Inferindo pontos de mudança em séries temporais com dados não rotulados: um breve estudo usando dados do NDT

Cleiton M. de Almeida[1], **Rosa M. M. Leão** [1], **Edmundo de Souza e Silva**[1]

[1]Programa de Engenharia de Sistemas e Computação
Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brasil

cleiton.almeida@coppe.ufrj.br, {rosam, edmundo}@land.ufrj.br

Abstract. *We use change-point detection algorithms in latency and throughput time series, collected using the NDT tool, to identify moments of statistical changes in these series. We examine three classical methods (Shewhart, EWMA, and CUSUM) and highlight that their straightforward implementations may not be suitable for detecting such points. We then present simple strategies to remedy this problem. We also introduce a novel change-point detection method that offers flexibility and interpretability to facilitate the decision-making process. We show a simple application that can be used to assess QoS, even when labels are not available.*

Resumo. *Utilizamos algoritmos de detecção de pontos de mudança em séries temporais de latência e throughput, coletados por meio da ferramenta NDT, para identificar momentos de mudanças estatísticas nessas séries. Examinamos três métodos clássicos (Shewhart, EWMA e CUSUM) e destacamos que suas implementações simples podem não ser adequadas para detectar tais pontos. Apresentamos então estratégias simples para remediar este problema. Também introduzimos um novo método de detecção de pontos de mudança que oferece flexibilidade e interpretabilidade para facilitar o processo de tomada de decisão. Mostramos uma aplicação simples que pode ser usada para avaliar QoS, mesmo quando os rótulos não estão disponíveis.*

1. Introdução

Identificar mudanças estatísticas significativas em dados de séries temporais tem sido um assunto de extensa pesquisa por décadas. A relevância deste campo abrange uma ampla gama de aplicações, incluindo monitoramento de desempenho de redes de computadores. Essas mudanças estatísticas podem se manifestar como alterações na média, variância, estrutura de correlação ou outras características dos dados. Apesar da importância crítica desta área e dos inúmeros algoritmos desenvolvidos ao longo dos anos, a implementação em cenários reais continua apresentando desafios significativos. Um dos principais desafios está relacionado à eficácia desses algoritmos que pode variar consideravelmente de acordo com o domínio de aplicação. Esta variabilidade requer uma abordagem adaptada às características e requisitos específicos de cada área onde o algoritmo será usado.

Neste trabalho, estamos interessados em determinar pontos de mudança e anomalias em séries temporais, em particular na aplicação de alguns métodos *online* em séries de vazão e latência coletadas em redes domésticas. Nosso objetivo é o de refinar metodologias simples e clássicas e desenvolver novas metodologias que possam ser usadas na

prática, especialmente no contexto de monitoramento e diagnóstico de desempenho de redes domésticas.

De forma geral, a detecção de pontos de mudança (*change-point detection*) envolve identificar momentos em que o comportamento estatístico de uma série temporal sofre uma alteração importante por um período nos descritores estatísticos da série, por exemplo, uma mudança no valor médio das amostras. A duração do período depende da aplicação específica e seus requisitos. Este conceito tem sido amplamente estudado em vários contextos, como destacado em trabalhos como [Tartakovsky et al. 2015, Tartakovsky et al. 2013, Aminikhanghahi and Cook 2017]. Um conceito intimamente relacionado, porém distinto, é o de detecções de "anomalias", também conhecido como detecção de *outliers*. Detecção de anomalias consiste em identificar instâncias onde amostras de dados desviam significativamente do que é considerado "normal" [Chandola et al. 2009]. Embora exista uma conexão intrínseca entre esses dois problemas, eles não são idênticos.

É importante notar que as características específicas dos dados de séries temporais em análise influenciam na escolha dos métodos de detecção. Por exemplo, a frequência de amostragem dos dados (e.g., medições por segundo, minuto ou hora) afeta a sensibilidade necessária no algoritmo de detecção. A alta variabilidade ou ruído nos dados pode tornar muito difícil a identificação dos reais pontos de mudança ou de anomalias. Consequentemente, métodos capazes de filtrar o ruído são preferidos nesses cenários. Além disso, lidar com grandes conjuntos de dados exige algoritmos que sejam precisos e computacionalmente eficientes [Aminikhanghahi and Cook 2017].

Neste artigo, o nosso foco principal é verificar o comportamento de métodos bem estabelecidos, comumente aplicados em vários domínios, mas aqui empregados especificamente para identificar tanto pontos de mudança quanto anomalias em dados de séries temporais obtidos do *Network Diagnosis Tool* (Ferramenta de Diagnóstico de Rede (NDT)) [M-Lab 2024]. Um objetivo chave é avaliar a adequação desses métodos convencionais para aplicação *online*, utilizando as séries temporais univariadas que coletamos.

Como não temos rótulos para os pontos de mudança ou as anomalias (*ground-truth*) em nossos dados, não é possível determinar os algoritmos que melhor se adequam ao problema. Então, fizemos uma comparação de como eles se comportam nas séries reais coletadas. Iniciamos aplicando os métodos Shewhart, EWMA e CUSUM em suas formas fundamentais. A escolha foi motivada por serem métodos clássicos da literatura, de fácil implementação e baixo custo computacional e, em seguida, propusemos algumas modificações para esses algoritmos básicos. É importante observar que, embora nosso estudo tenha sido feito usando essas três formulações básicas, várias extensões avançadas e métodos alternativos existem na literatura. Nossa análise visa obter *insights* sobre como métodos bem conhecidos, embora simples, se comportam no contexto da análise de dados de desempenho de redes, uma área que apresenta seu próprio conjunto de desafios e requisitos. Propusemos também um novo método para detecção de pontos de mudança que permite uma análise da confiança dos resultados. Essa nova proposta é simples mas flexível e baseada em conceitos de métodos de votação ponderada.

