

ANÁLISE DE QUALIDADE DE SERVIÇO EM REDES POR MEIO DE
DETECÇÃO DE PONTOS DE MUDANÇA E CLUSTERIZAÇÃO BASEADA
EM SOBREVIVÊNCIA

Ian José Agra Gomes

Dissertação de Mestrado apresentada ao
Programa de Pós-graduação em Engenharia
de Sistemas e Computação, COPPE, da
Universidade Federal do Rio de Janeiro, como
parte dos requisitos necessários à obtenção do
título de Mestre em Engenharia de Sistemas e
Computação.

Orientador: Edmundo Albuquerque de Souza
e Silva

Rio de Janeiro
Dezembro de 2025

ANÁLISE DE QUALIDADE DE SERVIÇO EM REDES POR MEIO DE
DETECÇÃO DE PONTOS DE MUDANÇA E CLUSTERIZAÇÃO BASEADA
EM SOBREVIVÊNCIA

Ian José Agra Gomes

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO
GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E
COMPUTAÇÃO.

Orientador: Edmundo Albuquerque de Souza e Silva

Aprovada por: Prof. Edmundo Albuquerque de Souza e Silva
Prof. Rosa Maria Meri Leão
Prof. Carlos Alberto Vieira Campos

RIO DE JANEIRO, RJ – BRASIL
DEZEMBRO DE 2025

Agra Gomes, Ian José

Análise de Qualidade de Serviço em Redes por Meio de Detecção de Pontos de Mudança e Clusterização Baseada em Sobrevivência/Ian José Agra Gomes. – Rio de Janeiro: UFRJ/COPPE, 2025.

XIII, 69 p.: il.; 29, 7cm.

Orientador: Edmundo Albuquerque de Souza e Silva

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2025.

Referências Bibliográficas: p. 63 – 69.

1. Redes de computadores. 2. Qualidade de Serviço (QoS). 3. Pontos de mudança. 4. Análise de sobrevivência. 5. Grandes Modelos de Linguagem (LLMs). I. Albuquerque de Souza e Silva, Edmundo. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

“Duc in altum”.

Agradecimentos

Agradeço a Deus por todas as graças imerecidas. A São José, pela intercessão que, atendendo a meu pedido, possibilitou o ingresso no Mestrado e a chegada a esta etapa final. Seu exemplo foi essencial nos momentos de dificuldade.

À minha esposa Ana Beatriz, minha mais profunda gratidão. Seu apoio incondicional, compreensão nas longas noites de estudo e incentivo constante foram fundamentais para que eu conseguisse conduzir os estudos no curso. Sem sua presença ao meu lado, este sonho não teria se concretizado.

Aos meus filhos Benício José, Maria Helena e Catarina, agradeço pelo amor incondicional que me deu forças e motivação para perseverar. Vocês foram o combustível que me impulsionou nos momentos mais desafiadores desta jornada acadêmica.

À minha família de origem: minha mãe Bárbara, pelo dom da vida; meu saudoso pai Lucrécio, que continua a me inspirar; e minha avó Eunice, cuja criação me proporcionou as oportunidades educacionais que me trouxeram até aqui.

À Marinha do Brasil, agradeço pela confiança para representá-la e por permitir que eu leve os conhecimentos obtidos à briosíssima instituição.

Ao meu orientador, professor Edmundo de Souza e Silva, e à professora Rosa Maria Meri Leão, professora do grupo de pesquisa do LAND, minha eterna gratidão pela orientação e ensinamentos atenciosos e pacientes ao longo de todo o curso. Suas contribuições acadêmicas e humanas transcenderam o papel de educadores, tornando-se verdadeiros mentores.

Gratidão especial aos alunos e ex-alunos do grupo de pesquisa do LAND, cujos conhecimentos e esforços pavimentaram o caminho que permitiu a realização deste estudo.

À Rede Nacional de Ensino e Pesquisa (RNP), agradeço pela disponibilização dos servidores para a realização dos testes de medição de rede, no âmbito de seu acordo de cooperação com o M-Lab. Esse suporte foi essencial à infraestrutura de experimentação deste trabalho.

Por fim, a todos que, direta ou indiretamente, contribuíram para esta realização acadêmica, seja com conselhos, palavras de incentivo ou gestos de apoio, minha eterna e sincera gratidão.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ANÁLISE DE QUALIDADE DE SERVIÇO EM REDES POR MEIO DE
DETECÇÃO DE PONTOS DE MUDANÇA E CLUSTERIZAÇÃO BASEADA
EM SOBREVIVÊNCIA

Ian José Agra Gomes

Dezembro/2025

Orientador: Edmundo Albuquerque de Souza e Silva

Programa: Engenharia de Sistemas e Computação

Este trabalho apresenta uma metodologia integrada para análise da qualidade de serviço (QoS) em redes de computadores que combina detecção de pontos de mudança em séries temporais e clusterização baseada em análise de sobrevivência. Desenvolvida como consequência da colaboração entre a Rede Nacional de Ensino e Pesquisa (RNP) e o Measurements Lab (M-Lab), a metodologia foi validada utilizando dados reais coletados durante três meses via protocolo *Network Diagnostic Tool* (NDT) em um provedor de internet brasileiro. A abordagem implementa duas estratégias analíticas complementares: uma baseada em alterações na vazão de download e outra em mudanças no tempo de resposta (RTT) de upload. Os resultados evidenciaram que intervalos com mudanças estatísticas frequentes apresentam métricas de desempenho significativamente inferiores. A análise dos coeficientes de um modelo de regressão logística, acoplado ao algoritmo de clusterização, permitiu identificar e quantificar sistematicamente os clientes e servidores problemáticos e as métricas que exercem maior impacto na instabilidade da rede. Por fim, a interpretação desses coeficientes é realizada por Grandes Modelos de Linguagem (LLMs), transformando a análise estatística em uma ferramenta de diagnóstico acessível a operadores de rede.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

QUALITY OF SERVICE ANALYSIS IN COMPUTER NETWORKS THROUGH CHANGEPOINT DETECTION AND SURVIVAL-BASED CLUSTERING

Ian José Agra Gomes

December/2025

Advisor: Edmundo Albuquerque de Souza e Silva

Department: Systems Engineering and Computer Science

This work presents an integrated methodology for Quality of Service (QoS) analysis in computer networks that combines changepoint detection in time series and survival-based clustering. Developed as a consequence of the collaboration between the Brazilian National Education and Research Network (RNP) and the Measurement Lab (M-Lab), the methodology was validated using real data collected over three months via the *Network Diagnostic Tool* (NDT) protocol from a Brazilian internet service provider. The approach implements two complementary analytical strategies: one based on changes in download throughput and another on variations in upload round-trip time (RTT). The results showed that intervals with frequent statistical changes exhibit significantly inferior performance metrics. The analysis of the coefficients from a logistic regression model, coupled with the clustering algorithm, allowed for the systematic identification and quantification of problematic clients and servers and the metrics that have the greatest impact on network instability. Finally, the interpretation of these coefficients is carried out by Large Language Models (LLMs), transforming the statistical analysis into a diagnostic tool accessible to network operators.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xii
Lista de Abreviaturas	xiii
1 Introdução	1
1.1 Contextualização	2
1.2 Problema de pesquisa	2
1.3 Hipótese	3
1.4 Objetivos	3
1.5 Contribuições	4
1.6 Organização	5
2 Fundamentação teórica e trabalhos relacionados	6
2.1 QoS em redes de computadores	6
2.1.1 Abordagens para análise de QoS em provedores de internet . .	7
2.1.2 Contribuição em relação a trabalhos existentes	8
2.2 Detecção de pontos de mudança	9
2.2.1 Métodos clássicos	9
2.2.2 Desafios em dados de redes de computadores	10
2.3 Análise de sobrevivência	11
2.3.1 Da medicina à engenharia de redes	11
2.3.2 Aprendizado de máquina para análise de sobrevivência	11
2.4 Síntese e posicionamento da pesquisa	12
3 Metodologia de pesquisa	14
3.1 Classificação da pesquisa	14
3.2 Procedimentos de coleta de dados	15
3.2.1 Infraestrutura de medição	15
3.2.2 Protocolo NDT	15
3.2.3 Servidores de teste	16

3.2.4	Distribuição dos medidores	16
3.2.5	Diferencial da abordagem de coleta	16
3.3	Caracterização dos dados coletados	17
3.3.1	Volume e distribuição dos dados	17
3.3.2	Estatísticas descritivas	17
3.4	Pré-processamento dos dados	18
3.4.1	Construção das séries temporais por par cliente-servidor . . .	18
3.5	Estruturação dos conjuntos de intervalos	19
3.5.1	Criação dos intervalos	19
3.5.2	Definição de censura	21
3.5.3	Características dos conjuntos de intervalos	21
3.6	Limitações do conjunto de dados	22
4	A abordagem proposta: detecção de mudanças e clusterização	23
4.1	Visão Geral da Abordagem	23
4.2	Detecção de pontos de mudança	25
4.2.1	Seleção das métricas de referência	25
4.2.2	Fundamentos do Algoritmo VWCD	26
4.2.3	Configuração do algoritmo VWCD	28
4.3	Preparação dos dados para clusterização	29
4.4	Clusterização baseada em análise de sobrevivência	30
4.4.1	Parametrização do SurvMixClust	32
4.5	Análise de fatores de risco via regressão logística	33
4.6	Interpretação dos coeficientes por modelos de linguagem	34
5	Resultados	36
5.1	Detecção de pontos de mudança	36
5.2	Análise de sobrevivência e clusterização	37
5.2.1	Caracterização dos <i>clusters</i>	37
5.2.2	Análise dos resultados nas séries temporais	40
5.3	Análise de fatores de risco e identificação de componentes críticos da rede	42
5.4	Interpretação automatizada dos coeficientes com LLM	45
5.4.1	Validação das respostas às questões de pesquisa	52
6	Conclusão	58
6.1	Contribuições principais	58
6.2	Validação das hipóteses da pesquisa	59
6.3	Limitações da metodologia	60
6.4	Trabalhos futuros	61

6.5	Considerações finais	62
	Referências Bibliográficas	63

Lista de Figuras

3.1	Ocorrência no tempo de testes NDT para cada cliente.	15
3.2	Matriz de correlação de Pearson das métricas de desempenho.	18
3.3	Exemplo fictício de construção das <i>features</i> dos intervalos para um par cliente-servidor. (a) Série temporal da métrica de referência (vazão de download, ou “THR-DL”) e pontos de mudança que definem os intervalos I_1 , I_2 e I_3 . (b) Séries temporais das quatro métricas de desempenho (vazão de download, RTT de download, vazão de upload e RTT de upload, respectivamente “THR-DL”, “RTT-DL”, “THR-UL” e “RTT-UL”) recortadas pelos mesmos pontos de mudança. (c) Esquematização das <i>features</i> formadas pelas médias e desvios padrão de cada métrica em cada intervalo.	20
4.1	Fluxo de coleta e processamento dos dados.	24
5.1	Exemplo de aplicação do algoritmo VWCD na série temporal de RTT de upload para o par <code>cliente16-gru03</code>	37
5.2	Funções de sobrevivência dos <i>clusters</i> definidos pelo algoritmo <i>SurvMixClust</i> , considerando as mudanças no <i>throughput</i> de download.	38
5.3	Funções de sobrevivência dos <i>clusters</i> definidos pelo algoritmo <i>SurvMixClust</i> , considerando as mudanças no RTT de upload.	38
5.4	Gráficos tipo caixa das <i>features</i> analisadas para cada <i>cluster</i> , considerando as mudanças no <i>throughput</i> de download.	41
5.5	Gráficos tipo caixa das <i>features</i> analisadas para cada <i>cluster</i> , considerando as mudanças no RTT de upload.	42
5.6	Séries Temporais das Métricas de Rede para pares de cliente servidor, considerando as mudanças no <i>throughput</i> de download.	43
5.7	Séries Temporais das Métricas de Rede para pares de cliente servidor, considerando as mudanças no RTT de upload.	44

Lista de Tabelas

2.1	Comparação de estudos recentes sobre QoS em redes.	7
3.1	Estatísticas descritivas das principais métricas coletadas (vazão em Mbps e RTT em ms).	17
3.2	Características dos conjuntos de dados de sobrevivência	22
5.1	Características estatísticas dos <i>clusters</i> identificados.	55
5.2	Coefficientes da regressão logística e <i>odds ratios</i> do modelo <i>SurvMix-Clust</i> calculadas nas duas abordagens.	56
5.3	Estatísticas agregadas das métricas de desempenho para os servidores <i>gru06</i> e <i>rnp_rj</i>	57

Lista de Abreviaturas

CUSUM	<i>Cumulative Sum</i>	9
EM	<i>Expectation-Maximization</i>	31
GLR	<i>Generalized Likelihood Ratio</i>	9
GPT	<i>Generative Pre-trained Transformer</i>	35
ISP	<i>Internet Service Provider</i>	4
JSON	<i>JavaScript Object Notation</i>	34
LLM	<i>Large Language Model</i>	2
MLE	<i>Maximum Likelihood Estimation</i>	26
NDT	<i>Network Diagnostic Tool</i>	1
NLP	<i>Natural Language Processing</i>	2
PoPs	<i>Points of Presence</i>	2
QoE	<i>Quality of Experience</i>	12
QoS	<i>Quality of Service</i>	1
RNP	Rede Nacional de Ensino e Pesquisa	2
SVM	<i>Support Vector Machine</i>	7
VWCD	<i>Voting Windows Change-point Detection</i>	1

Capítulo 1

Introdução

A Qualidade de Serviço (*Quality of Service* - QoS) em redes de computadores é um campo de estudo consolidado, essencial para garantir o desempenho e a confiabilidade de infraestruturas de comunicação. Tradicionalmente, a análise de QoS se apoia em modelos de confiabilidade e disponibilidade para otimizar a alocação de recursos e avaliar o desempenho de sistemas computacionais (DE SOUZA E SILVA e GAIL, 1989, 1986, 1990).

A análise de sobrevivência, campo relacionado, busca modelar o tempo até a ocorrência de um evento de interesse, com aplicações que vão desde estudos médicos a análises de falhas de equipamentos (DE SOUZA E SILVA *et al.*, 2022). Na área de redes de computadores, esses modelos têm sido progressivamente empregados para caracterizar a QoS. Aplicações incluem a previsão do impacto de interrupções no fluxo de vídeo na retenção de usuários, a estimativa de probabilidades de falha do sistema e a análise da eficácia de mecanismos de correção de erros (STREIT *et al.*, 2021b).

A detecção de pontos de mudança (*changepoint detection*) continua uma área de pesquisa ativa, com aplicação direta no monitoramento de QoS (DE ALMEIDA *et al.*, 2024). A identificação de mudanças estatísticas em métricas de desempenho de rede como latência e vazão (*throughput*) é dificultada por fatores como (i) a não-estacionariedade intrínseca das séries temporais devido à dinâmica de roteamento com mudanças de rota e balanceamento de carga (PUCHA *et al.*, 2007); (ii) a ausência de dados rotulados; e (iii) a rápida perda de relevância de rótulos históricos em ambientes de rede dinâmicos.

Nesse contexto, STREIT *et al.* (2021a) propôs o método VWCD (*Voting Windows Change-point Detection*), que DE ALMEIDA *et al.* (2025); DE ALMEIDA (2024) mostrou ser eficaz na identificação de alterações estatísticas em séries temporais de métricas de rede coletadas com o protocolo NDT (*Network Diagnostic Tool*) (M-LAB, 2025). O trabalho indicou também que a frequência de tais mudanças pode servir como um indicador da qualidade da experiência do usuário.

Paralelamente, avanços recentes em Processamento de Linguagem Natural (NLP), impulsionados pelos Grandes Modelos de Linguagem (LLMs), criaram novas possibilidades para a interpretação de dados técnicos (FANG *et al.*, 2024). Sua aplicação na interpretação de dados de rede tem o potencial de automatizar e simplificar análises complexas, tornando-as acessíveis a operadores sem formação especializada em estatística.

Este trabalho propõe e valida uma metodologia que combina detecção de pontos de mudança, análise de sobrevivência e interpretação assistida por modelos de linguagem para análise de QoS. A metodologia parte de trabalhos anteriores em detecção de pontos de mudança e análise de sobrevivência (BUGINGA e DE SOUZA E SILVA, 2024; BUGINGA, 2023; DE ALMEIDA *et al.*, 2024; STREIT *et al.*, 2021a). Os pontos de mudança são utilizados como delimitadores para segmentar as séries temporais em intervalos estatisticamente homogêneos. Em seguida, aplica-se análise de sobrevivência a esses intervalos para agrupá-los em *clusters* com curvas de sobrevivência distintas, considerando múltiplas métricas de rede conjuntamente. Cada curva de sobrevivência caracteriza a distribuição temporal dos intervalos entre mudanças estatísticas. Para complementar a análise, um LLM é utilizado para auxiliar na interpretação dos resultados, permitindo a extração de informações relevantes e a geração de relatórios em linguagem natural. Essa abordagem visa mitigar as barreiras técnicas na interpretação dos resultados, capacitando equipes operacionais a aplicar as análises de forma prática e eficaz.

1.1 Contextualização

O presente trabalho insere-se no contexto de um acordo de colaboração entre o *Measurement Lab* (M-Lab) e a Rede Nacional de Ensino e Pesquisa (RNP) (RNP, 2023). Essa parceria viabilizou a instalação de medidores de velocidade na infraestrutura da RNP e o compartilhamento dos dados resultantes para fins de pesquisa. Os dados provenientes de testes de velocidade realizados de forma sistemática entre clientes e os Pontos de Presença (PoPs) da RNP refletem a experiência real dos usuários e são fundamentais para a análise de QoS em um cenário operacional.

1.2 Problema de pesquisa

A análise de dados de QoS em larga escala, como os provenientes de medições NDT contínuas entre múltiplos clientes e servidores, apresenta desafios de complexidade e interpretabilidade. A extração de informações relevantes e de simples interpretação a partir desse volume massivo de dados requer uma metodologia sistemática e

automatizada. Nesse contexto, este trabalho busca responder às seguintes questões de pesquisa:

1. Como identificar, de forma não supervisionada, os clientes cujas métricas de desempenho apresentam padrões indicativos de qualidade insatisfatória?
2. Quais servidores possivelmente apresentam métricas de comunicação inferiores em suas interações com os clientes?
3. Quais métricas de rede exercem maior impacto no desempenho global da rede?
4. Como transformar análises estatísticas complexas em informações relevantes e compreensíveis para operadores de rede?

Responder a essas questões de forma eficiente e escalável, especialmente diante de grandes volumes de dados heterogêneos, é o cerne do problema abordado neste trabalho.

1.3 Hipótese

A hipótese central deste trabalho é que a instabilidade temporal das métricas de desempenho, quantificada pela frequência de mudanças estatísticas, é um indicador robusto da QoS. Especificamente, as hipóteses que norteiam esta pesquisa são:

1. A frequência das mudanças estatísticas em séries temporais de métricas de rede possuem uma forte relação com a percepção de degradação de serviço.
2. Os intervalos entre essas mudanças estatísticas podem ser agrupados em *clusters* distintos que representam diferentes níveis de qualidade de serviço, identificáveis por meio de análise de sobrevivência.
3. A aplicação de Grandes Modelos de Linguagem (LLMs) permite a tradução de resultados estatísticos complexos em diagnósticos e recomendações técnicas precisas e acionáveis, superando a necessidade de interpretação por especialistas.

1.4 Objetivos

O objetivo principal deste trabalho é desenvolver e validar uma metodologia robusta para a análise de qualidade de serviço em redes de computadores, integrando técnicas de detecção de pontos de mudança, análise de sobrevivência e interpretação assistida por modelos de linguagem.

Os objetivos específicos que norteiam esta pesquisa incluem:

1. Implementar e validar um algoritmo de detecção de pontos de mudança em séries temporais de métricas de rede.
2. Aplicar e avaliar um método de clusterização baseado em análise de sobrevivência para agrupar intervalos entre mudanças estatísticas, considerando múltiplas métricas simultaneamente.
3. Interpretar os resultados de forma assistida por LLM, traduzindo análises estatísticas em informações práticas para operadores de rede.
4. Validar a metodologia integrada utilizando dados reais coletados em parceria com um Provedor de Serviços de Internet (ISP) brasileiro, medindo a precisão na identificação de problemas e a utilidade das interpretações geradas.
5. Realizar uma análise comparativa de abordagens baseadas em diferentes métricas de referência (vazão de download e RTT de upload) para determinar suas complementaridades e particularidades na avaliação da QoS.

1.5 Contribuições

As principais contribuições desta pesquisa são:

1. **Metodologia inovadora:** o desenvolvimento de uma abordagem que integra, de forma inédita, a detecção de pontos de mudança (VWCD) com a análise de sobrevivência para clusterização (SurvMixClust) para caracterizar a estabilidade de conexões de rede.
2. **Análise quantitativa de fatores de risco:** a aplicação de uma regressão logística multinomial pelo algoritmo SurvMixClust para identificar e quantificar a influência de diferentes clientes, servidores e métricas de desempenho de rede na estabilidade da conexão, permitindo a criação de um *ranking* de criticidade para orientar ações de manutenção proativa.
3. **Validação em cenário real:** a aplicação e validação da metodologia em um conjunto de dados exclusivo, coletado na infraestrutura de um provedor de internet brasileiro em parceria com a RNP e o M-Lab.
4. **Análise comparativa de métricas:** a avaliação sistemática das diferenças e complementaridades entre abordagens baseadas em métricas de referência distintas (vazão de download vs. RTT de upload), fornecendo diretrizes sobre como utilizar cada abordagem para uma análise de QoS mais completa.

