torres da cidade de Maceió, Alagoas, na base fornecida pela [Telebrasil](#), que dispõe latitude e longitude de antenas de diversas operadoras no Brasil. Dez usuários foram dispostos arbitrariamente (e mantidos) em posições próximas das torres para que as distâncias fossem calculadas. Uma vez que dados de *handover* e custo de comunicação entre *data centers* não estão disponíveis, 30 instâncias foram geradas aleatoriamente. O tempo de execução de cada instância foi computado e ao final, a média foi calculada com os tempos das instâncias para as diferentes configurações de pesos.

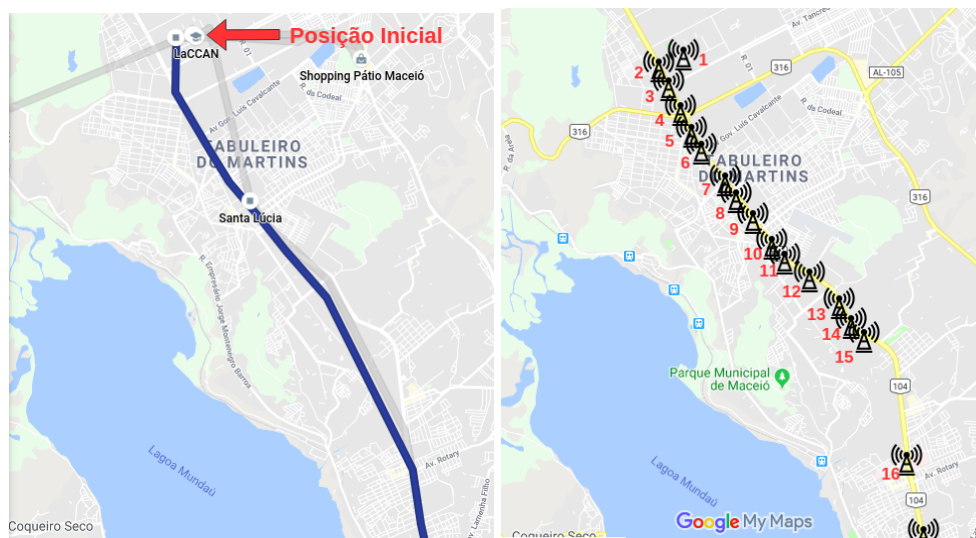
As configurações (1, 0, 0), (0, 1, 0) e (0, 0, 1) verificam o tempo de execução de $f(x_{ij}, h_{ij})$, $g(c_{is}, y_{is})$ e $z(b_{ki}, d_{ki}, \bar{h}_i)$ isoladamente na equação (3.17). Em um caso de atribuição de maior prioridade a um objetivo específico, é observável que o tempo de execução aumenta consideravelmente em (0.1, 0.8, 0.1) quando os custos de comunicação entre eNB e *data center* precisam ser estritamente menores. Nas configurações mais uniformes como (0.4, 0.3, 0.3), os tempos de execução são similares. As combinações de pesos (0.9, 0.1, 0) e (0.1, 0.9, 0) correspondem ao tempo médio de execução quando considerado apenas o modelo de [Taleb et al. \(2015\)](#).

w_1	w_2	w_3	Tempo médio de execução (segundos)
1	0	0	0.75
0	1	0	1.09
0	0	1	1.13
0.8	0.1	0.1	2.04
0.1	0.8	0.1	4.44
0.1	0.1	0.8	2.42
0.4	0.3	0.3	2.69
0.3	0.4	0.3	2.70
0.3	0.3	0.4	2.42
0.9	0.1	0	1,69
0.1	0.9	0	4,17

Tabela 4.1: Tempo médio de execução para diferentes combinações de pesos. Número de diferentes instâncias computadas = 30, $N = 60$, $DC = 10$, $U = 10$.

4.1.2 Alocação de usuários

Para alocação de usuários, foi utilizado um cenário baseado em dados reais de latitude e longitude de um UE em movimento. As figuras 4.1(a) e 4.1(b) exibem um deslocamento de 13.7 km, percorrido em um intervalo de 42 minutos a partir do ponto indicado na figura 4.1(a). O sentido do movimento é do ponto de partida para as torres com maior numeração, figura 4.1(b). A eNB que UE está alocado deve mudar conforme sua posição. A distância neste cenário é dada pela equação 3.14 (Haversine) e os dados de latitude e longitude foram coletados do histórico de localização do Google Maps. No total, 21 torres de teste (fictícias) foram posicionadas longo do percurso.



(a) Percurso percorrido pelo UE.

