

Correlacionando Dados de Monitoramento de Rede para Identificação de Causas de Problemas de Desempenho

Danielle S. Silva¹, Francisco V. J. Nobre¹, Maria C. Ferreira¹,
Thelmo P. Araújo¹, Rafael L. Gomes¹

¹Universidade Estadual do Ceará (UECE), Fortaleza, Ceará, Brasil.

{danielle.santos, valderlan.nobre, clara.mesquita,

{thelmo.araujo, rafa.lopes}@uece.br

Resumo. *Várias empresas e provedores de Internet (ISPs) possuem serviços de monitoramento de rede que abrangem avaliações regulares de desempenho, com foco principal na entrega de informações cruciais sobre a situação atual da infraestrutura de rede e, conseqüentemente, dos serviços que executam sobre ela. No entanto, estas ferramentas de monitoramento precisam de um desenvolvimento contínuo para incorporar tarefas mais complexas, como a detecção de problemas de desempenho. Dentro deste contexto, este artigo apresenta um mecanismo para identificar atrasos altos e os links de comunicação da rede que podem ser a causa desses problemas de desempenho, utilizando um Score de impacto formulado considerando aspectos temporais. Esse Score baseia-se em técnicas de correlação de dados aplicadas às informações coletadas por ferramentas de monitoramento. Os experimentos realizados com dados reais da RNP demonstram a eficácia do mecanismo proposto na identificação de links de rede que impactam a comunicação de dados, gerando os atrasos altos fim-a-fim.*

Abstract. *Several companies and Internet Service Providers (ISPs) offer network monitoring services that encompass regular performance evaluations, with a primary focus on delivering crucial information about the current state of the network infrastructure and, consequently, the services running on it. However, these monitoring tools require continuous development to incorporate more complex tasks, such as performance issue detection. Within this context, this article introduces a mechanism for identifying high delays and communication links in the network that may be the cause of these performance issues, using a temporally formulated Impact Score. This Score is based on data correlation techniques applied to information collected by monitoring tools. Experiments conducted with real data from the National Education and Research Network (RNP) demonstrate the effectiveness of the proposed mechanism in identifying network links impacting data communication, resulting in high end-to-end delays.*

1. Introdução

Na era moderna, é preciso que entidades como Provedores de Internet (ISPs), empresas e instituições adotem serviços de monitoramento de rede para coletar medições de desempenho e dados sobre o comportamento da infraestrutura de rede. Esta prática envolve a análise de diversos aspectos, como atraso, perda, vazão e outros [Gomes et al. 2020].

O atraso, que mede o tempo gasto de ponta a ponta na transmissão de bits pela rede, pode fornecer informações cruciais para tomadas de decisão na gestão de redes e serviços [Mok et al. 2021]. Essas ações de gestão incluem a expansão da infraestrutura de rede, o estabelecimento de expectativas de Qualidade de Serviço (QoS), a avaliação de Acordos de Nível de Serviço (SLA) e a análise da demanda por recursos de rede [Costa et al. 2021, Gomes et al. 2016].

Em relação a QoS, o atraso fim-a-fim possui um papel crucial na avaliação da eficiência dos serviços de rede que estão em operação [Gomes et al. 2017]. A ocorrência de casos de atraso alto pode resultar em uma taxa de transmissão de bits reduzida, congestionamento na rede e possíveis falhas na prestação de serviço, impactando negativamente a experiência do usuário e diminuindo a produtividade tanto da rede quanto dos clientes [Portela et al. 2024]. Por outro lado, a minimização dos atrasos contribui para aprimorar a experiência do usuário, resultando em maior produtividade e desempenho global da infraestrutura de rede. Portanto, é imprescindível monitorar e otimizar o desempenho da rede a fim de assegurar que as métricas de QoS dos SLAs sejam cumpridas [Moreira et al. 2021, Portela et al. 2023].

Ao monitorar o atraso fim-a-fim entre os nós da rede, torna-se possível manter um ambiente de rede adequado e lidar com as variações na demanda por acesso a Internet, especialmente causadas pelo comportamento elástico que influencia significativamente o desempenho do atraso. Este comportamento elástico manifesta-se quando os usuários acessam a Internet de maneiras distintas, dependendo de sua localização, resultando em flutuações na demanda por recursos de rede ao longo do dia [Silveira et al. 2023]. A negligência na gestão apropriada de cenários elásticos pode ocasionar interrupções de serviço, lentidão e desconexões, eventualmente resultando em violações dos requisitos do SLA firmado. Para manter métricas de QoS, é crucial adotar abordagens adaptativas que estejam alinhadas ao comportamento elástico, garantindo assim que o sistema opere de forma adequada. Apesar da importância das medições de rede, as ferramentas de monitoramento existentes ainda carecem de evolução em diversos aspectos, principalmente no que se refere a capacidade de identificar a ocorrência de casos de atraso alto em links de comunicação dentro da infraestrutura de rede [da Silva et al. 2020].

Assim, a identificação precoce desses atrasos nos links de comunicação resulta em benefícios aos ISPs, conforme destacado por estudos recentes [Scarpitta et al. 2023, BinSahaq et al. 2022]: (I) Planejamento de Capacidade: A identificação antecipada possibilita que as organizações antecipem as demandas futuras e aloquem recursos de rede de maneira apropriada, prevenindo congestionamentos na rede. (II) Monitoramento de Qualidade: A identificação contínua permite que os ISPs meçam e rastreiem o desempenho da rede ao longo do tempo, assegurando a conformidade com os SLAs. (III) Alocação de Recursos: A identificação eficiente permite que as organizações destinem mais recursos aos links críticos, reduzindo despesas desnecessárias em links subutilizados (que pode gerar um consumo de energia e custo de manutenção). (IV) Otimização de Desempenho: A identificação proporciona uma base para configurações avançadas de gerenciamento de rede, incluindo decisões de roteamento, balanceamento de carga e alocação de recursos, evitando a presença de links com atrasos elevados.