5. **Interpretação assistida por LLM:** a proposição e validação de uma estratégia que utiliza Grandes Modelos de Linguagem para traduzir os resultados da modelagem estatística em diagnósticos operacionais em linguagem natural, simplificando análises técnicas complexas e tornando-as acessíveis a operadores de rede.

1.6 Organização

Esta dissertação está estruturada da seguinte forma:

O Capítulo 2 apresenta a fundamentação teórica e os trabalhos relacionados, abordando os conceitos de detecção de pontos de mudança, análise de sobrevivência e uma revisão da literatura sobre qualidade de serviço em redes.

O Capítulo 3 detalha a metodologia de pesquisa adotada e descreve o conjunto de dados utilizado, incluindo o processo de coleta, os procedimentos de pré-processamento, a análise exploratória, as características estatísticas dos dados e a estruturação dos experimentos realizados.

O Capítulo 4 apresenta a abordagem proposta, pormenorizando os algoritmos de detecção de pontos de mudança e análise de sobrevivência e a interpretação assistida por modelos de linguagem.

O Capítulo 5 apresenta os resultados experimentais, a análise dos *clusters* identificados, e a interpretação dos resultados gerada por LLM.

O Capítulo 6 apresenta as conclusões do trabalho, com a discussão dos resultados obtidos, as limitações do estudo e as direções para pesquisas futuras.

Capítulo 2

Fundamentação teórica e trabalhos relacionados

Este capítulo estabelece a base teórica da pesquisa, realizando uma revisão da literatura sobre suas áreas centrais: avaliação da QoS em redes de computadores, detecção de pontos de mudança, análise de sobrevivência e aplicação de modelos de linguagem generativos na interpretação de dados. O objetivo é posicionar a pesquisa no estado da arte, mostrando como a metodologia proposta se diferencia dos trabalhos existentes.

2.1 QoS em redes de computadores

A literatura sobre avaliação de QoS em redes de computadores contempla desde metodologias voltadas à coleta e análise de indicadores de desempenho até estudos que investigam o impacto direto dessas métricas na experiência dos usuários finais. Trabalhos como os de DA HORA *et al.* (2018a,b); OLIVEIRA *et al.* (2024); SANTOS *et al.* (2022); SAVERIMOUTOU *et al.* (2019) propõem técnicas para avaliação ou predição de QoS em diferentes cenários.

Este trabalho, no entanto, diferencia-se das abordagens tradicionais, que frequentemente se concentram na estimativa direta da qualidade percebida por aplicações específicas. Apresentamos uma metodologia baseada em técnicas não supervisionadas de análise de séries temporais, cujo objetivo é identificar elementos da rede (clientes e servidores) que potencialmente exibem métricas de desempenho inadequadas ou fora dos parâmetros esperados.

2.1.1 Abordagens para análise de QoS em provedores de internet

Os estudos recentes sobre avaliação de QoS em ISPs concentram-se em três áreas principais: (1) análise de gargalos e componentes que limitam o desempenho; (2) detecção de anomalias e congestionamento; e (3) monitoramento e predição de desempenho. A Tabela 2.1 resume e contrasta trabalhos representativos com a abordagem aqui proposta.

Tabela 2.1: Comparação de estudos recentes sobre QoS em redes.

Estudo	Foco	Metodologia	Como nossa proposta se diferencia do referido trabalho
SHARMA <i>et al.</i> (2024a)	Gargalos em redes residenciais	Análise de vazão Wi-Fi vs. ISP	Análise macroscópica da estabilidade temporal da rede, em vez de focar em um único gargalo (Wi-Fi).
PI <i>et al.</i> (2024)	Desequilíbrios de congestionamento	Detecção de mudanças + SVM	Análise do desempenho agregado de clientes/servidores, em vez de rotas de rede específicas.
SHARMA <i>et al.</i> (2024b)	Clusterização espacial de desempenho	Agrupamento por padrões de latência	Foco na instabilidade temporal de séries individuais, complementar à análise de variabilidade espacial.
PAUL <i>et al.</i> (2022)	Contextualização de medições de velocidade	Inferência de planos de banda larga	Utiliza a análise de sobrevivência para agrupar dinamicamente por perfis de comportamento, sem depender da inferência de planos.
SUNDBERG <i>et al.</i> (2024)	Monitoramento passivo de latência	Instrumentação kernel com eBPF	Foco na análise de dados já coletados (NDT) para identificar padrões de instabilidade, sem instrumentação adicional.
Nossa proposta	Padrões de instabilidade temporal	Detecção de mudanças + Análise de sobrevivência + LLMs	Metodologia integrada que caracteriza perfis de estabilidade e traduz achados em <i>insights</i> acionáveis.

O trabalho de SHARMA *et al.* (2024a) realiza uma análise microscópica de gargalos em redes domésticas, concluindo que a rede Wi-Fi é frequentemente o principal fator limitante. Em contraste, nossa metodologia opera em um nível macroscópico, analisando dados de múltiplos clientes e servidores para identificar padrões temporais de instabilidade na rede do provedor como um todo, em vez de se concentrar em um único ponto de falha.

A ferramenta CongI, de PI *et al.* (2024), utiliza detecção de mudanças e classificadores SVM (*Support Vector Machine*) para identificar congestionamento em rotas de rede específicas com balanceamento de carga. Embora também se baseie em sé-

ries temporais de latência, o CongI está restrito à avaliação de rotas predefinidas. Nossa abordagem é mais ampla, pois avalia o desempenho agregado de clientes e servidores em suas interações com toda a rede, sem se limitar a caminhos específicos.

SHARMA *et al.* (2024b) propuseram uma clusterização temporal de latência aplicada a pequenas regiões geográficas, considerando o 95° percentil da latência como indicador de estabilidade. Embora compartilhem o objetivo de identificar problemas estruturais na rede, nossa abordagem foca na instabilidade observada em séries temporais individuais. As duas propostas são complementares, oferecendo visões espacial e temporal sobre a variabilidade da QoS.

A importância de contextualizar as medições de QoS foi destacada por PAUL *et al.* (2022), que desenvolveram um método para inferir os planos de assinatura de banda larga dos usuários a partir de dados coletivos de testes de velocidade. Essas informações permitiram interpretações mais precisas dos resultados de desempenho e mostraram que fatores como tecnologia de acesso e capacidades dos dispositivos afetam significativamente a taxa de transferência medida. A análise, baseada em dados do M-Lab, ressalta a importância de categorizar usuários conforme suas expectativas de serviço, conceito comparável à abordagem de agrupamento baseada em análise de sobrevivência proposta neste trabalho.

SUNDBERG *et al.* (2024) desenvolveram um sistema de monitoramento passivo de latência para ISPs sem fio. Implantado em uma rede operacional, o sistema mediu o tempo de ida e volta por sub-rede e revelou variações significativas de latência no segmento de última milha, principalmente durante horários de pico. Por outro lado, este estudo mostra o valor do monitoramento contínuo para identificar segmentos de baixo desempenho na infraestrutura do ISP.

Com foco em monitoramento adaptativo, LINDSTÅHL *et al.* (2024) propuseram um sistema para detecção adaptativa de pontos de mudança no desempenho da rede. O sistema programa medições dinamicamente com base nos padrões observados, permitindo detecção mais eficiente da degradação da QoS. Essa abordagem alinha-se conceitualmente com o nosso uso de detecção de pontos de mudança em séries temporais, embora nosso foco seja na análise retrospectiva para identificar padrões recorrentes, e não no agendamento adaptativo de medições.

2.1.2 Contribuição em relação a trabalhos existentes

Até onde sabemos, nenhuma proposta anterior combina sistematicamente a detecção de pontos de mudança em séries temporais com análise de sobrevivência, aplicadas a métricas de rede como latência e vazão, para identificar, de forma não supervisionada, padrões recorrentes de degradação de desempenho em larga escala. Adicionalmente, nossa abordagem se diferencia significativamente pela incorporação

de modelos de linguagem generativos para traduzir resultados técnicos complexos em informações práticas, facilitando sua interpretação por operadores de rede não especializados em análise de dados.

2.2 Detecção de pontos de mudança

A detecção de pontos de mudança é uma área ativa de pesquisa, com aplicações em diversos campos além de redes de computadores, como finanças, medicina, processamento de sinais e análise de mídias sociais. Esta seção apresenta uma visão geral das principais abordagens, contextualizando a escolha do algoritmo VWCD para a metodologia proposta.

2.2.1 Métodos clássicos

As abordagens tradicionais podem ser classificadas em *offline* (que processam toda a série temporal de uma vez) e *online* (que detectam mudanças sequencialmente à medida que novos dados surgem). Além disso, os métodos podem ser paramétricos ou não paramétricos, dependendo das suposições sobre a distribuição dos dados.

Entre os métodos paramétricos, destacam-se:

- **CUSUM** (*Cumulative Sum*): Proposto por PAGE (1954), monitora a soma cumulativa dos desvios em relação a um valor de referência e detecta uma mudança quando essa soma excede um limiar predefinido.
- **GLR** (*Generalized Likelihood Ratio*): Usa a razão de verossimilhança entre diferentes hipóteses sobre a distribuição dos dados para detectar mudanças, como descrito por LAI (2018).
- **Métodos Bayesianos**: Incorporam conhecimento prévio sobre a distribuição dos pontos de mudança, como no trabalho de FEARNEHEAD e LIU (2011), que propõe um método eficiente para detecção online assumindo dependência de Markov entre segmentos.

Entre os métodos não-paramétricos, que são mais flexíveis por não assumirem uma distribuição específica para os dados, pode-se citar:

- **Métodos baseados em *rank***: Usam a ordem relativa dos dados em vez de seus valores absolutos, como o teste de Wilcoxon-Mann-Whitney para detectar mudanças na mediana.
- **Métodos baseados em *kernel***: Transformam os dados para um espaço de características de alta dimensão, onde as mudanças podem ser detectadas mais facilmente, como proposto por HARCHAOU *et al.* (2008).

- **Métodos baseados em densidade:** Detectam mudanças na densidade de probabilidade dos dados, como o algoritmo KL-CPD proposto por KAWAHARA e SUGIYAMA (2012).
- **Abordagens baseadas em Aprendizado de Máquina:** Usam técnicas como redes neurais, *random forest* e *Support Vector Machine* (SVM) para aprender padrões de mudança em dados históricos, como no trabalho de LIU *et al.* (2013).

2.2.2 Desafios em dados de redes de computadores

A aplicação de métodos tradicionais de detecção de pontos de mudança em dados de redes apresenta desafios específicos, como destacado por DE ALMEIDA *et al.* (2024) e LINDSTÄHL *et al.* (2024):

- **Não-estacionariedade:** Séries temporais de métricas de rede geralmente apresentam comportamento não-estacionário, com tendências sazonais e cíclicas que podem gerar falsos positivos em métodos que assumem estacionariedade.
- **Heterogeneidade dos dados:** Diferentes métricas (vazão, latência, perda de pacotes) podem ter distribuições e características distintas, exigindo métodos flexíveis ou específicos para cada tipo de dado.
- **Ruídos e anomalias:** Medições de rede frequentemente contêm ruídos e anomalias transitórias que não representam mudanças estatísticas persistentes, mas podem confundir algoritmos sensíveis.
- **Necessidade de detecção em tempo real:** Em muitos cenários operacionais, é necessário detectar mudanças em tempo real para resposta imediata, impondo restrições de eficiência computacional.

Esses desafios motivaram o desenvolvimento de abordagens especializadas para análise de dados de rede, como o algoritmo VWCD (STREIT *et al.*, 2021a), adotado neste trabalho. A escolha pelo VWCD se deve, além das suas características técnicas, ao fato de ser um algoritmo desenvolvido e aprimorado pelo nosso grupo de pesquisa, tendo sido testado em trabalhos anteriores especificamente com dados de métricas de rede obtidos através do protocolo NDT (DE ALMEIDA *et al.*, 2024). O VWCD destaca-se por sua capacidade de lidar com séries não estacionárias, sua robustez a anomalias pontuais e, principalmente, por possibilitar a interpretação da confiabilidade associada a cada ponto de mudança detectado através de seu mecanismo de votação e probabilidades posteriores. O detalhamento técnico deste algoritmo será apresentado no Capítulo 4.

2.3 Análise de sobrevivência

Originalmente desenvolvida para estudos médicos, a análise de sobrevivência tem encontrado aplicações crescentes em diversos campos, incluindo engenharia para prever falhas de equipamentos (YANG *et al.*, 2022) e, mais recentemente, em análise de dados de rede (BUGINGA, 2023).

2.3.1 Da medicina à engenharia de redes

A análise de sobrevivência tradicional, representada por modelos como Kaplan-Meier (KAPLAN e MEIER, 1958) e Cox Proportional Hazards (COX, 2018), evoluiu significativamente com a incorporação de técnicas de aprendizado de máquina. WANG *et al.* (2019) apresentam uma revisão dessa evolução, destacando como modelos não paramétricos e técnicas de aprendizado profundo ampliaram a capacidade de análise em cenários complexos.

A adaptação dessas técnicas para análise de redes de computadores é uma tendência recente e promissora. Em vez de modelar o tempo até o óbito de pacientes, esses métodos podem ser aplicados para modelar o tempo até falhas em equipamentos de rede, interrupções de serviço ou, como na nossa abordagem, mudanças estatísticas significativas em métricas de desempenho.

2.3.2 Aprendizado de máquina para análise de sobrevivência

O campo de análise de sobrevivência tem se beneficiado da integração com técnicas de aprendizado de máquina nos últimos anos. Entre as principais contribuições nesta área, destacam-se:

- ***Random Survival Forests*** (ISHWARAN *et al.*, 2008): Extensão das florestas aleatórias (*random forests*) para análise de sobrevivência, permitindo modelagem não linear de efeitos e interações sem suposições paramétricas.
- **DeepSurv** (KATZMAN *et al.*, 2018): Implementação baseada em redes neurais do modelo de Cox, capaz de aprender representações complexas e não lineares para previsão de sobrevivência.
- **DeepHit** (LEE *et al.*, 2018): Modelo de aprendizado profundo para previsão de risco de múltiplos eventos, permitindo modelagem de riscos competitivos.
- **CoxPH Neural Network** (KVAMME e ØRNULF BORGAN, 2021): Integração de redes neurais com o modelo de Cox, incorporando dependência temporal nos coeficientes do modelo.

Mais recentemente, abordagens que combinam clusterização com análise de sobrevivência têm surgido, permitindo a identificação de grupos com perfis de sobrevivência distintos. O algoritmo SurvMixClust (BUGINGA e DE SOUZA E SILVA, 2024), adotado na metodologia proposta neste trabalho, representa um avanço nessa direção, integrando o agrupamento com a previsão da função de sobrevivência por meio de uma mistura de especialistas não paramétricos. Seus aspectos técnicos serão detalhados no Capítulo 4.

2.4 Síntese e posicionamento da pesquisa

Este trabalho consolida e expande uma linha de investigação contínua desenvolvida pelo nosso grupo de pesquisa, posicionando-se na interseção de duas áreas principais: análise de QoS em redes de computadores e técnicas de aprendizado não supervisionado para séries temporais. A metodologia proposta é o resultado evolutivo de trabalhos que pavimentaram o desenvolvimento de ferramentas para monitoramento e diagnóstico de redes.

O trabalho de XIMENES *et al.* (2018) introduziu um *framework* para detectar e localizar eventos na rede de um ISP utilizando a correlação espaço-temporal de séries temporais residenciais. Posteriormente, STREIT *et al.* (2021a) aplicaram decomposição de tensores para identificar anomalias de tráfego durante a pandemia, enquanto SANTOS *et al.* (2022) utilizaram Modelos Ocultos de Markov (HMM) combinados com dados de reclamações de usuários para inferir a Qualidade de Experiência (QoE).

As bases algorítmicas específicas deste trabalho fundamentam-se em contribuições mais recentes do grupo. BUGINGA (2023) desenvolveu o SurvMixClust, originalmente aplicado a dados clínicos de pacientes com doença arterial coronariana. O autor uniu análise de sobrevivência e clusterização para caracterizar perfis de risco e, ao final, mostrou a generalidade do método realizando um teste com dados reais de métricas de rede. DE ALMEIDA (2024) realizou um estudo comparativo abrangente de diversos métodos de detecção de pontos de mudança aplicados a séries temporais de métricas de desempenho de rede coletadas via protocolo NDT, validando o método VWCD, proposto anteriormente por STREIT *et al.* (2021a), como uma alternativa robusta e flexível baseada em votação ponderada.

Este trabalho se posiciona como a integração dessas duas frentes, diferenciando-se de abordagens existentes ao explorar a premissa de que a frequência de mudanças estatísticas nas métricas de desempenho da rede é um indicador direto da degradação do serviço. A metodologia proposta permite a identificação de padrões de instabilidade e a caracterização de perfis de desempenho com base nos pontos de mudança detectados nas séries temporais das métricas de desempenho. Adicional-

mente, utiliza-se um modelo de linguagem (LLM) para interpretação assistida dos resultados, traduzindo análises estatísticas complexas em informações simples e relevantes para operadores de rede.

Capítulo 3

Metodologia de pesquisa

Este capítulo detalha a abordagem metodológica adotada para a condução desta pesquisa. Descrevemos aqui o desenho do estudo, a natureza da pesquisa, os procedimentos de coleta de dados e a estruturação dos experimentos realizados para validar as hipóteses levantadas.

3.1 Classificação da pesquisa

A presente pesquisa classifica-se como **aplicada**, de natureza **quantitativa** e **experimental**. É aplicada pois busca resolver um problema concreto de gerenciamento de redes (a identificação de instabilidade). É quantitativa pois se baseia na coleta e análise estatística de métricas numéricas de desempenho (vazão e latência). Por fim, é experimental, pois envolve a intervenção no ambiente (instalação de sondas) e a observação sistemática dos fenômenos de rede sob condições controladas e reais.

O método experimental foi dividido em três fases:

1. **Coleta de dados:** monitoramento ativo da infraestrutura de um Provedor de Internet (ISP) parceiro.
2. **Pré-processamento e modelagem:** tratamento das séries temporais e aplicação dos algoritmos propostos.
3. **Validação e interpretação:** análise estatística dos *clusters* formados e validação qualitativa via LLMs.

Cabe ressaltar que este capítulo detalha os procedimentos da **Fase 1** e a etapa de pré-processamento da **Fase 2**. A modelagem algorítmica é apresentada no Capítulo 4, enquanto a **Fase 3** (Validação e Interpretação) é discutida nos Capítulos 5 e 6.

3.2 Procedimentos de coleta de dados

3.2.1 Infraestrutura de medição

Neste trabalho, usamos dados coletados por 16 dispositivos Raspberry Pi 4 modelo B, programados para executar testes NDT a cada 30 minutos durante três meses, de maio a julho de 2024. Os dispositivos foram distribuídos em duas categorias:

- **Residências de voluntários:** 4 dispositivos instalados em casas com diferentes planos de internet e tecnologias de acesso; e
- **Infraestrutura do ISP:** 12 dispositivos em pontos estratégicos da rede do provedor.

Essa distribuição permitiu capturar tanto a experiência do usuário final quanto o desempenho em pontos intermediários da rede.

A Figura 3.1 mostra a distribuição temporal das medições por cliente, onde os pontos na cor verde representam medições em residências e os pontos de cor laranja indicam medições na infraestrutura do ISP, perfazendo um total de 84 dias de observação.

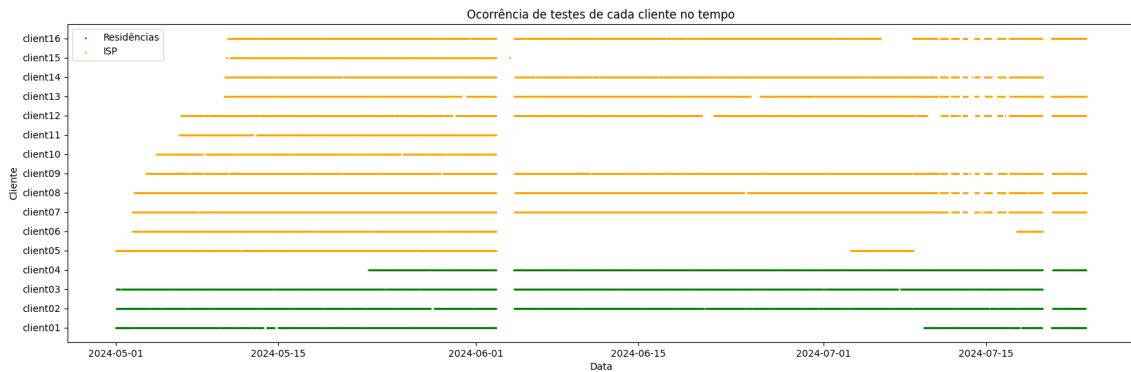


Figura 3.1: Ocorrência no tempo de testes NDT para cada cliente.

