



**INSTITUTO
FEDERAL**
Paraíba

Instituto Federal de Educação, Ciência e Tecnologia da Paraíba

Coordenação do Curso de Engenharia Elétrica

VINÍCIUS SILVA SIMÃO

**CONTROLE DE REDE EM MALHA FECHADA PARA COMPUTAÇÃO
DE BORDA INDUSTRIAL EM REDES 5G PRIVADAS: UMA
ABORDAGEM BASEADA EM IA E TELEMETRIA PARA
APLICAÇÕES SENSÍVEIS A LATÊNCIA**

JOÃO PESSOA

2026

Instituto Federal de Educação, Ciência e Tecnologia da Paraíba

Coordenação do Curso de Engenharia Elétrica

**Controle de Rede em Malha Fechada para Computação de
Borda Industrial em Redes 5G Privadas: Uma Abordagem
Baseada em IA e Telemetria para Aplicações Sensíveis a
Latência**

Trabalho de Conclusão de Curso submetido à
Coordenação do Curso de Bacharelado em En-
genharia Elétrica do Instituto Federal de Edu-
cação, Ciência e Tecnologia da Paraíba, como
parte dos requisitos para a obtenção do grau de
Engenheiro Eletricista.

JOÃO PESSOA

2026

Dados Internacionais de Catalogação na Publicação – CIP
Biblioteca Nilo Peçanha – IFPB, *Campus* João Pessoa

S588c Simão, Vinícius Silva.
 Controle de rede em malha fechada para computação de
 borda industrial em redes 5G privadas : uma abordagem
 baseada em IA e telemetria para aplicações sensíveis a latência
 / Vinícius Silva Simão. – 2026.
 35 f. : il.
 TCC (Curso Superior de Bacharelado em Engenharia
 Elétrica) – Instituto Federal de Educação da Paraíba / Unidade
 Acadêmica de Controle e Processos Industriais / Coordenação
 do Curso Superior de Bacharelado em Engenharia Elétrica,
 2026.
 Orientador: Profº Dr. Paulo Ditarso Maciel Jr.
 1. Redes 5G privadas. 2. Telemetria no plano de dados. 3.
 Modelo de predição baseado em IA. 4. Controle de redes em
 malha fechada. 5. Latência. I. Título.
CDU 004.7(043)

VINÍCIUS SILVA SIMÃO

POLO DE INOVAÇÃO - UNIDADE EMBAPII IF-PB

TRABALHO DE CONCLUSÃO DE CURSO submetido a Coordenação do Curso Superior de Bacharelado em Engenharia Elétrica, do Instituto Federal da Paraíba (IFPB), como parte dos requisitos institucionais para a obtenção do Título de **ENGENHEIRO ELETRICISTA**.

Aprovado em 10/02/2026.

Membros da Banca Examinadora:

Prof. Dr. PAULO DITARSO MACIEL JR.

Instituto Federal da Paraíba
Orientador

Prof. Dr. RUAN DELGADO GOMES

Instituto Federal da Paraíba
Examinador

Prof. Dr. LEANDRO CAVALCANTI DE ALMEIDA

Instituto Federal da Paraíba
Examinador

Documento assinado eletronicamente por:

- Paulo Ditarso Maciel Junior, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 12/02/2026 13:14:44.
- Ruan Delgado Gomes, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 12/02/2026 13:15:13.
- Leandro Cavalcanti de Almeida, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 12/02/2026 13:22:04.

Este documento foi emitido pelo SUAP em 12/02/2026. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifpb.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código 834930
Verificador: 7c0b0873a4
Código de Autenticação:



AGRADECIMENTOS

Agradeço a Deus, por me conceder saúde, força e perseverança ao longo de toda a minha trajetória acadêmica.

Agradeço à minha família, ao meu pai, à minha mãe e ao meu irmão, pelo apoio, incentivo e compreensão durante todo o período de realização deste curso.

Agradeço aos amigos e colegas de graduação, pelo companheirismo e pelas contribuições ao longo dessa caminhada acadêmica, com destaque para João Paulo, Mykaella Fernanda e Lucas Medeiros, pelo apoio, incentivo e parceria ao longo dessa trajetória.

Agradeço ao Instituto Federal da Paraíba (IFPB), pela formação acadêmica oferecida e por todas as oportunidades proporcionadas ao longo da graduação, fundamentais para a minha formação profissional.

Agradeço ao Polo de Inovação, pela oportunidade de participação no projeto de pesquisa EDGE5G, o qual contribuiu de forma significativa para o meu desenvolvimento técnico, científico e acadêmico.

Agradeço aos professores, pelos ensinamentos transmitidos e pela contribuição para a minha formação profissional.

Por fim, agradeço ao orientador Prof. Dr. Paulo Ditarso Maciel Junior, pela orientação, disponibilidade e pelas contribuições técnicas e acadêmicas essenciais para o desenvolvimento e a conclusão deste Trabalho de Conclusão de Curso.

RESUMO

A Indústria 4.0 demanda soluções de comunicação capazes de garantir conectividade confiável e de baixa latência para aplicações sensíveis ao tempo, como controle em tempo real, inspeção por vídeo e automação inteligente. Nesse contexto, as redes 5G privadas surgem como uma alternativa promissora, especialmente quando associadas à programabilidade de rede e à computação na borda. Este trabalho apresenta o desenvolvimento e a validação de um mecanismo de controle em malha fechada para ambientes industriais baseados em computação de borda e redes 5G privadas. A solução proposta integra telemetria em tempo real proveniente do plano de dados programável com um modelo de predição baseado em Inteligência Artificial, permitindo a detecção proativa de congestionamentos e a priorização dinâmica de fluxos sensíveis à latência. O sistema foi validado por meio de uma prova de conceito utilizando switches P4, núcleo 5G open source e uma aplicação real de vídeo industrial. Os resultados demonstraram reduções de até 41,04% na latência média e 75,17% no jitter em relação ao cenário sem controle, evidenciando a eficiência do controle orientado por telemetria aliado à IA para garantir requisitos de Qualidade de Serviço (QoS) em aplicações críticas da Indústria 4.0.

Palavras-chaves: Redes 5G Privadas, Telemetria no Plano de Dados, Modelo de Predição baseado em IA e Controle em Malha Fechada.

ABSTRACT

Industry 4.0 demands communication solutions capable of ensuring reliable, low-latency connectivity for time-sensitive applications such as real-time control, video inspection, and intelligent automation. In this context, private 5G networks emerge as a promising alternative, especially when combined with network programmability and edge computing. This work presents the development and validation of a closed-loop control mechanism for industrial environments based on edge computing and private 5G networks. The proposed solution integrates real-time telemetry from the programmable data plane with an Artificial Intelligence-based prediction model, enabling proactive congestion detection and dynamic prioritization of latency-sensitive flows. The system was validated through a proof of concept using P4 switches, an open-source 5G core, and a real industrial video application. The results demonstrated reductions of up to 41.04% in average latency and 75.17% in jitter compared to the baseline scenario, highlighting the effectiveness of telemetry-driven control combined with AI in ensuring Quality of Service (QoS) requirements for critical Industry 4.0 applications.

