

Orquestração Inteligente de *Network Slicing*: Revisão da Literatura e Prospecção para Redes 6G

Henrique V. de Lima¹, Rogério S. Silva^{1,2}, Cristiano B. Both⁴,
Antonio Oliveira-Jr^{1,3}, Kleber V. Cardoso¹, Sand L. Corrêa¹

¹Instituto de Informática - Universidade Federal de Goiás (UFG), GO, Brasil

²Instituto Federal de Goiás (IFG), GO, Brasil

³Fraunhofer Portugal AICOS, Porto, Portugal

⁴Universidade do Vale do Rio dos Sinos, São Leopoldo, RS, Brasil

{henriquedelima}@inf.ufg.br, {rogerio.sousa}@ifg.edu.br,

{cbboth@unisin.br, {sandluz,antoniojr,kleber}@ufg.br

Abstract. *In this paper, we investigate research challenges on intelligent management and orchestration of network slices in 5G networks and Beyond. In particular, we review the literature in order to understand the main problems involving this issue, as well as machine learning techniques usually employed to solve such problems. We also discuss open issues and new challenges on management and orchestration of network slices imposed by 6G networks.*

Resumo. *Neste trabalho, investiga-se os desafios de pesquisa relacionados à gerência e orquestração inteligente de network slices nas redes 5G e de próxima geração. Particularmente, a literatura é revisada com o objetivo de compreender os principais problemas abordados neste escopo, bem como as técnicas de Aprendizado de Máquina geralmente empregadas para a solução de tais problemas. Além disso, discute-se questões em aberto e novos desafios que as redes 6G imporão à gerência e orquestração de slices.*

1. Introdução

A Quinta Geração de Rede Móvel Celular (5G) é a primeira a introduzir o conceito de uma rede flexível, programável e capaz de atender serviços com requisitos díspares e desafiadores, acelerando a transformação digital em diversos segmentos verticais. Esse novo conceito de rede traz diversas vantagens para o operador, principalmente em termos de redução de custos em investimentos e operação. Diante destas vantagens, é esperado que tal conceito alcance sua plenitude na Sexta Geração de Rede Móvel Celular (6G).

Um elemento essencial para a implantação de redes móveis com tais características é o conceito de *network slicing*. Proposto pela *Next Generation Mobile Networks* (NGMN) como um conceito fim-a-fim, o termo *network slicing* refere-se à criação e operação de múltiplas redes logicamente independentes sobre uma infraestrutura de telecomunicação física e compartilhada [NGMN Alliance 2016]. Cada rede lógica, denominada *network slice* ou simplesmente *slice*, consiste em um conjunto de *Virtualized Network Functions* (VNFs), bem como recursos de computação, armazenamento e rede para executá-las. Essas funções e recursos formam uma rede lógica para atender, de forma customizada, as características específicas (e.g., *Service Level Agreement* (SLA)) do serviço representado pelo *slice*. Embora *network slice* aumente o nível de flexibilidade e eficiência das redes móveis de próxima geração, ele também traz a necessidade de mecanismos flexíveis

para implantação e gerenciamento do ciclo de vida do serviço, alocação de recursos otimizados para cada tipo de caso de uso e garantias que os requisitos acordados para cada serviço serão cumpridos. Essas exigências, por sua vez, trazem maior complexidade para a operação e o gerenciamento das redes. Mecanismos autônomos para a gerência e orquestração de *slices* se tornam essenciais. Inteligência Artificial (IA), realizada através de técnicas de Aprendizado de Máquina (AM), é a escolha natural para apoiar a gerência inteligente das redes de próxima geração. Nos últimos anos, alguns trabalhos foram propostos visando aplicar AM para resolver alguns problemas relacionados com a gerência e orquestração inteligente de *slices*. No entanto, tais soluções ainda estão em estágio inicial e diversas questões ainda estão em aberto para que essas soluções possam ser concretamente implantadas em uma rede em operação. Adicionalmente, características implícitas às redes 6G, como a integração de diferentes meios de comunicação (e.g., espaço, ar e terra), o suporte nativo a serviços de IA e o compromisso com a sustentabilidade, devem trazer novos desafios à realização do conceito de *network slicing*.