3.2.2 Protocolo NDT

O NDT, parte do conjunto de ferramentas do M-Lab (CLARK e WEDEMAN, 2021), é um protocolo aberto para avaliação de desempenho de redes. Ele fornece várias métricas, como:

- **Vazão de download e upload:** Taxa de transferência, em Mbps;
- **RTT (*Round-Trip Time*):** Tempo de ida e volta dos pacotes, em ms;
- **Perda de pacotes:** Percentual de pacotes perdidos;

- **Jitter**: Variação no tempo de resposta, em ms; e
- **Parâmetros TCP**: Tamanho da janela de congestionamento, número de re-conexões, etc.

Para cada teste, o protocolo estabelece uma conexão usando um fluxo TCP, realizando transmissões bidirecionais que simulam o comportamento real de aplicações. Os clientes NDT usam seleção automática de servidor baseada na proximidade geográfica e na disponibilidade.

3.2.3 Servidores de teste

Os testes NDT foram direcionados a medidores de velocidade da infraestrutura do M-Lab hospedados em Pontos de Presença (PoPs) da RNP, resultado de um acordo de colaboração técnica entre as instituições (RNP, 2023).

Foram utilizados cinco servidores: `gru03`, `gru05`, `gru06`, `rnp_sp` (localizados em São Paulo-SP) e `rnp_rj` (localizado no Rio de Janeiro-RJ).

A localização dos servidores em pontos estratégicos de troca de tráfego no Brasil é um fator relevante para a caracterização do desempenho da rede.

3.2.4 Distribuição dos medidores

Os 16 dispositivos de medição foram instalados em diferentes localidades na área de cobertura do ISP, com o objetivo de capturar uma diversidade de condições de rede.

A coleta de dados ocorreu de forma contínua durante três meses (maio a julho de 2024), abrangendo diferentes padrões de uso da rede, com variações diárias (horários de pico e de baixa demanda), semanais (dias úteis e fins de semana ou feriados) e mensais.

3.2.5 Diferencial da abordagem de coleta

A abordagem de medição deste trabalho difere da simples utilização de dados de testes NDT iniciados por usuários. Enquanto esses testes geralmente ocorrem quando os usuários percebem problemas e decidem avaliar sua conexão, a coleta realizada buscou capturar um panorama abrangente das condições da rede.

Ao realizar medições sistemáticas e independentes da percepção do usuário, reduz-se o viés introduzido pela tendência dos clientes de fazer testes apenas em momentos de degradação do serviço (GILL *et al.*, 2022). Isso permite uma análise mais imparcial da qualidade da conexão ao longo do tempo.

3.3 Caracterização dos dados coletados

3.3.1 Volume e distribuição dos dados

O conjunto de dados bruto apresenta as seguintes características:

- **Total de medições:** 33.313 testes NDT completos;
- **Média de medições por cliente:** 2.082 testes;
- **Média de medições por servidor:** 6.662,6 testes;
- **Média de testes por par cliente-servidor:** 475,9 testes;
- **Período total de coleta:** 83 dias (2.208 horas);
- **Intervalo teórico entre medições:** 30 minutos (48 medições diárias); e
- **Cobertura temporal média:** 52,3% do período total.

A cobertura temporal média de 52,3% indica a ocorrência de falhas na execução periódica dos testes. Tais falhas podem ser atribuídas a fatores como instabilidade na alimentação elétrica dos dispositivos de medição e perda de conectividade intermitente, que são condições inerentes a um ambiente de coleta de dados distribuído e não controlado. Não obstante essa esparsidade nos dados brutos, a metodologia foi projetada para ser robusta a tais interrupções, conforme detalhado no Capítulo 4.

Além disso, a distribuição das medições entre os diferentes pares cliente-servidor não é uniforme devido à seleção automática de servidores.

3.3.2 Estatísticas descritivas

A Tabela 3.1 apresenta as principais estatísticas das métricas coletadas, calculadas para todos os pares cliente-servidor, ou seja, considerando todas as medições.

Tabela 3.1: Estatísticas descritivas das principais métricas coletadas (vazão em Mbps e RTT em ms).

Métrica	Média	Mediana	Desvio Padrão	Mínimo	Máximo
Vazão download	624,43	662,67	226,64	0,08	934,68
Vazão upload	584,15	718,24	286,45	0,00	919,07
RTT download	29,01	18,31	39,25	2,28	446,25
RTT upload	34,70	17,45	46,82	2,08	450,06

A Figura 3.2 apresenta a matriz de correlação de Pearson entre as métricas. A correlação observada entre a vazão e o RTT correspondente em cada direção (download e upload) era esperada, uma vez que, no teste NDT, a vazão e o RTT são obtidos

de forma simultânea e refletem o comportamento da mesma rota de rede sob carga. Como ambas as métricas respondem de maneira conjunta às mesmas condições, é natural que apresentem maior valor de correlação. Esses valores fornecem subsídios iniciais para a interpretação dos resultados dos algoritmos usados.

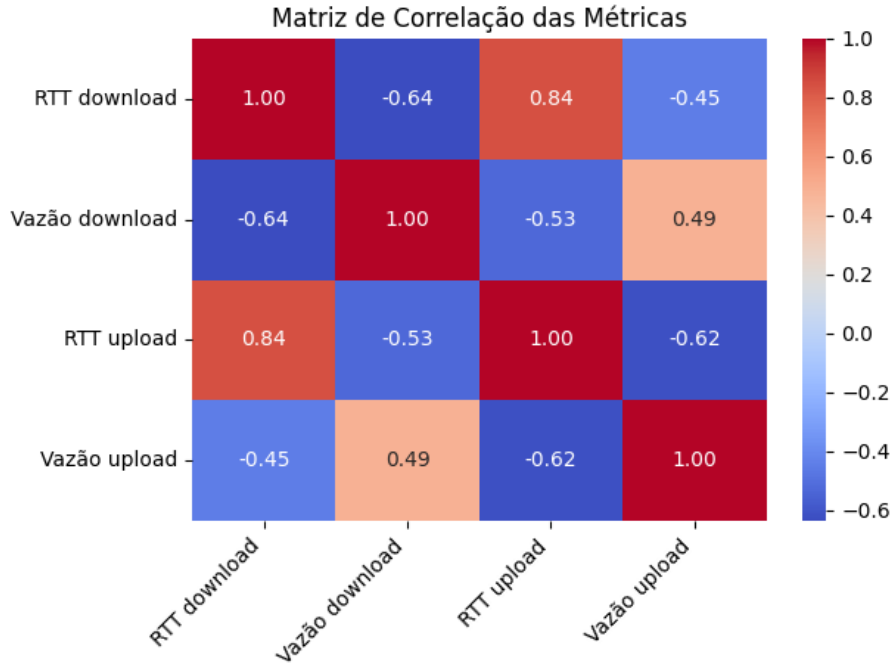


Figura 3.2: Matriz de correlação de Pearson das métricas de desempenho.

3.4 Pré-processamento dos dados

Os dados brutos passaram por um processo de pré-processamento para garantir sua qualidade e adequação às análises.

3.4.1 Construção das séries temporais por par cliente-servidor

A partir dos dados brutos, foram construídas séries temporais de cada métrica, agregadas por par cliente-servidor. Para assegurar uma amostra mínima para a análise subsequente de pontos de mudança, foram mantidos apenas os pares que continham ao menos 100 medições completas ao longo do período de observação. Este limiar foi estabelecido empiricamente como um balanço entre reter o maior número de séries possível e garantir a estabilidade das estimativas estatísticas do algoritmo utilizado para detecção de pontos de mudança.

Com o intuito de avaliar a robustez da metodologia em condições reais de operação, optou-se por não realizar nenhum tratamento para a remoção de *outliers* ou

anomalias das séries temporais. Dessa forma, o algoritmo de detecção de pontos de mudança foi exposto às condições não ideais encontradas nos dados brutos. As características das séries temporais resultantes estão sumarizadas a seguir:

- **Número de séries:** 70 pares cliente-servidor $\times$ 4 métricas = 280 séries temporais;
- **Comprimento médio das séries:** 118,97 pontos; e
- **Intervalo médio entre pontos:** 209,54 minutos (acima do intervalo nominal de 30 minutos, indicando interrupções ou falhas na execução periódica dos testes).

3.5 Estruturação dos conjuntos de intervalos

Conforme descrito no Capítulo 4, foi aplicado o método VWCD para rotular as séries temporais com os pontos de mudança detectados. Esses instantes no tempo foram utilizados para segmentar as séries e construir conjuntos de intervalos (em que cada intervalo constitui uma entrada), que serão usados como entrada para o algoritmo SurvMixClust.

3.5.1 Criação dos intervalos

Os pontos de mudança identificados nas séries temporais de vazão de download e de RTT de upload foram considerados como eventos de interesse para construir dois conjuntos distintos de intervalos:

1. **Dataset baseado em vazão de download:** Usa como eventos os pontos de mudança detectados nas séries de vazão de download.
2. **Dataset baseado em RTT de upload:** Usa como eventos os pontos de mudança detectados nas séries de RTT de upload.

Cada entrada corresponde a um intervalo temporal, delimitado por eventos (pontos de mudança) consecutivos ou por interrupções nas medições, conforme será detalhado na Subseção 3.5.2.

A Figura 3.3 ilustra, através de dados sintéticos, o procedimento de construção dos conjuntos de intervalos a partir de uma métrica de referência e das demais métricas coletadas para um par cliente-servidor.

No painel (a) da Figura 3.3, a série temporal da métrica de referência (vazão de download, ou “THR-DL”) é segmentada pelos pontos de mudança estatística detectados nos instantes 20 e 60. Esses instantes definem três intervalos I_1 , I_2 e I_3 ,

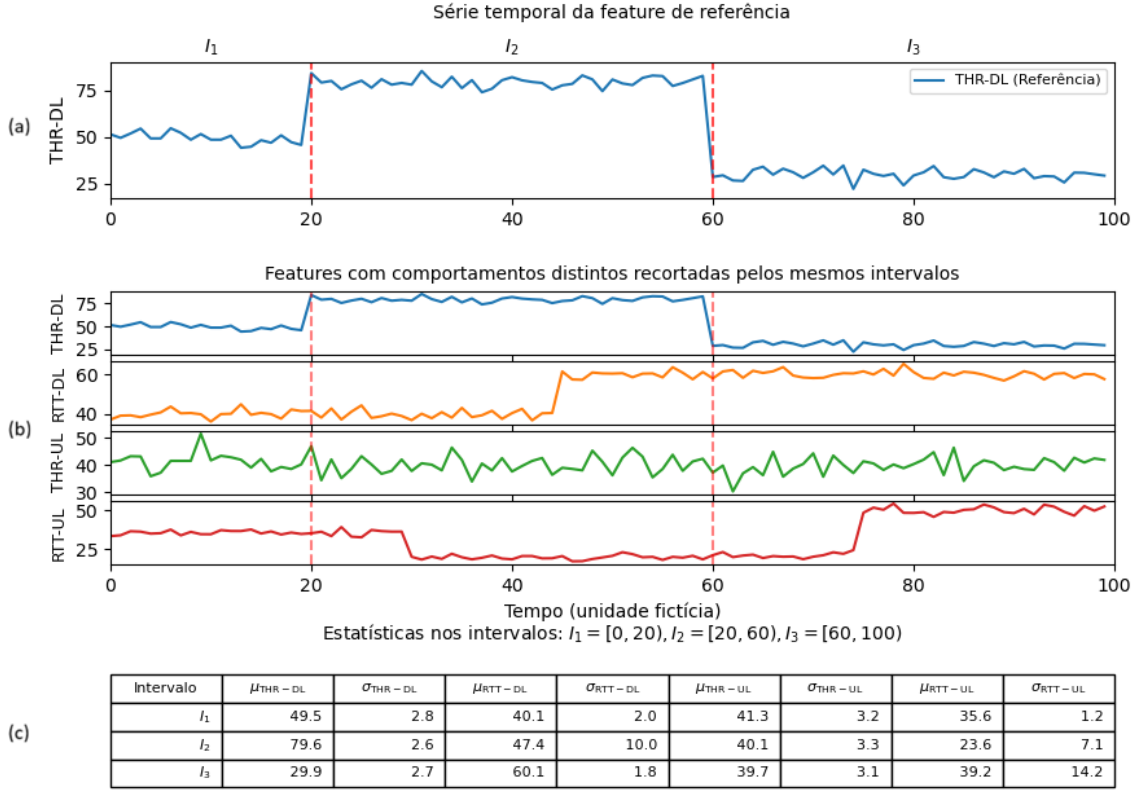


Figura 3.3: Exemplo fictício de construção das *features* dos intervalos para um par cliente-servidor. (a) Série temporal da métrica de referência (vazão de download, ou “THR-DL”) e pontos de mudança que definem os intervalos I_1 , I_2 e I_3 . (b) Séries temporais das quatro métricas de desempenho (vazão de download, RTT de download, vazão de upload e RTT de upload, respectivamente “THR-DL”, “RTT-DL”, “THR-UL” e “RTT-UL”) recortadas pelos mesmos pontos de mudança. (c) Esquematização das *features* formadas pelas médias e desvios padrão de cada métrica em cada intervalo.

cada um consistindo em uma entrada do conjunto de intervalos que servirá de base para a análise de sobrevivência.

Em seguida, como mostrado no painel (b), os mesmos limites temporais são aplicados às quatro métricas de desempenho de rede (vazão de download, RTT de download, vazão de upload e RTT de upload, respectivamente “THR-DL”, “RTT-DL”, “THR-UL” e “RTT-UL”). Embora apenas a vazão de download tenha sido utilizada diretamente na detecção dos pontos de mudança, os mesmos instantes são usados para segmentar, de forma consistente, as séries temporais das quatro métricas.

Para cada intervalo I_k , calculam-se então as estatísticas (média e desvio padrão) *locais*, isto é, para cada intervalo I_k e para cada uma das quatro métricas. O painel (c) da Figura 3.3 esquematiza o resultado desse processo: cada intervalo possui oito *features* de estatísticas locais (médias e desvios padrão de cada uma das quatro métricas). Dessa forma, ainda que os intervalos sejam definidos a partir de uma

única métrica de referência, cada intervalo incorpora informação agregada sobre o comportamento conjunto de vazão e latência em download e upload, o que permite que a etapa de clusterização considere a dinâmica de desempenho da rede.

Além das estatísticas locais das métricas, cada intervalo possui outras *features*, como a sua duração, um indicador de censura e identificadores de cliente e servidor, conforme detalhado na Subseção 3.5.3.

3.5.2 Definição de censura

Em análise de sobrevivência, a censura refere-se a uma situação em que o tempo exato de ocorrência de um evento de interesse não é conhecido precisamente, mas sabe-se que ocorreu dentro de um intervalo específico de tempo (ver referências em BUGINGA (2023)). Neste trabalho, o conjunto de intervalos é composto por intervalos com e sem censura, conforme se segue:

- **Intervalos não censurados:** Delimitados por dois pontos de mudança consecutivos.
- **Intervalos com censura à direita:** Iniciados por um ponto de mudança e terminados pelo fim da série temporal ou por uma interrupção das medições.
- **Intervalos com censura à esquerda:** Começados pelo início da série ou após uma interrupção e terminados em um ponto de mudança.
- **Intervalos duplamente censurados:** Iniciados após uma interrupção ou pelo início da série e terminados por outra interrupção ou pelo final da série, sem pontos de mudança.

Cada intervalo possui um indicador binário de censura: 1 para intervalos onde o evento ocorreu (não censurados) e 0 para intervalos censurados.

Para a definição de interrupção, foi estabelecido um limiar de três dias sem medições. Este valor foi uma escolha metodológica conservadora, por ser significativamente superior ao intervalo médio observado entre medições válidas e, portanto, menos provável de ocorrer por variações aleatórias no processo de medição.

3.5.3 Características dos conjuntos de intervalos

A Tabela 3.2 apresenta as características dos conjuntos de dados de intervalos resultantes das mudanças estatísticas detectadas em cada métrica de referência.

Cada intervalo nos conjuntos é caracterizado por:

Tabela 3.2: Características dos conjuntos de dados de sobrevivência

Característica	Vazão de download	RTT de upload
Número total de intervalos	257	195
Intervalos não censurados	88 (34,2%)	44 (22,6%)
Intervalos com censura à direita	101 (39,3%)	91 (46,7%)
Intervalos com censura à esquerda	51 (19,8%)	45 (23,1%)
Intervalos duplamente censurados	17 (6,6%)	15 (7,7%)
Duração média dos intervalos (dias)	18,8	24,8

1. **Variáveis categóricas:** Variáveis binárias representando cada cliente e cada servidor, derivadas da aplicação de *one-hot encoding* sobre os identificadores originais.
2. **Variáveis temporais:** Duração do intervalo em dias e *timestamps* de início e fim.
3. **Variáveis estatísticas:** Médias e desvios padrão locais das métricas analisadas.
4. **Indicador de evento:** Variável binária indicando a ocorrência (1) ou não (0) do evento.

3.6 Limitações do conjunto de dados

Os conjuntos de dados gerados a partir da coleta sistemática de medições NDT fornecem uma base empírica para a metodologia proposta. A distribuição dos medidores e a duração da coleta permitiram capturar uma variedade de condições de rede. A estruturação em dois conjuntos de dados distintos, um focado em mudanças na vazão de download e outro em RTT de upload, permite uma análise complementar da QoS, explorando tanto a capacidade de transferência quanto a responsividade da rede.

Entretanto, é necessário reconhecer as limitações inerentes a esse conjunto de dados. A cobertura temporal média de 52,3% e o intervalo médio entre medições superior ao nominal indicam a presença de esparsidade nos dados, uma condição realista em medições de campo, mas que impõe desafios à análise. Adicionalmente, a metodologia proposta define limiares para filtragem das séries (mínimo de 100 medições) e para a definição de interrupções (3 dias), cujos valores, embora justificados, podem influenciar os resultados.

Capítulo 4

A abordagem proposta: detecção de mudanças e clusterização

Conforme apresentado no Capítulo 1, a metodologia deste trabalho se aplica à análise de conjuntos massivos de séries temporais. A premissa central é que mudanças estatísticas frequentes afetam negativamente a qualidade dos serviços de internet oferecidos aos usuários, resultando em piores métricas de desempenho de rede.

As séries temporais utilizadas foram coletadas na rede de um ISP parceiro. As medições foram obtidas pelo protocolo NDT por meio de dispositivos Raspberry Pi distribuídos em diversos pontos da rede do provedor, conforme detalhado no Capítulo 3.

4.1 Visão Geral da Abordagem

A metodologia deste trabalho foi projetada para realizar uma análise da QoS a partir de grandes conjuntos de séries temporais de métricas de rede. A premissa central é que a instabilidade temporal das conexões, manifestada por mudanças estatísticas frequentes, está relacionada a um desempenho percebido inferior.

Para capturar a natureza multidimensional da QoS, a metodologia é executada em duas abordagens analíticas paralelas e complementares:

1. **Abordagem 1 (foco em vazão):** utiliza os pontos de mudança detectados nas séries temporais de **vazão de download** como eventos de interesse. Essa abordagem visa caracterizar a estabilidade da capacidade de transferência de dados da rede, que impacta diretamente atividades de consumo de conteúdo.
2. **Abordagem 2 (foco em latência):** utiliza os pontos de mudança detectados nas séries temporais de **RTT de upload** como eventos de interesse. Essa abordagem foca na estabilidade da responsividade e interatividade da rede, crucial para aplicações em tempo real.

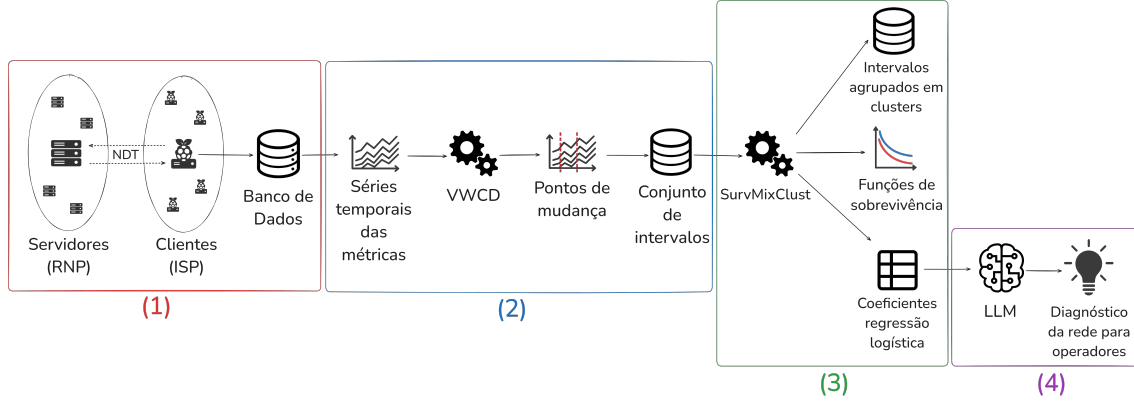


Figura 4.1: Fluxo de coleta e processamento dos dados.

