UNIVERSIDADE FEDERAL DO PARÁ
FACOMP – FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

IMPLEMENTAÇÃO E AVALIAÇÃO DE UM SERVIDOR DE MONITORAMENTO E ANÁLISE DE FLUXOS COM FERRAMENTAS OPEN-SOURCE EM UM CENÁRIO DE PROVEDOR DE INTERNET

BRUNO ALEXANDER JAQUES MOREIRA

Castanhal-PA
2024

UNIVERSIDADE FEDERAL DO PARÁ
FACOMP – FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

BRUNO ALEXANDER JAQUES MOREIRA

IMPLEMENTAÇÃO E AVALIAÇÃO DE UM SERVIDOR DE MONITORAMENTO E ANÁLISE DE FLUXOS COM FERRAMENTAS OPEN-SOURCE EM UM CENÁRIO DE PROVEDOR DE INTERNET

Trabalho de Conclusão de Curso submetido ao colegiado da Faculdade de Computação da Universidade Federal do Pará, como requisito parcial para a obtenção do grau de bacharel em ENGENHARIA DE COMPUTAÇÃO.

Orientador: Prof. Dr. José Jailton Júnior

Castanhal-PA
2024

UNIVERSIDADE FEDERAL DO PARÁ
FACOMP – FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

IMPLEMENTAÇÃO E AVALIAÇÃO DE UM SERVIDOR DE MONITORAMENTO E ANÁLISE DE FLUXOS COM FERRAMENTAS OPEN-SOURCE EM UM CENÁRIO DE PROVEDOR DE INTERNET

Trabalho de Conclusão de Curso apresentado ao Colegiado da Faculdade de Computação (FACOMP) da Universidade Federal do Pará do campus de Castanhal, como requisito parcial para a obtenção do Grau de bacharel em ENGENHARIA DE COMPUTAÇÃO.

_____
Prof. Dr. José Jailton Júnior
Orientador-UFPA/FACOMP

_____
Prof. Dr. Igor Ruiz Gomes
Membro da Banca – UFPA/FACOMP

_____
Prof. Dr. Tássio Costa de Carvalho
Membro da Banca – UFPA/FACOMP

_____
Profa. Dra. Yomara Pinheiro Pires
Diretor (a) da Faculdade de Computação - FACOMP

Castanhal-PA
2024

DEDICATÓRIA

Dedico este Trabalho de Conclusão de Curso à minha avó Graça e aos meus pais, Alcy e Eliane, que sempre se esforçaram ao máximo para me proporcionar a melhor educação possível, por muitas vezes sacrificando suas vontades e necessidades em prol do meu futuro. Sem vocês nada seria possível, gratidão eterna a todo carinho, amor e apoio que me deram.

AGRADECIMENTOS

Gostaria de agradecer primeiramente à Deus pela benção que é estar vivo e com saúde, o que me possibilita continuar lutando por meus objetivos e realizando sonhos. Agradeço a toda minha família, minha avó Graça, minha mãe Eliane, meu pai Alcy, e meus irmãos Paula e Ygor, pelo apoio e amor incondicionais.

Também sou grato a todos os amigos, que com sua presença, me ajudaram a chegar até este momento. Os amigos do colégio Darwin, Gabriel, Rafael e Wilton, a quem considero irmãos. Os amigos e colegas da UFPA, Adrielle, Andercley, Claudine, Kleu, Rose, Sara e Tatiane, que me apoiaram de todas as formas possíveis em toda a trajetória acadêmica. Os amigos e colegas de trabalho, Adal, Claudionor, Eduarda, Eliane, Emanuel, Gustavo, Luciana, Milena e Rafael, que ofereceram apoio tanto emocional quanto intelectual no desenvolvimento da minha carreira profissional e deste trabalho. E as amizades conquistadas em diferentes momentos da vida, e se fazem presentes tanto nos momentos de felicidade quanto nos de dificuldades, Antônio, Ariel, Brena, Dhamy, Ícaro, Renner, Rony e minha gêmea Hanna.

Agradeço a todo o corpo docente da Faculdade de Computação, por todo conhecimento adquirido dentro e fora das aulas. Em especial a meu orientador, professor JJJ, pela insistência e paciência investidas para o desenvolvimento e conclusão deste trabalho. Por fim agradeço ao José Paulo pelo apoio psicológico e ajuda profissional nos momentos difíceis, sem o qual não conseguiria me dedicar a finalizar este trabalho.

SUMÁRIO

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

RESUMO

Durante a pandemia de Covid-19, o tráfego de internet no Brasil aumentou significativamente. Mesmo anos após o impacto inicial, a demanda continua a crescer de forma constante. Nesse contexto, o monitoramento e a análise de tráfego são essenciais para atender a essa demanda de maneira eficaz e melhorar os serviços oferecidos pelos provedores de internet. A avaliação da eficácia das ferramentas utilizadas para o monitoramento e análise de tráfego em um cenário dinâmico de provedor de internet contribui para o aprimoramento da gestão de redes, resultando na oferta de serviços mais confiáveis e eficientes.

Este trabalho tem como objetivo implementar e avaliar a eficácia prática de ferramentas open-source para monitoramento e análise de tráfego em um cenário de provedor de internet. Inicialmente foi abordado o monitoramento e a análise de tráfego através do protocolo de exportação de fluxos NetFlow, e em seguida foram discutidas a implementação e a avaliação prática das ferramentas open-source flow-tools e FlowViewer.

Durante a implementação do conjunto de ferramentas no servidor, surgiram dificuldades relacionadas a erros de código nos scripts dos softwares, exigindo um esforço adicional na identificação e correção dos erros. Isso demandou um certo nível de conhecimento em Shell Script e Perl, além de tempo para ajustes. A superação das dificuldades demonstrou a viabilidade técnica das ferramentas, a interface amigável e de fácil utilização do FlowViewer facilitou a obtenção e a interpretação dos dados capturados pelo flow-tools.

A partir dos dados fornecidos pelos relatórios gerados pelo FlowViewer, foi possível realizar uma análise detalhada do tráfego coletado, permitindo a identificação de padrões. A avaliação foi focada no volume de tráfego, nas taxas de transmissão e nas portas mais utilizadas na rede. Essa experiência reforça a viabilidade da adoção dessas ferramentas em cenários de provedores de internet, comprovando que é possível monitorar e analisar de forma efetiva o tráfego de redes complexas utilizando soluções gratuitas.

PALAVRAS-CHAVES: tráfego de redes. monitoramento e análise de tráfego. exportação de fluxos. Netflow. open-source.

ABSTRACT

During the Covid-19 pandemic, internet traffic in Brazil increased significantly. Even years after the initial impact, the demand continues to grow steadily. In this context, traffic monitoring and analysis are essential to effectively meet this demand and improve the services offered by internet service providers. Evaluating the effectiveness of the tools used for traffic monitoring and analysis in a dynamic internet service provider scenario contributes to the enhancement of network management, resulting in more reliable and efficient services.

This study aims to implement and assess the practical effectiveness of open-source tools for traffic monitoring and analysis in an ISP scenario. Initially, traffic monitoring and analysis were conducted using the NetFlow flow export protocol, followed by the implementation and practical evaluation of the open-source tools flow-tools and FlowViewer.

During the implementation of the toolset on the server, challenges related to code errors in the software scripts arose, requiring additional effort to identify and correct the errors. This required a certain level of knowledge in Shell Script and Perl, as well as time for adjustments. Overcoming these difficulties demonstrated the technical feasibility of the tools; the user-friendly and easy-to-use interface of FlowViewer made it easier to obtain and interpret the data captured by flow-tools.

The detailed traffic analysis, based on data provided by FlowViewer's reports, allowed for the identification of traffic patterns, focusing on traffic volume, transmission rates, and the most frequently used network ports. This experience highlights the feasibility of adopting these tools in ISP environments, proving that it is possible to effectively monitor and analyze complex network traffic using free solutions.

KEYWORDS: network traffic. traffic monitoring and analysis. flow export. Netflow. open-source

1 INTRODUÇÃO

A pandemia de COVID-19 provocou uma mudança profunda no comportamento digital dos brasileiros. A transição massiva para o trabalho remoto, a educação online e a comunicação virtual impulsionaram o número de usuários da internet e consequentemente a demanda por serviços de provedores (NIC.BR, 2021). Segundo dados da pesquisa anual TIC Domicílios, realizada pelo Cetic.BR (departamento do Núcleo de Informação e Coordenação do Ponto BR), cerca de 84% da população do Brasil fez uso da internet em 2023 (156 milhões de usuários), 7% a mais de usuários em relação à 2019 (NIC.BR, 2023).

A crescente quantidade de usuários e as mudanças no modo em que eles utilizam a internet estabeleceram um alto padrão de crescimento na quantidade de dados trafegados, que se mantém mesmo após o fim da pandemia. Em março de 2020 o IX.br (projeto que promove e cria a infraestrutura necessária para a interconexão direta entre as redes que compõe a internet brasileira) registrou picos de tráfego de 11 Tb/s, 60% a mais que o maior pico do ano anterior. Em março de 2021 um novo recorde foi alcançado, com picos de tráfego de 16 Tb/s, em dezembro de 2021 o tráfego chegou à 20 TB/s, e em julho de 2023 o recorde de tráfego foi novamente quebrado, com a marca de 31 TB/s (IX.BR, 2020, 2021, 2023).

O monitoramento e o entendimento das crescentes demandas de tráfego são essenciais para entender as novas expectativas dos usuários, e orientar estratégias de expansão e aprimoramento dos serviços oferecidos pelos provedores de Internet. Autores como Feldmann et al. (2020) e Böttger et al. (2020) destacam a necessidade do monitoramento e análise de tráfego para que as redes sejam resilientes o suficiente a fim de atender a essa demanda dinâmica.

A crescente complexidade das redes dos provedores de internet demanda avaliações da eficácia prática de ferramentas utilizadas para monitorar e analisar seu tráfego. No panorama das ferramentas utilizadas para monitorar e analisar o tráfego de redes, os protocolos de exportação de fluxos, Netflow e IPFIX, destacam-se como os mais utilizados em comparação com as demais tecnologias, graças à sua grande escalabilidade, baixo ou inexistente investimento inicial, e fácil usabilidade.

Este trabalho visa abordar a implementação e avaliação da eficácia prática de um conjunto de ferramentas open-source que utiliza o protocolo Netflow para monitoramento e análise de fluxos em um contexto específico de provedor de internet. Diante da

importância crítica de manter operações estáveis e eficientes, a pesquisa busca avaliar a capacidade das ferramentas flow-tools e FlowViewer em identificar e analisar o tráfego de fluxos. A indagação sobre a eficácia destas ferramentas em um cenário dinâmico de provedor de internet contribui com o aprimoramento da gestão de redes, e consequentemente com a oferta de serviços mais confiáveis e eficientes.

Este trabalho tem como objetivo geral a implementação das ferramentas de monitoramento e análise de fluxos flow-tools e FlowViewer, com a finalidade de avaliar sua eficácia prática em um cenário específico de provedor de internet, para atingir este objetivo foram delimitados objetivos específicos:

- Contextualizar o monitoramento e análise de tráfego, e sua relevância para garantir a estabilidade e eficiência das operações de redes de computadores.
- Contextualizar a exportação de fluxos, e suas vantagens em comparação com outros métodos de monitoramento e análise de tráfego.
- Contextualizar o protocolo de exportação de fluxos utilizado nas ferramentas implementadas neste trabalho, Netflow.
- Apresentar as ferramentas open-source utilizadas, flow-tools e FlowViewer, fornecendo uma base teórica para a sua implementação prática.
- Abordar a implementação das ferramentas apresentadas em um cenário de provedor de internet.
- Abordar a eficácia prática das ferramentas implementadas em termos de análise do tráfego de redes.

Este trabalho está organizando em dez capítulos, sendo que o restante está organizado como segue:

Capítulo 2: Neste capítulo serão contextualizados o monitoramento de tráfego em um contexto de redes de computadores, seus dois tipos (ativo e passivo) e suas respectivas características, e a importância da análise de tráfego para o aprimoramento dos serviços de internet.

Capítulo 3: Neste capítulo serão abordados o que é a exportação de fluxos, suas vantagens em comparação com outros meios de monitorar tráfego, arquitetura e funcionamento.

Capítulo 4: Neste capítulo serão apresentados o protocolo de exportação de fluxos Netflow, sua relevância, funcionamento, modelos de implementação, ferramentas compatíveis e exemplos práticos de utilização.

Capítulo 5: Neste capítulo será abordada a arquitetura do sistema de monitoramento de fluxos implementado, apresentando o cenário onde foi instalado, e as especificações de hardware e software utilizados.

Capítulo 6: Este capítulo aborda de forma detalhada a implementação da ferramenta open-source flow-tools, conjunto de utilitários responsável pela coleta e armazenamento de fluxos no servidor.

Capítulo 7: Neste capítulo é detalhado o processo de instalação da ferramenta responsável pela emissão de relatórios FlowViewer, detalhando as dificuldades encontradas e soluções empregadas.

Capítulo 8: Este capítulo aborda a efetiva análise do tráfego de rede, realizada através dos dados fornecidos pelos relatórios do FlowViewer. Examinando especificamente atributos como o volume de tráfego, taxas de transmissão e portas mais utilizadas para o estabelecimento de padrões de tráfego.

Capítulo 9: Neste capítulo são apresentados os resultados deste trabalho, onde foram agregadas informações adquiridas através da análise de tráfego, contribuindo para a obtenção de insights sobre o comportamento da rede.

Capítulo 10: Neste capítulo será apresentada a conclusão deste trabalho, discutindo os resultados obtidos e possíveis trabalho futuros.

2 MONITORAMENTO E ANÁLISE DE TRÁFEGO

Redes de computadores são estruturas interconectadas que permitem a troca de dados e comunicação entre dispositivos, facilitando a colaboração e o compartilhamento de recursos. Segundo Tanenbaum et al. (2021), uma rede de computadores consiste em nós (dispositivos) interligados por meio de diferentes meios de comunicação, como cabos ou conexões sem fio. A complexidade dessas redes varia desde pequenas redes locais (LANs) até vastas redes globais, como a Internet. Autores como Kurose e Ross (2021) destacam que as redes desempenham um papel crucial na atual era da informação, facilitando a transmissão de dados, a colaboração remota e o acesso a recursos distribuídos.

A intensificação da dependência da sociedade em redes de computadores foi significativamente acelerada pela pandemia de 2020, devido ao aumento na demanda por teletrabalho, educação online e consumo de mídia digital. Esse crescimento substancial no volume de dados requer uma gestão estratégica dos recursos e uma otimização criteriosa da rede. Consequentemente, monitorar e avaliar métricas de tráfego de rede tornou-se uma responsabilidade crucial para provedores de serviços e gestores de redes.

A análise e o monitoramento do tráfego de rede são essenciais no contexto das telecomunicações, pois asseguram a eficiência da rede, a segurança e a qualidade da experiência do usuário. Esses processos são fundamentais para que provedores de serviços de internet (ISPs) e outras empresas de telecomunicações possam identificar obstáculos, otimizar o fluxo de tráfego e alocar a largura de banda de forma eficiente. Além disso, essas práticas permitem decodificar padrões complexos de tráfego, prever e mitigar problemas relacionados ao congestionamento da rede, gargalos e possíveis ameaças à segurança (Fowdur e Babooram, 2024).

Joshi et al. (2016) apresentam conceitos importantes a respeito do monitoramento e análise do tráfego de rede em "Network Traffic Analysis Measurement and Classification Using Hadoop". De acordo com os autores, os dados de rede são encapsulados em pacotes que compõem a carga na rede, e o tráfego de rede se refere à quantidade de dados que transitam pela rede em um determinado momento. Eles destacam que o tráfego de rede é um componente essencial para o controle, medição e simulação da rede, sendo que o monitoramento de tráfego tem como objetivo principal assegurar a disponibilidade e a estabilidade das operações em redes de computadores, e seu processo

envolve técnicas de exploração e captura de pacotes para acompanhar o comportamento de uma rede.

Diversos pesquisadores têm contribuído com definições e estudos sobre o monitoramento e análise do tráfego em redes de computadores. Abbasi et al. (2021) descrevem o monitoramento e análise de tráfego como um conjunto de técnicas que observam o tráfego da rede em diferentes níveis de granularidade. Conti et al. (2018) afirmam que a análise de tráfego de rede é uma área da ciência da computação que investiga métodos inferenciais, utilizando rastros de rede gerados por dispositivos como entrada para extrair informações sobre os dispositivos, seus usuários, aplicativos ou o próprio tráfego.

Mistry et al. (2016) destacam que o monitoramento do tráfego na Internet é o processo de observar a troca de dados entre dispositivos, enquanto sua análise envolve a captura e o exame detalhado do tráfego para identificar atividades de rede. Essa captura pode ser realizada em diferentes pontos da rede, como switches, roteadores e gateways. Shen et al. (2022) definem a análise de tráfego como a prática de monitorar atividades de rede, identificar padrões específicos e coletar informações relevantes.

As aplicações de monitoramento e análise de tráfego em redes abrangem desde a obtenção de uma visão detalhada do tráfego até a identificação de anomalias e ataques desconhecidos, alimentando sistemas responsáveis pelo monitoramento e contabilização do uso de rede (D'Alconzo et al., 2019). Dentre as métricas que devem ser consideradas na avaliação de uma rede, as principais são: latência, perda de pacotes, jitter, largura de banda, disponibilidade, segurança, confiabilidade e throughput. Essas métricas são essenciais para garantir o desempenho, a estabilidade e a segurança das operações em redes de computadores (Ponce et al., 2023).

Koumar e Čejka (2022) destacam que a análise de tráfego de rede é comumente realizada por meio de séries temporais, com o objetivo principal de identificar e reconhecer padrões periódicos no tráfego dentro de uma infraestrutura monitorada. Em seus experimentos, os autores observaram que a comunicação periódica detectada e suas características podem ser utilizadas para classificar o tráfego e identificar o aplicativo, serviço ou dispositivo responsável pela comunicação. Com base na evolução das séries temporais, os autores definem a comunicação periódica de rede como a transferência repetitiva de pacotes que possuem o mesmo propósito.

O monitoramento de tráfego de redes pode ser categorizado em duas abordagens: ativa e passiva (Liu et al., 2014). Ambos os tipos de monitoramento são essenciais para

uma gestão abrangente de redes, cada um oferecendo benefícios específicos em termos de detecção de problemas, otimização do desempenho e análise do tráfego. A escolha entre monitoramento passivo e ativo depende das metas específicas e das características da infraestrutura de rede. Essa distinção oferece uma compreensão mais refinada das técnicas disponíveis para os administradores de redes, contribuindo para a eficácia do monitoramento e aprimoramento contínuo da infraestrutura (Hofstede et al., 2014).

Wassermann (2022) destaca que as medições ativas são realizadas por meio da inserção de pacotes na rede, a partir dos quais o desempenho pode ser inferido. Esse processo de injeção de pacotes e observação das respostas, embora útil para a solução de problemas, não é escalável (D'Alconzo et al., 2019). Segundo Abbasi et al. (2021), essa “sonda ativa” inserida na rede simula o tráfego real de rede, medindo o desempenho ponta a ponta. Avaliar o estado da rede ao simular diferentes cenários, utilizando "sensores" ou "agentes", permite a coleta de informações valiosas, como tempos de resposta, latência e outras métricas relevantes (Fowdur e Babooram, 2024).

Al-Sbou (2020) discorre sobre o funcionamento de medições ativas em seu estudo "Wireless Networks Performance Monitoring Based On Passive-Active Quality Of Service Measurements". De acordo com o autor, a estrutura básica de uma sondagem ativa ponta a ponta é composta por duas entidades principais: o remetente e o receptor da sonda (ou sensor). Em cada medição, o remetente gera e envia um fluxo de sondagem que percorre uma rota na rede até atingir o receptor, também conhecido como coletor.

O resultado bruto da medição é definido pelos números de sequência da sonda contidos nas cargas úteis, bem como pelos carimbos de data/hora de partida e chegada dos pacotes, registrados pelos monitores tanto do remetente quanto do receptor. Ao configurar características específicas no remetente, como o tamanho do pacote, o horário de partida e a taxa de bits, é possível calcular diversas métricas, analisando o comportamento do fluxo da sonda (como o horário de chegada) no destino, com o objetivo de determinar métricas de desempenho de ponta a ponta (da origem ao destino).

A qualidade de serviço (QoS) e o desempenho desses pacotes de sondagem são monitorados para inferir, de forma direta, o comportamento da rede e o desempenho dos pacotes dos usuários. Através das medições ativas, é possível determinar a QoS experimentada pelo fluxo de sondagem em um caminho específico, o que permite avaliar a QoS percebida pelos aplicativos. De acordo com Ponce et al. (2023) um exemplo desse tipo de medição é a determinação da capacidade máxima de carga da rede, medida através

do envio de pacotes com aumento progressivo da velocidade de transmissão até que a rede atinja o estado de saturação.

As medições ativas oferecem diversas vantagens, conforme discutido por Al-Sbou (2020). Uma das principais é a flexibilidade na criação de fluxos de sondagem com características específicas, adaptando-os às necessidades de cada medição. Esses fluxos podem variar em tamanho de pacotes, tipos de tráfego e intervalos de envio. Além disso, as medições ativas resultam em uma menor quantidade de dados coletados em comparação às medições passivas, otimizando o processo de análise.

Outra vantagem é que elas proporcionam aos provedores de rede controle total sobre o que será medido e quando realizar a medição, permitindo avaliações proativas dos níveis de serviço, até mesmo antes do início do tráfego real dos usuários. Além disso, esse método evita a necessidade de monitorar diretamente o tráfego de produção, garantindo conformidade com regulamentos de privacidade, que proíbem a inspeção ou duplicação do tráfego de usuários para análise posterior. Isso também elimina a preocupação com a criptografia do tráfego ou o uso de tunelamento (Clemm et al., 2020).

Em termos de desvantagens, Al-Sbou (2020) destaca que o principal problema das medições ativas reside em seu caráter invasivo. Os pacotes de sondagem utilizados nesse processo podem interferir na rede, afetando as métricas de QoS do tráfego real dos usuários. Clemm et al. (2020) acrescentam que a sobrecarga associada ao tráfego de teste sintético consome uma quantidade significativa de recursos durante o envio e recebimento de sondas, além de sobrecarregar os refletores de tráfego. Esse processo também pode consumir uma porção considerável da largura de banda disponível na rede.

Diversas ferramentas baseadas em métodos de medição ativa estão disponíveis, como as mensagens de solicitação e resposta de eco (ping) do Protocolo de Mensagens de Controle da Internet (ICMP), conforme especificado na RFC 729. Além disso, ferramentas como traceroute e Surveyor também se destacam nesse contexto (Al-Sbou, 2020).

As medições passivas têm sido amplamente utilizadas desde o início dos anos 2000, com ferramentas modernas capazes de processar múltiplos fluxos de dados na ordem de dezenas de Gb/s utilizando hardware comum (Trevisan et al., 2020). Conforme apontado por Fowdur e Babooram (2024), o monitoramento passivo é ideal quando se requer um histórico detalhado do comportamento de uma rede. Além disso, Liu et al. (2014) destacam que esse tipo de medição é amplamente empregado no gerenciamento e planejamento de redes comerciais.

A observação contínua do tráfego e a extração de métricas em tempo real são a base para o monitoramento passivo (D'Alconzo et al., 2019). Segundo Ponce et al. (2023), essa técnica é realizada ao monitorar o tráfego que flui pela rede sem interferir ou afetar seu funcionamento. Para Fowdur e Babooram (2024) o monitoramento passivo atua como um observador silencioso, registrando todas as atividades da rede sem adicionar volume aos dados existente.

De acordo com Liu et al. (2014), o monitoramento passivo e a análise em redes comerciais são viabilizados por dispositivos que capturam tráfego de alta velocidade, hardware com grande capacidade de armazenamento e servidores com alto desempenho computacional. Esses dispositivos de rede geralmente incorporam algum nível de inteligência para identificar e registrar os pacotes que passam por eles (Wassermann, 2022). Abbasi et al. (2021) apontam que essas sondas passivas são colocadas nos enlaces da rede, capturando todo o tráfego que transita pela conexão monitorada.

