

UNIVERSIDADE FEDERAL DO PAMPA

Alexandre Aprato Ferreira da Costa

**Estratégias de Controle de Fluxo na Computação na Borda: Uma Revisão
Sistemática**

**Alegrete
Janeiro de 2025**

Alexandre Aprato Ferreira da Costa

**Estratégias de Controle de Fluxo na Computação na Borda: Uma Revisão
Sistemática**

Dissertação apresentada ao Programa de Pós-graduação Stricto Sensu em Engenharia Elétrica da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Fábio Diniz Rossi

Coorientador: Marcelo Caggiani Luizelli

Alegrete
Janeiro de 2025

Ficha catalográfica elaborada automaticamente com os dados fornecidos
pelo(a) autor(a) através do Módulo de Biblioteca do
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

C837e Costa, Alexandre Aprato Ferreira da
Estratégias de controle de fluxo na computação na borda:
uma revisão sistemática / Alexandre Aprato Ferreira da Costa.
70 p.

Dissertação(Mestrado)-- Universidade Federal do Pampa,
MESTRADO EM ENGENHARIA ELÉTRICA, 2025.

"Orientação: Fábio Diniz Rossi".

1. Computação na Borda. 2. Agendamento de Fluxos. 3.
Revisão Sistemática. I. Título.

ALEXANDRE APRATO FERREIRA DA COSTA

ESTRATÉGIAS DE CONTROLE DE FLUXO NA COMPUTAÇÃO NA BORDA: Uma Revisão Sistemática

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Mestre em Engenharia Elétrica.

Dissertação defendida e aprovada em: 22/01/2025

Banca examinadora:

Prof. Dr. Fábio Diniz Rossi

Orientador

(IFFar)

Prof. Dr. Marcelo Caggiani Luizelli

Coorientador

(Unipampa)

Prof. Dr. Paulo Silas Severo de Souza

(Unipampa)

Prof. Dr^a. Josiane Fontoura dos Anjos
(IFFar)



Assinado eletronicamente por **PAULO SILAS SEVERO DE SOUZA, PROFESSOR DO MAGISTERIO SUPERIOR**, em 18/02/2025, às 12:49, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **FABIO DINIZ ROSSI, PESSOAL VOLUNTÁRIO**, em 26/02/2025, às 17:10, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **Josiane Fontoura dos Anjos, Usuário Externo**, em 27/02/2025, às 16:15, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **MARCELO CAGGIANI LUIZELLI, PROFESSOR DO MAGISTERIO SUPERIOR**, em 20/03/2025, às 15:28, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



A autenticidade deste documento pode ser conferida no site https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1672857** e o código CRC **F335561A**.

RESUMO

A ascensão dos dispositivos móveis e da Internet das Coisas (em inglês: *Internet of Things - IoT*) impulsionou a demanda por casos de uso emergentes, caracterizados por requisitos rigorosos de desempenho. Nesse contexto, a Computação na Borda (em inglês: *Edge Computing*) surge como uma solução para equilibrar os recursos limitados dos dispositivos móveis com a escalabilidade da nuvem. Este trabalho visa apresentar uma revisão sistemática sobre agendamento de fluxos de rede (em inglês: *Flow Scheduling*) em infraestruturas de borda, dividido em uma análise quantitativa e uma análise qualitativa. A análise quantitativa explora a distribuição temporal e geográfica dos trabalhos publicados, bem como informações sobre afiliações dos autores, citações e outros indicadores bibliométricos, enquanto a análise qualitativa busca entender e categorizar as abordagens propostas na literatura por meio de uma taxonomia. Espera-se que esta pesquisa contribua para o avanço das soluções de *Flow Scheduling* em *Edge Computing*, ao sintetizar os estudos existentes sobre o tema, identificar lacunas na literatura e apontar oportunidades para futuras investigações.

Palavras-chave: Computação na Borda, Agendamento de Fluxos, Revisão Sistemática.

ABSTRACT

The rise of mobile devices and the Internet of Things has driven the demand for emerging use cases characterized by stringent performance requirements. In this context, Edge Computing emerges as a solution to balance the limited resources of mobile devices with the scalability of the cloud. This work aims to present a systematic review of flow scheduling in edge infrastructure, divided into a quantitative analysis and a qualitative analysis. The quantitative analysis explores the temporal and geographical distribution of published works, as well as information on authors' affiliations, citations, and other bibliometric indicators, while the qualitative analysis seeks to understand and categorize the approaches proposed in the literature through a taxonomy. It is expected that this research will contribute to the advancement of flow scheduling solutions in Edge Computing by synthesizing existing studies on the subject, identifying gaps in the literature, and pointing out opportunities for future investigations.

Keywords: Edge Computing, Flow Scheduling, Systematic Review.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo de serviços <i>Cloud Computing</i>	16
Figura 2 – Interação entre recursos de nuvem, servidores de borda e dispositivos finais.	19
Figura 3 – Arquitetura da tecnologia <i>SDN</i>	20
Figura 4 – Metodologia de revisão apresentada em Brereton et al.	25
Figura 5 – <i>String</i> de busca usada na revisão.	27
Figura 6 – Metodologia utilizada para filtrar artigos durante a revisão.	28
Figura 7 – Número de publicações ao longo do tempo.	29
Figura 8 – Investimento em P&D e produção científica dos países.	31
Figura 9 – Distribuição da autoria.	31
Figura 10 – Número de citações e de referências.	33
Figura 11 – Meios de publicação.	34
Figura 12 – Taxonomia proposta que organiza as pesquisas de <i>Flow Scheduling</i> em ambientes de <i>Edge Computing</i> e <i>Fog Computing</i>	39

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	15
2.1	<i>Cloud Computing</i>	15
2.2	<i>Edge Computing</i>	18
2.3	<i>Software Defined Networking</i>	19
3	METODOLOGIA	23
3.1	Motivação	23
3.2	Revisões Relacionadas	24
3.3	Protocolo de Busca	25
3.3.1	Planejamento	26
3.3.2	Execução	27
3.3.3	Relatório	28
4	REVISÃO QUANTITATIVA DA LITERATURA	29
4.1	Artigos Publicados	29
4.2	Distribuição Geográfica dos Trabalhos	29
4.3	Afiliação dos Autores	32
4.4	Citações e Referências	32
4.5	Forma de Publicação	33
4.6	Revistas e Conferências mais Relevantes	34
4.7	Palavras-chave mais Utilizadas	34
4.8	Trabalhos mais Influentes	37
5	REVISÃO QUALITATIVA DA LITERATURA	39
5.1	Taxonomia	39
5.2	Infraestrutura de Rede	40
5.2.1	Infraestruturas com Fio	40
5.2.2	Infraestruturas sem Fio	41
5.2.3	Infraestruturas Mistas	42
5.2.4	Resumo e Conclusões	43
5.3	Objetivo	44
5.3.1	Controle de Congestionamento	44
5.3.2	Conformidade com <i>SLAs</i>	46
5.3.3	Economizar Energia	47
5.3.4	Resumo e Conclusões	48
5.4	Métrica	49
5.4.1	Eficiência de Tempo	49

5.4.2	Utilização de Recursos	50
5.4.3	Resumo e Conclusões	51
5.5	Técnica	52
5.5.1	Modelo Matemático	52
5.5.2	Heurística	53
5.5.3	Meta-heurística	53
5.5.4	Aprendizado de Máquina	54
5.5.5	Resumo e Conclusões	54
5.6	Validação	55
5.6.1	Simulação/Emulação	55
5.6.2	Experimentação	57
5.6.3	Modelagem Formal	59
5.6.4	Resumo e Conclusões	59
5.7	Resumo e Conclusões Gerais	62
6	CONSIDERAÇÕES FINAIS	65
6.1	Direções Futuras	65
	REFERÊNCIAS	67

1 INTRODUÇÃO

Nos últimos anos, observou-se uma proliferação exponencial de dispositivos móveis juntamente com o crescimento da *IoT* [1]. Esse avanço tecnológico tem fomentado casos de uso com requisitos rigorosos de desempenho, como serviços de vídeo *streaming*, soluções para saúde digital, jogos *online* e a integração de tecnologias em cidades inteligentes [2].

A intensificação desses casos de uso tem gerado uma reflexão sobre a localização ideal para processar os dados gerados. Em um lado do espectro, temos a nuvem, oferecendo recursos virtualmente infinitos [3], porém, sujeitos a problemas de tempo de resposta, como alta latência e saturação do *backhaul*. Por outro lado, os dispositivos dos usuários finais proporcionam baixíssima latência, mas possuem recursos limitados, como capacidade de processamento e autonomia de bateria.

Diante desse dilema, surge a ideia de *Edge Computing* [4, 5], que visa aproximar o processamento dos usuários finais. Essa proximidade é alcançada por meio de recursos computacionais estrategicamente posicionados entre a nuvem e os dispositivos finais. Dentre as características de infraestruturas de borda, destacam-se a descentralização da infraestrutura, a limitação computacional de servidores e a utilização de infraestruturas de rede suscetíveis à instabilidade e falhas [6, 7].

A fragilidade inerente à rede de borda exige uma abordagem de controle refinado. A fim de assegurar que a proximidade não seja comprometida por uma utilização não otimizada dos recursos, adota-se uma estratégia baseada em *Software-Defined Networking (SDN)* [8]. Essa abordagem proporciona visibilidade total da rede, controle holístico e refinado dos recursos. Exemplos práticos incluem o roteamento para baixa latência, o balanceamento de carga para evitar sobrecarga em pontos específicos e uma rápida recuperabilidade, tudo viabilizado pela visibilidade oferecida por *SDN*.

Este trabalho de mestrado apresenta uma revisão sistemática sobre *Flow Scheduling* em *Edge Computing*, compreendendo: **(i) análise quantitativa**, que explora a distribuição temporal e geográfica dos trabalhos publicados, das afiliações dos autores, de informações sobre citações e outros indicadores bibliométricos, e **(ii) análise qualitativa**, que busca entender e categorizar as abordagens existentes através de uma taxonomia. Assim, este estudo visa fornecer uma visão ampla do campo, sintetizando os achados existentes, identificando lacunas e destacando direções potenciais para investigações futuras.

O restante deste documento está organizado como segue. O Capítulo 2 discute os fundamentos conceituais que sustentam este trabalho. O Capítulo 3 delinea o escopo da pesquisa e a metodologia adotada, enquanto nos Capítulos 4 e 5, são apresentadas, respectivamente, a Revisão Quantitativa da Literatura e a Revisão Qualitativa da Literatura, proporcionando uma análise das tendências e das abordagens dos estudos selecionados. Por fim, o Capítulo 6 reúne as Considerações Finais, que incluem uma visão geral dos principais achados, discussões pertinentes e as direções futuras para pesquisas nesta área.

2 REFERENCIAL TEÓRICO

Este capítulo fundamenta os conceitos essenciais que delineiam a base deste trabalho, abordando a Computação em Nuvem (em inglês: *Cloud Computing*) (Seção 2.1), *Edge Computing* (Seção 2.2) e *SDN* (Seção 2.3), destacando a interação entre esses paradigmas e enfatizando a relevância em como as políticas de controle de fluxo, integradas às *SDNs*, desempenham um papel fundamental na otimização da comunicação em ambientes que demandam requisitos estritos de latência e largura de banda.

2.1 CLOUD COMPUTING

A popularização do modelo baseado em serviços no âmbito de Tecnologia da Informação (TI) reflete a consolidação da *Cloud Computing* como a principal abordagem para hospedagem de aplicações [9]. A nuvem visa prover recursos computacionais sob demanda através da *Internet*, utilizando um modelo de serviços. Essa tendência é respaldada pelas características chave da *Cloud Computing*, tais como o *on-demand self-service* e a elasticidade, que permitem a disponibilização ágil e escalonável de recursos de TI, atendendo às necessidades variáveis das organizações de forma eficiente e econômica [10].

Os modelos de serviço de nuvem são uma formalização preestabelecida das diferentes abordagens para fornecer serviços de *Cloud Computing* [11]. Esses modelos incluem *Infrastructure as a Service (IaaS)*, *Platform as a Service (PaaS)* e *Software as a Service (SaaS)*, onde:

- O modelo *IaaS* é caracterizado pelo fornecimento de recursos de infraestrutura, como processamento, armazenamento e redes, permitindo que os usuários implantem e gerenciem seus próprios sistemas operacionais e aplicativos. Exemplos de provedores de *IaaS* incluem *Amazon Web Services (AWS)* e *Microsoft Azure*. Em comparação com o modelo de *on-premises*, onde a infraestrutura e os aplicativos são mantidos localmente pela organização, o modelo *IaaS* proporciona maior flexibilidade e escalabilidade, além de reduzir os custos operacionais, uma vez que a responsabilidade pela infraestrutura e manutenção é transferida para o provedor de nuvem.
- O modelo *PaaS* é caracterizado pelo fornecimento de plataformas de desenvolvimento e execução de aplicativos, permitindo que os usuários desenvolvam, testem e implantem aplicativos sem se preocuparem com a infraestrutura subjacente. Exemplos de provedores de *PaaS* incluem *Google App Engine* e *Heroku*. Em comparação com o modelo de *on-premises*, o modelo *PaaS* oferece maior agilidade e eficiência no desenvolvimento de aplicativos, além de reduzir os custos operacionais, uma vez que a responsabilidade pela infraestrutura e manutenção também é transferida para o provedor de nuvem.

- O modelo *SaaS* é caracterizado pelo fornecimento de aplicativos de software prontos para uso, acessíveis pela *Internet*, eliminando a necessidade de instalação e manutenção local. Exemplos de aplicativos *SaaS* incluem *Salesforce* e *Office 365*. Em comparação com o modelo de *on-premises*, o modelo *SaaS* oferece maior flexibilidade e escalabilidade, além de reduzir os custos operacionais, uma vez que a responsabilidade pela infraestrutura e manutenção mais uma vez é transferida para o provedor de nuvem.

Complementando a discussão teórica, a Figura 1, apresenta exemplos representativos de cada opção de modelo de serviço.

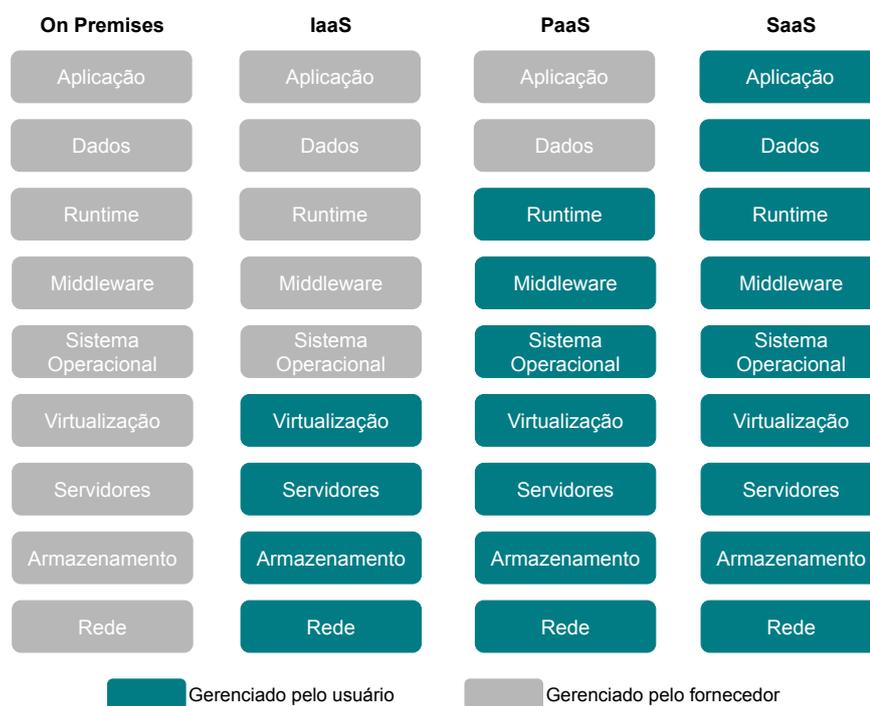


Figura 1 – Modelo de serviços *Cloud Computing*.

Explorando a relação intrínseca entre a virtualização e a *Cloud Computing*, destacamos a significativa implementação das propriedades dinâmicas do modelo de nuvem por meio dessas tecnologias. A ideia central da virtualização consiste em desacoplar os componentes de *hardware* dos componentes de *software* [12], alcançando isso por meio de sistemas especializados que gerenciam eficientemente os recursos computacionais. Em linhas gerais, as tecnologias de virtualização possibilitam a criação de ambientes computacionais isolados em nível de *software*. Essa abordagem oferece flexibilidade e escalabilidade, permitindo que recursos físicos sejam compartilhados e alocados dinamicamente conforme a demanda, sem a necessidade de configurações específicas de *hardware*. A virtualização torna possível a execução de múltiplos sistemas operacionais e aplicativos em uma única infraestrutura física, promovendo a eficiência na utilização de recursos e facilitando a implementação ágil de serviços na nuvem. O desacoplamento entre *hardware* e *software* proporcionado

pela virtualização é essencial para otimizar a alocação de recursos, melhorar a eficiência operacional e viabilizar a oferta de serviços de nuvem altamente flexíveis e adaptáveis às demandas variáveis dos usuários.

As principais tecnologias de virtualização, Máquinas Virtuais (VMs) e *Containers*, desempenham papéis distintos na implementação de ambientes virtualizados em *Cloud Computing*. VMs, em um conceito de alto nível, representam a virtualização de sistemas operacionais completos, sendo gerenciadas por um *hypervisor* que virtualiza os recursos físicos do *hardware*.

Uma vantagem significativa das VMs é o alto nível de isolamento proporcionado, permitindo a execução de diferentes sistemas operacionais e aplicativos de forma independente. No entanto, essa abordagem também apresenta desvantagens, destacando-se o *overhead* associado à virtualização completa de sistemas operacionais, o qual pode resultar em uma utilização menos eficiente dos recursos físicos. Por outro lado, os *Containers*, em um conceito de alto nível, representam uma forma mais leve de virtualização, encapsulando apenas o aplicativo e suas dependências, compartilhando o mesmo *kernel* do sistema operacional hospedeiro. *Container runtimes*, como *Docker*, desempenham um papel importante nesse contexto, facilitando a criação, execução e gerenciamento de *Containers*.

Uma das principais vantagens dos *Containers* é a velocidade de provisionamento, pois podem ser iniciados em questão de segundos, tornando-os ideais para ambientes dinâmicos e escaláveis. Contudo, a abordagem de *Containers* apresenta um nível de isolamento mais reduzido em comparação com VMs, o que pode representar uma preocupação em termos de segurança, especialmente em cenários nos quais é necessária uma separação mais robusta entre diferentes cargas de trabalho. Assim, a escolha entre VMs e *Containers* na implementação de ambientes virtualizados depende das características específicas de cada aplicação e dos requisitos de isolamento e eficiência operacional desejados.

A operação de infraestruturas de nuvem, especialmente em larga escala, demanda uma atenção para garantir sua eficiência e disponibilidade. Essas infraestruturas, devido à necessidade de lidar com enormes volumes de dados e tráfego, requerem um cuidado especial em termos de manutenção, englobando aspectos como refrigeração, suprimento de energia ininterrupta e gestão eficaz de recursos. Para atender a essas demandas, as empresas adotam a centralização por meio de instalações especializadas denominadas *data centers*. Os *data centers*, nesse contexto, funcionam como o núcleo vital para o armazenamento, processamento e distribuição de dados em larga escala. Suas características gerais destacam-se pela concentração de servidores, armazenamento e demais infraestruturas em um único local, proporcionando eficiência operacional, facilitando a manutenção centralizada e otimizando o uso de recursos. A centralização em *data centers* é fundamental para garantir a confiabilidade e a escalabilidade necessárias para sustentar as demandas crescentes das infraestruturas de nuvem, permitindo, assim, a entrega consistente e eficaz de serviços na era de *Cloud Computing*.

2.2 EDGE COMPUTING

A popularização de aplicações com requisitos estritos de desempenho, impulsionada pelo avanço de *IoT* e pelo uso generalizado de dispositivos móveis, tem desafiado as limitações do modelo de nuvem. A centralização dos recursos de computação em *data centers* remotos resulta em latências significativas devido ao número de saltos necessários para a comunicação entre as fontes de dados e os centros de dados em nuvem. Essa latência conflita com as demandas de tempo real dessas aplicações, impactando negativamente a experiência do usuário. Além disso, as limitações de largura de banda de entrada e saída nos *data centers* em nuvem reduzem o desempenho geral da rede. Diante desses desafios, o conceito emergente de *Edge Computing* surge como uma alternativa promissora.

Edge Computing, envolve a distribuição de dispositivos de computação em rede próximos aos dispositivos finais, superando as limitações de latência e largura de banda associadas à *Cloud Computing* centralizada. Além disso, *Edge Computing* permite que as aplicações sejam executadas mesmo quando a conexão com a nuvem é interrompida, garantindo a continuidade do serviço. Essa abordagem descentralizada oferece a promessa de atender aos requisitos de desempenho das aplicações, proporcionando uma experiência mais ágil e responsiva para os usuários finais.