Para ambas as abordagens, o fluxo de processamento, ilustrado na Figura 4.1, é idêntico e consiste nas seguintes etapas sequenciais:

- 1. Medição e coleta de dados:** cada medidor (cliente) instalado na rede do ISP dispara, a cada 30 minutos, um teste NDT, que é realizado a partir de um servidor da RNP e cujos resultados são armazenados em um banco de dados.
- 2. Detecção de pontos de mudança:** para cada par cliente-servidor, são extraídas do banco de dados as séries temporais de quatro métricas de desempenho de rede (vazão de download, RTT de download, vazão de upload e RTT de upload). As séries temporais da métrica de referência escolhida em cada abordagem são processadas para identificar os instantes de mudança estatística, que delimitarão os intervalos. Para cada abordagem, um conjunto de intervalos é criado, considerando como delimitadores os instantes de mudança detectados. É fundamental notar que, embora os intervalos sejam definidos pelos pontos de mudança das séries temporais de uma única métrica de referência, **cada intervalo possui *features* com as estatísticas locais (média e desvio padrão) das quatro métricas de rede monitoradas (vazão e RTT de download e upload)**. O processo de construção dos conjuntos de intervalos está detalhado na Subseção 3.5.1.
- 3. Clusterização baseada em análise de sobrevivência:** Cada conjunto de intervalos é então submetido ao algoritmo SurvMixClust para agrupar os intervalos em perfis de desempenho distintos com base em suas características de sobrevivência (duração e ocorrência do evento de interesse) e nas estatísticas locais das métricas de desempenho.
- 4. Interpretação assistida por LLM:** os coeficientes da regressão logística, calculados pelo SurvMixClust no processo de clusterização, são fornecidos a um LLM, que gera análises que não apenas descrevem os achados de cada

abordagem, mas também os comparam e sintetizam, identificando os clientes e servidores com problemas de desempenho e fornecendo uma visão integrada e acionável do estado da rede.

A fim de permitir a reprodução dos experimentos, disponibilizamos o código-fonte, os dados e os artefatos gerados em https://github.com/ianagra/masters_project.

As seções a seguir detalham os fundamentos e a implementação da metodologia proposta.

4.2 Detecção de pontos de mudança

No primeiro passo da metodologia, são identificados os momentos em que ocorreram mudanças estatísticas nas métricas de desempenho escolhidas. Para cada par cliente-servidor, foi usado o método VWCD, proposto por STREIT *et al.* (2021a), para identificar pontos de mudança nas séries temporais das métricas “vazão de download” e “RTT de upload”. Outras métricas poderiam ser consideradas, como, por exemplo, a taxa de perda. Além disso, o método de detecção permite analisar mudanças estatísticas em várias métricas simultaneamente, mas adotamos uma abordagem univariada para facilitar a interpretação dos resultados.

4.2.1 Seleção das métricas de referência

A escolha das métricas para definir os pontos de mudança é um passo metodológico fundamental. Foram selecionadas duas métricas ortogonais para constituir as duas abordagens de análise, com base em sua capacidade de representar aspectos complementares da QoS:

1. **Vazão de download (Abordagem 1):** representa a **capacidade** da rede de entregar dados ao usuário. Variações nessa métrica indicam instabilidade na entrega de conteúdo e afetam diretamente atividades como *streaming* de vídeo, downloads de arquivos e navegação web.
2. **RTT de upload (Abordagem 2):** reflete a **responsividade** e a latência da conexão do ponto de vista do envio de dados. Instabilidades nessa métrica são críticas para aplicações interativas como videoconferências, jogos online e serviços que dependem de comunicação bidirecional de baixa latência.

Ao analisar a rede sob estas duas perspectivas distintas, a metodologia busca fornecer uma visão mais completa, capaz de diagnosticar tanto problemas de capacidade quanto de latência.

4.2.2 Fundamentos do Algoritmo VWCD

O algoritmo *Voting Windows Change-point Detection* (VWCD) foi criado por STREIT *et al.* (2021a) para detectar mudanças estatísticas em séries temporais em tempo real, combinando eficiência computacional e simplicidade. O método usa janelas deslizantes que “votam” em potenciais pontos de mudança dentro de seu intervalo de observação.

O VWCD se baseia na ideia de que mudanças estatísticas importantes podem ser detectadas analisando segmentos consecutivos da série. O algoritmo opera sobre uma janela deslizante de tamanho fixo W , que examina os dados sequencialmente. Seja $\{x_t\}_{t=1}^T$ a série temporal observada, onde x_t representa a observação no instante t . A janela W_t no instante t contém as amostras $\{x_{t-W+1}, \dots, x_t\}$.

Para cada posição da janela, o algoritmo busca identificar se existe um ponto τ dentro do intervalo $[t-W+1, t]$ que representa uma mudança estatística significativa. Essa busca é feita através da análise da verossimilhança dos dados sob diferentes hipóteses de localização do ponto de mudança.

Mecanismo de Votação

O processo de votação é o núcleo do algoritmo e funciona da seguinte forma:

1. Para cada janela W_t , o algoritmo calcula a verossimilhança dos dados assumindo diferentes posições possíveis para o ponto de mudança τ ;
2. A posição τ^* que maximiza a verossimilhança dentro da janela recebe um voto; e
3. O peso do voto é determinado pela probabilidade posterior normalizada.

O algoritmo usa a distribuição Gaussiana como modelo para os dados, pois permite obter os parâmetros por máxima verossimilhança (MLE) de forma analítica e eficiente. Assim, a verossimilhança para uma posição candidata τ é calculada como:

$$L(\tau|W_t) = \max_{\theta_0, \theta_1} \prod_{i=t-W+1}^{\tau} f(x_i|\theta_0) \prod_{i=\tau+1}^t f(x_i|\theta_1) \quad (4.1)$$

onde θ_0 e θ_1 são os parâmetros estimados para os segmentos antes e depois de τ , respectivamente, e $f(\cdot|\theta)$ é a função densidade de probabilidade da distribuição normal assumida para os dados.

Uma função logística $y(x) = \frac{1}{1+e^{-(x-\mu)/s}}$ é usada para definir a probabilidade *a priori* sobre a posição do voto na janela. Com isso, são atribuídos pesos diferentes aos votos conforme sua posição dentro da janela deslizante, favorecendo votos com maior confiabilidade e evitando estimativas nos extremos.

Agregação dos Votos

O algoritmo oferece duas estratégias principais para agregar os votos e determinar os pontos de mudança:

1. **Agregação por limiar de confiança:** nessa abordagem, apenas os votos com probabilidade posterior acima de um limiar predefinido p_0 são considerados. Para cada instante t , a pontuação agregada é calculada como:

$$S_1(t) = \sum_{j=t-W+1}^t \mathbb{I}(P(\tau_j^* = t | W_j) > \alpha) \quad (4.2)$$

onde $\mathbb{I}(\cdot)$ é a função indicadora e $P(\tau_j^* = t | W_j)$ é a probabilidade posterior atribuída pela janela W_j ao instante t .

2. **Agregação ponderada:** esta estratégia considera todos os votos, ponderando-os por suas probabilidades posteriores:

$$S_2(t) = \sum_{j=t-W+1}^t P(\tau_j^* = t | W_j) \quad (4.3)$$

A probabilidade posterior $P(\tau | W_t)$ é calculada usando o teorema de Bayes:

$$P(\tau | W_t) = \frac{L(\tau | W_t) \pi(\tau)}{\sum_{\tau'} L(\tau' | W_t) \pi(\tau')} \quad (4.4)$$

onde $\pi(\tau)$ é a probabilidade *a priori*, que pode ser uniforme ou seguir uma distribuição beta-binomial que atribui menor peso a posições próximas às bordas da janela.

Crítérios de decisão

A decisão final sobre a presença de um ponto de mudança depende da estratégia de agregação escolhida:

- Para a agregação por limiar (S_1), um ponto de mudança é declarado quando o número de votos qualificados excede um limiar k ; e
- Para a agregação ponderada (S_2), além dos limiares p_0 e k , utiliza-se um limiar h sobre a soma ponderada das probabilidades posteriores.

Entretanto, é possível utilizar outras formas de agregação dos votos, que estão sendo estudadas por nosso grupo, por exemplo, em SIQUEIRA (2025).

Distinção entre ponto de mudança e anomalia

Uma característica importante do método VWCD é sua capacidade de distinguir mudanças persistentes de anomalias ao analisar o padrão temporal dos votos. As mudanças persistentes se caracterizam por votos consistentes (com alta probabilidade posterior) emitidos por várias janelas consecutivas. Por outro lado, as anomalias geralmente recebem votos com baixa probabilidade ou inconsistentes entre as janelas que as contêm, resultando em pontuações agregadas menores.

A escolha entre as estratégias de agregação pode ser feita com base nas características específicas da aplicação e no equilíbrio desejado entre sensibilidade e robustez. Geralmente, a agregação por limiar adota uma postura mais conservadora, enquanto a agregação por média ponderada permite detectar mudanças mais sutis, embora seja mais suscetível à interferência de ruídos.

Embora diversos métodos para detecção de pontos de mudança existam na literatura, o VWCD foi selecionado por duas razões principais. Primeiramente, ele foi especificamente projetado e validado para séries temporais de métricas de rede, demonstrando eficácia em cenários com as mesmas características dos dados utilizados neste trabalho, como não-estacionariedade e presença de ruído (DE ALMEIDA *et al.*, 2025). Em segundo lugar, o mecanismo de votação do VWCD permite interpretar a confiança associada a cada ponto de mudança detectado, o que confere maior robustez à análise subsequente. Ressalta-se, porém, que a metodologia proposta neste trabalho é **agnóstica ao método de detecção de pontos de mudança**, podendo incorporar outros algoritmos em trabalhos futuros.

4.2.3 Configuração do algoritmo VWCD

A eficácia do algoritmo VWCD está atrelada à sua configuração, que define a sensibilidade e a robustez da detecção. Uma característica central do método é sua flexibilidade, visto que ele oferece liberdade de escolha ao pesquisador, para que ajuste as suas configurações às características dos dados, aos objetivos da análise e da função de risco associada.

Para este trabalho, optou-se por adotar as mesmas definições já validadas por DE ALMEIDA (2024), que se mostraram otimizadas para séries temporais de métricas de rede NDT. As principais definições adotadas foram:

- **Tamanho da Janela ($W = 20$):** define o tamanho da janela deslizante de observação. Essa escolha busca equilibrar resolução temporal e robustez estatística, sendo um valor grande o suficiente para estimar com mais estabilidade os parâmetros das funções densidade de probabilidade, mas pequeno o bastante para permitir a detecção online.

- **Probabilidade a Priori** ($\alpha = 1, \beta = 1$): hiperparâmetros da distribuição beta-binomial usada como distribuição *a priori* sobre a posição do ponto de mudança dentro da janela. Essa configuração gera uma distribuição uniforme, atribuindo a mesma probabilidade a todas as posições da janela. Foi adotada por ser neutra e garantir simplicidade e estabilidade numérica.
- $p_0 = 0,8$: limiar de probabilidade individual para que o voto de uma janela seja contabilizado. Esse valor foi otimizado para reduzir falsos positivos mantendo alta sensibilidade.
- $h = 0,9$: limiar que determina a probabilidade mínima para definir um ponto de mudança após a agregação dos votos. O valor adotado prioriza precisão sobre sensibilidade.
- $k = 10$: número mínimo de votos para definir um ponto de mudança. Essa regra ajuda a filtrar anomalias e flutuações transitórias, garantindo consistência temporal.
- $\mu = 0,5$ e $s = 0,9$: parâmetros da função logística. Essa escolha permite dar maior ênfase aos votos dados com maior confiança.
- Agregação ponderada dos votos, proporcionando equilíbrio entre sensibilidade e robustez. Com essa escolha, foi priorizada a simplicidade e a interpretabilidade dos resultados.

Ressalta-se, entretanto, que não é necessário usar esses hiperparâmetros, e a agregação de votos pode ser feita usando técnicas conhecidas (SIQUEIRA, 2025). Neste trabalho, usamos agregações semelhantes às usadas em DE ALMEIDA *et al.* (2025), pois nosso foco é testar a metodologia (com os métodos de change-point e análise de sobrevivência, e ainda testar se o emprego de LLMs a partir dos resultados de clusterização fornece *insights* relevantes a operadores leigos).

4.3 Preparação dos dados para clusterização

Uma vez identificados os pontos de mudança, o passo seguinte da metodologia consiste em preparar os dados para a análise de sobrevivência. Para isso, as séries temporais de cada par cliente-servidor foram segmentadas, utilizando os pontos de mudança nas séries de uma métrica de referência como delimitadores para criar os intervalos. Dessa forma, foram criados dois conjuntos de intervalos distintos: um considerando os instantes de mudança detectados nas séries de vazão de download e outro considerando as mudanças na métrica RTT de upload. Independentemente da

métrica de referência usada, cada intervalo é caracterizado pelas estatísticas locais de todas as métricas de desempenho de rede, além de outras *features*.

A estruturação detalhada destes conjuntos, incluindo a definição dos intervalos, a modelagem da censura e a caracterização multidimensional de cada intervalo por meio de suas estatísticas locais, foi apresentada na Seção 3.5. Os dois conjuntos de intervalos resultantes, cujas características estão sumarizadas na Tabela 3.2, servem como entrada para o algoritmo SurvMixClust, descrito a seguir.

4.4 Clusterização baseada em análise de sobrevivência

O algoritmo SurvMixClust, proposto por BUGINGA (2023), integra clusterização com previsão de funções de sobrevivência. Esse método aprende representações latentes e prevê funções de sobrevivência individuais através de uma mistura de especialistas não paramétricos.

Dado um conjunto de intervalos, potencialmente censurados, e um número K de *clusters*, seu objetivo é determinar as curvas de sobrevivência que melhor representam os intervalos associados a cada uma delas. Para um dado ponto n , sua densidade de probabilidade é dada por:

$$f(t_n|\mathbf{x}_n) = \sum_{k=1}^K \tau_k(\mathbf{x}_n) f(t_n|\theta_k) \quad (4.5)$$

onde t_n é o tempo até o evento para o ponto n , $\mathbf{x}_n$ é o vetor de características do ponto n , K é o número total de *clusters*, $\tau_k(\mathbf{x}_n)$ é o peso da mistura para o *cluster* k , θ_k são os parâmetros da distribuição do *cluster* k e $f(t_n|\theta_k)$ é a função densidade de probabilidade do *cluster* k .

Os pesos da mistura $\tau_k(\mathbf{x}_n)$ são calculados por meio de uma regressão logística multinomial:

$$\tau_k(\mathbf{x}_n) = \frac{\exp(\boldsymbol{\beta}_k^T \cdot \mathbf{x}_n)}{\sum_{l=1}^K \exp(\boldsymbol{\beta}_l^T \cdot \mathbf{x}_n)} \quad (4.6)$$

onde $\boldsymbol{\beta}_k$ é o vetor de coeficientes da regressão para o *cluster* k .

A função de sobrevivência correspondente é:

$$S(t_n|\mathbf{x}_n) = \sum_{k=1}^K \tau_k(\mathbf{x}_n) S(t_n|\boldsymbol{\theta}_k) \quad (4.7)$$

onde $S(t_n|\boldsymbol{\theta}_k)$ é a função de sobrevivência do *cluster* k .

Treinamento do algoritmo

O SurvMixClust é treinado usando uma combinação de máxima verossimilhança e o algoritmo *Expectation-Maximization* (EM) (DEMPSTER *et al.*, 1977). O objetivo é descobrir *clusters* de dados de sobrevivência e, ao mesmo tempo, prever as funções de sobrevivência associadas a cada *cluster*.

O treinamento alterna entre dois passos principais:

Etapas de Expectativa (passo “E”): o algoritmo calcula a probabilidade posterior de um ponto n pertencer a um *cluster* k , dados os parâmetros atuais do modelo. Essa probabilidade é dada por:

$$r_{nk}^{(t)} = P(z_n = k \mid t_n, d_n, \mathbf{x}_n; \theta^{(t)}) = \frac{f(t_n | \theta_k)^{\mathbb{I}(d_n=1)} S(t_n | \theta_k)^{\mathbb{I}(d_n=0)} \tau_k^{(t)}(\mathbf{x}_n)}{\sum_{l=1}^K f(t_n | \theta_l)^{\mathbb{I}(d_n=1)} S(t_n | \theta_l)^{\mathbb{I}(d_n=0)} \tau_l^{(t)}(\mathbf{x}_n)} \quad (4.8)$$

onde $r_{nk}^{(t)}$ é a responsabilidade, ou seja, a probabilidade de pertencimento ao *cluster* k do ponto n , z_n é o *cluster* atribuído ao ponto n , $\mathbb{I}(\cdot)$ é a função indicadora, d_n é o indicador binário de ocorrência do evento para o ponto n e o sobrescrito (t) representa o tempo da iteração.

Uma atribuição é então realizada, associando cada ponto ao *cluster* com a maior probabilidade posterior:

$$q_n^{(t)} = \arg \max_k r_{nk}^{(t)}. \quad (4.9)$$

Etapas de Maximização (passo “M”): os parâmetros do modelo são ajustados para maximizar a verossimilhança dos dados, com base nas responsabilidades $r_{nk}^{(t)}$ obtidas na etapa anterior. A atualização envolve três componentes principais:

1. **Estimativa da função de sobrevivência:** para cada *cluster* k , a função de sobrevivência $S(t | \theta_k)$ é reestimada de forma não paramétrica usando o estimador de Kaplan-Meier (KAPLAN e MEIER, 1958) nos pontos atribuídos ao *cluster*.
2. **Atualização da regressão logística multinomial:** os pesos de atribuição de *cluster* $\tau_k(\mathbf{x}_n)$ são atualizados treinando uma regressão logística multinomial. As características de entrada $\mathbf{x}_n$ são usadas para prever as atribuições de *cluster* $q_n^{(t)}$ do passo “E”.
3. **Verificação de convergência:** o algoritmo itera até que a mudança na verossimilhança logarítmica ou nas atribuições de *cluster* caia abaixo de um limiar predefinido.

Previsão para Novos Dados

Após o treinamento, o modelo pode fazer previsões de sobrevivência para novos dados, baseado nos *clusters* identificados durante o treinamento. A inferência é realizada em duas etapas:

1. **Atribuição de cluster:** Cálculo dos pesos da mistura $\tau_k(\mathbf{x}_{\text{nov}})$ usando os coeficientes da regressão logística multinomial β_k , aprendidos durante o treinamento.
2. **Previsão da função de sobrevivência:** A função de sobrevivência para $\mathbf{x}_{\text{nov}}$ é obtida pela combinação das funções de sobrevivência $S(t | \theta_k)$ de cada *cluster*, ponderadas pelos pesos da mistura $\tau_k(\mathbf{x}_{\text{nov}})$.

4.4.1 Parametrização do SurvMixClust

O algoritmo SurvMixClust foi configurado com os seguintes parâmetros:

- **Número de *clusters* ($K = 2$):** a escolha de dois *clusters* foi uma decisão metodológica orientada pela interpretabilidade e pela capacidade de discriminação do modelo. Para definir este valor, foram realizados testes exaustivos variando K entre 3, 4 e 5. Para cada valor testado, observou-se que apenas duas curvas da função de sobrevivência eram claramente distintas. As demais curvas praticamente se sobrepunham às outras duas e também apresentavam demais características estatísticas próximas a uma delas. Por isso, escolheu-se $K = 2$, configuração que demonstrou a melhor capacidade de separar os intervalos em grupos funcionalmente distintos, o que foi validado qualitativamente pelas características distintas dos *clusters* resultantes, conforme será detalhado no Capítulo 5.
- **Número de execuções ($n_{\text{fits}} = 20$):** uma vez que o algoritmo EM pode convergir para mínimos locais, foram realizadas 20 execuções com inicializações aleatórias distintas para aumentar a probabilidade de encontrar um ajuste globalmente ótimo.
- **Número máximo de iterações do EM ($n_{\text{EM}} = 100$):** define o limite de iterações do algoritmo EM em cada execução, com um critério de parada antecipada para otimizar o tempo de treinamento.
- **Critério de convergência:** o processo iterativo é interrompido se a variação na log-verossimilhança entre duas iterações consecutivas for inferior a 10^{-4} ou se o número máximo de iterações (n_{EM}) for atingido.

4.5 Análise de fatores de risco via regressão logística

Uma das principais contribuições do algoritmo SurvMixClust é a incorporação de uma regressão logística multinomial para calcular os pesos da mistura $\tau_k(\mathbf{x}_n)$, conforme a Equação 4.6. Esses pesos representam a probabilidade de um intervalo com vetor de características $\mathbf{x}_n$ pertencer a cada um dos K *clusters*. Os coeficientes da regressão (β_k) podem ser diretamente interpretados para quantificar o impacto de cada *feature* do intervalo na probabilidade de pertencer a um determinado perfil de estabilidade.

Para $K = 2$, a regressão logística modela a probabilidade de pertencimento ao *Cluster* 1. Os coeficientes β_1 podem ser convertidos em *odds ratios* (razões de chances) através da exponenciação (e^{β_1}). A razão de chances de uma variável quantifica o quanto as chances de pertencer ao *Cluster* 1 aumentam (se > 1) ou diminuem (se < 1) para cada unidade de aumento na variável, mantendo constantes as demais.