Dentro desse contexto, este artigo apresenta um mecanismo para identificar casos de atraso alto em links de comunicação na infraestrutura de rede. A proposta modela uma

Pontuação (*Score*) de Impacto, que é utilizada sobre dados provenientes de ferramentas de monitoramento de rede. O *Score* de Impacto aplica técnicas de correlação de dados, que quantificam e analisam a relação estatística ou associação entre duas ou mais variáveis nas medições de rede, auxiliando na descoberta de padrões, dependências e indicações nos dados analisados. Na solução foram usados os dados de atraso em medições de Ping e o conjunto de links usados pelas medições através da ferramenta traceroute (IPs/Equipamentos que fazem parte do caminho usado). Estes dois dados possibilitam mapear os caminhos usados pelas diversas medições e correlacionar os links da infraestrutura de rede com os casos de atraso alto e atraso baixo, habilitando a identificação de possíveis causas.

A fim de validar a solução proposta e analisar sua eficiência em um cenário real, foram realizados experimentos usando dados reais do Serviço de monitoramento da rede Ipê (Monipê)¹ da Rede Nacional de Ensino e Pesquisa (RNP). Os resultados obtidos mostram que o método proposto é capaz de indicar os links de rede envolvidos em casos de atrasos fim-a-fim significativos.

O restante deste artigo está organizado da seguinte forma: A Seção 2 apresenta os trabalhos relacionados existentes. A Seção 3 apresenta a proposta deste artigo, enquanto a Seção 4 descreve os resultados dos experimentos realizados. Por fim, a Seção 5 conclui o artigo e apresenta trabalhos futuros.

2. Trabalhos Relacionados

Esta seção descreve os principais trabalhos relacionados, e recentemente publicados pela comunidade científica, sobre correlação de dados no contexto de redes de computadores, incluindo aspectos de desempenho e qualidade de serviço.

Scarpitta et al. [Scarpitta et al. 2023] propõem uma solução para Software-Defined Wide Area Networks (SD-WANs) que utiliza o Simple Two-way Active Measurement Protocol (STAMP) para monitorar o atraso de um caminho Segment Routing over IPv6 (SRv6) entre dois nós. Esta ferramenta de monitoramento provê insumos para soluções de gerenciamento de redes que usam dados de medição para dar suporte aos administradores de rede.

Gajewski et al. [Gajewski et al. 2022] apresentam uma estratégia para detecção de anomalias que fornece responsabilidade compartilhada entre um cliente de serviço e o provedor de rede no contexto de uma Rede residencial (*Home Area Network* - HAN). Os autores usam uma abordagem de aprendizado de máquina para classificar os dados de monitoramento e correlação na busca de comportamentos suspeitos nos recursos da rede. Desta forma, as técnicas de correlação são aplicadas sobre características dos fluxos de rede, a fim de agrupar anomalias similares. Este contexto difere da proposta deste artigo, visto que neste trabalho a correlação ocorre entre fontes de dados de monitoramento distintos e com uma análise dentro de um intervalo de tempo.

Arachchige et al. [Arachchige et al. 2023] realizaram um estudo, usando técnicas de correlação de dados, sobre as temperaturas atingidas em dispositivos usados para processar algoritmos blockchain e o consumo de energia de três algoritmos blockchain comumente usados, executados em microcontroladores de baixa potência no contexto de redes de sensores sem fio. Este estudo foi limitado apenas aos dados de temperatura do

¹monipe-central.rnp.br

dispositivos que executam as ações no blockchain e consumo de energia destes dispositivos. Assim, apesar de aplicar técnicas de correlação, o contexto do trabalho é diferente da proposta deste artigo, mas mostra a importância da correlação de dados no processo de identificação de eventos e suporte a tomada de decisão.

Imran et al. [Imran et al. 2023] investigam o tempo para detectar um ataque em redes IoT baseadas no protocolo Transporte de telemetria de Enfileiramento de Mensagens (Message Queuing Telemetry Transport - MQTT). Os autores usam abordagens de aprendizado de máquina e propõe uma abordagem que aplica análise de correlação para reduzir o tempo de treinamento e teste desses algoritmos. Os autores concluem que a análise de correlação é significativamente benéfica no processo de engenharia de recursos, principalmente para determinar os recursos mais relevantes no conjunto de dados MQTT. Similarmente, Wang et al. [Wang et al. 2024] aplicam uma variação de uma Rede Neural Recorrente (*Recurrent Neural Network* - RNN) baseado em correlação de dados para determinar o tamanho do reservatório que corresponde a uma determinada tarefa de treinamento. A proposta dos autores utiliza a teoria da probabilidade e a teoria da informação para medir a correlação entre os neurônios da RNN, removendo dinamicamente os neurônios com baixa correlação e construindo uma matriz de características mais concisa. A proposta em questão foca em otimizar o treinamento e uso de variações de rede neurais, sem um contexto específico em si.

A partir do levantamento bibliográfico realizado, nota-se que nenhum artigo da literatura se concentrou no desenvolvimento de um mecanismo de identificação de causas de problema de desempenho de rede que utilize dados de monitoramento de rede, que é o foco deste artigo. Nossa proposta realiza ações estatísticas para aprimorar o processo de detecção, criando um *Score* que facilita o processo de gerência de desempenho de rede.