Key-words: Private 5G Networks, Data Plane Telemetry, AI-Based Prediction Model, Closed Loop Control

LISTA DE FIGURAS

Figura 1	– Caso de uso do Espelho em um cenário da Indústria 4.0. (CHAVES et al., 2025)	16
Figura 2	– Operação de Telemetria de Rede INT. (CABRAL et al., 2025)	18
Figura 3	– Operação de Telemetria de Rede ONT. (CABRAL et al., 2025)	18
Figura 4	– Visão geral do ambiente experimental. (SIMÃO et al., 2025)	22
Figura 5	– Resultado da latência ao longo do tempo em cada cenário. (SIMÃO et al., 2025)	29
Figura 6	– Resultado do jitter ao longo do tempo em cada cenário. (SIMÃO et al., 2025)	29
Figura 7	– Boxplot da latência para os três cenários experimentais. (SIMÃO et al., 2025)	30
Figura 8	– Boxplot do jitter para os três cenários experimentais. (SIMÃO et al., 2025) .	31

LISTA DE TABELAS

Tabela 1 – Análise comparativa entre os modelos SNARIMAX e AR	20
Tabela 2 – Análise dos dados dos resultados experimentais.	31

LISTA DE ABREVIATURAS E SIGLAS

5G	Quinta Geração de Redes Móveis
5GC	5G Core
AGVs	Automated Guided Vehicles
AR	Autoregressivo
BMv2	Behavioral Model version 2
CPU	Central Processing Unit
CU	Central Unit
DU	Distributed Unit
ESXi	VMware ESXi Hypervisor
IA	Inteligência Artificial
IIoT	Industrial Internet of Things
INT	In-band Network Telemetry
IQR	Interquartile Range
LT	Learning Time
LTP	Learning Time during Prediction
NFV	Network Functions Virtualization
ONT	Out-of-band Network Telemetry
P4	Programming Protocol-Independent Packet Processors
QoS	Quality of Service
RAN	Radio Access Network
RMSE	Root Mean Square Error
RU	Radio Unit
SBA	Service-Based Architecture
SDN	Software-Defined Networking

SNARIMAX Seasonal Nonlinear AutoRegressive Integrated Moving Average with eXogenous variables

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Objetivo Geral	13
1.2	Objetivos Específicos	13
1.3	Estrutura do Trabalho	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Redes 5G Privadas na Indústria 4.0	15
2.2	Aplicação Distribuída de Vídeo	15
2.3	Redes Programáveis	16
2.4	Telemetria de Plano de Dados	17
2.5	Modelos de Predição baseados em IA	19
3	MECANISMO DE CONTROLE DE REDES EM MALHA FECHADA	21
3.1	Visão Geral do Sistema	21
3.2	Coleta de Métricas	21
3.3	Integração com o Modelo de Predição	22
3.4	Arquitetura do Controlador	23
3.5	Fluxo de Operação	23
4	METODOLOGIA	24
4.1	Configuração do Ambiente de Testes (Testbed)	24
4.2	Ferramentas e Tecnologias Utilizadas	24
4.3	Cenários de Avaliação	25
4.3.1	Cenário 1 – Sem Controle	25
4.3.2	Cenário 2 – Controle Reativo	25
4.3.3	Cenário 3 – Controle Preditivo com IA	26
4.4	Métricas de Desempenho	26
4.5	Procedimentos Experimentais	26
5	RESULTADOS E DISCUSSÕES	28
5.1	Análise dos Resultados	28
5.2	Comparação Estatística entre os Cenários	28
5.3	Avaliação do Modelo de Predição	30
5.4	Ganhos de Desempenho	31
6	CONSIDERAÇÕES FINAIS	32
6.1	Trabalhos Futuros	32

REFERÊNCIAS BIBLIOGRÁFICAS	33
---	-----------

1 INTRODUÇÃO

A Quarta Revolução Industrial (Indústria 4.0) tem impulsionado a integração entre automação, Internet das Coisas Industrial (IIoT) e sistemas baseados em Inteligência Artificial (IA), demandando conectividade confiável e comunicação em tempo real para garantir o funcionamento adequado de aplicações críticas. Nesse contexto, as redes 5G privadas surgem como um elemento essencial, pois oferecem alta velocidade, baixa latência e comunicação segura, adaptada a ambientes industriais (JOHN et al., 2024). Essas redes possibilitam aplicações de vídeo em alta resolução para monitoramento, inspeção, controle remoto e análise em tempo real, as quais exigem requisitos rigorosos de Qualidade de Serviço (QoS), como baixa latência, alta confiabilidade e mínima perda de pacotes. A violação desses requisitos pode comprometer a segurança, a tomada de decisão e a qualidade dos produtos fabricados.

Para atender às exigências da Indústria 4.0, as redes 5G privadas combinam tecnologias como Software-Defined Networking (SDN), Network Function Virtualization (NFV) e Network Slicing, permitindo a alocação flexível e isolada de recursos. Contudo, ainda existem desafios relacionados à orquestração dinâmica de recursos e à manutenção do desempenho sob condições de tráfego intenso, uma vez que o compartilhamento de infraestrutura entre fluxos críticos e não críticos pode gerar contenção e atrasos, impactando negativamente aplicações sensíveis à latência (OCHONU; VIDAL, 2024).

O gerenciamento eficiente de tráfego sensível à latência envolve lidar com condições de rede variáveis, priorização dinâmica de fluxos e requisitos estritos de QoS. Abordagens tradicionais, como otimização de protocolos e agendamento estático, tornam-se ineficazes em cenários modernos com alto dinamismo de tráfego (MIRZAEINIA; MIRZAEINIA; REZGUI, 2020). Nesse sentido, as redes programáveis, habilitadas por SDN e linguagens como P4 (Programming Protocol-independent Packet Processors), permitem separar os planos de controle e dados, favorecendo a adaptação em tempo real e a inserção de lógicas de encaminhamento diretamente no hardware. Essa programabilidade também viabiliza a telemetria em rede, que coleta métricas em tempo real e fornece subsídios para decisões baseadas em IA, como detecção de anomalias e controle adaptativo.

Diante desse cenário, este trabalho propõe um mecanismo de controle de redes em malha fechada que combina telemetria em tempo real proveniente do plano de dados programável com um modelo preditivo baseado em IA para antecipar congestionamentos e ajustar dinamicamente os recursos da rede. O mecanismo foi validado por meio de uma prova de conceito utilizando uma aplicação real de distribuição de vídeo industrial, implementada sobre uma infraestrutura virtualizada com switches P4 e núcleo 5G open source, demonstrando reduções de até 41,04% na latência e 75,17% no jitter em comparação ao cenário sem controle.

1.1 Objetivo Geral

Desenvolver e validar um mecanismo de controle em malha fechada para redes 5G privadas em ambientes de computação de borda industrial, integrando telemetria em tempo real e predição baseada em Inteligência Artificial para otimizar o desempenho de aplicações sensíveis à latência.

1.2 Objetivos Específicos

- Implementar um ambiente experimental (testbed) com switches programáveis em P4 e núcleo 5G open source;
- Integrar um sistema de telemetria de plano de dados para coleta contínua de métricas de rede;
- Desenvolver e treinar um modelo preditivo baseado em IA para estimar o nível de congestionamento das filas de rede;
- Projetar um controlador capaz de atuar de forma reativa ou proativa com base nos dados coletados;
- Avaliar o impacto do mecanismo proposto sobre a latência e o jitter em cenários de transmissão de vídeo industrial.

1.3 Estrutura do Trabalho

O Capítulo 2 (Fundamentação Teórica) apresenta os conceitos essenciais que sustentam o desenvolvimento deste trabalho, incluindo Indústria 4.0, redes 5G privadas, computação de borda, redes programáveis, telemetria de plano de dados e modelos de predição baseados em IA. Esses tópicos fornecem a base teórica para o entendimento do mecanismo proposto.

O Capítulo 3 (Mecanismo de Controle de Redes em Malha Fechada) descreve a arquitetura desenvolvida, abordando a integração entre telemetria e IA no controle proativo e reativo de tráfego. São detalhados o funcionamento do sistema, o papel de cada componente e as estratégias adotadas para a priorização dinâmica de fluxos sensíveis à latência.

O Capítulo 4 (Metodologia Experimental) detalha a configuração do ambiente virtualizado (testbed) utilizado nos experimentos, incluindo os componentes de software e hardware empregados, a geração dos cenários de avaliação, as métricas analisadas (latência e jitter) e os procedimentos adotados para coleta e análise dos dados.

O Capítulo 5 (Resultados e Discussões) apresenta a análise dos resultados obtidos a partir dos experimentos realizados, comparando o comportamento da rede sob três cenários distintos: sem controle, com controle reativo e com controle proativo. Também são discutidos os ganhos

de desempenho alcançados e as contribuições do modelo proposto para aplicações industriais sensíveis à latência.

Por fim, o Capítulo 6 (Considerações Finais) sintetiza as principais conclusões e contribuições do trabalho, destacando os benefícios da integração entre telemetria e IA em redes programáveis. Além disso, são apontadas as limitações identificadas e sugeridas direções para trabalhos futuros, como o aprimoramento do modelo de predição e a avaliação do mecanismo em topologias mais complexas e escaláveis.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Redes 5G Privadas na Indústria 4.0

As redes 5G privadas são componentes essenciais da transformação digital impulsionada pela Indústria 4.0, pois oferecem alta velocidade, baixa latência e elevada densidade de conexões, características fundamentais para aplicações industriais críticas, como sistemas de controle em malha fechada, robôs móveis e veículos autônomos guiados (AGVs) (5G-ACIA, 2020). Diferentemente das redes públicas, as redes 5G privadas permitem que indústrias implementem sua própria infraestrutura, com slicing de rede, segurança reforçada e operações em espectro sem interferência, garantindo isolamento e confiabilidade nas comunicações (AIJAZ, 2020).