Essa abordagem permite uma análise de QoS multifacetada e quantitativa:

1. **Identificação de clientes e servidores críticos:** as variáveis binárias que representam cada cliente e servidor terão suas próprias *odds ratios*. Elementos com *odds ratios* significativamente abaixo de 1 são aqueles mais fortemente associados ao *Cluster* 0, sendo, portanto, candidatos prioritários para investigação.
2. **Identificação das métricas de maior impacto:** os coeficientes associados às estatísticas locais das métricas de rede (médias e desvios padrão de vazão e RTT) revelam quais indicadores de desempenho são os preditores mais fortes de instabilidade. Um coeficiente negativo para a média de RTT de download, por exemplo, indica que um intervalo com maior latência de download possui menor chance de ser classificado como estável (*Cluster* 1).

Essa análise direta dos coeficientes da regressão logística fornece uma ferramenta poderosa e estatisticamente fundamentada para o diagnóstico de problemas de QoS, permitindo que os operadores de rede priorizem suas ações com base em evidências quantitativas extraídas diretamente dos dados, sem a necessidade de conhecimento aprofundado em análise de dados.

4.6 Interpretação dos coeficientes por modelos de linguagem

Uma contribuição fundamental da metodologia é sua capacidade de sintetizar a complexa dinâmica da QoS da rede em tabelas de coeficientes, um formato compacto que não apenas viabiliza a escalabilidade da análise para redes e provedores de maior porte, mas também se mostra diretamente interpretável. Contudo, a interpretação manual dessas tabelas de coeficientes e *odds ratios* pode não ser trivial para equipes operacionais de um provedor de internet sem formação especializada em estatística.

Para mostrar que os coeficientes, por si sós, contêm a informação necessária para uma análise de alto nível e para validar a hipótese de que esta análise pode ser automatizada, foi realizado um experimento com um Grande Modelo de Linguagem (LLM). O objetivo foi mostrar que, com o contexto adequado, um LLM pode interpretar os resultados e responder diretamente às questões de pesquisa que motivaram este trabalho.

Para isso, foi elaborado um *prompt* único e detalhado, apresentado no Quadro 1 da Seção 5.4, que contextualiza a metodologia, seus objetivos e todo o fluxo de coleta e processamento dos dados. Foi anexada às instruções a tabela contendo os coeficientes das duas abordagens de análise (considerando as mudanças na vazão de download e no RTT de upload). Para serializar os dados tabulares e, dessa forma, facilitar seu processamento pelos modelos de linguagem (FANG *et al.*, 2024), as tabelas de coeficientes foram convertidas para formato JSON (*JavaScript Object Notation*). Por fim, o modelo foi instruído a analisar os dados e a gerar um relatório em linguagem natural.

O experimento utilizou o modelo *GPT-5* (OPENAI, 2025), um LLM proprietário de fronteira e baseado em nuvem. Para analisar a viabilidade da abordagem em um cenário em que a privacidade dos dados é uma questão central, o experimento foi replicado com o *gpt-oss-20b* (OPENAI *et al.*, 2025), um modelo de pesos abertos (*open-weights*), de 20 bilhões de parâmetros, executado localmente.

Visando maximizar a performance e a consistência dos LLMs utilizados, o *prompt* foi escrito no idioma inglês, pois os modelos, embora possuam capacidade de receber instruções em outros idiomas, foram treinados e otimizados predominantemente para o processamento de texto em inglês. Isso também elimina potenciais ambiguidades na interpretação de terminologia técnica específica da área de redes de computadores. O modelo foi instruído a gerar a resposta também em inglês.

Além disso, utilizamos formatação em *markdown* para separar instruções, exemplos e dados tabulares enviados nos *prompts*. Apesar de não haver um formato universalmente ótimo que seja eficaz em todos os modelos de linguagem, o uso do *markdown* levou a ganhos de desempenho com modelos GPT (*Generative Pre-trained*

Transformer) em *benchmarks* (HE *et al.*, 2024).

As instruções fornecidas aos modelos e as respostas geradas estão detalhadas na Seção 5.4.

Capítulo 5

Resultados

Este capítulo apresenta os resultados obtidos com a aplicação da metodologia aos dados coletados e pré-processados, conforme descrito nos capítulos anteriores. A organização segue a sequência metodológica: primeiro, são apresentados os resultados da detecção de pontos de mudança; depois, os resultados da análise de sobrevivência e clusterização; e, por fim, os resultados da interpretação assistida por modelos de linguagem.

5.1 Detecção de pontos de mudança

O algoritmo VWCD (STREIT *et al.*, 2021a) foi aplicado para detectar pontos de mudança nas séries temporais de vazão de download e de RTT de upload para cada par cliente-servidor. Embora se tenha optado pelo método VWCD por sua simplicidade e demais vantagens descritas na Subseção 4.2.2, é importante ressaltar que a metodologia proposta neste trabalho é independente do método de detecção de pontos de mudança empregado.

A Figura 5.1 mostra um exemplo de detecção de pontos de mudança na série temporal de RTT de upload do par `cliente16-gru03`. O gráfico superior exibe a série original (cor cinza) e os pontos de mudança detectados (linhas verticais tracejadas em vermelho). O gráfico central mostra o número de votos (cor verde) e o inferior mostra a probabilidade posterior agregada (cor âmbar) para cada ponto da série que obteve ao menos 10 votos. Os pontos de mudança declarados correspondem aos pontos cuja probabilidade posterior agregada ultrapassou o limiar de 0,9.

Pode-se notar que os pontos identificados pelo algoritmo correspondem a mudanças visualmente perceptíveis no comportamento da série. Percebe-se, também, a robustez do algoritmo a anomalias, pois, apesar de alguns votos serem atribuídos a pontos de dados visualmente discrepantes, estes não acumulam votos suficientes ao longo das janelas deslizantes para atingir os limiares de confiança (h) e de votos (k), não sendo, portanto, classificados como pontos de mudança. Isso mostra a robustez

do método em relação a eventos pontuais atípicos.

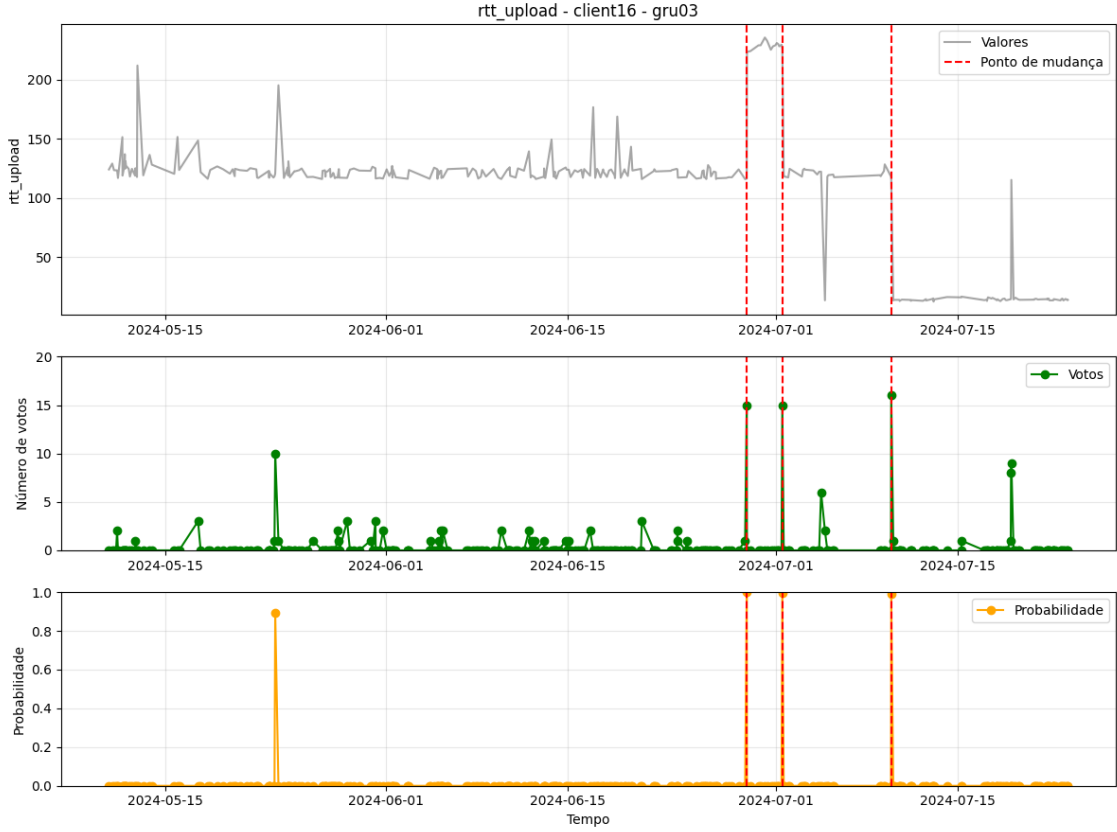


Figura 5.1: Exemplo de aplicação do algoritmo VWCD na série temporal de RTT de upload para o par `cliente16-gru03`.

A detecção adequada dos pontos de mudança é fundamental para a próxima etapa da metodologia, que consiste na análise de sobrevivência e agrupamento dos intervalos entre esses pontos.

5.2 Análise de sobrevivência e clusterização

O algoritmo `SurvMixClust` (BUGINGA e DE SOUZA E SILVA, 2024) foi aplicado aos conjuntos de intervalos criados a partir dos pontos de mudança detectados isoladamente nas séries temporais de vazão de download e RTT de upload. Como detalhado na Seção 4.4, o algoritmo foi configurado para identificar $K = 2$ clusters.

5.2.1 Caracterização dos *clusters*

Em ambas as abordagens, o algoritmo identificou dois *clusters* com características de sobrevivência bem diferentes, como mostram as Figuras 5.2 e 5.3.

Em ambos os casos, o *Cluster 0* (vermelho) mostra uma queda mais acentuada na probabilidade de sobrevivência ao longo do tempo, indicando maior frequência

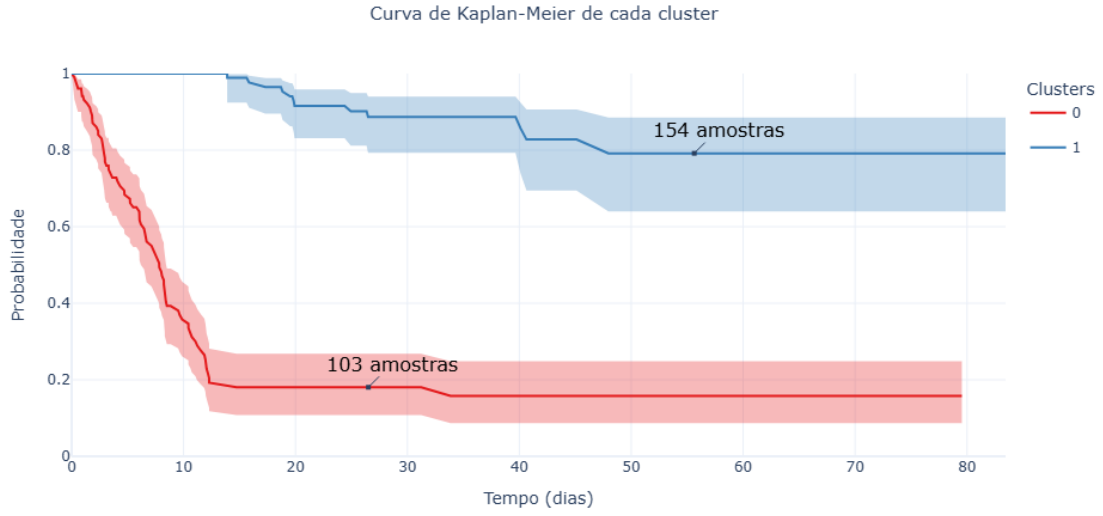


Figura 5.2: Funções de sobrevivência dos *clusters* definidos pelo algoritmo *SurvMix-Clust*, considerando as mudanças no *throughput* de download.

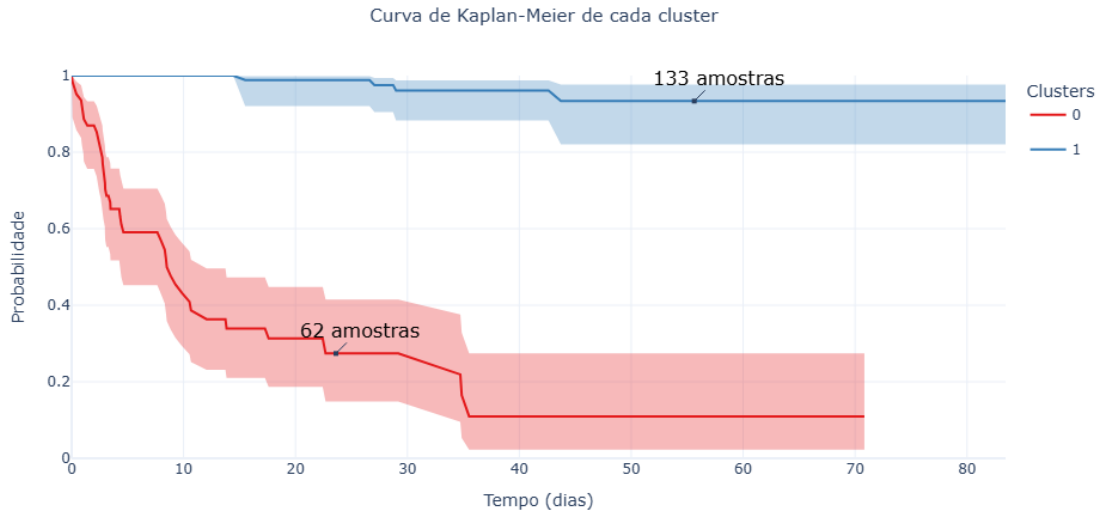


Figura 5.3: Funções de sobrevivência dos *clusters* definidos pelo algoritmo *SurvMix-Clust*, considerando as mudanças no RTT de upload.

de alterações estatísticas na rede. Já o *Cluster 1* (azul) apresenta uma função de sobrevivência mais elevada, caracterizando intervalos mais longos entre mudanças, o que sugere maior estabilidade na rede. É importante lembrar que todas as *features* associadas aos intervalos são consideradas na clusterização, não apenas o tamanho do intervalo. Essas características estão detalhadas na Subseção 3.5.3.

Para uma caracterização quantitativa de cada perfil de desempenho, foram calculadas as estatísticas descritivas das métricas de rede associadas a cada *cluster*. É crucial destacar que essas estatísticas **não são calculadas a partir das estatís-**

ísticas locais dos intervalos, mas sim diretamente sobre os dados brutos (séries temporais) originais.

O procedimento adotado foi o seguinte: para cada *cluster*, foram agregadas todas as medições das séries temporais originais (por exemplo, todos os valores medidos de vazão de download) cujos *timestamps* se encontravam dentro dos períodos de tempo dos intervalos classificados naquele *cluster*. Em seguida, a média, a mediana e o desvio padrão foram calculados sobre esse conjunto agregado de medições brutas.

Considere, por exemplo, um *cluster* hipotético que contenha apenas dois pontos (intervalos): um com 20 e outro com 200 medições. Cada intervalo é representado pelas estatísticas locais calculadas a partir das suas medições. Se fosse calculada a média aritmética simples dessas estatísticas locais, estaríamos, na prática, atribuindo o mesmo peso a um intervalo baseado em 20 observações e a outro baseado em 200, o que geraria um resultado para o *cluster* que desconsidera a quantidade de dados de cada intervalo.

Por isso, para calcular as estatísticas do *cluster* de exemplo, utilizaríamos todos os 220 valores medidos (20 do primeiro intervalo e 200 do segundo), de modo que cada observação contribua igualmente para o resultado. Dessa forma, obtém-se uma representação mais fidedigna do comportamento da rede ao longo do tempo para cada perfil de estabilidade identificado. A Tabela 5.1 sumariza os resultados dessa análise.

Para avaliar se os *clusters* 0 e 1 diferem estatisticamente em todas as métricas nas duas abordagens, aplicamos o teste t de Welch (WELCH, 1947), que não pressupõe igualdade de variâncias e ajusta automaticamente os graus de liberdade em função da dispersão amostral, oferecendo maior robustez quando as variâncias ou tamanhos de amostra diferem entre os grupos:

- **Abordagem 1 (intervalos definidos pelas mudanças no throughput de download):**

- Vazão de download: $t = -5,9052$ e $p = 1,79 \times 10^{-8}$;
- Vazão de upload: $t = -7,6031$ e $p = 1,43 \times 10^{-12}$;
- RTT de download: $t = 6,9745$ e $p = 3,07 \times 10^{-10}$;
- RTT de upload: $t = 6,2875$ e $p = 6,10 \times 10^{-9}$.

- **Abordagem 2 (intervalos definidos pelas mudanças no RTT de upload):**

- Vazão de download: $t = -11,3044$ e $p = 5,78 \times 10^{-19}$;
- Vazão de upload: $t = -10,3036$ e $p = 2,59 \times 10^{-18}$;
- RTT de download: $t = 8,5827$ e $p = 3,03 \times 10^{-12}$;

– RTT de upload: $t = 7,9708$ e $p = 2,63 \times 10^{-11}$.

Em todos os casos, $p \ll 0,05$, permitindo rejeitar a hipótese nula de médias iguais e confirmando que os *clusters* apresentam diferenças estatisticamente significativas em todas as métricas de desempenho. O *Cluster* 1 apresenta consistentemente métricas de comunicação mais favoráveis, com:

- Intervalos mais longos entre mudanças (128,3% maior na abordagem baseada em vazão e 198,1% maior na abordagem baseada em RTT);
- Maior vazão média de download (50,5% maior na abordagem baseada em vazão e 132,8% maior na abordagem baseada em RTT);
- Maior vazão média de upload (63,5% maior na abordagem baseada em vazão e 184,0% maior na abordagem baseada em RTT);
- Menor RTT médio de download (70,4% menor na abordagem baseada em vazão e 80,2% menor na abordagem baseada em RTT);
- Menor RTT médio de upload (59,9% menor na abordagem baseada em vazão e 75,0% menor na abordagem baseada em RTT);
- Menor variabilidade em todas as métricas (exceto vazão de upload, na abordagem baseada em RTT), indicando maior consistência geral nas métricas de desempenho.

Esses resultados **confirmam a hipótese inicial de que uma quantidade maior de mudanças estatísticas nas séries temporais das métricas de desempenho de rede está associada a métricas “piores”** (menor vazão e maior latência).

Uma análise mais detalhada das características dos *clusters* é mostrada pelos gráficos tipo caixa das métricas de desempenho apresentados nas Figuras 5.4 e 5.5.

Os *boxplots* mostram a separação entre os *clusters* em todas as métricas analisadas. Pode-se observar que (i) a distinção entre *clusters* é mais acentuada na abordagem baseada em RTT de upload, com menor sobreposição entre as distribuições; e (ii) o *Cluster* 0 apresenta maior dispersão em todas as métricas (exceto vazão de upload, na abordagem baseada em RTT), indicando, além de pior desempenho médio, maior instabilidade.

5.2.2 Análise dos resultados nas séries temporais

Para ilustrar como a atribuição de *clusters* se manifesta nas séries temporais, foram selecionados dois pares cliente-servidor representativos: `client08-rnp_sp` e

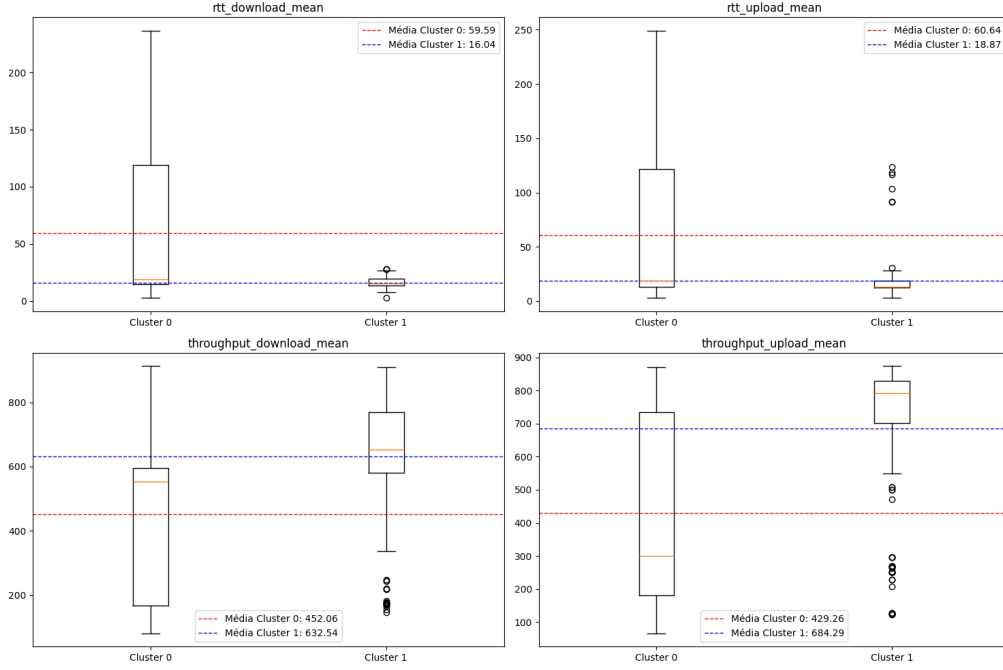


Figura 5.4: Gráficos tipo caixa das *features* analisadas para cada *cluster*, considerando as mudanças no *throughput* de download.

client14-gru03. As Figuras 5.6 e 5.7 exibem as séries temporais das quatro métricas de desempenho de rede para esses pares, com os intervalos entre os pontos de mudança coloridos de acordo com o *cluster* ao qual foram atribuídos (vermelho para *Cluster 0*, azul para *Cluster 1*) e os instantes de mudança estatística detectados (linhas verticais tracejadas na cor verde). A função de sobrevivência individual dos intervalos (gráficos na cor roxa na parte inferior das figuras) é definida por uma mistura das funções de sobrevivência dos *clusters*, cuja proporção é definida por meio de uma regressão logística, conforme detalhado na Seção 4.4.