3. Proposta

O mecanismo proposto para identificar causas de atrasos altos em infraestruturas de rede requer duas informações para realizar sua tarefa: atraso fim-a-fim entre os pontos de medição e um conjunto de links usados na comunicação fim-a-fim (ou seja, o resultado da ferramenta traceroute que rastreia a rota que os pacotes de dados seguem ao viajar de um computador/dispositivo para outro). Em geral, o mecanismo proposto executa as seguintes atividades: Coleta de Dados, Separação dos dados de Traceroute e Atraso, Identificação de casos de Atraso Altos e Baixos, Geração da Matriz de Correlação, Cálculo do *Score* de Impacto (em horário único e por período de tempo) e no final faz uma comparação entre os dados para descobrir os links com maior probabilidade de problemas de desempenho. Uma visão geral da solução é apresentada na Figura 1 e na notação das formulações usadas ao decorrer do artigo são resumidas na Tabela 1.

Inicialmente, os dados sobre a infraestrutura de rede são inseridos no método, onde esse processo de coleta de dados é realizado a cada minuto, obtendo todos os dados sobre o atraso fim-a-fim (representado por M) e o traceroute da comunicação entre os nós (links das rotas utilizadas na comunicação, que foram o conjunto N). Ferramentas de medição e monitoramento de rede geram esses dados para auxiliar no diagnóstico e na solução de problemas de desempenho de rede.

Após receber o conjunto de dados, o mecanismo busca por casos de medição com atraso alto (ou seja, atrasos fim-a-fim superiores a $60ms$ que formam o conjunto

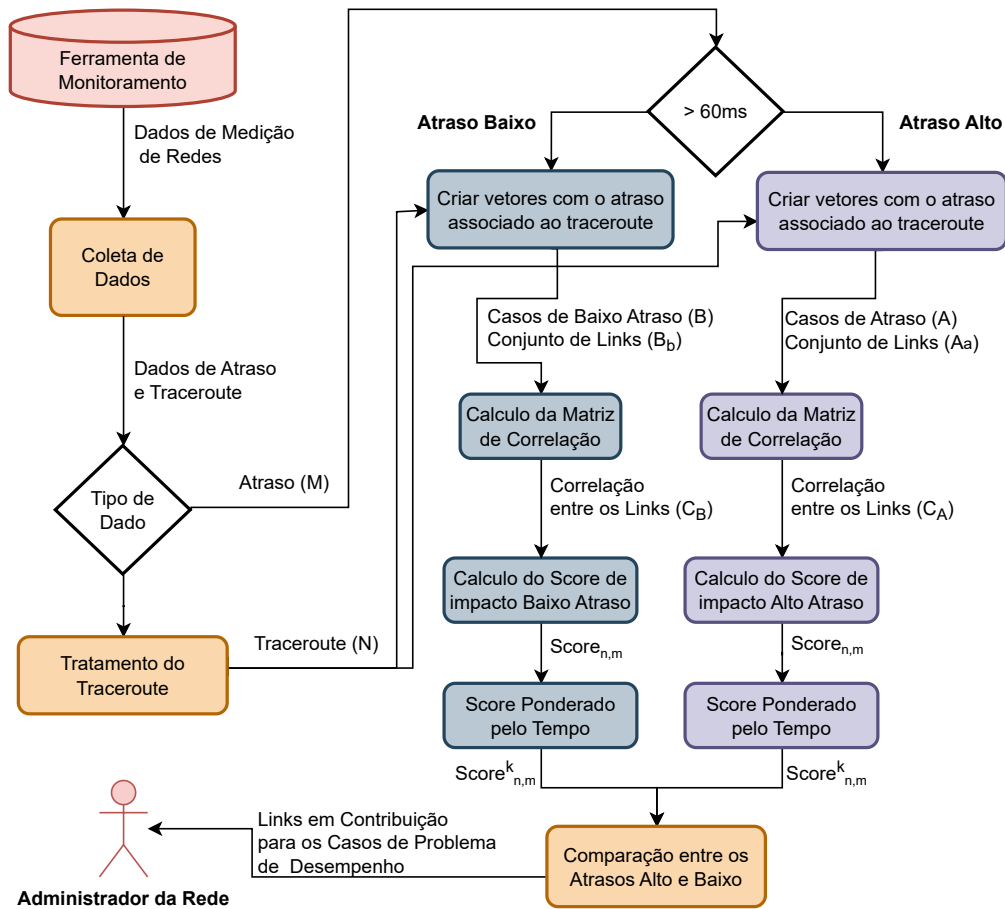


Figura 1. Visão geral da solução.

Tabela 1. Notação

| Símbolo | Descrição |
|------------------------|------------------------------------------------------------------------|
| M | Dados de medição de atraso. |
| N | Conjunto de links de comunicação usados na comunicação. |
| A | Casos de Alto Atraso. |
| A_a | Links de comunicação que fazem parte do caso de Alto Atraso a . |
| B | Casos de Baixo Atraso. |
| B_b | Links de comunicação que fazem parte dos caso de Baixo Atraso b . |
| C_A | Matriz de Correlação dos Casos de Alto Atraso. |
| C_B | Matriz de Correlação dos Casos de Baixo Atraso. |
| $C_{i,j}$ ou $r(i, j)$ | Coefficiente de Correlação de Pearson entre i e j . |
| $Score^k_{n,m}$ | Pontuação de Impacto do link n como causa de Caso m no tempo k . |

A, pois esse é um limite adequado para os serviços dos ISPs [Scarpitta et al. 2023, BinSahaq et al. 2022]) e os demais casos são considerados como atraso baixo (que fazem parte do conjunto B). Em seguida, a partir dos dados tratados do traceroute (filtragem dos links), o método seleciona os links que fazem parte dos caminhos desses casos de atraso alto e baixo. Assim, com essas etapas, tem-se os dados que relacionam os casos de atraso

alto e baixo, bem como os links relacionados a cada um desses casos.

