PREVISÃO DE TRÁFEGO DE REDE EM AMBIENTE ACADÊMICO COM A FERRAMENTA PROPHET

João Guilherme Cabianchi Pires (PIC/UEM), Alisson Renan Svaigen (Orientador).
E-mail: ra134747@uem.br

Universidade Estadual de Maringá, Centro de Tecnologia, Maringá, PR.

Área e subárea do CNPq: Ciência da Computação / Sistemas de Computação

Palavras-chave: Monitoramento de redes; Séries temporais; Tráfego TCP/UDP.

RESUMO

Este trabalho investiga o fluxo em redes em um ambiente acadêmico a partir da hipótese de que o tráfego de pacotes apresenta padrões temporais regulares, passíveis de previsão por modelos de séries temporais. O conjunto de dados utilizado foi constituído por registros de tráfego coletados de forma contínua ao longo de aproximadamente 60 dias, consolidados em séries temporais com granularidade de duas horas. Inicialmente, foi realizada uma análise exploratória por meio da visualização em gráficos de pequenos múltiplos, que evidenciou indícios de ciclos diários e semanais. Para realizar a previsão de tráfego, foi utilizada a ferramenta Prophet. Os resultados obtidos indicam uma potencialidade de sazonalidade no tráfego de rede monitorado, sobretudo no protocolo UDP, o que reforça a viabilidade de modelos preditivos como ferramenta de apoio ao planejamento e ao gerenciamento proativo de redes universitárias. Como trabalhos futuros, planeja-se expandir o conjunto de dados por meio de uma coleta com maior janela temporal, além de investigar os padrões temporais em diferentes ferramentas, além do Prophet.

INTRODUÇÃO

O congestionamento em redes universitárias é um problema recorrente, podendo levar o enlace de acesso à internet a operar com mais de 90% de utilização em períodos de pico, o que resulta em formação de filas, atrasos e perda de pacotes, prejudicando a experiência dos usuários (VAZQUEZ; MORABITO; MARCONDES, 2015). Sem previsibilidade sobre o comportamento do tráfego, o gerenciamento da infraestrutura tende a ser reativo, respondendo a problemas somente depois que eles ocorrem.

Este trabalho parte da seguinte questão de pesquisa: “o tráfego de rede, considerando um ambiente universitário, apresenta padrões temporais regulares que possibilitam sua previsão por meio de modelos de séries temporais?”

De modo a investigar esta pergunta, o objetivo deste trabalho é analisar e realizar a previsão de padrões temporais do tráfego de rede em um ambiente universitário, a partir do monitoramento de tráfego de um recorte da rede cabeada da Universidade

Estadual de Maringá (UEM). Como principais contribuições, este trabalho apresenta uma análise do tráfego a partir da visualização de gráficos de pequenos múltiplos e, a partir desta análise, aplica os dados coletados em uma ferramenta de predição de séries temporais, de modo a tentar prever o fluxo de rede futuro.

MATERIAIS E MÉTODOS

Este trabalho consistiu numa metodologia de caráter exploratório, constituído pelas seguintes etapas: definição do ambiente de captura de tráfego de rede; captura e tratamento dos dados; análise empírica dos dados coletados; previsão de séries temporais. Cada etapa é descrita a seguir.