(b) Disposição das eNBs ao longo do percurso.

Figura 4.1: Rota seguida pelo UE ao longo do tempo, partindo do ponto inicial e seguindo no sentido das eNBs de maior numeração.

A figura 4.2 exibe as eNBs onde o UE foi alocado para o percurso de 13.7 km. De acordo com suas informações de GPS, no instante inicial sua posição era próxima de eNB₁ eNB₂, com distâncias respectivas de 0.08 km e 0.45 km. A média de *handover* destas torres é de 5.3 para a primeira e 3.75 para segunda. O modelo proposto considera menor distância e média de *handover*, logo, é observável no gráfico a alocação do usuário em eNB₂. Durante os 10 últimos minutos, eNB₁₉, eNB₂₀ e eNB₂₁ se encontram mais próximas, a uma distância respectiva de 0.78, 0.91 e 0.53 quilômetros. Na mesma ordem, as médias de *handover* são 3.8, 6.25, 5.5, o que justifica a permanência de UE em eNB₁₉ durante esse trecho do percurso.

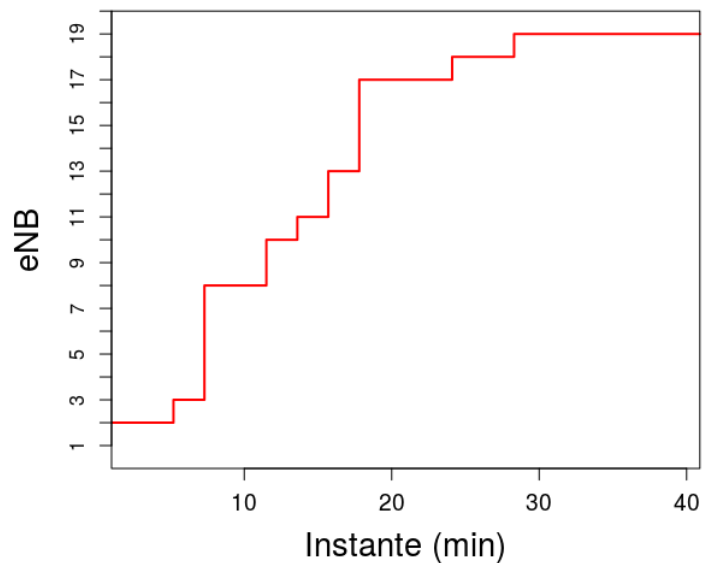


Figura 4.2: Alocação do usuário nas eNBs ao longo do tempo.

4.1.3 Gerenciamento de largura de banda

Para verificar as restrições de largura de banda definidas pelo modelo, simularemos um segundo cenário de mobilidade para um UE se deslocando linearmente e 19 UEs em posições estáticas (20 em total), além de 5 eNBs. As restrições de largura de banda para eNBs e UEs são respectivamente $L_i = 20$ Mbps e $I_k = 3$ Mbps (valores hipotéticos). O modelo não contempla diretamente tempo, mas as posições (que influenciam o valor de d_{ki}) variam com o passar do mesmo, e neste exemplo a posição x será modificada pela função ($x = t$), de modo que t é um instante qualquer — em outras palavras, o modelo matemático é executado a cada unidade de tempo. O gráfico na figura 4.3 mostra a torre eNB _{i} que um usuário UE (nomeadamente UE₁) foi alocado. A equação (3.8) leva em consideração a posição atual de UE₁ e da eNB mais próxima. A partir do momento que se torna muito custoso se manter conectado a eNB de origem, o processo de *handover* é realizado. Como pode ser observado na figura 4.3, a transição de eNB₁ para eNB₂ é realizada nas proximidades da posição $x = 30$. As transições se repetem conforme o usuário se dirige para as outras torres.