Nossas contribuições podem ser resumidas da seguinte forma: **(a)** Aplicando métodos clássicos de mudança de ponto (especificamente Shewhart, EWMA e CUSUM)

a um conjunto de dados reais de medições em rede, mostramos que suas implementações básicas não são suficientes para distinguir entre pontos de mudança e anomalias, o que evidencia os desafios na aplicação de métodos tradicionais a dados complexos do mundo real. Motivados pelas limitações observadas, desenvolvemos e implementamos estratégias muito simples para lidar com esses desafios na prática. **(b)** Introduzimos um novo algoritmo baseado em um sistema de votação ponderada. Esta abordagem mostrou ser robusta na detecção de pontos de mudança. O algoritmo proposto também oferece vantagens quanto à interpretabilidade e flexibilidade, permitindo que seja ajustado de acordo com as características específicas dos dados de séries temporais sendo analisados. **(c)** Aplicamos esses métodos a um conjunto de dados reais, fornecendo *insights* sobre sua utilidade no contexto de monitoramento de qualidade de rede.

Na seção 2, apresentamos alguns trabalhos relacionados. Na seção 3, apresentamos uma breve descrição dos métodos clássicos em estudo. Usando um exemplo simples extraído de nosso conjunto de dados reais, ilustramos na seção 4 alguns dos desafios inerentes à implementação padrão desses métodos, e então propomos modificações simples, objetivando aprimorar seus desempenhos. Na seção 5, descrevemos um novo método que desenvolvemos e discutimos as potenciais vantagens que o mesmo oferece. Descrevemos o conjunto de dados que coletamos na seção 6, analisamos os resultados da aplicação dos métodos estudados a esses dados e fornecemos *insights* sobre a utilidade prática desses métodos em cenários do mundo real. A seção 7 apresenta nossas conclusões.

2. Trabalhos relacionados

Nas últimas décadas, inúmeros algoritmos foram desenvolvidos para os problemas de detecção de pontos de mudança e/ou anomalias. No caso de Redes de Computadores, métodos sequenciais de pontos de mudança são frequentemente utilizados para identificar padrões de tráfego anômalos, conforme detalhado em estudos como [Vasantam et al. 2021] e [Tartakovsky et al. 2013]. Além disso, vários esforços focam na detecção de mudanças que influenciam na Qualidade de Serviço (QoS) fornecida por ISPs ou afetam a Qualidade de Experiência (QoE) do usuário. Por limitação de espaço e por haver vários *surveys* na literatura, referenciamos apenas alguns trabalhos mais ligados ao nossos interesses.

Nossas pesquisas tem lidado com problemas relacionados à detecção de pontos de mudança e anomalias. Por exemplo, em [Ximenes et al. 2018], introduzimos um novo *framework* para detectar mudanças em dados de séries temporais coletados de roteadores domésticos. Este *framework* utiliza correlação espaço-temporal de pontos de mudança em diferentes residências. O objetivo principal foi identificar o local mais provável na rede do provedor de acesso à internet (ISP) que poderia estar impactando várias séries temporais simultaneamente. Para determinar esses pontos de mudança, foi utilizado um método de otimização *offline*. No entanto, nosso foco atual é em métodos *online*, sem ênfase na localização do equipamento específico que potencialmente impacta várias séries temporais simultaneamente.

Recentemente em [Santos et al. 2022], usamos Modelos Ocultos de Markov (HMM) para detectar pontos de mudança em séries temporais de perda de pacotes. Este estudo também incorpora dados de reclamações de usuários a um *call center*, com o objetivo final de inferir a QoE da rede. Em [Streit et al. 2021], anomalias de rede foram

identificadas usando amostras de tráfego coletadas a cada minuto de milhares de roteadores domésticos. Neste caso, a decomposição tensorial foi empregada como a principal ferramenta analítica. Dentre os resultados desta abordagem destacamos a detecção de ataques DDoS e vários eventos de rede que afetam negativamente a qualidade de serviço (QoS).

Apesar de existir uma extensa literatura para a detecção de anomalias utilizando métodos de aprendizado de máquina clássico e aprendizado profundo, estudos recentes mostram que os métodos estatísticos ainda apresentam melhor desempenho [Braei and Wagner 2020, Schmidl et al. 2022]. Na tarefa de detecção de pontos de mudanças, a maioria dos métodos não-supervisionados recentes ainda possuem abordagem estatística. Um trabalho recente [Li et al. 2024] na área de aprendizado profundo, reportou desempenho competitivo com os métodos clássicos. No entanto, o método proposto é *offline* e supervisionado, o que o diferencia do nosso objetivo que é a detecção *online* e não-supervisionada de pontos de mudança e anomalias.

Com relação a rótulos, um aspecto importante a mencionar é que os conjuntos de dados tradicionais, como de imagens de animais bem diferentes, são normalmente fáceis de serem rotulados por humanos. Por outro lado, rotular séries temporais não é normalmente uma tarefa trivial, como no caso dos dados de redes analisados. Na fig. 4b, por exemplo, a série exibe um comportamento não estacionário de variância, tornando a tarefa de rotular os pontos de mudanças subjetivo.

3. Métodos tradicionais de detecção de mudanças

Considere uma sequência de observações $(X_1, X_2, \ldots)$ e um instante ν tal que as amostras $(X_1, X_2, \ldots, X_\nu)$ e $(X_{\nu+1}, X_{\nu+2}, \ldots)$ pertencem à densidades distintas $f_0(x)$ e $f_1(x)$, respectivamente. O instante ν é chamado de ponto de mudança na sequência e o problema em questão é determinar ν na série temporal $\mathbf{X} = \{X_t\}_{t\geq 1}$. As séries temporais que usaremos neste trabalho são provenientes de medidas de latência e vazão, são univariadas e não-estacionárias. Entretanto, em geral, os métodos de detecção de pontos de mudança são projetados para encontrar instantes onde ocorrem alterações significativas no processo com base na suposição de que o processo é estacionário entre esses pontos. Semelhantemente à maioria dos trabalhos, adotamos essa suposição e também a de que as observações são *iid*.

