

MEDICAL IMAGE DATAFLOW IN A PUBLIC HOSPITAL

Higor Viana Viana De Morais - CENTRO UNIVERSITÁRIO SENAC SANTO AMARO - Orcid:
<https://orcid.org/0000-0001-9785-1925>

Josué Batista Matos Deschamps De Melo - CENTRO UNIVERSITÁRIO SENAC SANTO AMARO - Orcid:
<https://orcid.org/0000-0001-8978-1664>

Clarice Gameiro Da Fonseca Pachi - CENTRO UNIVERSITÁRIO SENAC SANTO AMARO - Orcid:
<https://orcid.org/0000-0001-7833-6215>

Jorge Futoshi Yamamoto - HOSPITAL DAS CLÍNICAS DA FMUSP - Orcid: <https://orcid.org/0000-0003-0360-2800>

We studied the data flow within the computer network of the Hospital das Clínicas Faculdade de Medicina da Universidade (HCFMUSP) in order to describe the traffic of the medical images, analyzing all the equipment connections and the Input/Output of the PACS servers, through the SFlow protocol. We analyse the time series and mapped the data flow of the equipment to see if the obtained time series fit the ARIMA model and, therefore, to forecast the behavior of the network. The partial results refer to the first stage of the Scientific Initiation project at the SENAC University Center - Santo Amaro campus in partnership with the Innovation group of the Núcleo Especializado de Tecnologia da Informação (NETI) at HCFMUSP.

Keywords: Time Series, Computer Network, Medical Images, Python, Hospital

FLUXO DE DADOS DE IMAGENS MÉDICAS EM UM HOSPITAL PÚBLICO

Nesse trabalho estudamos o fluxo de dados dentro da rede de computadores do Hospital das Clínicas Faculdade de Medicina da Universidade (HCFMUSP) com o objetivo de descrever o tráfego de imagens médicas, analisando todas as conexões dos equipamentos e do Input/Output dos servidores PACS, por meio do protocolo SFlow. Desenvolvemos uma análise das séries temporais e mapeamos o fluxo de dados dos equipamentos para verificar se o modelo ARIMA se adapta a série temporal obtida e, dessa forma, consideramos fazer predições do comportamento futuro da rede. Os resultados parciais apresentados se referem a primeira etapa do projeto de Iniciação Científica do Centro Universitário SENAC- campus Santo Amaro em parceria com o grupo de Inovação do Núcleo Especializado de Tecnologia da Informação (NETI) do HCFMUSP.

Palavras-chave: Séries Temporais, Rede de Computadores, Imagens Médicas, Python, Hospital

FLUXO DE DADOS DE IMAGENS MÉDICAS EM UM HOSPITAL PÚBLICO

MEDICAL IMAGE DATAFLOW IN A PUBLIC HOSPITAL

Higor Viana de Moraes

<https://orcid.org/0000-0001-9785-1925>

437.763.228-02

Centro Universitário Senac-campus Santo Amaro

higor.2100@outlook.com

Josué Batista Matos Deschamps de Melo

<https://orcid.org/0000-0001-8978-1664>

473952358-27

Centro Universitário Senac-campus Santo Amaro

josue.melomatos@gmail.com

Profa. Dra. Clarice Gameiro da Fonseca Pachi

<https://orcid.org/0000-0001-7833-6215>

057068838-82

Centro Universitário Senac-campus Santo Amaro

claricgfp@gmail.com

Prof. Dr. Jorge Futoshi Yamamoto

<https://orcid.org/0000-0003-0360-2800>

035368678-63

insilicium@gmail.com

Hospital das Clínicas da FMUSP

RESUMO:

Nesse trabalho estudamos o fluxo de dados dentro da rede de computadores do Hospital das Clínicas Faculdade de Medicina da Universidade (HCFMUSP) com o objetivo de descrever o tráfego de imagens médicas, analisando todas as conexões dos equipamentos e do Input/Output dos servidores PACS, por meio do protocolo SFlow.