- **Definição do ambiente de captura:** a captura dos dados de tráfego de rede foi realizada em parceria com o Núcleo de Processamento de Dados da UEM, na qual foi destinada uma porta específica para captura de um recorte do tráfego de rede da universidade. Os dados coletados consistiram unicamente no protocolo do pacote e a sua quantidade, sendo mantido o encapsulamento e encriptação dos dados, de modo a preservar a privacidade das informações trafegadas.
- **Captura e tratamento dos dados:** os dados foram coletados por meio de um *sniffer* desenvolvido em Python com a biblioteca Scapy, que capturava continuamente os pacotes trafegados pela rede monitorada, classificando-os por protocolo (TCP, UDP ou ICMP) e por fluxo de origem e destino. A captura ocorreu em uma thread dedicada, enquanto uma segunda thread processava os pacotes de forma assíncrona, evitando perda de dados durante picos de tráfego. A cada duas horas, os contadores agregados eram exportados para arquivos CSV, posteriormente consolidados em uma série temporal com granularidade de duas horas e exportados em formato JSON, contendo data, horário, tipo e quantidade de pacotes.
- **Análise empírica dos dados coletados:** a partir dos dados coletados, foram geradas visualizações exploratórias em formato *pequenos múltiplos*, que evidenciaram indícios de ciclos diários e semanais no volume de tráfego. Como houveram momentos de falha na captura (e.g., quedas de energia na universidade), as lacunas de captura foram preenchidas com a média histórica de cada horário do dia, técnica que preserva melhor o padrão cíclico diário.
- **Previsão de séries temporais:** para realizar a previsão, foi utilizada a ferramenta Prophet, ferramenta de código aberto criada pela empresa Meta (TAYLOR; LETHAM, 2018). Os dados foram separados em conjunto de treino, com 720 pontos, cerca de 60 dias, e conjunto de teste, com 36 pontos, referentes aos últimos três dias. O modelo foi configurado com sazonalidade diária e semanal ativas, sazonalidade anual desativada, dado o curto período de coleta, e *changepoint_prior_scale* de 0,05, valor conservador que restringe mudanças abruptas na tendência estimada. O desempenho foi avaliado pelas métricas MAPE, que mede, em média, o quanto a previsão do modelo errou em relação ao valor real, bias, cobertura do intervalo de confiança de 95% e

MASE, que compara o erro do modelo ao de um baseline ingênuo (HYNDMAN; KOEHLER, 2006), além de validação cruzada temporal com janelas de três dias, o que confere maior robustez estatística do que um único corte de treino/teste.

RESULTADOS E DISCUSSÃO

A Figura 1 apresenta os resultados da previsão realizada pelo Prophet para os protocolos TCP e UDP, com a linha de previsão futura destacada em laranja e a área sombreada em azul representando o intervalo de confiança de 95% (IC 95%). Observa-se que a previsão se aproxima consideravelmente dos valores reais, obtendo uma cobertura do intervalo de confiança (IC 95%) de 97,2% para o TCP e de 100% para o UDP.