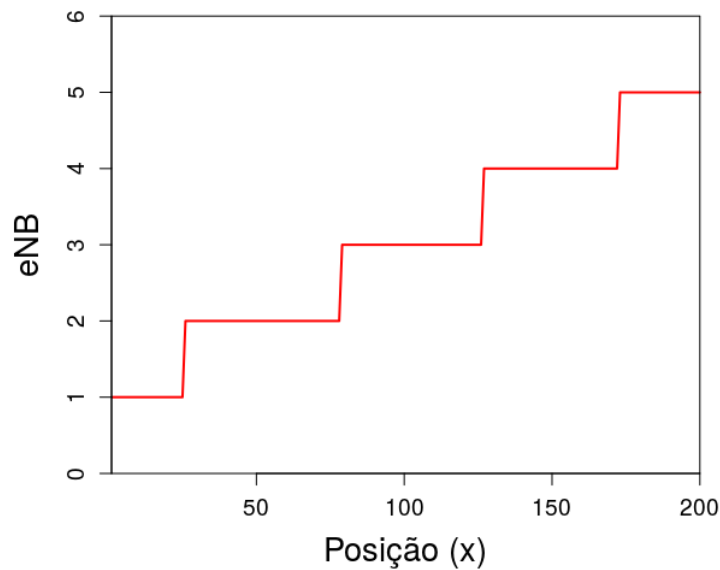
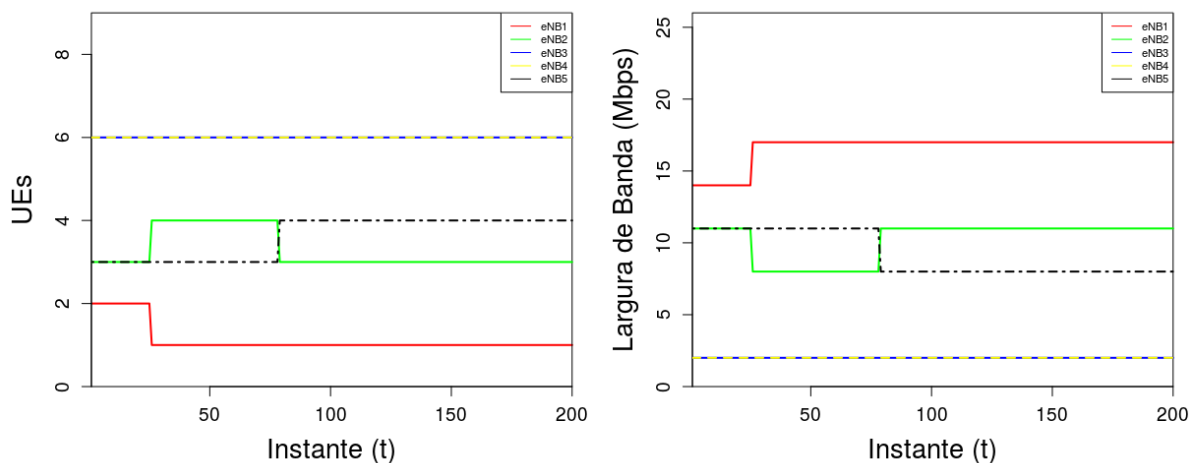


Figura 4.3: Torre (eNB_i) cujo usuário (UE_1) está alocado em função de sua posição (x).

Os gráficos nas figuras 4.4(a) e 4.4(b) apresentam respectivamente a quantidade de UEs alocados e largura de banda disponível em cada eNB no instante t . Os valores mudam de acordo com o deslocamento de UE_1 . Podemos observar na figura 4.4(a) que inicialmente a torre eNB_1 possui 2 UEs conectados. Sabemos também que UE_1 se desloca linearmente, com sua posição determinada por $x = t$. Logo, para $t = 0$, UE_1 está em $x = 0$, e conseqüentemente alocado em eNB_1 . Observando a disponibilidade de largura de banda de eNB_1 na figura 4.4(b) no mesmo instante $t = 0$, para 2 usuários conectados, cada um com uma exigência de 3 Mbps, é verificável que a torre dispõe de apenas 14 Mbps, de um total $L_1 = 20$ Mbps.



(a) UEs alocados em uma dada eNB no instante (t). (b) Largura de banda disponível em cada eNB no instante (t).

Figura 4.4: Quantidade de usuários alocados e largura de banda disponível em cada eNB. Um usuário está se deslocando e alterando a configuração da rede para cada instante (t).

A reorganização da rede pode ser verificada a medida que UE_1 se desloca. Como constatado previamente, a transferência de eNB_1 para eNB_2 ocorre nas proximidades de $x = 30$, de modo que se verifica na figura 4.4(a) para $t \approx 30$ a perda de um UE em eNB_1 (de 2 para 1) e o ganho na torre eNB_2 (de 3 para 4). O ganho ou perda de largura de banda pode se verificar na figura 4.4(b) para o mesmo instante de referência.

A disponibilidade de largura de banda de eNB_4 e eNB_5 está a todo momento baixa, uma vez que cada uma está operando com 6 usuários. Ao se aproximar destas torres, e é observável na figura 4.3 que de fato isso acontece, UE_1 certamente deverá ser atendido. Mesmo operando na sua capacidade limite em algumas torres, a rede realiza comutação de UEs, garantindo que todos sejam atendidos. Para qualquer instante t , a soma total dos usuários conectados em todas $eNBs$ é igual a 20.