Após a organização de casos de atraso e seus respectivos links, é possível calcular a matriz de correlação desses dois tipos caso, onde é definida uma matriz de correlação C_A para os Casos de Alto Atraso e uma matriz de correlação C_B para os Casos de Baixo Atraso. Nota-se que cada uma dessas matrizes terá a dimensão de acordo com a quantidade de links que faz parte dos seus respectivos casos, ou seja, faz-se a união dos links ($\cup_a A_a$ e $\cup_b B_b$). A partir da matriz de correlação, calcula-se o Índice de Impacto de um link, chamado de $Score_{n,m}^k$, representando o *Score* de Impacto do link n como causa de um caso m no tempo k . Inicialmente apenas com os dados da matriz gerada, e posteriormente é calculado dando um peso de acordo com o período de tempo. Estas etapas são detalhadas nas próximas subseções.

Por fim, é válido ressaltar que esta abordagem de analisar o impacto dos links de caso (Atraso Alto e Baixo) possibilita a solução verificar se um determinado link tem uma impacto alto em casos de alto atraso e um baixo impacto em casos de baixo atraso, representando que este link possui grandes indícios de ser uma das causas do baixo desempenho da rede em um determinado momento.

3.1. Correlação de dados

Técnicas de correlação de dados envolvem a identificação e quantificação de relações, dependências ou associações entre diferentes variáveis ou conjuntos de dados [Li et al. 2021]. Essas técnicas são comumente utilizadas em diversas áreas, incluindo estatísticas, análise de dados, aprendizado de máquina e ciência de dados. O valor de correlação aumenta à medida que a relação entre as variáveis se torna mais forte. É importante lembrar que correlação não implica causalidade, ou seja, apenas porque duas variáveis estão correlacionadas, não significa que uma causa a outra. A correlação de dados é útil para identificar relações entre variáveis na análise de dados, mas é necessário considerar outras informações para estabelecer causalidade [Gottwalt et al. 2019].

Uma das técnicas mais populares é o Coeficiente de Correlação de Pearson, frequentemente denotado como r . É uma medida estatística que quantifica a força e direção da relação linear entre duas variáveis contínuas. Avalia quão bem a relação entre duas variáveis pode ser descrita por uma linha reta. A correlação de Pearson é calculada usando a Equação 1. Na equação, x_i e y_i são pontos de dados das duas variáveis sendo correlacionadas, enquanto \bar{x} e \bar{y} são os valores médios (média) das duas variáveis. Posteriormente, é gerada uma matriz de correlação, chamada C , onde $C_{i,j} = r(x, y)$.

$$C_{i,j} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Quanto às propriedades da correlação de Pearson [Pires and Mascarenhas 2023], seu o coeficiente gera um valor entre -1 e 1, onde: (i) $C_{i,j} = 1$ indica uma correlação linear positiva perfeita; (ii) $C_{i,j} = -1$ indica uma correlação linear negativa perfeita; e (iii) $C_{i,j} = 0$ indica nenhuma correlação linear. Assim, $C_{i,j} > 0$ indica uma correlação positiva, ou seja, à medida que uma variável aumenta, a outra tende a aumentar também. O caso de $C_{i,j} < 0$ indica uma correlação negativa, ou seja, à medida que uma variável aumenta, a outra tende a diminuir. Ainda assim, $C_{i,j} = 0$ indica nenhuma correlação

linear, ou seja, as variáveis não têm uma relação linear. Outras propriedades incluem a simetria, o que significa que $C_{i,j} = C_{j,i}$, e a linearidade, que mede relações lineares entre variáveis. Apesar disso, a correlação de Pearson é sensível a valores atípicos, que podem influenciar desproporcionalmente o coeficiente [Kim et al. 2015]. No contexto deste trabalho, tem-se então uma matriz de correlação com os links de comunicação da infraestrutura de rede que fazem dos casos de problema de desempenho. Cada elemento da matriz irá ser usado no cálculo do *Score* de Impacto, assim como descrito a seguir.

3.2. Score de Impacto

Após o cálculo da matriz de correlação dos dados, é necessário extrair conhecimento sobre a situação atual da infraestrutura de rede. Para permitir a identificação de links de comunicação que afetam o atraso de ponta a ponta, desenvolvemos o *Score* de Impacto, que representa o grau de responsabilidade de um link de comunicação em casos de alto atraso ou baixo atraso na infraestrutura de rede. O cálculo do *Score* é realizado em duas fases: inicialmente calcula-se o *Score* de um tempo específico e, posteriormente, pondera-se o *Score* de acordo com o tempo, onde quanto mais próximo o *Score* do período de análise maior é o peso. Assim, o *Score* de Impacto em um dado momento é definido como a Equação 2, onde $[C_{i,j} > 0]$ é uma função que retorna 1 se $C_{i,j}$ for positivo (maior que 0) e 0 caso contrário. Dessa forma, o *Score* de Impacto visa agregar os valores de correlação de um link específico que faz parte de caminhos em casos de alto atraso, indicando um grau de responsabilidade para a situação atual.

$$Score_{n,m} = \sum_{i=1}^m \sum_{j=1}^n (C_{ij} \cdot [C_{ij} > 0]) \quad (2)$$

Quanto mais forte for a relação entre os links, maior será o valor da correlação. Consequentemente, o *Score* de Impacto proposto concentra-se em identificar o conjunto de links que têm uma correlação mais alta com outros links quando ocorre uma situação de alto ou baixo atraso. Dessa forma, quanto maior o *Score* de Impacto, maior a probabilidade de esse link ser uma das causas desse problema de desempenho.