Na Figura 5.6 (abordagem 1), o algoritmo VWCD detectou um único ponto de mudança na série temporal de vazão de download do par *client08-rnp_sp*, aproximadamente em 17/07/2024. O intervalo anterior a esse ponto foi atribuído ao *Cluster 1*, devido à sua longa duração, às médias elevadas de vazão e aos baixos RTT médios. Isso resultou em uma curva de função de sobrevivência com menor inclinação no intervalo considerado. Já o segmento após o instante de mudança foi atribuído ao *Cluster 0*, devido à queda no valor médio de vazão de download e à sua menor duração. Na Figura 5.7 (abordagem 2), como o VWCD não identificou nenhum ponto de mudança na série temporal de RTT de upload, todo o intervalo foi atribuído ao *Cluster 1*. Esta diferença mostra como as duas abordagens são complementares e podem capturar características distintas do desempenho da rede.

Em relação ao par *client14-gru03*, em ambas as abordagens (Figuras 5.6 e 5.7), foram detectados os mesmos três pontos de mudança. Os três primeiros intervalos

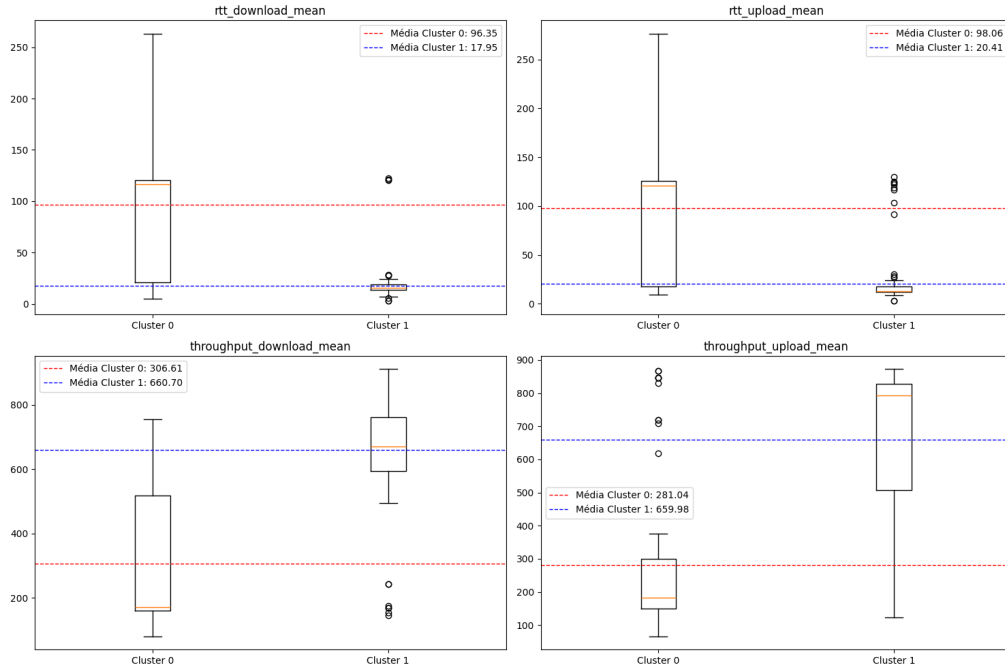


Figura 5.5: Gráficos tipo caixa das *features* analisadas para cada *cluster*, considerando as mudanças no RTT de upload.

foram atribuídos ao *Cluster* 0 devido aos baixos valores médios de vazão (inferiores a 200 Mbps) e às médias elevadas de RTT (aproximadamente 130 ms), resultando em uma curva da função de sobrevivência com decaimento mais acentuado. A partir do instante de mudança por volta de 09/07/2024, observou-se uma melhora significativa nas métricas, com aumento na vazão e uma queda na latência média, fazendo com que o segmento final fosse atribuído ao *Cluster* 1. Esse caso mostra a capacidade da metodologia de detectar e caracterizar mudanças positivas no desempenho da rede, não apenas degradações.

5.3 Análise de fatores de risco e identificação de componentes críticos da rede

A Tabela 5.2 apresenta os coeficientes da regressão logística calculados pelo algoritmo SurvMixClust e as respectivas *odds ratios* para as *features* dos conjuntos de intervalos, considerando as duas abordagens. Esses dados, referentes ao *Cluster* 1, permitem interpretar a influência de cada variável na proporção de mistura desse *cluster*, que corresponde à probabilidade de pertencimento a esse grupo. Assim, coeficientes positivos das variáveis contínuas (como médias e desvios padrão das métricas) indicam que um aumento nessas *features* está associado a um aumento nos logaritmos das razões de chance (*log odds*) de pertencer ao *Cluster* 1, associado a intervalos mais longos entre eventos, considerando todas as demais *features*

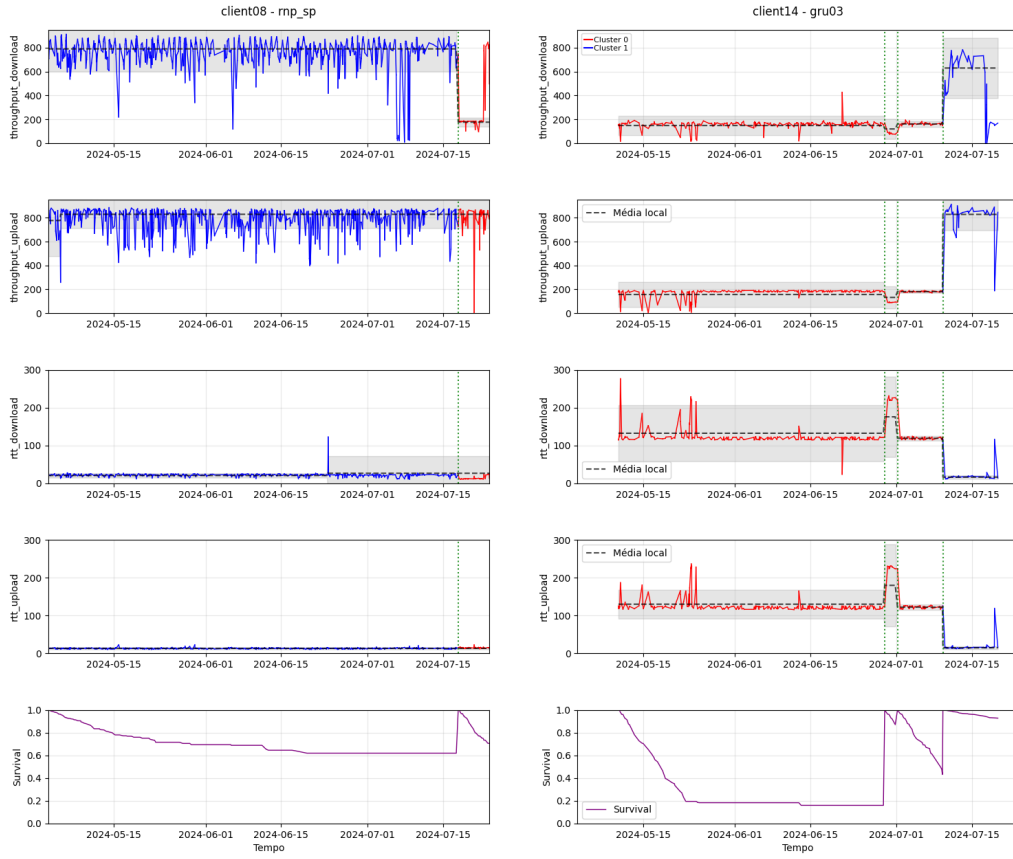


Figura 5.6: Séries Temporais das Métricas de Rede para pares de cliente servidor, considerando as mudanças no *throughput* de download.

constantes. Para as variáveis binárias que identificam os clientes e servidores, um coeficiente positivo indica que o respectivo cliente ou servidor tem maior probabilidade de pertencer ao *Cluster 1*.

A razão de chances de cada *feature* é calculada pela exponencial $e^{\text{coeficiente}}$, e quantifica o efeito da variável como um fator multiplicativo sobre as chances de o evento ocorrer. Valores superiores a 1 indicam que a *feature* aumenta as chances de pertencer ao *Cluster 1*, enquanto valores inferiores a 1 a diminuem e valores iguais a 1 indicam que a *feature* não influencia as chances de pertencimento.

Na Tabela 5.2, observa-se que as duas abordagens revelam tanto consistências quanto variações significativas entre os perfis de clientes e servidores. O cliente *client05*, por exemplo, manteve valores elevados de *odds ratio*, sugerindo uma associação forte com o *Cluster 1* em ambas as abordagens. Por outro lado, *client04*, *client12* e *client13* apresentaram variação considerável de coeficiente entre abordagens, inclusive com mudança de sinal, o que pode ser atribuído a particularidades no tráfego e na dinâmica de cada conexão.

Entre os servidores, enquanto *rnp_rj*, *gru05* e *gru06* mostraram consistência entre as abordagens, *gru03* e *rnp_sp* apresentam diferenças consideráveis, refletindo possíveis variações na resposta aos eventos quando se avalia a vazão *versus* o RTT.

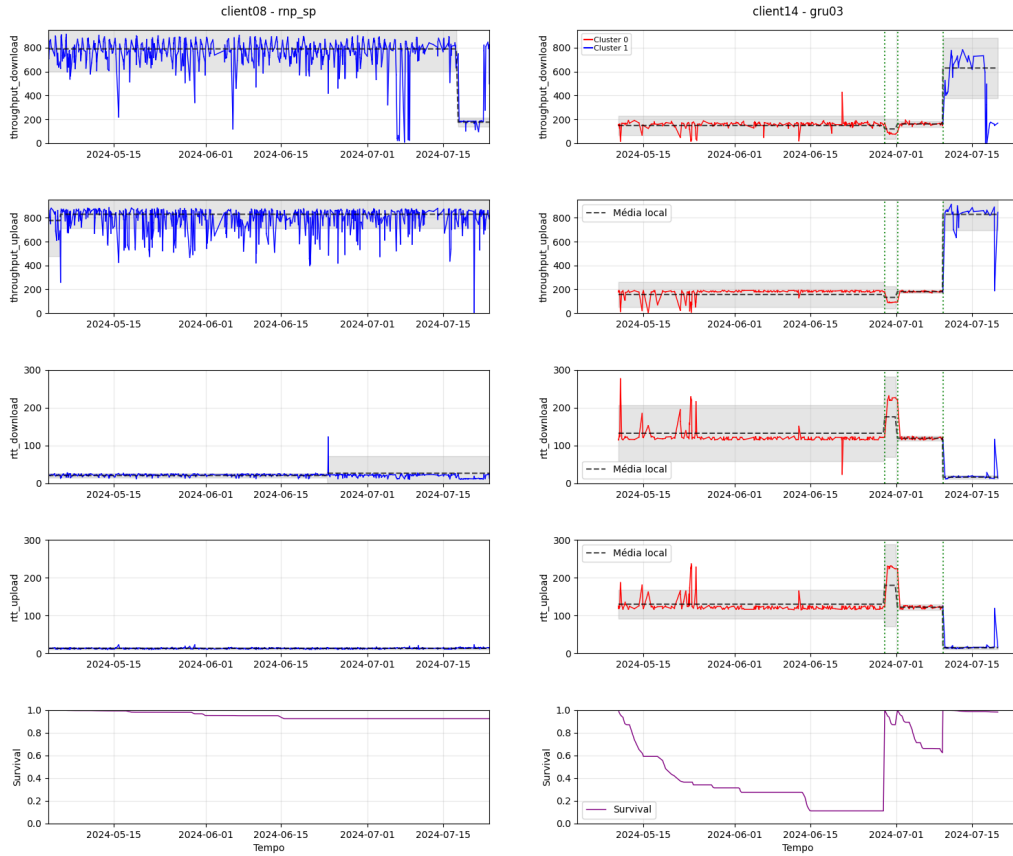


Figura 5.7: Séries Temporais das Métricas de Rede para pares de cliente servidor, considerando as mudanças no RTT de upload.

De maneira semelhante, as *features* de vazão e RTT mostram comportamentos distintos: a média de vazão de download, por exemplo, apresenta um coeficiente próximo a zero na primeira abordagem e torna-se fortemente positivo na segunda, indicando que as mudanças no RTT de upload podem ter grande impacto nas mudanças de vazão de download. Já os RTTs, tanto de download quanto de upload, apresentam diferença elevada nas *odds ratios*, destacando a sensibilidade diferenciada de cada métrica em relação aos eventos analisados.

A partir dessas observações, pode-se imaginar que um método simples baseado nos valores da média e do desvio padrão das métricas, seria suficiente para inferir a QoS dos servidores e clientes. A Tabela 5.3 ilustra os valores da média e desvio padrão das métricas para dois servidores, calculados usando todas as amostras das séries temporais agregadas para cada um. O servidor `gru06` possui coeficientes da regressão logística positivos, o que indica uma maior chance de pertencer ao *Cluster* 1. Já o servidor `rnp_rj` possui coeficientes negativos, indicando maior chance de pertencer ao *Cluster* 0. No entanto, a média e desvio padrão das métricas dos servidores apresentadas na Tabela 5.3 possuem valores semelhantes, não permitindo diferenciar a QoS entre esses servidores.

Conforme detalhado na Seção 4.4, para realizar a clusterização, o algoritmo

SurvMixClust considera tanto o resultado da regressão logística quanto outras características dos intervalos, como a sua duração, a ocorrência de evento e a função de sobrevivência. Por isso, para realizar uma análise robusta da qualidade do serviço de rede, seria necessária uma avaliação criteriosa e multifacetada dos resultados do algoritmo. Do ponto de vista de um provedor de internet, essa compreensão completa dos dados de rede para uma grande quantidade de clientes e servidores exigiria investimentos em operadores especializados em aprendizado de máquina, estatística, entre outros. Mesmo especialistas encontrariam dificuldades para realizar essa análise em tempo hábil para atender à demanda de clientes insatisfeitos, por exemplo, o que tornaria a metodologia impraticável em um cenário operacional real.

Os coeficientes da regressão logística permitem aplicações práticas imediatas, como a identificação prévia de clientes ou servidores críticos, ou seja, com coeficientes negativos e com alto valor absoluto. Isso permite que os operadores direcionem as intervenções para melhorar as condições da conexão desses elementos de forma proativa. A instabilidade consistente dos servidores `rnp_rj` e `rnp_sp`, por exemplo, sugere um problema estrutural que afeta múltiplos clientes, enquanto o comportamento dos clientes `client02` e `client07` indica um problema que pode estar mais próximo ao usuário final.

Por outro lado, o estudo das características de clientes e servidores com coeficientes altamente positivos, como o cliente `client05` e o servidor `gru05`, permite replicá-las em outros pontos e, assim, otimizar o desempenho geral da rede.

5.4 Interpretação automatizada dos coeficientes com LLM

Para validar a hipótese de que a análise dos coeficientes da regressão logística pode ser automatizada e traduzida em respostas às questões de pesquisa apresentadas na Seção 1.2, foi conduzido um experimento utilizando Grandes Modelos de Linguagem (LLM), um proprietário e baseado em nuvem, o *GPT-5* (OPENAI, 2025) e um de pesos abertos (*open-weights*), rodando *on-premise*, o *gpt-oss-20b* OPENAI *et al.* (2025).

Foi elaborado um *prompt* que contextualiza o problema, descreve a metodologia, apresenta os dados dos coeficientes e solicita respostas para as quatro questões de pesquisa definidas no Capítulo 1. O Quadro 1 a seguir apresenta o *prompt* fornecido. A Tabela 5.2, contendo os coeficientes calculados em ambas as abordagens, foi inserida em formato JSON a um LLM, juntamente com as instruções e separada por abordagem.

Quadro 1 – Prompt para o LLM

Persona

You are a senior network performance analyst at a major Brazilian ISP. Your task is to diagnose Quality of Service (QoS) issues by interpreting measurement data from network probes.

Experiment Context

You will analyze data from two parallel approaches, each defined by a different reference metric for detecting instability (statistical changes).

Data Collection

- Probes: 16 Raspberry Pi devices ran Network Diagnostic Tool (NDT) tests every 30 minutes.
- Raw Metrics: Throughput (download/upload) and RTT (download/upload).
- Targets: RNP (Brazilian National Research and Education Network) servers.

Methodological Pipeline

1. Change-Point Detection: For each client-server pair, statistical change-points were detected for the reference metric, slicing the timeline into "intervals" of homogeneous behavior.
2. Multivariate Interval Characterization: While each interval is defined by a single reference metric (either throughput or RTT), every interval was characterized by the local mean and standard deviation of ALL FOUR metrics.
3. Clustering and Logistic Regression:
 - A clustering algorithm grouped all intervals into exactly two clusters: Cluster 0 and Cluster 1.
 - The analysis revealed two distinct operational profiles: an "unstable" cluster (Cluster 0), characterized by frequent statistical changes and degraded performance metrics such as higher latency and lower throughput, and a "stable" cluster (Cluster 1), defined by extended periods of consistent performance and superior quality of service.
 - The clustering algorithm uses ALL interval's features to define the clusters: time, event occurrence, client/server ID, and the local mean/std of all the four metrics.
 - The model uses multinomial logistic regression with Cluster 0 as the baseline to define the probability of an interval

belonging to one cluster or the other.

- CRUCIAL: The logistic regression coefficients you will see are for the log-odds of an interval belonging to Cluster 1. A positive coefficient for a variable increases the odds of being in Cluster 1; a negative coefficient increases the odds of being in Cluster 0.

Logistic Regression Coefficients

1. Approach 1 (based on changepoints in download throughput):

```
```json
{coeficientes_abordagem_1}
```
```

2. Approach 2 (based on changepoints in upload RTT):

```
```json
{coeficientes_abordagem_2}
```
```

Your Mission

1. Identification of the most influential metrics: use the logistic regression coefficients to identify the metrics that have the greatest impact on the definition of the clusters.
2. Critical Clients Identification: Use the logistic regression coefficients to identify clients most strongly associated with Cluster 0.
3. Critical Servers Identification: Use the logistic regression coefficients to identify servers most strongly associated with Cluster 0.
4. Technical Report per Approach: Consolidate findings for each approach (throughput and RTT) into a technical report.
8. Consolidated Operational Report: Synthesize both technical reports into a single, actionable document for network operations teams.

Core Directives

- Data Grounding: Base ALL conclusions strictly on the data provided. Do not extrapolate or assume.
- Output Language: Brazilian Portuguese (PT-BR), keeping technical

- terms like "RTT" and "throughput" in English.
- Formatting: Use Markdown '##' titles for structure. DO NOT write tables.

Fonte: elaborado pelo autor (2025).

O Quadro 2 apresenta a resposta gerada pelo LLM, que traduz os dados quantitativos em uma análise diagnóstica.

Quadro 2 – Resposta gerada pelo LLM

```
## Executive Summary
- Cluster 0 = unstable; Cluster 1 = stable. Coefficients are log-odds toward Cluster 1 (positive stable; negative unstable).
- Across both approaches, download-side RTT (level and variability) is the dominant driver toward instability, while higher download throughput drives stability.
- Critical clients (most associated with Cluster 0): Approach 1 client02, client07 (very strong), then client13, client08, client09, client10. Approach 2 client09, client12, client02, client07, client08. Consistent cross-approach signals: client02, client07, client08, client09.
- Critical servers: rnp_rj and rnp_sp consistently push toward Cluster 0 in both approaches; gru05 and gru06 push toward Cluster 1.

## How to read the coefficients
- Positive coefficient: increases odds the interval is in Cluster 1 (stable).
- Negative coefficient: increases odds the interval is in Cluster 0 (unstable).
- Larger absolute values indicate stronger influence within the same model.

## Approach 1 (changepoints in download throughput)
1) Most influential metrics (by absolute coefficient magnitude)
- Strong drivers toward Cluster 0 (negative):
  - rtt_download_mean: -1.4541
  - rtt_download_std: -0.9938
```

- rtt_upload_mean: -0.6237
- throughput_download_std: -0.5446
- Strong drivers toward Cluster 1 (positive):
 - rtt_upload_std: +0.8974
 - throughput_upload_std: +0.4793
 - throughput_upload_mean: +0.4126
 - throughput_download_mean: +0.1128

Interpretation anchored to data:

- Higher download RTT (both mean and variability) is the largest signal toward instability.
- Higher upload throughput and higher upload RTT variability correlate with stability in this model.
- Higher variability in download throughput correlates with instability.