A medição passiva pode ser realizada por meio da coleta de dados de fluxo de tráfego de roteadores, switches ou hosts de ponto final. Um dos métodos para implementar essa coleta envolve a adição de um servidor autônomo na região de interesse da rede, como o núcleo ou a borda, que funciona como um medidor de tráfego ou dispositivo de monitoramento, armazenando dados sobre o tráfego passado. O nível de detalhe e precisão das informações coletadas nos pontos de medição varia conforme a quantidade de métricas processadas e o volume de tráfego que atravessa o dispositivo de monitoramento (Al-Sbou, 2020).

D’Alconzo et al. (2019), em “A Survey on Big Data for Network Traffic Monitoring and Analysis”, discutem o processo de medições passivas. Os autores explicam que, inicialmente, um exportador de rede captura o tráfego nos pontos de monitoramento, como roteadores que agregam múltiplos clientes. Esse exportador, então, envia uma cópia dos pacotes observados ou estatísticas de tráfego para um coletor. Os dados coletados são armazenados em formatos apropriados para serem utilizados por aplicativos de monitoramento e análise de tráfego. Esses aplicativos acessam os dados para gerar informações úteis. Além disso, os autores destacam que os componentes dessa arquitetura podem ser integrados em um único dispositivo, como um roteador ou monitor de rede dedicado, ou distribuídos em uma arquitetura mais ampla.

Conforme dissertado por Fowdur e Babooram (2024), a coleta de dados para monitoramento e análise de tráfego geralmente se baseia em três fontes principais: captura de pacotes, dados de fluxo e arquivos de log. No campo das técnicas de monitoramento

passivo, a captura de pacotes se destaca como uma das mais avançadas para essas atividades. Este método intercepta e registra pacotes de dados enquanto eles transitam pela rede, proporcionando um nível detalhado de análise, uma vez que captura cada transmissão individual. Essa técnica oferece uma visão abrangente do tráfego de rede, incluindo elementos como o conteúdo da carga útil, cabeçalhos e padrões de comunicação.

Por meio da captura de pacotes, os especialistas em rede têm acesso direto às informações contidas nos pacotes, desde os cabeçalhos até os dados transmitidos. Essa abordagem é crucial para identificar padrões de comunicação, avaliar o comportamento do fluxo de dados e, mais significativamente, detectar anomalias que possam indicar falhas de segurança ou irregularidades na conexão. Técnicas como a inspeção profunda de pacotes (DPI) são exemplos dessa abordagem, permitindo que os administradores de rede examinem o tráfego em sua forma original e não modificada.

A análise de pacotes, comumente chamada de "farejamento de pacotes", é realizada por meio de um farejador de pacotes, um utilitário projetado para capturar o tráfego bruto da rede. Existem diversas ferramentas disponíveis, que podem ser gratuitas ou comerciais, baseadas tanto em interface de linha de comando (CLI) quanto em interface gráfica do usuário (GUI) (Jain, 2022).

Abaixo estão algumas das ferramentas mais utilizadas para captura de pacotes:

- Hardware Probe: Dispositivo de rede capaz de capturar e analisar pacotes na camada física. As sondas de hardware fornecem dados mais detalhados, incluindo informações da camada física, sem sobrecarregar os recursos da rede. Embora seja mais eficiente em comparação com ferramentas baseadas em software, a solução de hardware tende a ser mais cara e menos flexível (Shen et al., 2022).
- Libpcap: Biblioteca popular de captura de pacotes, desenvolvida em C/C++, que opera na camada de enlace de dados, permitindo a visualização dos pacotes transmitidos pela rede. O formato de armazenamento dos pacotes capturados, o "pcap", é amplamente utilizado por diversas ferramentas de captura de tráfego, como Tcpdump e Wireshark (Shen et al., 2022).
- Omnipeek: Ferramenta comercial de análise de pacotes com interface gráfica (GUI), comercializada pela LiveAction, utilizada para captura e análise de tráfego de rede (Jain, 2022).

- Tcpdump: Ferramenta poderosa de captura e análise de pacotes baseada em CLI, amplamente utilizada e gratuita, funciona em sistemas operacionais baseados em UNIX, como Linux (Jain, 2022). Utiliza a biblioteca Libpcap para capturar e registrar pacotes em arquivos pcap, permite filtrar pacotes utilizando expressões regulares e ajustar parâmetros de captura, como a limitação do número de pacotes armazenados (Shen et al., 2022).
- Wireshark: Ferramenta gratuita e de código aberto para análise de pacotes, com interface gráfica, disponível para vários sistemas operacionais (Jain, 2022). Assim como o Tcpdump, permite capturar pacotes de acordo com parâmetros específicos do usuário. Além disso, oferece uma visualização mais rica, exibindo diferentes informações de protocolo com cores distintas. Comparado ao Tcpdump, oferece uma melhor experiência de usuário, embora consuma mais processamento e memória (Shen et al., 2022).

De acordo com Lee e Lee (2013), outras ferramentas de captura de pacotes incluem:

- Bro: Sistema de monitoramento de segurança de rede baseado na captura de pacotes.
- CoralReef: Desenvolvido pela CAIDA, oferece funcionalidades flexíveis para captura, análise e geração de relatórios de tráfego de rede.
- Snort: Ferramenta de detecção de intrusão em tempo real, de código aberto, que suporta análise de pacotes de rede.
- Tstat: Ferramenta de análise que disponibiliza funcionalidades avançadas para análise de métricas de desempenho TCP, classificação de aplicativos e características de VoIP.

A análise no nível de pacotes é fundamental para a resolução de problemas complexos de rede e para investigações forenses. No entanto, essa técnica pode gerar um grande volume de dados, o que exige consideráveis recursos de armazenamento e processamento. Ferramentas como Wireshark e tcpdump são amplamente utilizadas por administradores e analistas de rede para obter informações detalhadas e em tempo real sobre pacotes específicos. Contudo, o uso dessas ferramentas demanda uma quantidade significativa de recursos, incluindo bateria, processador e memória.

De acordo com Ehrlich et al. (2017) o aumento das velocidades de conexão e os elevados requisitos de processamento aliados a técnica de captura de pacotes, resultaram em volumes massivos de dados que se mostraram difíceis de gerenciar e analisar adequadamente. Hofstede et al. (2014) sugerem que uma alternativa mais escalável, adequada para redes de alta velocidade, é a exportação de fluxos. Nesse método, os pacotes são agrupados em fluxos e exportados para armazenamento e análise.

O monitoramento baseado em fluxo é amplamente reconhecido como a principal técnica de monitoramento passivo. Apenas um resumo de cada fluxo de tráfego é exportado para coletores, simplificando o processo de análise (D'Alconzo et al., 2019). Ponce et al. (2023) destacam que essa técnica é utilizada para medir o fluxo de dados na rede, contabilizando o número de pacotes e bytes transmitidos pelos dispositivos. Isso é realizado por meio de ferramentas especializadas que capturam pacotes e extraem informações como endereços IP de origem e destino, protocolo utilizado e tamanho dos pacotes. Essas ferramentas podem ser configuradas para reportar métricas específicas, como volume de tráfego, taxa de transmissão de dados e largura de banda consumida.

Tanto a captura de pacotes quanto o monitoramento de fluxo exigem sistemas robustos de armazenamento e servidores de alto desempenho para processar e analisar os dados provenientes de múltiplos dispositivos de rede. No nível de pacotes, as informações como tamanho em bytes e horários de chegada são coletadas para gerar relatórios estatísticos. Já no nível de fluxo, os dados fornecem detalhes sobre cada fluxo de comunicação, como sua duração e o número de pacotes envolvidos. Esses dados podem ser agregados e analisados sob diferentes perspectivas, gerando relatórios mais abrangentes que podem incluir informações por usuário ou classe de serviço (Liu et al., 2014).

Em relação às vantagens das técnicas de monitoramento passivo, Joshi et al. (2016) destacam que essas técnicas geram menor sobrecarga na rede, permitindo sua execução em segundo plano, sendo ideais para ativar ações de gerenciamento de rede. De acordo com Al-Sbou (2020), além da redução da sobrecarga, as medições passivas superam as limitações das medições ativas, especialmente no que se refere aos atrasos. Isso ocorre porque o monitoramento de fluxos em medições passivas oferece uma avaliação mais precisa do desempenho do tráfego do usuário em comparação às medições ativas.

Alkenani e Nassar (2022) realizaram um comparativo entre os dois tipos de monitoramento de tráfego, apontando que a medição passiva é mais indicada em cenários

que os locais de captura podem ser escolhidos livremente (permitindo a captura dos dados trafegados entre remetente e destinatário), e a medição ativa é mais indicada em cenários que não é possível escolher pontos de captura.

Autores como Ehrlich et al. (2017) ressaltam que as soluções de monitoramento passivo de rede se destacam por poderem ser implementadas amplamente sem a necessidade de hardware adicional, sendo aplicáveis a diversos domínios. Com o objetivo de obter uma visão ampla da rede sem a necessidade de altos investimentos em hardware, optou-se pela adoção de técnicas de monitoramento passivo neste trabalho. Considerando que o ambiente do servidor utilizado neste estudo opera em uma rede de alta velocidade, escolheu-se a exportação de fluxos como método, o qual será detalhado no próximo capítulo.

3 EXPORTAÇÃO DE FLUXOS

Um fluxo é definido como um conjunto de pacotes IP que atravessam um ponto de observação na rede durante um determinado período, sendo que todos os pacotes pertencentes a um fluxo compartilham um conjunto de propriedades comuns. Essas propriedades incluem, por exemplo, informações do cabeçalho do pacote, como os endereços de origem e destino ou os números de porta (Ehrlich et al., 2017).

De acordo com Fowdur e Babooram (2024), os fluxos representam sessões de comunicação unidirecional, permitindo que os administradores de rede tomem decisões mais embasadas ao considerar informações acumuladas em vez de analisar pacotes individualmente, o que proporciona uma visão mais abrangente do comportamento da rede. Koumar et al. (2023) destacam que, entre as informações e estatísticas agregadas sobre a comunicação, as métricas mais simples e comumente usadas incluem a contagem de pacotes e de bytes transmitidos. Esse tipo de representação de tráfego é suficientemente genérico para oferecer uma visão geral de redes de grande porte, com altos volumes de tráfego, incluindo o tráfego criptografado.

Os protocolos e tecnologias de exportação de fluxos, além de serem apropriados para redes de alta velocidade, apresentam diversas vantagens em relação à captura convencional de pacotes. Primeiramente, esses protocolos são amplamente utilizados, especialmente pela sua integração em dispositivos avançados de encaminhamento de pacotes, como roteadores, switches e firewalls. Outra vantagem é que o monitoramento de fluxos não exige a instalação de dispositivos adicionais para captura, tornando essa abordagem menos custosa em comparação com a captura tradicional de pacotes.

Embora a exportação de fluxos reduza consideravelmente o volume de dados a ser analisado em relação às técnicas de captura de pacotes, os repositórios de dados gerados ainda podem atingir volumes substanciais, chegando a dezenas de terabytes (Hofstede et al., 2014). Segundo Joshi et al. (2016), devido ao menor volume de dados processados e a ampla disponibilidade de ferramentas, a medição e análise de tráfego baseadas em fluxos são amplamente implementadas por provedores ISPs.

O monitoramento de fluxos envolve a coleta de estatísticas sobre os fluxos de dados utilizando os dispositivos que o atravessam. Essas estatísticas consolidam a atividade da rede em registros de fluxos, refletindo métricas como a frequência de pacotes transmitidos e recebidos por segundo e as variações na largura de banda (Fowdur e Babooram, 2024).

Autores como Ehrlich et al. (2017) explicam que os registros de fluxo representam a comunicação entre dois pontos, oferecendo uma visão detalhada do tráfego na rede. Esses registros fornecem informações sobre o estado dos dispositivos de rede e dos protocolos em diferentes camadas, bem como sobre o próprio tráfego que flui. De acordo com Hofstede et al. (2014), esses registros podem ser pensados como linhas em um banco de dados, cada uma representando um fluxo com suas propriedades específicas.

Para identificar fluxos individuais, os pacotes são classificados com base em sete características principais: endereço de origem, número da porta de origem, endereço de destino, número da porta de destino, tipo de protocolo, Tipo de Serviço (ToS) e interface lógica de entrada. Esses são conhecidos como "campos-chave".

Os caches de fluxo, que representam fluxos ativos em um determinado momento, agregam novos pacotes de entrada conforme suas propriedades correspondem aos campos-chave. Se houver uma correspondência, os contadores de pacotes e bytes do cache são atualizados; se não houver correspondência, um novo cache de fluxo é criado. Esses caches são exportados periodicamente em forma de registros de fluxo, e sua capacidade depende dos recursos de memória do sistema e das configurações de expiração.

O gerenciador de fluxo é o responsável por criar e atualizar os registros de fluxo no cache, enquanto o exportador de fluxo transmite os registros de fluxos finalizados, que são encerrados por pacotes TCP FIN ou eventos de timeout, em um formato binário predefinido (Liu et al., 2014). Ehrlich et al. (2017) também ressaltam que a análise dos dados é a fase final do processo de monitoramento de fluxo, e nessa fase ferramentas automatizadas são amplamente utilizadas para gerar relatórios, definir limites e configurar alertas e notificações.

Na fase de análise, as funções principais incluem correlação, agregação, classificação de tráfego, detecção de anomalias e buscas forenses. Todas as etapas do monitoramento de fluxo estão interligadas, e a compreensão completa de cada uma delas é essencial para garantir a precisão das medições. Do contrário, podem ocorrer perda de dados e artefatos sem serem detectados (Hofstede et al., 2014).

Lucas (2010) define que um típico sistema de gerenciamento baseado em fluxos compreende três elementos essenciais, um sensor (ou sensores), um coletor e um sistema de relatórios. O sensor, também conhecido como sonda, é o responsável por monitorar a rede, capturar e transmitir os dados de tráfego. Ele pode se manifestar como um dispositivo de hardware, como um switch, roteador ou firewall com a funcionalidade de

exportação de fluxos integrada, neste cenário encaminhar informações de fluxos para um servidor de gravação implica em uma sobrecarga mínima, sem a necessidade de instalar qualquer software adicional.

Alternativamente, o sensor pode ser um software designado para escutar um tap Ethernet ou uma porta de switch no modo monitor, porém na maioria dos casos, sensores já integrados em hardware são as opções mais adequadas e diretas. A implementação de sensores de fluxo pode ser considerada uma etapa desafiadora devido a evidente necessidade de uma abordagem criteriosa na escolha da distribuição dos sensores, especialmente em ambientes de rede geograficamente distribuídos. Ao planejar a localização para a captura dos dados, é crucial considerar o padrão de tráfego entre os dispositivos e configurar a exportação de fluxos exclusivamente nos pontos de "estrangulamento" centralizados na rede.

A maioria dos roteadores e switches de nível intermediário a avançado, possuem a capacidade de armazenar dados de fluxos, porém não oferecem uma interface direta para que os dados de fluxos sejam visualizados localmente por um usuário. Em vez disso, para examinar os registros de fluxos, é necessário exportá-los do hardware para um computador.

O coletor é o componente de software encarregado de receber os registros provenientes do sensor e armazená-los no disco. Ele representa o ponto focal para a maioria das atividades relacionadas a um sistema de gerenciamento de fluxos, devido desempenhar um papel central na recepção e processamento dos registros provenientes dos equipamentos de rede. Diante disso, é imperativo aplicar princípios sólidos de administração de sistemas no dispositivo coletor, embora haja uma ampla flexibilidade em relação ao hardware, sistema operacional e softwares que podem ser utilizados.

A prática da exportação de fluxos foi inicialmente implementada em roteadores com recursos de hardware bastante limitados. Em muitos desses dispositivos, à medida que as larguras de banda das interfaces aumentavam, o acompanhamento de cada pacote exigia mais capacidade do que o próprio roteador ou tap poderia oferecer. Diante disso, o hardware recorria à amostragem de pacotes para gerar dados de fluxos, capturando e exportando apenas uma parcela específica do tráfego que atravessava o dispositivo. Esses dados de fluxos eram, por natureza, incompletos.

Idealmente, é recomendável registrar todo o tráfego que percorre a rede, a amostragem só deve ser considerada quando o hardware não consegue lidar com o rastreamento completo do fluxo. Embora a posse de dados amostrados seja preferível à

ausência de dados, registrar o máximo de detalhes possível é consideravelmente mais vantajoso para a solução de problemas.

O registro de dados pelo coletor é iniciado apenas após o sensor começa a enviar informações. Quando os dados são recebidos pelo coletor, ele os organiza em um arquivo de log. Assim que uma instância de captura de fluxos está operacional, é preciso decidir como lidar com os dados recebidos. Uma opção é direcionar todos os sensores para alimentar dados para um único coletor, enquanto outra é permitir que cada sensor alimente dados para seu próprio coletor. Configurar todos os sensores para enviarem registros para um único coletor é uma abordagem simples, basta configurar um coletor único e não restringir os endereços que podem enviar dados para ele, com essa configuração, o coletor irá consolidar todos os registros de fluxos de todos os sensores em um único arquivo de log.

A menos que haja razões convincentes para fazer o contrário, é recomendável executar um coletor separado para cada sensor de fluxo, o que facilita a manutenção dos dados separados. Todas as instâncias de captura de fluxos podem ser executadas no mesmo servidor e utilizar o mesmo endereço IP. Cada processo de captura de fluxos pode ser atribuído a sua própria porta UDP e diretório de dados, permitindo a análise do tráfego de cada segmento de rede de forma independente.

Mesmo ao selecionar cuidadosamente os arquivos de fluxos para examinar e filtrar seu conteúdo, o resultado ainda é um vasto conjunto de dados para integrar, agregar e analisar, mesmo em redes de menor escala. Devido a isso, é necessária uma ferramenta capaz de agregar, classificar e apresentar dados de fluxos de forma cumulativa.

Além disso, é essencial que a ferramenta de análise seja personalizável para atender a diversas necessidades, seja para gerar relatórios instantâneos, explorar visualizações inesperadas do tráfego ou obter uma análise detalhada do histórico de tráfego. Nesse contexto, surge a necessidade de ferramentas de software especializadas para coletar, armazenar e facilitar a compreensão de fluxos. O sistema de relatórios é o responsável por extrair informações dos arquivos armazenados pelo coletor e gerar relatórios compreensíveis e úteis, é fundamental que ele trabalhe com o mesmo formato de arquivo utilizado pelo coletor.

No que diz respeito ao desempenho de sistemas de monitoramento de fluxos, Yahyaoui e Zhani (2020) destacam que o atraso na recepção de estatísticas pelo coletor é influenciado principalmente pelo intervalo de sondagem, que representa o tempo entre mensagens consecutivas de monitoramento, e pelo atraso de relatório, que corresponde

ao tempo necessário para que um pacote de monitoramento deixe o dispositivo gerador de estatísticas e percorra a rede até chegar ao coletor.

Os autores também ressaltam que para garantir estatísticas de fluxos precisas e em tempo hábil, tanto o intervalo de sondagem quanto o atraso de relatório devem ser mantidos em níveis mínimos. Embora um intervalo de sondagem reduzido aumente a precisão das estatísticas e permita capturar pequenas variações no tráfego, isso pode resultar em muitas mensagens de monitoramento sendo enviadas entre o dispositivo e o coletor. Essa situação apresenta problemas de escalabilidade, especialmente em cenários realistas onde milhões de fluxos precisam ser monitorados simultaneamente.

Além disso, quando o dispositivo monitorado está localizado a vários saltos de distância do coletor, as mensagens de monitoramento utilizam grandes quantidades de largura de banda ao longo do caminho, aumentando o atraso de relatório em situações com alta latência de propagação. Velan e Jirsik (2020) analisam em sua pesquisa como a configuração do sistema de monitoramento de fluxos influencia os dados resultantes. Eles enfatizam que as variáveis de configuração mais fundamentais são os tempos limite de expiração de fluxo, os quais afetam significativamente a análise de dados de rede.

Segundo os autores, o tempo limite de expiração influencia diretamente a criação de registros de fluxos e, por consequência, a maneira como os fluxos são descritos. E o tempo limite ativo afeta as métricas extraídas desses registros, como o número de pacotes, bytes e a duração da conexão. A escolha da configuração adequada é crucial, pois diferentes ajustes resultam em diferentes registros de fluxo para o mesmo tráfego.

Registros mais curtos, criados a partir de tempos limite menores, carregam mais detalhes, enquanto registros mais longos tendem a ser menos informativos. Em casos extremos, a criação de um registro para cada pacote maximiza o valor informativo, enquanto a criação de um único registro para cada conexão reduz a quantidade de informações. Ademais, o processamento de grandes volumes de registros de fluxos, especialmente quando se empregam algoritmos sofisticados de análise, pode ser computacionalmente dispendioso. Assim, a criação excessiva de registros de fluxos pode sobrecarregar o mecanismo de análise.

Além disso, os registros de fluxos precisam ser verificados periodicamente quanto à inatividade, de modo que possam ser exportados dentro de um prazo adequado. O RFC 5470 define três condições principais para a expiração de fluxo: tempo limite inativo, tempo limite ativo e restrições de recursos. Na prática, também se observam outros fatores que influenciam a expiração, como o término do fluxo e o desligamento do exportador.

O tempo limite inativo é um mecanismo primário que garante a remoção de registros de fluxos do cache após determinado período de inatividade. Um tempo limite inativo muito extenso faz com que os registros permaneçam por mais tempo no cache, consumindo mais recursos e atrasando o processamento. Por outro lado, um tempo limite inativo muito curto pode dividir fluxos extensos de forma desnecessária.

O tempo limite ativo, por sua vez, força a expiração regular de fluxos de longa duração. Ele garante que fluxos de execução prolongada sejam observados e processados em tempo hábil, evitando que fiquem invisíveis por muito tempo, o que poderia comprometer a segurança e a contabilização do tráfego. Portanto, a configuração de tempos limite exige um equilíbrio cuidadoso entre a completude das informações, a pontualidade da análise e a utilização eficiente de recursos.

De acordo com Lucas (2010) os sensores de fluxo realizam a exportação dos registros primordialmente quando a atividade de rede correspondente é finalizada ou quando o tempo limite estabelecido é alcançado. Quando uma conexão persiste até o tempo limite estabelecido, o dispositivo exporta um registro de fluxo e inicia um novo.

Os tempos limite são também aplicados na gestão de registros de fluxos para protocolos como UDP, ICMP e outros que não são baseados em TCP. Cada transação desencadeia a criação de um registro de fluxo, que será exportado quando o tempo limite expirar. Embora os dispositivos de rede não possam determinar precisamente o término de um fluxo UDP, o uso de tempos limite garante que o registro correspondente seja eventualmente exportado.

Diversos estudos tratam da exportação de fluxos e das ferramentas associadas a essa prática. Fowdur e Babooram (2024) destacam que as primeiras pesquisas sobre o tema remontam à década de 1990, estabelecendo as bases para protocolos modernos, como o NetFlow e o IPFIX. De acordo com Lee e Lee (2013), há diversas ferramentas de análise de fluxos, tanto de código aberto quanto comerciais, entre elas o flow-tools, flowscan, argus e Peakflow.

Segundo D’Alconzo et al. (2019), a Cisco estimou, em 2013, que aproximadamente 90% das análises de tráfego de rede eram baseadas em fluxos, enquanto apenas 10% utilizavam abordagens baseadas em pacotes. Em relação aos protocolos de análise de fluxos mais utilizados, Ehrlich et al. (2017) apontam que as tecnologias mais difundidas são as versões v5 e v9 do NetFlow, além da técnica mais recente IPFIX. No próximo capítulo, será explorado o protocolo NetFlow, tecnologia empregada pela ferramenta flow-tools, utilizada para este estudo.

4 NETFLOW

O NetFlow foi desenvolvido pela Cisco como uma técnica de cache flexível e leve para melhorar o desempenho de roteamento de seus dispositivos de rede e coletar informações estatísticas passivamente, somente posteriormente teve sua utilização voltada para o monitoramento de redes (Ehrlich et al., 2017).

Segundo Santos (2016), o Netflow originalmente atendia os propósitos específicos de contabilização do uso de rede e avaliação de diversos aspectos do tráfego IP, incluindo a utilização da largura de banda e o desempenho de aplicativos. Mistry et al. (2016) apontam que atualmente ele tem como propósito específico coletar e monitorar o tráfego de fluxo de rede gerado pelos roteadores e switches habilitados para NetFlow, e é usado principalmente para analisar o tráfego de rede, determinar o tráfego de entrada e saída, e a quantidade de tráfego sendo gerado.