Para realizar uma análise temporal do comportamento do *Score* de Impacto ao longo de uma medição contínua, foi implementado um sistema de peso para cada horário de medição. Essa solução foi aplicada para ser mais intuitivo a influência que uma medição possui no atraso final de uma rota, seja ele alto ou baixo. Esse peso é atribuído de forma que a última medição registrada será a de maior influência sobre o resultado (peso $w = 1$), enquanto as anteriores seguirão uma fórmula conforme descrito na Equação 3, que representa o *Score* de um link específico em uma medição feita no tempo t_i , considerando t_0 como início da medição dos atrasos e t_n como final. Com isso, a primeira medição será a que possui menos influência sobre o atraso final da rota enquanto a última medição feita, terá a maior influência no resultado.

$$Score_{n,m}^k = \sum_{i=1}^k \frac{1}{2(t_k - t_i) + 1} * Score_{n,m} \quad (3)$$

É importante lembrar que correlação não implica causalidade, o que significa que o fato de duas variáveis estarem correlacionadas não implica que uma causa a outra. Por-

todos os atrasos altos, a fim de criar uma matriz de correção com os dados de traceroute dos links utilizados (links pelos quais a rota passou) em comparação com a rota completa (ponta a ponta). Em outro momento foi selecionado apenas os dados de atrasos baixo e repetido o método acima para gerar os mesmos dados obtidos com ataso alto para fins de comparação e extrair novas informações.

4.2. Resultados

Nesta seção serão apresentados três resultados dos experimentos realizados: Matrizes de Correlação definidas, Scores de Impacto dos Links considerando apenas um Horário e o Score de Impacto considerando um intervalo de tempo, quais são discutidos nas Seções 4.2.1, 4.2.2 e 4.2.3, respectivamente.

4.2.1. Definição de Matriz de Correlação

Inicialmente nos experimentos realizados, coletamos dados em dois momentos distintos no dia 29 de maio de 2023 as 00h12min e as 00h13min Esses horários foram selecionados devido à programação que apresentava mais medidas de atrasos altos. Dessa forma, as matrizes geradas têm tamanhos diferentes devido às quantidades variadas de atrasos considerados altos e baixos nos momentos escolhidos.

A Figura 3 exibe as matrizes de correlação para esses casos. Os valores $C_{i,j}$ nas matrizes indicam o coeficiente de correlação entre os links utilizados em uma rota e os links diretos onde foram registradas medições de atraso alto durante um determinado período de tempo. Conforme mencionado anteriormente, a Rede Ipê possui Pontos de Presença (PoPs) nas capitais de todos os estados do Brasil, incluindo o Distrito Federal, e os links são indicados pelas siglas dos estados de origem e destino.

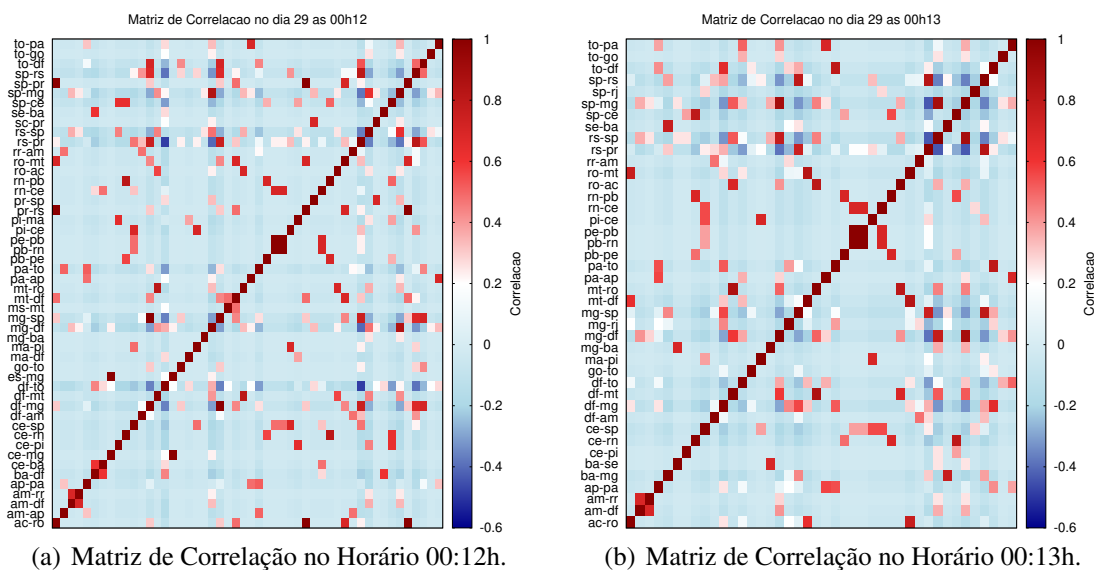


Figura 3. Resultados das Matrizes de Correlação

A partir das matrizes de correlação, percebe-se que alguns links de comunicação possuem uma alta correlação (valores em vermelho) com diversos casos de atraso alto em ambos os horários de medição (tais como os links $CE - SP$, $AP - PA$, $RS - SP$,

$RS - PR$, $DF - MG$ e $SP - MG$), enquanto que outros links aparecem como uma correlação alto em uma medição mas com uma correlação baixo em outra (como nos casos dos links $TO - GO$, $DF - TO$ e $MG - BA$). Desta forma, conclui-se que apenas a matriz de correlação não apresenta indícios fortes para a identificação de causas de problema de desempenho.

4.2.2. Score de Impacto em Horário Único

Embora a matriz de correlação possa fornecer informações valiosas sobre a possível causa do problema, utilizá-la como única fonte de análise pode não oferecer uma imagem relevante do desempenho entre os links e das possíveis causas de atraso alto. Com base nas informações coletadas da matriz de correlação, os resultados na Figura 4 apresentam os *Scores* em relação aos links utilizados durante os atrasos altos e baixos, onde quanto maior o *Score*, maior a probabilidade de o link estar contribuindo para o caso em questão (alto ou baixo). Ao adicionar as informações dos *Scores* de Impacto, é possível oferecer uma indicação mais precisa de qual dos links pode ser responsável por esses atrasos elevados. Ao observar a Figura 4, identificamos os mesmos links com os maiores *Scores* de Impacto, o que nos leva a concluir onde o problema pode estar presente, indicando uma solução mais direcionada para esses casos.