Neste trabalho, investiga-se os desafios de pesquisa relacionados à gerência e orquestração inteligente de *network slices* nas redes 5G e de próxima geração. Particularmente, literatura é revisada, tendo em vista três objetivos principais, descritos a seguir. Na Seção 2, revisa-se como a gerência e orquestração de *slices* foi desenhada na arquitetura 5G, apresentando as principais funções de rede da arquitetura envolvidas nessa atividade. Na Seção 3, investiga-se os principais problemas abordados na literatura envolvendo a gerência e orquestração inteligente de *slices*, bem como discute-se as técnicas de AM, geralmente empregadas para a solução tais problemas. Na Seção 4, discute-se problemas em aberto e novos desafios para gerência e orquestração inteligente de *slices*, tendo em vista os requisitos esperados para as futuras redes 6G. Finalmente, a Seção 5 traz as considerações finais deste trabalho.

2. *Network Slicing* na Arquitetura 5G

Seguindo a nomenclatura proposta pelo *3rd Generation Partnership Project* (3GPP) em [3GPP 2018], um *slice* é uma entidade cujo ciclo de vida deve ser gerenciado dentro da rede do operador. Particularmente, o ciclo de vida de um *slice* é composto por quatro fases principais: preparação, instanciação, execução e descomissionamento. Na fase de preparação, o *slice* ainda não existe. Nesta fase, os requisitos do serviço fim-a-fim associado ao *slice* são descritos através de *template*. Na fase de instanciação, os requisitos do serviço fim-a-fim são traduzidos para requisitos do *slice*, criando um descritor de *slice*. O *slice* é então criado, configurado e implantado sobre a infraestrutura do operador, de forma a atender aos requisitos descritos no seu descritor. Na fase de execução, o *slice* entra em operação e seu comportamento é monitorado para garantir o cumprimento dos requisitos especificados em seu descritor. Finalmente, no descomissionamento, o *slice* é encerrado e os recursos a ele atribuídos são liberados.

Para gerenciar e orquestrar o ciclo de vida de *slices*, o 3GPP propõe três funções de rede principais, inseridas na arquitetura de serviço 5G: *Communication Service Management Function* (CSMF), *Network Slice Management Function* (NSMF) e *Network Slice Subnet Management Function* (NSSMF). Essas funções estão representadas na Figura 1. A função CSMF é responsável pela tradução dos requisitos de um serviço fim-a-fim em requisitos de *slice*, gerando o descritor de *slice*. A função NSMF recebe o descritor gerado pela CSMF e o decompõe em requisitos de sub-rede. As funções NSSMF, presentes nas diferentes sub-redes (e.g., *Radio Access Network* (RAN) e *Core Network* (CN)), utilizam

os requisitos recebidos da NSMF e instanciam e gerenciam as VNFs relacionadas à sua sub-rede. Para isso, a NSSMF depende de um arcabouço de gerenciamento e orquestração (e.g., ETSI NFV-MANO [ETSI 2014]) implantado na sua sub-rede.

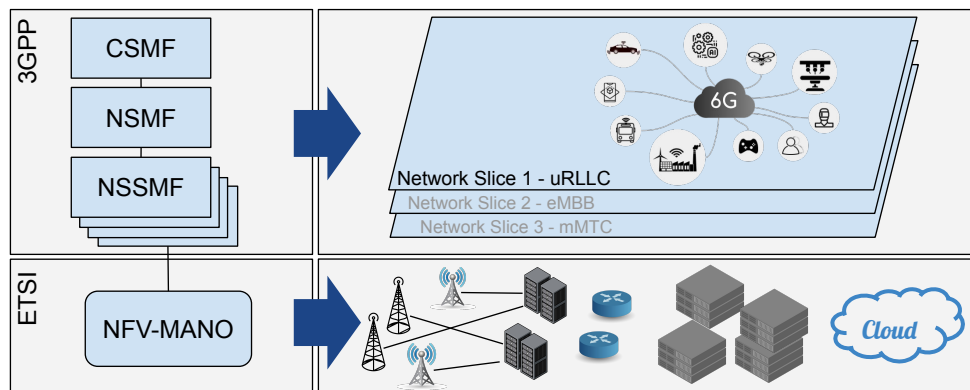


Figura 1. Arquitetura para gerenciamento e orquestração de *network slicing*.