5

Considerações Finais

Este trabalho apresentou um modelo de alocação de usuários em torres para redes de telefonia móveis considerando processos de *handover* e redes definidas por *software*. Os parâmetros utilizados foram: distancia entre torre e usuário, média de *handover*, largura de banda disponível e largura de banda exigida. O modelo expande uma modelagem de minimização de custo de comunicação entre torres e *data centers* e minimização dos processos de *handover* entre áreas de cobertura. A finalidade é o aprimoramento de experiência de conectividade e mobilidade, com possíveis aplicações em redes 5G. Nos resultados para alocação de usuários em torres foi verificado que as conexões de fato consideram menores distâncias, média de *handover* e disponibilidade de largura de banda, mesmo em situação mobilidade.

Como trabalhos futuros, elencam-se o aprimoramento do modelo com a introdução de novas restrições e incorporação de custos de operações da rede, integração com simuladores como o Mininet para redes definidas por *software*, e a implementação de diferentes heurísticas para a busca das soluções de alocação.

Referências bibliográficas

- I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393 – 422, 2002.
- Zoraze Al, Nicola Baldo, Josep Mangues-Bafalluy, and Lorenza Giupponi. Machine learning based handover management for improved QoE in LTE. *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, jul 2016.
- Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Communications Surveys & Tutorials*, 17(4):2347–2376, jun 2015.
- Federico Boccardi, Jeffrey Andrews, Hisham Elshaer, Mischa Dohler, Stefan Parkvall, Petar Popovski, and Sarabjot Singh. Why to Decouple the Uplink and Downlink in Cellular Networks and How To Do It. *IEEE Communications Magazine*, 54(3):110–117, mar 2016.
- Min Chen, Shiwen Mao, and Yunhao Liu. Big Data: A Survey. *Mobile Networks & Applications*, 19(2):171–209, APR 2014.
- CISCO. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022. <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.pdf>, feb 2019. White paper, acessado em maio de 2019.
- Luis M. Correia, Dietrich Zeller, Oliver Blume, Dieter Ferling, Ylva Jading, István Gódor, Gunther Auer, and Liesbet Van Der Perre. Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks. *IEEE Communications Magazine*, 48(11):66–72, nov 2010.
- Aleksandar Damnjanovic, Juan Montojo, Yongbin Wei, Tingfang Ji, Tao Luo, Madhavan Vajapeyam, Taesang Yoo, Osok Song, and Durga Mallad. A Survey on 3GPP Heterogeneous Networks. *IEEE Wireless Communications*, 18(3):10–21, jun 2011.
- Xiaoyu Duan and Xianbin Wang. Authentication handover and privacy protection in 5G hetnets using software-defined networking. *IEEE Communications Magazine*, 53(4):28–35, apr 2015.

Ericsson. More than 50 billion connected devices.

https://www.akos-rs.si/files/Telekomunikacije/Digitalna_agenda/Internetni_protokol_Ipv6/More-than-50-billion-connected-devices.pdf, feb 2011. White paper, acessado em maio de 2019.

GSMA Intelligence. Understanding 5G: Perspectives on future technological advancements in mobile.

<https://www.gsmainelligence.com/research/?file=141208-5g.pdf&download>, dec 2014. White paper, acessado em maio de 2019.

Bo Han, Vijay Gopalakrishnan, Lusheng Ji, and Seungjoon Lee. Network function virtualization: Challenges and opportunities for innovations. *IEEE Communications Magazine*, 53(2): 90–97, feb 2015.

Chi-Yao Hong, Srikanth Kandula, Ratul Mahajan, Ming Zhang, Vijay Gill, Mohan Nanduri, and Roger Wattenhofer. Achieving High Utilization with Software-driven WAN. *SIGCOMM Comput. Commun. Rev.*, 43(4):15–26, aug 2013.

IBM. IBM ILOG CPLEX Optimization Studio Getting Started with CPLEX.

https://www.ibm.com/support/knowledgecenter/SSSA5P_12.8.0/ilog.odms.studio.help/pdf/gscplex.pdf, apr 2017. Online, acessado em maio de 2019.

Sushant Jain, Alok Kumar, Subhasree Mandal, Joon Ong, Leon Poutievski, Arjun Singh, Subbaiah Venkata, Jim Wanderer, Junlan Zhou, Min Zhu, Jonathan Zolla, Urs Hölzle, Stephen Stuart, and Amin Vahdat. B4: Experience with a Globally Deployed Software Defined WAN. In *Proceedings of the ACM SIGCOMM Conference*, 2013.