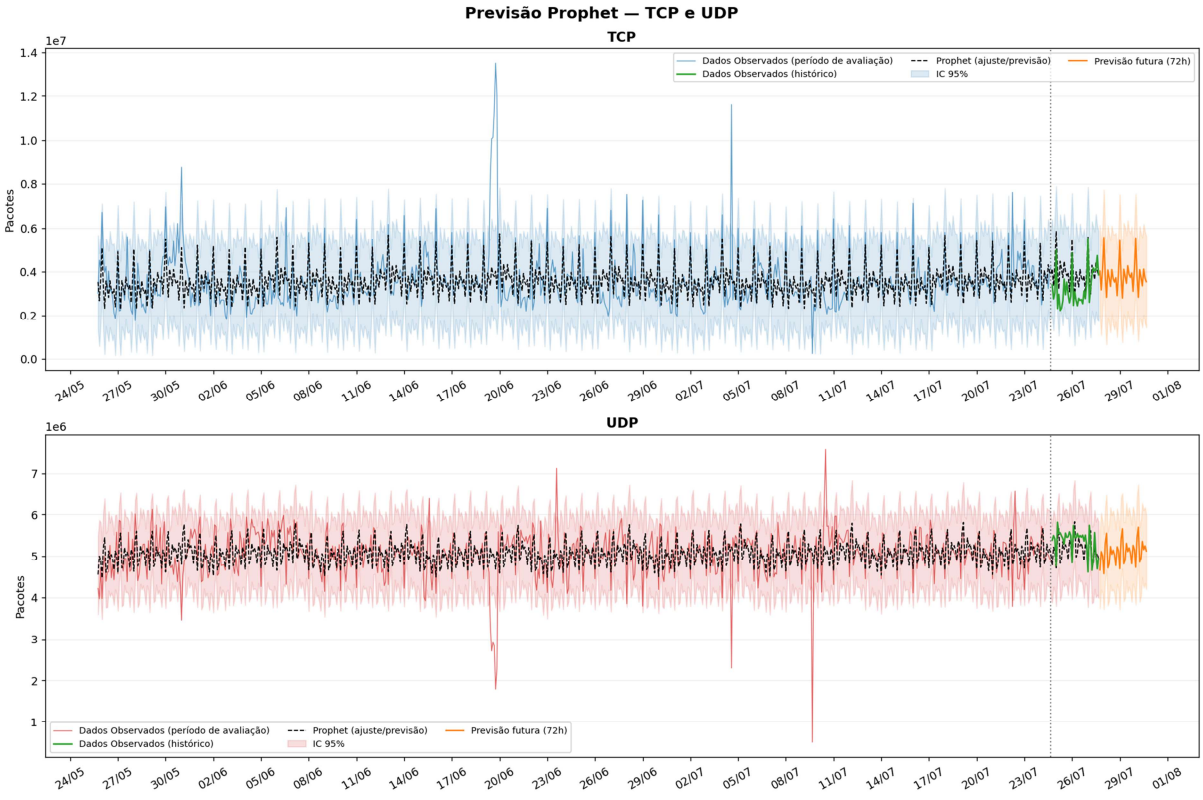


Figura 1 – Previsão de tráfego de rede (TCP e UDP) obtida pelo modelo Prophet.

Na linha azul é possível observar o dado real que foi medido e utilizado como histórico para que o modelo aprendesse o comportamento do tráfego. Em relação à linha verde estão os dados coletados que foram comparados com os dados gerados pelo Prophet para que pudéssemos fazer os cálculos das métricas de desempenho.

Por fim, à linha tracejada, observa-se a linha que o próprio Prophet está gerando, representando a previsão feita a partir do que foi aprendido com o histórico.

Em relação ao erro percentual médio (MAPE), a previsão apresentou um erro médio de 29,08% para o tráfego TCP e de 4,87% para o tráfego UDP. Já em relação ao bias, o TCP apresentou um valor negativo (-711.535 pacotes), indicando que o modelo superestimou o tráfego real, enquanto o UDP apresentou um valor positivo (189.636 pacotes), indicando subestimação por parte do modelo. Esse resultado evidencia que os dois protocolos apresentaram padrões opostos de erro direcional.

Por fim, quanto ao MASE, ambos os protocolos obtiveram desempenho superior ao baseline ingênuo, com valores de 0,56 para o UDP e 0,90 para o TCP. Destaca-se, no entanto, que o resultado do TCP se aproxima bastante da pontuação do baseline ingênuo (valor igual a 1), o que sugere um ganho preditivo mais limitado para esse protocolo em comparação ao UDP.

Conclusões

Este trabalho demonstrou a viabilidade da previsão de tráfego de rede em ambiente universitário por meio de modelos de séries temporais, com desempenho satisfatório para o protocolo UDP (MAPE de 4,87% e MASE de 0,56), superando de forma consistente o baseline ingênuo. Já para o protocolo TCP, a previsão apresentou desempenho mais limitado (MAPE de 29,08% e MASE de 0,90), próximo ao baseline, evidenciando que a previsibilidade do tráfego varia conforme a natureza do protocolo analisado. Como trabalhos futuros, pretende-se realizar uma coleta de tráfego com uma maior janela de tempo, visto que, conforme maior a série temporal coletada, melhor tende a ser o modelo de previsão. Além disso, espera-se expandir o corpo de estratégias de previsão testada, não limitando-se apenas a ferramenta Prophet, de modo a comparar sistematicamente o desempenho entre diferentes sistemas.

Referências

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting*, v. 22, n. 4, p. 679-688, 2006.

TAYLOR, S. J.; LETHAM, B. Forecasting at scale. *The American Statistician*, v. 72, n. 1, p. 37-45, 2018. Disponível em: https://doi.org/10.1080/00031305.2017.1380080

VAZQUEZ, M.; MORABITO, R.; MARCONDES, C. A. C. Caracterização, modelagem e simulação de enlace congestionado de uma universidade. Gestão & Produção, v. 22, n. 4, p. 1-12, 2015. ISSN 1806-9649. Disponível em: https://doi.org/10.1590/0104-530X1429-14