Ao identificar os fluxos, o NetFlow é capaz de criar os registros de fluxo necessários, que são coleções de todos os pacotes contidos. Algumas vantagens do Netflow incluem a baixa ou inexistente necessidade de investimento de capital porque quase todos os dispositivos de rede de vários fornecedores já contêm recursos do NetFlow. E além das razões financeiras, o NetFlow é capaz de medir e relatar todo o tráfego do Protocolo de Internet (IP) de forma automática e dinâmica para diferentes tamanhos de rede (Ehrlich et al., 2017). Além disso, Splunder (2015) ressalta que outra vantagem é a simples configuração do Netflow devido sua coleta ser um recurso padrão de roteadores já usados em muitas redes.

Cisco Systems (2006) fornece alguns preceitos a respeito da implementação do Netflow, onde para configurar a captura e exportação de dados do NetFlow, as seguintes etapas são seguidas:

- Configurar o software coletor para receber os dados Netflow;
- Configurar o sensor NetFlow para capturar fluxos no cache do dispositivo;
- Configurar a exportação do dispositivo NetFlow para enviar os fluxos capturados ao coletor.

Após o fim das configurações, o cache do dispositivo NetFlow verifica constantemente se há fluxos finalizados, os fluxos finalizados são agrupados e transportados para o servidor coletor NetFlow, então o software do coletor NetFlow gera relatórios conforme solicitado pelo usuário.

Segundo a fabricante, existem duas abordagens principais para acessar os dados do NetFlow: a utilização da Interface de Linha de Comando (CLI) do próprio software coletor, que exibe comandos diretos, ou o emprego de uma aplicação de relatórios. Para uma visão imediata das atividades na rede, a CLI é preferível, e é especialmente útil para solução de problemas. Em contrapartida, a utilização de ferramentas de relatórios oferece visuais mais fáceis de serem compreendidos e melhor usabilidade para usuários com menos conhecimento técnico.

Em relação a visualização dos dados de fluxos, Tremel et al. (2022) apontam que o NetFlow fornece entradas para sistemas de monitoramento de rede que normalmente mostram visualizações de séries temporais ao longo de diferentes dimensões de dados. A quantidade de dados NetFlow coletados e processados é enorme, mesmo para redes de médio porte e curtos períodos. Portanto é comum a presença de abordagens de visualização interativa em produtos comerciais para análise de dados NetFlow.

Essas abordagens normalmente empregam estratégias diferentes para lidar com grandes volumes de dados, como limitar a análise interativa a curtos períodos. Isso permite visualizar os dados em detalhes e evita problemas de escalabilidade durante a exploração interativa. Outras abordagens focam em uma tarefa específica, como a identificação de ameaças à segurança. Ao investigar um incidente de segurança específico, a filtragem para os dispositivos relevantes também reduz a quantidade de dados que precisam ser processados para análise interativa. Como os padrões de tráfego diários são bastante regulares em redes de grandes empresas, desvios desse comportamento regular são bons pontos de partida para exploração e inspeção detalhada.

Em geral, o protocolo NetFlow está disponível em várias versões, começando com a v1 até a versão atual v9, fornecendo diferentes funcionalidades e dando suporte a vários casos de uso ou aplicativos. As versões v5 e v9 são as mais utilizadas, algumas versões intermediárias nem foram lançadas ou dificilmente são usadas em aplicativos reais (Ehrlich et al., 2017). Santos (2016) descreve em “Network Security with NetFlow and IPFIX” as diferentes versões do Netflow e suas respectivas características:

- Versão 1 - A primeira versão do NetFlow é restrita a redes IPv4 e não incluía máscaras de rede IP, nem números de Sistema Autônomo (ASN’s). Esta versão se tornou obsoleta;
- Versão 2 - Embora tenha sido desenvolvida, a versão 2 do NetFlow nunca foi lançada oficialmente;

- Versão 3 - Da mesma forma que a versão 2, o NetFlow v3 também não foi lançado;
- Versão 4 - Assim como as duas versões anteriores, o NetFlow v4 não chegou a ser lançado;
- Versão 5- Esta versão do NetFlow foi amplamente adotada em muitos roteadores de diversos fabricantes. No entanto, é limitada a fluxos IPv4;
- Versão 6 - Considerada obsoleta e não mais suportada pela Cisco;
- Versão 7 - Similar à versão 5, mas com a adição de um campo de roteador de origem. Também obsoleta;
- Versão 8 - Embora oferecesse várias formas de agregação, estava limitado a informações já presentes nos registros da versão 5. Também considerada obsoleta;
- Versão 9 - Esta é a versão mais utilizada atualmente, ficou disponível a partir de 2009, e apresenta uma mudança na forma em que o protocolo opera devido possuir a utilização de modelos como base de seu funcionamento. Os modelos oferecem uma estrutura flexível para o formato dos registros, possibilitando futuras modificações e aprimoramentos nos serviços sem a necessidade de alterações substanciais no formato dos registros de fluxo utilizados anteriores. O Netflow v9 é usado principalmente para relatar fluxos IPv6, MPLS ou até mesmo IPv4 com BGP no próximo salto.

A partir dos anos 2000, um grupo de trabalho foi definido com o propósito de definir padrões de formatos de fluxos e evitar uma maior fragmentação nessa área. Este grupo desenvolveu um protocolo padrão de fluxo de rede com o nome de IP Flow Information eXport (IPFIX), que é baseado na versão 9 do Netflow, mas com ajustes mínimos visando aprimorar sua acessibilidade. Embora a Cisco continue envolvida nesse processo, sua influência não é mais exclusiva, sendo agora apenas um dos membros do grupo (Lucas, 2010).

De acordo com Santos (2016), o IPFIX é um protocolo liderado pela Internet Engineering Task Force (IETF), destinado a estabelecer uma norma universal e consistente para a exportação de informações de fluxos provenientes de dispositivos de rede. Em essência, tanto o NetFlow versão 9 quanto o IPFIX se baseiam em princípios

similares para a exportação de informações de fluxo, embora possam apresentar algumas alterações terminológicas.

O cenário de softwares open-source que possuem suporte ao Netflow oferece uma gama diversificada de ferramentas, dentre elas coletores como:

- flow-tools;
- flowd;
- IPFlow;
- SiLK.

E ferramentas de análise e visualização de dados como:

- Cflowd;
- FlowViewer;
- NFdump;
- NFsen;
- Stager;
- iSiLK;
- ELK.

O autor também relata algumas práticas recomendadas e diretrizes gerais para a preparação e implementação do NetFlow em uma organização, sendo essencial realizar um planejamento minucioso ao ativar o NetFlow em áreas críticas da rede. Uma opção é iniciar a ativação do NetFlow em determinadas regiões da rede para compreender o impacto preliminar antes de expandir a implementação, outra possibilidade é utilizar ferramentas disponíveis para ajudar a prever o impacto da ativação do NetFlow na infraestrutura de rede.

É recomendado habilitar o dispositivo NetFlow na proximidade imediata da camada de acesso, seja na camada de acesso do usuário, na camada de acesso ao data center, nos pontos terminais VPN, ou outros pontos de entrada. Também é possível habilitar o NetFlow em dispositivos de rede em todas as camadas, permitindo registrar e examinar abrangentemente o tráfego de rede. Outra recomendação é que a coleta de dados do NetFlow seja realizada o mais próximo possível do ponto de origem do tráfego para reduzir a utilização da largura de banda e minimizar o impacto da sobrecarga na rede.

Segundo diretrizes da fabricante Cisco Systems (2006), a localização do NetFlow depende da localização da ferramenta de relatórios utilizada e da topologia da rede, certos sistemas de relatórios adotam uma arquitetura de dois níveis, caracterizada pela

distribuição de coletores em pontos estratégicos da rede, responsáveis por consolidar e encaminhar dados para um servidor central de relatórios. Por outro lado, implementações menores podem optar por um único servidor responsável tanto pela geração de relatórios quanto pela coleta de dados.

De acordo com Santos (2016), no modelo Netflow de implantação distribuída, os coletores são dispersos em múltiplos pontos da rede, geralmente posicionados próximos às fontes que geram o maior volume de registros NetFlow. Essa estratégia visa reduzir a carga imposta pela coleta de NetFlow, distribuindo os coletores de forma estratégica ao longo da infraestrutura. No modelo de implantação centralizada todos os coletores NetFlow são concentrados em um único local. Essa abordagem oferece a vantagem de centralizar a coleta de dados, possivelmente utilizando um único endereço IP global para a coleta de NetFlow. Essa estratégia é mais adequada para ambientes onde os geradores NetFlow estão distantes entre si.

Em relação a ambientes onde o roteamento é assimétrico, todos os dispositivos ao longo de uma rota assimétrica devem encaminhar os registros NetFlow para o mesmo coletor, garantindo a uniformidade na coleta de dados, evitando a dispersão dos registros para coletores distintos. Ao lidar com múltiplos pontos geograficamente distribuídos, é crucial considerar as restrições de largura de banda entre esses pontos.

Recomenda-se, como boa prática, utilizar um único coletor para capturar o máximo de tráfego relacionado possível. No entanto, os ganhos associados à centralização da coleta diminuem quando o tráfego em questão não é uniforme ou similar entre os pontos. Não é aconselhável coletar dados NetFlow em conexões de rede de longa distância (WAN), devido às possíveis limitações de largura de banda entre os diferentes locais. Portanto, é essencial planejar com antecedência e identificar os pontos de monitoramento que são mais relevantes para o ambiente específico de cada organização.

Uma das fases cruciais no planejamento e desenho das implementações de NetFlow é a avaliação e mensuração do volume de fluxos por segundo (fps) que serão gerados nos pontos de monitoramento. O volume, ou fps, indica a capacidade necessária dos coletores para receber e processar os registros. Diversos elementos impactam o volume de registros NetFlow produzidos pelos dispositivos de infraestrutura de rede, devido a isso estimar um número exato pode ser desafiador.

Geralmente, um dispositivo NetFlow pode gerar entre 1.000 e 5.000 fps para cada gigabit por segundo (Gbps) de tráfego que atravessa o dispositivo. O volume de fps é influenciado pela quantidade de fluxos únicos que transitam pelo dispositivo NetFlow,

número de novas conexões estabelecidas por segundo e duração de cada fluxo individual. A carga adicional imposta à rede pela atividade do NetFlow, também é afetada pelo número de fps e pelo tamanho dos registros NetFlow. A Cisco recomenda a utilização do NetFlow v9, o qual resulta em uma média de 34 registros NetFlow por pacote de 1.500 bytes.

Em relação ao uso de amostragem aleatória no NetFlow, o número de pacotes ou o nível de "aleatoriedade" é ajustável pelo usuário. Essa abordagem estatística de amostragem reduz significativamente a carga sobre a CPU e outros recursos, enquanto ainda fornece dados úteis de NetFlow. A amostragem aleatória é frequentemente utilizada em engenharia de tráfego, planejamento de capacidade e outras aplicações em que uma visão completa do tráfego de rede não é essencial.

Além de recomendações para uma eficiente implementação do Netflow, Santos (2016) também aborda exemplos práticos da utilização do Netflow. Segundo o autor, o NetFlow desempenha um papel fundamental na segurança da rede. Dados do Netflow podem ser integrados com dados de telemetria de rede durante investigações de incidentes de segurança e análises forenses de rede. Esses dados podem incluir:

- Registros do protocolo DHCP;
- Registros de conexões VPN;
- Informações de tradução de endereço de rede (NAT);
- Registros de autenticação 802.1x;
- Registros do servidor (syslog);
- Registros de acesso de proxy da web.

Os registros históricos fornecidos pelo NetFlow podem também revelar a origem de uma infecção, identificar o canal de comando e controle utilizado por um malware, indicar quais outros dispositivos na rede interna foram acessados pelo host infectado e determinar se outros hosts na rede foram comprometidos pelo mesmo atacante ou sistema de comando e controle.

O NetFlow também constitui uma ferramenta valiosa para supervisionar atividades de usuários convidados e terceirizados na rede. Utilizar análises nos dados obtidos através do NetFlow pode também ajudar os especialistas em segurança a identificar grandes quantidades de dados incomuns que deixam a organização, além de identificar padrões de tráfego anômalos dentro dela. O NetFlow pode ser empregado em

conjunto com registros DNS para auxiliar na detecção de tráfego potencialmente suspeito e malicioso, como:

- Requisições inexplicáveis a domínios como .gov, .mil e .edu, mesmo sem qualquer interação prévia com essas entidades;
- Elevada atividade de saída da rede durante as horas noturnas direcionada a sites de procedência duvidosa;
- Fluxo direcionado a países sob embargo (os quais não possuem relações comerciais ou transações previstas);
- Solicitações e tráfegos suspeitos relacionados a redes privadas virtuais (VPN);
- Interações com sites desprovidos de conteúdo legítimo ou transações significativas;
- Acesso a páginas que violam a política corporativa e utilização de sites ilegais para compartilhamento de arquivos.

Outras áreas em que o Netflow também atua são a engenharia de tráfego e o planejamento de redes, permitindo uma compreensão detalhada dos padrões de tráfego que podem ser ajustados e otimizados na infraestrutura da rede. Isso inclui a identificação de tráfego menos crítico que consome recursos de rede, como downloads de arquivos de mídia pelos usuários finais ou acesso a plataformas de redes sociais como Facebook e Twitter.

Ele também pode revelar problemas de configuração na infraestrutura de rede, possibilitando a identificação de gargalos e a otimização do desempenho. Para provedores de serviços, a correlação entre NetFlow e protocolos como BGP pode melhorar a eficiência do peering e fornecer insights valiosos sobre a rentabilidade, custos e possíveis violações de tráfego IP. Da mesma forma, em redes corporativas, a correlação com protocolos de roteamento interno, como OSPF, EIGRP e RIP, pode oferecer benefícios semelhantes na otimização da infraestrutura de rede.

Os administradores de rede podem também empregar o NetFlow como recurso de planejamento de capacidade, facilitando ajustes e previsões de demanda futura de forma mais eficiente. Eles enfrentam o desafio contínuo de monitorar o tráfego de rede para garantir o funcionamento ininterrupto das operações essenciais, através da coleta de dados do NetFlow em dispositivos de infraestrutura, é possível identificar:

- Serviços em nuvem acessados pelos usuários da rede;

- Quantidade de endereços IP únicos em uso;
- Volume de tráfego direcionado a esses provedores de serviços em nuvem e seu impacto na largura de banda por aplicativo na nuvem;
- Presença de uso não autorizado de protocolos e aplicativos.

No próximo capítulo será abordada a arquitetura do servidor Netflow implementado, onde serão tratados o cenário da rede do provedor, e as especificações de hardware e softwares utilizados.

5 ARQUITETURA DO SISTEMA

No que tange à seleção de ferramentas para um sistema de monitoramento de fluxos, Fowdur e Babooram (2024) discutem as vantagens e desvantagens entre soluções comerciais e de código aberto. Segundo os autores, a escolha entre esses dois tipos de solução deve ser feita com cautela, considerando as necessidades específicas de cada organização, suas limitações orçamentárias e os recursos disponíveis.

Soluções comerciais geralmente envolvem custos financeiros, mas oferecem uma vasta gama de recursos, suporte técnico robusto e desempenho confiável. Essas ferramentas são particularmente adequadas para grandes empresas com arquiteturas de rede complexas, que exigem funcionalidades avançadas, alta escalabilidade e suporte especializado. Além disso, sistemas comerciais frequentemente dispõem de interfaces intuitivas, atendimento ao cliente dedicado e um conjunto abrangente de funcionalidades, o que os torna atraentes para organizações com grandes exigências de monitoramento.

Por outro lado, soluções de código aberto proporcionam flexibilidade, personalização e economia, pois são oferecidas gratuitamente. Pequenas empresas, instituições de ensino e usuários tecnicamente proficientes, que possuem as habilidades necessárias para configurar e manter essas tecnologias, tendem a preferir essas alternativas. Soluções de código aberto também fomentam a criação de uma comunidade colaborativa, permitindo que os usuários participem ativamente no desenvolvimento e adaptação do software conforme suas necessidades específicas.

Manohar (2020) destaca que no monitoramento de redes, os aplicativos de código aberto são bem construídos e fornecem uma base sólida para futuras modificações. Entretanto, o autor adverte que, por serem amplamente acessíveis, esses sistemas podem ser explorados por indivíduos mal-intencionados, o que pode comprometer sua confiabilidade.

Em “Network Flow Analysis”, Lucas (2010) oferece diretrizes para a construção de um sistema de gerenciamento de rede baseado em fluxos, onde são utilizados um S.O gratuito do tipo Unix, softwares open-source, e hardwares de rede já existentes, proporcionando uma implementação econômica. Neste trabalho foram utilizados o mesmo sistema e conjunto de softwares abordados pelo autor, e hardwares de rede já existentes no provedor em que foram implementados.

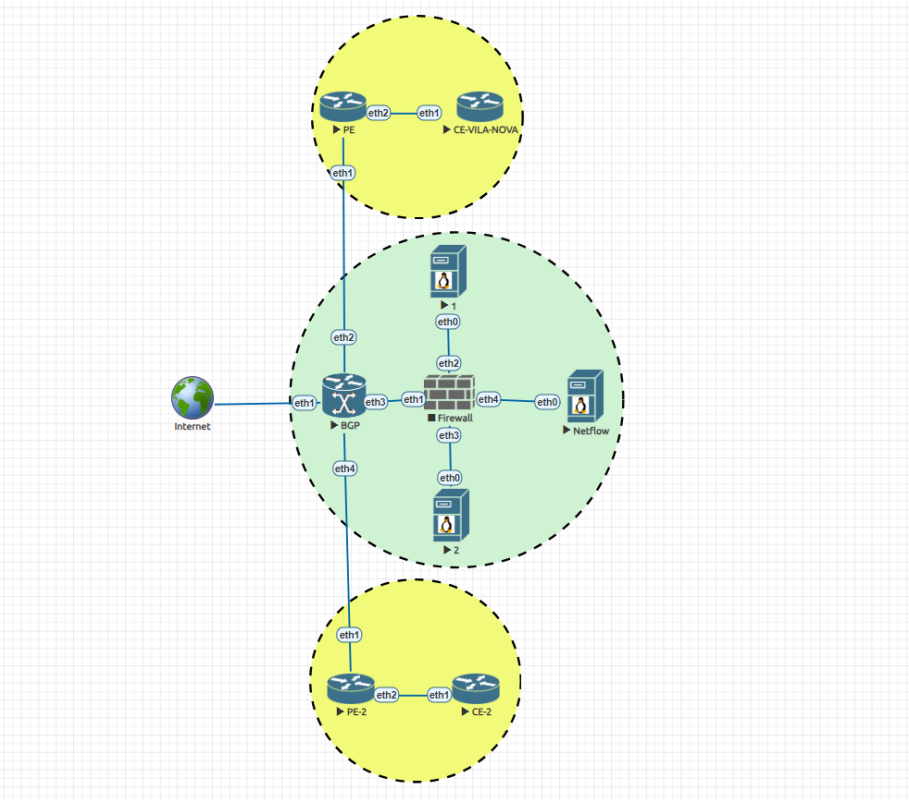
A implementação do servidor de monitoramento e análise de fluxos foi realizada em um provedor de internet que atua na região nordeste do estado do Pará, com apoio

logístico e material da equipe de engenharia da empresa, responsáveis por habilitar a exportação de fluxos em um dos roteadores da rede e conceder o acesso em uma máquina virtual para implantação do servidor.

Para o experimento foi optado por utilizar um roteador com baixo tráfego em comparação com os demais da rede do provedor, evitando assim possíveis maiores problemas de sobrecarga na própria rede ou no servidor. A ativação da exportação de fluxos foi realizada em um roteador Mikrotik RB750Gr3, nomeado “CE-VILA-NOVA”, equipamento responsável pela autenticação da conexão e tráfego de internet dos clientes da região de "Vila Nova”, localizada na cidade de São João da Ponta.

A figura 5.1 representa a arquitetura de redes do provedor, em que o círculo verde representa o POP central, onde fica o BGP, equipamento responsável pela gerência de todos os dispositivos da rede do ISP e conexão direta com os links de internet. O roteador disponibilizado para exportação de fluxos está localizado geograficamente fora do POP central, especificamente próximo aos clientes finais atendidos por ele. No entanto, mesmo que geograficamente distante, em termos de arquitetura de rede, apenas um salto separa o roteador CE-VILA-NOVA do BGP.

Figura 5.1 - Representação da arquitetura de redes do provedor



Fonte: O autor

As máquinas que abrigam os servidores responsáveis por diversos serviços do provedor, incluindo a que abriga o servidor de monitoramento Netflow, estão localizadas no POP central e conectadas diretamente no dispositivo de Firewall da empresa, equipamento responsável pela segurança dos serviços utilizados no provedor e que estabelece a conexão direta dos servidores com o BGP. Em relação a arquitetura do próprio servidor de monitoramento e análise de fluxos, a abordagem utilizada é simples, onde o software coletor e o software utilizado para análise de dados ocupam a mesma máquina virtual.

Em relação ao sistema e softwares empregados, Lucas (2010) utiliza como referência em seu trabalho o sistema operacional FreeBSD, o kit de ferramentas flow-tools, responsável pela coleta dos fluxos, e a ferramenta FlowViewer, responsável pela emissão de relatórios.

Em relação à escolha do sistema operacional utilizado, o autor aponta que praticamente qualquer sistema operacional moderno do tipo Unix atenderá aos requisitos mínimos do coletor, recomendando preferencialmente um sistema operacional BSD, embora alternativas como Linux, OpenSolaris, ou outros sistemas Unix padrão de 32 ou 64 bits, equipados com um compilador GCC recente e bibliotecas atualizadas, também sejam adequados. No entanto, é prudente optar por sistemas operacionais conhecidos por sua compatibilidade e estabilidade, minimizando assim possíveis complicações com o coletor e o sistema de relatórios.

Outra observação feita pelo autor é que alguns sistemas operacionais comerciais do tipo Unix podem apresentar comportamentos singulares que, embora tecnicamente estejam em conformidade com os padrões, podem divergir de maneira peculiar ou imprevista dos demais sistemas Unix. Portanto, é sensato escolher uma opção conhecida por sua interoperabilidade. Em termos de segurança, é crucial que o dispositivo coletor seja dedicado exclusivamente à função de gerenciamento de fluxos, evitando assim a prestação de outros serviços que possam comprometer a segurança do sistema

O FreeBSD é uma plataforma totalmente gratuita e de código livre, desenvolvida e mantida através da colaboração de uma grande comunidade de usuários. Ele é derivado do sistema BSD (uma das versões do UNIX®), e tem como foco oferecer recursos, velocidade e estabilidade aos seus usuários, sendo ideal para utilização em servidores de Internet ou Intranet devido aos robustos serviços de redes que fornece, sendo capaz de lidar com altas cargas de dados, usando a memória de maneira efetiva para trabalhar com milhares de processos simultâneos. Atualmente o sistema conta com mais de 33.000

bibliotecas e aplicativos, oferecendo suporte para desktops, servidores, dispositivos e ambientes integrados (FreeBSD, 2024).

A versão do sistema utilizada na implementação do servidor foi a 13.0, a mais recente atualmente disponível no próprio site da organização, o FreeBSD foi instalado em uma máquina virtual com as configurações da tabela 5.1.

Tabela 5.1 - Configurações de hardware do servidor

Configurações de hardware	
Sistema Operacional	FreeBSD 13.0 (amd64)
Processador	Intel(R) Xeon (R) CPU E7440 @ 2.40 GHz
Núcleos CPU	8
Processadores Lógicos	16
Memória RAM	8 GB
Disco Rígido	100 GB

Fonte: O autor

Segundo Lucas (2010), há uma diversidade de implementações disponíveis para cada componente de um sistema de gerenciamento baseado em fluxos. No entanto, muitos dos softwares disponíveis são desatualizados, embora ainda sejam recomendados em antigos fóruns online, e muitas combinações de software e hardware podem ser sutilmente incompatíveis. Dentro da variedade de coletores de fluxos disponíveis gratuitamente, destacam-se o cflowd, flowd e flow-tools. Entretanto, o cflowd está desatualizado e não pode ser compilado em sistemas de 64 bits, e o flowd é uma ferramenta que não possui amplo suporte entre usuários ou software de terceiros.