A interconexão entre os recursos de *Cloud Computing*, os dispositivos de borda e os dispositivos finais, representada pela Figura 2, forma um modelo de computação em três camadas. Nos *data centers* em nuvem (Camada 1), a escalabilidade de computação e armazenamento é virtualmente infinita, com uma infraestrutura de rede estável e mecanismos de segurança robustos. As instalações de borda (Camada 2) estão distribuídas próximas às fontes de dados, oferecendo menor latência em comparação com a Camada 1, embora com escalabilidade limitada devido a restrições de espaço, refrigeração e fornecimento de energia. Por fim, a Camada 3 abriga os produtores e consumidores de dados, como dispositivos móveis e sensores *IoT*. A coordenação entre esses recursos permite que as aplicações sejam executadas mais próximas dos dispositivos finais, reduzindo a latência e melhorando a eficiência do sistema.

Na borda, os recursos computacionais são distribuídos geograficamente, permitindo que o processamento de dados ocorra mais próximo das fontes de informação, reduzindo a latência e melhorando a eficiência da rede. No entanto, os recursos computacionais na borda são limitados em comparação com os *data centers* em nuvem, resultando em capacidades de processamento e largura de banda mais restritas. Além disso, a rede na borda é geralmente menos robusta do que a infraestrutura de rede estável encontrada nos *data centers* em nuvem. Em contraste, os *data centers* em nuvem oferecem recursos computacionais centralizados e escaláveis, com uma infraestrutura de rede robusta e segura. Os recursos na nuvem são geralmente mais poderosos e podem ser facilmente escalados para atender às demandas de processamento e armazenamento. No entanto, as limitações da borda requerem uma gestão mais detalhada dos recursos computacionais, levando em

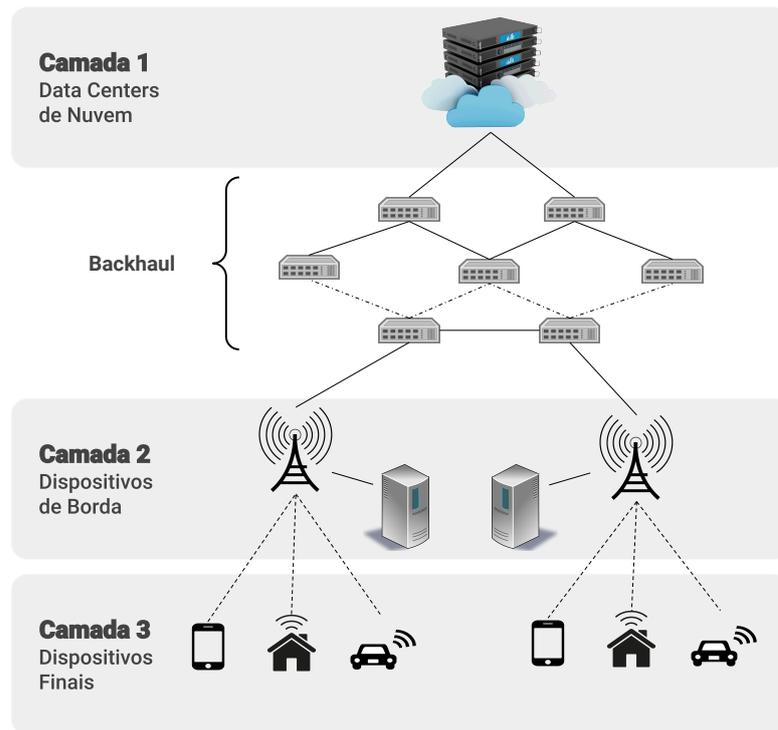


Figura 2 – Interação entre recursos de nuvem, servidores de borda e dispositivos finais.

consideração as restrições de espaço, refrigeração e fornecimento de energia, bem como a necessidade de garantir a robustez da rede em locais remotos. A gerência fina dos recursos computacionais na borda é essencial para garantir a eficiência e a segurança do sistema.

Além das abordagens tradicionais de *Edge Computing*, também existe o conceito de *Fog Computing*, que expande *Edge Computing* ao adicionar uma camada intermediária entre os dispositivos de borda e os *data centers* em nuvem. Enquanto *Edge Computing* processa dados diretamente em dispositivos de borda próximos à fonte de origem, a *Fog Computing* estende essa capacidade ao distribuir recursos de processamento e armazenamento em nós intermediários, conhecidos como pontos de presença de *fog*, localizados em locais estratégicos. No entanto, neste trabalho, optamos por tratar *Edge Computing* e *Fog Computing* como conceitos semelhantes, uma vez que ambos compartilham o objetivo de processar dados próximos à fonte e são considerados por alguns autores como implementações alternativas do mesmo conceito geral [13]. Portanto, ao longo desta revisão sistemática, não faremos distinção entre *Edge Computing* e *Fog Computing*, concentrando-nos em examinar as estratégias de *Flow Scheduling* em ambientes de borda, independentemente da terminologia utilizada.

2.3 SOFTWARE DEFINED NETWORKING

As infraestruturas de rede tradicionais apresentam limitações significativas devido ao design estático e predefinido das funcionalidades dos componentes de rede, muitas vezes

configurados de acordo com as preferências dos fornecedores, resultando em dispositivos opacos e de difícil gerenciamento. Além disso, a impossibilidade de adaptação *on-the-fly* das políticas de gerência de recursos, conforme as necessidades dinâmicas do ambiente, é uma consequência direta da natureza “caixa preta” desses dispositivos, dificultando a flexibilidade e a eficiência operacional.

Com a adoção das *SDNs* [8], ocorre o desacoplamento entre o plano de dados e o plano de controle (Figura 3). Anteriormente, ambos estavam intimamente integrados aos dispositivos de encaminhamento, enquanto, com a adoção do *SDN*, os dispositivos de encaminhamento passam a se concentrar exclusivamente no encaminhamento de dados. As decisões de gerência de recursos, anteriormente realizadas localmente, são transferidas para entidades externas denominadas controladores, por meio das interfaces *Southbound* e *Northbound*, permitindo uma abordagem mais dinâmica e centralizada para a administração da rede.

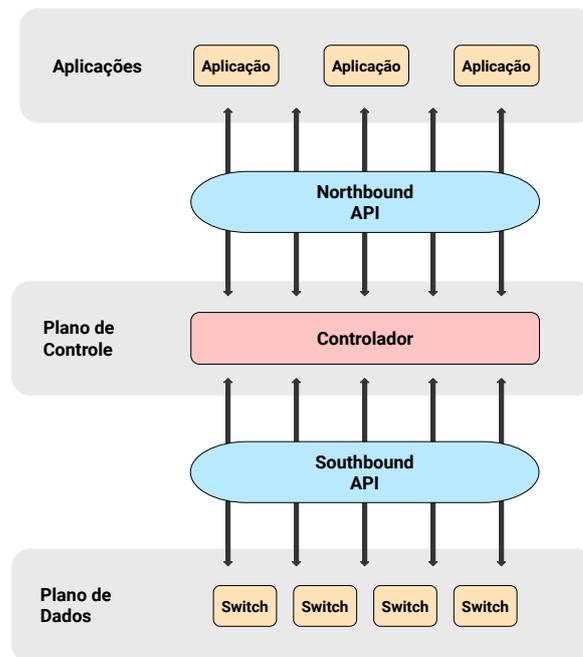


Figura 3 – Arquitetura da tecnologia *SDN*.

A camada controladora desempenha um papel fundamental na arquitetura *SDN*, fornecendo serviços que possibilitam um plano de controle distribuído, além de promover os conceitos de gerenciamento de estado e centralização. Adicionalmente, as *SDNs* oferecem diversas oportunidades para otimizar recursos por meio de políticas aprimoradas de controle de fluxo.

O controle de fluxos, que envolve a seleção e encaminhamento de pacotes de dados, é essencial para garantir a Qualidade de Serviço (*QoS*) para diferentes tipos de tráfego. Em cenários onde aplicações possuem requisitos estritos de latência e largura de banda, as políticas de controle de fluxo podem otimizar a comunicação, garantindo a alocação justa

e eficiente dos recursos de rede, como a largura de banda disponível e o espaço de *buffer*, para diferentes classes de serviço. Essas políticas visam assegurar a justa distribuição dos recursos, além de proteger a rede contra comportamentos não conformes, garantindo que usuários bem-comportados não sejam prejudicados por usuários que excedem suas taxas de dados acordadas.

O conceito de *SDN* emergiu como uma solução inovadora para superar as limitações de infraestruturas de rede tradicionais. A capacidade de desacoplar o plano de dados do plano de controle proporciona uma flexibilidade sem precedentes, permitindo que os operadores de TI configurem e ajustem dinamicamente políticas de gerenciamento de rede de acordo com as necessidades específicas da borda. Essa configurabilidade avançada de redes que operam sob o modelo de *SDN* possibilita a aplicação de políticas otimizadas de controle de fluxos e outras estratégias de gerenciamento de rede, melhorando significativamente o desempenho e a eficiência operacional da infraestrutura como um todo. Como resultado, as *SDNs* apresentam um grande potencial para infraestruturas de borda, fornecendo uma base sólida para a evolução e inovação contínuas no campo de *Edge Computing*.

3 METODOLOGIA

Este capítulo apresenta a motivação, as pesquisas relacionadas e a metodologia adotada no estudo. Primeiramente, são discutidos os fatores que justificam a pesquisa, seguidos por uma análise de importantes trabalhos da literatura em áreas correlatas, que fornecem uma base conceitual para este estudo. Por fim, descreve-se a abordagem metodológica utilizada para alcançar os objetivos propostos, detalhando os caminhos percorridos ao longo da pesquisa.

3.1 MOTIVAÇÃO

A crescente popularização de aplicações com requisitos estritos de desempenho tem evidenciado as limitações do modelo de nuvem, especialmente em relação à latência, devido ao distanciamento entre os *data centers* e os dispositivos finais dos usuários. *Edge Computing* surge como uma alternativa promissora, expandindo o modelo de nuvem e aproximando os recursos computacionais dos dispositivos finais, o que se mostra essencial para atender às demandas por baixa latência e alta eficiência de rede.

Os desafios de gerência de redes na borda são significativos, uma vez que as infraestruturas de borda apresentam características que as tornam mais suscetíveis a falhas em comparação com os *data centers* em nuvem. Isso se deve à presença de servidores com menor capacidade computacional e em menor número, infraestrutura descentralizada, conexão por meio de redes, mesmo sem recursos necessários para assegurar sua robustez, e restrições de espaço, ventilação e suprimento de energia em algumas instalações. Essas limitações tornam a gerência de redes na borda um desafio, exigindo soluções inovadoras para garantir a eficácia e a confiabilidade das infraestruturas de borda.

Este trabalho de mestrado se baseia em duas observações. Primeiramente, reconhecemos que as políticas de gerenciamento de recursos de rede, incluindo *Flow Scheduling*, desempenham um papel fundamental na garantia da eficácia das infraestruturas de borda, mesmo que estas sejam mais propensas a falhas em comparação com os *data centers* em nuvem. Além disso, observamos que a quantidade de trabalhos existentes na literatura e a complexa relação entre *Edge Computing* e outros paradigmas de computação representam uma barreira significativa para novos pesquisadores e desenvolvedores. Identificar as políticas de gerenciamento de recursos mais recentes, as métricas relevantes, os principais grupos de pesquisa e os problemas em aberto torna-se uma tarefa desafiadora.

Baseado nesse contexto, torna-se evidente a necessidade de um estudo sistemático que mapeie de forma compreensiva a literatura sobre políticas de gerenciamento de recursos de rede, com foco especial no *Flow Scheduling* em *Edge Computing*. Esse estudo não apenas preencherá lacunas no conhecimento existente, mas também fornecerá uma base sólida para futuras pesquisas e desenvolvimento de soluções inovadoras para melhorar a eficiência e confiabilidade das redes na borda.

3.2 REVISÕES RELACIONADAS

Neste momento, discutiremos algumas revisões relacionadas ao tema. Recentemente, tem sido observado um aumento significativo no número de pesquisas sobre *Edge Computing*, abordando o assunto a partir de diferentes perspectivas. Diversos estudos e análises foram publicados, refletindo o crescente interesse e importância desse campo emergente. Essas pesquisas têm contribuído para o avanço do conhecimento e para a identificação de desafios e oportunidades na área.

Alguns trabalhos abordam os novos paradigmas de *Edge Computing* e orquestração *IoT* com *SDN* e borda. Enquanto Tang et al. [14] destacam a importância da integração entre computação e rede na Convergência de Computação e Rede, enfatizando a necessidade de estratégias inteligentes de agendamento para lidar com os desafios de recursos heterogêneos, otimização de múltiplos objetivos e tarefas diversas, Rafique et al. [15] discutem a mudança dos modelos de computação centralizada para a *Edge Computing* distribuída. Essa transição busca atender aos requisitos de serviços sensíveis à latência e contextuais das aplicações *IoT*. Além disso, são apresentadas análises sobre a orquestração de *IoT* com *SDN* e borda, destacando requisitos-chave, esforços de padronização, estudos de caso e classificação de avanços em *SDIoT-Edge Computing* com base em múltiplos parâmetros de desempenho. Vulnerabilidades de segurança e privacidade são abordadas, e direções para pesquisas futuras são fornecidas.

Outros estudos exploram diferentes aspectos de *Edge Computing*, *SDN* e *IoT*. Mamane et al. [16] analisam o estado da arte da alocação de recursos de rádio em redes 5G e futuras, oferecendo uma visão da literatura existente e destacando trabalhos futuros. Luo et al. [17] abordam o agendamento de recursos em *Edge Computing*, discutindo arquiteturas, modelos unificados, técnicas de agendamento e desafios futuros. Vijayalayan et al. [18] apresentam esquemas de agendamento distribuído para redes *mesh* sem fio, enquanto Wang et al. [19] exploram como a *SDN* pode aprimorar *Edge Computing*, destacando desafios e pesquisas futuras.

Jazaeri et al. [20] fornecem uma revisão sistemática de questões, desafios e soluções em *Edge Computing* em redes *SDN-IoT*, destacando a interdependência entre *SDN* e *IoT*. A pesquisa explora como esses dois desenvolvimentos fundamentais na rede se relacionam e propõe soluções para melhorar a eficiência na prestação de serviços *IoT* nesse contexto. O estudo oferece uma visão das questões atuais, desafios e direções futuras para pesquisas eficazes em serviços *IoT* no paradigma *SDIoT-Edge*.

Taleb et al. [21] apresentam uma pesquisa sobre o *Edge Computing* e se concentram nas principais tecnologias facilitadoras. Elabora a orquestração do *Edge Computing* considerando tanto serviços individuais quanto uma rede de plataformas *Edge Computing* que suportam a mobilidade, fornecendo *insights* sobre as diferentes opções de implantação de orquestração. Além disso, analisam a arquitetura de referência do *Edge Computing* e os principais cenários de implantação, os quais oferecem suporte multilocalização para

desenvolvedores de aplicações, provedores de conteúdo e terceiros. O estudo também aprofunda os desafios de pesquisa em aberto, fornecendo uma visão do cenário atual e das oportunidades futuras de pesquisa.

Durante esta revisão, foram abordadas diversas pesquisas sobre *Edge Computing*. No entanto, é importante ressaltar que nenhuma delas se compara à singularidade e abordagem apresentadas em nosso estudo, como apresentado na Tabela 1. Ao destacar as contribuições únicas e perspectivas diferenciadas em nossa pesquisa, evidenciamos a relevância e o potencial impacto significativo que nosso trabalho pode trazer para o campo de *Flow Scheduling* em *Edge Computing*.

Tabela 1 – Comparação do escopo da revisão e estudos anteriores.

Trabalho	Paradigma	Tipo de Rede		Foco	Análise Cientométrica
		Sem Fio	Cabeada		
Tang et al.[14]	Genérico	✓	✗	Scheduling (Geral)	✗
Mamane et al.[16]	<i>Edge Computing</i>	✓	✗	Scheduling (Geral)	✗
Luo et al.[17]	<i>Edge Computing</i>	✓	✗	Scheduling (Geral)	✗
Vijayalayan, Harwood e Karunasekera[18]	Genérico	✓	✗	Scheduling (Geral)	✗
Wang et al.[19]	<i>Edge Computing</i>	✗	✓	Geral	✗
Rafique et al.[15]	Genérico	✓	✗	Geral	✗
Jazaeri et al.[20]	<i>Edge Computing</i>	✓	✗	Geral	✗
Taleb et al.[21]	<i>Edge Computing</i>	✓	✗	Geral	✗
Este Trabalho	<i>Edge Computing</i>	✓	✓	Scheduling (Fluxos)	✓

3.3 PROTOCOLO DE BUSCA

Esta revisão sistemática adota a metodologia de pesquisa apresentada por Brereton et al. [22], a qual segmenta o processo de revisão em três fases: Planejamento, Execução e Relatório, como apresentado na Figura 4. Embora essa abordagem tenha sido inicialmente concebida para o campo de Engenharia de Software, ela oferece um guia genérico para a identificação e síntese de estudos de pesquisa pertinentes. Como resultado, ela foi incorporada em diversas áreas, abrangendo domínios que se entrelaçam com a abrangência deste trabalho, tais como *Cloud Computing* e *Edge Computing*.

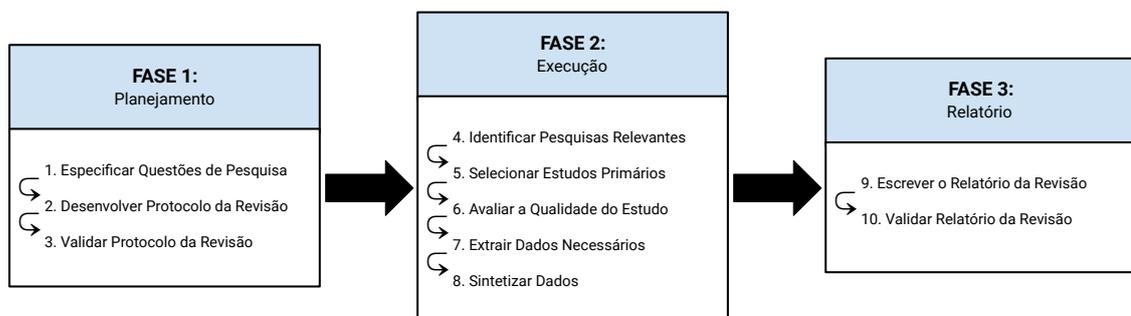


Figura 4 – Metodologia de revisão apresentada em Brereton et al.

3.3.1 PLANEJAMENTO

Esta revisão adota três questões de pesquisa (QPs), apresentadas na Tabela 2. Com relação à QP1, abordada durante a análise quantitativa, buscamos mensurar o número de trabalhos publicados nos temas de pesquisa abordados. No tocante à QP2, o objetivo foi identificar as características relevantes dos estudos selecionados, onde engloba uma taxonomia organizando a pesquisa selecionada de acordo com suas características. Por fim, a QP3 busca identificar oportunidades de pesquisa, destacando os desafios em aberto no âmbito do tema investigado.

Tabela 2 – Lista de questões de pesquisa abordadas nesta revisão.

Identificador	Questão de Pesquisa
QP1	Quantos artigos no tópico abordado foram publicados entre 2017 e 2023?
QP2	Quais são as métricas de interesse, estratégias e principais abordagens de avaliação no âmbito do tópico abordado?
QP3	Quais são os desafios e questões em aberto no âmbito do tópico abordado?

Esta análise se concentra em investigar estudos acadêmicos catalogados no mecanismo de busca *Scopus* (www.scopus.com), o qual apresenta um banco de dados extenso que indexa diversas editoras, incluindo *Springer*, *ScienceDirect* e *Wiley*. A base de dados *Scopus* apresenta vantagens significativas em comparação com outras fontes bibliográficas. Em primeiro lugar, a quantidade de artigos relacionados à *Cloud Computing* englobados pelo *Scopus* é quase o dobro da encontrada no *WoS*, incluindo inclusive publicações em pré-impressão de uma variedade considerável de periódicos. Em segundo lugar, o *Scopus* oferece recursos avançados para exportação de dados estruturados, incluindo citações e informações bibliográficas, juntamente com resumos, palavras-chave e referências.

A composição da *string* de busca projetada para recuperar trabalhos de pesquisa na base de dados selecionada é ilustrada na Figura 5. A formulação dessa *string* reflete a atenção dada à inclusão de termos, com o objetivo de identificar abordagens no contexto de *Flow Scheduling* em *Edge Computing*. A seleção criteriosa de termos, como *Flow Management*, *Flow Shaping*, *Flow Control*, *Flow Scheduling*, *Traffic Management*, *Traffic Shaping*, *Traffic Control*, *Traffic Scheduling*, *Bandwidth Management*, *Bandwidth Shaping*, *Bandwidth Control*, *Bandwidth Scheduling*, *Congestion Management*, *Congestion Shaping*, *Congestion Control* e *Congestion Scheduling*, visa capturar as diversas nuances desses campos específicos.