A arquitetura do 5G Core (5GC) é baseada em serviços independentes (Service-Based Architecture – SBA), enquanto a Rede de Acesso por Rádio (RAN) é dividida em Unidade de Rádio (RU), Unidade Centralizada (CU) e Unidade Distribuída (DU), o que proporciona maior flexibilidade no gerenciamento dos recursos (DAHLMAN; PARKVALL; SKÖLD, 2020). O uso de implementações open source, como o free5GC¹ e o OpenAirInterface², tem viabilizado experimentações e implantações reais de redes 5G privadas em ambientes industriais.

Entretanto, mesmo com tais avanços, garantir a Qualidade de Serviço (QoS) em toda a cadeia de comunicação, especialmente na rede de transporte compartilhada, ainda representa um desafio. Isso ocorre porque diferentes fluxos, como tráfego administrativo e de produção, competem pelos mesmos recursos, podendo gerar contenção e degradação de desempenho (MUZAFFAR et al., 2023).

As redes 5G se destacam pela capacidade de atender simultaneamente a aplicações com demandas heterogêneas. Enquanto a manufatura contínua requer ampla cobertura e maior tolerância à latência, a manufatura discreta exige baixa latência em áreas reduzidas (VARGA et al., 2020). Para aplicações industriais baseadas em vídeo em tempo real, o 5G é particularmente vantajoso, pois oferece alta taxa de transferência e confiabilidade (ZOU et al., 2023). Estudos experimentais comprovam que o 5G melhora significativamente o desempenho de comunicação, reduzindo latência e aumentando a eficiência de controle (BASARAS et al., 2023), (NAKIMULI et al., 2021).

2.2 Aplicação Distribuída de Vídeo

A distribuição de vídeo em tempo real é uma das aplicações mais críticas em ambientes industriais, usada para inspeção automatizada, monitoramento de processos, controle remoto e

¹ <https://free5gc.org/>

² <https://openairinterface.org/>

análise baseada em visão computacional.

Neste contexto, o middleware Espelho (NETO et al., 2024; CHAVES et al., 2025), foi utilizado como base para os experimentos deste trabalho. Desenvolvido para ambientes de computação de borda (edge computing), o Espelho abstrai a complexidade da transmissão de vídeo, permitindo que aplicações acessem fluxos de vídeo de forma transparente por meio de dispositivos virtuais Linux (`/dev/video*`), como se estivessem conectadas a câmeras locais, conforme ilustrado na Figura 1. Sua arquitetura segue o modelo publicador/assinante (pub/sub), composta por três módulos: Capture (captura o vídeo da câmera e envia à rede), Servidor Espelho (distribui os fluxos aos assinantes) e Decapture (recebe o fluxo e o expõe às aplicações).

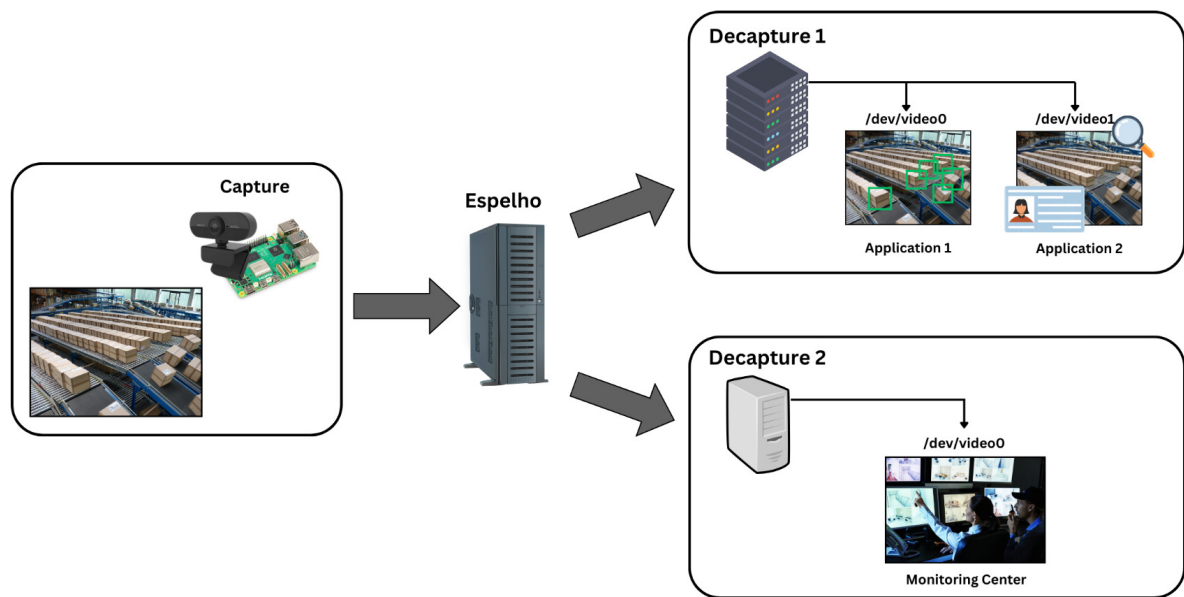


Figura 1 – Caso de uso do Espelho em um cenário da Indústria 4.0. (CHAVES et al., 2025)

Os estudos conduzidos em (NETO et al., 2024) demonstraram que o Espelho suporta acesso paralelo de múltiplos clientes com latência gerenciável e taxa de quadros completa em resoluções de até 2560×1440. Em resoluções superiores (3840×2160 e 4096×2160), observou-se que a limitação estava na capacidade de processamento da CPU, e não no middleware. Assim, o Espelho mostra-se adequado para aplicações de vídeo industrial em tempo real, especialmente quando integrado a redes 5G privadas e arquiteturas de controle baseadas em telemetria.

2.3 Redes Programáveis

As redes programáveis representam um novo paradigma na engenharia de redes, caracterizado pela capacidade de controlar e adaptar dinamicamente o comportamento da infraestrutura de comunicação por meio de software. Diferentemente das redes tradicionais, nas quais o plano de controle e o plano de dados são rigidamente integrados em dispositivos físicos, as Software-Defined Networks (SDN) promovem a separação lógica entre esses planos, possibi-

tando gerenciamento centralizado, automação e reconfiguração dinâmica dos fluxos de tráfego (RAY; KUMAR, 2021).

Essa arquitetura permite que o plano de controle (responsável pelas decisões de roteamento) seja implementado em controladores centrais, enquanto o plano de dados (responsável pelo encaminhamento dos pacotes) é executado nos switches e roteadores. Tal separação amplia a flexibilidade operacional, permitindo que as políticas de encaminhamento sejam ajustadas conforme as condições da rede, algo essencial em aplicações sensíveis à latência.

Com o avanço das SDNs, surgiu a necessidade de programabilidade diretamente no plano de dados, o que levou ao desenvolvimento da linguagem P4 (Programming Protocol-Independent Packet Processors)³. Essa linguagem foi projetada para descrever de forma independente de protocolo como os pacotes devem ser processados dentro dos dispositivos de rede. O P4 possibilita implementar lógicas de encaminhamento diretamente no hardware, sem depender de controladores externos, o que reduz a latência de resposta e aumenta a eficiência do controle.

Por meio do P4, os switches tornam-se inteligentes e configuráveis em tempo de execução, capazes de realizar operações como marcação de pacotes, classificação dinâmica de fluxos e coleta de metadados, funcionalidades fundamentais para o controle adaptativo de tráfego em redes 5G privadas e ambientes industriais da Indústria 4.0. Essa capacidade de processamento distribuído no plano de dados permite construir mecanismos de controle de malha fechada, em que decisões são tomadas localmente com base em métricas atualizadas de rede.

2.4 Telemetria de Plano de Dados

A evolução dos planos de dados programáveis também trouxe avanços na forma como as métricas de rede são coletadas e analisadas. A telemetria de plano de dados fornece visibilidade em tempo real sobre o estado da rede, permitindo monitorar congestionamentos, atrasos, perdas e ocupação de filas diretamente nos dispositivos de encaminhamento. Com isso, é possível reagir de forma rápida e precisa a variações nas condições do tráfego (ARSLAN; MCKEOWN, 2019).