O flow-tools, software utilizado neste trabalho, é o kit de ferramentas de gerenciamento de fluxos mais amplamente utilizado, ele foi desenvolvido para sistemas do tipo Unix, é compatível com boa parte dos sensores, e diversas ferramentas de relatórios utilizam seu formato de dados. Grande parte de seu código utiliza a linguagem Perl, para sua utilização é recomendado conhecimento mínimo em Perl, para que seja possível editar os scripts necessários para sua configuração e utilização.

Embora tenha sido lançado inicialmente por Mark Fullmer em 2005, uma comunidade de usuários assumiu a responsabilidade pelo desenvolvimento do flow-tools em 2007, esses usuários continuaram a aprimorar e corrigir bugs, lançando versões

atualizadas conforme necessário. Embora possa haver bugs ocasionais, como em qualquer software, o flow-tools possui uma base de usuários significativa.

Além do utilitário responsável pela captura e armazenamento dos fluxos chamado flow-capture, o kit flow-tools fornece ferramentas que podem ser usadas em conjunto para visualização dos dados e geração de relatórios em formato de texto. Dentre essas ferramentas, destacam-se o flow-print, flow-cat, flow-nfilter e flow-report. O flow-print é o responsável por exibir os dados de forma "crua" no terminal, o flow-cat concatena arquivos de fluxo para uma visualização mais polida, o flow-nfilter permite a utilização de filtros de forma dinâmica para que dados mais direcionados possam ser exibidos, e o flow-report exibe relatórios baseados em indicadores pré-configurados.

O flow-cat pode ser utilizado para combinar dados que serão exibidos em utilização conjunta com o flow-print, a filtragem de dados pode ser feita no meio deste processo, bastando utilizar o flow-nfilter em conjunto, onde os dados repassados pelo flow-cat serão filtrados e então repassados ao flow-print para exibição. O flow-report é muito útil nos casos em que os mesmos relatórios precisam ser gerados repetidas vezes, facilitando acesso à dados agregados, classificados e resumidos.

Embora dados em texto exibidos na interface de linha de comando sejam muito úteis para administradores de redes, pessoas com menos conhecimento técnico podem não interpretar facilmente os dados apresentados desta maneira, o que torna vantajoso possuir uma ferramenta de relatórios onde os dados possam ser apresentados através de gráficos. A linguagem Perl é muito utilizada para criar ferramentas de relatórios e interfaces web, dentre elas, estão o FlowScan e FlowViewer. O FlowScan possui uma implementação mais difícil de ser realizada em comparação ao FlowViewer, que porventura, é uma ferramenta mais personalizável, e permite que relatórios baseados na web sejam criados e acessados rapidamente (Lucas, 2010).

O FlowViewer é uma solução de código aberto, inicialmente desenvolvida para as redes do Sistema de Dados e Informações de Ciências da Terra (ESDIS) da NASA, ele proporciona uma interface de usuário web (HTML/CSS) adequada para conjuntos de ferramentas de fluxo de dados, incluindo o flow-tools, e a partir da versão 4.0, o SiLK. A ferramenta oferece suporte total às versões 5 e 9 do Netflow, e sua integração ao SiLK permite aos usuários que a utilizem em conjunto com o protocolo de dados de fluxo mais recente, o IPFIX.

Algumas funcionalidades principais do FlowViewer incluem a visualização contínua do tráfego de rede através de Dashboards, representação gráfica de conjuntos de

tráfego filtrados em períodos específicos, rastreamento a longo prazo e em segundo plano de tráfego filtrado, armazenamento de filtros e relatórios para referência futura, e alertas por e-mail para situações anormais de tráfego.

A estrutura do FlowViewer está dividida em três componentes principais, chamados FlowViewer, FlowGrapher e FlowMonitor. Os três utilitários são configurados por meio de um arquivo de configuração compartilhado, e são capazes de utilizar a interface web para coletar informações, e aplicar filtros aos dados de fluxo capturados e armazenados pelo coletor que residem no mesmo host. Os componentes do FlowViewer oferecem também recursos adicionais de gráficos e monitoramento, utilizando softwares de código aberto, como gd de Thomas Boutrell, GD de Lincoln Stein, GD::Graph de Martien Verbruggen e pacotes RRDtool de Tobias Oetiker (FlowViewer, 2024).

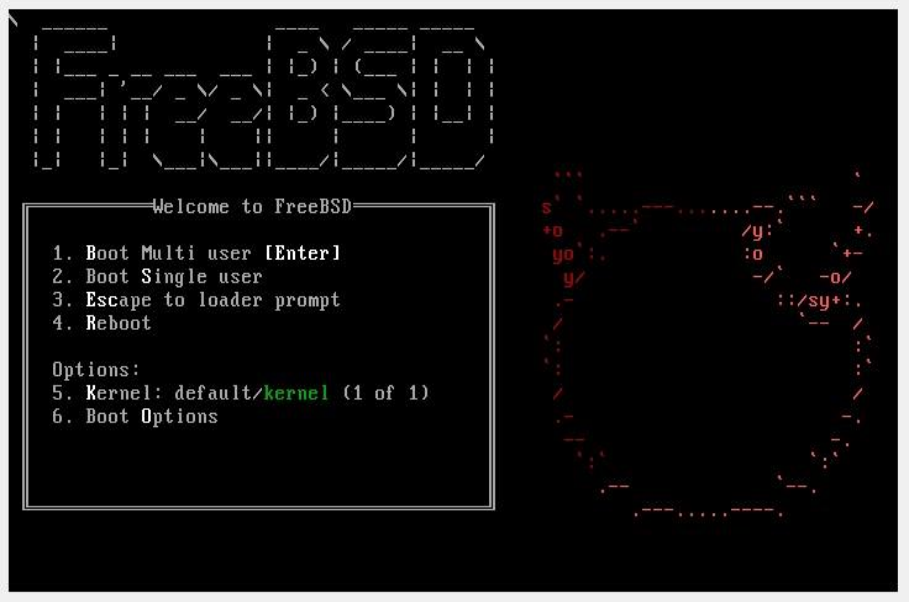
De acordo com Lucas (2010), o utilitário FlowViewer serve como uma interface web para o flow-print e flow-nfilter do conjunto flow-tools, possibilitando a definição de filtros por meio de um formulário web para exibir os fluxos correspondentes. O FlowGrapher permite que os subconjuntos de dados filtrados sejam representados graficamente. Por fim o FlowMonitor possibilita o monitoramento contínuo de tráfego, através da construção e visualização de gráficos RRD (dados round-robin), permitindo por exemplo monitorar continuamente o tráfego de um protocolo específico para um servidor específico.

O uso combinado dos três utilitários abordados acima proporciona uma visibilidade ampla da rede no FlowViewer, permitindo uma análise mais aprofundada do tráfego. Para que os dados de tráfego possam ser visualizados no FlowViewer, é necessário inicialmente configurar o servidor responsável pela coleta dessas informações. No capítulo seguinte, será descrito o processo de implementação do flow-tools, responsável pela captura dos fluxos, em uma máquina virtual com o sistema operacional FreeBSD.

6 IMPLANTAÇÃO DO FLOW-TOOLS

O primeiro passo para que se possa coletar e armazenar dados de fluxos é a devida implementação do coletor, assim que o coletor estiver apto, o sensor pode ser habilitado para que os dados sejam capturados e enviados ao coletor. A instalação do sistema operacional foi realizada utilizando um arquivo ISO, obtido através do site oficial da própria organização, a figura 6.1 exibe a tela inicial de instalação do FreeBSD. Após a finalização da instalação, o sistema foi preparado para que as ferramentas utilizadas na coleta e tratamento de fluxos sejam instaladas.

Figura 6.1 - Tela inicial de instalação do sistema FreeBSD



Fonte: O autor

Para agilizar o processo de instalação das ferramentas a serem utilizadas e suas respectivas dependências, foi feito login com o usuário "root" no sistema, evitando assim a necessidade de dar permissão para cada ação a ser realizada futuramente. Alguns programas necessários para auxiliar no processo de implementação do servidor foram instalados através dos repositórios do próprio FreeBSD. O primeiro programa instalado foi o wget, através dele é possível adquirir arquivos por meio da internet, também foi instalado o editor de textos vi e o interpretador bash, que foi definido como shell padrão do sistema.

Por fim foi instalado o programa tcpdump, utilizado para capturar pacotes trafegados em uma rede, sendo útil para testar o funcionamento da captura de fluxos no servidor. A sincronização de data e hora do sistema é essencial para o pleno funcionamento do servidor de captura e interpretação de fluxos, pois qualquer falha de sincronização pode ocasionar erros na aquisição ou leitura dos dados. Através do comando "date" é possível verificar e caso necessário corrigir as atuais data e hora do sistema.

A instalação do flow-tools pode ser feita por meio de pacotes disponíveis no próprio sistema operacional ou no servidor do fornecedor. A maioria dos sistemas Unix oferece versões pré-compiladas do software gratuitamente, no entanto, é importante observar que muitos sistemas operacionais incluem apenas a versão 0.68 do flow-tools, em vez da versão mais recente com correções de bugs. Alguns sistemas operacionais, como o FreeBSD, fornecem o software mais atualizado em um pacote denominado flow-tools-ng (Lucas, 2010).

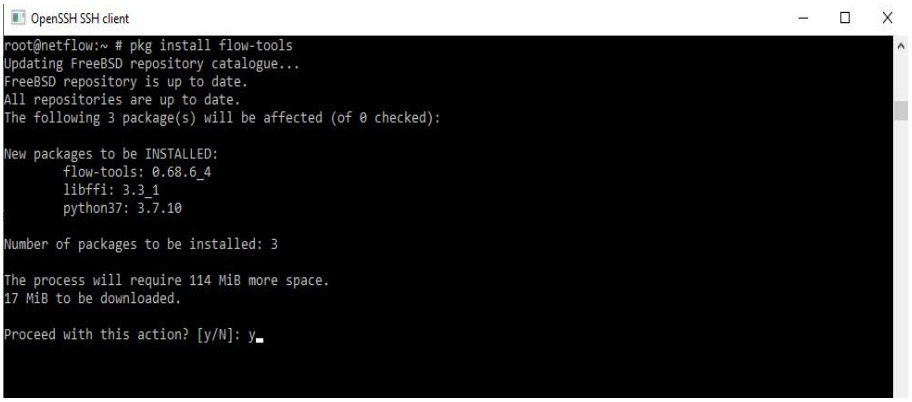
Uma das dependências para a utilização da ferramenta flow-tools é a biblioteca GCC, um conjunto de compiladores responsável pela interpretação de diversos comandos do programa, ele foi instalado através dos repositórios do próprio sistema operacional. Outro passo importante na preparação do servidor é a criação do diretório onde os fluxos serão armazenados, a figura 6.2 exibe a criação do diretório “flows”.

Figura 6.2 - Criação do diretório para armazenamento dos fluxos

Fonte: O autor

Após a instalação da biblioteca GCC e criação da pasta onde os fluxos serão armazenados, o flow-tools foi instalado através do repositório de dados do FreeBSD, conforme exibe a figura 6.3.

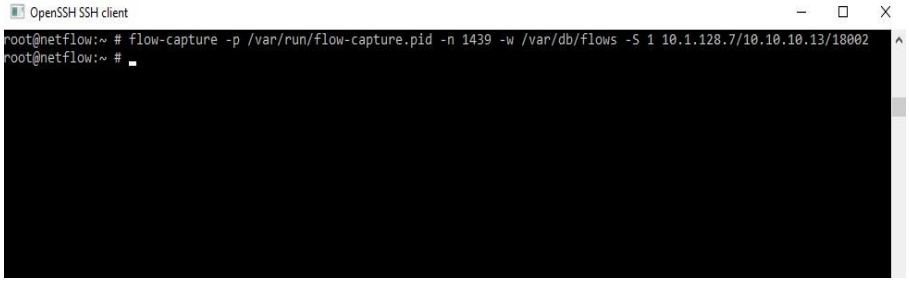
Figura 6.3 – Instalação do flow-tools



Fonte: O autor

O flow-tools é um kit de ferramentas onde cada utilitário possui uma função específica nos processos de captura, armazenamento, exportação e leitura de fluxos. O flow-capture, um dos utilitários do kit, é o responsável por capturar e armazenar dados Netflow no disco. O seguinte comando foi utilizado para ativar a captura de dados realizada pelo flow-capture: "flow-capture - p /var/run/flow-capture.pid -n 1439 -w /var/db/flows -S 1 10.1.128.7/10.10.10.13/18002". A figura 6.4 exibe a utilização do comando para captura dos fluxos no servidor.

Figura 6.4 – Comando responsável pela captura de fluxos



Fonte: O autor

No comando acima, o argumento "-p" tem como função informar para o flow-capture onde os arquivos PID (IDs dos processos) devem ser armazenados, neste cenário foi apontado o caminho "/var/run/flow-capture.pid".

O argumento "-n" tem como função informar ao flow-capture a quantidade de logs que devem ser gerados em um período de 24 horas. Ao iniciar seu processo, o flow-capture cria um arquivo de log e então rotaciona para a criação de um novo arquivo na quantidade de vezes informada no argumento, o valor 1439 resulta na criação de um log

a cada minuto, gerando um total de 1440 arquivos em um dia, mesma quantidade de minutos que possui um dia.

A variável “-w” diz ao flow-capture onde armazenar os arquivos dos fluxos, os arquivos foram armazenados no diretório "/var/db/flows".

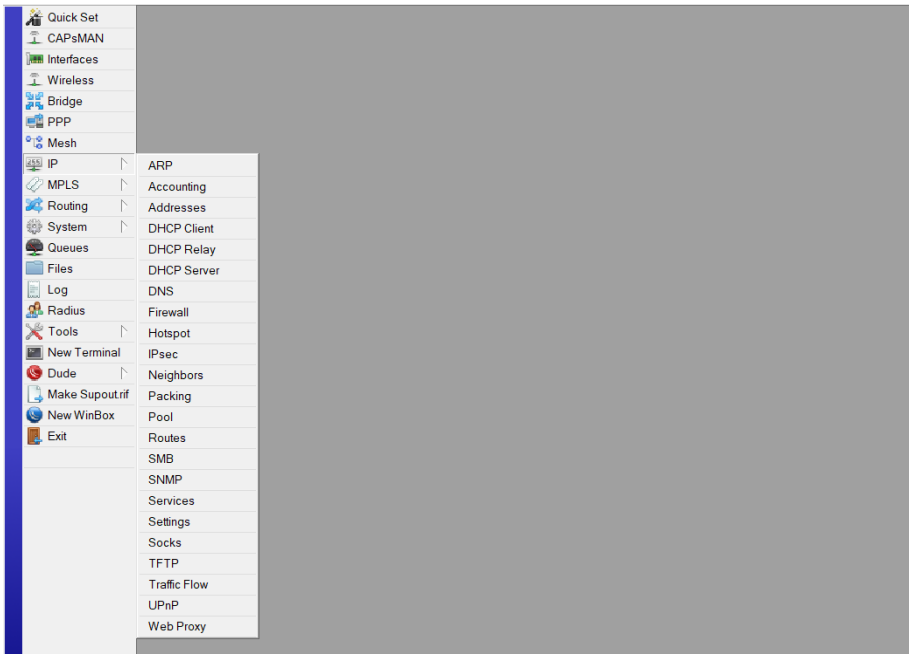
A instrução “-S” diz ao flow-capture para registrar mensagens no syslog, informando quantos fluxos foram processados e abandonados, e quantos pacotes foram recebidos, o valor “1” indica que as mensagens de log serão registradas a cada um minuto.

O último argumento do comando, “10.1.128.7/10.10.10.13/18002”, representa a configuração de rede do flow-capture. O primeiro endereço “10.1.128.7”, é o IP da interface onde o flow-capture realizará a escuta, neste cenário o IP da própria máquina virtual. O segundo endereço “10.10.10.13”, é o IP do sensor que enviará dados ao coletor, se o valor neste espaço for ’0’ ou vazio, a captura irá aceitar dados de qualquer endereço IP, aumentando o risco de algum intruso enviar dados falsos para o coletor, mas também permitindo que sejam aceitos fluxos de fontes diferentes simultaneamente. Por fim, o último atributo é o número da porta UDP por onde será realizada a escuta. De acordo com Santos (2016), o Netflow não possui um padrão de porta especificado em sua documentação (RFC 3954), a porta utilizada nesta implementação foi a “18002”.

Mesmo após sua ativação, o coletor não registrará nada até que um sensor comece a enviar dados, a configuração em roteadores da marca Mikrotik pode ser realizada através da interface de linha de comando do próprio equipamento ou através do utilitário de configuração da própria fornecedora chamado Winbox, que oferece uma interface gráfica simples e de fácil compreensão. Neste trabalho foi optado por ativar a exportação de fluxos através do Winbox, para configurar o tráfego de fluxos na interface gráfica basta selecionar a opção de configuração “IP”, e em seguida “Traffic Flow”, conforme exibido na figura 6.5.

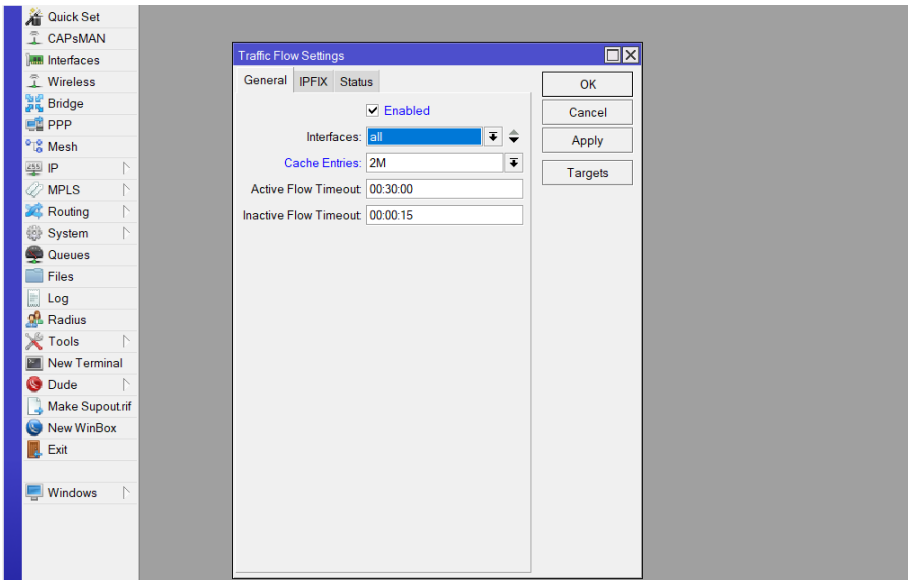
Para habilitar o tráfego de fluxos basta marcar a caixa apresentada na tela de configuração, conforme a figura 6.6, onde também é possível configurar qual interface do equipamento será monitorada e exportará fluxos, a quantidade de memória cache utilizada, o tempo máximo para exportação de fluxos de longa duração (padrão de 30 minutos), e o tempo máximo para exportação de fluxos inativos (padrão de 15 segundos).

Figura 6.5 – Tela inicial do Winbox para habilitar o tráfego de fluxo



Fonte: O autor

Figura 6.6 – Tela do Winbox com a ativação do tráfego de fluxo

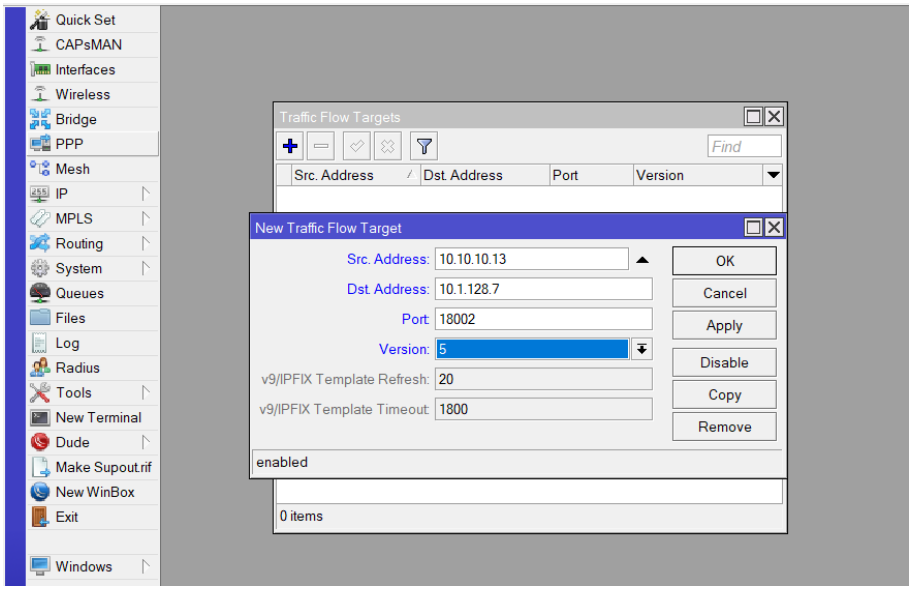


Fonte: O autor

Após ativar o tráfego de fluxos no roteador, é necessário configurar os “alvos” com a configuração de rede para onde o tráfego será exportado, selecionando o botão “Targets”. A figura 6.7 exibe a configuração de um novo alvo, onde é preciso informar o IP de origem, que neste ambiente se refere ao endereço do próprio roteador, o IP de

destino, que neste caso é o endereço da máquina virtual onde o servidor coletor está hospedado. É preciso informar também a porta UDP que será utilizada na comunicação entre roteador e servidor, e o número da versão do Netflow que será utilizada, neste cenário será a versão 5, devido ser a mais recente utilizada pelo flow-tools.

Figura 6.7 – Tela do Winbox com configuração da exportação de fluxo



Fonte: O autor

Após a configuração e ativação da exportação de fluxos, todos os fluxos gerados pelo roteador já devem ser capturados e registrados pelo flow-capture no servidor. Para verificar se os dados enviados pelo sensor estão chegando ao servidor, foi utilizada a ferramenta tcpdump, responsável por capturar os pacotes trafegados pela interface de rede do servidor, possibilitando averiguar se os pacotes enviados pelo roteador com direção ao servidor estão chegando em sua interface, para isso, foi utilizado o seguinte comando: "tcpdump -p -n -i em0 udp and port 18002".

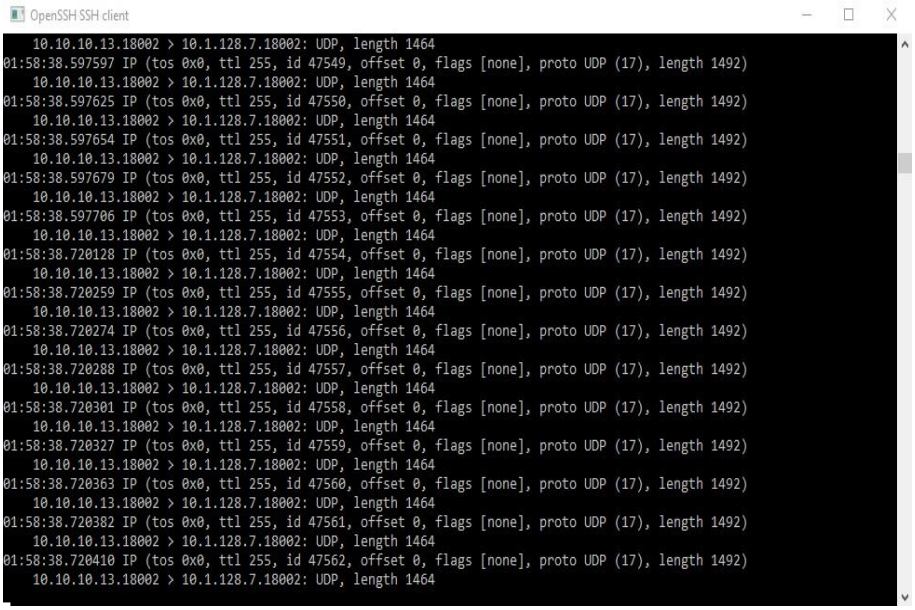
A figura 6.8 exibe o comando para dar início na captura de pacotes, onde o argumento “-p” desabilita o "promiscuous mode", para que o tcpdump rastreie apenas o tráfego da interface indicada, o argumento “-i” indica o nome da interface que será ouvida, neste cenário a interface chamada "em0", por fim deve ser informado o protocolo e porta utilizados no coletor, neste caso "udp" e "180002" respectivamente.