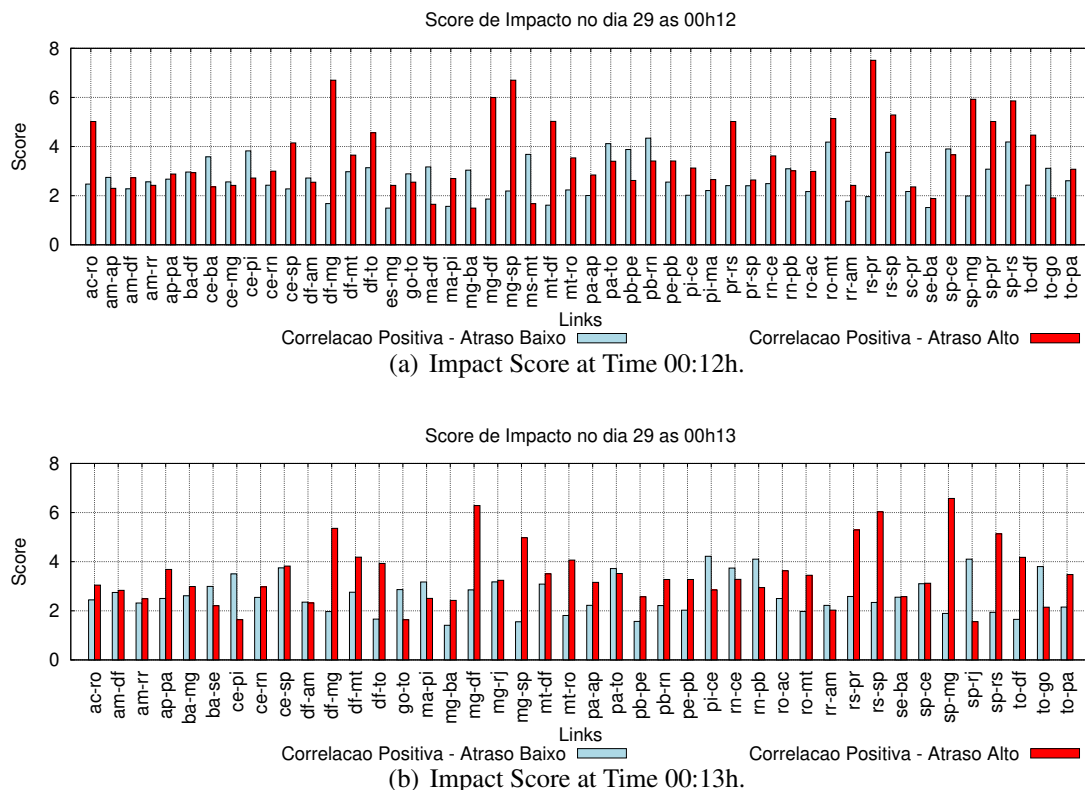


Figura 4. Resultados do Score de Impacto

Adicionalmente, quando um link possui um *Score* no caso de Atraso Alto maior que no caso de Atraso Baixo, tem-se um forte indício que este link é um causador de problema de desempenho, visto que sua contribuição maior ocorre nos casos de baixo desempenho da rede. Um exemplo desta situação ocorre nos links $AC - RO$, $DF - MG$,

$MG - SP$, $RS - PR$, dentre outros, onde a diferença do *Score* de Atraso Alto e Atraso baixo é muito grande. Da mesma forma, percebe-se nos links $TO - GO$, $PA - TO$, $PB - PE$, $CE - BA$ e outros, que o *Score* de Atraso Baixo é superior ao de Atraso Alto, indicando que estes links não possuem responsabilidade sobre os casos de atraso alto.

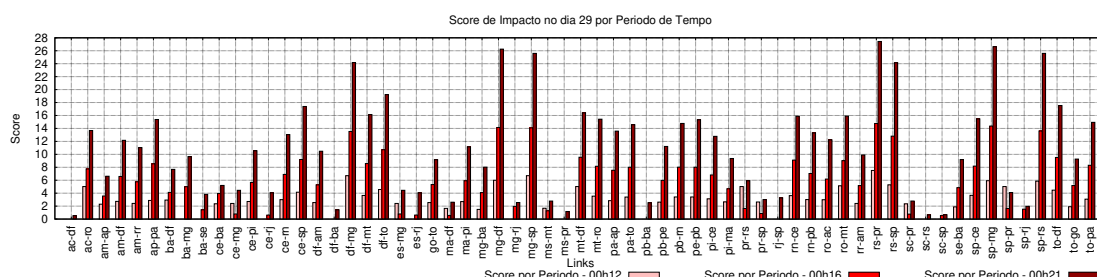
Contudo, ao analisar os dados da Figura 4, é possível detectar que há casos com alta variação de *Score* de uma medição para a outra (ou seja, em um tempo de 1 minuto), como ocorre nos links $AC - RO$, $CE - SP$, $MA - PI$, $RR - AM$ e outros. Este fato mostra que uma análise singular em relação a um momento de medição pode gerar conclusões precipitadas, porque uma medição pode capturar uma instabilidade passageira na infraestrutura de rede. Portanto, uma análise considerando um período de tempo maior tende a dar uma maior eficiência na identificação de causas de problemas de desempenho.