Além disso, destaca-se que a busca foi conduzida considerando tanto a *Edge Computing* quanto a *Fog Computing*. Esta abordagem ampliada visa assegurar a inclusão de trabalhos que explorem essas facetas interconectadas da computação distribuída, proporcionando uma visão holística sobre *Flow Scheduling* em ambientes de borda e névoa.

```
("Flow Management" OR "Flow Shaping" OR "Flow Control" OR "Flow Scheduling" OR "Traffic Management" OR "Traffic Shaping" OR "Traffic Control" OR "Traffic Scheduling" OR "Bandwidth Management" OR "Bandwidth Shaping" OR "Bandwidth Control" OR "Bandwidth Scheduling" OR "Congestion Management" OR "Congestion Shaping" OR "Congestion Control" OR "Congestion Scheduling") AND ("Edge Computing" OR "Fog Computing")
```

Figura 5 – *String* de busca usada na revisão.

A complexidade inerente a essas disciplinas exige uma busca criteriosa, e a *string* foi concebida com o intuito de alcançar o máximo possível de contribuições relevantes para a revisão sistemática em questão.

Após a coleta dos artigos recuperados pelos mecanismos de busca, foram adotados um conjunto de critérios de inclusão e exclusão, visando refinar e selecionar os resultados obtidos. Os critérios de inclusão adotados priorizaram estudos primários, aqueles que oferecem contribuições diretas e inovadoras à área, em contraposição à mera revisão de pesquisas anteriores. Estabelecemos o período de publicação entre os anos de 2017 e 2023, visto que nenhum estudo anterior a 2017 foi encontrado. Os estudos foram considerados elegíveis se apresentassem, no mínimo, quatro páginas escritas em inglês.

Para garantir a integridade da revisão, foram implementados critérios de exclusão que eliminaram documentos não enquadrados como artigos de pesquisa regulares, incluindo capítulos de livros, relatórios técnicos e patentes. Duplicatas do mesmo artigo também foram excluídas para evitar redundâncias. Além disso, foram excluídos estudos que não apresentassem uma validação explícita das soluções propostas, assegurando a qualidade e confiabilidade dos resultados considerados. Finalmente, foram excluídos documentos cujo foco principal não estava centrado em *Flow Scheduling* em *Edge Computing*, garantindo uma seleção precisa e alinhada aos objetivos específicos desta revisão sistemática.

3.3.2 EXECUÇÃO

A etapa subsequente do processo de pesquisa resultou na identificação de 540 entradas. Após a coleta da amostra inicial, a filtragem dos artigos foi conduzida, seguindo a metodologia detalhada na Figura 6. Na fase inicial de triagem, os critérios de inclusão e exclusão foram aplicados aos títulos dos artigos, promovendo uma redução significativa no número de estudos selecionados, que passou a ser de 83. Posteriormente, os mesmos critérios foram empregados na análise dos resumos dos artigos, resultando em uma redução adicional para 56 estudos. Por fim, uma análise de todo o manuscrito foi realizada em relação aos critérios de inclusão e exclusão, culminando na compilação da lista final de revisão, composta por 43 artigos que atendiam integralmente aos requisitos estabelecidos.

Essa abordagem escalonada visa assegurar a qualidade e relevância dos estudos considerados nesta revisão sistemática.

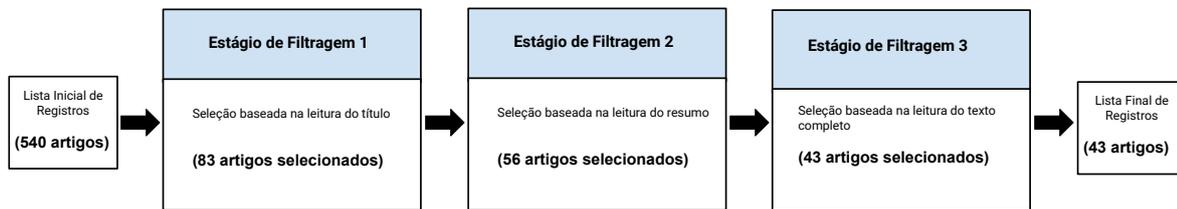


Figura 6 – Metodologia utilizada para filtrar artigos durante a revisão.

3.3.3 RELATÓRIO

A presente dissertação avança para uma análise dos resultados obtidos, dividida em duas abordagens complementares: quantitativa e qualitativa. Esses aspectos serão explorados em detalhes nos próximos capítulos, proporcionando uma compreensão dos dados coletados.

No Capítulo 4, será apresentada a análise quantitativa, com foco na mensuração e interpretação de dados numéricos obtidos durante a pesquisa. Essa análise busca identificar padrões, tendências e correlações relevantes, apoiando-se em métodos estatísticos.

Já o Capítulo 5 trará a análise qualitativa, que tem como objetivo interpretar os aspectos mais subjetivos dos estudos revisados, oferecendo uma visão mais profunda sobre o conteúdo e as contribuições das publicações. Essa abordagem permitirá discutir não apenas os números, mas também o significado e impacto dos resultados no contexto da área investigada.

Essas duas análises, apresentadas nos capítulos seguintes, oferecem uma base para a interpretação crítica dos dados, contribuindo para o fortalecimento das conclusões desta revisão sistemática.

4 REVISÃO QUANTITATIVA DA LITERATURA

Nesta seção, exploraremos uma análise dos aspectos quantitativos relacionados aos estudos examinados. Investigaremos a distribuição dos artigos publicados ao longo do tempo e em diferentes regiões geográficas, além de examinar a afiliação dos autores para compreender o contexto institucional das pesquisas. Também analisaremos o número de citações e referências em cada estudo, fornecendo *insights* sobre sua relevância e impacto. Além disso, investigaremos a forma de publicação dos artigos, considerando diferentes tipos de periódicos e conferências. Essa análise quantitativa nos permitirá obter uma compreensão da literatura existente sobre o tema em questão, incluindo aspectos como as principais palavras-chave e os artigos mais influentes.

4.1 ARTIGOS PUBLICADOS

Entre os anos de 2017 e 2023, foram identificados 43 artigos, abrangendo o escopo desta pesquisa sobre *Flow Scheduling* em *Edge Computing*. Esses números respondem à (QP1). A Figura 7 ilustra a distribuição anual desses artigos, oferecendo uma visão concisa da evolução temporal das contribuições.

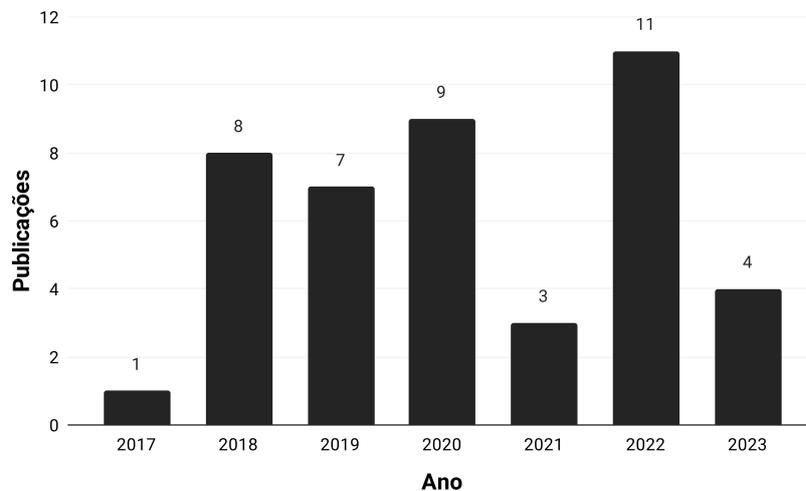


Figura 7 – Número de publicações ao longo do tempo.

4.2 DISTRIBUIÇÃO GEOGRÁFICA DOS TRABALHOS

Com o intuito de aprofundar a compreensão dos padrões de contribuição, analisamos de maneira mais detalhada a distribuição das publicações por país e os padrões de autoria. Para essa análise, foram consideradas as nacionalidades de cada autor, e, ao final, foram identificadas as dez nacionalidades mais frequentes. Com base nesses dados, calculou-se a porcentagem de autores de cada nacionalidade em relação ao total das dez principais. Conforme apresentado na Tabela 3 (R.: *Rank*), a maioria das pesquisas sobre

Flow Scheduling em *Edge Computing* é conduzida por pesquisadores chineses (38,73%) e do Reino Unido (8,09%).

Dado que esses números, por si só, não proporcionam uma análise completa da relevância das contribuições por nacionalidade, também elaboramos uma classificação de países contribuintes para as publicações que são citadas com maior frequência. Os valores na extremidade direita da Tabela 3 indicam que a maioria dos autores da China (45,24%) e da Índia (11,90%) contribuíram com publicações amplamente reconhecidas.

Tabela 3 – Países contribuintes (Esquerda: Top 10 países contribuintes; Direita: Publicações citadas pelo menos 18 vezes.

R.	País	%	Gasto P&D (bi USD)	R.	País	%	Gasto P&D (bi USD)
1	China	38,73	514.798	1	China	45,24	514.798
2	Reino Unido	8,09	51.702	2	Índia	11,90	58.691
3	Coréia do Sul	6,94	100.055	3	Alemanha	9,52	131.932
4	Índia	6,36	58.691	3	Coréia do Sul	9,52	100.055
5	Egito	5,78	6.2	5	Reino Unido	7,14	51.702
6	Austrália	5,20	23.3	6	Portugal	4,76	3.6
6	Paquistão	5,20	2.4	7	Japão	2,38	172.614
8	Estados Unidos	2,89	612.714	7	Áustria	2,38	14.945
8	Taiwan	2,89	42.945	7	Estados Unidos	2,38	612.714
10	Alemanha	2,31	131.932	7	Austrália	2,38	23.3

Os números de investimento em pesquisa e desenvolvimento (P&D) para os países citados foram obtidos de diferentes fontes reconhecidas. Para a China, Reino Unido, Estados Unidos, Japão, Alemanha, Coreia do Sul, Taiwan e Áustria, os dados foram extraídos da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), especificamente do indicador de Gastos Internos Brutos em P&D (*Gross domestic spending on R&D*), referentes ao ano de 2019, enquanto para Portugal os dados sobre gastos internos brutos em P&D também foram provenientes da OCDE, porém referentes ao ano de 2014¹. No caso da Índia, o investimento em P&D como percentual do Produto Interno Bruto (GDP) é baseado em dados de 2018². Para a Austrália, os números foram obtidos do *Australian Bureau of Statistics*, relacionados ao Gasto Bruto em P&D em 2014³. Para o Egito e o Paquistão, os números de investimento em P&D como percentual do GDP correspondem a dados de 2013⁴.

Ao investigar a correlação entre o investimento em P&D, realizado por esses países, e o número de artigos científicos publicados sobre o tema, observa-se uma correlação moderada entre o montante investido e o volume de publicações, conforme evidenciado

¹ Disponível em: <<<https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm>>>. Acesso em: 26/04/2024

² Disponível em: <<<https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS?locations=IN>>>. Acesso em: 26/04/2024.

³ Disponível em: <<<https://www.abs.gov.au/ausstats/abs@.nsf/Previousproducts/8104.0Main%20Features42013-14>>>. Acesso em: 26/04/2024.

⁴ Disponível em: <<<http://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS>>>. Acesso em: 26/04/2024.

na Figura 8[A]. Ao analisar o investimento em P&D e sua possível correlação com as publicações mais citadas, observamos através da Figura 8[B] outra correlação moderada entre esses indicadores. Esses resultados sugerem que o simples investimento financeiro em P&D por parte destes países não garantiu automaticamente um aumento no número de publicações nem necessariamente a produção de publicações de maior impacto ou relevância no contexto deste trabalho.

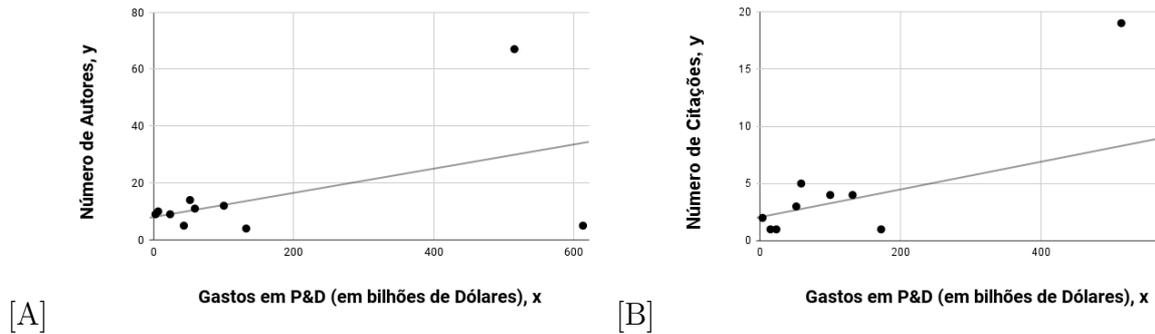


Figura 8 – Investimento em P&D e produção científica dos países.

No que diz respeito à autoria das publicações, a média de autores por publicação nos últimos sete anos está retratada na Figura 9. A distribuição da autoria revela que, para mais da metade das publicações, o número de coautores varia entre dois e quatro. Juntamente com a percentagem relativamente significativa de publicações com cinco ou mais autores, a distribuição evidencia que a colaboração pode apresentar algumas vantagens sobre a pesquisa realizada por investigadores individuais. A natureza interdisciplinar da investigação em *Cloud Computing* pode ser apontada como uma das razões primárias para o predomínio dos trabalhos conjuntos.

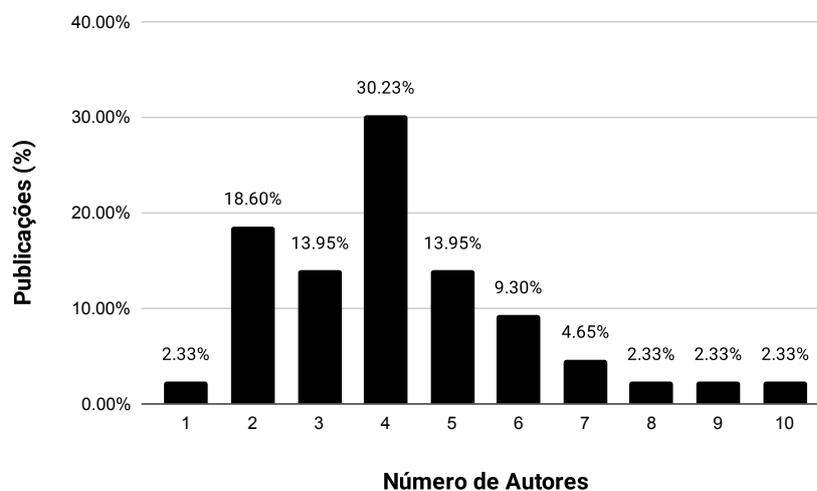


Figura 9 – Distribuição da autoria.

4.3 AFILIAÇÃO DOS AUTORES

A análise da distribuição institucional dos pesquisadores envolvidos nos estudos revisados é fundamental para compreender o panorama da pesquisa em *Flow Scheduling* em *Edge Computing*. Esta análise nos permite identificar padrões de colaboração entre instituições e países, destacar centros de pesquisa líderes neste campo e até mesmo revelar áreas geográficas emergentes de interesse. Entender a afiliação dos autores fornece contexto sobre o ambiente institucional em que o trabalho foi realizado, o que pode influenciar na interpretação dos resultados e na avaliação da credibilidade e relevância do estudo. Por exemplo, pesquisas conduzidas em instituições de renome ou em parceria com empresas reconhecidas podem ser vistas como mais confiáveis ou impactantes. Esta análise da afiliação dos autores enriquece nossa compreensão da rede de colaboração, do contexto institucional e da qualidade da pesquisa realizada neste campo promissor, e pode ser observada na Tabela 4.

Tabela 4 – Principais instituições de pesquisa.

R.	Instituição	País	Pesquisadores
1	Hong Kong Polytechnic University	China	8
2	Newcastle University	Reino Unido	6
3	Menoufia University	Egito	5
3	Nanjing University	China	5
3	National University of Defense Technology	China	5
3	National Yang Ming Chiao Tung University	Taiwan	5

Isso revela os principais centros de pesquisa envolvidos na investigação sobre *Flow Scheduling* em *Edge Computing*, destacando a dominância da China na área, seguida pelo Reino Unido, conforme já evidenciado na subseção 4.2. Em destaque, a Hong Kong Polytechnic University, da China, lidera com o maior número de pesquisadores contribuindo para este campo. Em seguida, a Newcastle University, do Reino Unido, também se destaca como uma instituição relevante. O terceiro lugar é compartilhado pela Nanjing University e pela National University of Defense Technology, ambas chinesas, além da National Yang Ming Chiao Tung University, de Taiwan, e da Menoufia University, do Egito, todas elas com um igual número de colaboradores. Esses dados destacam a diversidade geográfica das instituições envolvidas e fornecem *insights* sobre os principais *players* na pesquisa em *Flow Scheduling* em *Edge Computing*.

4.4 CITAÇÕES E REFERÊNCIAS

A seguir, exploramos os padrões de referências em publicações com uma bibliografia substancial (Figura 10). Cada barra representa a relação entre o número total de publicações e a média de referências dessas publicações, divididas em segmentos ao longo do eixo x

que representam o número mínimo de citações. Por exemplo, uma publicação citada por 100 ou mais publicações possui, em média, 31 referências.

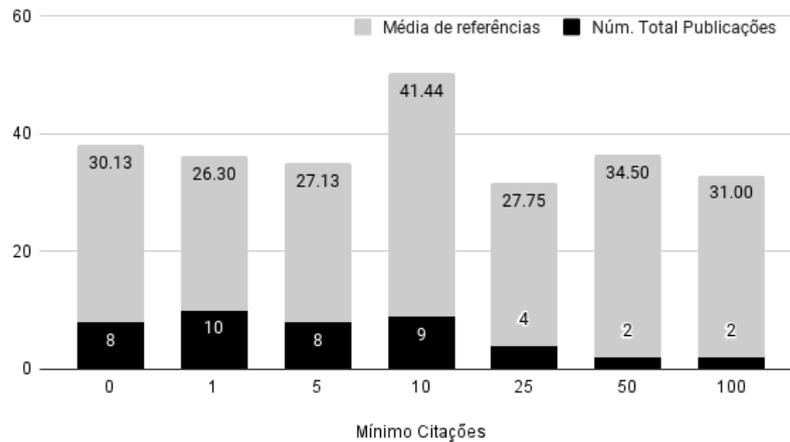


Figura 10 – Número de citações e de referências.

Observamos que, em geral, as publicações mantiveram uma média de referências bastante consistente em todas as faixas de citações consideradas. Entretanto, notamos uma exceção significativa nos trabalhos que possuem entre 10 e 24 citações, onde a média de referências foi notavelmente mais elevada, com destaque para um artigo em particular que apresentou um número substancialmente maior de referências em comparação com os demais. A abrangência de uma revisão bibliográfica substancial é reconhecida como um critério fundamental para pesquisas de alta qualidade [23]. Embora essas conclusões se baseiem em evidências empíricas robustas, é essencial considerar que as restrições frequentes de espaço em periódicos podem impactar o número máximo de páginas por publicação, influenciando, assim, o número de referências.

4.5 FORMA DE PUBLICAÇÃO

A escolha do meio de publicação muitas vezes desempenha um papel importante na visibilidade e no impacto de um artigo. Dessa forma, é relevante examinar as preferências dos pesquisadores quanto ao tipo de veículo utilizado para comunicar suas ideias e descobertas à comunidade científica. Com a especificação do tipo de documento para todos os artigos de *Flow Scheduling* em *Edge Computing* analisados, torna-se possível explorar a distribuição dos tipos de documentos em um determinado campo de pesquisa. De acordo com a Figura 11, a maioria dos artigos sobre o assunto, 62.79%, é preferencialmente publicada por meio de revistas científicas.

A preferência dos pesquisadores por publicar em revistas em vez de conferências pode ser atribuída principalmente ao prestígio associado às publicações em revistas de alto impacto e aos requisitos institucionais que valorizam esse tipo de publicação. Publicar

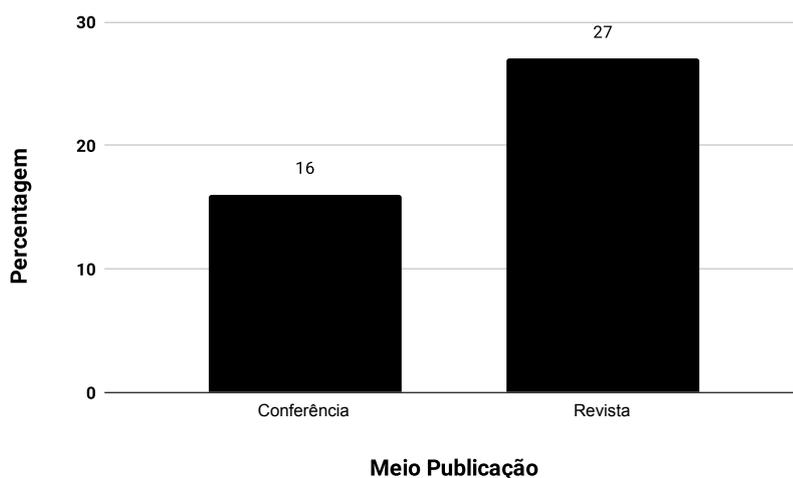


Figura 11 – Meios de publicação.

em revistas renomadas dentro de uma área acadêmica específica é reconhecido como um indicador de excelência e contribuição significativa para o campo [24]. Tais publicações tendem a ter um público mais amplo e uma maior influência na comunidade científica, o que pode aumentar a visibilidade e a reputação dos pesquisadores. Além disso, muitas instituições acadêmicas e programas de pesquisa consideram as publicações em revistas de renome um critério para avaliar o progresso e o mérito dos pesquisadores, o que influencia diretamente oportunidades de financiamento, promoções e reconhecimento profissional.