Figura 6.8 – Comando responsável pela captura de pacotes do tcpdump

Fonte: O autor

O resultado da captura de dados feita pelo tcpdump é mostrado na figura 6.9, onde é exibido o tráfego capturado na porta UDP 18002. Todos os pacotes capturados possuem como origem o endereço IP do roteador com o Netflow habilitado, comprovando que a comunicação entre roteador e servidor foi estabelecida com sucesso.

Figura 6.9 – Tráfego na porta 18002 capturado pelo tcpdump



Fonte: O autor

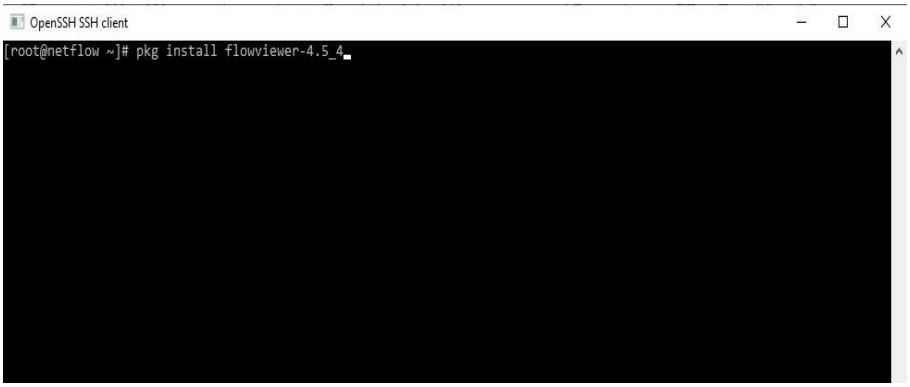
Após garantir que o servidor está recebendo os dados transmitidos pelo sensor, também foi confirmado que o coletor está realizando o registro dos fluxos, através da visualização dos arquivos salvos na pasta configurada para armazená-los. Os dados de fluxo podem então ser explorados, para posterior análise e geração de relatórios, no entanto, os arquivos de fluxo contêm dados binários que mesmo compactados ainda são

muitos extensos, e sua visualização sem a utilização das ferramentas corretas pode resultar em uma série de informações confusas no terminal. Então foi feita a instalação da ferramenta de relatórios para visualizar os dados de forma mais efetiva, o próximo capítulo abordará a implementação do FlowViewer.

7 IMPLANTAÇÃO DO FLOWVIEWER

Para que a ferramenta de emissão de relatórios FlowViewer seja implementada, é necessário que alguns pacotes de software sejam instalados como pré-requisitos, dentre eles, um servidor Web atual como o Apache, e a biblioteca Perl, responsável pela interpretação dos comandos utilizados na ferramenta. Os demais pacotes são as bibliotecas e módulos específicos para geração de gráficos, RRDtool, GD Graphics Library, gd::graph, e GDBM (Lucas, 2010). Os pré-requisitos e posteriormente o próprio FlowViewer foram instalados no sistema através do repositório de dados, a figura 7.1 exibe o comando para a instalação do FlowViewer.

Figura 7.1 – Comando para instalação do FlowViewer



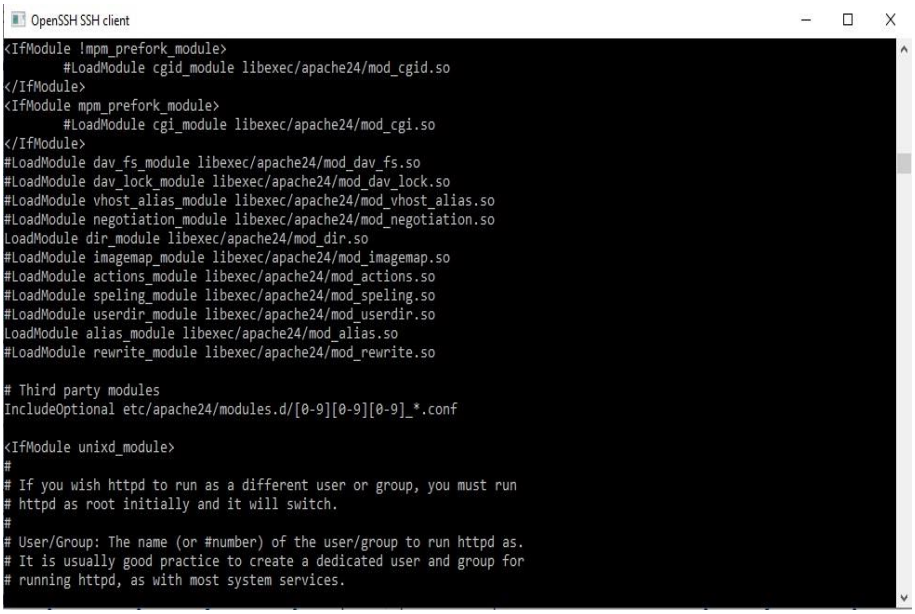
Fonte: O autor

Após a instalação do FlowViewer, é preciso configurar o servidor web utilizado, que nesta implementação foi o Apache. A configuração dele é realizada através da edição do arquivo httpd.conf, localizado na pasta do próprio servidor web. O FlowViewer utiliza a execução de scripts CGI nos conteúdos de suas páginas WEB, e para que isso seja possível no servidor web Apache, é necessário ativar os módulos CGI em seu arquivo de configuração, removendo da seção de comentários do script original as duas linhas com o código abaixo:

"#LoadModule cgid_module libexec/apache24/mod_cgid.so".

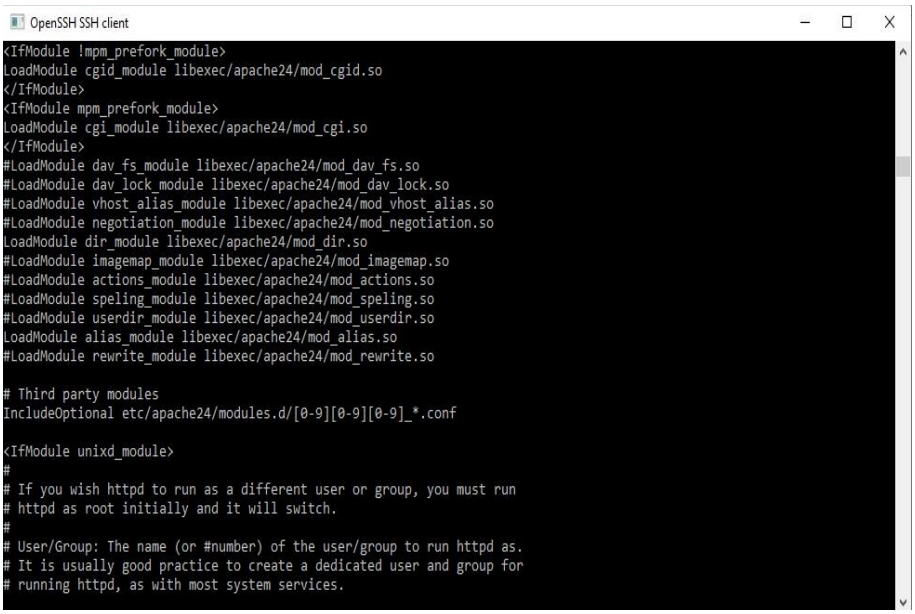
Para remover as linhas da sessão de comentários do script, basta remover o "#" do seu início. As figuras 7.2 e 7.3 exibem respectivamente o script do arquivo de configuração do Apache antes e depois da edição realizada para ativação dos módulos CGI.

Figura 7.2 – Arquivo de configuração do Apache com módulos CGI desativados



Fonte: O autor

Figura 7.3 – Arquivo de configuração do Apache com módulos CGI ativados



Fonte: O autor

Após a ativação dos módulos CGI, é preciso configurar o nome do servidor, adicionando uma linha com o início "ServerName", seguida do nome desejado, neste cenário foi alocado o endereço IP do servidor como o seu nome, deixando a linha desta maneira:

"ServerName 10.1.128.7".

Após configurar o nome do servidor, foi configurado o “apelido” para o caminho de seu diretório, possibilitando realizar acesso web ao diretório do FlowViewer com mais praticidade. O “apelido” pode ser configurado adicionando uma linha com o comando “Alias”, seguido do apelido desejado, e depois o caminho do diretório do FlowViewer entre aspas, nesta implementação a linha do script ficou desta forma:

“Alias /FlowViewer/ "/usr/local/www/flowviewer/"”.

É necessário também conceder ao servidor web permissões para a execução dos scripts CGI presentes no diretório do FlowViewer, adicionando as linhas abaixo em seu script de configuração:

“<Directory "/usr/local/www/flowviewer/»

Options +ExecCGI

AddHandler cgi-script .cgi

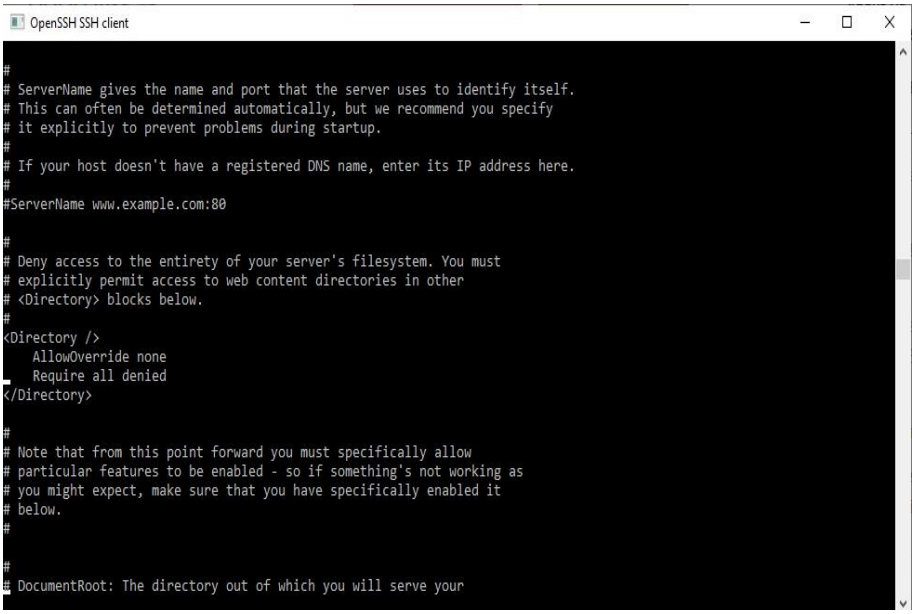
AllowOverride All

Require all granted

</Directory>”.

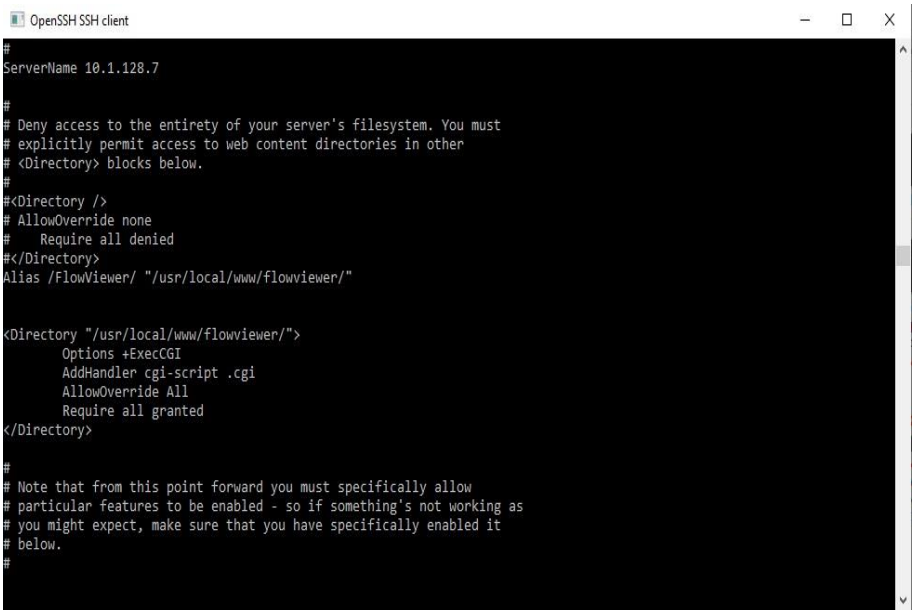
A figura 7.4 exibe o script de configuração do Apache antes das alterações de nome, apelido e permissões de execução dos scripts CGI do FlowViewer. A figura 7.5 exibe o script de configuração do Apache após as alterações.

Figura 7.4 – Arquivo de configuração do Apache sem dados do FlowViewer



Fonte: O autor

Figura 7.5 – Arquivo de configuração do Apache com dados do FlowViewer



Fonte: O autor

Após finalizar a configuração do servidor Web, o Apache foi definido como proprietário do diretório onde o FlowViewer foi instalado, evitando assim futuros problemas de permissão entre o servidor Web e o FlowViewer. Conforme mostra a figura 7.6, este foi o comando utilizado:

"chown -R www /usr/local/www/".

Figura 7.6 – Definindo o Apache como proprietário do diretório do FlowViewer

Fonte: O autor

Após finalizar as configurações necessárias para o acesso web do FlowViewer, o serviço Apache foi reiniciado para que as alterações fossem validadas e o acesso

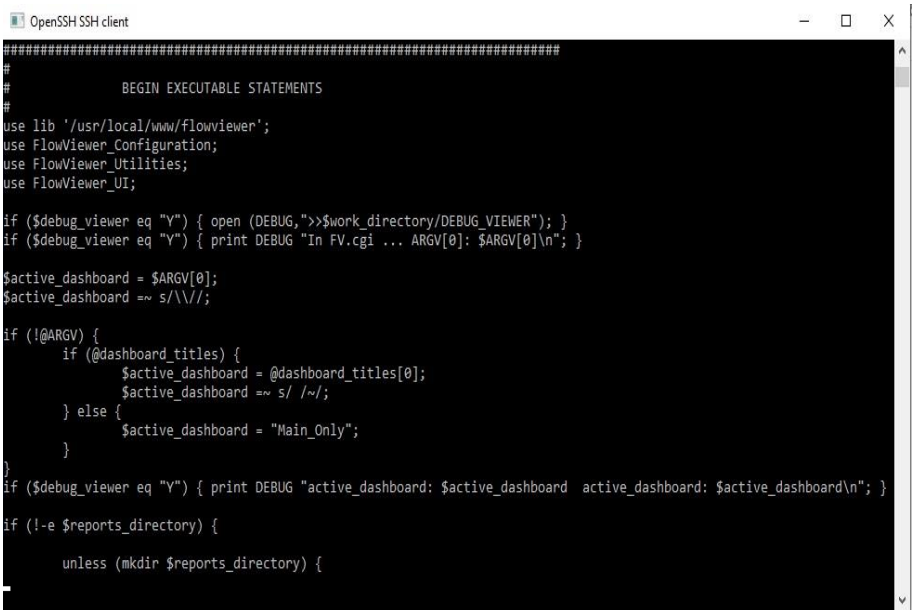
devidamente habilitado. No entanto, ao tentar acessar a página Web do servidor, ela permaneceu inacessível.

Analisando os registros de log do Apache e do próprio sistema operacional, foram notados erros que ocorreram devido o script FV.cgi presente na pasta do FlowViewer não estar conseguindo localizar no sistema os arquivos apontados dentro do próprio código. Este problema foi solucionado editando o script FV.cgi, com uma linha de script que obriga o código a utilizar os arquivos localizados na pasta do FlowViewer. Após a sessão de comentários do início do script, foi adicionada a linha de código abaixo:

"use lib ’/usr/local/www/flowviewer’".

Esta linha de código é a responsável por forçar o script a utilizar os arquivos apontados na pasta do FlowViewer. Na figura 7.7 é exibido o script do arquivo FV.cgi após a adição da linha de código.

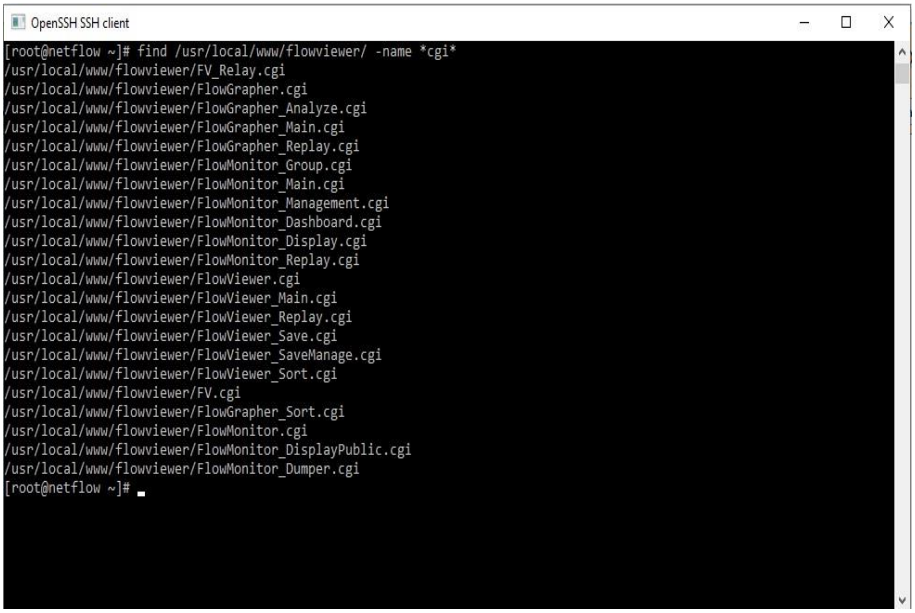
Figura 7.7 – Arquivo FV.cgi após edição



Fonte: O autor

A mesma linha de código adicionada no arquivo FV.cgi, também foi acrescentada em todos os demais scripts CGI do FlowViewer, como meio de precaução para que futuros problemas em suas respectivas execuções sejam evitados. A figura 7.8 exibe a lista de todos os scripts CGI presentes no diretório do FlowViewer.

Figura 7.8 – Lista de scripts CGI presentes no FlowViewer



Fonte: O autor

Após finalizar a edição dos scripts CGI, foi parametrizado o arquivo “FlowViewer_Configuration.pm”. Para evitar erros em sua execução, foi adicionada a linha usada nos arquivos CGI para apontar o caminho do FlowViewer, em seguida foi definido o nome do servidor, que neste caso foi o seu próprio IP. Então foram revisados e atualizados os diretórios presentes no script, a figura 7.9 exibe o script após alterações.

Figura 7.9 – Arquivo de configuração do FlowViewer após alterações

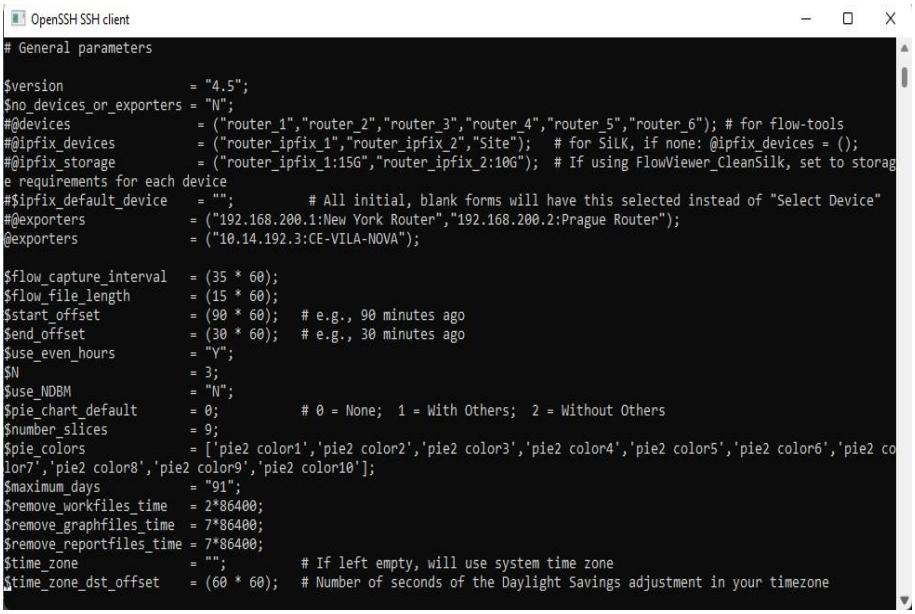
Fonte: O autor

Após confirmar que não há mais diretórios com caminhos incorretos no código, enfim foi realizada a configuração do sensor responsável pela transmissão dos fluxos no script do FlowViewer. Para isso, foi adicionada no script a linha abaixo:

"@exporters = ("10.10.10.13:CE-VILA-NOVA");".

Esta linha de código registra um exportador de fluxos, recebendo como parâmetros o endereço IP e o nome do dispositivo, respectivamente. A figura 7.10 exibe a configuração do sensor exportador de fluxos no script do FlowViewer.

Figura 7.10 – Configuração do sensor exportador de fluxos no script do FlowViewer

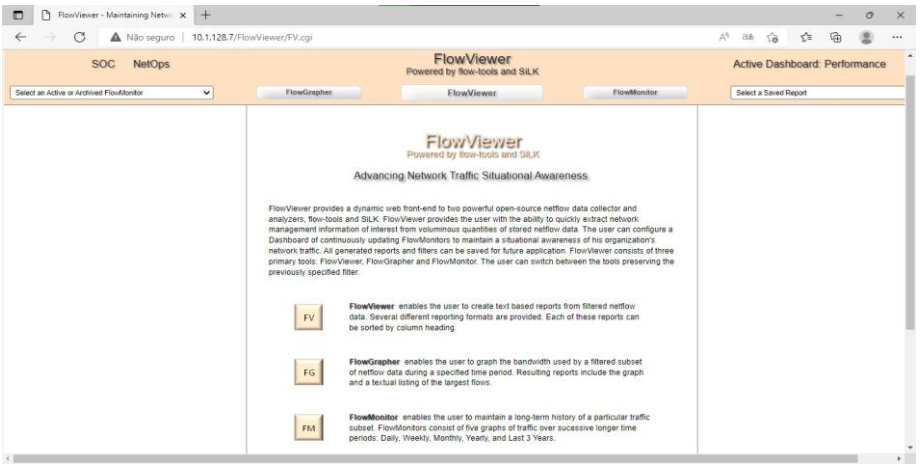


Fonte: O autor

Após finalizar a edição do arquivo de configuração do FlowViewer, ele está apto para ler os dados de fluxos armazenados pelo coletor. Conforme configurado no servidor web, o acesso web ao FlowViewer será realizado através do IP do servidor, seguido do caminho onde está o arquivo CGI responsável pela página inicial da interface web da ferramenta ("10.1.128.7/FlowViewer/FV.cgi").

Vale ressaltar que o domínio de acesso pode ser configurado conforme preferência do administrador, bastando alterar as configurações no arquivo de configuração do servidor apache. A figura 7.11 exibe a interface web do FlowViewer após sua implementação.

Figura 7.11 – Página inicial da interface web do FlowViewer



Fonte: O autor

No próximo capítulo será abordada a análise do tráfego feita através dos relatórios emitidos pelo FlowViewer, com foco na identificação de padrões através de atributos como volume de tráfego, taxas de transmissão e portas utilizadas.

8 ANÁLISE DO TRÁFEGO

Para constatar a eficácia prática das ferramentas implementadas, foram coletados fluxos do dispositivo nomeado "CE-VILA-NOVA”. Os fluxos coletados e analisados neste trabalho pertencem ao período com início às 00:00 horas do dia 16/09/2021, e fim às 00:00 horas do dia 23/09/2021, totalizando em uma semana completa de dados colhidos e analisados. A partir destes fluxos, foram realizadas as seguintes análises:

- Total de tráfego do roteador em todo o período;
- Distribuição da direção do tráfego (tráfego de entrada e saída para a internet, e tráfego na rede interna) em todo o período;
- Distribuição por serviços do tráfego entre roteador e servidores de distribuição de conteúdo do provedor em todo o período;
- Distribuição da direção (entrada e saída) e por serviços do tráfego entre roteador e servidores de distribuição de conteúdo do provedor em todo o período;
- Taxas de transmissão diárias máxima, média e mínima;
- Horários diários com maiores e menores taxas de transmissão;
- TOP 10 portas de origem, destino e geral com mais tráfego em todo o período;
- TOP 10 portas de origem, destino e geral mais presentes em fluxos em todo o período.