Método de Shewhart [Shewhart 1929, Tartakovsky et al. 2015]: Esse método foi desenvolvido no contexto de Controle Estatístico de Processos e, para o caso gaussiano, pode ser definido em termos da média e do desvio padrão de um processo. Cada amostra de dados é comparada com limites com o objetivo de detectar mudanças no processo. Supondo que a média e o desvio padrão da sequência sejam conhecidos, pode-se formular o problema como um teste de hipóteses, onde a hipótese nula $\mathcal{H}_0$ é uma amostra X_t pertencer a um processo com média μ_0 e variância σ_0 conhecidas ($\mathbf{X} \sim \mathcal{N}(\mu_0, \sigma_0)$), e a hipótese alternativa $\mathcal{H}_1$ é X_t não pertencer a essa distribuição. Neste caso, é fácil obter os limites para o teste:

$$|X_t - \mu_0| \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \kappa\sigma_0, \tag{1}$$

onde κ é o número de desvio padrões aceitáveis [Montgomery 2013]. Esse teste considera amostras unitárias e, embora muito simples, em alguns casos pode possuir desempenho

melhor do que outros métodos clássicos [Moustakides 2014, Vasantam et al. 2021].

Amostras de tamanho $w > 1$: Consideremos agora uma janela de tamanho $w > 1$. O problema é determinar se as amostras da janela pertencem ou não a uma distribuição com média conhecida. Esse problema pode também ser formulado como o teste de hipótese $\mathcal{H}_0 : \mu = \mu_0$ e $\mathcal{H}_1 : \mu \neq \mu_0$. Neste caso, o teste estatístico é o *log-likelihood ratio* (LLR) *test*. É fácil mostrar que o LLR *test* é equivalente ao *t-test*.

EWMA [Roberts 1959]: O método média móvel exponencialmente ponderada (EWMA) (também chamado de média móvel geométrica), utiliza a estatística

$$z_t = \lambda X_t + (1 - \lambda) z_{t-1}, \tag{2}$$

onde $0 < \lambda < 1$ é um hiperparâmetro que pondera as observações passadas. Quanto maior λ, mais importância é dada às observações recentes em relação às amostras passadas. Para $X_n \sim \mathcal{N}(\mu_0, \sigma^2)$, a variância de Z_n é dada por [Montgomery 2013]

$$\sigma^2_{Z_t} = \sigma^2 \left(\frac{\lambda}{2 - \lambda} \right) \left[1 - (1 - \lambda)^{2t} \right], \tag{3}$$

$\sigma^2_{Z_t}$ é usada para definir os limites de controle $\pm \kappa_d \sigma_{Z_t}$.

CUSUM: O algoritmo Cumulative Sum (CUSUM) é provavelmente o mais popular para detecção de mudanças [Page 1954]. Considere as amostras $(X_1, X_2, \ldots, X_\nu)$ e $(X_{\nu+1}, X_{\nu+2}, \ldots)$ e seja X_t uma amostra em t, e Z_t o *log-likelihood ratio* (LLR) entre a hipótese $\mathcal{H}_1$ de que X_t seja uma amostra cuja densidade é $f_1(x)$ e a hipótese $\mathcal{H}_0$ de que X_t tenha sido gerada por uma densidade $f_0(x)$. Então, as seguintes expressões são usadas para detectar mudanças:

$$Z_t = \log \frac{\mathcal{L}(X_t|\mathcal{H}_1)}{\mathcal{L}(X_t|\mathcal{H}_0)} \text{ e } \mathcal{S}_t = \max(0, \mathcal{S}_{t-1} + Z_t), \tag{4}$$

onde $\mathcal{L}(X_t|\mathcal{H}_v)$ é a verosimilhança de $\mathbf{X}_t$ considerando a hipótese $\mathcal{H}_v$. Note que a estatística $\mathcal{S}_t$ do CUSUM tem uma tendência de crescimento após um ponto de mudança, uma vez que passa a ser mais provável que, a partir deste instante, as amostras sejam obtidas da distribuição $f_1(x)$, pois $Z_t > 1$. Quando S_t ultrapassa um determinado limiar h, o ponto de mudança é detectado. Deste modo, no CUSUM, é feita uma comparação da soma cumulativa $\mathcal{S}_t$ com um limite. Esta é a principal diferença em relação ao método de Shewhart, que independe do histórico do processo.

Para o caso particular de mudança da média de um processo Gaussiano, o método é também chamado de *tabular* CUSUM ou two-sided CUSUM (2S-CUSUM) [Montgomery 2013] e [Tartakovsky et al. 2015].

WL-CUSUM: O CUSUM tradicional assume que os parâmetros do processo antes e depois de um ponto de mudança são conhecidos, por exemplo, a média e variância das densidades $f_0(x)$ e $f_1(x)$. Esta informação é importante para que se possa fazer uma boa estimativa do limite a ser usado para o teste de detecção de mudança. Entretanto, os parâmetros pós mudança podem não ser conhecidos, como no nosso caso onde nem mesmo os parâmetros de $f_0(x)$ são disponíveis.

Na literatura, em geral, os métodos usam uma janela W de amostras recentes para estimar os parâmetros de $f_1(x)$. Por exemplo, o método *Generalized likelihood*

ratio (GLR) otimiza os parâmetros da distribuição $f_1(x)$ dentro da janela e utiliza o LLR para determinar se houve mudança de parâmetros em relação aos de f_0 a cada janela [Basseville and Nikiforov 1993]. Por outro lado, o WL-CUSUM [Xie et al. 2023], emprega a estatística do CUSUM (equação (4)) e, como o GLR, usa uma estimativa para os parâmetros de $f_1(x)$ (por exemplo o MLE) a partir das amostras da janela. Portanto, acumulando evidências a favor ou contra uma mudança dentro de uma janela com base na observação atual. Tanto para o método GLR quanto para o WL-CUSUM, é um desafio estimar o tamanho da janela e também o valor do limite a ser usado.