4.6 REVISTAS E CONFERÊNCIAS MAIS RELEVANTES

Neste momento, analisamos os padrões de citações em conferências e periódicos relevantes na área de estudo. Utilizando um método de contagem direta, investigamos como as citações estão distribuídas entre diferentes fontes acadêmicas, visando distinguir claramente entre o impacto da pesquisa e a produtividade. Concentramo-nos em identificar as conferências e periódicos que recebem um maior número de citações, proporcionando uma visão sobre os locais onde a pesquisa nesta área é mais influente e reconhecida pela comunidade científica. A Tabela 5 compreende os resultados da apuração.

4.7 PALAVRAS-CHAVE MAIS UTILIZADAS

As palavras-chave desempenham um papel importante na representação e categorização do conteúdo abordado em um artigo científico. Elas servem como uma ferramenta eficaz para encapsular os principais temas e aspectos relacionados a uma área específica de pesquisa. A análise das palavras-chave oferece *insights* valiosos sobre a evolução e os tópicos emergentes dentro de um determinado campo ao longo do tempo. Além disso, ao investigar as coocorrências entre as palavras-chave, é possível identificar conexões e

Tabela 5 – Revistas e conferências mais influentes.

R.	Local de Publicação	Forma	Citações
1	IEEE Internet of Things Journal	Revista	133
1	IEEE Transactions on Industrial Informatics	Revista	133
3	IEEE Transactions on Parallel and Distributed Systems	Revista	80
4	IEEE Journal on Selected Areas in Communications	Revista	68
5	Journal of Ambient Intelligence and Humanized Computing	Revista	42
6	IEEE International Conference on Communications	Conferência	40
7	Sensors (Switzerland)	Revista	36
8	International Symposium on Wireless Communication Systems	Conferência	31
9	Wireless Communications and Mobile Computing	Revista	24
10	IEEE/ACM Transactions on Networking	Revista	20

relações significativas entre diferentes aspectos do estudo.

Durante a análise dos dados, foram identificadas 147 palavras-chave exclusivas nos campos de palavras-chave do autor e de índice dos artigos originais analisados. É relevante observar que nem sempre os autores fornecem palavras-chave para seus artigos. Em casos em que as palavras-chave estão ausentes, profissionais especializados em indexação atribuem palavras-chave com base em tesouros específicos. Isso revela que, para 85,11% das publicações analisadas, as palavras-chave foram especificadas pelos autores. Em contraste, 14,89% das publicações foram identificadas com palavras-chave atribuídas pela *Scopus* ou não receberam nenhuma atribuição de palavras-chave. Quando as palavras-chave dos autores não estão disponíveis, as palavras-chave do índice são utilizadas para análise de palavras-chave comuns e agrupamentos.

Quanto à distribuição média do número de palavras-chave por publicação, observa-se que é comum o uso de quatro a cinco palavras-chave para representar o tema central de um artigo. Essa prática não se limita à *Edge Computing*, já que editores geralmente estabelecem um número mínimo e/ou máximo de palavras-chave por publicação. Para mitigar a variabilidade nos termos das palavras-chave, editores frequentemente fornecem uma lista predefinida de palavras-chave padronizadas que são relevantes para um periódico ou conferência específica.

Na Tabela 6 são apresentadas as palavras-chave utilizadas com maior frequência. Os resultados da revisão destacam as áreas-chave de investigação nas pesquisas recentes. Evidentemente, os estudos concentram-se principalmente em *Edge Computing* e suas aplicações, como *IoT* e *Cloud Computing*. A análise revela um foco significativo na otimização de recursos e no agendamento de tarefas, visando melhorar a eficiência e a *QoS*. Além disso, questões relacionadas a eficiência energética emergem como preocupações fundamentais. A conexão entre esses temas e outros campos de pesquisa, como *SDN* e tecnologias 5G, sugere uma abordagem integrada e interdisciplinar para resolver os desafios em *Edge Computing*.

Na Tabela 7, é apresentada a distribuição temporal das palavras-chave mais recorrentes nos estudos analisados. Essa análise longitudinal permite observar as tendências

Tabela 6 – Palavras-chave mais utilizadas.

R.	Palavra-Chave	f	R.	Palavra-Chave	f
1	edge computing	10	16	traffic scheduling	3
2	internet of things	7	18	lyapunov optimization	2
2	SDN	7	18	software defined networking	2
4	fog computing	6	18	collaborative edge computing	2
4	MEC	6	18	software defined networking (sdn)	2
4	internet of things (IoT)	6	18	bandwidth allocation	2
7	task scheduling	4	18	openflow	2
8	QoS	4	18	congestion control	2
8	multi-access edge computing	3	18	data center networks	2
8	resource allocation	3	18	network resource management	2
8	computation offloading	3	18	software-defined networking	2
8	5G	3	18	traffic management	2
8	cloud	3	18	IoT	2
8	mobile edge computing	3	18	energy efficiency	2
8	network flow scheduling	3	18	fairness	2
8	quality of service	3			

de pesquisa em *Flow Scheduling* em *Edge Computing*, evidenciando como as áreas de interesse evoluíram ao longo dos anos.

Tabela 7 – Palavras-chave por ano.

Palavra-chave	2018	2019	2020	2021	2022	2023
edge computing	-	3	1	-	5	1
internet of things	1	3	1	2	-	-
SND	2	1	3	1	-	-
fog computing	2	1	1	1	1	-
MEC	2	-	2	-	2	-
internet of things (IoT)	-	1	-	-	3	2
task scheduling	-	1	1	-	2	-
QoS	2	1	-	-	-	1
mobile edge computing	1	1	1	-	1	-
multi-access edge computing	-	1	1	-	1	-
computation offloading	-	1	-	-	1	1
5G	-	2	1	-	-	-
cloud	-	-	-	1	2	-
network flow scheduling	-	1	-	1	1	-
traffic scheduling	-	-	1	-	1	1
quality of service	-	1	-	2	-	-
software-defined networking	1	-	-	1	-	-
resource allocation	-	2	1	-	-	-
collaborative edge computing	-	1	-	1	-	-
software defined networking (SDN)	-	-	-	-	1	1
bandwidth allocation	-	-	-	-	2	-
openflow	-	-	-	-	2	-
congestion control	1	1	-	-	-	-
data center networks	1	1	-	-	-	-
network resource management	1	-	1	-	-	-
traffic management	1	1	-	-	-	-
IoT	-	1	1	-	-	-
energy efficiency	-	1	1	-	-	-
fairness	-	1	-	-	-	1

4.8 TRABALHOS MAIS INFLUENTES

Na sequência, exploramos os artigos identificados em nossa revisão que receberam um número significativo de citações, fornecendo uma análise do impacto alcançado por essas contribuições na literatura científica. Para avaliar a relevância e influência desses estudos, utilizamos índices baseados em citações individuais, considerando tanto citações de periódicos quanto de conferências. A Tabela 8 retrata estes estudos.

Tabela 8 – Publicações mais citadas.

R	Referência	Título	Citações
1	Ning et al.[25]	Joint Computation Offloading, Power Allocation, and Channel Assignment for 5G-Enabled Traffic Management Systems	123
2	Sahni, Cao e Yang[26]	Data-aware task allocation for achieving low latency in collaborative edge computing	122
3	Sahni et al.[27]	Multi-Hop Multi-Task Partial Computation Offloading in Collaborative Edge Computing	80
4	Ge et al.[28]	QoE-Assured 4K HTTP live streaming via transient segment holding at mobile edge	68
5	Diro, Reda e Chilamkurti[29]	Differential flow space allocation scheme in SDN based fog computing for IoT applications	42
6	Aujla et al.[30]	An Ensembled Scheme for QoS-Aware Traffic Flow Management in Software Defined Networks	40
7	Hong et al.[31]	QCon: QoS-aware network resource management for fog computing	36
8	Nasimi et al.[32]	Edge-Assisted Congestion Control Mechanism for 5G Network Using Software-Defined Networking	31
9	Zhang et al.[33]	Adaptive Configuration Selection and Bandwidth Allocation for Edge-Based Video Analytics	20
10	Sun et al.[34]	Efficient algorithm for traffic engineering in Cloud-of-Things and edge computing	18

Durante a análise dos trabalhos mais citados, percebemos certas similaridades em relação às soluções propostas. Em linhas gerais, é possível dividir tais artigos em três grupos de contribuições:

- **Gerenciamento de tarefas e *offloading* de computação em *Collaborative Edge Computing*:** Alguns artigos abordam a alocação de tarefas e o *offloading* de computação em contextos de *Collaborative Edge Computing*. Sahni et al. [26] propõe uma alocação de tarefas consciente de dados para minimizar a latência em *Collaborative Edge Computing*, enquanto Sahni et al. [27] explora o *offloading* parcial de computação multi-salto e agendamento de fluxo de rede em *Collaborative Edge Computing*.
- **Gerenciamento de recursos de rede em *Fog Computing* e *SDN* para suportar aplicações *IoT*:** Outros trabalhos discutem o gerenciamento de recursos de rede em *Fog Computing* e *SDN* para suportar aplicações *IoT*. Diro et al. [29] propõe um esquema de alocação diferencial de espaço de fluxo para aplicações heterogêneas de *IoT* em *Fog Computing* com *SDN*, Hong et al. [31] apresentam um

framework para gerenciamento de recursos de rede em *Fog Computing*, e Ning et al. [25] et al. abordam o *offloading* de computação, alocação de energia e atribuição de canais para sistemas de gerenciamento de tráfego habilitados para 5G.

- **Controle de congestionamento e alocação de largura de banda em *Edge Computing* para melhorar a QoS em tempo real:** Outras abordagens se destacam por tratarem do controle de congestionamento e da alocação de largura de banda em *Edge Computing* para melhorar a QoS em tempo real. Nasimi et al. [32] apresentam um mecanismo de controle de congestionamento em *Edge Computing* para gerenciar o tráfego de dados em tempo real, Zang et al. [33] abordam a seleção de configuração e alocação de largura de banda para otimizar a análise de vídeo em tempo real na borda e Ning et al. abordam o *offloading* de computação, alocação de energia e atribuição de canais para sistemas de gerenciamento de tráfego habilitados para 5G.

Além disso, as pesquisas restantes também apresentam contribuições importantes para o gerenciamento de tráfego e *Edge Computing*, como o esquema de gerenciamento de fluxos de vídeo em 4K via *HTTP* proposto por Ge et al. [28], a abordagem combinada proposta por Aujla et al. [30] para gerenciar o fluxo de tráfego de forma eficiente e garantir a QoS em *SDNs*, e o esquema proposto por Sun et al. [34] para otimizar o gerenciamento de tráfego em redes de vários domínios, incluindo *Edge Computing*.

5 REVISÃO QUALITATIVA DA LITERATURA

Nesta seção, realizaremos uma análise qualitativa dos estudos sobre *Flow Scheduling* em *Edge Computing*, complementando os aspectos quantitativos abordados anteriormente. Nosso objetivo é categorizar e compreender as estratégias, metodologias e resultados apresentados por esses estudos. Essa análise qualitativa nos permitirá identificar padrões, tendências e contribuições significativas dos estudos revisados na área de *Flow Scheduling* em *Edge Computing*.

5.1 TAXONOMIA

Flow Scheduling em *Edge Computing* abrange uma diversidade de atividades e estratégias. Para estruturar e compreender as contribuições existentes, categorizamos os estudos revisados de acordo com uma taxonomia representada na Figura 12. Essa taxonomia agrupa os esforços de pesquisa com base em características fundamentais:

- **Infraestrutura de Rede:** Indica o tipo de rede considerado nos estudos.
- **Objetivo:** Reflete os focos principais dos estudos de *Flow Scheduling*.
- **Métrica:** Indica as medidas de desempenho utilizadas nos estudos.
- **Técnica:** Descreve os métodos empregados para implementar o *Flow Scheduling*.
- **Validação:** Indica como os estudos foram validados ou testados.

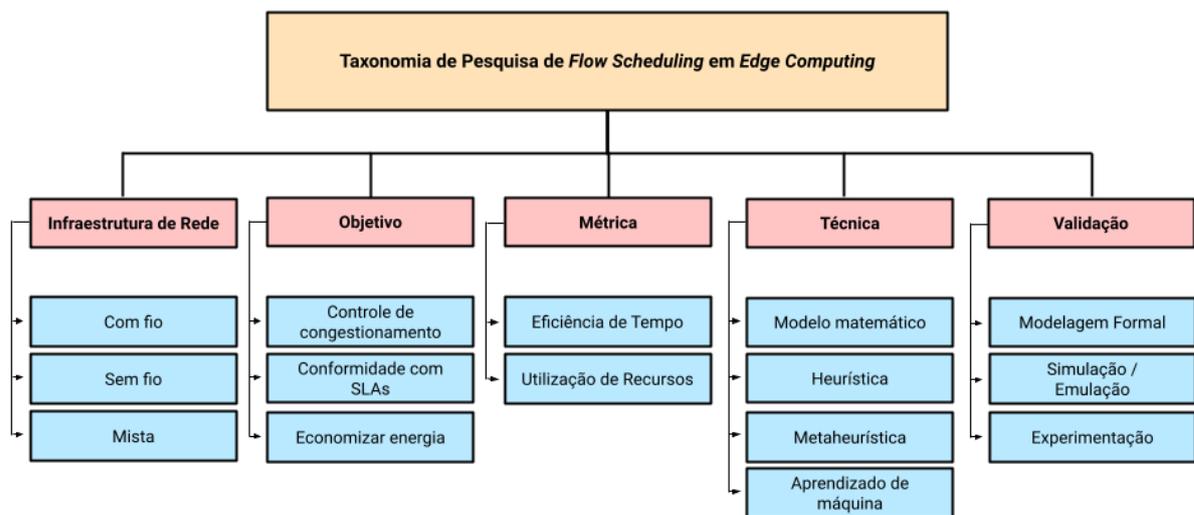


Figura 12 – Taxonomia proposta que organiza as pesquisas de *Flow Scheduling* em ambientes de *Edge Computing* e *Fog Computing*.

A abordagem de taxonomia utilizada adota o modelo facetado para analisar estudos de *Flow Scheduling* em *Edge Computing* [35]. Nesta estrutura, diferentes facetas capturam aspectos específicos dos estudos revisados. Por exemplo, um artigo de pesquisa pode apresentar uma estratégia de *Flow Scheduling* voltada para a Conformidade com *SLAs* (em inglês: *Service Level Agreement*) como seu objetivo principal, avaliada por métricas como eficiência de tempo ou utilização de recursos. A implementação dessa estratégia pode envolver o uso de heurísticas como técnica principal e foi validada por meio de experimentação. Essa abordagem facilita uma categorização dos estudos revisados e está alinhada com práticas adotadas por estudos relacionados, como mencionado por [36] e [37]. Na sequência desta seção, serão abordadas estratégias específicas de *Flow Scheduling* para *Edge Computing*, classificando-as conforme os elementos definidos na taxonomia.

O restante desta seção discute estratégias de *Flow Scheduling* para *Edge Computing* e *Fog Computing*, categorizando-as de acordo com os ramos da taxonomia proposta.

5.2 INFRAESTRUTURA DE REDE

A infraestrutura de rede é um aspecto importante no *Flow Scheduling* em *Edge Computing*, pois define o ambiente no qual as soluções são implementadas e operam.

5.2.1 INFRAESTRUTURAS COM FIO

As redes com fio são aquelas em que os dispositivos são conectados por meio de cabos físicos, como *Ethernet* e fibra óptica. Essa configuração é frequentemente utilizada em ambientes que necessitam de conexões estáveis e de alta velocidade, essenciais para aplicações que requerem baixa latência e alta largura de banda.

A infraestrutura de *SDN* em redes com fio tem sido explorada em determinados estudos. Alharbi e Fei [38] propuseram uma arquitetura *SDN* para melhorar o *throughput* de grandes fluxos utilizando *Multipath TCP (MPTCP)*. De forma semelhante, Aujla et al. [30] abordaram a gestão de fluxo de tráfego em redes *SDN*, enquanto Wang et al. [39] focaram no gerenciamento de fluxo em *switches SDN* usando tabelas de fluxo múltiplas. Estes estudos destacaram a aplicação de redes com fio para otimizar o desempenho de redes *SDN*.

Publicações focadas no controle de congestionamento em redes com fio também foram realizados. Selvi et al. [40] exploraram a aplicação de um *framework* híbrido para controle de congestionamento *TCP Incast* em redes de *data center*, que são tipicamente ambientes com fio. De maneira semelhante, Yuan et al. [41] aplicaram um algoritmo de controle de congestionamento em sistemas de *Cloud Computing* empresariais, que tradicionalmente utilizam infraestrutura com fio.

5.2.2 INFRAESTRUTURAS SEM FIO

As redes sem fio são aquelas em que os dispositivos se conectam através de sinais de rádio ou outras formas de transmissão sem cabos físicos, como *Wi-Fi*, *Bluetooth* e redes celulares. Essa configuração é utilizada em ambientes que demandam mobilidade e flexibilidade, permitindo que os dispositivos se conectem à rede de forma dinâmica e sem a necessidade de infraestrutura física extensa. As redes sem fio são essenciais para suportar a conectividade de dispositivos móveis e aplicações que requerem acesso rápido e onipresente à *Internet* e a outros recursos de rede.

A aplicação de *Edge Computing* em redes 5G tem sido rotineiramente explorada, destacando a sua utilização em redes sem fio. Nasimi et al. [32] discutiram mecanismos de controle de congestionamento em redes 5G utilizando *Edge Computing* e *SDN*. De forma semelhante, Ismail et al. [42] focaram na implementação de um modelo *Edge Computing* eficiente em termos de energia para redes 5G, enquanto Salama et al. [43] abordaram a colaboração de *Edge Computing* com um modelo baseado em gerenciamento ativo de fila, *Active Queue Management-based Green Cloud Model*. Huang et al. [44] discutiram a integração de *Edge Computing* em redes celulares 4G/5G, e Wang et al. [45] mencionaram a aplicação do modelo *Queueing with Linear Reward and Nonlinear Workload Fairness* em sistemas *Edge Computing*, redes sem fio 5G e *crowdsourcing* móvel.

Estudos focados em *IoT* e *Fog Computing* também destacaram a importância das redes sem fio. Diro et al. [29] discutiram a integração de *SDN* e *Fog Computing* para aplicações de *IoT*, enquanto Han et al. [46] abordaram a aplicação do protocolo *Parallel UDT (PaUDT)*, que é uma extensão do protocolo *UDP-based Data Transfer Protocol (UDT)*, em *Fog Computing*, focando em dispositivos *IoT*. Moraes e Martins [47] exploraram a orquestração de tráfego *IoT* utilizando um *framework* integrado *SDN* em ambientes de *Fog Computing*, e Yin et al. [48] analisaram a transmissão cooperativa em sistemas de *Edge Computing* envolvendo dispositivos *IoT*.

A alocação de tarefas e o *offloading* em ambientes de computação colaborativa na borda foram temas recorrentes em diversos estudos. Sahni et al. [26] discutiram a alocação de tarefas em ambientes de computação colaborativa na borda, utilizando dispositivos *IoT* conectados via redes sem fio. De maneira semelhante, Sahni et al. [27] abordaram o *offloading* de computação parcial *multi-hop* em computação colaborativa na borda, e Gao et al. [49] focaram no *offloading* parcial de tarefas em um ambiente de computação colaborativa na borda utilizando *SDN*. Enquanto Yun e Chen [50] discutem uma infraestrutura de rede predominantemente sem fio, para otimizar o processamento de tarefas em *Edge Computing* no contexto da *IoT*.

Trabalhos direcionados para *streaming* e gerenciamento de tráfego em redes sem fio foram explorados. Ge et al. [28] trataram a entrega de *streaming* ao vivo em 4K usando *cache* na borda, focados em redes de acesso via rádio. Ning et al. [25] investigaram a gestão de tráfego em sistemas de gerenciamento de tráfego habilitados para 5G, enquanto Yan et

al. [51] priorizaram a programação de tráfego para serviços de *streaming* em um ambiente de *Edge Computing*.

A aplicação de *Edge Computing* em redes sem fio foi o tema central em determinados estudos. Hong et al. [31] direcionaram seus esforços ao *framework* de gerenciamento de recursos de rede consciente de *QoS*, *qCon*, em dispositivos limitados como pontos de acesso *Wi-Fi* e dispositivos móveis. Pencheva e Atanasov [52] abordaram a implementação de gerenciamento de largura de banda em ambientes de *Edge Computing*, e Xiang et al. [53] discutiram a provisão de serviços móveis utilizando *Edge Computing*, onde os servidores de borda são acessados via rede sem fio. Li et al. [54] e Dlamini e Vilakati [55] apresentam, respectivamente, um sistema baseado em uma configuração de comunicação móvel, onde os dispositivos móveis inteligentes se comunicam com a estação base através de uma conexão sem fio, e um sistema de borda que integra pequenas células inteligentes e estações base para gerenciar o tráfego de dados e a alocação de recursos de forma eficiente.