Na prática, existem duas abordagens principais para a coleta de dados de telemetria: In-band Network Telemetry (INT) e Out-of-band Network Telemetry (ONT).

A INT é definida pela especificação do P4 Consortium (P4, 2021) e consiste na inserção de instruções de telemetria diretamente nos pacotes da aplicação. À medida que os pacotes atravessam a rede, cada *switch* insere informações de telemetria, tais como identificação do dispositivo, *timestamp*, atraso, profundidade de fila e portas de entrada e saída, conforme ilustrado na Figura 2. Essa abordagem fornece dados de alta granularidade, permitindo análises detalhadas do caminho percorrido por cada fluxo. No entanto, a principal limitação do INT está no overhead introduzido, uma vez que cada pacote transporta metadados adicionais, há aumento no tamanho total dos pacotes e potencial degradação do desempenho (CABRAL et al., 2025).

³ <https://p4.org/>

sensíveis à latência.

2.5 Modelos de Predição baseados em IA

Modelos de predição consistem em técnicas de aprendizado de máquina capazes de estimar valores futuros a partir de padrões observados em dados históricos. Esses modelos podem empregar tanto abordagens baseadas em redes neurais quanto técnicas clássicas de aprendizado de máquina, como Random Forest e Máquinas de Vetores de Suporte, dependendo das características dos dados e do objetivo da análise. Em diversos contextos, tais modelos são aplicados a séries temporais, nas quais as informações variam ao longo do tempo e apresentam dependência entre amostras consecutivas.

No contexto deste trabalho, foi adotado um modelo baseado em combinação linear, que utiliza autoregressão e diferenciação para capturar tendências e padrões presentes nos dados históricos. Esse modelo pertence à classe dos modelos autorregressivos (AR) e faz parte da suíte SNARIMAX, disponibilizada pela biblioteca River⁴, a qual é projetada especificamente para aprendizado online. Essa característica permite que o modelo seja continuamente atualizado à medida que novos dados são recebidos durante a execução do sistema, sem a necessidade de reprocessar todo o conjunto de dados previamente coletado.

A principal vantagem da biblioteca River reside justamente em sua capacidade de aprendizado incremental, tornando-a especialmente adequada para cenários dinâmicos, como redes de comunicação, nos quais as condições de tráfego podem variar de forma frequente e imprevisível. Diferentemente dos modelos tradicionais de aprendizado offline, que exigem etapas explícitas de reentrenamento sempre que novos dados são incorporados, o aprendizado online possibilita uma adaptação contínua com baixo custo computacional.

A escolha do modelo autorregressivo com integração (SNARIMAX) deve-se principalmente à sua capacidade de operação em tempo real, bem como ao desempenho superior observado em comparação com outros modelos avaliados preliminarmente, como Random Forest e Regressão Linear. Esses modelos alternativos, embora apresentem bom desempenho em cenários estáticos, não oferecem suporte nativo a aprendizado contínuo e demandam reentrenamento frequente, o que resulta em tempos de execução elevados e os torna inviáveis para aplicações sensíveis à latência.

Para validar essa escolha, foi realizada uma análise comparativa entre o modelo SNARIMAX da biblioteca River e um modelo AR implementado na biblioteca Statsmodels. A avaliação considerou métricas como o erro quadrático médio (RMSE), o tempo de aprendizado (LT) e o tempo de aprendizado durante a predição (LTP), utilizando uma janela de dados previamente definida. O experimento foi conduzido em um cenário de predição da ocupação de filas de

⁴ <https://riverml.xyz/latest/api/time-series/SNARIMAX/>

roteadores em uma rede 5G, considerando um total de 217.281 amostras para treinamento e 500 amostras para predição.

Os resultados dessa comparação, apresentados na Tabela 1, indicam que, embora o modelo AR apresente um RMSE ligeiramente inferior, seu tempo de processamento é significativamente maior, especialmente durante a fase de predição, tornando-o impraticável para uso online. Em contraste, o modelo SNARIMAX demonstrou tempos de aprendizado e predição compatíveis com os requisitos de sistemas em malha fechada, mesmo apresentando um erro de predição marginalmente superior.

Tabela 1 – Análise comparativa entre os modelos SNARIMAX e AR

Modelo	RMSE	LT	LTP
SNARIMAX	22.89	29.94 s	0.08 s
AR	17.52	40.59 s	1439.83 s

Dessa forma, o modelo SNARIMAX mostrou-se mais adequado para a aplicação proposta neste trabalho, uma vez que alia desempenho preditivo satisfatório à capacidade de adaptação contínua e baixa latência computacional, características essenciais para a antecipação de congestionamentos e a tomada de decisão proativa em redes programáveis.

3 MECANISMO DE CONTROLE DE REDES EM MALHA FECHADA

A garantia de requisitos rigorosos de QoS em aplicações industriais sensíveis à latência demanda mecanismos de controle capazes de monitorar continuamente o estado da rede e reagir de forma rápida e precisa às variações de tráfego. Nesse contexto, este capítulo apresenta o mecanismo de controle em malha fechada proposto neste trabalho, que combina telemetria de plano de dados e predição baseada em Inteligência Artificial para a gestão dinâmica de recursos em redes 5G privadas. A solução foi concebida para atuar diretamente sobre a rede de transporte compartilhada, mitigando congestionamentos e priorizando fluxos críticos de forma reativa e preditiva, conforme detalhado a seguir.

3.1 Visão Geral do Sistema

O mecanismo proposto adota uma arquitetura de controle em malha fechada, na qual informações coletadas em tempo real no plano de dados são utilizadas para tomar decisões de controle que impactam diretamente o comportamento da rede. A solução integra três componentes principais: (i) telemetria de rede, responsável pela coleta contínua de métricas; (ii) modelo de predição baseado em IA, que antecipa condições futuras de congestionamento; e (iii) controlador de rede, encarregado de aplicar políticas de priorização e ajuste de filas nos dispositivos programáveis.

O ambiente considerado é composto por uma rede 5G privada, na qual o tráfego gerado por aplicações industriais, após atravessar a RAN e o núcleo 5G, é encaminhado por uma rede de transporte compartilhada, sujeita à concorrência com fluxos prioritários e não prioritários. Essa coexistência de tráfegos com diferentes requisitos torna necessária a aplicação de mecanismos inteligentes de controle para evitar degradações de desempenho em fluxos sensíveis à latência.

A Figura 4 ilustra a arquitetura geral do sistema proposto, destacando a separação entre o tráfego de aplicação, tráfego de fundo e o tráfego de telemetria, bem como o fluxo de informações que caracteriza o ciclo de controle em malha fechada.

3.2 Coleta de Métricas

A coleta de informações sobre o estado da rede é realizada por meio de telemetria no plano de dados, explorando as capacidades oferecidas por switches programáveis em P4. Essa abordagem permite que métricas relevantes sejam obtidas diretamente nos dispositivos de encaminhamento, sem a necessidade de consultas frequentes ao plano de controle, reduzindo a latência de resposta do sistema.

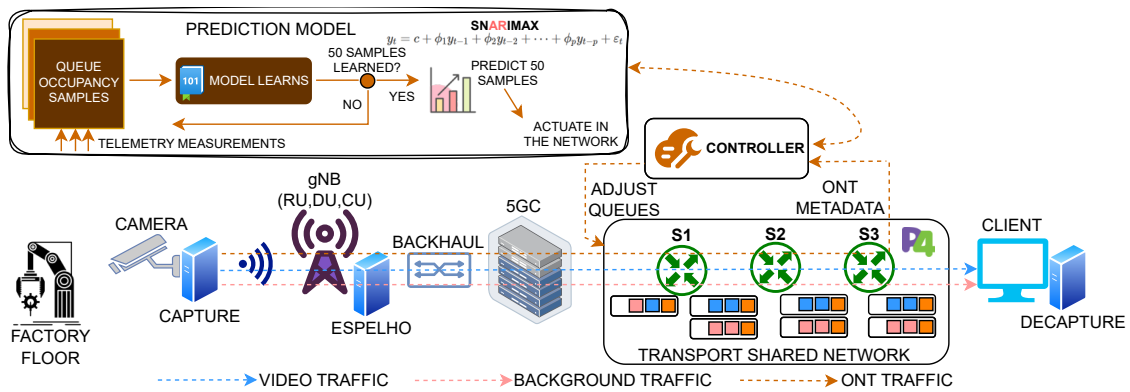


Figura 4 – Visão geral do ambiente experimental. (SIMÃO et al., 2025)

Neste trabalho, foi adotada a abordagem de Out-of-band Network Telemetry (ONT), na qual pacotes de sondagem dedicados percorrem a rede coletando informações como ocupação de filas, timestamps e identificação dos dispositivos atravessados. Diferentemente da In-band Network Telemetry (INT), essa estratégia não altera os pacotes da aplicação, evitando aumento de overhead, fragmentação e impactos negativos sobre o tráfego de produção.