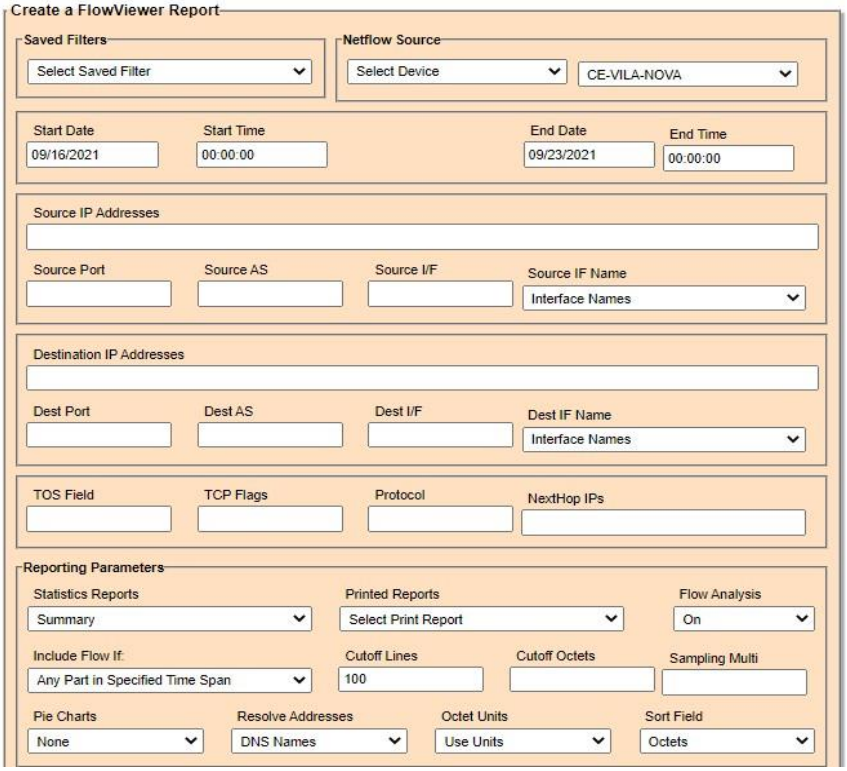
Azab et al. (2024) afirmam que a extração de características a partir de fluxos de rede gera atributos como o número total de pacotes, total de bytes e a duração dos fluxos. Diversos estudos utilizaram esses dados para analisar o tráfego de redes. Benes et al. (2023), por exemplo, realizaram uma análise empírica desses componentes transmitidos, com o objetivo de identificar o volume total de tráfego em diferentes intervalos de tempo. Além disso, parâmetros relevantes como as taxas de upload e download podem ser derivados desses atributos. Segundo Fowdur e Babooram (2024), a identificação dessas taxas é essencial, pois afetam diretamente a Qualidade de Serviço (QoS).

Para averiguar a quantidade de tráfego total em todo o período foi utilizado o componente "FlowViewer” da ferramenta de relatórios, nele foram definidos como entradas os parâmetros: exportador (selecionando o já configurado "CE-VILA-NOVA"),

data inicial, hora inicial, data final e hora final. Definindo o período de 00:00 horas do dia 16/09/2021 à 00:00 horas do dia 23/09/2021, totalizando os 7 dias completos de fluxos coletados.

Para visualizar um "resumo" dos dados coletados no período definido e consequentemente a contabilidade total do tráfego, foi utilizada a opção de relatórios "summary", o restante das opções de filtros e relatórios da interface permaneceram no padrão. Na figura 8.1 é exibida a tela da interface do componente FlowViewer, e sua configuração de filtros e relatórios para obtenção do relatório de dados resumidos.

Figura 8.1 – Configuração de filtros e relatórios para relatório resumido do tráfego



Fonte: O autor

A figura 8.2 mostra o resultado do relatório solicitado acima. Onde exibe o número total de fluxos, octetos e pacotes trafegados, além das informações de duração e taxas médias de transmissão no período selecionado. Segundo o relatório resumido, foram transmitidos pelo roteador 19718434642338 octetos, ou cerca de 17,93 Terabytes nos 7 dias de fluxos coletados, o que representa o volume total do tráfego do roteador CE-VILA-NOVA.

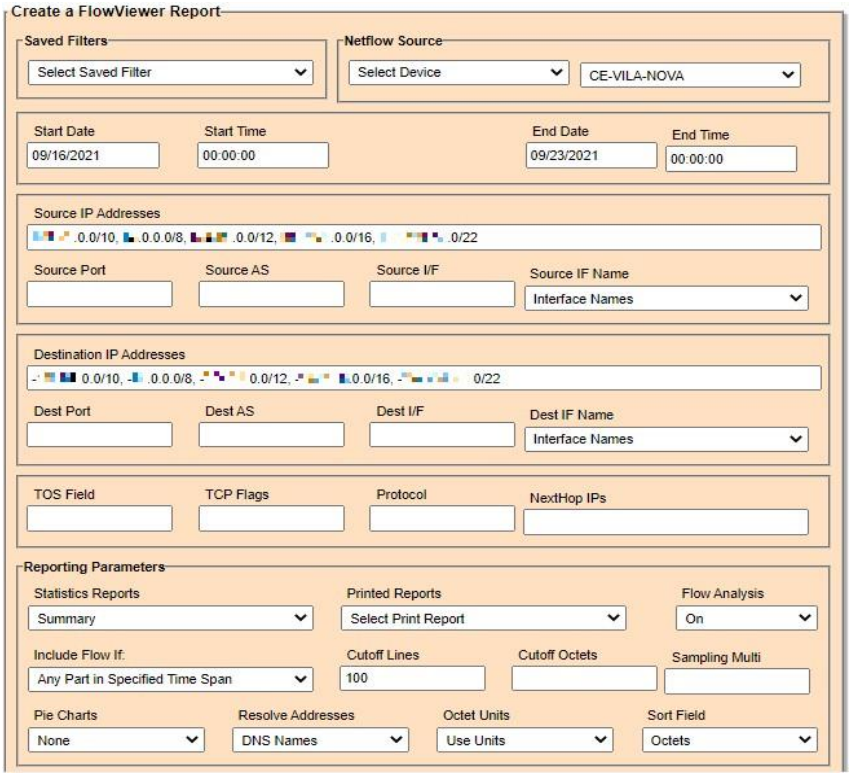
Figura 8.2 – Resultado do relatório resumido do tráfego total

```
Total Flows                    : 248870072
Total Octets                   : 19718434642338
Total Packets                  : 20587572315
Total Time (1/1000 secs) (flows): 1411670073410
Duration of data  (realtime)   : 606591
Duration of data (1/1000 secs) : 608378350
Average flow time (1/1000 secs) : 5672.3174
Average packet size (octets)   : 957.7833
Average flow size (octets)     : 79231.8359
Average packets per flow       : 82.7242
Average flows / second (flow)  : 409.0715
Average flows / second (real)  : 410.2766
Average Kbits / second (flow)  : 259291.8906
Average Kbits / second (real)  : 260055.7500
```

Fonte: O autor

Em seguida foi identificado o volume do tráfego direcionado, ou seja, da entrada e saída de dados do roteador. Para identificar o tráfego originado na rede do ISP e destinado à internet (saída), o campo endereço IP de origem (Source IP Address) foi preenchido com os grupos de IPs do ISP separados por vírgulas. No campo endereço IP de destino (Destination IP Address) foram adicionados os grupos de IPs do provedor com um sinal de subtração na frente e separados por vírgulas. A figura 8.3 mostra os parâmetros para relatório resumido do tráfego de saída para a internet.

Figura 8.3 – Tela com parâmetros para relatório resumido de tráfego de saída



Fonte: O autor

O resultado do relatório resumido mostrou que o tráfego total do roteador originado na rede do provedor e destinado à internet, foi de cerca de 1,078 Terabytes, que representa o consumo total de upload do link de internet naquele período.

Para averiguar o tráfego do roteador originado na internet e destinado à rede do ISP, foram invertidos os parâmetros de IP de origem e IP de destino utilizados no relatório anterior. O relatório de tráfego oriundo da internet e com destino à rede do ISP (entrada) resultou em um volume de cerca de 11,16 Terabytes, que representa o consumo total de download do link de internet naquele período.

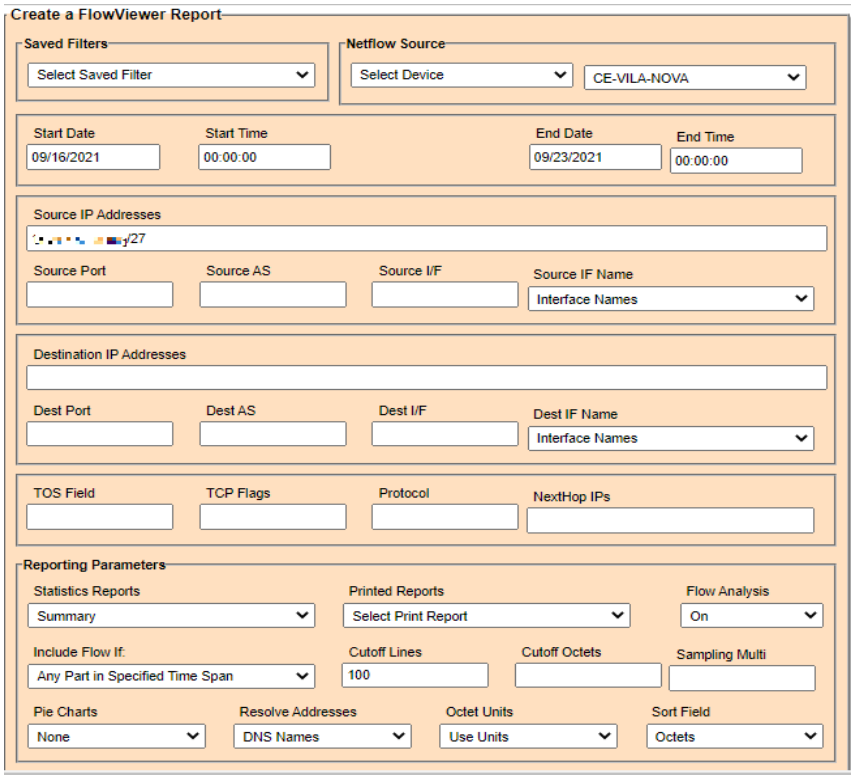
Para identificar o tráfego do roteador que possui como origem e destino a rede do próprio provedor, foram alterados os parâmetros de IP de origem e IP de destino, adicionando todos os grupos de IPs do provedor separados por vírgulas em ambos os campos. O que resultou em um tráfego de cerca de 5,68 Terabytes, que representa o tráfego total entre o roteador e a rede interna do ISP que não consumiram banda do link de internet.

O FlowViewer também foi empregado para analisar o tráfego entre o roteador e os servidores de distribuição de conteúdo (CDNs) do provedor. Conforme descrito por Trevisan et al. (2020), as CDNs surgiram nos anos 1990 com o objetivo de reduzir a sobrecarga nos servidores centralizados responsáveis pela transmissão de dados e minimizar o tempo de resposta para o usuário ao acessar os conteúdos. Atualmente, tanto CDNs públicas quanto privadas têm viabilizado a expansão da distribuição de conteúdo na Internet, permitindo que os usuários acessem informações de servidores substitutos mais próximos.

No ambiente onde os dados foram coletados, o provedor contava com servidores de distribuição de três serviços distintos: Facebook, Google e Netflix. Para identificar o volume de tráfego originado em um dos servidores CDNs com destino ao roteador, foram alterados os parâmetros de filtros e relatórios utilizados anteriormente, inserindo o grupo de IPs do respectivo servidor CDN no campo de IP de origem, e deixando o campo de IP de destino vazio. A figura 8.4 exibe a tela com os parâmetros para emissão de relatório resumido do tráfego originado em um dos servidores CDNs com destino ao roteador.

O procedimento inverso foi realizado para identificar o tráfego originado no roteador com destino à um dos CDNs, inserindo o grupo de IPs do CDN no campo de IP de destino, e deixando vazio o campo de IP de origem.

Figura 8.4 - Tela com parâmetros para relatório resumido do tráfego originado em um dos servidores CDNs



Fonte: O autor

O FlowViewer também foi utilizado para identificação de padrões de periodicidade do tráfego. Haffey et al. (2018) destacam a importância de os operadores de rede compreenderem esses padrões para distinguir entre comportamentos normais e anômalos em suas redes. Além disso, enfatizam que a detecção da ausência de periodicidade é igualmente relevante, especialmente quando sistemas que deveriam gerar tráfego periódico deixam de fazê-lo.

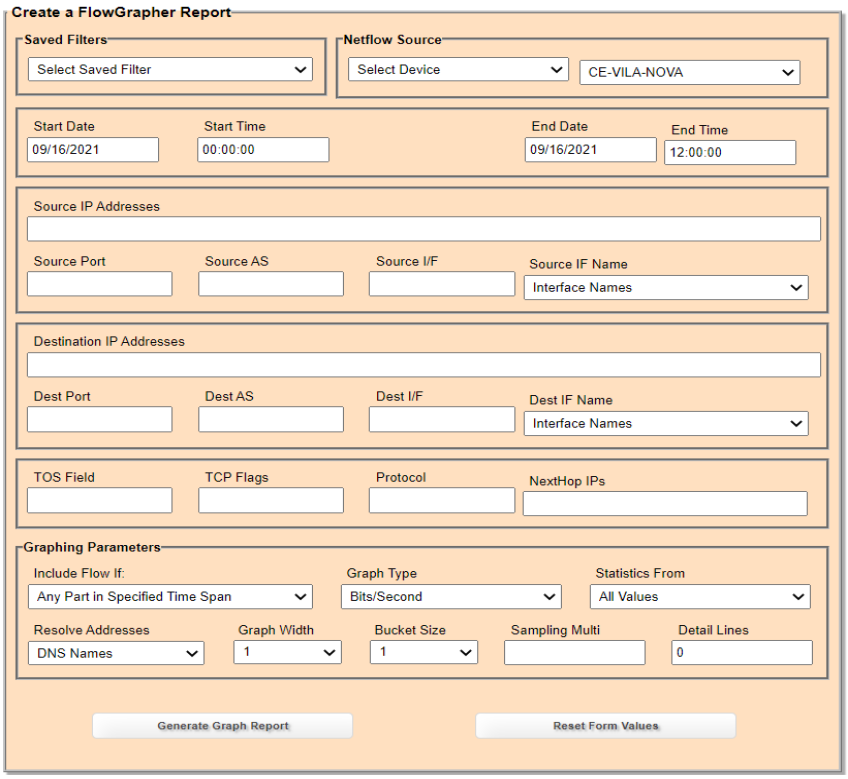
Os autores definem tráfego periódico como a ocorrência repetida de um "evento" em intervalos regulares. Para caracterizar corretamente esse tipo de tráfego, é necessário considerar o número e o tipo de eventos de rede a serem detectados, a duração da janela de detecção, os períodos mínimo e máximo a serem observados e a margem de tolerância quanto à periodicidade.

Um exemplo de análise de padrões de periodicidade no tráfego de rede pode ser encontrado no estudo de Padilla (2020), que examinou o comportamento do tráfego durante a quarentena, estabelecendo padrões para o volume diário e os horários de pico durante a pandemia. De forma semelhante, Benmusa et al. (2022) realizaram medições em intervalos curtos, permitindo o cálculo da taxa média de transmissão e dos valores de

pico para cada hora. Além disso, calcularam a utilização da capacidade total do link diariamente, proporcionando uma visão clara do desempenho do sistema, o que é essencial para a tomada de decisões empresariais.

O FlowViewer foi utilizado para gerar gráficos com as taxas de transmissão do roteador em períodos específicos através do componente FlowGrapher, onde foram definidas as entradas: exportador, data inicial, hora inicial, data final e hora final. Foi optado por realizar análises em turnos de 12 horas, ou seja, cada dia completo de dados coletados foram divididos em dois períodos. A figura 8.5 mostra os parâmetros do FlowGrapher para o tráfego no período de 00:00 horas a 12:00 horas do dia 16/09/2021.

Figura 8.5 – Parâmetros do FlowGrapher para tráfego de 00:00 a 12:00 horas do dia 16/09/2021

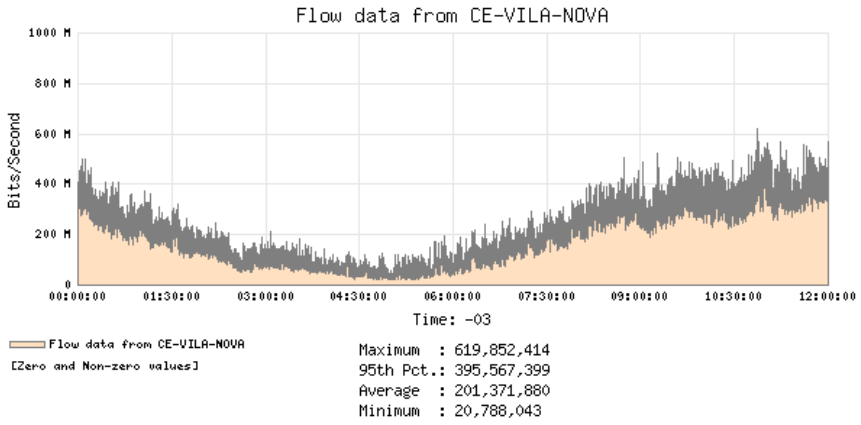


Fonte: O autor

Nos parâmetros da exibição de informações nos gráficos, foi selecionada a opção “Any Part In Specified Time Span” no campo Include Flow If, para que todas as partes dos fluxos contidos no período especificado fossem contabilizadas (independente se os fluxos foram ou não finalizados). No campo Graph Type foi selecionada a opção “Bits/Second”, para que a quantidade de dados trafegados em relação ao tempo seja exibida. Em Statistics From foi selecionada a opção “All Values”, para que todos os

valores fossem contabilizados (inclusive zeros). O tamanho do gráfico selecionado em Graph Width foi “1”, e o tempo de amostragem em segundos definido em Bucket Size foi “1”. A figura 8.6 exibe o gráfico gerado pelo FlowGrapher do tráfego de 00:00 horas a 12:00 horas do dia 16/09/2021.

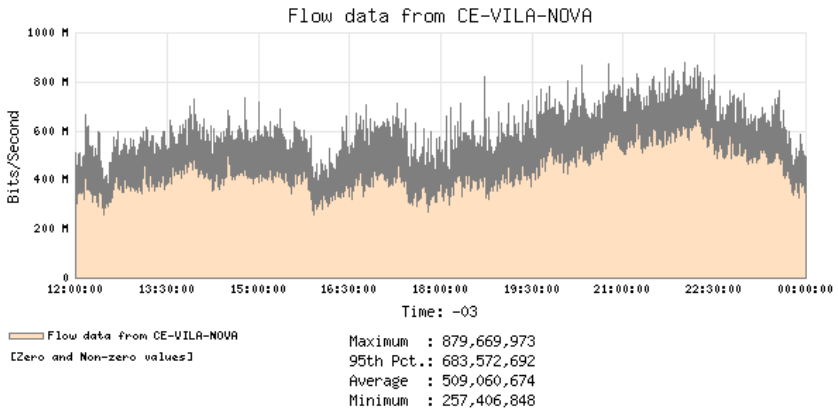
Figura 8.6 – Gráfico do tráfego de 00:00 horas a 12:00 horas do dia 16/09/2021



Fonte: O autor

No período de 00:00 a 12:00 horas do dia 16/09/2021, o intervalo com maiores taxas de transmissão foi de 09:00 a 12:00 horas, com um pico de cerca de 619 Mbps. As menores taxas ocorreram no período de 02:00 a 07:00 horas, com a menor taxa em torno de 20 Mbps. E a taxa média total do período foi de cerca de 201 Mbps. A figura 8.7 mostra o tráfego de 12:00 horas do dia 16/09/2021 à 00:00 horas do dia 17/09/2021.

Figura 8.7 – Gráfico do tráfego de 12:00 horas do dia 16/09/2021 à 00:00 horas do dia 17/09/2021



Fonte: O autor

É possível observar no gráfico da figura 8.7 que a taxa de transmissão de dados variou menos que no gráfico do período anterior, com taxa média de 509 Mbps, mais do que o dobro da taxa média do período anterior. A taxa mínima de transmissão foi de 257 Mbps, cerca de dez vezes mais do que a taxa mínima do período anterior. O intervalo com maior transmissão de dados foi de 19:30 horas do dia 16/09/2021 à 00:00 horas do dia 17/09/202, com uma taxa máxima de transmissão de 879 Mbps. Analisando as taxas de transmissão do dia 16/09/2021, podemos notar que o segundo turno possui uma quantidade maior de dados trafegados e picos de transmissão maiores que o primeiro.

Outro aspecto importante do campo da análise do tráfego de rede é a classificação de tráfego, e, conforme exposto por Shahraki et al. (2022), refere-se ao conjunto de técnicas empregadas para categorizar o tráfego em diferentes classes, com base em suas características. Os autores destacam que essa prática oferece grandes benefícios em diversas áreas dos serviços de rede, como na avaliação da qualidade de serviço (QoS), na detecção de malware e na prevenção de intrusões. Santos et al. (2018) ressaltam que classificar o tráfego gerado por diferentes dispositivos possibilita a criação de perfis de tráfego, permitindo a identificação de padrões de comunicação.

Essa classificação é crucial para que provedores de serviços de internet (ISPs) e organizações monitorem e administrem o tráfego conforme suas necessidades. Entre as metodologias disponíveis, as técnicas baseadas em portas têm sido amplamente utilizadas para classificar aplicativos tradicionais, cujas portas são atribuídas pela Internet Assigned Numbers Authority (IANA). Esses métodos são eficientes para aplicativos que utilizam protocolos conhecidos, como HTTP, FTP, POP3 e CMP (Yamansavascilar et al., 2017).

Azab et al. (2024) abordam os pontos fortes e fracos da classificação baseada em portas, destacando entre suas vantagens, a simplicidade de implementação, os baixos requisitos de processamento e a alta velocidade na classificação do tráfego. Entretanto, suas principais desvantagens incluem a prática de mascaramento por diversos aplicativos, que utilizam portas padrão para transmitir outros tipos de tráfego, como o tráfego malicioso por meio de HTTP. Além disso, muitos aplicativos modernos utilizam portas dinâmicas ou não padronizadas, como ocorre com aplicativos de VoIP, o que reduz a eficácia dessa abordagem.

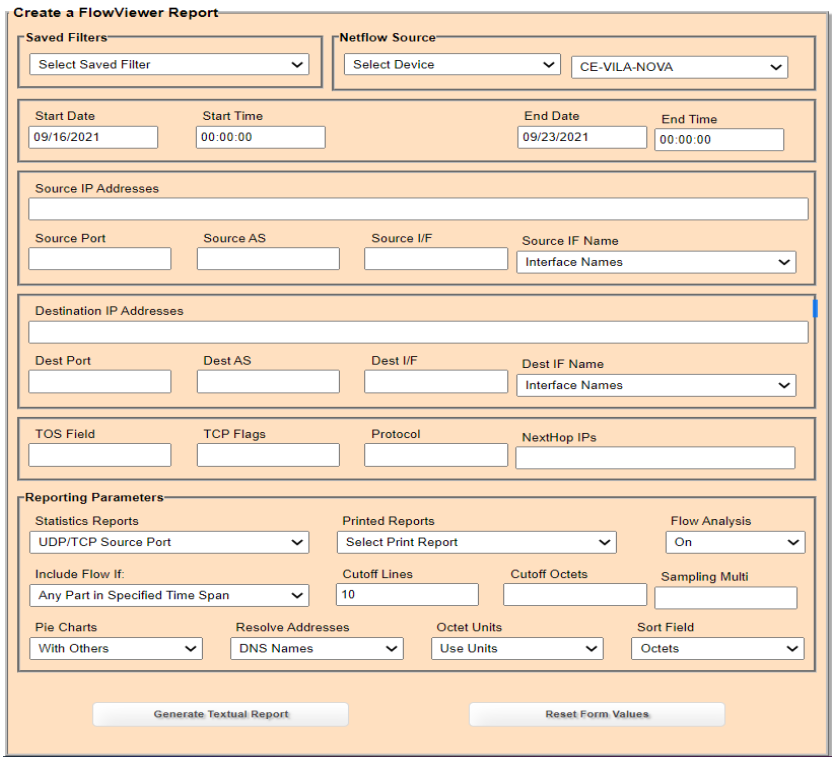
Os autores também enfatizam que a IANA aloca números de porta padrão para diferenciar serviços e protocolos de tráfego. Esses números são distribuídos em três categorias: portas do sistema (0–1023), portas do usuário (1024–49151) e portas dinâmicas (49152–65535). As portas reservadas para a implementação de protocolos

padrão pertencem ao intervalo de 0 a 1023. O processo de classificação, geralmente, envolve a inspeção dos números de porta dos pacotes dos protocolos Transmission Control Protocol (TCP) e User Datagram Protocol (UDP), que são mapeados para as portas predefinidas pela IANA com o objetivo de categorizar o tráfego de rede.