Complementando, Sellami et al. [56] projetaram uma infraestrutura para lidar com a variabilidade do tráfego *IoT*, utilizando um controlador *SDN* que coordena a alocação de recursos e o agendamento de tarefas em tempo real, utilizando rede sem fio. E Latif et al. [57] modelaram uma rede como um grafo, onde os nós representam pontos de acesso, servidores de borda e dispositivos *IoT*, e as arestas representam as conexões entre eles. A arquitetura *SDN* proporciona uma visão global da rede, facilitando a tomada de decisões sobre o *offloading* de tarefas e a programação de fluxos, enquanto a *blockchain* é utilizada para garantir a segurança e a integridade das transações de recursos. E Saranya e Sasikala [58] mencionam que a proposta assume que os dispositivos de borda se comunicam via uma rede sem fio, implicando em desafios como a variação do estado do canal sem fio e a interferência entre transmissores e receptores.

5.2.3 INFRAESTRUTURAS MISTAS

As redes mistas combinam elementos de redes com fio e sem fio, aproveitando as vantagens de ambas as tecnologias. Essa configuração é frequentemente utilizada em ambientes onde há a necessidade de conexões estáveis e de alta velocidade proporcionadas pelas redes com fio, juntamente com a mobilidade e flexibilidade das redes sem fio. Redes mistas são ideais para infraestruturas que exigem um balanceamento entre desempenho, cobertura e conveniência, suportando uma ampla gama de dispositivos e aplicações em diferentes cenários operacionais.

Algumas pesquisas têm explorado a infraestrutura de redes mistas em contextos de *Fog Computing* e *IoT*. Jiang et al. [59] discutiram o controle de tráfego em redes de névoa, abrangendo componentes com fio e sem fio. Da mesma forma, Sethon [60] abordou o gerenciamento de fluxo em redes de *Fog Computing*, utilizando redes mistas para distribuir dados mais próximos ao usuário final. Serdaroglu e Baydere [61], por sua vez, propuseram uma arquitetura escalável de *Fog Computing* para gerenciar o tráfego de pacotes de dados

IoT, integrando tanto redes com fio e quanto sem fio para melhorar a *QoS*.

Estudos focados em *Edge Computing* e *IoT* também utilizaram redes mistas para gestão de tráfego e recursos. Sun et al. [34] abordaram a gestão de tráfego em redes de *IoT* e *Edge Computing*, enquanto Habeeb et al. [62] investigaram a gestão de tráfego em ambientes de *Cloud Computing* e *Cloud Computing*, integrando redes com fio e sem fio. Zhang et al. [63] focaram no gerenciamento de tarefas em um ambiente de computação colaborativa na borda, envolvendo recursos de redes mistas para melhorar a eficiência e a distribuição de dados.

A integração de serviços de *Cloud Computing* e na borda foi um tema comum em alguns estudos. Kwak et al. [64] discutiram a gestão de tráfego em ambientes de *Fog Computing* e *data centers* distribuídos, utilizando redes com fio e sem fio. De maneira semelhante, Zhao et al. [65] abordaram a integração de serviços de *Cloud Computing* e na borda, indicando a utilização de redes mistas na arquitetura do sistema. Arshed et al. [66] propuseram um agendamento de tarefas em um ambiente de *Cloud Computing* e névoa, utilizando recursos de redes com fio e sem fio para otimizar o desempenho do sistema.

Trabalhos direcionados ao agendamento de tarefas e alocação de recursos em ambientes de *Fog Computing* e *Edge Computing* também reportaram a utilização de redes mistas. Lei e Minzhi [67] discutiram a alocação de tarefas em um ambiente de armazém inteligente utilizando *Fog Computing*, integrando tanto redes com fio quanto redes sem fio. Da mesma forma, Fahad et al. [68] abordaram o agendamento de tarefas em um ambiente de *Fog Computing*, enquanto Zang et al. [33] focaram na análise de vídeo baseada em *Edge Computing*, utilizando redes mistas para transmitir fluxos de vídeo para servidores de borda.

A combinação de *Fog Computing* e *SDN* para melhorar o *QoS* em sistemas *IoT* foi discutida por Ahammad et al. [69]. Esse estudo salienta a utilização de redes mistas na arquitetura proposta para integrar *fog nodes* com redes *SDN*, visando melhorar a *QoS* em sistemas *IoT*. De maneira complementar, Wen et al. [70] abordam a infraestrutura de rede em um contexto de *Edge Computing*, enfatizando a gestão de *streaming* de dados em ambientes de *Edge Computing*. Ele destaca a utilização de redes mistas, incluindo tanto conexões cabeadas quanto sem fio, para otimizar a alocação de largura de banda e reduzir a latência em aplicações sensíveis ao tempo, promovendo um balanceamento de carga eficaz entre nós de borda e melhorando o desempenho geral dos sistemas *IoT*.

5.2.4 RESUMO E CONCLUSÕES

Os diferentes tipos de redes apresentam vantagens e desafios distintos, adequados a diversas aplicações e necessidades tecnológicas. Redes com fio são conhecidas por suas conexões estáveis e de alta velocidade, características essenciais para ambientes que demandam baixa latência e alta largura de banda, como *data centers* e serviços de *Cloud Computing*. Por outro lado, as redes sem fio destacam-se por sua flexibilidade e mobilidade,

suportando a dinâmica de dispositivos *IoT* e ambientes móveis, embora enfrentem desafios significativos na gestão de tráfego e eficiência energética.

Redes mistas, que combinam elementos de redes com fio e sem fio, oferecem uma solução equilibrada, beneficiando-se da estabilidade das conexões com fio e da adaptabilidade das conexões sem fio. Essa combinação é ideal para infraestruturas que necessitam de alta performance e cobertura ampla, acomodando uma grande variedade de dispositivos e aplicações.

Observa-se uma tendência crescente em direção às redes sem fio, refletindo a popularidade dos dispositivos móveis e a necessidade de suportar a mobilidade em infraestruturas de rede. Este cenário sublinha a importância de pesquisas voltadas para tecnologias emergentes de redes sem fio, como 5G, 6G e redes satelitais. Estas tecnologias prometem transformar a conectividade global, mas ainda são subexploradas na literatura científica, principalmente quando comparadas às redes tradicionais e especialmente no contexto de controle de fluxos e gestão de recursos.

A adoção de tecnologias como *SDN* e *Edge Computing* tem se mostrado essencial para aprimorar a orquestração de tráfego, a eficiência energética e a capacidade de adaptação às demandas crescentes das redes contemporâneas. A continuidade e ampliação de pesquisas nessas áreas emergentes serão fundamentais para enfrentar os desafios futuros da conectividade global, assegurando que as redes possam suportar de maneira eficiente e sustentável uma gama diversificada de dispositivos e aplicações. A Tabela 9 resume os tipos de rede considerados nos estudos selecionados.

Tabela 9 – Resumo dos tipos de rede considerados nos artigos selecionados

Infraestrutura de Rede	Referência
Com Fio	[38] [30] [39] [40] [41]
Sem Fio	[31] [32] [42] [54] [55] [43] [56] [44] [45] [29] [46] [47] [48] [26] [27] [28] [25] [52] [53] [57] [58] [51] [49] [50]
Mista	[59] [60] [61] [34] [62] [63] [64] [65] [66] [67] [68] [33] [69] [70]

5.3 OBJETIVO

Os objetivos dos estudos direcionam o desenvolvimento e a avaliação das soluções propostas. No contexto de *Flow Scheduling* em *Edge Computing*, os objetivos determinam as prioridades, como o controle de congestionamento, a conformidade com *SLAs* e a eficiência energética.

5.3.1 CONTROLE DE CONGESTIONAMENTO

O controle de congestionamento visa garantir a eficiência e a estabilidade das redes de *Edge Computing*. Ele envolve técnicas que gerenciam o fluxo de dados para evitar sobrecargas, minimizar atrasos e garantir a *QoS*. Estudos acessados utilizam desde algorit-

mos de roteamento até alocação de recursos para mitigar os efeitos do congestionamento e melhorar o desempenho geral da rede.

A melhoria do *throughput* e a eficiência do controle de tráfego em redes foram o foco de algumas pesquisas. Jiang et al. [59] descreveram a modelagem de técnicas de controle de tráfego em redes de névoa utilizando o modelo *Leaky Bucket*. Similarmente, Alharbi e Fei [38] concentraram em melhorar o *throughput* de grandes fluxos em arquiteturas *SDN* utilizando *TCP multipath*. Enquanto Han et al. [46] exploraram a eficiência do controle de congestionamento com o protocolo *UDT* baseado em *UDP* em ambientes de *Fog Computing*, e Wang et al. [39] se concentraram na redução do tempo de processamento dos fluxos priorizando entradas populares visando diminuir a latência de correspondência.

Pesquisas sobre controle de congestionamento em redes de acesso e *Edge Computing* exploraram diversas técnicas para mitigar sobrecargas e melhorar o desempenho. Hong et al. [31] implementaram um mecanismo de controle de congestionamento em redes de acesso via rádio utilizando servidores *Edge Computing*. Selvi et al. [40] mitigaram o congestionamento *TCP Incast* em *data centers* combinando controle orientado pelo receptor e *Edge Computing*. Li et al. [54] focaram na estabilidade das filas através do controle de congestionamento e alocação de recursos, enquanto Salama et al. [43] abordaram a eficiência energética e latência em redes 5G utilizando um modelo colaborativo de *Edge Computing* e um modelo de nuvem verde baseado em gerenciamento ativo de filas. Ning et al. [25] incluem a maximização da taxa de *offloading* e a otimização da alocação de recursos em um sistema de gerenciamento de tráfego habilitado para 5G, que embora possa ser classificado como Controle de Congestionamento, a abordagem também pode ser vista como uma forma de Conformidade com *SLAs*, pois visa atender às exigências de latência e confiabilidade de aplicações críticas em tempo real. Na mesma linha, Zhao et al. [65] e Wen et al. [70] concentram, respectivamente, na otimização do tráfego de rede e na alocação eficiente de recursos para evitar congestionamentos e melhorar a performance da rede, embora também aborde aspectos relacionados à conformidade com *SLAs* ao otimizar a *QoS*, e no gerenciamento a alocação de largura de banda e a distribuição de tráfego para evitar a saturação dos nós de borda.

A alocação de recursos e o agendamento de tarefas foram objetivos principais em alguns estudos. Sahni et al. [26] procuraram minimizar o tempo de conclusão das aplicações através da alocação sensível aos dados e *Flow Scheduling*. Sahni et al. [27] visaram minimizar o tempo médio de conclusão das tarefas com escalonamento de fluxos e *offloading* parcial. Similarmente, Gao et al. [49] otimizaram a alocação de recursos de computação e rotas de dados, e o Wang et al. [45] abordaram a alocação de recursos entre múltiplas filas com uma garantia de equidade para resolver a congestão de dados.

A gestão de tráfego em ambientes de *Edge Computing* foi amplamente explorada. Xiang et al. [53] procuraram minimizar o tempo médio de resposta dos serviços otimizando a alocação de recursos de computação e o agendamento de tráfego. Huang et

al. [44] garantiram baixa latência para fluxos de tráfego *Edge Computing* controlando a priorização do tráfego. Latif et al. [57] focaram na otimização do *offloading* de tarefas e no balanceamento de carga entre múltiplos dispositivos de borda, enquanto o Serdaroglu e Baydere [61] abordaram o desafio do congestionamento causado pela programação de tráfego de dados multiprioritário em um sistema de névoa escalável.

Estudos voltados para o roteamento e o balanceamento de carga destacaram a importância de uma gestão eficaz do tráfego para evitar congestionamentos. Sun et al. [34] buscaram otimizar o roteamento do tráfego para resolver problemas de inconsistência entre o roteamento de sobreposição e a política de roteamento *Border Gateway Protocol*. Yan et al. [51] visaram melhorar o balanceamento de carga para otimizar a experiência do usuário e reduzir os custos operacionais dos provedores de serviços. Enquanto Yuan et al. [41] desenvolveram um algoritmo de controle de congestionamento tolerante a atrasos para melhorar o uso de recursos de rede e minimizar a perda de pacotes.

5.3.2 CONFORMIDADE COM SLAS

A conformidade com *SLAs* é essencial para assegurar que as redes de *Edge Computing* atendam aos requisitos de desempenho e qualidade esperados. Isso inclui garantir a disponibilidade, latência, e *throughput* conforme estipulado nos contratos de serviço. Estudos exploram diferentes técnicas, desde otimização de alocação de recursos até gerenciamento de tráfego, para garantir que as redes operem dentro dos parâmetros acordados e ofereçam uma experiência consistente e confiável aos usuários.

O foco na garantia da *QoS* e Qualidade de Experiência (*QoE*) para diferentes aplicações, diretamente relacionadas à conformidade com *SLAs*, foram identificados de alguns estudos. Por exemplo, Aujla et al. [30] abordaram a gestão do fluxo de tráfego para garantir a *QoS*, enquanto Diro et al. [29] se concentraram em gerenciar o espaço de fluxo para atender os requisitos de *QoS* das aplicações *IoT*. Ge et al. [28] visaram garantir a *QoE* para *streaming* ao vivo em 4K, e Hong et al. [31] focaram na gestão de recursos de rede para assegurar a *QoS* em ambientes de *Fog Computing*. De forma similar, Moraes e Martins [47] trataram da gestão de recursos de rede para garantir a *QoS* para o tráfego *IoT*.

Pesquisas voltadas para a otimização de recursos e priorização de tráfego destacaram a importância de uma gestão eficaz para o cumprimento dos *SLAs*. Yin et al. [48] visaram garantir a transmissão com prioridade e justiça, melhorando a utilização dos recursos de rede. Kwak et al. [64] focaram em garantir a *QoS* e maximizar a utilização da rede em um ambiente de *Internet of Everything*, enquanto Pencheva e Atanasov [52] propõem a extensão dos serviços de gerenciamento de largura de banda para melhorar a *QoS*. Ahammad et al. [69], por sua vez, buscou melhorar o *QoS* no sistema *IoT* através da otimização do ponto de acesso e local de processamento de dados.

A garantia de processamento em tempo real e a alocação de tarefas foram objetivos

principais em alguns trabalhos. Arshed et al. [66] se concentraram em garantir que as tarefas fossem processadas em tempo real para atender aos requisitos de tempo de atraso. Sethon [60] abordou a garantia da *QoS* em aplicações de rede através de um algoritmo de otimização de colônia de formigas, enquanto o Fahad et al. [68] focaram no equilíbrio da alocação de tarefas entre aquelas sensíveis à latência e aquelas que podem tolerar algum atraso. Saranya e Sasikala [58] visaram garantir a conformidade com os *SLAs* ao otimizar o sequenciamento de tarefas e o *offloading* parcial.

A comunicação em tempo crítico e o desempenho das aplicações também foram explorados. Habeeb et al. [62] visaram garantir a conformidade com os *SLAs* fornecendo garantias de comunicação em tempo crítico para aplicações *IoT* através do fatiamento dinâmico de largura de banda. Zhang et al. [63] buscaram otimizar o desempenho das aplicações nativas de borda, garantindo *throughput* e latência baixos conforme os requisitos de *SLA*. Zang et al. [33], por sua vez, focaram na otimização da seleção de configuração e da alocação de largura de banda para maximizar a precisão da análise de vídeo e minimizar o consumo de energia, atendendo aos requisitos de latência de serviço.

5.3.3 ECONOMIZAR ENERGIA

Economizar energia é um dos objetivos nas redes de *Edge Computing*, especialmente dado o aumento do número de dispositivos conectados e a demanda por processamento eficiente. As técnicas para redução do consumo de energia incluem desde algoritmos de gerenciamento de energia até a otimização da alocação de recursos. Pesquisas investigam como balancear o desempenho da rede com a eficiência energética, garantindo que as operações de rede sejam sustentáveis sem comprometer a *QoS*.

Estudos focados na redução do consumo de energia em redes 5G e *Edge Computing* elaboraram técnicas para melhorar a eficiência energética. Ismail et al. [42] visaram reduzir o consumo de energia e a latência em redes 5G utilizando o modelo de gerenciamento de filas ativas, combinado com *Edge Computing*. De forma semelhante, Dlamini e Vilakati [55] procuraram minimizar o consumo de energia geral e garantir a *QoS* dentro da rede através de um algoritmo de gerenciamento de rede baseado em energia verde.

A otimização da alocação de recursos e a redução do consumo de energia foram objetivos principais em alguns estudos. Sellami et al. [56] utilizaram um algoritmo de Aprendizado por Reforço Profundo para alocação eficiente de tarefas, visando minimizar o consumo de energia e a latência na rede. De forma similar, Arshed et al. [66] focaram na otimização da alocação de recursos para minimizar o tempo de execução e os custos financeiros, além de reduzir o consumo de largura de banda e energia. De forma semelhante, Yun e Chen [50] focaram na redução do custo energético associado ao processamento de tarefas em servidores *Edge Computing*.

5.3.4 RESUMO E CONCLUSÕES

A análise dos objetivos dos estudos revela uma clara orientação em torno de três principais metas: controle de congestionamento, conformidade com *SLAs* e otimização da eficiência energética.

O controle de congestionamento é essencial para garantir a eficiência e a estabilidade das redes de *Edge Computing*. Estudos exploraram desde algoritmos de roteamento até técnicas de alocação de recursos para evitar sobrecargas e melhorar o desempenho da rede. Este enfoque é fundamental para manter a *QoS* e a experiência do usuário em ambientes com alta demanda de tráfego.

A conformidade com *SLAs* assegura que as redes atendam aos requisitos de desempenho e qualidade estipulados. Pesquisas nesta área empregam a otimização de recursos e o gerenciamento de tráfego para garantir que os parâmetros acordados sejam mantidos, proporcionando uma operação consistente e confiável das redes. A atenção à *QoS* e *QoE* é destacada, evidenciando a importância de atender às expectativas dos usuários e contratos de serviço.

Economizar energia é outro objetivo crítico, especialmente dado o aumento do número de dispositivos conectados e a crescente demanda por processamento eficiente. Estudos focaram em algoritmos de gerenciamento de energia e otimização da alocação de recursos para reduzir o consumo de energia, balanceando o desempenho da rede com a sustentabilidade das operações.

Os resultados observados são coerentes com os desafios contemporâneos das infraestruturas de rede. A maioria dos artigos selecionados concentra-se na otimização do desempenho, com uma menor proporção focando na eficiência do uso de recursos, particularmente em termos de energia. Este padrão indica uma área promissora para futuras pesquisas, destacando a necessidade de se aprofundar na eficiência energética das infraestruturas de rede.

Para futuras pesquisas, é importante reiterar a importância da eficiência energética, dado seu impacto significativo na sustentabilidade e custos operacionais das redes. Incentivar estudos focados neste aspecto pode promover o desenvolvimento de soluções mais equilibradas e sustentáveis. Mais investigação é necessária para explorar como as novas tecnologias, como 5G, 6G e redes satelitais, podem ser otimizadas para eficiência energética, garantindo que o avanço da conectividade global também seja ecologicamente responsável. A Tabela 10 resume os objetivos dos estudos selecionados.

Tabela 10 – Resumo dos objetivos dos estudos selecionados

Objetivo dos Estudos	Referência
Controle de Congestionamento	[59] [38] [46] [39] [31] [40] [54] [25] [43] [26] [27] [65] [49] [45] [53] [44] [57] [61] [34] [51] [41] [70]
Conformidade com SLAs	[30] [29] [28] [31] [47] [25] [48] [64] [52] [69] [67] [60] [68] [58] [62] [63] [33]
Economizar Energia	[42] [55] [56] [66] [50]

5.4 MÉTRICA

As métricas utilizadas na pesquisa científica são essenciais para a avaliação e validação das soluções propostas. No contexto de *Flow Scheduling* em *Edge Computing*, a escolha das métricas apropriadas assegura que o desempenho dos modelos e algoritmos desenvolvidos seja medido de forma precisa e relevante. Essas métricas, que abrangem desde a eficiência de tempo até a utilização de recursos, fornecem os indicadores necessários para avaliar a eficácia e a eficiência das soluções em enfrentar os desafios dinâmicos encontrados nesse campo.

5.4.1 EFICIÊNCIA DE TEMPO

A latência, *delay*, é uma métrica amplamente utilizada para avaliar a eficiência de tempo em muitos estudos. Jiang et al. [59], Diro et al. [29], Han et al. [46], Hong et al. [31], Nasimi et al. [32], Sun et al. [34], Selvi et al. [40], Sahni et al. [26], Sahni et al. [27], Xiang et al. [53], Ahammad et al. [69], Lei e Minzhi [67], Fahad et al. [68], Huang et al. [44], Latif et al. [57], Saranya e Sasikala [58], Serdaroglu e Baydere [61], Yuan et al. [41], Gao et al. [49] e Wang et al. [45] utilizam a latência para medir a rapidez com que os dados são transmitidos e processados, destacando a importância de minimizar atrasos para melhorar o desempenho dos sistemas de *Edge Computing*. Wen et al. [70] também aprofundaram especialmente a métrica de latência, analisando seu impacto em diferentes tipos de cargas de trabalho e condições de rede, além de abordarem *throughput*, capacidade de largura de banda e atraso de enfileiramento.