Diego Kreutz, Fernando M. V. Ramos, Paulo Verissimo, Christian Esteve Rothenberg, Siamak Azodolmolky, and Steve Uhlig. Software-Defined Networking: A Comprehensive Survey. *Proceedings of the IEEE*, 103(1):14–76, jan 2015.

Slawomir Kuklinski, Yuhong Li, and Khoa Truong Dinh. Handover management in SDN-based mobile networks. *2014 IEEE Globecom Workshops, GC Wkshps 2014*, pages 194–200, mar 2015.

Jiseong Lee and Younghwan Yoo. Handover cell selection using user mobility information in a 5G SDN-based network. *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 697–702, jul 2017.

R.T. Marler and J.S. Arora. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6):369–395, apr 2004.

Peter Mell and Tim Grance. The NIST Definition of Cloud Computing. *Communications of The ACM*, 53(6):50, jun 2010.

- Saraju P. Mohanty, Uma Choppali, and Elias Kougianos. Everything you wanted to know about smart cities: The Internet of things is the backbone. *IEEE Consumer Electronics Magazine*, 5(3):60–10, jul 2016.
- Bruno Astuto A. Nunes, Marc Mendonca, Xuan-Nam Nguyen, Katia Obraczka, and Thierry Turletti. A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks. *IEEE Communications Surveys & Tutorials*, 16(3):1617–1634, feb 2014.
- ONF. OpenFlow Switch Specification. <https://www.opennetworking.org/wp-content/uploads/2014/10/openflow-switch-v1.5.1.pdf>, apr 2015. Online; Acessado em maio de 2019.
- Nisha Panwar, Shantanu Sharma, and Awadhesh Kumar Singh. A survey on 5G: The next generation of mobile communication. *Physical Communication*, 18:64 – 84, 2016. Special Issue on Radio Access Network Architectures and Resource Management for 5G.
- Jonathan Prados-Garzon, Oscar Adamuz-Hinojosa, Pablo Ameigeiras, Juan J. Ramos-Munoz, Pilar Andres-Maldonado, and Juan M. Lopez-Soler. Handover implementation in a 5G SDN-based mobile network architecture. pages 1–6, sep 2016.
- Li Qiang, Jie Li, and Changcheng Huang. A software-defined network based vertical handoff scheme for heterogeneous wireless networks. *2014 IEEE Global Communications Conference*, 2014.
- Li Qiang, Jie Li, and Corinne Touati. A User Centered Multi-Objective Handoff Scheme for Hybrid 5G Environments. *IEEE Transactions on Emerging Topics in Computing*, 5(3): 380–390, apr 2016.
- Alexander Shalimov, Dmitry Zuikov, Daria Zimarina, Vasily Pashkov, and Ruslan Smeliansky. Advanced Study of SDN/OpenFlow Controllers. In *Proceedings of the 9th Central & Eastern European Software Engineering Conference in Russia*, pages 1:1–1:6. ACM, 2013.
- Vishal Sharma, Fei Song, Ilsun You, and Han-Chieh Chao. Efficient Management and Fast Handovers in Software Defined Wireless Networks Using UAVs. *IEEE Network*, 31(6):78–85, nov 2017.
- Tarik Taleb, Miloud Bagaa, and Adlen Ksentini. User mobility-aware Virtual Network Function placement for Virtual 5G Network Infrastructure. In *2015 IEEE International Conference on Communications (ICC)*, 2015.
- Telebrasil. Mapa de ERBs Brasil (antenas). <http://www.telecocare.com.br/telebrasil/erbs/>. Online, acessado em maio de 2019.

- Ms. Lopa J. Vora. Evolution of Mobile Generation Technology: 1G to 5G and Review of Upcoming Wireless Technology 5G. *International Journal of Modern Trends in Engineering and Research*, 2(10):281–290, oct 2015.
- Jun Wu, Zhifeng Zhang, Yu Hong, and Yonggang Wen. Cloud Radio Access Network (C-RAN): A Primer. *IEEE Network*, 29(1):35–41, jan 2015.
- Xiaohuan Yan, Y. Ahmet Şekercioğlu, and Sathya Narayanan. A survey of vertical handover decision algorithms in Fourth Generation heterogeneous wireless networks. *Computer Networks*, 54(11):1848–1863, aug 2010.
- Haijun Zhang, Na Liu, Xiaoli Chu, Keping Long, Abdol-Hamid Aghvami, and Victor C. M. Leung. Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges. *IEEE Communications Magazine*, 55(8):138–145, aug 2017.