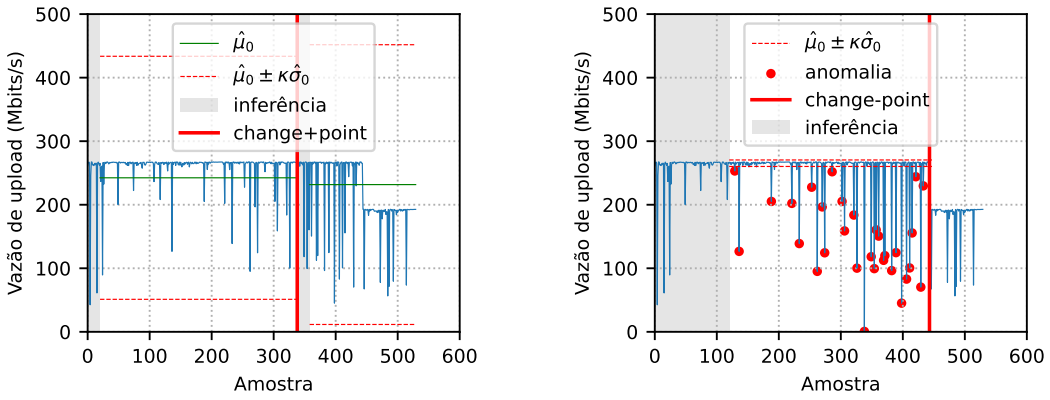
4. Propostas de modificações

Nesta seção apresentamos propostas para a implementação dos algoritmos Shewhart, EWMA e CUSUM para que possam ser utilizados de forma eficaz nas séries que analisamos.

4.1. Estimativa de parâmetros

Para a implementação dos métodos clássicos, uma prática usual é estimar os parâmetros de $f_0(x)$ através da estimativa de máxima verossimilhança (MLE). Porém, se os dados de treinamento apresentarem *outliers*, ou se o processo estiver em um transiente no momento da amostragem, as estimativas podem não refletir o estado normal do processo. Isto é ilustrado no exemplo da fig. 1a, onde método de Shewhart não foi capaz de detectar um ponto de mudança próximo a $n = 500$.

Inspirado no trabalho de [Farkas 2015], mas tendo como objetivo uma alternativa mais simples e adequada para métodos *online*, propomos os passos que se seguem para estimar os parâmetros de $f_0(x)$. Para cada amostra são verificadas duas condições para a estimativa dos parâmetros de $f_0(x)$: se a hipótese de normalidade do teste de Shapiro-Wilker não é rejeitada e se a variância da amostra não aumentou acima de um valor limite percentual. Se uma das condições não for verdadeira, o teste é repetido nas janelas subsequentes, até um determinado limiar. Após um determinado limiar, é aplicado um filtro percentil para remover possíveis *outliers*, e em seguida os parâmetros de $f_0(x)$ são estimados. Quando os critérios de normalidade e variância necessitam de um número de passos (janelas) maior do que um para serem atingidos, ou quando o número de passos



(a) Método básico **(b) Método proposto**

Figura 1. Método de Shewhart - Cliente 5, servidor NDT gig03

excede o limiar definido realizamos ainda o teste de estacionariedade *Augmented Dick Fuller*. Caso o processo tenha feito uma transição de não estacionário para estacionário, declaramos um ponto de mudança.

A fig. 1 apresenta um exemplo da detecção de pontos de mudança em uma série de vazão de *upload* com o método de Shewhart clássico e a implementação proposta. Na figura, o período em cinza (inferência), refere-se as janelas necessárias para a estimativa dos parâmetros de $f_0(x)$. As linhas horizontais vermelhas da fig. 1a indicam um variabilidade alta (estimada) em torno da média devido ao conjunto de amostras usado para a estimativa dos parâmetros de $f_0(x)$, o que implica que o ponto de mudança próximo a $n = 450$ não seja detectado utilizando o método básico. Já na fig. 1b, com a implementação proposta, a variabilidade em torno da média é bem menor, e portanto o método foi capaz de identificar o ponto de mudança próximo a $n = 450$.

4.2. Diferenciando anomalias de pontos de mudanças

Quando ocorre uma anomalia, também ocorrem mudanças abruptas nas propriedades dos dados, ainda que transitórias. De modo geral, os métodos sequenciais de *change-point* não são capazes de diferenciar esses de uma anomalia.

De acordo com [Aggarwal 2017], um *outlier* pode ser classificado como uma anomalia caso exceda um determinado número de desvios padrão: $X_n = \mu_0 \pm \kappa_a \sigma_0$. Assim, para diferenciar anomalias de *change-points*, propomos o seguinte esquema simples para os métodos Shewhart, EWMA e CUSUM: quando a estatística do teste se desvia do limite de controle durante um determinado número de observações consecutivas, um ponto de mudança é confirmado. Caso contrário, o ponto é classificado como uma anomalia.A fig. 1b ilustra o método proposto. Para ilustrar, na fig. 1a, o método básico identificou o *outlier* próximo a $n = 350$ como um ponto de mudança, enquanto que o método proposto (fig. 1b) foi capaz de classificar este *outlier* como uma anomalia (segundo a definição de anomalia em [Aggarwal 2017]), e não como um ponto de mudança.

Os métodos CUSUM e EWMA possuem memória, ou seja, consideram amostras atuais e passadas. Desta forma, para que a proposta possa ser aplicada também a esses métodos, reinicializamos a estatística do método ao último valor antes de ocorrer o desvio.

5. *Voting Windows Change-point Detection* (VWCD)

Nesta seção apresentamos um novo método em estudo por nosso grupo: o VWCD. O método foi idealizado para ser utilizado de forma *online*, não supõe o conhecimento prévio de nenhum parâmetro (seja anteriormente ou após um instante de mudança) e, principalmente, para que os resultados sejam de fácil interpretação para a tomada de decisão e, desta forma, facilitar ajustes de acordo com o problema estudado.

O VWCD, emprega uma janela deslizante, assim como outros métodos *online*. Usa também noções de *ensemble*, comum em algoritmos de aprendizado de máquina. No nosso algoritmo, cada janela (com tamanho W) escolhe um dentre os W instantes dentro da janela como um possível ponto de mudança e "**vota**" neste ponto. A esse voto é atribuída uma probabilidade que representa a confiança que a janela possui na escolha do ponto. Essa probabilidade é proveniente de uma abordagem Bayesiana. Deve-se notar que, como a janela desliza a cada amostra, um instante de tempo pode receber um voto

de cada uma das W janelas que visitam o instante ao deslizar. Esses votos são utilizados na decisão final para determinar se o instante constitui ou não um ponto de mudança. Variantes do algoritmo incluem a maneira de se computar o resultado final da votação dos W votos recebidos, assim como ajustes que podem ser aplicados dentro da janela, por exemplo, para dar pesos diferentes a votos com maior "confiança" ou ainda para evitar a estimativa de parâmetros com poucas amostras. No que se segue, por limitações de espaço, faremos apenas uma breve descrição da versão do algoritmo utilizada.