Além da latência, outros aspectos da eficiência de tempo também foram considerados. Alharbi e Fei [38] utilizaram o *throughput* como métrica principal para avaliar a eficácia da proposta, enquanto Aujla et al. [30] incluíram tanto a latência quanto o *jitter* para medir a qualidade da transmissão de dados. Ge et al. [28] avaliaram a latência e o *buffering* para assegurar uma entrega eficiente de *streaming* ao vivo, destacando como diferentes métricas de tempo podem ser combinadas para obter uma visão mais completa do desempenho do sistema. Wang et al. [39], por sua vez, utilizaram o tempo de processamento de fluxo como métrica para avaliar a eficiência do esquema proposto.

Alguns estudos combinaram a métrica de eficiência de tempo com a de utilização de recursos para uma avaliação de desempenho. Ismail et al. [42], Li et al. [54], Ning et al. [25], Salama et al. [43], Sellami et al. [56] e Zang et al. [33] avaliaram tanto *delay* quanto o consumo energético, destacando a importância de otimizar a velocidade e a eficiência energética simultaneamente. Essas métricas permitem uma visão de como os sistemas gerenciam a troca entre desempenho e baixo consumo de energia.

Outros estudos, como Moraes e Martins [47], Yin et al. [48], Kwak et al. [64], Pencheva e Atanasov [52], Sethon [60], Habeeb et al. [62], Yan et al. [51] e Zhang et al. [63], combinaram latência com largura de banda para avaliar a eficiência de tempo e a utilização

de recursos. Essa combinação é importante para entender como os sistemas mantêm a *QoS* e utilizam a capacidade de rede de maneira eficiente. A análise da latência e da largura de banda alocada ajuda a identificar maneiras de otimizar o fluxo de dados e garantir que os recursos de rede sejam usados de forma eficaz.

Huang et al. [44] obtiveram destaque por combinar o tempo de execução como métrica de eficiência de tempo com o consumo energético. Esta abordagem fornece uma perspectiva sobre como os sistemas podem ser otimizados para executar tarefas rapidamente enquanto minimizam o consumo de energia, mostrando a importância de equilibrar a eficiência temporal e energética para alcançar um desempenho ideal.

5.4.2 UTILIZAÇÃO DE RECURSOS

Alguns estudos priorizaram principalmente a utilização de recursos como métrica para avaliar o desempenho dos sistemas. Dlamini e Vilakati [55] utilizaram o consumo energético como a principal métrica de avaliação, enfatizando a utilização de recursos. Este estudo destacou a importância crítica de minimizar o consumo de energia para aumentar a eficiência geral do sistema, demonstrando como a gestão eficaz dos recursos energéticos pode levar a melhorias significativas no desempenho da rede. Por outro lado, Zhao et al. [65] focaram na otimização da largura de banda, avaliando como a utilização eficaz deste recurso pode melhorar a eficiência geral do sistema. Ambos os estudos ilustram como diferentes aspectos da utilização de recursos podem ser fundamentais para o desempenho eficiente de redes de *Edge Computing*. Yun e Chen [50] abordaram principalmente duas métricas de desempenho: custo total e custo de entrega, para avaliar os algoritmos de agendamento de tráfego em *Edge Computing*. Essas métricas foram usadas para comparar o desempenho do algoritmo proposto com métodos tradicionais, demonstrando a eficácia do algoritmo baseado em Aprendizado Profundo em reduzir custos em diferentes cenários. Ambos os estudos ilustram como diferentes aspectos da utilização de recursos podem ser fundamentais para o desempenho eficiente de redes de *Edge Computing*.

Outros, combinaram a métrica de utilização de recursos como a de eficiência de tempo. Ismail et al. [42] e Li et al. [54] destacam a importância da redução da latência e do consumo de energia, mostrando que um sistema mais eficiente requer menos recursos para processar e transmitir dados. O modelo baseado em gerenciamento ativo de fila busca minimizar a quantidade de retransmissões, reduzindo o desperdício de energia, enquanto o esquema de alocação de recursos proposto otimiza a utilização dos recursos disponíveis, equilibrando a carga de trabalho nos servidores *Edge Computing* e dispositivos móveis. Nessa mesma linha, alguns trabalhos que focaram no gerenciamento de tráfego habilitado para 5G, como Ning et al. [25] e Salama et al. [43] que usaram métricas de utilização de recursos, como a taxa de *offloading* e a taxa de acesso, para avaliar a eficiência dos sistemas propostos. Essas métricas refletem a capacidade do sistema em maximizar a utilização dos recursos de computação e comunicação disponíveis, garantindo um desempenho otimizado

em ambientes de tráfego dinâmico. Estudos focados em sistemas de *IoT* e análise de vídeo baseados em borda, como Sellami et al. [56] e Zang et al. [33], enfatizam a importância da eficiência energética e da largura de banda. A eficiência energética avalia como os nós de *Fog Computing* gerenciam e consomem energia, essencial para ambientes alimentados por baterias. A latência de rede, por outro lado, indica a rapidez com que as tarefas são processadas e os resultados comunicados, evidenciando a eficácia na utilização dos recursos de rede e computação. A gestão dessas métricas permite uma alocação otimizada de recursos, resultando em sistemas mais eficientes e responsivos às demandas dinâmicas.

5.4.3 RESUMO E CONCLUSÕES

A análise das métricas de desempenho utilizadas nos estudos revela importantes considerações para a avaliação de soluções em *Edge Computing*. A grande maioria dos artigos selecionados foca em otimizar aspectos de tempo, o que é bastante coerente com o contexto de infraestruturas de rede, dada a direta relação entre eficiência de tempo e a *QoS* entregue aos usuários. Métricas como latência e *throughput* são amplamente utilizadas, destacando a necessidade de minimizar atrasos e otimizar a transmissão de dados, essenciais para garantir a *QoS* em redes de alta demanda.

Não obstante, há também outros fatores importantes, como a eficiência no uso de recursos, que visam objetivos econômicos e de sustentabilidade. A utilização de recursos, incluindo o consumo energético e a largura de banda, também é fundamental para a avaliação de desempenho. Estudos que combinaram métricas de tempo com métricas de utilização de recursos forneceram uma visão pertinente do desempenho do sistema, permitindo uma análise equilibrada entre a rapidez do processamento e a eficiência no uso dos recursos.

Ainda assim, a quantidade de trabalhos que focam em otimizar o uso de recursos é bastante baixa, o que pode ser um indicativo de que mais pesquisas são necessárias com foco nesse aspecto. A gestão eficaz dos recursos é importante para a sustentabilidade e eficiência das operações de rede, e futuras pesquisas poderiam beneficiar-se ao explorar mais profundamente este aspecto.

Essas observações indicam que, enquanto a otimização do desempenho temporal é bem coberta na literatura, há uma oportunidade significativa para expandir o foco nas métricas de utilização de recursos. A integração de múltiplas métricas permite uma avaliação mais completa e pode conduzir a melhorias significativas no desempenho geral das redes de *Edge Computing*, assegurando que as soluções sejam tanto rápidas quanto eficientes em termos de recursos. A Tabela 11 resume as medidas de desempenho utilizadas nos estudos selecionados.

Tabela 11 – Resumo das medidas de desempenho utilizadas pelos artigos selecionados

Métrica Utilizada	Referência
Eficiência de Tempo	[59] [29] [46] [31] [32] [34] [40] [26] [27] [53] [67] [68] [44] [57] [58] [61] [41] [49] [45] [38] [30] [28] [39] [42] [54] [25] [43] [56] [33] [47] [48] [64] [52] [60] [62] [51] [63] [44] [69]
Utilização de Recursos	[55] [65] [50] [42] [54]

5.5 TÉCNICA

As técnicas utilizadas na pesquisa científica são fundamentais para o desenvolvimento e implementação de soluções competentes. No contexto de *Flow Scheduling* em *Edge Computing*, a escolha das técnicas apropriadas assegura que os modelos e algoritmos desenvolvidos sejam eficientes e eficazes. Essas técnicas, que vão desde modelos matemáticos até abordagens adaptativas de aprendizado de máquina, fornecem as ferramentas necessárias para enfrentar os desafios encontrados nesse campo.

5.5.1 MODELO MATEMÁTICO

Modelos matemáticos foram utilizados para abordar problemas em redes de *Edge Computing*. Alharbi e Fei [38], Diro et al. [29] e Han et al. [46] aplicaram modelos matemáticos para controle de tráfego e congestionamento. Enquanto Jiang et al. [59] utilizaram técnicas como o modelo *Leaky Bucket* para modelar e analisar o controle de tráfego, Nasimi et al. [32] optaram por equações diferenciais para controle de congestionamento e Selvi et al. [40] envolveram modelo matemático para analisar e implementar o controle de congestionamento orientado pelo receptor.

Alguns estudos também utilizaram modelos matemáticos que focaram na otimização de recursos e alocação de largura de banda: Ismail et al. [42], Ning et al. [25], Kwak et al. [64], Pencheva e Atanasov [52], Xiang et al. [53] e Zhao et al. [65]. Outros, de mesmo foco, utilizaram modelos matemáticos para resolver problemas de otimização, combinando técnicas como a teoria de otimização de *Lyapunov* para decompor o problema de otimização estocástica, como abordado por Li et al. [54], modelos de *Markov* para predição de popularidade vistos em Wang et al. [39], teoria de filas descritos por Salama et al. [43] e algoritmos de alocação de largura de banda propostos por Moraes e Martins [47].

Ahammad et al. [69] e Wang et al. [45] utilizaram modelos matemáticos para abordar questões específicas de redes de *Edge Computing*. Além destes, recuperamos o uso de um modelo matemático baseado no conceito de curva de serviço para adaptar a taxa de serviço em tempo real proposto por Huang et al. [44], o gerenciamento de recursos via *SDN* e *Blockchain* descritos por Latif et al. [57], e a alocação de tarefas e *Flow Scheduling* para maximizar o *throughput* de Zhang et al. [63]. Técnicas como algoritmos de colônia de formigas em Yuan et al. [41], otimização de *Lyapunov* em Li et al. [54], Xiang et al. [53] e Zang et al. [33] e teoria de filas de prioridade múltipla de Serdaroglu e Baydere [61] também foram aplicadas para resolver problemas, demonstrando a versatilidade e eficácia

dos modelos matemáticos em cenários de rede.

5.5.2 HEURÍSTICA

Heurísticas foram utilizadas para resolver problemas em redes de *Edge Computing*. Aujla et al. [30] envolveram o uso de heurísticas para ordenação linear, classificação de tráfego e gerenciamento de filas, Ge et al. [28] utilizaram heurísticas para a retenção temporária de segmentos de vídeo na borda móvel e Hong et al. [31] para gerenciamento de recursos de rede, aplicando políticas como agendamento de compartilhamento proporcional, reserva mínima de largura de banda e limitação máxima de largura de banda.

Outros estudos focaram na otimização do roteamento e alocação de recursos através de heurísticas. Sun et al. [34] utilizaram uma abordagem de busca em vizinhança variável para otimizar o peso dos *links* e o roteamento de tráfego, enquanto Sahni et al. [26] aplicaram o algoritmo *Multistage Greedy Adjustment* para ajustar o posicionamento das tarefas e a largura de banda dos fluxos. E Yin et al. [48] focaram na seleção de nós de retransmissão e alocação de largura de banda com base na prioridade dos fluxos utilizando o mecanismo *Collaborative Transmission Optimization Mechanism*.

Heurísticas também foram empregadas para melhorar a eficiência do *offloading* e agendamento de tarefas. Sahni et al. [27] envolveram o uso de uma heurística denominada *Joint Partial Offloading and Flow Scheduling Heuristic* para *offloading* parcial e escalonamento de fluxos. Fahad et al. [68] fizeram uso de um algoritmo heurístico de agendamento de tarefas baseado em prioridade multi-fila que categoriza tarefas como curtas e longas com base no tempo de *burst* e ajusta dinamicamente o valor do intervalo de tempo para preempção. Habeeb et al. [62] em um sistema de filas múltiplas baseado em *QoS*, utilizaram heurísticas para otimizar o desempenho em ambientes de borda colaborativa. Adicionalmente, Saranya e Sasikala [58] utilizaram o algoritmo *Improved Offloading and Heuristic Flow Optimization* para otimizar o sequenciamento de tarefas e o *offloading* parcial, utilizando técnicas de relaxação semidefinida.

5.5.3 META-HEURÍSTICA

Meta-heurísticas são técnicas utilizadas para lidar com problemas de otimização em redes de *Edge Computing*. Lei e Minzhi [67] utilizaram um algoritmo baseado em colônia de abelhas artificiais, melhorado com aprendizagem reversa, para otimizar o escalonamento das tarefas e a alocação de dados de computação. Esta abordagem permitiu um ajuste dinâmico e eficiente dos recursos, garantindo um melhor desempenho do sistema.

A otimização de caminhos de rede foi abordada de forma eficiente por Sethon [60], que empregou um algoritmo de otimização de colônia de formigas. As formigas virtuais exploraram a rede e marcaram os melhores caminhos com feromônios, otimizando o tráfego de rede com base em diferentes métricas de *QoS*. Em Yuan et al. [41] também foram

utilizados algoritmos de colônia de formigas, demonstrando que esta técnica é eficaz na descoberta de rotas otimizadas, melhorando a *QoS*.

Arshed et al. [66] utilizaram um algoritmo genético para otimizar a alocação de tarefas e o agendamento de módulos de aplicação em um ambiente de *Edge Computing* e *Fog Computing*. O algoritmo *GA-IRACE* utilizou operadores de cruzamento e mutação para evoluir as soluções, mostrando-se eficiente na busca por soluções ótimas em um vasto espaço de solução.

5.5.4 APRENDIZADO DE MÁQUINA

Aprendizado de Máquina tem sido aplicado com sucesso em redes de *Edge Computing* para prever cargas de tráfego e otimizar o gerenciamento de recursos. Dlamini e Vilakati [55] utilizaram uma rede neural recorrente para prever a carga de tráfego, permitindo um gerenciamento dinâmico e eficiente dos recursos. Esta abordagem permitiu antecipar variações na demanda e ajustar os recursos de rede de acordo, melhorando a eficiência e a *QoS*.

A otimização da alocação de tarefas e gerenciamento de recursos também foi abordada por Sellami et al. [56], que empregaram um algoritmo de aprendizado por reforço profundo. Esta técnica de aprendizado profundo permitiu ao sistema aprender e adaptar-se a diferentes condições de rede, otimizando a alocação de tarefas e melhorando a utilização dos recursos. O uso do aprendizado por reforço profundo mostrou-se eficaz na tomada de decisões em tempo real, ajustando-se dinamicamente às mudanças na rede.

Yun e Chen [50] utilizaram de redes neurais para analisar e classificar as solicitações de tarefas com base em informações sobre o estado do servidor e características das solicitações. O modelo é projetado para aprender a partir de dados históricos, permitindo que ele se adapte e melhore sua precisão ao longo do tempo.

5.5.5 RESUMO E CONCLUSÕES

A análise dos estudos que utilizaram diferentes técnicas para *Flow Scheduling* em *Edge Computing* revela diversas abordagens com suas próprias vantagens e limitações.

Modelos matemáticos fornecem uma base rigorosa para a formulação e resolução de problemas, permitindo uma análise precisa do comportamento do sistema e facilitando a implementação de soluções otimizadas. Eles são amplamente aplicados para controle de tráfego, otimização de recursos e alocação de largura de banda.

Heurísticas oferecem uma abordagem prática e flexível, capaz de resolver problemas de forma eficiente, especialmente em cenários onde a velocidade de processamento é fundamental. Elas são usadas para gerenciamento de recursos de rede, otimização de roteamento e alocação de recursos.

Meta-heurísticas, como algoritmos de colônia de formigas, algoritmos genéticos e otimização de colônia de abelhas, são eficazes para explorar grandes espaços de solução

e encontrar soluções próximas do ótimo em problemas de otimização. Essas técnicas demonstram versatilidade e robustez em diversas aplicações de rede.

Técnicas de aprendizado de máquina trazem uma camada adaptativa e inteligente, permitindo previsões precisas e ajustes dinâmicos em tempo real, essenciais para a gestão eficiente de recursos em ambientes dinâmicos. No entanto, é notável que apenas três dos estudos selecionados utilizam técnicas de aprendizado de máquina. Enquanto dois trabalhos focam na capacidade de aprendizado ao longo do tempo para decisões de alocação de recursos, um terceiro artigo utiliza técnicas de aprendizado de máquina para prever a carga de trabalho. Dada a ampla gama de técnicas de aprendizado de máquina disponíveis e o potencial para otimizar decisões em infraestruturas de rede, é surpreendente que tão poucos artigos tenham explorado o uso dessa abordagem.

Essas abordagens complementares ressaltam a importância de uma estratégia multifacetada na aplicação de técnicas. A combinação de precisão teórica, eficiência prática e adaptabilidade inteligente é importante para desenvolver soluções robustas e eficientes em redes de *Edge Computing*. Futuras pesquisas podem beneficiar-se ao expandir o uso de técnicas de aprendizado de máquina, explorando seu potencial para melhorar ainda mais a gestão de recursos e o desempenho de redes de *Edge Computing*. A Tabela 12 resume as técnicas empregados para implementar o *Flow Scheduling*.

Tabela 12 – Resumo das técnicas empregadas pelos artigos selecionados

Técnica Utilizada	Referência
Modelo Matemático	[38] [29] [46] [59] [32] [40] [42] [25] [64] [52] [53] [65] [54] [39] [43] [47] [69] [45] [44] [57] [63] [41] [54] [53] [33] [61] [70]
Heurística	[30] [28] [31] [34] [26] [48] [27] [68] [62] [58]
Meta-heurística	[41] [67] [60] [66]
Aprendizado de Máquina	[55] [56] [50]

5.6 VALIDAÇÃO

A validação é uma etapa na pesquisa científica que permite verificar e confirmar a eficácia das soluções propostas. No contexto de *Flow Scheduling* em *Edge Computing*, as técnicas de validação asseguram que os modelos e algoritmos são robustos e funcionais.

5.6.1 SIMULAÇÃO/EMULAÇÃO

A validação é uma etapa importante na pesquisa científica, permitindo verificar e confirmar a eficácia das soluções propostas. No contexto de *Flow Scheduling* em *Edge Computing*, as técnicas de validação asseguram que os modelos e algoritmos desenvolvidos são robustos e funcionais. A validação por meio de Simulação/Emulação é especialmente importante, pois permite criar cenários controlados para testar o desempenho dos sistemas. A simulação utiliza modelos abstratos para prever o comportamento sob diferentes condições, enquanto a emulação replica fielmente ambientes reais usando ferramentas específicas.

Essas abordagens complementares fornecem uma análise do funcionamento dos algoritmos e modelos em ambientes que imitam tanto condições teóricas quanto práticas.

Estudos focados na validação de modelos em redes *Fog Computing* e *IoT* utilizaram simulações para testar suas propostas. Moraes e Martins [47] utilizaram simulações para avaliar o desempenho do *framework PSIoT-SDN*, enquanto Ning et al. [25] utilizaram dados reais de trajetórias de táxis em *Shanghai* para simulações que avaliaram o desempenho do sistema proposto. Eles compararam o desempenho do sistema com outras soluções existentes, como o esquema baseado em *Non-Orthogonal Multiple Access* e *offloading* aleatório. Em Xiang et al. [53], o modelo proposto foi simulado em diversos cenários de carga de trabalho e diferentes topologias de rede, demonstrando que a abordagem pode melhorar significativamente a utilização dos recursos e reduzir a latência do sistema. Ahammad et al. [69], a política proposta foi simulada em diversos cenários de rede, demonstrando que a integração de *Fog Computing* com *SDN* pode otimizar significativamente a utilização dos recursos e melhorar a latência e o *throughput* do sistema. Arshed et al. [66], o agendador *GA-IRACE* foi simulado em vários cenários de carga de trabalho, evidenciando que a abordagem proposta melhora a eficiência do uso dos recursos e reduz os custos operacionais. Nos estudos de Wen et al. [70] e Yun e Chen [50], os modelos propostos foram simulados em diferentes cenários de tráfego e topologias de rede. No primeiro, o algoritmo *JANUS* demonstrou ser eficaz na redução da latência e na melhoria do *throughput* em ambientes de borda. Similarmente, no segundo, o algoritmo baseado em Aprendizado Profundo mostrou-se eficiente na otimização da alocação de recursos e na redução do consumo de energia em redes *IoT*, mantendo níveis aceitáveis de latência.

Algoritmos de controle de congestionamento e alocação de recursos também foram validados por meio de simulação. Selvi et al. [40] implementaram o modelo matemático *Receiver-oriented Congestion Control* em um simulador de rede para avaliar o desempenho da abordagem em comparação com outras. Li et al. [54] utilizaram simulações para comparar o desempenho de seu algoritmo de alocação de recursos com outros algoritmos, enquanto Sahni et al. [26] utilizaram simulações para comparar o algoritmo *Multistage Greedy Adjustment* com soluções de referência. Dlamini e Vilakati [55] e Sahni et al. [27] também utilizaram simulações para validar suas estratégias de balanceamento de carga e heurísticas.