2) Clients most associated with Cluster 0 (negative coefficients)

- Very strong: client02 (-1.9521), client07 (-1.7344)
- Additional: client13 (-0.6060), client08 (-0.4008), client09 (-0.3170), client10 (-0.2823)

3) Servers most associated with Cluster 0 (negative coefficients)

- rnp_rj (-1.3675), rnp_sp (-1.2327), gru03 (-0.4635)
- Servers associated with Cluster 1: gru05 (+1.8158), gru06 (+1.2647)

Approach 2 (changepoints in upload RTT)

1) Most influential metrics (by absolute coefficient magnitude)

- Strong drivers toward Cluster 0 (negative):
 - rtt_download_std: -1.9935
 - rtt_download_mean: -0.6625
 - throughput_download_std: -0.5894
 - rtt_upload_std: -0.1243
- Strong drivers toward Cluster 1 (positive):
 - throughput_download_mean: +1.3197
 - throughput_upload_std: +0.8281
 - throughput_upload_mean: +0.6422
 - rtt_upload_mean: +0.2166

Interpretation anchored to data:

- Variability in download RTT is the single strongest driver toward

```

    instability.
- Higher download throughput (mean) is the single strongest driver
  toward stability.
- Higher variability in download throughput pushes toward
  instability; higher variability in upload throughput correlates
  with stability in this model.

2) Clients most associated with Cluster 0 (negative coefficients)
- Strong: client09 (-0.9923), client12 (-0.7821), client02 (-0.6515),
  client07 (-0.6036)
- Additional: client08 (-0.4562), client04 (-0.2015), client10
  (-0.0109)

3) Servers most associated with Cluster 0 (negative coefficients)
- rnp_rj (-0.9031), rnp_sp (-0.4217)
- Servers associated with Cluster 1: gru05 (+0.5908), gru06
  (+0.5873), gru03 (+0.1500)

## Cross-Approach Consensus and Priorities
- Metrics with consistent, strong association to instability (Cluster
  0):
  - rtt_download_std: strongest negative in Approach 2 and strong
    negative in Approach 1.
  - rtt_download_mean: strong negative in both approaches.
  - throughput_download_std: negative in both approaches (moderate).
- Metrics with consistent association to stability (Cluster 1):
  - throughput_download_mean: strongly positive in Approach 2;
    positive (smaller) in Approach 1.
  - throughput_upload_mean: positive in both.
  - throughput_upload_std: positive in both.
- Critical clients (consistent negative across both approaches):
  - client02, client07, client08, client09.
- Critical servers (consistent negative across both approaches):
  - rnp_rj (strongest), rnp_sp (clear), mixed/weak for gru03 (neg in
    A1, slight pos in A2).
  - Servers associated with stability across both: gru05, gru06.

## Operational Actions (data-driven)
1) Focus monitoring on download-side latency and its variability
- Treat rtt_download_std and rtt_download_mean as primary QoS risk

```

indicators.

- Increase sampling/alerting sensitivity on these metrics for all probes, especially when targeting rnp_rj and rnp_sp.

2) Prioritize investigation on critical clients

- Tier 1 (highest priority): client02, client07.
- Tier 2: client09, client08.
- Actions: increase test frequency, compare paths to rnp_rj vs rnp_sp vs gru05/gru06, and correlate with time-of-day to isolate patterns within the intervals already flagged by changepoints.

3) Prioritize investigation on critical servers/targets

- rnp_rj (primary) and rnp_sp (secondary): schedule intensified probing from multiple clients, including those not flagged as critical, to separate endpoint-specific from destination-specific effects.
- Use gru05 and gru06 as control targets (associated with stability) to benchmark expected stable performance levels for the same clients.

4) Stabilization levers to track (guided by model signals)

- Reduce download RTT variation: any change that lowers rtt_download_std or rtt_download_mean should move intervals toward Cluster 1, given the strong negative coefficients.
- Maintain/increase download throughput: improvements in throughput_download_mean are aligned with stability (especially in Approach 2).
- Watch download throughput variability: reductions in throughput_download_std are favorable.

5) Reporting and alerting

- Flag any interval where rtt_download_std and/or rtt_download_mean rise concurrently, especially when targeting rnp_rj or rnp_sp.
- Maintain a watchlist dashboard for the named clients and servers with per-interval cluster assignment probabilities, using these coefficients as the explanatory signals.

Summary of Most Influential Metrics (by approach)

- Approach 1: rtt_download_mean (), rtt_download_std (), rtt_upload_std (+) are the top three by magnitude.

```

- Approach 2: rtt_download_std (), throughput_download_mean (+),
  throughput_upload_std (+) are the top three by magnitude.

## Summary of Critical Entities for Cluster 0
- Clients (Approach 1): client02, client07, then client13, client08,
  client09, client10.
- Clients (Approach 2): client09, client12, client02, client07,
  client08.
- Cross-approach client focus: client02, client07, client08, client09.
- Servers (both approaches): rnp_rj (strongest), rnp_sp;
  control/comparison servers associated with stability: gru05,
  gru06.

```

Fonte: gerado pelo modelo *GPT-5* a partir do prompt do Quadro 1 (2025).

Este experimento demonstra que a metodologia proposta não apenas identifica padrões complexos nos dados, mas também que seus resultados, quando processados por um LLM, podem ser traduzidos em um formato acessível e diretamente aplicável por equipes operacionais, diminuindo a barreira entre a análise de dados e a tomada de decisão.

O objetivo era verificar se o modelo conseguiria, sem intervenção humana, interpretar os valores numéricos e responder diretamente às questões de pesquisa que orientaram este trabalho.

É fundamental ressaltar que a mesma análise foi realizada com o modelo *gpt-oss-20b* (OPENAI *et al.*, 2025), um LLM de menor porte executado localmente. Os resultados foram funcionalmente os mesmos, identificando os mesmos clientes, servidores e métricas críticas, com variações apenas no nível de detalhamento e no estilo da resposta. Este resultado corrobora a viabilidade de integrar a metodologia a modelos de linguagem menores e de código aberto, permitindo a sua implementação em ambientes operacionais sem a dependência de APIs externas e proprietárias.

5.4.1 Validação das respostas às questões de pesquisa

O relatório gerado pelo LLM, apresentado no Quadro 2, valida a terceira hipótese apresentada na Seção 1.3: a de que a análise dos coeficientes da regressão logística pode ser automatizada para responder diretamente às questões de pesquisa propostas. O experimento demonstra que, ao ser devidamente contextualizado, o modelo de linguagem é capaz de traduzir os resultados estatísticos em diagnósticos acioná-

veis. A seguir, detalha-se como a resposta do LLM aborda cada uma das questões de pesquisa.

1. **Como identificar clientes com qualidade insatisfatória?** O LLM interpretou corretamente os coeficientes negativos como indicadores de instabilidade (associação com o *Cluster* 0). Ao analisar os resultados de ambas as abordagens, o modelo foi capaz de realizar uma análise de consenso (“*Cross-Approach Consensus*”), identificando os clientes com associação mais forte e consistente com o perfil de desempenho insatisfatório. A resposta do modelo destacou, de forma acertada, os clientes `client02` e `client07` como os mais críticos, seguidos por `client09` e `client08`, o que corresponde a uma análise precisa dos dados da Tabela 5.2. Essa capacidade de hierarquizar a criticidade transforma a análise estatística em uma lista de prioridades para a equipe de operações.
2. **Quais servidores apresentam desempenho inferior?** De forma análoga, o LLM identificou os servidores com os coeficientes mais negativos. O relatório gerado apontou os servidores `rnp_rj` e `rnp_sp` como os mais críticos em ambas as abordagens, ressaltando a consistência do sinal de instabilidade. Adicionalmente, o modelo utilizou os servidores com coeficientes positivos (`gru05` e `gru06`) como um ponto de referência para o comportamento estável, demonstrando uma capacidade de análise comparativa que vai além da simples listagem dos componentes problemáticos.
3. **Quais métricas de rede exercem maior impacto no desempenho?** O modelo *GPT-5* interpretou com precisão tanto os sinais quanto as magnitudes dos coeficientes das métricas de rede. O relatório destacou corretamente a média e, principalmente, o desvio padrão do RTT de download (coeficientes de $-1,4541$ e $-1,9935$ nas abordagens 1 e 2, respectivamente) como os mais fortes indicadores de instabilidade. Em contrapartida, identificou a média de *throughput* de download (coeficiente de $+1,3197$ na abordagem 2) como um fator determinante para a estabilidade da conexão. Essa análise valida que o LLM é capaz de discernir as nuances da influência de cada métrica, respondendo à questão de forma direta.
4. **Como transformar análises estatísticas em informações relevantes?** Esta questão é respondida pela própria estrutura do relatório gerado pelo modelo. Ele não se limitou a extrair e listar os dados, mas os sintetizou em uma seção de “*Operational Actions*”, que traduz os achados estatísticos em um plano de ação claro. As recomendações, como a priorização de monitoramento do RTT de download e a investigação focada nos clientes e servidores críticos

identificados, exemplificam a transformação de dados complexos em *insights* operacionais. Isso confirma o potencial da metodologia para democratizar o acesso a análises de dados avançadas, convertendo os resultados do modelo SurvMixClust em diretrizes práticas para a gestão da rede.

Este experimento, portanto, demonstra que a metodologia proposta não apenas identifica padrões complexos nos dados, mas também que seus resultados, quando processados por um LLM, podem ser traduzidos em um formato acessível e diretamente aplicável por equipes operacionais, diminuindo a barreira entre a análise de dados e a tomada de decisão. É fundamental ressaltar que a mesma análise foi realizada com o modelo *gpt-oss-20b* (OPENAI *et al.*, 2025), um LLM de menor porte executado localmente. Os resultados foram funcionalmente os mesmos, identificando os mesmos clientes, servidores e métricas críticas, com variações apenas no nível de detalhamento e no estilo da resposta. Este resultado corrobora a viabilidade de integrar a metodologia a modelos de linguagem menores e de código aberto, permitindo a sua implementação em ambientes operacionais sem a dependência de APIs externas e proprietárias.

A análise confirma que os coeficientes da regressão logística, quando devidamente contextualizados, contêm a informação necessária para uma avaliação abrangente da QoS. Além disso, evidencia que os LLMs são uma ferramenta viável para automatizar a interpretação desses resultados, tornando a metodologia proposta não apenas estatisticamente robusta, mas também operacionalmente aplicável.

Tabela 5.1: Características estatísticas dos *clusters* identificados.

| Característica | Cluster 0 ¹ | Cluster 1 ¹ | Cluster 0 ² | Cluster 1 ² |
|--|------------------------|------------------------|------------------------|------------------------|
| Nº de intervalos | 103 | 154 | 62 | 133 |
| Duração (dias) | | | | |
| Média | 10,6 | 24,3 | 10,6 | 31,5 |
| Mediana | 6,7 | 19,6 | 4,4 | 29,0 |
| Desvio padrão | 13,8 | 22,5 | 13,1 | 24,9 |
| Freq. relativa de eventos | 0,738 | 0,078 | 0,645 | 0,030 |
| Probabilidades de sobrevivência | | | | |
| 1 dia | 0,946 | 1,000 | 0,910 | 1,000 |
| 7 dias | 0,579 | 1,000 | 0,586 | 1,000 |
| 15 dias | 0,226 | 1,000 | 0,354 | 0,990 |
| 30 dias | 0,218 | 0,901 | 0,216 | 0,990 |
| 60 dias | 0,171 | 0,832 | 0,089 | 0,961 |
| 90 dias | 0,171 | 0,832 | 0,089 | 0,961 |
| Vazão download (Mbps) | | | | |
| Média | 455,0 | 684,9 | 288,9 | 672,6 |
| Mediana | 577,7 | 703,8 | 172,5 | 692,9 |
| Desvio padrão | 278,6 | 167,7 | 211,2 | 184,1 |
| Vazão upload (Mbps) | | | | |
| Média | 398,3 | 651,2 | 223,9 | 635,9 |
| Mediana | 299,5 | 781,0 | 187,3 | 766,5 |
| Desvio padrão | 274,4 | 259,9 | 149,9 | 263,4 |
| RTT download (ms) | | | | |
| Média | 60,2 | 17,8 | 97,2 | 19,3 |
| Mediana | 18,9 | 18,2 | 115,8 | 17,9 |
| Desvio padrão | 66,2 | 7,0 | 73,8 | 15,0 |
| RTT upload (ms) | | | | |
| Média | 62,0 | 24,9 | 100,8 | 25,2 |
| Mediana | 18,6 | 16,8 | 118,1 | 16,9 |
| Desvio padrão | 68,4 | 30,5 | 76,1 | 31,1 |

¹ Abordagem considerando como eventos de interesse as mudanças na vazão de download.² Abordagem considerando como eventos de interesse as mudanças no RTT de upload.

Tabela 5.2: Coeficientes da regressão logística e *odds ratios* do modelo *SurvMixClust* calculadas nas duas abordagens.

| Tipo | Feature | Coef. ¹ | Odds Ratio ¹ | Coef. ² | Odds Ratio ² |
|---------------|----------------|--------------------|-------------------------|--------------------|-------------------------|
| Clientes | client01 | 0,8326 | 2,299 | 0,0250 | 1,025 |
| | client02 | -1,9521 | 0,142 | -0,6515 | 0,521 |
| | client03 | 0,4735 | 1,606 | 0,6016 | 1,825 |
| | client04 | 0,1233 | 1,131 | -0,2015 | 0,818 |
| | client05 | 1,3414 | 3,825 | 0,8722 | 2,392 |
| | client06 | 0,3746 | 1,454 | 0,4000 | 1,492 |
| | client07 | -1,7344 | 0,176 | -0,6036 | 0,547 |
| | client08 | -0,4008 | 0,670 | -0,4562 | 0,634 |
| | client09 | -0,3170 | 0,728 | -0,9923 | 0,371 |
| | client10 | -0,2823 | 0,754 | -0,0109 | 0,989 |
| | client11 | 0,6815 | 1,977 | 0,0927 | 1,097 |
| | client12 | 0,4860 | 1,626 | -0,7821 | 0,457 |
| | client13 | -0,6060 | 0,546 | 0,5795 | 1,785 |
| | client14 | 0,5455 | 1,726 | 0,1137 | 1,120 |
| | client15 | 0,4048 | 1,499 | 0,0988 | 1,104 |
| | client16 | 0,0461 | 1,047 | 0,9178 | 2,504 |
| Servidores | gru03 | -0,4635 | 0,629 | 0,1500 | 1,162 |
| | gru05 | 1,8158 | 6,146 | 0,5908 | 1,805 |
| | gru06 | 1,2647 | 3,542 | 0,5873 | 1,799 |
| | rnp_rj | -1,3675 | 0,255 | -0,9031 | 0,405 |
| | rnp_sp | -1,2327 | 0,292 | -0,4217 | 0,656 |
| Média | Vazão download | 0,1128 | 1,119 | 1,3197 | 3,742 |
| | Vazão upload | 0,4126 | 1,511 | 0,6422 | 1,901 |
| | RTT download | -1,4541 | 0,234 | -0,6625 | 0,516 |
| | RTT upload | -0,6237 | 0,536 | 0,2166 | 1,242 |
| Desvio Padrão | Vazão download | -0,5446 | 0,580 | -0,5894 | 0,555 |
| | Vazão upload | 0,4793 | 1,615 | 0,8281 | 2,289 |
| | RTT download | -0,9938 | 0,370 | -1,9935 | 0,136 |
| | RTT upload | 0,8974 | 2,453 | -0,1243 | 0,883 |

¹ Abordagem considerando como eventos de interesse as mudanças na vazão de download.

² Abordagem considerando como eventos de interesse as mudanças no RTT de upload.

Tabela 5.3: Estatísticas agregadas das métricas de desempenho para os servidores gru06 e rnp_rj.

| Métrica | gru06 | rnp_rj |
|-----------------------------------|--------|---------|
| RTT download (média) | 16,90 | 18,68 |
| RTT download (desvio padrão) | 10,11 | 5,95 |
| Vazão download (média) | 708,07 | 664,74 |
| Vazão download (desvio padrão) | 169,63 | 165,39 |
| RTT upload (média) | 21,86 | 24,46 |
| RTT upload (desvio padrão) | 31,91 | 25,10 |
| Vazão upload (média) | 684,58 | 628,95 |
| Vazão upload (desvio padrão) | 256,49 | 257,89 |
| Coef. Reg. Logística ¹ | 1,2647 | -1,3675 |
| Coef. Reg. Logística ² | 0,5873 | -0,9031 |

¹ Abordagem considerando como eventos de interesse as mudanças na vazão de download.

² Abordagem considerando como eventos de interesse as mudanças no RTT de upload.

Capítulo 6

Conclusão

Este trabalho desenvolveu e validou uma metodologia integrada para análise de Qualidade de Serviço (QoS) que se distingue por três contribuições para a área de gerenciamento de redes. A abordagem proposta representa uma mudança de paradigma, por meio de uma análise proativa fundamentada na identificação de perfis de instabilidade baseados na distribuição dos pontos de mudança nas séries temporais das métricas de desempenho de rede. Ao validar esta abordagem em um ambiente operacional real, mostrou-se ser possível não apenas identificar com precisão os elementos de rede que necessitam de atenção, mas também interpretar os resultados por meio de Grandes Modelos de Linguagem (LLMs).

6.1 Contribuições principais

As contribuições centrais desta dissertação podem ser resumidas em três pilares:

1. **Caracterização de perfis de instabilidade de conexão:** Foi proposta uma nova abordagem para análise de QoS que utiliza a distribuição de mudanças estatísticas como indicador de estabilidade da conexão. Ao contrário de métodos que se baseiam em limiares fixos de desempenho, nossa metodologia combina a detecção de pontos de mudança com análise de sobrevivência para clusterização. O resultado é a identificação não supervisionada de perfis de comportamento de rede, permitindo uma caracterização multidimensional que considera simultaneamente vazão, latência e estabilidade temporal.
2. **Validação empírica em cenário operacional:** A metodologia foi validada com um conjunto de dados exclusivo, gerado por medições sistemáticas em um provedor de internet brasileiro ao longo de três meses. Os resultados comprovaram a hipótese central de que a instabilidade temporal é um indicador robusto de degradação de QoS. O *cluster* instável (*Cluster 0*), caracterizado por mudanças mais frequentes, apresentou desempenho significativamente inferior,

com RTT médio significativamente maior e vazão média menor em comparação com o *cluster* estável (*Cluster 1*), conforme detalhado na Seção 4.4.

3. **Análise quantitativa de fatores de risco para operadores de rede:** Foi mostrado como os coeficientes da regressão logística, inerentes ao modelo de clusterização, podem ser diretamente utilizados como uma ferramenta de diagnóstico. Essa abordagem permite ranquear clientes e servidores por criticidade e identificar as métricas de desempenho que mais impactam a instabilidade da rede, traduzindo um resultado estatístico complexo em um guia prático e acionável para a manutenção proativa da infraestrutura.
4. **Disseminação científica dos resultados:** Parte da metodologia e dos resultados apresentados neste trabalho foi previamente publicada em forma de artigo em GOMES *et al.* (2025). O estudo apresenta uma versão preliminar da proposta de caracterização de perfis de instabilidade temporal e dos resultados experimentais, que foram aprofundados e estendidos neste trabalho.

6.2 Validação das hipóteses da pesquisa

Os resultados apresentados no Capítulo 5 permitiram validar as três hipóteses que nortearam esta pesquisa:

1. **A instabilidade temporal das métricas de rede é um indicador de degradação de serviço.** Esta hipótese foi ratificada pela análise dos *clusters*. O *Cluster 0*, definido por intervalos mais curtos entre mudanças (duração média de 10,6 dias em ambas as abordagens), apresentou métricas de desempenho estatisticamente piores em todas as dimensões analisadas quando comparado ao *Cluster 1* (que apresentou intervalos com duração média de 24,3 e 31,5 dias). Especificamente, o RTT médio do *Cluster 0* foi até 80,2% superior, enquanto a vazão média de download foi até 184,0% inferior, confirmando a correlação direta entre frequência de mudanças e degradação da QoS.
2. **Intervalos entre mudanças estatísticas podem ser agrupados para representar diferentes níveis de QoS.** A aplicação do algoritmo SurvMix-Clust validou esta hipótese ao identificar dois *clusters* distintos e estatisticamente separáveis ($p \ll 0,05$ no teste t de Welch WELCH (1947) para todas as métricas). Cada *cluster* demonstrou não apenas uma curva de sobrevivência distinta, refletindo diferentes padrões de estabilidade, mas também características de desempenho (vazão e RTT) marcadamente diferentes, funcionando como perfis empíricos de “conexão estável” (*Cluster 1*) e “conexão instável” (*Cluster 0*).

3. **Os coeficientes gerados pelo modelo podem traduzir a análise em informações acionáveis.** Esta hipótese foi confirmada pela análise dos coeficientes da regressão logística, calculados pelo algoritmo SurvMixClust. A análise permitiu identificar os componentes da rede mais propensos a problemas de desempenho, como, por exemplo, o servidor `rnp_rj` e o cliente `client02`), além das métricas que exercem maior impacto, como a média e o desvio padrão do RTT de download. Isso demonstra que o próprio modelo, sem necessidade de camadas de interpretação adicionais, fornece os *insights* acionáveis para os operadores.