Desenvolvemos uma análise das séries temporais e mapeamos o fluxo de dados dos equipamentos para verificar se o modelo ARIMA se adapta a série temporal obtida e, dessa forma, consideramos fazer previsões do comportamento futuro da rede.

Os resultados parciais apresentados se referem a primeira etapa do projeto de Iniciação Científica do Centro Universitário SENAC- *campus* Santo Amaro em parceria com o grupo de Inovação do Núcleo Especializado de Tecnologia da Informação (NETI) do HCFMUSP.

ABSTRACT:

We studied the data flow within the computer network of the Hospital das Clínicas Faculdade de Medicina da Universidade (HCFMUSP) in order to describe the traffic of the medical images, analyzing all the equipment connections and the Input/Output of the PACS servers, through the SFlow protocol.

We analyse the time series and mapped the data flow of the equipment to see if the obtained time series fit the ARIMA model and, therefore, to forecast the behavior of the network.

The partial results refer to the first stage of the Scientific Initiation project at the SENAC University Center - Santo Amaro campus in partnership with the Innovation group of the Núcleo Especializado de Tecnologia da Informação (NETI) at HCFMUSP.

PALAVRAS-CHAVE: Séries Temporais, Rede de Computadores, Imagens Médicas, Python, Hospital

KEYWORD: Time Series, Computer Network, Medical Images, Python, Hospital

1 INTRODUÇÃO

Neste trabalho serão abordadas as experiências, estudos e resultados parciais conquistados por meio do projeto “*Desenvolvimento de ferramentas estatísticas de análise de uma rede de computadores em um complexo hospitalar*”, que faz parte da pesquisa de Iniciação Científica realizada no Centro Universitário SENAC – *campus* Santo Amaro em parceria com o Hospital das Clínicas Faculdade de Medicina da Universidade (HCFMUSP).

Considerado um dos mais importantes polos brasileiros de disseminação de informações técnico-científicas, o HCFMUSP é um Centro de excelência e referência no campo de ensino, pesquisa e assistência.

Com área construída de cerca de 380 mil metros quadrados, conta com dois mil leitos e 15 mil profissionais nas mais diversas atividades. É formado por sete institutos, dois hospitais auxiliares, laboratórios de investigação médica, unidades especializadas e demais áreas de apoio como o Prédio da Administração e Anexos, o Centro de Convenções Rebouças e a Escola de Educação Permanente.

Em relação ao parque tecnológico, o HCFMUSP dispõe de um total de máquinas em torno de 9000 computadores distribuídos entre os usuários, além de alguns data centers, sendo que o principal está localizado em um ponto central do complexo hospitalar, com diversas fibras óticas em topologia de estrela-anel interligando todos os institutos.

Diante da complexidade dessa rede e do volume de tráfego do HCFMUSP, pode-se destacar a importância de se manter um sistema íntegro e funcional que evite falhas que comprometam a integridade dos dados (PACHI et al., 2021).

Existem muitas soluções exclusivas para ajudar nas tarefas de análise de tráfego de redes de computador. Em geral, essas ferramentas permitem coletar e armazenar dados para análise em tempo real e posterior. Alguns desenvolvedores criam algoritmos baseados

em modelos matemáticos que permitem análises por simuladores ou robôs que interagem com relatórios na rede. Ao usar essas ferramentas em pontos estratégicos de uma rede é possível capturar dados globais ou fragmentados, dependendo de suas necessidades analíticas. Portanto, fazer a escolha de um ponto de coleta de tráfego é fator essencial para projetos de análise de rede e, conforme a rede cresce, a tarefa de coletar dados de tráfego se torna cada vez mais complexa (MACEDO, 2015).

Atualmente o mundo todo lida com diversos tipos de informações que são espalhadas através da internet. Com isso a tecnologia para coleta e análise de dados ganha um espaço importante em grande parte das empresas, permitindo que a informação seja obtida, armazenada, compartilhada e visualizada de forma mais otimizada.