Estudos focados em eficiência energética e desempenho de redes utilizaram simulações para validar suas propostas. Nasimi et al. [32] apresentaram uma avaliação analítica e simulação do desempenho do mecanismo proposto, enquanto Salama et al. [43] e Sellami et al. [56] demonstraram melhorias em atraso, consumo de energia e *throughput* através de simulações.

A implementação de algoritmos em simuladores de rede foi uma abordagem comum para validação. Ismail et al. [42] utilizaram o *Green Cloud Simulator* para simular o desempenho do modelo de nuvem verde, e Yin et al. [48] utilizaram o simulador *NS-3*

para avaliar o desempenho da abordagem proposta. De forma semelhante, Fahad et al. [68] e Zang et al. [33] utilizaram simuladores como *iFogSim* e outros para validar suas propostas.

Estudos que validaram algoritmos em ambientes específicos, como simulações utilizando *MATLAB* e emuladores, foram conduzidos para testar a eficácia das propostas. Lei e Minzhi [67] utilizaram o *MATLAB* para avaliar um algoritmo de colônia de abelhas artificiais, enquanto o Latif et al. [57] utilizaram emuladores como *Mininet Wi-Fi* para simulações. Saranya e Sasikala [58] e Wang et al. [45] também utilizaram simulações em diferentes ambientes para validar a eficácia de suas soluções.

Algumas pesquisas em *Edge Computing* utilizaram tanto simulação quanto experimentação para validar suas propostas. Habeeb et al. [62] utilizaram a ferramenta *IoTSim-Osmosis* para modelar diversos cenários com variações no número de dispositivos *IoT* e condições de rede, analisando a eficácia da abordagem proposta. Huang et al. [44] utilizaram simulações para avaliar o desempenho do *Prioritized Traffic Shaping for MEC* em cenários controlados, demonstrando sua capacidade de atender aos objetivos de latência para fluxos *Edge Computing* mesmo em situações de congestionamento. Serdaroglu e Baydere [61] complementaram os testes práticos com simulações para observar mudanças nos valores de tempo de espera em diferentes sistemas de filas, avaliando a eficácia do modelo proposto. Yan et al. [51] conduziram simulações baseadas em uma série de cenários para avaliar o desempenho do sistema em termos de alocação de largura de banda, uso da rede e consumo energético. Yuan et al. [41] utilizaram a plataforma *OPNET* para simular o ambiente de rede da nuvem empresarial, avaliando o desempenho do algoritmo de evitação de congestionamento em diferentes cenários. Zhang et al. [63] desenvolveram um *testbed* híbrido com nós de borda físicos e virtuais para avaliar o sistema em grande escala. E Gao et al. [49] criaram instâncias de teste em uma rede complexa com parâmetros variados, gerando 24.000 resultados experimentais para avaliar o desempenho dos algoritmos comparativos.

5.6.2 EXPERIMENTAÇÃO

A validação experimental busca assegurar a eficácia das propostas teóricas em cenários reais, utilizando uma variedade de abordagens de experimentação para validar suas contribuições.

Alguns estudos realizaram experimentos em ambientes controlados e redes de teste. Alharbi e Fei [38] utilizaram o ambiente de teste *GENI* para avaliar a arquitetura *SDN* proposta, implantando uma rede *cFat Tree* e executando testes com diferentes configurações de tráfego, como *TCP* regular, *MPTCP* com roteamento *ECMP* e *MPTCP* com roteamento auxiliar *SDN*. Aujla et. al. [30] empregaram um ambiente experimental com um *switch OpenFlow* e um controlador *SDN*, utilizando quatro nós sob diferentes cargas de tráfego, variando de 10 a 100 *Mbps*, com diversos tamanhos de pacotes e tipos de fluxo (voz, vídeo e dados). A eficácia do esquema foi comparada com um esquema

First-In-First-Out e um esquema proposto sem ordenação linear. Diro et al. [29] realizaram análises e avaliações do desempenho do sistema proposto em comparação com sistemas *SDN* sem *QoS* em termos de probabilidade de bloqueio de pacotes, *throughput* e utilização da rede, além de discutirem a eficiência energética do sistema. Ge et al. [28] implementaram o sistema em um cenário de rede realista, injetando diferentes tipos de tráfego para testar a robustez e eficiência do sistema em um ambiente de teste *Long-Term Evolution Advanced*, utilizando fontes de vídeo reais distribuídas globalmente.

Estudos focados na eficiência de protocolos e plataformas de transmissão de dados também utilizaram experimentação para validar suas propostas. Han et al. [46] realizaram testes empíricos para comparar a eficiência do protocolo *UDP-based Data Transfer Protocol* modificado com o protocolo *Real-Time Media Flow Protocol* em um ambiente de rede local e em um ambiente de *IoT* e *Fog Computing*. De forma semelhante, Sun et al. [34] implementaram uma plataforma de *IoT* e borda em um ambiente hospitalar e universitário, comparando o desempenho da plataforma com a abordagem proposta no artigo.

Outros utilizaram da implementação de protótipos em ambientes de rede física. Hong et al. [31] validaram suas funcionalidades implementando o controlador *SDN qCon* em um cenário real. O ambiente experimental consistia em múltiplos *switches* e controladores conectados, simulando uma rede de *data centers*. Foram realizados testes para verificar a capacidade do controlador em gerenciar o tráfego de forma eficiente, adaptando-se dinamicamente às mudanças na demanda de rede. Kwak et al. [64] descreveram a implementação do controlador *IntelliTC* em um ambiente de teste real, integrado com *OpenStack* e *ONOS*, demonstrando a eficácia do controle inteligente do tráfego entre *data centers* distribuídos.

Embora Zhao et al. [65] tivessem utilizado simulações dentro do protótipo, o desenvolvimento e a avaliação em um ambiente protótipo funcional indicam que a validação foi realizada predominantemente por meio de experimentação. O protótipo funcional implementou as principais características do sistema proposto e foi submetido a testes para medir seu desempenho em condições de uso real. As simulações complementaram os testes experimentais, fornecendo uma visão do comportamento do sistema em vários cenários. E Huang et al. [44] implementaram um protótipo da solução em uma plataforma *Edge Computing* habilitada para *Long Term Evolution* e realizaram testes práticos para avaliar seu desempenho em cenários de congestionamento.

Algumas pesquisas em *Edge Computing* utilizaram tanto experimentação quanto simulação para validar suas propostas. Sethon [60] conduziu experimentos práticos em ambientes reais de *Fog Computing*, coletando dados como *throughput* e taxa de entrega de pacotes para avaliar o desempenho do algoritmo. Habeeb et al. [62] realizaram testes com dispositivos *IoT* reais em ambientes controlados, analisando dados de sensores processados nos nós de borda e nuvem. Serdaroglu e Baydere [61] conduziram testes práticos em um ambiente de emulação para verificar a eficiência de um modelo de filas proposto. Yan et al. [51] realizaram análises experimentais comparando o algoritmo proposto com

abordagens de referência, avaliando alocação de largura de banda e balanceamento de carga. Yuan et. al. [41] implementaram e testaram um algoritmo em um ambiente real de nuvem empresarial, validando seu desempenho em condições práticas. Zhang et. al. [63] avaliaram o desempenho do sistema por meio de experimentos comparativos com o estado da arte, utilizando métricas como *throughput* e tempo médio de espera. Gao et. al. [49] apresentaram resultados de experimentos em tabelas, comparando a eficácia da abordagem proposta com algoritmos comparativos para o problema de *offloading* de tarefas parciais em sistemas de *Edge Computing*.

5.6.3 MODELAGEM FORMAL

A modelagem formal visa garantir a precisão das propostas teóricas por meio de técnicas matemáticas, utilizando uma variedade de métodos formais para verificar e validar suas contribuições.

Desta forma, Pencheva e Atanasov [52] utilizaram modelagem formal para validar sua proposta de extensão do serviço de gerenciamento de banda em um ambiente de *Edge Computing*. Os autores descreveram formalmente e verificaram modelos que representam o estado da alocação de banda larga. Essa verificação formal permitiu demonstrar matematicamente que a extensão proposta é implementável de forma consistente, assegurando a eficácia do sistema na gestão de banda em redes de acesso rádio. Jiang et al. [59] avaliaram o desempenho do modelo na modelagem e controle do tráfego de dados em redes de névoa. Eles basearam-se em modelos teóricos, como o modelo de *Leaky Bucket*, para derivar métricas de desempenho e analisar o comportamento do tráfego. Nasimi et al. [32], utilizaram equações diferenciais e aproximações determinísticas indicando que houve um desenvolvimento de modelos formais para analisar o comportamento do sistema proposto.

Li et al. [54] validaram sua proposta através de uma combinação de modelagem formal e simulação. Inicialmente, desenvolveram modelos matemáticos para representar a transmissão de dados, o processamento computacional e a dinâmica das filas de tarefas no ambiente de *Edge Computing*. Foi desenvolvido o algoritmo *Delay-Aware Task Congestion Control and Resource Allocation*, cujos resultados foram analisados por meio de simulações. Porém os autores não mencionam o uso de um simulador específico pelo nome.

5.6.4 RESUMO E CONCLUSÕES

A análise dos métodos de validação utilizados nos estudos revela a abordagem dominante e as práticas de verificação da eficácia das soluções propostas. A Simulação/Emulação emerge como a técnica de validação predominante, evidenciando sua importância em testar hipóteses e modelos em ambientes controlados antes da implementação real. Simulações utilizam modelos abstratos para prever o comportamento sob diferentes condições, enquanto emulações replicam fielmente ambientes reais usando ferramentas específicas. Ambas as

abordagens permitem ajustar e otimizar as soluções, garantindo que os modelos funcionem conforme o esperado sob diversas condições de rede.

No entanto, a validação experimental também desempenha um papel importante ao testar as soluções em cenários reais. Experimentos práticos confirmam a eficácia das propostas em condições de operação realistas, destacando a necessidade de validar o desempenho e a confiabilidade em ambientes reais de rede. A combinação de simulação e experimentação oferece uma abordagem robusta, permitindo uma validação inicial em ambientes simulados, seguida por testes práticos. Esse equilíbrio entre precisão teórica e aplicabilidade prática é essencial para desenvolver soluções confiáveis e eficientes.

A modelagem formal, embora menos comum, proporciona uma base matemática precisa para verificar e validar as propostas teóricas. Este método assegura que os fundamentos teóricos estejam corretos antes de avançar para a simulação ou experimentação, adicionando uma camada de rigorosidade à validação.

A dominância da Simulação/Emulação como técnica de validação é evidente. Para fornecer uma visão das ferramentas preferidas pelos pesquisadores e como elas são utilizadas, destacamos na Tabela 13 as ferramentas de simulação e emulação mais empregadas nos estudos selecionados. Adicionalmente, a Tabela 14 resume como os estudos foram validados ou testados de maneira clara e detalhada.

Como visto anteriormente, nem todos os estudos explicitam o uso de um simulador ou emulador pelo nome. Alguns trabalhos mencionam o uso de simulações para validar suas propostas, mas não fornecem detalhes sobre a ferramenta específica utilizada.

Nasimi et al. [32] mencionam resultados de avaliação de desempenho, que geralmente são obtidos por meio de simulações para analisar o comportamento do mecanismo proposto em diferentes condições de rede. De maneira semelhante, Li et al. [54] descrevem a utilização de simulações para avaliar o desempenho do algoritmo proposto, chamado *Delay-Aware Task Congestion Control and Resource Allocation*.

Gao et al. [49] não mencionam explicitamente o uso de um simulador específico para validar a proposta. No entanto, eles indicam que os algoritmos foram implementados em *Java* e executados em um *notebook* com especificações de *hardware*, sugerindo que a validação foi realizada por meio de simulações computacionais personalizadas. As simulações demonstram melhorias significativas em termos de eficiência e desempenho, confirmando a viabilidade e a eficiência da abordagem proposta. De maneira semelhante, Saranya e Sasikala [58] utilizaram simulações em um ambiente *offline* utilizando um *MacBook Pro* com um processador *Intel Core i5* para validar a proposta. Sahni et al. [27] também seguem essa linha, mencionando simulações realizadas em um outro *MacBook Pro* também com um processador *Intel Core i5*.

Wang et al. [45] discorrem sobre a realização de simulações para validar a eficácia do algoritmo *QLBF-UCB* proposto. No entanto, o artigo não especifica um simulador particular utilizado para essas simulações. De modo similar, Yun e Chen [50] utilizaram

Tabela 13 – Ferramentas de simulação e emulação utilizadas nos trabalhos avaliados.

Ferramenta	Principais Funcionalidades	Tipos de Redes / Cenários Suportados	Referências
<i>NS-3</i>	Modelagem de redes, suporte a múltiplos protocolos e extensibilidade através de módulos.	Redes com fio e sem fio, <i>IoT</i> , <i>SDN</i> , 5G	[40] [48]
<i>Green Cloud Simulator</i>	Avaliação de consumo energético, simulação de <i>data centers</i> e algoritmos de gerenciamento de energia.	<i>Data centers</i> , redes de nuvem	[42] [43]
<i>iFogSim</i>	Suporte a ambientes de <i>Fog Computing</i> , modelagem de <i>IoT</i> , e simulação de aplicações sensíveis à latência.	<i>Fog Computing</i> , <i>IoT</i>	[69] [66] [68]
<i>IoTsim-Osmosis</i>	Avaliação de políticas de alocação de recursos, estratégias de <i>offloading</i> de tarefas, e eficiência energética.	<i>IoT</i> , <i>Fog Computing</i>	[62]
<i>EdgeSimTools</i>	Avaliação de estratégias de alocação de recursos, balanceamento de carga e eficiência de rede.	<i>Edge Computing</i>	[53]
<i>Mininet-WiFi</i>	Emulação de redes <i>Wi-Fi</i> , integração com <i>OpenFlow</i> , suporte a vários protocolos de rede sem fio.	Redes sem fio, <i>SDN</i> , <i>IoT</i>	[57]
<i>OPNET</i>	Análise de desempenho de rede, simulação de protocolos, e modelagem de tráfego de rede.	Redes corporativas, nuvem empresarial	[41]
MATLAB	Análise matemática, modelagem de sistemas, e simulações complexas.	Vários, incluindo redes e sistemas de comunicação	[26] [55] [39] [67]
<i>Esper CEP Engine</i>	Detecção de padrões, correlações e tendências em tempo real, processamento de eventos.	Redes em tempo real, <i>IoT</i> , sistemas de eventos	[61]

dados de *Google trace* para simular diferentes cenários e avaliar o desempenho do algoritmo *Deep Learning-Based Traffic Scheduling Approach*. As simulações demonstram melhorias significativas em termos de custo de energia e eficiência de alocação de recursos, mas o nome do simulador específico não é mencionado.

Ning et al. [25] descrevem simulações em termos de configuração de rede, parâmetros e métricas de desempenho, mas não nomeiam o simulador utilizado no texto. Sellami et al. [56] relataram a condução de simulações em um ambiente controlado, utilizando um *testbed* para avaliar o desempenho do algoritmo proposto, mas novamente, o simulador utilizado não é mencionado.

Yan et al. [51] utilizaram simulações para validar a eficácia do algoritmo *LBS*. As simulações demonstram melhorias significativas em termos de balanceamento de carga e

uso de largura de banda, confirmando a viabilidade e a eficiência da abordagem proposta. Entretanto, não há menção de um simulador específico utilizado nesse estudo. Zang et al. [33] também não especificaram o simulador utilizado, mas descrevem a validação da eficácia do algoritmo *JCAB* através de simulações, demonstrando melhorias significativas em termos de precisão, latência e consumo de energia.

Por fim, Wen et al. [70] utilizaram simulações para validar a eficácia do sistema *JANUS*. As validações demonstraram melhorias significativas em termos de latência e *throughput*, confirmando a viabilidade e a eficiência da abordagem proposta, mas o artigo não menciona o uso de um simulador específico.

Tabela 14 – Resumo das abordagens de validação utilizadas pelos artigos selecionados

Abordagem de Validação	Referências
Simulação	[47] [25] [40] [26] [55] [27] [43] [56] [42] [48] [68] [33] [67] [39] [57] [58] [45] [69] [66] [62] [53] [61] [51] [41] [63] [49] [70] [50]
Experimentação	[38] [30] [29] [28] [46] [34] [31] [64] [65] [60] [62] [44] [61] [51] [41] [63] [49]
Modelagem formal	[52] [59] [32] [54]

5.7 RESUMO E CONCLUSÕES GERAIS

A análise das diferentes categorias da taxonomia proposta de estudos em *Flow Scheduling* em *Edge Computing* revela uma abordagem multifacetada necessária para enfrentar os desafios dinâmicos deste campo. Cada categoria trouxe percepções sobre os componentes essenciais para o desenvolvimento de soluções eficientes.

Os diferentes tipos de infraestrutura de rede, incluindo redes com fio, sem fio e mistas, apresentam vantagens e desafios distintos que atendem a diversas necessidades tecnológicas. Redes com fio oferecem conexões estáveis e de alta velocidade, ideais para ambientes que requerem baixa latência e alta largura de banda. Redes sem fio destacam-se pela flexibilidade e mobilidade, sendo cruciais para dispositivos *IoT* e ambientes móveis, embora enfrentem desafios na gestão de tráfego e eficiência energética. Redes mistas combinam as vantagens de ambas, proporcionando uma solução equilibrada. A crescente tendência para redes sem fio e o desenvolvimento de tecnologias como 5G, 6G e redes satelitais sublinham a importância de pesquisas voltadas para essas áreas emergentes.

A análise dos objetivos dos estudos mostra uma ênfase clara em três metas principais: controle de congestionamento, conformidade com *SLAs* e eficiência energética. Controle de congestionamento é vital para manter a eficiência e estabilidade das redes, enquanto a conformidade com *SLAs* assegura que os parâmetros de desempenho e qualidade sejam mantidos. A eficiência energética é importante devido ao aumento de dispositivos conectados e à necessidade de operações sustentáveis. Embora a maioria dos estudos foque na otimização do desempenho, há uma necessidade crescente de explorar a eficiência energética, especialmente com a introdução de novas tecnologias de rede.

A eficiência de tempo, medida por latência e *throughput*, é uma métrica importante para garantir a QoS. A utilização de recursos, incluindo consumo energético e largura de banda, também é fundamental, embora menos explorada. A combinação dessas métricas permite uma avaliação ampla do desempenho, balanceando rapidez de processamento com eficiência no uso dos recursos. Observa-se uma oportunidade significativa para expandir o foco nas métricas de utilização de recursos, o que é essencial para a sustentabilidade das operações de rede.

Diversas técnicas são utilizadas para *Flow Scheduling*, cada uma com suas próprias vantagens. Modelos matemáticos oferecem uma base rigorosa para a análise precisa e implementação de soluções otimizadas. Heurísticas são práticas e flexíveis, eficazes para resolver problemas difíceis rapidamente. Meta-heurísticas exploram grandes espaços de solução, encontrando soluções próximas do ótimo em problemas de otimização. Técnicas de aprendizado de máquina trazem uma camada adaptativa e inteligente, embora ainda subexploradas, mostrando um potencial significativo para melhorar a gestão de recursos e o desempenho de redes. A combinação dessas abordagens permite desenvolver soluções robustas e eficientes.

Simulação/Emulação é a técnica predominante de validação, permitindo testar hipóteses e modelos em ambientes controlados antes da implementação real. A validação experimental em cenários reais confirma a eficácia das soluções sob condições práticas. A modelagem formal, embora menos comum, assegura a precisão matemática das propostas teóricas. A combinação de simulação/emulação, experimentação e modelagem formal proporciona uma abordagem robusta, equilibrando precisão teórica e aplicabilidade prática, essencial para desenvolver soluções confiáveis e eficientes.

A integração de abordagens diversas e complementares em termos de infraestrutura, objetivos, métricas, técnicas e métodos de validação é importante para desenvolver soluções robustas, eficientes e adaptáveis em redes de borda. A pesquisa contínua e a exploração de novas tecnologias, como 5G e redes satelitais, são fundamentais para enfrentar os desafios futuros da conectividade global, garantindo que as redes possam suportar de maneira eficiente e sustentável uma gama diversificada de dispositivos e aplicações.

6 CONSIDERAÇÕES FINAIS

A crescente atenção direcionada à *Edge Computing*, impulsionada pela sua capacidade de oferecer benefícios como redução da latência, a melhoria da eficiência da rede e o suporte a aplicações de tempo real, como veículos autônomos e saúde digital, reflete a urgência de encontrar soluções inovadoras para os desafios emergentes. *Edge Computing* herda diversas tendências de *Cloud Computing*, como a virtualização da infraestrutura de rede visando maior flexibilidade e desempenho, por meio de políticas de gerenciamento de recursos como o *Flow Scheduling*.