Além das funções ilustradas na Figura 1, a arquitetura de serviço 5G inclui uma função de rede, denominada *Network Data Analytic Function* (NWDAF), responsável pela coleta e análise de dados de vários módulos do sistema 5G, incluindo dados de sessões de usuários e informações relacionadas ao estado global do sistema. Tais dados podem ser processados para fornecer informações sobre tráfego da rede, utilização de recursos, métricas de desempenho e mobilidade de usuários, os quais podem ser usados como entrada para algoritmos de AM para prover a gerência e orquestração inteligente de *slices*.

3. Gerência e Orquestração Inteligente de *Slices*

Em geral, a gerência e orquestração inteligente de *slices* está relacionada à alocação de recursos nos diferentes momentos do ciclo de vida do *slice*. Neste sentido, três problemas têm sido comumente abordados na literatura [Debbabi et al. 2020], como discutido a seguir.

Posicionamento de VNFs: um *slice* envolve um conjunto de VNFs, as quais podem se estender desde a RAN até a CN. Uma decisão a ser tomada na fase de instanciação do *slice* consiste no posicionamento dessas VNFs ao longo dos diferentes segmentos de rede. Esse problema, normalmente denominado *Virtual Network Embedding* (VNE), deve ser tratado pela NSMF e consiste em dois sub-problemas: (i) o mapeamento de nós virtuais em nós físicos e (ii) o mapeamento de links virtuais (conectando nós virtuais) em caminhos físicos. Dois grafos são tomados como entrada do problema, um representando a infraestrutura física e outro representando o conjunto de VNFs a ser instanciado na infraestrutura. Nós da infraestrutura possuem recursos (e.g., CPU e largura de banda) que serão usados para hospedar os elementos virtuais. O propósito do problema é posicionar as VNFs que compõem o *slice* em nós da infraestrutura e mapear os *links* lógicos entre elas em caminhos físicos, de acordo com diferentes objetivos e respeitando os requisitos do *slice*. Os trabalhos propostos na literatura geralmente têm como objetivo maximizar o lucro do provedor de rede (operador). Neste caso, a receita e o custo incorridos com o mapeamento, bem como a taxa de aceitação de requisições, são os fatores que influenciam diretamente no lucro final do provedor. O problema de VNE pode ser formulado como um *Markov Decision Process* (MDP), onde um agente recebe as requisições

para criação de *slice* e ganha uma recompensa quando a requisição é mapeada para a infraestrutura física [Haeri and Trajković 2018]. O objetivo do agente é maximizar as recompensas recebidas. Recentemente, uma solução baseada em *Deep Reinforcement Learning* (DRL) [Quang et al. 2019] foi proposta para tratar o problema.

Controle de admissão: um mecanismo de controle de admissão de *slice* é um algoritmo que executa periodicamente na NSMF, sendo invocado na fase de instanciação com o objetivo de aceitar ou rejeitar requisições para criação de *slices* na infraestrutura do operador. Mecanismos dessa natureza são importante para evitar que os recursos do provedor sejam sobrecarregados em um cenário de alta demanda por requisições de criação de *slices*. Portanto, é função do controle de admissão de *slice* selecionar as requisições para criação de *slices* que serão aceitas em um determinado instante do tempo e provisionar recursos para os *slices* admitidos. A estratégia empregada, por sua vez, terá um grande impacto no lucro do provedor de rede e na utilização geral dos recursos. Uma estratégia muito conservadora pode levar a uma baixa utilização da infraestrutura e, conseqüentemente, redução de receita. Por outro lado, uma estratégia excessivamente agressiva pode levar a violações de SLA em um cenário de pico de demanda em múltiplos *slices* simultaneamente. Grande parte dos trabalhos relacionados com controle de admissão de *slices* utilizam estratégias de decisão baseadas em políticas para tratar esse problema [Bin and Schotten 2019]. Nessas estratégias, o provedor de rede apresenta uma lista de custos para os diferentes tipos de *slices*. Os *slices* são tipificados por *templates* definidos especificamente para cada tipo de serviço (e.g., *Enhanced Mobile Broadband* (eMBB), *Massive Machine Type Communications* (mMTC) e *Ultra-Reliable Low-Latency Communications* (URLLC)). Cada *template* define um conjunto de recursos (e.g. CPU e/ou largura de banda) necessários para instanciar o *slice*, o tempo de vida do *slice* e o SLA associado. Ao requisitar a criação de um *slice*, o inquilino escolhe o *template* desejado. A decisão sobre admitir ou não um *slice* é feita de acordo com a política do provedor de rede definida sobre o estado corrente do sistema. Esse último é normalmente definido a partir de informações como quantidade de recursos ociosos, conjunto de *slices* ativos e a fila de requisições em espera. Decisões baseadas em políticas são tipicamente MDP, onde uma política mapeia cada estado do sistema em uma ação e uma recompensa correspondente. Particularmente, para o problema de admissão de *slices*, os autores em [Han et al. 2019] mostram que a função de recompensa é geralmente não convexa sobre um grande espaço de solução. Nesse cenário, *Reinforcement Learning* (RL) é conhecidamente eficiente. Portanto, técnicas como *Q-Learning* [Bega et al. 2017], *Multi-armed Bandit* [Zanzi et al. 2021] e DRL [Bega et al. 2019] foram utilizadas.