Seja X_t a amostra no instante t da série em observação, e W_t a janela de observação em t e que inclui o conjunto de pontos $\{X_{t-W_t+1}, \ldots X_t\}$. Suponha que um ponto de mudança ocorre no instante τ ($t - W_t + 1 \leq \tau \leq t$) e seja $f_0^{(t)}$ ($f_1^{(t)}$) a função de densidade antes (respectivamente depois) de τ. As amostras anteriores e posteriores a τ são usadas para parametrizar as funções $f_0^{(t)}$ e $f_1^{(t)}$ usando MLE. É importante observar que existe uma série de funções parametrizáveis cujo cálculo usando MLE é uma fórmula simples em função das amostras. Esse é o caso, por exemplo de uma Gaussiana. Além do mais, apesar de nossas séries serem univariadas, a mesma abordagem pode ser usada para séries multivariadas. Seja $\mathcal{L}^{(t)}(\tau)$ a verossimilhança do modelo da janela W_t onde as amostras anteriores (posteriores) a τ seguem densidade $f_0^{(t)}$ ($f_1^{(t)}$). Então:

$$\mathcal{L}^{(t)}(\tau) = \arg\max_{\mathbf{\Theta}} \{p_t(\mathcal{D}|\tau, \mathbf{\Theta})\} = \arg\max_{\mathbf{\Theta}} \prod_{i=t-W_t+1}^{\tau} f_0^{(t)}(X_i) \prod_{i=\tau+1}^{t} f_1^{(t)}(X_i)$$

onde $\mathbf{\Theta}$ é um possível conjunto de valores para os parâmetros do modelo usado ($f_0^{(t)}$ e $f_1^{(t)}$) e $p_t(\mathcal{D}|\tau, \mathbf{\Theta})$ é a densidade de probabilidade dos dados $\mathcal{D}$ da série da janela W_t condicionada a τ e a $\mathbf{\Theta}$. (No caso de uma Gaussiana, por exemplo, é trivial obter os parâmetros através de uma fórmula simples.) Seja τ_t^* o instante onde $\mathcal{L}^{(t)}(\tau)$ é máximo.

$$\tau_t^* = \arg\max_{\tau} \mathcal{L}^{(t)}(\tau). \tag{5}$$

τ_t^* é o valor potencialmente escolhido pela janela W_t como ponto de mudança da janela, e $\mathbf{\Theta}^*$ os parâmetros correspondentes. Resta então calcular a confiança que a janela tem na escolha, o que pode ser feito usando Bayes.

A *posterior* para τ_t^* (equação (5)), i.e., a estimativa MAP (*maximum posterior*) é simplesmente a *likelihood* vezes a *prior* para τ_t^* da janela W_t, normalizada (Bayes). Omitindo $\mathbf{\Theta}^*$ da notação:

$$p_t(\tau_t^*|\mathcal{D}) = \frac{p_t(\mathcal{D}|\tau_t^*)p_t(\tau_t^*)}{\sum_{\tau=t-W_t+1}^{t} p_t(\mathcal{D}|\tau)p_t(\tau)} \tag{6}$$

É natural supor uma *prior* uniforme para τ_t^*. Desta forma, (6) é apenas uma normalização. Entretanto, podemos dar menor peso a escolha de pontos de mudança nos extremos do intervalo da janela W_t evitando o cálculo de $\mathbf{\Theta}$ com poucos pontos. Para isso usamos uma distribuição beta-binomial.

Como indicado acima, cada instante t recebe entre zero e W votos, com as respectivas probabilidades dadas por (6). O objetivo então é a tomada de decisão, baseando-se no voto individual com pesos. Esse é um problema geral tratado em diversas áreas

(e.g., [Nordmann and Pham 1999, Liu et al. 2021]), e há várias maneiras de lidar com o problema em um sistema desse tipo. Esta é uma vantagem do método, pois é fácil interpretar o resultado de uma "eleição" pelas probabilidades atribuídas aos instantes τ_t^*. Nesse artigo, nosso objetivo não é o de explorar esse tópico, mas ilustrar uma dentre várias possibilidade da proposta no contexto estudado. Por exemplo, podemos atribuir um limiar para a probabilidade (confiança) do voto de uma janela, de forma a contabilizar apenas votos com alta confiança. Além disso, é também possível usar uma função logística ($y(x) = \frac{1}{1+\mathrm{e}^{-(x-\mu/s)}}$) que nos permite dar maior ênfase aos votos dados com maior confiança, i.e., com maior probabilidade, e ainda combinar as estratégias.

6. Descrição do experimento e resultados

Com o objetivo de exemplificar os métodos em estudo, usamos um conjunto de dados reais de rede por nós coletados. A métricas foram obtidas da ferramenta NDT/M-Lab a partir de um conjunto de 6 raspberry-pi's (os clientes) conectados ao roteador residencial de voluntários em diferentes bairros da cidade do Rio e clientes de diferentes provedores de acesso. Os dados desse breve estudo são a vazão e latência de *download* e *upload*.

As amostras foram coletadas automaticamente a intervalos com média de 30 minutos e distribuição exponencial (portanto, obtendo em média 48 amostras diárias de cada métrica). Os testes foram realizados no período de seis meses, entre 01/06/2023 a 30/11/2023, para 9 servidores: gig01, gig02, gig03, gig04, rnp-rj (localizados no RJ); e gru02, gru03, gru04 e rnp-sp em SP. (Note que o NDT pode redirecionar um teste automaticamente para outro servidor dependendo de critérios de balanceamento de carga, etc.) Cada série temporal é composta por amostras de uma das quatro métricas para um par cliente-servidor durante os seis meses, perfazendo um total de 212 séries do total de possíveis $6 \times 4 \times 9 = 216$, pois descartamos séries com menos de 100 medições. A Fig. 2 ilustra as séries temporais de vazão de *upload* do Cliente 5.