A exploração do tráfego de dados dentro de um complexo hospitalar ganha relevância ao entender como funciona o processo de transferência de dados, quais as medidas de segurança para que os dados transmitidos não sejam expostos e como eles poderão ajudar aquele que irá analisá-los.

As redes de computadores permitem enviar informações como imagens, sons, vídeos e muitos tipos de arquivos. Essa transmissão pode ser usada para vários fins, incluindo a transmissão de imagens médicas, usando tecnologias específicas como o DICOM (Digital Imaging and Communications in Medicine) e o PACS (Picture Archiving and Communication System) entre outras. Em resumo, podemos dizer que as imagens são obtidas através dos equipamentos médicos (raios-X, tomografia etc.) e armazenadas no PACS são transmitidas pela rede TCP/IP existente em grandes complexos, obedecendo ao protocolo denominado DICOM.

Há vários protocolos que permitem a integração entre diferentes sistemas de obtenção de imagens médicas e o mais conhecido é o HL7 (Health Level 7), que permite a troca de informações entre o sistema de informações de pacientes e o DICOM, material de estudo desse trabalho (PIANYKH, 2012).

O DICOM é um sistema de comunicação para imagens digitais na medicina, que, tem como representação uma série de regras criadas em 1983 no American College of Radiology, que visam facilitar a armazenagem e comunicação de imagens. Desde sua criação, já sofreu três alterações para que houvesse uma melhoria no sistema de imagens e melhor expansão do PACS para conversação com sistemas e versões diferentes, melhorando a experiência e acesso para profissionais visualizarem exames em qualquer dispositivo (PIANYKH, 2012).

O PACs é um tipo de tecnologia que facilita a comunicação e arquivamento de clínicas diagnósticas através de estudos por imagens. Trata-se de uma ferramenta muito útil para evitar a perda de laudos, redução de custos e possibilita uma melhor flexibilidade dos dados armazenados (STRICKLAND, 2000).

Considerando a importância dessas tecnologias para a rede em estudo, surgiu a proposta de pesquisar, compreender e propor mecanismos de previsões para o comportamento dos dados e estudá-los por meio de algoritmos desenvolvidos na linguagem de programação Python com uso de suas diversas bibliotecas.

2 METODOLOGIA

Os dados da rede foram obtidos por meio de uma ferramenta de captura de código aberto do protocolo Sflow (Sample Flow), disponibilizada pela empresa InMon (www.inmon.com), que o criou. Detalhes sobre o Sflow podem ser encontrados no site <https://sflow.org/>.

O Sflow trata-se de uma simplificação do NetFlow criado pela Cisco. O Sflow vem com a vantagem de exigir pouco uso de processamento em um computador e ainda permite coletar os dados de uma rede de forma detalhada, o que justifica a necessidade de ser usado na rede em questão para que a pesquisa possa ser bem elaborada.

A ferramenta de captura permite a geração de arquivos ASCII, JSON, ou em formato binário, compatível com a ferramenta TCPDUMP, frequentemente utilizadas por aplicações em distribuições de Linux.

Foi utilizado o formato binário, por ocupar menos espaço e por poder utilizar o software Wireshark, que possui licença OpenSource (Fig 1).