Os dados coletados pelos pacotes de telemetria são enviados ao controlador, que pode utilizá-los tanto para decisões imediatas (controle reativo) quanto como entrada para o modelo de predição. A separação entre tráfego de aplicação e tráfego de monitoramento torna a ONT especialmente adequada para ambientes industriais, nos quais a previsibilidade e a estabilidade do desempenho são fundamentais.

3.3 Integração com o Modelo de Predição

Além do monitoramento reativo, o mecanismo proposto incorpora um modelo de predição baseado em Inteligência Artificial, com o objetivo de antecipar situações de congestionamento antes que impactem as aplicações críticas. O modelo utilizado é o SNARIMAX, pertencente à biblioteca River, escolhido por sua capacidade de aprendizado contínuo e baixa latência de predição.

O modelo recebe como entrada séries temporais provenientes da telemetria, em especial a ocupação das filas dos switches da rede de transporte. A partir de uma janela de amostras recentes, o SNARIMAX é capaz de prever o comportamento futuro da fila em um horizonte de tempo pré-definido, permitindo ao controlador atuar de forma proativa.

Essa integração entre telemetria e predição transforma o controle de rede em um processo antecipativo, no qual as decisões não se baseiam apenas no estado atual, mas em estimativas confiáveis do estado futuro da rede. Essa característica é essencial para aplicações industriais sensíveis à latência, nas quais ações tardias podem comprometer a estabilidade do sistema.

3.4 Arquitetura do Controlador

O controlador de rede é responsável por analisar as métricas coletadas e as previsões geradas pelo modelo de IA, tomando decisões de ajuste nos dispositivos de encaminhamento. A estratégia adotada baseia-se na priorização dinâmica de filas, explorando os recursos de programação do plano de dados dos switches P4 com BMv2.

Cada switch da rede de transporte opera com múltiplas filas, sendo uma destinada ao tráfego crítico (por exemplo, vídeo industrial em tempo real) e outra ao tráfego de fundo. O controlador ajusta dinamicamente os parâmetros de serviço dessas filas, como taxas de atendimento, de acordo com o nível de ocupação observado ou previsto.

Quando um limiar pré-definido de ocupação de fila é atingido ou previsto, o controlador aplica ajustes somente no dispositivo afetado, evitando ações globais desnecessárias e reduzindo o impacto sobre o restante da rede. Essa abordagem localizada contribui para a escalabilidade do mecanismo e para a preservação da estabilidade do sistema.

3.5 Fluxo de Operação

O funcionamento do mecanismo de controle ocorre de forma cíclica, caracterizando uma malha fechada contínua. Inicialmente, os switches enviam informações de estado por meio da telemetria ONT. Esses dados são analisados pelo controlador, que pode atuar de duas formas distintas: reativa ou preditiva.

No modo reativo, o controlador atua apenas quando as métricas atuais indicam degradação de desempenho, ajustando os parâmetros das filas após a ocorrência do congestionamento. Já no modo preditivo, os dados de telemetria alimentam o modelo SNARIMAX, que antecipa possíveis congestionamentos, permitindo que os ajustes sejam realizados antes que o problema se manifeste.

Essa abordagem permite avaliar os ganhos proporcionados pela predição baseada em IA, tema que será explorado nos capítulos seguintes por meio da metodologia experimental e da análise dos resultados obtidos.

4 METODOLOGIA

Neste capítulo, é apresentada a metodologia adotada para a validação do mecanismo de controle em malha fechada proposto. São descritos o ambiente experimental, as ferramentas utilizadas, os cenários de avaliação definidos, as métricas analisadas e os procedimentos experimentais seguidos para garantir a reprodutibilidade e a consistência dos resultados.

4.1 Configuração do Ambiente de Testes (Testbed)

O ambiente de testes foi implementado em um testbed virtualizado, projetado para reproduzir um cenário realista de computação de borda industrial sobre redes privadas 5G. A infraestrutura foi executada sobre a plataforma de virtualização VMware ESXi, permitindo a instânciação flexível dos componentes de rede, computação e aplicação.

O núcleo da rede 5G foi implementado utilizando o free5GC, enquanto a Rede de Acesso Rádio (RAN) foi emulada por meio do UERANSIM, possibilitando a geração de tráfego compatível com arquiteturas 5G reais. Após atravessar o domínio 5G, o tráfego segue para a rede de transporte industrial, a qual foi modelada com switches programáveis BMv2, compatíveis com a linguagem P4.

Essa rede de transporte é compartilhada por fluxos críticos (vídeo industrial em tempo real) e fluxos de fundo, reproduzindo condições de contenção típicas de ambientes industriais. A separação lógica de filas nos switches permitiu avaliar o impacto de diferentes estratégias de controle e priorização.

4.2 Ferramentas e Tecnologias Utilizadas

A metodologia experimental baseou-se na integração de múltiplas ferramentas e tecnologias, escolhidas por sua aderência a cenários reais de redes programáveis e computação de borda:

- **free5GC:** implementação open-source do núcleo de rede 5G, responsável pelo plano de controle e de dados do domínio móvel;
- **UERANIM:** emulador da RAN 5G, utilizado para simular dispositivos de usuários e enlaces rádio;
- **BMv2 (Behavioral Model v2):** switch P4 utilizado para emular a rede de transporte programável;

- **P4:** linguagem de programação do plano de dados, empregada para implementar a coleta de telemetria e o gerenciamento de filas;
- **Telemetria ONT:** mecanismo adotado para a coleta periódica do estado das filas, sem impactar o tráfego da aplicação;
- **River (SNARIMAX):** biblioteca de aprendizado de máquina online, utilizada para a implementação do modelo preditivo;
- **Espelho:** middleware de distribuição de vídeo em tempo real, empregado como aplicação industrial representativa;
- **iPerf3:** ferramenta usada para geração de tráfego de fundo controlado.

4.3 Cenários de Avaliação

Para avaliar o impacto do mecanismo proposto, foram definidos três cenários experimentais distintos. Todos os cenários compartilham a mesma topologia de rede, aplicação de vídeo e geração de tráfego de fundo, diferindo exclusivamente na presença e no tipo de mecanismo de controle adotado.

4.3.1 Cenário 1 – Sem Controle

No primeiro cenário, utilizado como linha de base, nenhum mecanismo de controle ou priorização dinâmica é aplicado. O tráfego de vídeo e o tráfego de fundo compartilham a rede de transporte sem ajustes.

Embora filas distintas estejam configuradas nos switches, ambas competem igualmente pelos recursos disponíveis, o que permite observar o impacto da contenção de tráfego sobre aplicações sensíveis à latência em um ambiente não controlado.

4.3.2 Cenário 2 – Controle Reativo

No segundo cenário, ativa-se o mecanismo de controle em malha fechada, operando de forma reativa. O tráfego de vídeo é alocado em uma fila de maior prioridade, enquanto o tráfego de fundo permanece em uma fila secundária.

As decisões de controle são baseadas diretamente nos dados coletados pela telemetria ONT. Sempre que a ocupação de uma fila ultrapassa um limiar pré-definido (50%), o controlador ajusta dinamicamente as taxas de serviço da fila correspondente. As ações são locais, afetando apenas o dispositivo congestionado, sem interferir nos demais elementos da rede.

4.3.3 Cenário 3 – Controle Preditivo com IA

O terceiro cenário estende o controle reativo ao incorporar um modelo de predição baseado em aprendizado de máquina. Os dados coletados pela telemetria ONT são encaminhados ao modelo SNARIMAX, que realiza previsões sobre a ocupação futura das filas.

Com base em uma janela de 50 amostras (equivalente a 10 segundos), o modelo prediz o comportamento das próximas 50 amostras. O controlador utiliza essas previsões para atuar de forma proativa, ajustando os parâmetros das filas antes que a congestão efetivamente ocorra.