Em todos algoritmos, utilizamos janelas com tamanho de 20 amostras e, para os métodos clássicos, empregamos parâmetros citados como usuais [Montgomery 2013]. Disponibilizamos em https://github.com/cleitonmoya/sbrc24 o código-fonte, a base de dados e instruções para quem desejar melhor avaliar ou reproduzir nosso experimento.

É importante ressaltar que não dispomos de rótulos que identifiquem os pontos de mudança, o que impossibilita calcular taxas de acerto e falso positivo dos métodos para identificar pontos de mudança e anomalias, ou mesmo fazer uma comparação entre os métodos. Conforme indicado na Seção 1, nosso objetivo é verificar o comportamento dos métodos em estudo quando aplicados a séries temporais reais obtidas pelo NDT, e fazer uma breve comparação dos instantes identificados no período como pontos de mudanças ou anomalias. Mostramos também que, mesmo sem rótulos, é possível obter alguns *insights* quanto à qualidade de acesso.

O número total de *change-points* identificados por cada método é mostrado na fig. 3. As implementações básicas dos métodos clássicos (Shewhart, CUSUM e EWMA) não distinguem *change-points* de anomalias, portanto elas identificaram um número significativamente maior de mudanças em comparação com as implementação propostas. Na fig. 4a, as modificações dos métodos clássicos e o novo método proposto VWCD foram capazes de identificar uma clara mudança de comportamento da série temporal entre $n = 800$ e 1000: a média e a variância aumentaram significativamente a partir deste

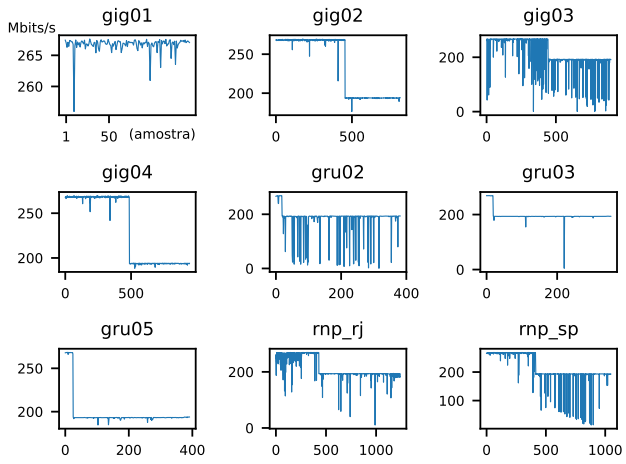


Figura 2. Vazão de *upload* do Cliente 5 para cada servidor NDT

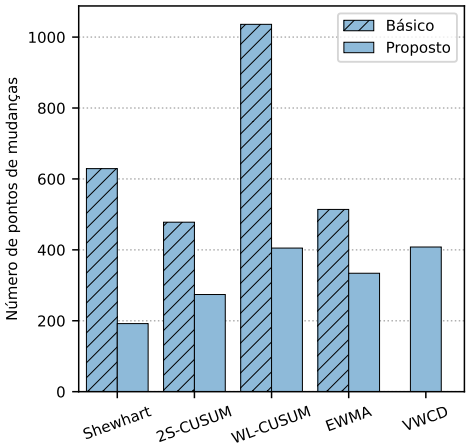


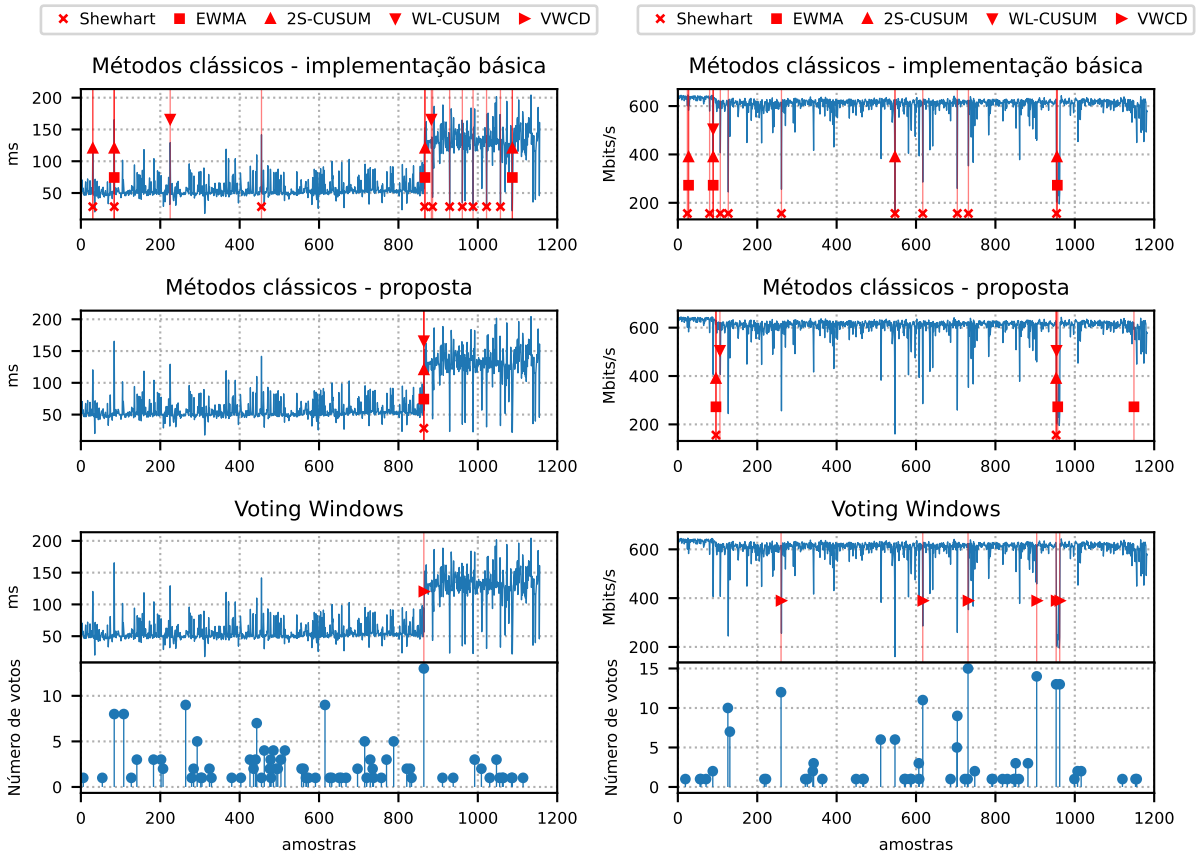
Figura 3. Número de pontos de mudança identificados por método