4.2.3. Score de Impacto com Período de Tempo

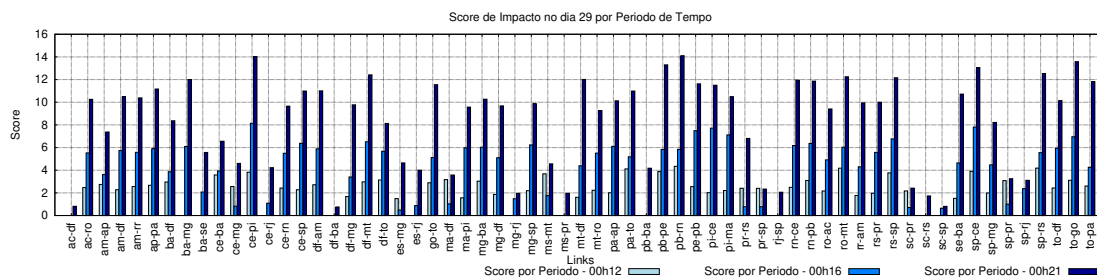
A fim de possibilitar uma análise dos links a partir de um aspecto temporal, a proposta deste artigo formulou um *Score* de Impacto considerando um período de tempo (como descrito na Seção 3). Nos experimentos realizados foi aplicada uma janela de análise de 10 minutos, ou seja, foram consideradas 10 medições (incluindo os dados de 29 de Maio das 00h12 até as 00h21) para a determinação do *Score* de um link para os casos de Atraso Alto e Atraso Baixo. Estes resultados são ilustrados na Figura 5, apresentando o comportamento do *Score* dos links em três momentos distintos (Figuras 5(a) e 5(b)): um medição (00h12), após cinco medições (00h16) e ao final de dez medições (00h21). Adicionalmente, são mostrados quatro casos específicos (nas Figuras 5(c) e 5(d)), a fim de discutir as conclusões oriundas da aplicação da solução proposta neste artigo.

As Figuras 5(a) e 5(b) mostram que, ao se considerar um período de tempo, link que possuíam *Scores* similares acabam se distanciando com o passar do tempo, como ocorre com os link $AP - PA$ e $CE - BA$, que na primeira medição do período (00h12) tinham *Scores* similares mas que ao final do período (00h21) o *Score* de $AP - PA$ acabou três vezes maior que de $CE - BA$ para o caso de Atraso Alto. Por outro lado, a análise do período também revela os casos de instabilidade de links (tais como $CE - MG$, $MS - MT$ e $PR - SP$), onde os *Scores* variam positivamente e negativamente. É válido ressaltar que um *Score* reduz de certo um período para o próximo quando este link em questão não foi utilizado em nenhuma rota de comunicação no momento da respectiva medição, sugerindo que este tem um impacto reduzido na situação de desempenho.

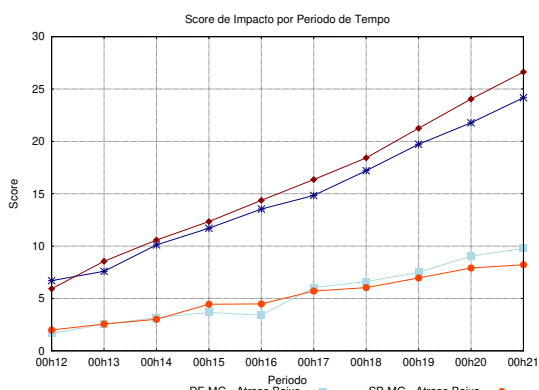
Dentre as situações identificadas, a mais importante é aquela dos links que possuem um crescimento do *Score* muito significativo ao longo do período para o caso de Atraso Alto, mas que o *Score* do caso de Atraso Baixo cresce de maneira muito discreta. Esta situação fica muito evidente nos links $DF - MG$ e $SP - MG$, que são mostrados com maior clareza na Figura 5(c). Percebe-se que estes links tem a diferença entre os *Scores* (de atraso alto e atraso baixo) se distanciando exponencialmente, indicando que estes links possuem um alto grau de participação em problemas de desempenho e, consequentemente, podem ser a causa (total ou parcial). No entanto, a Figura 5(d) apresenta dois casos onde a possibilidade dos links estarem relacionados com problemas de desempenho é muito baixo. Conclui-se isso devido ao fato dos *Scores* (baixo e alto atraso) serem similares ao longo do período de tempo (como ocorre no link $RR - AM$), bem como



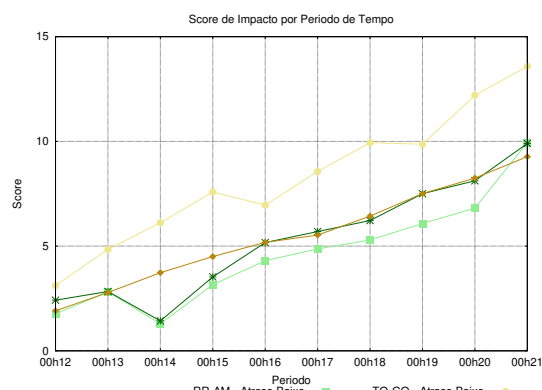
(a) Score considerando os Casos de Atraso Alto.



(b) Score considerando os Casos de Atraso Baixo.



(c) Exemplo de Índicio de Problema.



(d) Exemplo de Estabilidade de Desempenho.

Figura 5. Score por Período de Tempo

pele fato do *Score* de Atraso Baixo estar superior ao *Score* de Atraso Alto (caso do link $TO - GO$).

A partir da análise experimental conduzida neste trabalho, concluímos que o mecanismo proposto oferece uma significativa vantagem para a gestão de ISPs e o desempenho de redes. A identificação mais precisa das possíveis causas de baixo desempenho da rede possibilita aos gestores otimizar o desempenho, reduzir custos e aprimorar a satisfação do usuário. Assim, nosso mecanismo capacita os administradores de rede a tomar decisões mais informadas e abordar proativamente questões antes que impactem a rede e seus usuários.

5. Conclusão

Este artigo apresenta uma abordagem para a identificação de atrasos elevados nas conexões de rede, introduzindo o *Score* de Impacto por Período de Tempo, que se baseia em técnicas de correlação de dados e informações provenientes de ferramentas de monitoramento de rede. De maneira geral, a solução desenvolvida permite que administradores

de rede possam identificar se os links de comunicação da rede possuem uma participação ativa e impactante em casos de problema de desempenho, considerando que eles fazem parte das rotas de comunicação da rede. Assim, consegue-se distinguir se links que são muito utilizados na rede acabam afetando o desempenho da rede como um todo, bem como indica possíveis parte da infraestrutura de rede que precisam ser expandidas ou terem sua capacidade de transmissão de dados incrementada.