Esse cenário permite avaliar os ganhos obtidos com a antecipação de eventos de degradação em comparação com a abordagem puramente reativa.

4.4 Métricas de Desempenho

A avaliação de desempenho foi conduzida com foco em métricas diretamente relacionadas à qualidade de serviço de aplicações sensíveis a atraso. As principais métricas analisadas foram:

- **Latência:** tempo total decorrido entre o envio e o recebimento dos pacotes de vídeo;
- **Jitter:** variação da latência entre pacotes consecutivos, um indicador crítico para aplicações de vídeo em tempo real.

Os dados foram coletados continuamente durante a execução dos experimentos, abrangendo períodos com e sem tráfego de fundo. As métricas foram extraídas tanto da aplicação de vídeo quanto dos mecanismos de monitoramento de rede, garantindo uma visão consistente do comportamento do sistema.

4.5 Procedimentos Experimentais

Os experimentos foram conduzidos de forma controlada e repetitiva, seguindo uma sequência padronizada e idêntica para todos os cenários avaliados. Inicialmente, o ambiente experimental foi inicializado sem a presença de tráfego de fundo, momento em que o fluxo de vídeo industrial foi ativado e monitorado continuamente.

Após esse período inicial, com duração de 1 minuto, o tráfego de fundo foi introduzido de maneira controlada por um intervalo de 2 minutos, utilizando 8 fluxos paralelos gerados pela ferramenta *iperf3*. Essa etapa teve como objetivo induzir condições de carga elevada na rede de transporte, reproduzindo cenários de contenção típicos de ambientes industriais.

O ciclo composto pelas fases sem tráfego de fundo e com tráfego de fundo foi repetido sucessivamente até completar um tempo total de execução de 10 minutos para cada cenário.

Todos os cenários experimentais foram executados sob as mesmas condições temporais e de carga, assegurando a comparabilidade direta entre os resultados obtidos.

Os dados coletados durante os experimentos foram posteriormente processados para a geração de séries temporais, gráficos comparativos e análises estatísticas, possibilitando a quantificação dos ganhos proporcionados pelos mecanismos de controle reativo e preditivo em relação ao cenário sem controle.

5 RESULTADOS E DISCUSSÕES

Este capítulo apresenta e analisa os resultados obtidos a partir da validação experimental do mecanismo de controle em malha fechada proposto. A avaliação foi conduzida considerando os três cenários definidos no Capítulo 4, com foco nas métricas de latência e jitter, fundamentais para aplicações industriais sensíveis ao tempo, como a distribuição de vídeo em tempo real.

5.1 Análise dos Resultados

A análise inicial considera o comportamento temporal das métricas de latência e jitter ao longo da execução dos experimentos. Os resultados evidenciam o impacto direto da introdução de tráfego de fundo na rede de transporte, bem como a capacidade dos mecanismos de controle em mitigar os efeitos da contenção de recursos. O comportamento da latência ao longo do tempo para os três cenários avaliados é apresentado na Figura 5.

No Cenário 1, em que não há qualquer mecanismo de controle, observa-se uma elevação significativa tanto da latência quanto do jitter durante os períodos de carga elevada. A ausência de priorização faz com que o tráfego de vídeo industrial concorra diretamente com o tráfego de fundo, resultando em oscilações acentuadas e comportamento instável, inadequado para aplicações críticas.

Já nos Cenários 2 e 3, que incorporam controle reativo e preditivo, respectivamente, as curvas de latência apresentam um comportamento substancialmente mais estável. Mesmo durante os intervalos em que o tráfego de fundo está ativo, os valores permanecem controlados, indicando a efetividade do mecanismo proposto.

De forma complementar, a Figura 6 apresenta o comportamento do jitter ao longo do tempo para os três cenários. Observa-se que, no cenário sem controle, o jitter sofre variações significativas durante os períodos de carga elevada, enquanto os cenários com controle reativo e preditivo apresentam menor variabilidade, evidenciando maior estabilidade na entrega dos pacotes e melhor adequação às aplicações sensíveis à latência.

5.2 Comparação Estatística entre os Cenários

A comparação entre os cenários foi aprofundada por meio da análise estatística das distribuições das métricas de latência e jitter, utilizando diagramas de caixa (*boxplots*). Essa abordagem permite avaliar, de forma mais detalhada, não apenas os valores médios, mas também a mediana, o intervalo interquartil (IQR), a dispersão dos dados e a presença de valores extremos. As Figuras 7 e 8 apresentam, respectivamente, a distribuição da latência e do jitter para os três cenários experimentais.

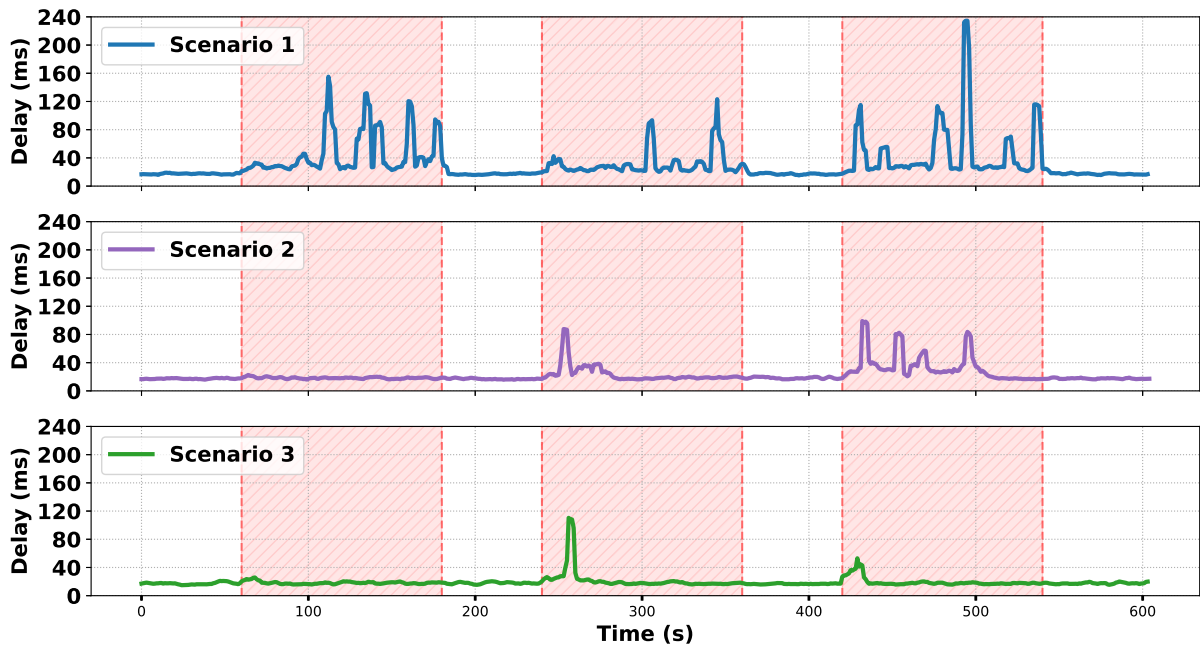


Figura 5 – Resultado da latência ao longo do tempo em cada cenário. (SIMÃO et al., 2025)