ponto. As implementações básicas foram capazes de identificar a mudança mas, por outro lado, marcaram pontos que podem ser apenas resultado da alta variabilidade, e portanto não seria adequado ”alarmar“ frequentemente caso esses métodos fossem empregados. Conclusões semelhantes podem ser tiradas da fig. 4b em relação aos métodos clássicos e as modificações propostas. O VWCD, nesta figura, não “alarmou” em torno da amostra 100, como os outros métodos. Por outro lado, esse ponto apresentou apenas uma pequena queda média da vazão ($\approx 5\%$) e alguns *outliers* que foram detectados pelos outros métodos. Além disso, detectou a maior variabilidade em torno da amostra 600. Note que seria possível utilizar mais de um método diferente para aumentar a confiança nos resultados.

As figuras 4a e 4b mostram ainda o número de votos para cada instante para o método VWCD, ilustrando como podemos usar esta informação extra para a tomada de decisão. Por exemplo, na fig. 4b, dois instantes próximos a amostra 100 receberam mais de 7 votos com alta confiança e que poderiam também ser considerados como um *change-point*. Portanto, o método pode ser adaptado de acordo com os objetivos da aplicação.

A fig. 5 mostra o *boxplot* do número de anomalias identificadas por todos os métodos clássicos usando as implementações propostas. Neste gráfico, as anomalias estão agrupadas apenas por clientes e tipo de métrica, e não pelos métodos. Os *boxplots* em laranja (azul) representam o número de anomalias que resultaram em uma diminuição (aumento) na média da métrica analisada. O objetivo é avaliar estatísticas básicas (mediana, quartis, mín., máx. e *outliers*) do número destas mudanças independente do método. Na figura, é evidente que a vazão de *upload* do Cliente 5, mostrou um número significativamente maior de instantes onde a vazão diminuiu em comparação aos outros clientes. Isso sugere uma QoS relativamente pior para o Cliente 5. De fato, verifica-se na fig. 2 instabilidade no *upload* para os servidores gig03, gru02, rnp-rj e rnp-sp.

Pode-se facilmente identificar os parâmetros dos modelos dos métodos após uma mudança e, em tempo real, verificar se houve uma piora nas métricas de QoS. A fig. 6, ilustra a utilidade de se usar os parâmetros de cada um dos modelos como forma de avaliar



(a) Cliente 3, rnp-sp, latência de *upload*

(b) Cliente 6, rnp-rj, vazão de *download*

Figura 4. Exemplos de aplicação dos métodos

a QoS após uma mudança. A figura mostra no eixo y o número de pontos de mudança que acarretaram piora na vazão de *upload* e no eixo x a magnitude da diminuição (em megabits por segundo). O objetivo é avaliar quais clientes tiveram maior degradação da qualidade de rede, pois quanto maior o número de mudanças com maior queda de magnitude, maior o impacto na QoS. Para gerar este gráfico, utilizamos todas as séries temporais de vazão de *upload* disponíveis para cada cliente (todas as séries possuem o mesmo período e cada série corresponde a um servidor NDT). Nota-se que, para o Cliente 5, todos os métodos identificaram ao menos 5 mudanças com diminuição superior a 50 Mbits/s, indicando uma potencial piora na qualidade da rede. A fig. 2 deixa claro que o Client 5 sofreu uma degradação da vazão de *upload* no período observado. Esse exemplo serve para mostrar a utilidade desse tipo de gráfico, que pode ser obtido em tempo real, **mesmo sem a existência de rótulos**.

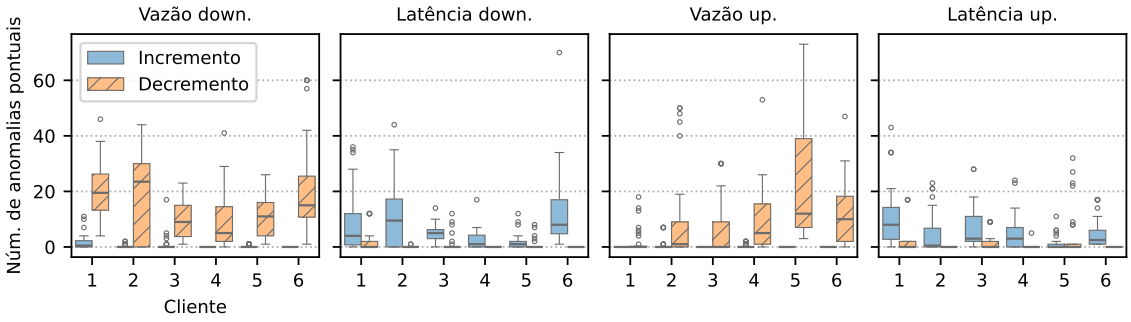


Figura 5. Número de anomalias (incremento/decremento na média da métrica) identificado pelos métodos clássicos propostos