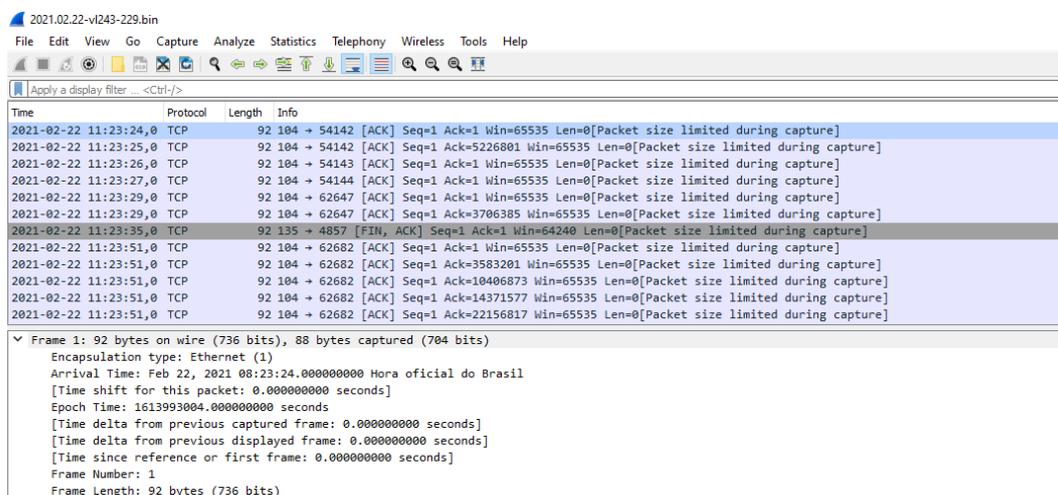
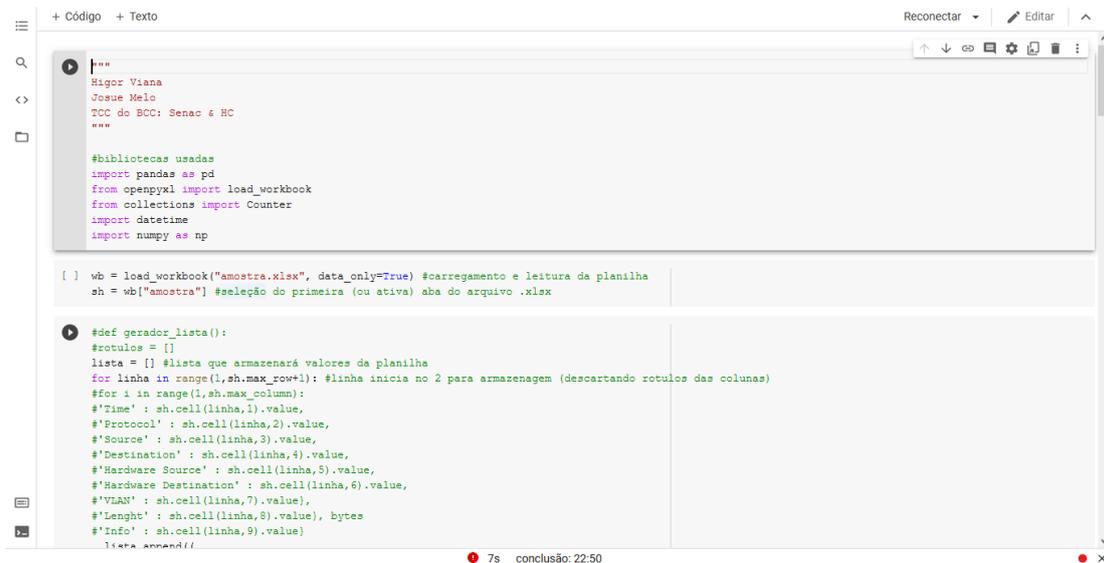


Figura 1 – Dados visualizados no Wireshark, usado para analisar toda a informação de forma bruta, sem que exista uma manipulação direta das informações coletadas de acordo com a Figura 1.