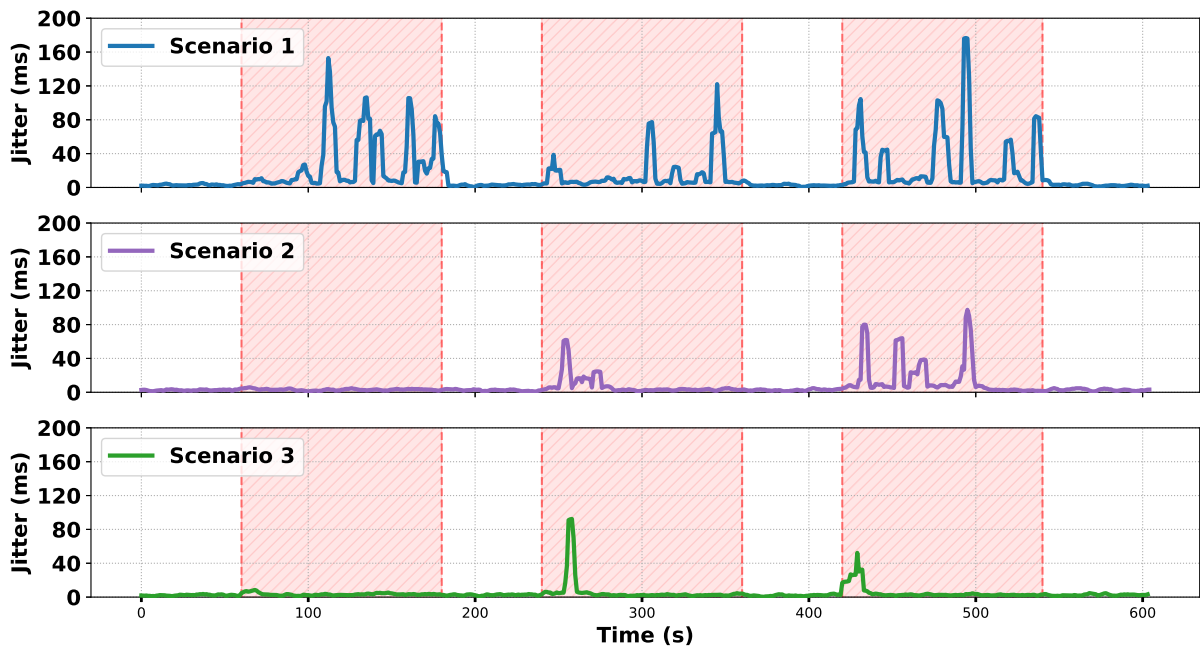


Figura 6 – Resultado do jitter ao longo do tempo em cada cenário. (SIMÃO et al., 2025)

No que se refere à latência, conforme ilustrado na Figura 7, observa-se que o Cenário 1 apresenta os maiores valores, com mediana em torno de 23 ms, um intervalo interquartil amplo e diversos valores extremos que se estendem além de 45 ms. Esse comportamento reflete a ausência de qualquer mecanismo de controle, resultando em elevada variabilidade e desempenho inconsistente. O Cenário 2 apresenta uma redução significativa da latência, com a mediana situada aproximadamente entre 17 e 18 ms, além de um IQR mais estreito e menor quantidade de

outliers quando comparado ao cenário sem controle. O Cenário 3 apresenta o melhor desempenho, alcançando a menor mediana de latência, em torno de 16 a 17 ms, e um intervalo interquartil bastante compacto, indicando maior previsibilidade e estabilidade no comportamento da rede.

De forma análoga, a Figura 8 evidencia o comportamento do jitter nos três cenários. O Cenário 1 apresenta os maiores valores, com a mediana situada entre 6 e 7 ms, ampla dispersão e diversos valores extremos que ultrapassam 20 ms, caracterizando uma elevada instabilidade na entrega dos pacotes. No Cenário 2, observa-se uma redução expressiva do jitter, com a mediana reduzida para aproximadamente 3 a 4 ms, além de uma distribuição mais concentrada e um menor número de outliers. O Cenário 3 novamente se destaca, apresentando a menor variabilidade entre os cenários, com a mediana inferior a 3 ms, intervalo interquartil bastante reduzido e comportamento significativamente mais estável.

De modo geral, os resultados dos boxplots confirmam que a introdução de mecanismos de controle melhora substancialmente o desempenho da rede em termos de latência e jitter. Além disso, a estratégia preditiva baseada em Inteligência Artificial apresenta ganhos adicionais em relação ao controle puramente reativo, proporcionando maior estabilidade e previsibilidade, características fundamentais para aplicações industriais sensíveis ao atraso.

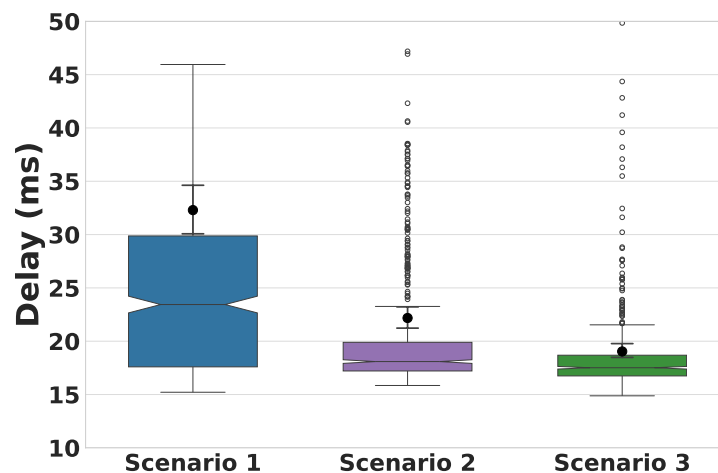


Figura 7 – Boxplot da latência para os três cenários experimentais. (SIMÃO et al., 2025)

5.3 Avaliação do Modelo de Predição

A avaliação do modelo de predição empregado no Cenário 3 evidencia sua adequação para operação em tempo real. O modelo SNARIMAX demonstrou capacidade de prever com antecedência o comportamento da ocupação das filas da rede de transporte, permitindo que o controlador atuasse antes da ocorrência efetiva do congestionamento.

A predição correta dos períodos de alta ocupação foi essencial para o desempenho superior observado no Cenário 3, pois possibilitou ajustes proativos nos parâmetros das filas, reduzindo a necessidade de ações corretivas tardias.

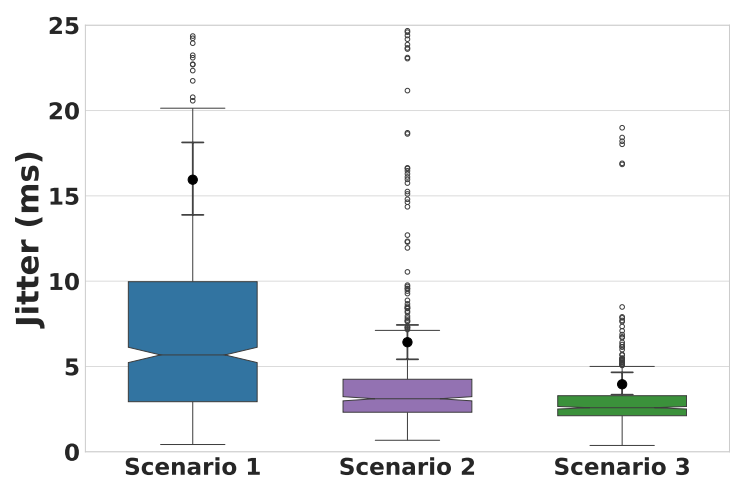


Figura 8 – Boxplot do jitter para os três cenários experimentais. (SIMÃO et al., 2025)

5.4 Ganhos de Desempenho

A Tabela 2 sintetiza os ganhos obtidos nos três cenários avaliados, considerando os valores médios de latência e jitter.

Tabela 2 – Análise dos dados dos resultados experimentais.

Cenário	Latência Média	Jitter Médio
1	32.29 ms	15.94 ms
2	22.17 ms (-31.33%)	6.42 ms (-59.72%)
3	19.04 ms (-41.04%)	3.95 ms (-75.17%)

Os resultados mostram que o Cenário 2 alcançou reduções de aproximadamente 31,33% na latência média e 59,72% no jitter em relação ao cenário sem controle, evidenciando que a simples introdução do controle reativo e da priorização dinâmica já traz benefícios significativos.

Entretanto, o Cenário 3, que incorpora predição baseada em Inteligência Artificial, apresentou os melhores resultados, com reduções de até 41,04% na latência média e 75,17% no jitter. Esses ganhos demonstram que a antecipação de eventos de congestionamento é um fator determinante para a estabilidade e a qualidade do serviço em aplicações industriais sensíveis ao tempo.

De forma geral, os resultados confirmam que a combinação de telemetria de plano de dados, redes programáveis e modelos preditivos constituiu uma abordagem eficaz para garantir a QoS em redes 5G privadas voltadas à Indústria 4.0, superando significativamente soluções sem controle ou puramente reativas.

6 CONSIDERAÇÕES FINAIS

Este trabalho apresentou o desenvolvimento, a implementação e a avaliação de um mecanismo de controle de redes em malha fechada voltado para aplicações sensíveis à latência em ambientes industriais baseados em computação de borda e redes 5G privadas. Motivado pelos requisitos rigorosos da Indústria 4.0, o mecanismo proposto integrou telemetria em tempo real do plano de dados programável com um modelo de predição baseado em Inteligência Artificial, permitindo a detecção proativa de congestionamentos e a priorização dinâmica de fluxos críticos.