Elasticidade: estratégias de elasticidade de *slices* focam no provimento de recursos para *slices* já admitidos e ativos na infraestrutura, sendo aplicadas na fase de execução dos *slices*, geralmente pela função NSMF. A propriedade de elasticidade é definida como a capacidade de um *slice* aumentar ou diminuir dinamicamente a quantidade de recursos alocada, de acordo com a carga de trabalho percebida. Tal propriedade é importante, uma vez que a carga de trabalho de um *slice* pode variar significativamente ao longo do tempo. A elasticidade de *slice* é geralmente classificada em dois tipos: vertical e horizontal. A maioria dos trabalhos definem elasticidade vertical como sendo a capacidade de aumentar ou diminuir dinamicamente os recursos atribuídos a uma máquina virtual ou contêiner, enquanto a elasticidade horizontal refere-se à habilidade de aumentar ou diminuir dinamicamente instâncias de máquinas virtuais ou contêineres. Operações de elasticidade podem ser implementadas como decisões baseadas em políticas. Nesse caso, o provedor de rede oferece um conjunto de recursos e a decisão de quais recursos alocar para cada *slice* em execução

depende da quantidade de recursos disponível, da demanda instantânea de cada *slice* ativo e sua respectiva taxa de serviço. Na maioria dos trabalhos, a elasticidade é abordada com relação aos recursos de computação, devido à cloudificação da CN (completamente) e da RAN (parcialmente). Muitos trabalhos [Li et al. 2018, Gutierrez-Estevez et al. 2019] utilizam DRL para resolver esse problema.

4. Problemas em Aberto e Proposição para 6G

Algoritmos de IA/AM são essenciais para prover autonomicidade às funções de gerenciamento e orquestração de *slices* e para avançar em direção a uma rede gerenciada e operada com intervenção humana mínima. No entanto, três desafios principais precisam ser superados para que esse objetivo seja alcançado [Benzaid and Taleb 2020]. O primeiro refere-se à falta de *datasets* de alta qualidade que possam ser utilizados no processo de aprendizado. A validação e acurácia dos modelos aprendidos dependem diretamente da disponibilidade de *datasets*. O segundo desafio refere-se à interpretabilidade dos modelos aprendidos. A adoção de modelos de IA para gerência e operação das redes de próxima geração depende, em grande parte, da compreensão da relação causa-e-efeito entre decisões tomadas e dados de entrada que levaram a tais tomadas de decisões. Infelizmente, os algoritmos de IA mais promissores para diversos problemas relacionados à gerência e orquestração de *slices*, como algoritmos de DRL, geram modelos com alta complexidade e pouca interpretabilidade. Finalmente, para suportar latências muito baixas e alta confiabilidade, funções de gerenciamento e orquestração de *slices* devem tomar decisões acuradas em tempo real ou quase em tempo real. Embora técnicas emergentes de IA/AM, como DRL, apresentem alta acurácia, o tempo de treinamento requerido por essas técnicas ainda é alto.