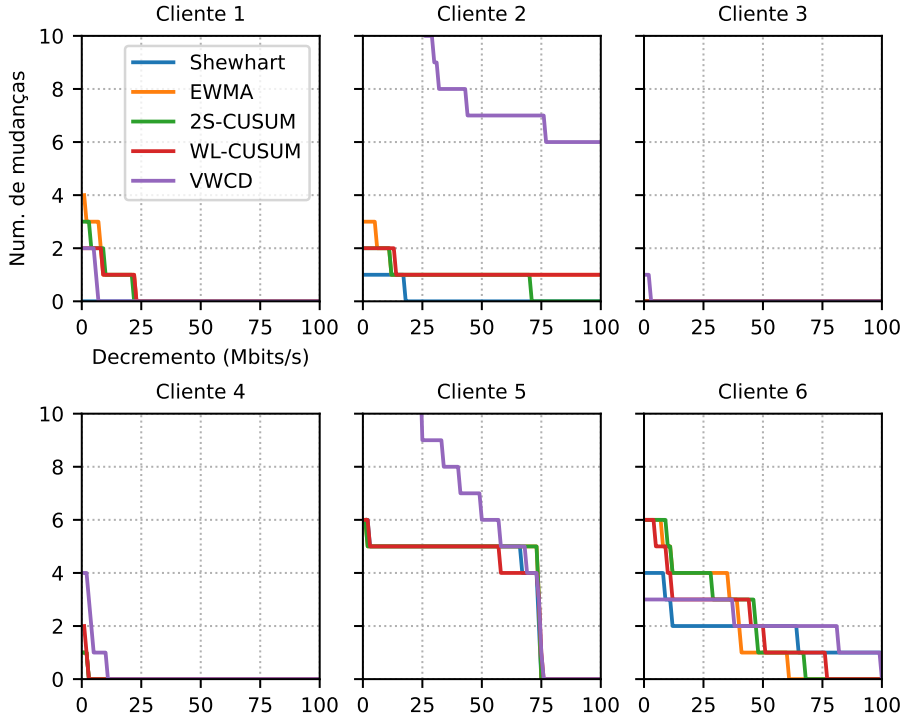


Figura 6. Número de pontos de mudança de decremento na vazão de *upload* em função da magnitude do decremento

7. Conclusões

Neste artigo exploramos métodos para detecção *online* de pontos de mudança e anomalias em séries temporais de latência e vazão de *download* e *upload*. Desenvolvemos estratégias simples que possibilitaram o uso adequado dos métodos clássicos Shewhart, EWMA e CUSUM de forma *online* e permitiram a distinção entre anomalias e pontos de mudanças. Também introduzimos um novo novo algoritmo baseado em um sistema de votação ponderada, o VWCD, que oferece vantagens quanto à flexibilidade e interpretabilidade para a tomada de decisão.

Para estudar os algoritmos, construímos um conjunto de dados com testes reais usando o *Network Diagnostic Tool* (NDT) realizados em redes de usuários residenciais e por seis meses. Como não dispomos de rótulos, não foi possível fazer uma análise comparativa dos métodos. Porém, avaliamos o comportamento dos métodos e mostramos as diferenças de cada um. Mostramos também como podemos utilizar os algoritmos propostos para identificar, de forma *online*, a degradação de QoS das redes dos usuários, mesmo sem a existência de rótulos. Esperamos que esse trabalho possa vir a contribuir para o monitoramento da qualidade de redes residenciais.

Atualmente dispomos de um conjunto de raspberry-pi's que estão sendo instalados na rede de um provedor parceiro. Com a entrada eminente de servidores da RNP como servidores do M-Lab, esperamos que a coleta de dados seja significativamente ampliada e que possamos obter rótulos do provedor parceiro para o *ajuste-fino* dos métodos. Com mais dados, e através de algoritmos como os estudados neste trabalho, será possível ter uma visão ampla da qualidade de redes residenciais no Brasil. Ressaltamos que os métodos aqui estudados são gerais e poderiam ser usados em outras séries temporais.

Referências

Aggarwal, C. C. (2017). *Outlier Analysis*. Springer, New York, 2 edition.

Aminikhanghahi, S. and Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367.

Basseville, M. and Nikiforov, I. V. (1993). *Detection of abrupt changes: theory and application*. Prentice Hall.

Braei, M. and Wagner, S. (2020). Anomaly detection in univariate time-series: A survey on the state-of-the-art. *arXiv preprint arXiv:2004.00433*.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58.

Farkas, K. (2015). Cusum anomaly detection. `https://www.measurementlab.net/publications/CUSUMAnomalyDetection.pdf`. [Online; accessed 12-January-2024].

Li, J., Fearnhead, P., Fryzlewicz, P., and Wang, T. (2024). Automatic Change-Point Detection in Time Series via Deep Learning. *Journal of the Royal Statistical Society Series B: Statistical Methodology*.

Liu, Z., Zhang, Z., and Liu, Y. (2021). Power grid security risk assessment method based on weighted voting ensemble machine learning algorithm. In *2021 6th International Conference on Power and Renewable Energy (ICPRE)*, pages 607–613.

M-Lab (2024). NDT - network diagnostic tool.

Montgomery, D. C. (2013). *Introduction to Statistical Quality Control*. Wiley, New York, 7 edition.

Moustakides, G. V. (2014). Multiple optimality properties of the shewhart test. *Sequential Analysis*, 33(3):318–344.

Nordmann, L. and Pham, H. (1999). Weighted voting systems. *IEEE Transactions on Reliability*, 48(1):42–49.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

Roberts, S. (1959). Control chart tests based on geometric moving averages. *Technometrics*, 1(3):239–250.

Santos, G., Mendonça, G., Leão, R., and e Silva, E. S. (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*, pages 98–111, Porto Alegre, RS, Brasil. SBC.

Schmidl, S., Wenig, P., and Papenbrock, T. (2022). Anomaly detection in time series: a comprehensive evaluation. *Proceedings of the VLDB Endowment*, 15(9):1779–1797.

Shewhart, W. A. (1929). Control of quality of manufactured product.

Streit, A., Santos, G. H., Leão, R. M., e Silva, E. d. S., Menasché, D., and Towsley, D. (2021). Network anomaly detection based on tensor decomposition. *Computer Networks*, 200:108503.

Tartakovsky, A., Nikiforov, I., and Basseville, M. (2015). *Sequential analysis: Hypothesis testing and changepoint detection*. CRC press.

Tartakovsky, A. G., Polunchenko, A. S., and Sokolov, G. (2013). Efficient computer network anomaly detection by changepoint detection methods. *IEEE Journal of Selected Topics in Signal Processing*, 7(1):4–11.

Vasantam, T., Towsley, D., and Veeravalli, V. V. (2021). Quickest change detection in the presence of transient adversarial attacks. In *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE.

Xie, L., Moustakides, G. V., and Xie, Y. (2023). Window-limited cusum for sequential change detection. *IEEE Transactions on Information Theory*.

Ximenes, D., Mendonça, G., Santos, G. H., de Souza, E., Leão, R. M., Menasché, D. S., 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*. SBC.