2 DESENVOLVIMENTO

Por meio do uso da linguagem de programação Python, conhecida pela simplicidade e fácil para compreensão humana, foi realizada exportação e organização dos dados em planilhas para serem lidas, manipuladas e visualizadas pela ferramenta gratuita da Google Colab (Fig 2).



```
"""
Rigor Viana
Josue Melo
TCC do BCC: Senac e HC
"""

#bibliotecas usadas
import pandas as pd
from openpyxl import load_workbook
from collections import Counter
import datetime
import numpy as np

wb = load_workbook("amostra.xlsx", data_only=True) #carregamento e leitura da planilha
sh = wb["amostra"] #seleção do primeira (ou ativa) aba do arquivo .xlsx

def gerador_lista():
    #rotulos = []
    lista = [] #lista que armazenará valores da planilha
    for linha in range(1,sh.max_row+1): #linha inicia no 2 para armazenagem (descartando rotulos das colunas)
        for i in range(1,sh.max_column):
            #'Time' : sh.cell(linha,i).value,
            #'Protocol' : sh.cell(linha,2).value,
            #'Source' : sh.cell(linha,3).value,
            #'Destination' : sh.cell(linha,4).value,
            #'Hardware Source' : sh.cell(linha,5).value,
            #'Hardware Destination' : sh.cell(linha,6).value,
            #'VLAN' : sh.cell(linha,7).value),
            #'Lenght' : sh.cell(linha,8).value), bytes
            #'Info' : sh.cell(linha,9).value)
            lista.append((
```

Figura 2 – Trecho do código que faz a manipulação com linguagem Python das informações capturadas, usando bibliotecas da linguagem.

Com o objetivo de melhorar o fluxo de trabalho e colaboração entre os indivíduos envolvidos na pesquisa, foi utilizada a ferramenta GitHub que armazena dados de forma gratuita, onde todos os códigos de programação estão disponíveis (Fig 3).

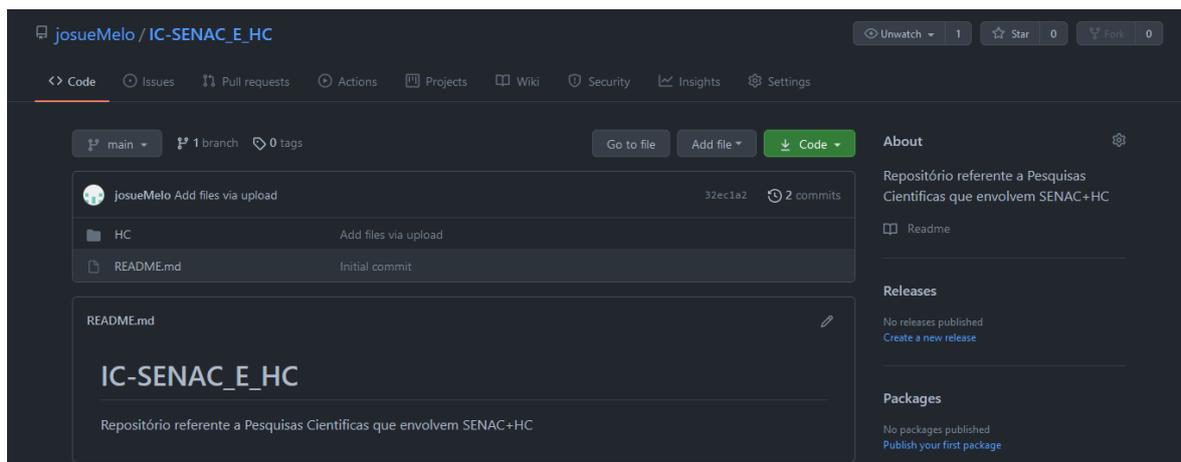


Figura 3 – Repositório no GitHub que armazena todo o código elaborado em Python para a visualização e manipulação dos dados.

4 RESULTADOS OBTIDOS

A coleta e manipulação garantiu várias formas de visualizar as informações, como por exemplo, a criação de uma série temporal que mostra a quantidade de comunicações que ocorreram na rede em um determinado período (Fig 4).