Redes 6G devem utilizar de forma integrada, diferentes meios de comunicação (e.g., espaço, ar e terra), provendo suporte nativo a serviços de IA e grande compromisso com a sustentabilidade [Saad et al. 2020]. Esses requisitos devem trazer novos desafios à realização do conceito de *network slicing*. No primeiro caso, o gerenciamento de *slices* sobre redes terrestres, aerotransportadas e de satélites não apenas aumentará a quantidade de recursos a serem gerenciados, como também exigirá uma coordenação criteriosa de segmentos de rede altamente heterogêneos. No segundo caso, a crescente demanda por IA em diferentes tipos de aplicações exigirá que a rede ofereça IA como um serviço. Dessa forma, serviços de IA também devem fazer parte de um ou múltiplos *slices*, como os serviços de rede e serviços MEC. Serviços de IA, por sua vez, apresentam novos requisitos (e.g., acurácia do modelo e velocidade de aprendizado) que diferem dos requisitos tradicionais de redes. Finalmente, modelos acurados de IA exigem um grande consumo de processamento e memória e, portanto, um alto consumo de energia elétrica. Portanto, além do tradicional compromisso entre multiplexação dos ganhos do provedor de rede e atendimento aos SLAs acordados, mecanismos de *network slicing* nas redes 6G devem levar em consideração também uma dimensão relacionada à eficiência energética provendo serviços de IA de forma sustentável.

5. Conclusão

Neste trabalho, discutiu-se os desafios de pesquisa relacionados à gerência e orquestração inteligente de *network slices* nas redes 5G e de próxima geração. A literatura especializada foi revisada para compreender os principais problemas abordados neste escopo e as técnicas de IA/am, geralmente empregadas para a solução de tais problemas. Por fim, este

artigo fornece uma discussão inicial sobre os desafios que as redes 6G devem trazer para o gerenciamento e orquestração de *slices*.

Agradecimentos

O presente trabalho foi realizado com apoio do Projeto Brasil 6G, financiado pelo Ministério da Ciência, Tecnologia e Inovações (MCTI) e Rede Nacional de Pesquisa (RNP).

Referências

- 3GPP (2018). Telecommunication Management; Study on Management and Orchestration of Network Slicing for Next Generation Network.
- Bega, D. et al. (2017). Optimising 5G infrastructure markets: The business of network slicing. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9.
- Bega, D. et al. (2019). DeepCog: Cognitive Network Management in Sliced 5G Networks with Deep Learning. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 280–288.
- Benzaid, C. and Taleb, T. (2020). AI-Driven Zero Touch Network and Service Management in 5G and Beyond: Challenges and Research Directions. *IEEE Network*, 34(2):186–194.
- Bin, H. and Schotten, H. D. (2019). Machine Learning for Network Slicing Resource Management: A Comprehensive Survey. *ZTE COMMUNICATIONS*, 17(4):27–32.
- Debbabi, F., Jmal, R., Fourati, L. C., and Ksentini, A. (2020). Algorithmics and Modeling Aspects of Network Slicing in 5G and Beyonds Network: Survey. *IEEE Access*, 8:162748–162762.
- ETSI (2014). Network Functions Virtualisation (NFV) Management and Orchestration.
- Gutierrez-Estevez et al. (2019). Artificial Intelligence for Elastic Management and Orchestration of 5G Networks. *IEEE Wireless Communications*, 26(5):134–141.
- Haeri, S. and Trajković, L. (2018). Virtual Network Embedding via Monte Carlo Tree Search. *IEEE Transactions on Cybernetics*, 48(2):510–521.
- Han, B., Feng, D., and Schotten, H. D. (2019). A Markov Model of Slice Admission Control. *IEEE Networking Letters*, 1(1):2–5.
- Li, R. et al. (2018). Deep Reinforcement Learning for Resource Management in Network Slicing. *IEEE Access*, 6:74429–74441.
- NGMN Alliance (2016). Description of Network Slicing Concept.
- Quang, P. T. A., Hadjadj-Aoul, Y., and Outtagarts, A. (2019). A Deep Reinforcement Learning Approach for VNF Forwarding Graph Embedding. *IEEE Transactions on Network and Service Management*, 16(4):1318–1331.
- Saad, W., Bennis, M., and Chen, M. (2020). A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems. *IEEE Network*, 34(3):134–142.
- Zanzi, L. et al. (2021). LACO: A Latency-Driven Network Slicing Orchestration in Beyond-5G Networks. *IEEE Transactions on Wireless Communications*, 20(1):667–682.