Os experimentos conduzidos com dados reais, da Rede Ipê da RNP, evidenciam a eficácia do mecanismo proposto em identificar os links da rede que contribuem para os atrasos altos fim-a-fim e, conseqüentemente, causam problemas de desempenho na rede e seus serviços. Esses resultados reforçam a utilidade da abordagem proposta, sinalizando seu potencial impacto na melhoria do desempenho e na otimização na gerência de redes.

Como trabalhos futuros, espera-se realizar uma análise mais aprofundada da solução buscando extrair informações mais precisas que permitam identificar os possíveis links responsáveis pelo baixo desempenho em infraestruturas de rede globais, como por exemplo a rede GEANT que disponibiliza dados de medição de desempenho e monitoramento de rede.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) do Brasil (*N^o* 303877/2021-9) e Rede Nacional de Pesquisa (RNP) através do Programa de Monitoramento de Redes (PMON 2023) e do Comitê Técnico de Ciência de Dados e Inteligência Artificial (CT-CDIA 2023) pelo apoio financeiro.

Referências

- Arachchige, K. G., Branch, P., and But, J. (2023). Evaluation of correlation between temperature of iot microcontroller devices and blockchain energy consumption in wireless sensor networks. *Sensors*, 23(14).
- BinSahaq, A., Sheltami, T., Mahmoud, A., and Nasser, N. (2022). Fast and efficient algorithm for delay-sensitive qos provisioning in sdn networks. *Wireless Networks*, pages 1–22.
- Costa, W. L., Portela, A. L., and Gomes, R. L. (2021). Features-aware ddos detection in heterogeneous smart environments based on fog and cloud computing. *International Journal of Communication Networks and Information Security*, 13(3):491–498.
- da Silva, G., Oliveira, D., Gomes, R. L., Bittencourt, L. F., and Madeira, E. R. M. (2020). Reliable network slices based on elastic network resource demand. In *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*, pages 1–9.
- Gajewski, M., Mongay Batalla, J., Mastorakis, G., and Mavromoustakis, C. X. (2022). Anomaly traffic detection and correlation in smart home automation iot systems. *Transactions on Emerging Telecommunications Technologies*, 33(6):e4053.
- Gomes, R., Bittencourt, L., Madeira, E., Cerqueira, E., and Gerla, M. (2017). Management of virtual network resources for multimedia applications. *Multimedia Systems*, 23(4):405–419.
- Gomes, R. L., Bittencourt, L. F., and Madeira, E. R. (2020). Reliability-aware network slicing in elastic demand scenarios. *IEEE Communications Magazine*, 58(10):29–34.

- Gomes, R. L., Bittencourt, L. F., Madeira, E. R., Cerqueira, E., and Gerla, M. (2016). A combined energy-bandwidth approach to allocate resilient virtual software defined networks. *Journal of Network and Computer Applications*, 69:98–106.
- Gottwalt, F., Chang, E., and Dillon, T. (2019). Corrcorr: A feature selection method for multivariate correlation network anomaly detection techniques. *Computers & Security*, 83:234–245.
- Imran, Zuhairi, M. F. A., Ali, S. M., Shahid, Z., Alam, M. M., and Su'ud, M. M. (2023). Improving reliability for detecting anomalies in the mqtt network by applying correlation analysis for feature selection using machine learning techniques. *Applied Sciences*, 13(11).
- Kim, Y., Kim, T.-H., and Ergün, T. (2015). The instability of the pearson correlation coefficient in the presence of coincidental outliers. *Finance Research Letters*, 13:243–257.
- Li, W., Wang, X., Zhang, Y., and Wu, Q. (2021). Traffic flow prediction over muti-sensor data correlation with graph convolution network. *Neurocomputing*, 427:50–63.
- Mok, R. K. P., Zou, H., Yang, R., Koch, T., Katz-Bassett, E., and Claffy, K. C. (2021). Measuring the network performance of google cloud platform. In *Proceedings of the 21st ACM Internet Measurement Conference, IMC '21*, page 54–61, New York, NY, USA. Association for Computing Machinery.
- Moreira, D. A., Marques, H. P., Costa, W. L., Celestino, J., Gomes, R. L., and Nogueira, M. (2021). Anomaly detection in smart environments using ai over fog and cloud computing. In *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–2. IEEE.
- Pires, S. and Mascarenhas, C. (2023). Cyber threat analysis using pearson and spearman correlation via exploratory data analysis. In *2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC)*, pages 257–262.
- Portela, A. L., Menezes, R. A., Costa, W. L., Silveira, M. M., Bittencourt, L. F., and Gomes, R. L. (2023). Detection of iot devices and network anomalies based on anonymized network traffic. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, pages 1–6.
- Portela, A. L. C., Ribeiro, S. E. S. B., Menezes, R. A., de Araujo, T., and Gomes, R. L. (2024). T-for: An adaptable forecasting model for throughput performance. *IEEE Transactions on Network and Service Management*, pages 1–1.
- Scarpitta, C., Sidoretti, G., Mayer, A., Salsano, S., Abdelsalam, A., and Filsfils, C. (2023). High performance delay monitoring for srv6 based sd-wans. *IEEE Transactions on Network and Service Management*, pages 1–1.
- Silveira, M. M., Portela, A. L., Menezes, R. A., Souza, M. S., Silva, D. S., Mesquita, M. C., and Gomes, R. L. (2023). Data protection based on searchable encryption and anonymization techniques. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, pages 1–5.
- Wang, B., Lun, S., Li, M., and Lu, X. (2024). Echo state network structure optimization algorithm based on correlation analysis. *Applied Soft Computing*, 152:111214.