Um exemplo de porta pré-definida para serviços ou aplicações em redes de computadores é o serviço WEB (protocolo HTTP) que utiliza a porta 80 (Do Carmo, 2019). Embora a identificação de quais são as portas mais utilizadas em uma rede não aponte precisamente quais serviços estão sendo mais utilizados, ela apresenta uma boa ideia de quais aplicações estão sendo utilizadas na rede, além de alertarem sobre possíveis falhas de segurança.

O FlowViewer possibilita identificar quais portas estão sendo mais utilizadas na rede através da quantidade total de dados trafegados por porta ou pela quantidade de fluxos por porta. Para identificar as portas mais utilizadas, foi utilizado o componente FlowViewer, a figura 8.8 mostra os parâmetros de filtros e relatórios utilizados para obter a lista das 10 portas de origem, ou seja, portas de onde foram originados os fluxos, que mais trafegaram dados no período total de coleta deste trabalho.

Figura 8.8 – Parâmetros para relatório das 10 portas de origem com mais tráfego



Fonte: O autor

Foram definidos como parâmetros dos filtros, o exportador “CE-VILA-NOVA” e o período de 00:00 horas do dia 16/09/2021 à 00:00 horas do dia 22/09/2021. Após a configuração de filtros foram selecionadas as entradas de relatórios do FlowViewer, na opção de relatórios estatísticos (Statistics Reports) foi selecionada a opção “UDP/TCP Source Port”, para que o relatório utilize como parâmetro apenas as portas de origem dos fluxos. Na opção linhas de corte (Cutoff lines) foi definida como entrada “10”, para que apenas as 10 portas de origem com maior tráfego de dados sejam exibidas no relatório.

A opção campo de classificação (Sort Field) foi mantida com a entrada já pré-estabelecida “octets”, o que classifica de forma decrescente as portas de origem levando em consideração o tráfego de dados, o restante dos parâmetros de relatórios permaneceu conforme o padrão inicial definido na interface do sistema. A figura 8.9 exibe o resultado com a lista das 10 portas de origem com mais tráfego de dados, onde a porta 443 foi a que obteve maior tráfego com cerca de 13,03 Terabytes, a porta 80 foi a segunda porta da lista com cerca de 2,07 Terabytes, e as demais portas apresentaram um tráfego bem abaixo das duas primeiras, com tráfegos entre 21 e 56 Gigabytes.

Figura 8.9 - Lista das 10 portas de origem com mais tráfego

Src Port	Flows	Octets	Packets	Octs/Flow	Pkts/Flow
443	58570477	13.03 TB	11126361517	244514	189
80	16971636	2.07 TB	1628633259	134425	95
2082	17159	55.84 GB	41047629	3494296	2392
41000	21453	46.93 GB	36828429	2348929	1716
8080	32863	44.06 GB	32807932	1439515	998
3478	120498	43.71 GB	70400903	389491	584
4520	1601	29.70 GB	23724473	19919985	14818
4659	1559	28.44 GB	22755059	19589367	14595
4340	1672	26.47 GB	21411458	16999731	12805
1194	1367	21.92 GB	30155389	17213680	22059

Fonte: O autor

Além das portas de origem com mais tráfego de dados, foi emitido um relatório com a lista de portas de origem mais presentes em fluxos, possibilitando averiguar quais portas mais iniciaram solicitações. Para alterar a emissão de relatórios da lista de portas de origem com maior tráfego de dados para as mais presentes em fluxos, foi alterada a entrada da opção campo de classificação (Sort Field) de “octets” para “flows”.

Na figura 8.10 é exibida a lista das 10 portas de origem mais presentes em fluxos (portas que mais iniciaram conexões). As portas que mais originaram fluxos foram as portas 443, 53 e 80, atribuídas respectivamente ao protocolo HTTPS, serviço de DNS e protocolo HTTP.

Figura 8.10 - Lista das 10 portas de origem mais presentes em fluxos

Src Port	Flows	Octets	Packets	Octs/Flow	Pkts/Flow
443	58570477	13.03 TB	11126361517	244514	189
53	26454405	5.38 GB	32997712	218	1
80	16971636	2.07 TB	1628633259	134425	95
161	4481557	386.82 MB	5084220	90	1
123	2794866	2.32 GB	32565867	890	11
5222	1343778	5.22 GB	23877404	4174	17
5228	1312662	1.80 GB	4577587	1475	3
1813	1298163	60.20 MB	1305380	48	1
6881	783103	803.87 MB	1836705	1076	2
853	365947	1.82 GB	5728960	5338	15

Fonte: O autor

Para identificar as portas de destino que estiveram presentes em mais fluxos, e as que apresentaram maior tráfego na rede, foram utilizados os mesmos parâmetros para geração das listas de portas de origem, sendo alterada apenas a entrada do campo de relatórios estatísticos (Statistics Reports) para “UDP/TCP Destination Port”. A figura 8.11 exibe a lista das 10 portas de destino com mais tráfego de dados. A porta de destino dos fluxos coletados que mais trafegou dados foi a porta 443, com um tráfego total acima de 627 Gigabytes, um volume mais de 10 vezes maior que o da segunda porta da lista.

Figura 8.11 - Lista das 10 portas de destino com mais tráfego

Dst Port	Flows	Octets	Packets	Octs/Flow	Pkts/Flow
443	61047919	627.26 GB	3882695760	11032	63
3478	150591	41.39 GB	67772312	295091	450
80	20374619	36.06 GB	613082384	1900	30
37260	5741	17.33 GB	13972734	3241102	2433
35374	90029	17.18 GB	18708312	204845	207
37839	63437	17.08 GB	16456644	289167	259
36071	2147	12.42 GB	9939876	6213684	4629
18002	1109	11.45 GB	8269967	11090456	7457
39280	5488	10.69 GB	8570226	2090903	1561
35929	61612	9.31 GB	10320774	162258	167

Fonte: O autor

A figura 8.12 exibe a lista das 10 portas de destino mais presentes em fluxos, onde se repete a mesma lista das portas de origem mais presentes em fluxos. Que são portas atribuídas à serviços e protocolos essenciais para o acesso WEB (443, 53, 80 e 853), à protocolos de grande importância para o funcionamento da rede e seus dispositivos (161, 123 e 1813), e à serviços de comunicação (5222 e 5228).

Figura 8.12 - Lista das 10 portas de destino mais presentes em fluxos

Dst Port	Flows	Octets	Packets	Octs/Flow	Pkts/Flow
443	61047919	627.26 GB	3882695760	11032	63
53	26691914	2.13 GB	32567822	85	1
80	20374619	36.06 GB	613082384	1900	30
161	4651860	450.66 MB	6122458	101	1
123	3158448	2.48 GB	34889962	843	11
5222	1376612	2.52 GB	21553437	1962	15
1813	1306741	397.13 MB	1342479	318	1
5228	1244299	385.79 MB	4044991	325	3
6881	668678	247.93 MB	1611900	388	2
853	346266	684.00 MB	5359660	2071	15

Fonte: O autor

Para identificar as portas que de forma geral estiveram presentes em mais fluxos, e apresentaram maior tráfego na rede, foi feita uma alteração nos mesmos parâmetros usados anteriormente, trocando a entrada de relatórios estatísticos (Statistics Reports) para “UDP/TCP Port”. A figura 8.13 exibe a lista das 10 portas com mais tráfego, onde a porta 443 foi a que obteve maior tráfego com cerca de 13,64 Terabytes, a porta 80 foi a segunda porta da lista com cerca de 2,11 Terabytes, e as demais portas apresentaram um tráfego bem abaixo das duas primeiras, com tráfegos entre 27 e 86 Gigabytes.

Figura 8.13 - Lista das 10 portas com mais tráfego

Port	Flows	Octets	Packets	Octs/Flow	Pkts/Flow
443	119618396	13.64 TB	15009057277	125355	125
80	37346255	2.11 TB	2241715643	62124	60
3478	271089	85.10 GB	138173215	337051	509
2082	34923	56.62 GB	55150978	1740901	1579
41000	32551	48.15 GB	41955895	1588204	1288
8080	77856	45.70 GB	46909760	630283	602
4520	4052	30.29 GB	25272521	8026145	6237
4659	4641	28.95 GB	23840747	6698224	5136
4340	5834	27.04 GB	23914178	4976337	4099
40771	85733	27.01 GB	43335815	338241	505

Fonte: O autor

Na figura 8.14 é exibida a lista das 10 portas mais presentes em fluxos de forma geral, onde as portas apresentadas são as mesmas da lista de portas de origem mais presentes em fluxos. As portas que ocuparam as primeiras posições da lista foram 443, 53 e 80.

Figura 8.14 - Lista das 10 portas mais presentes em fluxos

Port	Flows	Octets	Packets	Octs/Flow	Pkts/Flow
443	119618396	13.64 TB	15009057277	125355	125
53	53146319	7.50 GB	65565534	151	1
80	37346255	2.11 TB	2241715643	62124	60
161	9133417	837.48 MB	11206678	96	1
123	5953314	4.80 GB	67455829	865	11
5222	2720390	7.74 GB	45430841	3055	16
1813	2604904	457.34 MB	2647859	184	1
5228	2556961	2.18 GB	8622578	915	3
6881	1451781	1.03 GB	3448605	759	2
853	712213	2.49 GB	11088620	3750	15

Fonte: O autor

No próximo capítulo serão abordados os resultados das análises de tráfego baseadas nos dados obtidos através do FlowViewer, discorrendo sobre os insights obtidos através do volume de tráfego, taxas de transmissão e portas mais utilizadas na rede.

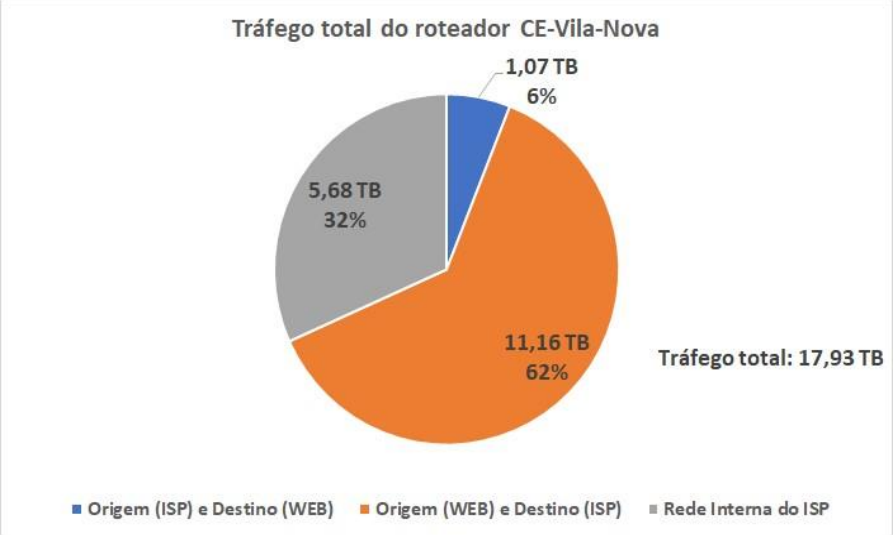
9 RESULTADOS

Conforme apontado por Ji et al. (2021), a visualização de dados desempenha um papel fundamental na extração de informações significativas dentro do contexto da análise de tráfego de rede. Com base nessa perspectiva, foram elaborados gráficos que visam melhorar a obtenção de insights sobre o comportamento do tráfego na rede.

Através dos relatórios de tráfego resumidos obtidos pelo FlowViewer, foi analisada a direção (entrada e saída) do tráfego do roteador. A Figura 9.1 exibe o gráfico com a distribuição da direção do tráfego do roteador, onde 62% dos dados trafegados foram originados na internet com destino à rede do ISP (cerca de 11,16 Terabytes), 6% tiveram como origem a rede do provedor e destino a internet (cerca de 1,07 Terabytes), e 32% do tráfego tiveram como origem e destino a rede do ISP (cerca de 5,68 Terabytes).

Constata-se então que o consumo de download dos links de internet utilizados no roteador foi cerca de 10 vezes maior que o consumo de upload, e cerca de um terço do tráfego passou somente pela rede interna do ISP, poupando banda dos links de internet.

Figura 9.1 – Gráfico com distribuição da direção do tráfego do roteador

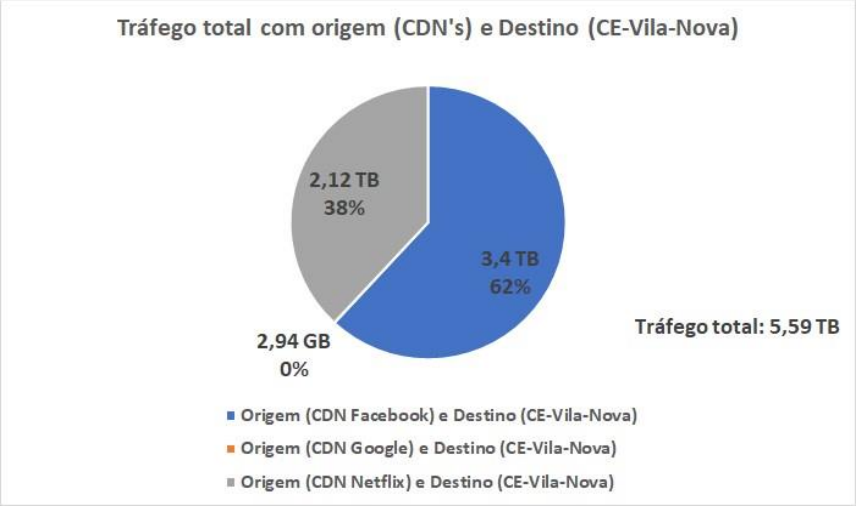


Fonte: O autor

A figura 9.2 exibe o gráfico que mostra a distribuição de serviços do tráfego originado em todos os CDNs com destino ao roteador, o CDN que enviou mais dados ao roteador foi o do Facebook com um total de 3,4 Terabytes enviados (cerca de 62% do

total), seguido pelo da Netflix com um total de 2,12 Terabytes (cerca de 38% do total), e por último o do Google com 2,94 Gigabytes de dados enviados (menos de 1% do total).

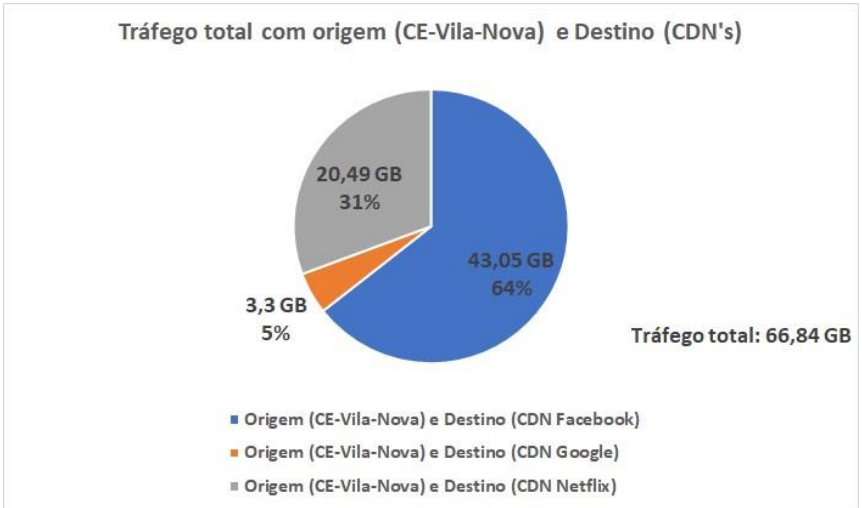
Figura 9.2 – Gráfico com distribuição de serviços do tráfego originado nos CDNs e destinados ao roteador



Fonte: O autor

Na figura 9.3 é mostrado o gráfico da distribuição do tráfego originado no roteador com destino aos CDNs, o CDN que mais recebeu dados foi o do Facebook com 43,05 Gigabytes (cerca de 64% do total), seguido pelo da Netflix com 20,49 Gigabytes (cerca de 31% do total) e o do Google com 3,3 Gigabytes (cerca de 5% do total).

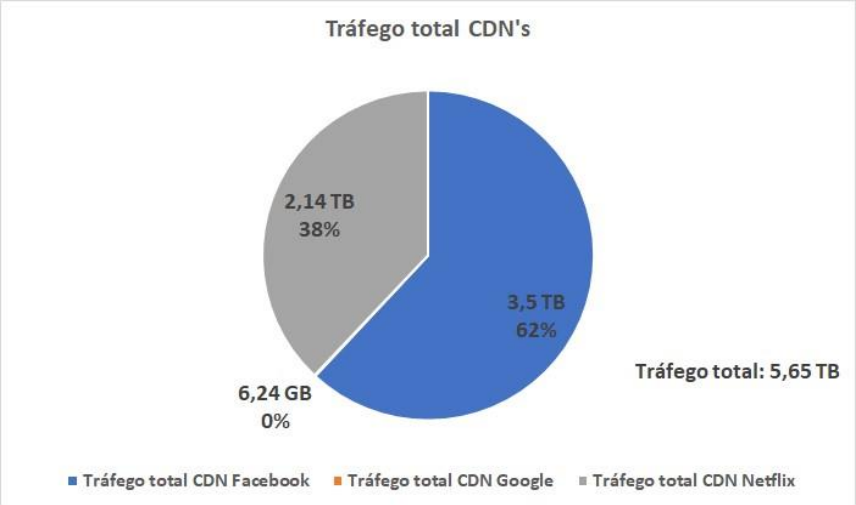
Figura 9.3 – Gráfico com distribuição de serviços do tráfego originado no roteador com destino aos CDNs



Fonte: O autor

Na figura 9.4 é exibido o gráfico com a distribuição de serviços do tráfego total entre o roteador e os CDNs, onde mostra que 62% do tráfego foi do CDN Facebook, 38% do CDN Netflix e menos de 1% do CDN Google. É possível então observar que o CDN Facebook é o servidor que mais proporciona economia de banda no consumo dos links de internet do provedor. A soma do tráfego total entre o roteador e todos os CDNs é de cerca de 5,65 Terabytes, que representa 99% do tráfego total entre o roteador e a rede interna do provedor, de cerca de 5,68 Terabytes, ou seja, o tráfego entre o roteador e a rede interna do ISP é quase inteiramente proveniente da comunicação com os CDNs.

Figura 9.4 – Gráfico com distribuição de serviços do tráfego total entre o roteador e os servidores CDNs

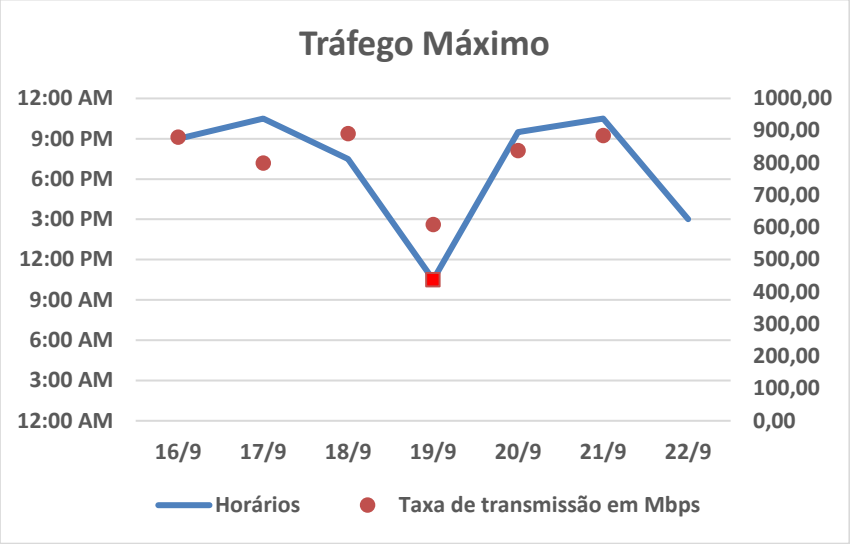


Fonte: O autor

Com base nos gráficos gerados no FlowGrapher, foi realizado um comparativo de como sucedeu o tráfego em cada dia, relacionando os horários com as maiores e menores taxas de transmissão.

A figura 9.5 exibe o gráfico com a relação dos horários com maiores taxas de transmissão em cada dia e seus respectivos valores em Mbps, onde em 5 dos 7 dias monitorados, os maiores picos de transmissão ocorreram no período da noite, entre 19:00 e 23:00 horas. As exceções ocorreram nos dias 19/09/2021 e 22/09/2021. No dia 19/09/2021 não houve tráfego no período da noite devido problemas técnicos na conexão da região, neste dia o horário com maior taxa de transmissão foi no período da manhã (marcador vermelho no gráfico, entre 10:00 e 11:00 horas). No dia 22/09/2021, o pico mais alto de tráfego ocorreu durante a tarde, por volta das 15:00 horas.

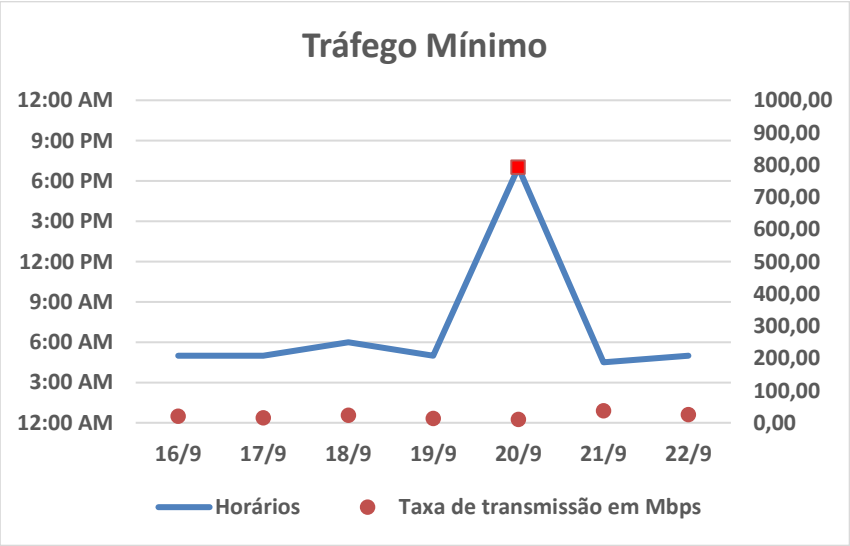
Figura 9.5 – Gráfico diário dos horários com maiores taxas de transmissão



Fonte: O autor

A figura 9.6 exibe o gráfico dos horários de menor tráfego em cada dia e suas respectivas taxas. Não foram considerados horários em que o tráfego foi afetado por incidentes, exceto no dia 20/09/2021 (marcador vermelho no gráfico), onde não houve tráfego por um longo período devido um incidente na rede, o que afetou os resultados deste dia. Nota-se um padrão de baixo tráfego no início do dia, entre 04:00 e 07:00 horas, exceto no dia 20/09/2021, aonde a conexão voltou apenas a noite após o fim do incidente.