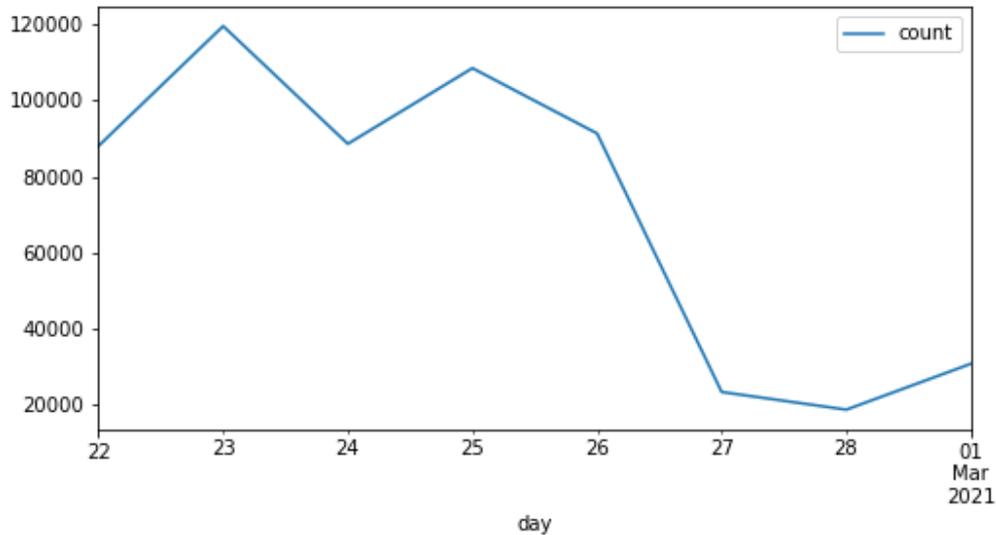


Figura 4 – Gráfico de séries temporais onde “count” refere-se à quantidade de comunicações que ocorreram na rede durante o período de 22 de fevereiro de 2021 a 01 de março de 2021.

Também foi verificada a quantidade de ocorrências de endereços IP’s, e quais foram os outros endereços que tiveram comunicação entre si e diversas outras informações em um determinado período (Fig 5).

```
In [*]: menu()

Vamos começar! Quantos IPs gostaria de visualizar? ex.: 5
> 5
Escolha como deseja filtrar a visualização:
0 - Time
1 - Protocol
2 - IP Source
3 - IP Destination
4 - Hardware Source
5 - Hardware Destination
6 - VLAN
7 - Length
8 - Info
> 3

Os 5 IP Destination mais frequentes foram:
1 - 10.65.137.46, usado 89059 vezes.
2 - 10.65.143.5, usado 57750 vezes.
3 - 10.65.143.4, usado 53205 vezes.
4 - 10.65.143.7, usado 41554 vezes.
5 - 10.65.142.81, usado 41489 vezes.

Para visualizar mais informações sobre um IP específico, digite seu ID e pressione ENTER:
> 17
Deseja visualizar apenas ocorrências que contenham o protocolo DICOM? (S/N)
> S
Resumos do IP 10.65.143.6:
1 - HewlettP_55:8f:6e-10.65.137.46-BrocadeC_ef:3c:6f-10.65.143.6-2021-02-22 13:43:26-DICOM-[TCP Previous segment not captur
ed] P-DATA[Packet size limited during capture]-234-362
2 - HewlettP_55:8f:6e-10.65.137.46-BrocadeC_ef:3c:6f-10.65.143.6-2021-02-22 15:35:16-DICOM-[TCP Previous segment not captur
ed] P-DATA[Packet size limited during capture]-234-362
3 - HewlettP_55:8f:6e-10.65.137.46-BrocadeC_ef:3c:6f-10.65.143.6-2021-02-22 18:51:18-DICOM-[TCP Previous segment not captur
ed] P-DATA[Packet size limited during capture]-234-362
4 - HewlettP_55:8f:6e-10.65.137.46-BrocadeC_ef:3c:6f-10.65.143.6-2021-02-22 20:04:34-DICOM-[TCP Previous segment not captur
ed] P-DATA[Packet size limited during capture]-234-362
5 - HewlettP_55:8f:6e-10.65.137.46-BrocadeC_ef:3c:6f-10.65.143.6-2021-02-22 20:49:04-DICOM-[TCP Previous segment not captur
ed] P-DATA[Packet size limited during capture]-234-362
6 - HewlettP_55:8f:6e-10.65.137.46-BrocadeC_ef:3c:6f-10.65.143.6-2021-02-22 21:04:10-DICOM-[TCP Previous segment not captur
ed] P-DATA[Packet size limited during capture]-234-362
```