A metodologia aqui proposta se diferencia das abordagens existentes ao integrar a análise de estabilidade temporal com a caracterização de desempenho. Diferentemente de trabalhos que se concentram em gargalos específicos da rede, como a infraestrutura Wi-Fi residencial (SHARMA *et al.*, 2024a), nossa abordagem avalia o desempenho *end-to-end*. Em contraste com métodos que focam na detecção de anomalias ou congestionamento em rotas específicas (PI *et al.*, 2024; SHARMA *et al.*, 2024b), a presente abordagem não se limita a detectar eventos de degradação, mas os utiliza para construir perfis de comportamento ao longo do tempo.

6.3 Limitações da metodologia

Apesar dos resultados positivos, a validade e aplicabilidade dos resultados apresentados estão sujeitas a certas condições, que definem o escopo da metodologia:

- **Dependência da densidade dos dados:** A eficácia da análise depende da disponibilidade de séries temporais com uma quantidade mínima de medições. O limiar de 100 medições por par cliente-servidor, adotado neste estudo para garantir a robustez estatística, pode limitar a aplicabilidade em redes com padrões de coleta mais esparsos.
- **Generalização dos resultados:** A metodologia foi validada com dados reais coletados durante um período específico a partir de 16 medidores (clientes), sendo 12 instalados em diferentes pontos na rede de um provedor de internet brasileiro e 4 instalados em residências de voluntários, em localidades e provedores distintos. A generalização dos perfis de desempenho identificados para outros contextos, como redes móveis ou corporativas, bem como para diferentes regiões geográficas, requereria um novo treinamento do modelo SurvMixClust, além de estudos adicionais de validação.

6.4 Trabalhos futuros

Com base nos resultados e limitações identificadas, várias direções se apresentam para pesquisas futuras, agrupadas em diferentes categorias.

Para aprimoramentos metodológicos, destaca-se a possibilidade de explorar configurações alternativas do algoritmo VWCD, investigando diferentes estratégias de votação e a utilização de configurações adaptativas. Também seria importante implementar e avaliar a detecção multivariada de pontos de mudança, considerando múltiplas métricas simultaneamente, em contraste com as abordagens univariadas adotadas.

É importante ressaltar que, embora a implementação atual utilize o algoritmo VWCD, a metodologia é agnóstica em relação ao método de detecção de pontos de mudança. Essa modularidade permite a incorporação futura de outros algoritmos, sem alterar as etapas subsequentes de clusterização e interpretação.

Adicionalmente, a expansão do conjunto de métricas para incluir perda de pacotes, *jitter* e variabilidade de RTT pode enriquecer a identificação de problemas de QoS. A integração de informações contextuais, como dados sobre tecnologias de acesso e tipos de planos de internet, pode melhorar a contextualização dos resultados.

A abordagem de interpretação dos resultados por LLM pode ser aprimorada para receber outros artefatos produzidos pelo algoritmo SurvMixClust, como, por exemplo, os dados estatísticos dos *clusters* para atribuição automatizada de perfis de risco de cada um, além de dados dos intervalos de um cliente ou servidor específico, que permitiriam um diagnóstico mais detalhado.

No campo das aplicações, a metodologia poderia ser estendida para:

1. Desenvolver modelos preditivos para antecipar degradações futuras de desempenho, baseados nos padrões históricos de mudanças;
2. Ampliar o sistema de interpretação para identificação automática de causas raiz dos problemas detectados;
3. Criar um sistema de recomendação que sugira intervenções específicas baseadas no histórico de casos semelhantes;
4. Desenvolver interfaces e APIs para integração com sistemas existentes de gerenciamento de rede, facilitando a adoção prática;
5. Implementar interfaces de visualização interativas (*dashboards*) para operadores explorarem os resultados; e
6. Investigar o uso de outros LLMs para traduzir os resultados em relatórios em linguagem natural e sistemas de perguntas e respostas.

A validação da metodologia em outros contextos também é importante. A adaptação para redes móveis, considerando métricas específicas como intensidade de sinal, poderia expandir seu uso. Ambientes corporativos, com seus requisitos específicos de QoS, representam outro campo de aplicação. Redes de IoT, com suas restrições e padrões característicos de comunicação, demandariam adaptações da metodologia. A extensão para análise de desempenho de serviços em nuvem permitiria considerar métricas como tempo de resposta de APIs e utilização de recursos.

Os modelos de linguagem usados na interpretação também oferecem possibilidades de melhoria. O ajuste fino de modelos para a tarefa específica de interpretação de dados de rede poderia melhorar a precisão das análises. A implementação de técnicas de *Retrieval-Augmented Generation* (RAG) poderia enriquecer as interpretações com conhecimento de documentos privados. O desenvolvimento de mecanismos de execução automática de intervenções, com integração de agentes com ferramentas de gerenciamento de redes, também seria útil. Por fim, a criação de sistemas adaptativos que aprendam com *feedbacks* dos usuários poderia melhorar a qualidade das análises.

Estas direções de pesquisa podem expandir o alcance da metodologia proposta, contribuindo para o avanço das técnicas de monitoramento e gestão de qualidade de serviço em redes.

6.5 Considerações finais

Este trabalho demonstrou que a estabilidade temporal, mensurada pela frequência de mudanças estatísticas, é um indicador de desempenho de rede tão ou mais relevante que os valores absolutos de vazão e latência. A metodologia desenvolvida, que integra detecção de pontos de mudança, análise de sobrevivência e interpretação por modelos de linguagem, representa uma mudança de paradigma: da monitoração reativa baseada em limiares para uma análise proativa baseada na identificação de perfis de instabilidade.

Ao validar esta abordagem em um ambiente operacional real, validou-se que é possível não apenas identificar com precisão os elementos de rede que necessitam de atenção, mas também democratizar o acesso a essa análise por meio da interpretação automatizada por LLM. A integração de algoritmos estatísticos rigorosos com a capacidade interpretativa da inteligência artificial generativa abre uma nova fronteira para o gerenciamento de redes, onde a complexidade dos dados deixa de ser uma barreira para se tornar a base de uma operação mais inteligente, eficiente e preditiva.

Referências Bibliográficas

- BUGINGA, G., DE SOUZA E SILVA, E., 2024, “Clustering Survival Data using a Mixture of Non-parametric Experts”, *arXiv preprint*, (maio). Disponível em: <<https://doi.org/10.48550/arXiv.2405.15934>>. Acesso em: 26 jan. 2025.
- BUGINGA, G. C., 2023, *Algoritmos de Aprendizado de Máquina Aplicados a Dados Censoriados para Previsão de Mortalidade de Pacientes com Doença Arterial Coronariana*. Dissertação de mestrado, COPPE - Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil.
- CLARK, D., WEDEMAN, S., 2021, “Measurement, Meaning and Purpose: Exploring the M-Lab NDT Dataset”. In: *TPRC49: The 49th Research Conference on Communication, Information and Internet Policy*, ago. Disponível em: <<http://dx.doi.org/10.2139/ssrn.3898339>>.
- COX, D. R., 2018, “Regression Models and Life-Tables”, *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 34, n. 2 (12), pp. 187–202. ISSN: 0035-9246. doi: 10.1111/j.2517-6161.1972.tb00899.x. Disponível em: <<https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>>.
- DA HORA, D., VAN DOORSELAER, K., VAN OOST, K., et al., 2018a, “Predicting the effect of home Wi-Fi quality on QoE”, *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pp. 944–952.
- DA HORA, D. N., ASRESE, A. S., CHRISTOPHIDES, V., et al., 2018b, “Narrowing the Gap Between QoS Metrics and Web QoE Using Above-the-fold Metrics”, *Passive and Active Measurement, PAM 2018*, pp. 31–43.
- DE ALMEIDA, C. M., LEÃO, R. M. M., DE SOUZA E SILVA, E., 2024, “Inferindo pontos de mudança em séries temporais com dados não rotulados: um breve estudo usando dados do NDT”. In: *Anais do XLII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pp. 686–699, Porto Alegre, RS, Brasil, maio.

- DE ALMEIDA, C. M., LEÃO, R. M. M., DE SOUZA E SILVA, E., 2025, “Inferring change points in unlabelled time series data collected from the network diagnosis tool”, *Ann. des Télécommunications*, v. 80, n. 9, pp. 775–792. doi: 10.1007/S12243-025-01103-2. Disponível em: <<https://doi.org/10.1007/s12243-025-01103-2>>.
- DE ALMEIDA, C. M., 2024, *Change-point detection in time series: a study of online methods applied to computer network measurements*. Dissertação de mestrado, COPPE - Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil.
- DE SOUZA E SILVA, C. G., BUGINGA, G. C., DE SOUZA E SILVA, E. A., et al., 2022, “Prediction of Mortality in Coronary Artery Disease: Role of Machine Learning and Maximal Exercise Capacity”, *Mayo Clinic Proceedings*, v. 97, n. 8, pp. 1472–1482.
- DE SOUZA E SILVA, E., GAIL, H. R., 1989, “Calculating availability and performability measures of repairable computer systems using randomization”, *J. ACM*, v. 36, n. 1, pp. 171–193. doi: 10.1145/58562.59307. Disponível em: <<https://doi.org/10.1145/58562.59307>>.
- DE SOUZA E SILVA, E., GAIL, H. R., 1986, “Calculating Cumulative Operational Time Distributions of Repairable Computer Systems”, *IEEE Trans. Computers*, v. 35, n. 4, pp. 302–332. doi: 10.1109/TC.1986.1676765. Disponível em: <<https://doi.org/10.1109/TC.1986.1676765>>.
- DE SOUZA E SILVA, E., GAIL, H. R., 1990, “Analyzing Scheduled Maintenance Policies for Repairable Computer Systems”, *IEEE Trans. Computers*, v. 39, n. 11, pp. 1309–1324. doi: 10.1109/12.61041. Disponível em: <<https://doi.org/10.1109/12.61041>>.
- DEMPSTER, A. P., LAIRD, N. M., RUBIN, D. B., 1977, “Maximum Likelihood from Incomplete Data via the EM Algorithm”, *Journal of the Royal Statistical Society*, v. 39, n. 1, pp. 1–38.
- FANG, X., XU, W., TAN, F. A., et al., 2024, “Large language models (LLMs) on tabular data: Prediction, generation, and understanding — a survey”, *Transactions on Machine Learning Research*. Disponível em: <<https://www.amazon.science/publications/large-language-models-llms-on-tabular-data-prediction-generation-and-understanding-a-survey>>. Acesso em: 15 fev. 2025.

- FEARNHEAD, P., LIU, Z., 2011, “Efficient Bayesian analysis of multiple changepoint models with dependence across segments”, *Statistics and Computing*, v. 21, n. 2 (abr.), pp. 217–229. ISSN: 1573-1375. doi: 10.1007/s11222-009-9163-6. Disponível em: <<https://doi.org/10.1007/s11222-009-9163-6>>.
- GILL, P., DIOT, C., OHLSEN, L. Y., et al., 2022, “M-Lab: user initiated internet data for the research community”, *SIGCOMM Comput. Commun. Rev.*, v. 52, n. 1, pp. 34–37. ISSN: 0146-4833. doi: 10.1145/3523230.3523236. Disponível em: <<https://doi.org/10.1145/3523230.3523236>>.
- GOMES, I. J. A., BUGINGA, G., E SILVA, E. S., et al., 2025, “Diagnóstico de QoS em Redes a partir de Análise conjunta de Sobrevivência e Mudanças Estatísticas em Séries Temporais de Desempenho”. In: *Anais do XXIV Workshop em Desempenho de Sistemas Computacionais e de Comunicação*, pp. 49–60, Porto Alegre, RS, Brasil. SBC. doi: 10.5753/wperformance.2025.8604. Disponível em: <<https://sol.sbc.org.br/index.php/wperformance/article/view/36280>>.
- HARCHAOUI, Z., MOULINES, E., BACH, F., 2008, “Kernel Change-point Analysis”. In: Koller, D., Schuurmans, D., Bengio, Y., et al. (Eds.), *Advances in Neural Information Processing Systems*, v. 21. Curran Associates, Inc. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2008/file/08b255a5d42b89b0585260b6f2360bdd-Paper.pdf>.
- HE, J., RUNGTA, M., KOLECZEK, D., et al., 2024. “Does Prompt Formatting Have Any Impact on LLM Performance?” Disponível em: <<https://arxiv.org/abs/2411.10541>>. Acesso em: 26 ago. 2025.
- ISHWARAN, H., KOGALUR, U. B., BLACKSTONE, E. H., et al., 2008, “Random survival forests”, *The Annals of Applied Statistics*, v. 2, n. 3, pp. 841 – 860. doi: 10.1214/08-AOAS169. Disponível em: <<https://doi.org/10.1214/08-AOAS169>>.
- KAPLAN, E. L., MEIER, P., 1958, “Nonparametric Estimation from Incomplete Observations”, *Journal of the American Statistical Association*, v. 53, n. 282, pp. 457–481. ISSN: 01621459, 1537274X.
- KATZMAN, J. L., SHAHAM, U., CLONINGER, A., et al., 2018, “DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network”, *BMC Medical Research Methodology*, v. 18, n. 1, pp. 24. ISSN: 1471-2288. doi: 10.1186/s12874-018-0482-1. Disponível em: <<https://doi.org/10.1186/s12874-018-0482-1>>.

- KAWAHARA, Y., SUGIYAMA, M., 2012, “Sequential change-point detection based on direct density-ratio estimation”, *Stat. Anal. Data Min.*, v. 5, n. 2 (abr.), pp. 114–127. ISSN: 1932-1864. doi: 10.1002/sam.10124. Disponível em: <<https://doi.org/10.1002/sam.10124>>.
- KVAMME, H., ØRNULF BORGAN, 2021, “Continuous and discrete-time survival prediction with neural networks”, *Lifetime Data Analysis*, v. 27, n. 4, pp. 710–736. ISSN: 1572-9249. doi: 10.1007/s10985-021-09532-6. Disponível em: <<https://doi.org/10.1007/s10985-021-09532-6>>.
- LAI, T. L., 2018, “Sequential Changepoint Detection in Quality Control and Dynamical Systems”, *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 57, n. 4 (12), pp. 613–644. ISSN: 0035-9246. doi: 10.1111/j.2517-6161.1995.tb02052.x. Disponível em: <<https://doi.org/10.1111/j.2517-6161.1995.tb02052.x>>.
- LEE, C., ZAME, W., YOON, J., et al., 2018, “DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks”, *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 32, n. 1 (Apr.). doi: 10.1609/aaai.v32i1.11842. Disponível em: <<https://ojs.aaai.org/index.php/AAAI/article/view/11842>>.
- LINDSTÅHL, A., AVRACHENKOV, K., POPOVIC, M., 2024, “Change Point Detection with Adaptive Measurement Schedules for Network Performance Verification”, *ACM SIGMETRICS Performance Evaluation Review*, v. 51, n. 1.
- LIU, S., YAMADA, M., COLLIER, N., et al., 2013, “Change-point detection in time-series data by relative density-ratio estimation”, *Neural Networks*, v. 43, pp. 72–83. ISSN: 0893-6080. doi: <https://doi.org/10.1016/j.neunet.2013.01.012>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0893608013000270>>.
- M-LAB, 2025. “NDT - Network Diagnostic Tool”. Disponível em: <<https://www.measurementlab.net/tests/ndt/>>. Acesso em: 27 jan. 2025.
- OLIVEIRA, D. A., LEÃO, R. M. M., DE SOUZA E SILVA, E., 2024, “A Multi-Dimensional Approach to Understanding the Effect of Page Content and Infrastructure on Page Load Time”. In: *Anais do XLII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pp. 756–769, Porto Alegre, RS, Brasil, maio.

- OPENAI, 2025, *GPT-5 System Card*. Relatório técnico, OpenAI, ago. Disponível em: <<https://cdn.openai.com/pdf/18a02b5d-6b67-4cec-ab64-68cdfbddebcd/preparedness-framework-v2.pdf>>. Acesso em: 31 ago. 2025.
- OPENAI, :, AGARWAL, S., et al., 2025. “gpt-oss-120b gpt-oss-20b Model Card”. Disponível em: <<https://arxiv.org/abs/2508.10925>>. Acesso em: 31 ago. 2025.
- PAGE, E. S., 1954, “Continuous Inspection Schemes”, *Biometrika*, v. 41, n. 1-2 (06), pp. 100–115. ISSN: 0006-3444. doi: 10.1093/biomet/41.1-2.100. Disponível em: <<https://doi.org/10.1093/biomet/41.1-2.100>>.
- PAUL, B., BISCHOF, Z. S., SCHULMAN, A., 2022, “The Importance of Contextualization of Crowdsourced Active Speed Test Measurements”. In: *Proceedings of the ACM Internet Measurement Conference (IMC)*.
- PI, Y., JAMIN, S., WEI, L., 2024, “Measuring Congestion-Induced Performance Imbalance in Internet Load Balancing at Scale”. In: *Proceedings of the ACM SIGCOMM*. Disponível em: <<https://yibopi.github.io/assets/papers/congi24.pdf>>.
- PUCHA, H., ZHANG, Y., MAO, Z. M., et al., 2007, “Understanding network delay changes caused by routing events”, *SIGMETRICS Perform. Eval. Rev.*, v. 35, n. 1 (jun.), pp. 73–84. ISSN: 0163-5999. doi: 10.1145/1269899.1254891. Disponível em: <<https://doi.org/10.1145/1269899.1254891>>.
- RNP, 2023. “RNP e M-Lab assinam MoU para disponibilização de medições de desempenho de rede para fins de pesquisa”. Disponível em: <<https://www.rnp.br/rnp-e-m-lab-assinam-mou-para-disponibilizacao-de-medies-de-desempenho-de-rede-para-fins-de-pesquisa>>. Acesso em: 26 jan. 2025.
- SANTOS, G. H. A., MENDONÇA, G., LEÃO, R. M. M., et al., 2022, “Detecção de anomalias em redes baseada em medições de QoS e rótulos de QoE com ruído”. In: *Anais do XL Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pp. 98–111, Porto Alegre, RS, Brasil, maio.
- SAVERIMOUTOU, A., MATHIEU, B., VATON, S., 2019, “A 6-month Analysis of Factors Impacting Web Browsing Quality for QoE Prediction”, *Computer Networks*, v. 164.

- SHARMA, A., SUNDARESAN, S., MADHYASTHA, H. V., 2024a, “Measuring the Prevalence of WiFi Bottlenecks in Home Access Networks”. In: *Proceedings of the ACM Internet Measurement Conference (IMC)*, a.
- SHARMA, T., PADAKI, R., FEAMSTER, N., 2024b, “Beyond Data Points: Regionalizing Crowdsourced Latency Measurements”. In: *Proceedings of the ACM SIGMETRICS*. ACM, b. Disponível em: <https://taveeshsharma.com/assets/pdf/SIGMETRICS_2025.pdf>.
- SIQUEIRA, I. G., 2025, *Agregação Probabilística na Detecção Não Supervisionada de Eventos em Séries Temporais*. Exame de qualificação de doutorado, COPPE - Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil.
- STREIT, A., RIBEIRO, M., LEÃO, R., et al., 2021a, “Efeito do confinamento causado pela pandemia Covid-19 nos perfis de tráfego residencial”. In: *Anais do XXXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pp. 238–251, Porto Alegre, RS, Brasil, a. SBC. doi: 10.5753/sbrc.2021.16724. Disponível em: <<https://sol.sbc.org.br/index.php/sbrc/article/view/16724>>.
- STREIT, A., SANTOS, G. H. A., LEÃO, R. M. M., et al., 2021b, “Network anomaly detection based on tensor decomposition”, *Computer Networks*, v. 200.
- SUNDBERG, P., O’DONNELL, S., SCHLINKER, B., 2024, “Measuring Network Latency from a Wireless ISP: Variations Within and Across Subnets”. In: *Proceedings of the ACM Internet Measurement Conference (IMC)*.
- WANG, P., LI, Y., REDDY, C. K., 2019, “Machine Learning for Survival Analysis: A Survey”, *ACM Comput. Surv.*, v. 51, n. 6 (fev.). ISSN: 0360-0300. doi: 10.1145/3214306. Disponível em: <<https://doi.org/10.1145/3214306>>.
- WELCH, B. L., 1947, “The Generalization of ‘Student’s’ Problem when Several Different Population Variances are Involved”, *Biometrika*, v. 34, n. 1/2, pp. 28–35. ISSN: 00063444. Disponível em: <<http://www.jstor.org/stable/2332510>>.
- XIMENES, D., MENDONÇA, G., SANTOS, G. H. A., et al., 2018, “O Problema de Detecção e Localização de Eventos em Séries Temporais Aplicado a Redes de Computadores”. In: *Anais do XVII Workshop em Desempenho de Sistemas Computacionais e de Comunicação*, Porto Alegre, RS, Brasil.

SBC. doi: 10.5753/wperformance.2018.3323. Disponível em: <<https://sol.sbc.org.br/index.php/wperformance/article/view/3323>>.

YANG, Z., KANNIAINEN, J., KROGERUS, T., et al., 2022, “Prognostic modeling of predictive maintenance with survival analysis for mobile work equipment”, *Scientific Reports*, v. 12, n. 1 (maio), pp. 8529. ISSN: 2045-2322. doi: 10.1038/s41598-022-12572-z. Disponível em: <<https://doi.org/10.1038/s41598-022-12572-z>>.