A validação experimental, realizada por meio de uma prova de conceito com switches P4, núcleo 5G open source e uma aplicação real de vídeo industrial, demonstrou que a abordagem proposta é capaz de melhorar significativamente o desempenho da rede. Os resultados evidenciaram reduções expressivas na latência média e no jitter quando comparados ao cenário sem controle, confirmando que a combinação entre programabilidade de rede, telemetria orientada ao plano de dados e modelos preditivos é uma solução eficaz para garantir requisitos de Qualidade de Serviço em aplicações críticas.

Além disso, a comparação entre os cenários avaliados mostrou que, embora o controle reativo já apresente ganhos relevantes, a incorporação de inteligência preditiva permite antecipar condições adversas da rede, resultando em maior estabilidade e desempenho superior. Dessa forma, o trabalho contribui ao demonstrar, de forma prática, como técnicas modernas de redes programáveis e aprendizado de máquina podem ser aplicadas de maneira integrada em ambientes industriais reais.

6.1 Trabalhos Futuros

Como trabalhos futuros, pretende-se expandir o mecanismo proposto para cenários de maior escala, envolvendo topologias com múltiplos switches programáveis e um número maior de fluxos heterogêneos, a fim de avaliar aspectos de escalabilidade e robustez. Também é de interesse investigar a incorporação de novas métricas de rede ao modelo de predição, como perda de pacotes e vazão, visando aumentar a acurácia das previsões e aprimorar a tomada de decisão do controlador.

Outra direção relevante consiste na avaliação do mecanismo em ambientes físicos de redes 5G privadas, substituindo parcialmente o testbed virtualizado, de modo a analisar o impacto de variabilidades reais do meio sem fio. Por fim, a exploração de modelos de aprendizado mais avançados ou híbridos, bem como a adaptação do controle para múltiplas classes de serviço simultâneas, representa um passo importante para tornar a solução ainda mais flexível e aplicável a diferentes cenários da Indústria 4.0.

REFERÊNCIAS BIBLIOGRÁFICAS

- 5G-ACIA. *Key 5G Use Cases and Requirements*. [S.l.], 2020. Disponível em: <<https://5g-acia.org/whitepapers/key-5g-use-cases-and-requirements/>>. Citado na página 15.
- AIJAZ, A. Private 5G: The future of industrial wireless. *IEEE Industrial Electronics Magazine*, IEEE, v. 14, n. 4, p. 136–145, 2020. Citado na página 15.
- ARSLAN, S.; MCKEOWN, N. Switches Know the Exact Amount of Congestion. In: *Proceedings of the 2019 Workshop on Buffer Sizing*. New York, NY, USA: Association for Computing Machinery, 2019. (BS '19). ISBN 9781450377454. Disponível em: <<https://doi.org/10.1145/3375235.3375245>>. Citado na página 17.
- BASARAS, P. et al. Experimentally Assessing Deployment Tradeoffs for AI-enabled Video Analytics Services in the 5G Compute Continuum. In: *2023 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*. [S.l.: s.n.], 2023. p. 99–104. Citado na página 15.
- CABRAL, R. et al. Explorando o overhead de telemetria em redes programáveis na estimativa de qos com aprendizado de máquina. In: *Anais do XLIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. Porto Alegre, RS, Brasil: SBC, 2025. p. 518–531. ISSN 2177-9384. Disponível em: <<https://sol.sbc.org.br/index.php/sbrc/article/view/35155>>. Citado 3 vezes nas páginas 6, 17 e 18.
- CHAVES, R. et al. Enabling Parallel Processing at the Edge for Real-Time Video Analysis Applications. In: *Anais do XVI Workshop de Pesquisa Experimental da Internet do Futuro*. Porto Alegre, RS, Brasil: SBC, 2025. p. 58–65. ISSN 2595-2692. Disponível em: <<https://sol.sbc.org.br/index.php/wpeif/article/view/35278>>. Citado 2 vezes nas páginas 6 e 16.
- DAHLMAN, E.; PARKVALL, S.; SKÖLD, J. *5G NR The Next Generation Wireless Access Technology*. 2nd. ed. [S.l.]: Academic Press, 2020. 79–106 p. Citado na página 15.
- JOHN, J. et al. Industry 4.0 and Beyond: The Role of 5G, WiFi 7, and Time-Sensitive Networking (TSN) in Enabling Smart Manufacturing. *Future Internet*, v. 16, n. 9, 2024. ISSN 1999-5903. Disponível em: <<https://www.mdpi.com/1999-5903/16/9/345>>. Citado na página 12.
- MIRZAEINIA, A.; MIRZAEINIA, M.; REZGUI, A. Latency and Throughput Optimization in Modern Networks: A Comprehensive Survey. *CoRR*, abs/2009.03715, 2020. Disponível em: <<https://arxiv.org/abs/2009.03715>>. Citado na página 12.
- MUZAFFAR, R. et al. 5G Deployment Models and Configuration Choices for Industrial Cyber-Physical Systems – A State of Art Overview. *IEEE Trans. on Industrial Cyber-Physical Systems*, v. 1, p. 236–256, 2023. Citado na página 15.
- NAKIMULI, W. et al. Deployment and Evaluation of an Industry 4.0 Use Case over 5G. *IEEE Communications Magazine*, v. 59, n. 7, p. 14–20, 2021. Citado na página 15.
- NETO, O. R. et al. Middleware para aplicações distribuídas de vídeo com suporte à computação na borda na indústria 4.0. In: *Proceedings of the 30th Brazilian Symposium on Multimedia and*

the Web. Porto Alegre, RS, Brasil: SBC, 2024. p. 215–222. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/webmedia/article/view/30315>>. Citado na página 16.

OCHONU, R.; VIDAL, J. *Slicing for Dense Smart Factory Network: Current State, Scenarios, Challenges and Expectations*. 2024. Disponível em: <<https://arxiv.org/abs/2405.03230>>. Citado na página 12.

P4. *In-band Network Telemetry (INT) Dataplane Specification*. [S.l.], 2021. Citado na página 17.

RAY, P. P.; KUMAR, N. SDN/NFV architectures for edge-cloud oriented IoT: A systematic review. *Computer Communications*, Elsevier, v. 169, p. 129–153, 2021. Citado na página 17.

SIMÃO, V. S. et al. Closed-loop network control for industrial edge computing: A telemetry-driven and ai-based approach for latency-critical applications in private 5g. *IEEE Networking Letters*, IEEE, 2025. Citado 5 vezes nas páginas 6, 22, 29, 30 e 31.

VARGA, P. et al. 5G support for Industrial IoT Applications— Challenges, Solutions, and Research gaps. *Sensors*, v. 20, n. 3, 2020. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/20/3/828>>. Citado na página 15.

ZOU, X. et al. Robust Edge AI for Real-Time Industry 4.0 Applications in 5G Environment. *IEEE Communications Standards Magazine*, v. 7, n. 2, p. 64–70, 2023. Citado na página 15.

	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA
	Campus João Pessoa - Código INEP: 25096850
	Av. Primeiro de Maio, 720, Jaguaribe, CEP 58015-435, João Pessoa (PB)
	CNPJ: 10.783.898/0002-56 - Telefone: (83) 3612.1200

Documento Digitalizado Ostensivo (Público)

Trabalho de Conclusão de Curso - Corrigido (Já com Ficha Catalográfica e Folha de Aprovação)

Assunto:	Trabalho de Conclusão de Curso - Corrigido (Já com Ficha Catalográfica e Folha de Aprovação)
Assinado por:	Vinicius Simao
Tipo do Documento:	Anexo
Situação:	Finalizado
Nível de Acesso:	Ostensivo (Público)
Tipo do Conferência:	Cópia Simples

Documento assinado eletronicamente por:

- Vinicius Silva Simao, DISCENTE (20211610006) DE BACHARELADO EM ENGENHARIA ELÉTRICA - JOÃO PESSOA, em 14/02/2026 17:26:59.

Este documento foi armazenado no SUAP em 14/02/2026. Para comprovar sua integridade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifpb.edu.br/verificar-documento-externo/> e forneça os dados abaixo:

Código Verificador: 1768499
Código de Autenticação: d85b9d8ec4