Figura 5 – Impressão das informações através de um prompt da ferramenta Google Colab. As informações mostram a manipulação feita que gerou os principais endereços de IP encontrados na amostra e demais detalhes que os envolvem.

Fonte: Elaborado pelos autores

De forma análoga toda a manipulação dos dados foi armazenada em planilhas, para facilitar sua visualização.

A Tabela 1 a seguir mostra os 5 maiores endereços IP remetentes e endereços IP destinatários obtidos pela leitura dos dados e a quantidade de vezes em que estas comunicações ocorreram e seu respectivo percentual em relação ao total de ocorrências da amostra coletada.

Tabela 1: 5 maiores endereços de IP remetentes e endereços IP destinatários obtidos pela leitura dos dados

	Endereços IP	Quantidade	Quantidade (%)
Remetentes	10.65.137.46	89059	0,156603
	10.65.143.5	57750	0,101549
	10.65.143.4	53205	0,093557
	10.65.143.7	41554	0,07307
	10.65.142.81	41489	0,072955
Destinatários	10.65.137.46	423348	0,744425
	10.65.142.81	27353	0,048098
	10.65.142.82	20529	0,036099
	10.65.142.83	8493	0,014934
	10.65.136.126	7259	0,012764

5 CONSIDERAÇÕES FINAIS

Embora os resultados obtidos ainda sejam preliminares, a pesquisa segue em desenvolvimento, com objetivo de aprimorar a coleta de dados de todas as conexões dos equipamentos da rede do complexo HCFMUSP.

Ao analisar o fluxo de dados dos switches que fazem a transmissão dos dados, é possível avaliar o tráfego das imagens médicas dentro da rede e, dessa forma, fazer predições úteis.

REFERÊNCIAS

CHECHI, L. and Bayer, F. M. (2012). *Modelos univariados de séries temporais para previsão das temperaturas médias mensais de Erechim, RS*, SciELO - Scientific Electronic Library Online.

COMER, D. E.4^a (Ed.), 2007. *Redes de Computadores e Internet*. Bookman.

COUTO, A. V. (2012). *Uma abordagem de Gerenciamento de Redes baseado no Monitoramento de Fluxos de Tráfego Netflow com o suporte de Técnicas de Business Intelligence*, Universidade de Brasília.

MACEDO, E. L. C. (2015). *PREVISÃO DE TRÁFEGO EM ENLACES DE REDES UTILIZANDO SÉRIES TEMPORAIS*, Universidade Federal do Rio de Janeiro.

MORETTIN, P.A., TOLOI, C.M.C 1^a (Ed.), **2006.** *Análise de séries temporais*. Edgard Blucher.

PACHI, C.G.F., VIDO, K.C., TAVARES. S., BARBOZA, S.H.I. 2021. (Projeto de Pesquisa) Desenvolvimento de ferramentas estatísticas de análise de uma rede de computadores em um complexo hospitalar.

PIANYKH, O. S. 2^a (Ed.), **2012.** *Digital Imaging and Communications in Medicine (DICOM)*. Springer, Berlin, Heidelberg.

SILVA, I. R. S.; Silva, R. O. *Linguagem de Programação Python, Revista Tecnologias em Projeção*, v10, n^o1, ano 2019. p.55-71.

STRICKLAND, N. (2000a). *PACS (picture archiving and communication systems): filmless radiology*, BMJ Publishing Group.