Neste trabalho, foi apresentada uma revisão sistemática da literatura sobre *Flow Scheduling* em *Edge Computing*. Um dos principais objetivos desta pesquisa foi realizar uma análise quantitativa da evolução e das características dessa área em constante desenvolvimento. A análise quantitativa trouxe percepções sobre a evolução temporal e geográfica das publicações, além de destacar os principais veículos de publicação, autores e instituições que dominam esse campo. Esses dados são fundamentais para compreender o panorama atual da área e identificar as maiores contribuições científicas.

Além disso, a análise qualitativa proporcionou uma categorização dos estudos, organizada por meio de uma taxonomia que compreende aspectos como infraestrutura de rede, métricas de avaliação e técnicas empregadas. Essa análise aprofundou a compreensão das diferentes estratégias de *Flow Scheduling*, destacando as contribuições mais relevantes e identificando lacunas e desafios ainda não resolvidos.

A integração das análises quantitativa e qualitativa permitiu não apenas mapear o estado atual da pesquisa, mas também identificar oportunidades de avanço, tanto no desenvolvimento de novas abordagens técnicas quanto na adaptação a ambientes mais diversificados de *Edge Computing*.

6.1 DIREÇÕES FUTURAS

Com base na síntese dos estudos revisados, várias direções futuras emergem como áreas promissoras de investigação. Primeiramente, destaca-se a necessidade de desenvolver novos algoritmos de agendamento que conciliem melhor os requisitos de baixa latência e eficiência energética, que são aspectos críticos em ambientes de *Edge Computing*. Além disso, é evidente a falta de experimentações em cenários reais que contemplem a diversidade de aplicações *IoT* e as complexidades inerentes a redes heterogêneas.

Outra oportunidade reside na exploração de técnicas como aprendizado de máquina e meta-heurísticas, visando otimizar o desempenho em tempo real do *Flow Scheduling*. Pesquisas futuras também poderiam se concentrar em cenários de redes 5G/6G, onde as demandas impostas ao *Flow Scheduling* são ainda mais rigorosas. Por fim, com a popularização de *Edge Computing*, será fundamental investigar métodos que garantam maior resiliência e segurança, especialmente em contextos descentralizados e sujeitos a falhas, garantindo que as soluções propostas sejam viáveis e eficazes em cenários práticos.

REFERÊNCIAS

- 1 GUBBI, J. et al. Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, ELSEVIER, v. 29, n. 7, p. 1645–1660, 2013. Citado na página 13.
- 2 LIU, Y. et al. Toward edge intelligence: Multiaccess edge computing for 5g and internet of things. *IEEE Internet of Things Journal*, IEEE, v. 7, n. 8, p. 6722–6747, 2020. Citado na página 13.
- 3 BUYYA, R.; BROBERG, J.; GOSCINSKI, A. *Cloud computing: Principles and paradigms*. [S.l.]: John Wiley & Sons, 2011. v. 87. ISBN 978-0-470-88799-8. Citado na página 13.
- 4 SHI, W.; DUSTDAR, S. The promise of edge computing. *Computer*, IEEE, v. 49, n. 5, p. 78–81, 2016. Citado na página 13.
- 5 SATYANARAYANAN, M. The emergence of edge computing. *Computer*, IEEE, v. 50, n. 1, p. 30–39, 2017. Citado na página 13.
- 6 JIANG, C. et al. Energy aware edge computing: A survey. *Comput. Commun.*, Elsevier Science Publishers B. V., v. 151, n. C, p. 556–580, 2020. Citado na página 13.
- 7 ARAL, A.; BRANDIC, I. Learning spatiotemporal failure dependencies for resilient edge computing services. *IEEE Transactions on Parallel and Distributed Systems*, IEEE, v. 32, n. 7, p. 1578–1590, 2021. Citado na página 13.
- 8 GORANSSON, P.; BLACK, C.; CULVER, T. *Software Defined Networks: A Comprehensive Approach*. [S.l.]: Morgan Kaufmann, 2016. v. 2. ISBN 9780128045558. Citado 2 vezes nas páginas 13 e 20.
- 9 BUYYA, R. et al. Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems*, ELSEVIER, v. 25, n. 6, p. 599–616, 2009. Citado na página 15.
- 10 ARMBRUST, M. et al. A view of cloud computing. *Commun. ACM*, Association for Computing Machinery, v. 53, n. 4, p. 50–58, 2010. Citado na página 15.
- 11 MELL, P.; GRANCE, T. *The nist definition of cloud computing*. [S.l.]: NIST Special Publication, 2011. v. 800. Citado na página 15.
- 12 XING, Y.; ZHAN, Y. *Virtualization and Cloud Computing*. [S.l.]: Springer Berlin Heidelberg, 2012. Citado na página 16.
- 13 DOLUI, K.; DATTA, S. K. Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing. In: IEEE. *Global Internet of Things Summit*. [S.l.], 2017. p. 1–6. Citado na página 19.
- 14 TANG, S. et al. A survey on scheduling techniques in computing and network convergence. *IEEE Communications Surveys & Tutorials*, IEEE, v. 26, n. 1, p. 160–195, 2023. Citado 2 vezes nas páginas 24 e 25.
- 15 RAFIQUE, W. et al. Complementing iot services through software defined networking and edge computing: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, IEEE, v. 22, n. 3, p. 1761–1804, 2020. Citado 2 vezes nas páginas 24 e 25.

- 16 MAMANE, A. et al. Scheduling algorithms for 5g networks and beyond: Classification and survey. *IEEE Access*, IEEE, v. 10, p. 51643–51661, 2022. Citado 2 vezes nas páginas 24 e 25.
- 17 LUO, Q. et al. Resource scheduling in edge computing: A survey. *IEEE Communications Surveys & Tutorials*, IEEE, v. 23, n. 4, p. 2131–2165, 2021. Citado 2 vezes nas páginas 24 e 25.
- 18 VIJAYALAYAN, K. S.; HARWOOD, A.; KARUNASEKERA, S. Distributed scheduling schemes for wireless mesh networks: A survey. *ACM Computing Surveys*, ACM, v. 46, n. 1, p. 1–34, 2013. Citado 2 vezes nas páginas 24 e 25.
- 19 WANG, A. et al. Software-defined networking enhanced edge computing: A network-centric survey. *Proceedings of the IEEE*, IEEE, v. 107, n. 8, p. 1500–1519, 2019. Citado 2 vezes nas páginas 24 e 25.
- 20 JAZAERI, S. S. et al. Edge computing in sdn-iot networks: a systematic review of issues, challenges and solutions. *Cluster Computing*, Springer, v. 24, p. 3187–3228, 2021. Citado 2 vezes nas páginas 24 e 25.
- 21 TALEB, T. et al. On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration. *IEEE Communications Surveys & Tutorials*, IEEE, v. 19, n. 3, p. 1657–1681, 2017. Citado 2 vezes nas páginas 24 e 25.
- 22 BRERETON, P. et al. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, ELSEVIER, v. 80, n. 4, p. 571–583, 2007. Citado na página 25.
- 23 STRAUB, D. W. et al. Normative standards for is research. *SIGMIS Database*, ACM, v. 25, n. 1, p. 21–34, 1994. Citado na página 33.
- 24 FREYNE, J. et al. Relative status of journal and conference publications in computer science. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 53, n. 11, p. 124–132, nov 2010. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/1839676.1839701>>. Citado na página 34.
- 25 NING, Z. et al. Joint computation offloading, power allocation, and channel assignment for 5g-enabled traffic management systems. *IEEE Transactions on Industrial Informatics*, IEEE, v. 15, n. 5, p. 3058–3067, 2019. Citado 13 vezes nas páginas 37, 38, 41, 44, 45, 48, 49, 50, 52, 55, 56, 61 e 62.
- 26 SAHNI, Y.; CAO, J.; YANG, L. Data-aware task allocation for achieving low latency in collaborative edge computing. *IEEE Internet of Things Journal*, IEEE, v. 6, n. 2, p. 3512–3524, 2019. Citado 12 vezes nas páginas 37, 41, 44, 45, 48, 49, 52, 53, 55, 56, 61 e 62.
- 27 SAHNI, Y. et al. Multi-hop multi-task partial computation offloading in collaborative edge computing. *IEEE Transactions on Parallel and Distributed Systems*, IEEE, v. 32, n. 5, p. 1133–1145, 2021. Citado 12 vezes nas páginas 37, 41, 44, 45, 48, 49, 52, 53, 55, 56, 60 e 62.
- 28 GE, C. et al. Qoe-assured 4k http live streaming via transient segment holding at mobile edge. *IEEE Journal on Selected Areas in Communications*, IEEE, v. 36, n. 8, p. 1816–1830, 2018. Citado 12 vezes nas páginas 37, 38, 41, 44, 46, 48, 49, 52, 53, 55, 58 e 62.

- 29 DIRO, A. A.; REDA, H. T.; CHILAMKURTI, N. Differential flow space allocation scheme in sdn based fog computing for iot applications. *Journal of Ambient Intelligence and Humanized Computing*, Springer, 2018. Citado 10 vezes nas páginas 37, 41, 44, 46, 48, 49, 52, 55, 58 e 62.
- 30 AUJLA, G. S. et al. An ensembled scheme for qos-aware traffic flow management in software defined networks. *2018 IEEE International Conference on Communications (ICC)*, IEEE, p. 1–7, 2018. Citado 12 vezes nas páginas 37, 38, 40, 44, 46, 48, 49, 52, 53, 55, 57 e 62.
- 31 HONG, C.-H. et al. qcon: Qos-aware network resource management for fog computing. *Sensors*, MDPI, v. 18, n. 10, 2018. Citado 12 vezes nas páginas 37, 42, 44, 45, 46, 48, 49, 52, 53, 55, 58 e 62.
- 32 NASIMI, M. et al. Edge-assisted congestion control mechanism for 5g network using software-defined networking. *15th International Symposium on Wireless Communication Systems*, IEEE, p. 1–5, 2018. Citado 11 vezes nas páginas 37, 38, 41, 44, 49, 52, 55, 56, 59, 60 e 62.
- 33 ZHANG, S. et al. Adaptive configuration selection and bandwidth allocation for edge-based video analytics. *IEEE/ACM Transactions on Networking*, IEEE, v. 30, n. 1, p. 285–298, 2022. Citado 12 vezes nas páginas 37, 38, 43, 44, 47, 48, 49, 51, 52, 55, 57 e 62.
- 34 SUN, J. et al. Efficient algorithm for traffic engineering in cloud-of-things and edge computing. *Computers & Electrical Engineering*, ELSEVIER, v. 69, p. 610–627, 2018. Citado 12 vezes nas páginas 37, 38, 43, 44, 46, 48, 49, 52, 53, 55, 58 e 62.
- 35 USMAN, M. et al. Taxonomies in software engineering: A systematic mapping study and a revised taxonomy development method. *Information and Software Technology*, v. 85, p. 43–59, 2017. ISSN 0950-5849. Citado na página 40.
- 36 QU, C.; CALHEIROS, R. N.; BUYYA, R. Auto-scaling web applications in clouds: A taxonomy and survey. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 51, n. 4, jul 2018. ISSN 0360-0300. Citado na página 40.
- 37 LIU, X.; BUYYA, R. Resource management and scheduling in distributed stream processing systems: A taxonomy, review, and future directions. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 53, n. 3, may 2020. ISSN 0360-0300. Disponível em: <<https://doi.org/10.1145/3355399>>. Citado na página 40.
- 38 ALHARBI, F.; FEI, Z. An sdn architecture for improving throughput of large flows using multipath tcp. p. 111–116, 2018. Citado 9 vezes nas páginas 40, 44, 45, 48, 49, 52, 55, 57 e 62.
- 39 WANG, C.; KIM, K.; YOUN, H. Popflow: a novel flow management scheme for sdn switch of multiple flow tables based on flow popularity. v. 14, 2020. Citado 9 vezes nas páginas 40, 44, 45, 48, 49, 52, 55, 61 e 62.
- 40 SELVI, M. et al. A hybrid framework for tcp incast congestion control in data center networks. *International Journal of Recent Technology and Engineering (IJRTE)*, v. 8, p. 797–806, 07 2019. Citado 10 vezes nas páginas 40, 44, 45, 48, 49, 52, 55, 56, 61 e 62.

- 41 YUAN, K.; WANG, F.; MARSZALEK, Z. A delay-tolerant data congestion avoidance algorithm for enterprise cloud system based on modular computing. p. 617–627, 2022. Citado 12 vezes nas páginas 40, 44, 46, 48, 49, 52, 53, 55, 57, 59, 61 e 62.
- 42 ISMAIL, A. H. et al. Congestion-aware and energy-efficient mec model with low latency for 5g. p. 156–159, 2019. Citado 11 vezes nas páginas 41, 44, 47, 48, 49, 50, 52, 55, 56, 61 e 62.
- 43 SALAMA, G. M. et al. Congestion-aware multiaccess edge computing collaboration model for 5g. *International Journal of Communication Systems*, v. 33, n. 12, p. e4446, 2020. Citado 11 vezes nas páginas 41, 44, 45, 48, 49, 50, 52, 55, 56, 61 e 62.
- 44 HUANG, P.-H. et al. Prioritized traffic shaping for low-latency mec flows in mec-enabled cellular networks. p. 120–125, 2022. Citado 11 vezes nas páginas 41, 44, 46, 48, 49, 50, 52, 55, 57, 58 e 62.
- 45 WANG, X.; XIE, H.; LUI, J. C. Analyzing queueing problems via bandits with linear reward & nonlinear workload fairness. *IEEE Transactions on Mobile Computing*, v. 23, n. 4, p. 3410–3423, 2023. Citado 10 vezes nas páginas 41, 44, 45, 48, 49, 52, 55, 57, 60 e 62.
- 46 HAN, Z. et al. A novel udt-based transfer speed-up protocol for fog computing. *Wirel. Commun. Mob. Comput.*, John Wiley and Sons Ltd., GBR, v. 2018, jan 2018. ISSN 1530-8669. Disponível em: <<https://doi.org/10.1155/2018/3681270>>. Citado 9 vezes nas páginas 41, 44, 45, 48, 49, 52, 55, 58 e 62.
- 47 MORAES, P. F.; MARTINS, J. S. B. A pub/sub sdn-integrated framework for iot traffic orchestration. Association for Computing Machinery, New York, NY, USA, 2019. Disponível em: <<https://doi.org/10.1145/3341325.3342001>>. Citado 9 vezes nas páginas 41, 44, 46, 48, 49, 52, 55, 56 e 62.
- 48 YIN, X. et al. Cooperative transmission with priority and fairness in edge computing systems. *IEEE Access*, v. 7, p. 44059–44069, 2019. Citado 11 vezes nas páginas 41, 44, 46, 48, 49, 52, 53, 55, 56, 61 e 62.
- 49 GAO, Y. et al. Latency-aware partial task offloading in collaborative edge computing. p. 1826–1831, 2023. Citado 10 vezes nas páginas 41, 44, 45, 48, 49, 52, 57, 59, 60 e 62.
- 50 YUN, S.; CHEN, Y. Intelligent traffic scheduling for mobile edge computing in iot via deep learning. *CMES - Computer Modeling in Engineering and Sciences*, v. 134, n. 3, p. 1815–1835, 2022. ISSN 1526-1492. Citado 11 vezes nas páginas 41, 44, 47, 48, 50, 52, 54, 55, 56, 60 e 62.
- 51 YAN, H. et al. Load balance traffic scheduling for live streaming services in edge. p. 25–32, 2022. Citado 10 vezes nas páginas 42, 44, 46, 48, 49, 52, 57, 58, 61 e 62.
- 52 ATANASOV, I.; PENCHEVA, E. Application level extension of bandwidth management in radio access network. *International Journal of Reasoning-based Intelligent Systems*, v. 12, p. 158, 01 2020. Citado 9 vezes nas páginas 42, 44, 46, 48, 49, 52, 55, 59 e 62.
- 53 XIANG, Z. et al. Computing power allocation and traffic scheduling for edge service provisioning. p. 394–403, 2020. Citado 10 vezes nas páginas 42, 44, 45, 48, 49, 52, 55, 56, 61 e 62.

- 54 LI, S. et al. Delay-aware task congestion control and resource allocation in mobile edge computing. p. 1–6, 2019. Citado 12 vezes nas páginas 42, 44, 45, 48, 49, 50, 52, 55, 56, 59, 60 e 62.
- 55 DLAMINI, T.; VILAKATI, S. Lstm-based traffic load balancing and resource allocation for an edge system. *Wireless Communications and Mobile Computing*, v. 2020, n. 1, p. 8825396, 2020. Citado 11 vezes nas páginas 42, 44, 47, 48, 50, 52, 54, 55, 56, 61 e 62.
- 56 SELLAMI, B. et al. Deep reinforcement learning for energy-efficient task scheduling in sdn-based iot network. p. 1–4, 2020. Citado 12 vezes nas páginas 42, 44, 47, 48, 49, 51, 52, 54, 55, 56, 61 e 62.
- 57 LATIF, Z. et al. Sdblockedge: Sdn-blockchain enabled multihop task offloading in collaborative edge computing. *IEEE Sensors Journal*, v. 22, n. 15, p. 15537–15548, 2022. Citado 10 vezes nas páginas 42, 44, 46, 48, 49, 52, 55, 57, 61 e 62.
- 58 SARANYA, G.; SASIKALA, E. Task sequencing in heterogeneous device for improved offloading decision using optimization technique. *Measurement: Sensors*, v. 24, p. 100446, 2022. ISSN 2665-9174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2665917422000800>>. Citado 11 vezes nas páginas 42, 44, 47, 48, 49, 52, 53, 55, 57, 60 e 62.
- 59 JIANG, J. et al. Towards an analysis of traffic shaping and policing in fog networks using stochastic fluid models. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, p. 196–204, 2017. Disponível em: <<https://doi.org/10.1145/3144457.3144496>>. Citado 9 vezes nas páginas 42, 44, 45, 48, 49, 52, 55, 59 e 62.
- 60 SETHOM, K. Colored ant for flow management in fog computing. p. 1–5, 2021. Citado 10 vezes nas páginas 42, 44, 47, 48, 49, 52, 53, 55, 58 e 62.
- 61 SERDAROGLU, K. C.; BAYDERE, S. An efficient multipriority data packet traffic scheduling approach for fog of things. *IEEE Internet of Things Journal*, v. 9, n. 1, p. 525–534, 2022. Citado 11 vezes nas páginas 42, 44, 46, 48, 49, 52, 55, 57, 58, 61 e 62.
- 62 HABEEB, F. et al. Dynamic bandwidth slicing for time-critical iot data streams in the edge-cloud continuum. *IEEE Transactions on Industrial Informatics*, v. 18, n. 11, p. 8017–8026, 2022. Citado 12 vezes nas páginas 43, 44, 47, 48, 49, 52, 53, 55, 57, 58, 61 e 62.
- 63 ZHANG, M. et al. Ents: An edge-native task scheduling system for collaborative edge computing. Los Alamitos, CA, USA, p. 149–161, dec 2022. Citado 10 vezes nas páginas 43, 44, 47, 48, 49, 52, 55, 57, 59 e 62.
- 64 KWAK, J. et al. Intellitc: Intelligent inter-datacenter traffic controller for internet of everything service based on fog computing. *IET Communications*, v. 14, 01 2020. Citado 9 vezes nas páginas 43, 44, 46, 48, 49, 52, 55, 58 e 62.
- 65 ZHAO, Y. et al. Network resource scheduling for cloud/edge data centers. p. 1–4, 2020. Citado 9 vezes nas páginas 43, 44, 45, 48, 50, 52, 55, 58 e 62.

- 66 ARSHED, J. U. et al. Ga-irace: Genetic algorithm-based improved resource aware cost-efficient scheduler for cloud fog computing environment. *Wireless Communications and Mobile Computing*, v. 2022, n. 1, p. 6355192, 2022. Citado 9 vezes nas páginas 43, 44, 47, 48, 54, 55, 56, 61 e 62.
- 67 LEI, C.; MINZHI, J. Intelligent warehouse task information flow scheduling under fog computing environment. *Journal of Physics: Conference Series*, v. 1939, p. 012065, 05 2021. Citado 10 vezes nas páginas 43, 44, 48, 49, 52, 53, 55, 57, 61 e 62.
- 68 FAHAD, M. et al. A multi-queue priority-based task scheduling algorithm in fog computing environment. *Concurrency and Computation: Practice and Experience*, v. 34, n. 28, p. e7376, 2022. Citado 11 vezes nas páginas 43, 44, 47, 48, 49, 52, 53, 55, 57, 61 e 62.
- 69 AHAMMAD, I.; KHAN, M. A. R.; SALEHIN, Z. U. Qos performance enhancement policy through combining fog and sdn. *Simulation Modelling Practice and Theory*, v. 109, p. 102292, 2021. ISSN 1569-190X. Citado 10 vezes nas páginas 43, 44, 46, 48, 49, 52, 55, 56, 61 e 62.
- 70 WEN, Z. et al. Janus: Latency-aware traffic scheduling for iot data streaming in edge environments. *IEEE Transactions on Services Computing*, IEEE Computer Society, Los Alamitos, CA, USA, v. 16, n. 06, p. 4302–4316, nov 2023. ISSN 1939-1374. Citado 8 vezes nas páginas 43, 44, 45, 48, 49, 55, 56 e 62.