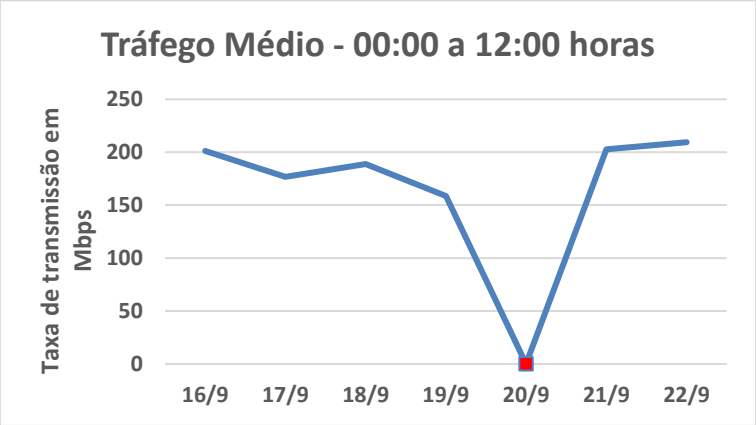
Figura 9.6 – Gráfico diário dos horários com menores taxas de transmissão



Fonte: O autor

A figura 9.7 exibe o gráfico da taxa média no período de 00:00 a 12:00 horas de cada dia. A taxa se manteve entre 150 Mbps e 250 Mbps, exceto em 16/09/2021 (marcador vermelho no gráfico), onde não houve tráfego devido a um incidente na rede.

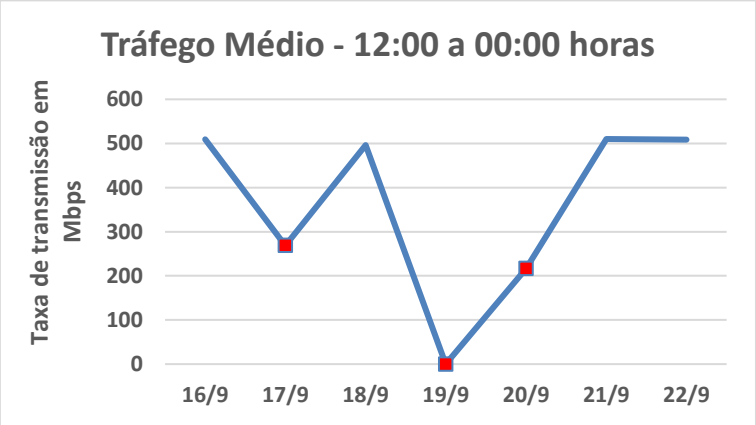
Figura 9.7 – Gráfico diário da taxa média de transmissão entre 00:00 e 12:00 horas



Fonte: O autor

Na figura 9.8 é exibido o gráfico com a taxa média de 12:00 a 00:00 horas, onde nos dias em que não ocorreram incidentes que perduraram por um longo tempo, a taxa média de transmissão se manteve entre 400 Mbps e 600 Mbps, mais do que o dobro da taxa média do período anterior. Nos dias 17/09/2021 e 20/09/2021 os valores da taxa média ficaram entre 200 Mbps e 300 Mbps, devido a incidentes na rede que ocasionaram ausência de tráfego por longos períodos. No dia 19/09/2021 o valor médio de tráfego foi 0 Mbps, devido a um incidente que ocasionou ausência de tráfego em todo o período.

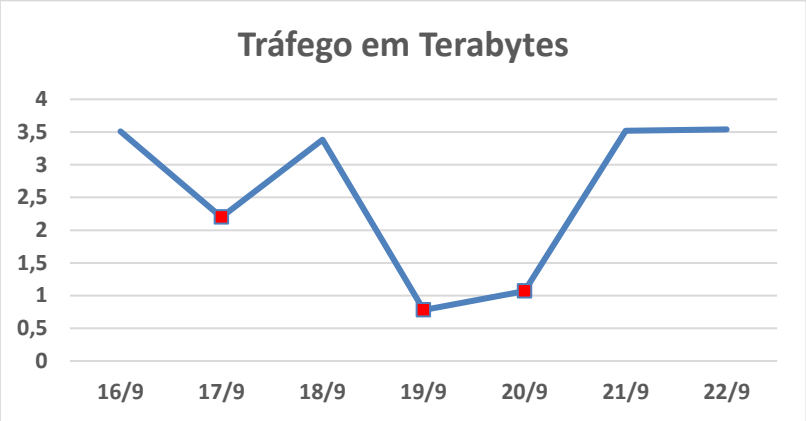
Figura 9.8 – Gráfico com média de tráfego diária entre 12:00 horas e 00:00 horas



Fonte: O autor

A figura 9.9 mostra um gráfico com o tráfego total em Terabytes de todos os dias analisados, onde é possível notar um padrão que varia entre 3 e 4 Terabytes, com exceção dos dias afetados por longos incidentes na rede (marcadores em vermelho).

Figura 9.9 – Gráfico diário do tráfego total no roteador



Fonte: O autor

A partir dos gráficos apresentados, observa-se que o menor volume de tráfego ocorre nas primeiras horas da manhã, seguido por um aumento gradual ao longo do dia, atingindo o pico de transmissão durante o período noturno. Essas informações são valiosas para os administradores de rede, permitindo decisões mais assertivas, como a seleção de horários adequados para a realização de manutenções e a priorização da disponibilidade para o monitoramento da rede e suporte aos clientes em momentos de maior demanda.

As informações referentes às taxas médias de transmissão e ao volume total de tráfego diário são fundamentais para a adequada manutenção da rede. Elas permitem uma gestão eficiente da capacidade total dos equipamentos e dos links de internet, prevenindo possíveis instabilidades decorrentes de gargalos na infraestrutura de rede.

Em relação as portas mais utilizadas, através de uma pesquisa no site da organização responsável pela designação do número de portas paras as aplicações chamada “IANA” (Autoridade para Atribuição de Números da Internet), foi verificado que algumas portas listadas são oficialmente atribuídas ao uso de protocolos e serviços. Segue abaixo a lista de portas de origem que mais trafegaram dados e seus respectivos protocolos e serviços atribuídos:

- 443 – Porta oficial do protocolo HTTPS, versão mais segura do HTTP;

- 80 – Porta oficial do protocolo HTTP, responsável pelo acesso WEB;
- 2082 – Porta atribuída para o serviço “Infowave Mobility Server”, pertencente a empresa “Infowave”, responsável por serviços de T.I voltados para os setores da saúde, seguros e serviços públicos;
- 41000 – Não atribuída oficialmente a serviços ou aplicações conforme o portal da organização “IANA”;
- 8080 – Utilizada como porta alternativa para o protocolo HTTP, geralmente usada em servidores web;
- 3478 – Porta atribuída ao conjunto padronizado de métodos STUN (“Session Traversal Utilities for NAT”), muito utilizado em aplicações de voz, vídeo, mensagens e demais serviços de comunicação em tempo real;
- 4520 – Não atribuída oficialmente a serviços ou aplicações conforme o portal da organização “IANA”;
- 4659 – Porta registrada como a oficial do serviço “PlayStation2 Lobby”;
- 4340 – Porta atribuída ao serviço “Gaia”;
- 1194 – Porta pré-estabelecida como padrão do serviço “OpenVPN”, um software livre e open-source utilizado na criação de redes privadas.

Segue abaixo a lista de portas de origem mais presentes em fluxos e suas respectivas atribuições segundo o “IANA”:

- 443 – Porta oficial do protocolo HTTPS, versão mais segura do HTTP;
- 53 – Porta atribuída oficialmente ao DNS (Sistema de Nomes de Domínio), utilizado em navegadores web para conversão de domínios em IP’s;
- 80 – Porta oficial do protocolo HTTP, responsável pelo acesso WEB;
- 161 – Porta atribuída ao SNMP (Protocolo Simples de Gerência de Rede), protocolo responsável pelo gerenciamento de dispositivos em redes IP;
- 123 – Porta atrelada ao NTP (Protocolo de Tempo para Redes), protocolo responsável pela sincronização dos relógios dos dispositivos de uma rede;
- 5222 – Porta atribuída ao “xmpp-client”, o XMPP é um protocolo para sistemas de mensagens instantâneas utilizado em algumas aplicações como Google Talk e serviços de comunicação instantânea do Gmail;
- 5228 – Porta atribuída ao serviço “HP Virtual Room”;

- 1813 – Porta atribuída ao RADIUS Accouting, protocolo de rede utilizado por serviços de email, VPN e por provedores de internet no gerenciamento de acesso à internet ou intranet;
- 6881 – Porta não atribuída oficialmente no “IANA”;
- 853 – Porta atribuída ao DNS sobre TLS, um dos padrões de criptografia para consultas DNS.

Comparando a lista das 10 portas de origem com maior tráfego e a lista das 10 portas de origem mais presentes em fluxos, pode-se notar que as únicas portas em ambas as listas foram a 443 e 80, oficialmente atribuídas a serviços WEB, com a porta 443 na primeira posição em ambas as listas, e a porta 80 na segunda posição das portas de origem com mais tráfego e terceira posição nas portas de origem mais presentes em fluxos.

Mesmo com a possibilidade de outros serviços também utilizarem estas portas já pré-estabelecidas respectivamente para os protocolos HTTPS e HTTP, o alto tráfego e requisições para essas portas indicam o acesso WEB como o principal serviço utilizado pelos usuários. A grande utilização da internet para acessar páginas WEB justifica também a presença das portas 53 (atribuída ao DNS) e 853 (atribuída ao DNS sobre TLS) na lista das portas de origem que mais fizeram requisição na rede, devido a necessidade da utilização desse serviço na tradução dos domínios para endereços IPs. Vale ressaltar que outro padrão de criptografia do DNS, o DNS sobre HTTPS, também utiliza a já citada anteriormente porta 443, logo, todas as portas atreladas a serviços de DNS e suas versões criptografadas estão presentes na lista de portas de origem mais presentes nos fluxos.

Na lista de portas de origem mais presentes em fluxos, há portas atribuídas a protocolos básicos para o funcionamento da rede e seus dispositivos, como as portas 161 (protocolo SNMP) e 123 (protocolo NTP), o que indica que muitas requisições foram realizadas por meio delas devido a importância desses protocolos, porém elas não possuem tráfego considerável para figurar na lista de portas de origem com mais tráfego.

Na lista abaixo estão presentes as 10 portas de destino que mais trafegaram dados e suas respectivas atribuições segundo o “IANA”:

- 443 – Porta oficial do protocolo HTTPS, versão mais segura do protocolo HTTP;
- 3478 – Porta atribuída ao conjunto padronizado de métodos STUN (“Session Traversal Utilities for NAT”), muito utilizado em aplicações de voz, vídeo, mensagens e demais serviços de comunicação em tempo real;

- 80 – Porta oficial do protocolo HTTP, responsável pelo acesso WEB;
- 37260 – Porta não atribuída oficialmente a nenhum serviço conforme o "IANA";
- 35374 – Porta não atribuída oficialmente a nenhum serviço conforme o "IANA";
- 37839 – Porta não atribuída oficialmente a nenhum serviço conforme o "IANA";
- 36071 – Porta não atribuída oficialmente a nenhum serviço conforme o "IANA";
- 18002 – Porta não atribuída oficialmente a nenhum serviço conforme o "IANA";
- 39280 – Porta não atribuída oficialmente a nenhum serviço conforme o "IANA";
- 35929 – Porta não atribuída oficialmente a nenhum serviço conforme o "IANA".

Segue abaixo a lista de 10 portas de destino que estiveram presentes em mais fluxos e suas respectivas atribuições segundo o "IANA":

- 443 – Porta oficial do protocolo HTTPS, versão mais segura do protocolo HTTP;
- 53 – Porta atribuída oficialmente ao DNS (Sistema de Nomes de Domínio), utilizado em navegadores web para conversão de domínios em endereços IP;
- 80 – Porta oficial do protocolo HTTP, responsável pelo acesso WEB;
- 161 – Porta atribuída ao SNMP (Protocolo Simples de Gerência de Rede), protocolo responsável pelo gerenciamento de dispositivos em redes IP;
- 123 – Porta atrelada ao NTP (Protocolo de Tempo para Redes), protocolo responsável pela sincronização dos relógios dos dispositivos de uma rede;
- 5222 – Porta atribuída ao "xmpp-client", o XMPP é um protocolo para sistemas de mensagens instantâneas utilizado em algumas aplicações como Google Talk e serviços de comunicação instantânea do Gmail;
- 1813 – Porta atribuída ao RADIUS Accouting, protocolo de rede utilizado por serviços de email, VPN e por provedores de internet no gerenciamento de acesso à internet ou intranet;

- 5228 – Porta atribuída ao serviço “HP Virtual Room”;
- 6881 – Porta não atribuída oficialmente a nenhum serviço conforme o “IANA”;
- 853 – Porta atribuída ao DNS sobre TLS, um dos padrões de criptografia para consultas DNS.

Na lista das portas de destino com mais volume de tráfego, as três primeiras portas foram a 443 (atribuída ao protocolo HTTPS), 3478 (atribuída ao STUN), e 80 (atribuída ao protocolo HTTP) respectivamente, todas elas também estão na lista das 10 portas de origem com mais tráfego. No entanto, o restante das portas, 37260, 35374, 37839, 36071, 18002, 39280 e 35929, estão presentes somente na lista de portas de destino com mais tráfego, e não estão atribuídas a nenhuma aplicação ou serviço pelo “IANA”.

É preciso ressaltar que a porta 18002, oitava porta da lista, com tráfego de cerca de 11,45 Gigabytes, foi a porta utilizada para a coleta de fluxos no servidor. A presença desta porta na lista das portas de destino com mais tráfego é justificada devido o roteador ter sido configurado para enviar dados constantemente ao flow-tools, ocasionando assim um alto tráfego de dados com destino a esta porta.

Assim como nas listas das portas de origem, as únicas portas de destino que estão presente nas listas de portas com maior volume de tráfego e mais presentes em fluxos, foram as portas atribuídas a serviços WEB (443 e 80), ressaltando o acesso WEB como o serviço mais utilizado.

Em relação as portas que estiveram mais presentes em fluxos, as mesmas portas que estão na lista de portas de origem também aparecem na lista de portas de destino, com diferença apenas na ordem de duas das portas. Além das portas atribuídas a serviços WEB, estão presentes na lista portas atribuídas a protocolos básicos para o funcionamento da rede e portas atribuídas a serviços de DNS.

Também foi feito o levantamento das portas que de forma geral mais trafegaram dados, segue abaixo a lista das portas e suas respectivas atribuições de acordo com o “IANA”:

- 443 – Porta oficial do protocolo HTTPS, versão mais segura do protocolo HTTP;
- 80 – Porta oficial do protocolo HTTP, responsável pelo acesso WEB;
- 3478 – Porta atribuída ao conjunto padronizado de métodos STUN (“Session Traversal Utilities for NAT”), muito utilizado em aplicações de

voz, vídeo, mensagens e demais serviços que demandam comunicação em tempo real;

- 2082 – Porta atribuída para o serviço “Infowave Mobility Server”, pertencente a empresa “Infowave”, responsável por serviços de Tecnologia da Informação voltados para os setores da saúde, seguros e serviços públicos;
- 41000 – Não atribuída oficialmente a serviços ou aplicações conforme o portal da organização “IANA”;
- 8080 – Utilizada como porta alternativa para o protocolo HTTP, geralmente usada em servidores web;
- 4520 – Não atribuída oficialmente a serviços ou aplicações conforme o portal da organização “IANA”;
- 4659 – Porta registrada como a oficial do serviço “PlayStation2 Lobby”;
- 4340 – Porta atribuída ao serviço “Gaia”;
- 407771 – Não atribuída oficialmente a serviços ou aplicações conforme o portal da organização “IANA”.

Também foram verificadas as 10 portas que de forma geral estiveram presentes em mais fluxos, segue abaixo a lista de portas e suas respectivas atribuições conforme o “IANA”:

- 443 – Porta oficial do protocolo HTTPS, versão mais segura do protocolo HTTP;
- 53 – Porta atribuída oficialmente ao DNS (Sistema de Nomes de Domínio), utilizado em navegadores web para conversão de domínios em endereços IP;
- 80 – Porta oficial do protocolo HTTP, responsável pelo acesso WEB;
- 161 – Porta atribuída ao SNMP (Protocolo Simples de Gerência de Rede), protocolo responsável pelo gerenciamento de dispositivos em redes IP;
- 123 – Porta atrelada ao NTP (Protocolo de Tempo para Redes), protocolo responsável pela sincronização dos relógios dos dispositivos de uma rede;
- 5222 – Porta atribuída ao “xmpp-client”, o XMPP é um protocolo para sistemas de mensagens instantâneas utilizado em algumas aplicações como Google Talk e serviços de comunicação instantânea do Gmail;

- 1813 – Porta atribuída ao RADIUS Accouting, protocolo de rede utilizado por serviços de email, VPN e por provedores de internet no gerenciamento de acesso à internet ou intranet;
- 5228 – Porta atribuída ao serviço “HP Virtual Room”;
- 6881 – Porta não atribuída oficialmente a nenhum serviço conforme o “IANA”;
- 853 – Porta atribuída ao DNS sobre TLS, um dos padrões de criptografia para consultas DNS.

De forma geral, o mesmo padrão estabelecido nas listas de portas de origem e destino se repete, as portas atribuídas a serviços de acesso WEB (80 e 443) se destacam nas primeiras posições das listas de portas com maior volume de tráfego e maior presença em fluxos, evidenciando esse tipo de serviço como um dos quais realiza mais solicitações e trafega mais dados. As demais portas da lista com mais tráfego de dados são atribuídas a variados serviços (3478, 2082, 4659 e 4340) ou não estão atribuídas oficialmente a serviços ou protocolos (41000, 4520 e 47771).

Em relação ao restante das portas que estão na lista das que possuem maior presença em fluxos, são acentuadas portas que são atribuídas ao serviço de DNS (53 e 853), à protocolos que auxiliam na operação da rede (161, 123 e 1813), e à serviços de comunicação (5222 e 5228).

O conhecimento sobre quais portas são mais utilizadas na rede, tanto em termos de volume de tráfego quanto na quantidade de requisições, é um recurso valioso para os administradores de rede. Isso não apenas facilita a identificação de padrões de tráfego e serviços em uso, como também auxilia na detecção de possíveis vulnerabilidades de segurança. As portas identificadas nas listas anteriores revelam o uso predominante de serviços de comunicação e acesso WEB no tráfego dos clientes, o que é um comportamento esperado. No entanto, também foi observada uma utilização significativa de portas que não estão oficialmente registradas para nenhum serviço ou protocolo, o que exige uma análise mais aprofundada para identificar possíveis ataques ou falhas de segurança.

10 CONCLUSÃO

O monitoramento do tráfego em redes de computadores é uma área de estudo fundamental para a manutenção e segurança de suas infraestruturas. Este trabalho destacou a relevância da tecnologia de exportação de fluxos para identificação e análise de padrões de tráfego, com ênfase no protocolo NetFlow. Através de sua capacidade de coletar dados detalhados sobre o comportamento da rede, o NetFlow se revela uma importante tecnologia para detectar anomalias, gerenciar a capacidade de rede e assegurar a qualidade do serviço em ambientes de alta demanda.

A avaliação prática da utilização de ferramentas NetFlow em cenários reais demonstra sua eficácia no auxílio da gestão de redes complexas. A implementação das ferramentas flow-tools e FlowViewer em um servidor para monitoramento e análise do tráfego de um provedor de internet, embora bem-sucedida, enfrentou dificuldades relacionadas a erros de código nos scripts dos softwares. Esses problemas demandaram um esforço adicional na identificação e correção dos erros, o que exigiu certo nível de conhecimento nas linguagens Shell Script e Perl, e tempo para ajustes. Apesar desses desafios, a superação das dificuldades demonstrou a viabilidade técnica das ferramentas, ressaltando a importância de uma fase de testes e validação minuciosa durante a implantação para garantir a sua eficácia.

A utilização do flow-tools em conjunto com o FlowViewer, se provou eficaz ao fornecer valiosas informações para análise de tráfego, especificamente dados como volume de tráfego, taxas de transmissão e portas utilizadas. A interface amigável e de fácil utilização do FlowViewer facilitou a obtenção e interpretação dos dados capturados pelo flow-tools, comprovando que é possível monitorar e analisar de forma efetiva o tráfego de redes complexas utilizando soluções gratuitas. Essa experiência reforça a viabilidade de adoção dessas ferramentas em cenários de provedores de internet, oferecendo uma alternativa acessível e eficiente para o monitoramento e gerenciamento do tráfego de rede.

Embora a eficiência das ferramentas flow-tools e FlowViewer para monitoramento e análise do tráfego de redes em cenários reais tenha sido comprovada, ainda existem muitas possibilidades a serem exploradas, dadas as vastas capacidades dessas ferramentas. O grande potencial oferecido por elas abre caminhos para novas implementações e estudos aprofundados, que podem expandir ainda mais suas funcionalidades e adaptabilidade a diferentes ambientes de rede. Portanto, futuros

trabalhos podem focar na exploração de outras aplicações e na ampliação do uso dessas ferramentas, visando otimizar o monitoramento de redes complexas e maximizar os benefícios obtidos com essas soluções. Nos tópicos abaixo estão algumas possibilidades a serem exploradas com o flow-tools e FlowViewer em trabalhos futuros:

- Explorar mais afundo as possibilidades de relatórios e gráficos do FlowViewer e do FlowGrapher.
- Configuração do utilitário FlowTracker para avaliar a eficácia prática dos Dashboards de monitoramento de tráfego em tempo real.
- Avaliar a eficácia das ferramentas em cenários com tráfegos maiores, por exemplo, através da coleta de fluxos de um roteador de borda ou coleta de fluxos de mais de um dispositivo exportador.
- Avaliar a eficácia prática das ferramentas na identificação e resolução de problemas de segurança.
- Explorar o potencial do FlowViewer enquanto ferramenta de código aberto para melhoria de sua interface e implementação de possíveis upgrades em suas funções.

11 REFERÊNCIAS

ABBASI, M., SHAHRAKI, A. and TAHERKORDI, A., “Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey”. **Computer Communications**, Amsterdam, Netherlands, v. 170, pp. 20-21, mar./2021. Disponível em: https://www.sciencedirect.com/science/article/pii/S0140366421000426. Acesso em: 16 de outubro de 2024.

ALKENANI, J. and NASSAR, K., “Network Monitoring Measurements for Quality of Service: A Review”. **Iraqi Journal for Electrical and Electronic Engineering**, Basrah, Iraq, v. 18, n2, pp. 37-38, jun./dez.2022. Disponível em: https://ijeee.edu.iq/Papers/Vol18-Issue2/1570801485.pdf. Acesso em: 22 de maio de 2024.

AL-SBOU, Y. A., “Wireless Networks Performance Monitoring Based on Passive-Active Quality of Service Measurements”. **International Journal of Computer Networks & Communications (IJCNC)**, Chennai, India, v. 12, n6, pp. 16-17, nov./2020. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3745449. Acesso em: 17 de outubro de 2024.

AZAB, et al., “Network Traffic Classification: Techniques, Datasets, and Challenges”. **Digital Communications and Networks**, Beijing, China, v. 10, n3, pp. 677-679, jun./2024. Disponível em: https://www.sciencedirect.com/science/article/pii/S2352864822001845. Acesso em: 18 de outubro de 2024.

BENES, T., PESEK, J. and ČEJKA, T., **"Look at my Network: An Insight into the ISP Backbone Traffic"**. In: 2023 19th International Conference on Network and Service Management (CNSM), Niagara Falls, ON, Canada, pp. 4-7 (2023). Disponível em: https://dl.ifip.org/db/conf/cnsm/cnsm2023/1570928339.pdf. Acesso em: 18 de outubro de 2024.

BENMUSA, T., LAYAS, A. and ASBALI, A., **“Internet Usage Patterns and Traffic Analysis in Libya using Deep Packet Inspection Tools”**. In: The International Libyan Conference for Information and Communications Technologies (ILCICT 2022), Tripoli, Libya, pp. 39-43 (2022). Disponível em: https://www.researchgate.net/publication/360212941_Internet_Usage_Patterns_and_Traffic_Analysis_in_Libya_using_Deep_Packet_Inspection_Tools. Acesso em: 18 de outubro de 2024.

BÖTTGER, T., IBRAHIM, G. and VALLIS, B., **“How the Internet reacted to Covid-19 – A perspective from Facebook’s Edge Network”**. In: Proceedings of the ACM Internet Measurement Conference (ACM IMC 2020), Virtual Event, USA, p. 39 (2020). Disponível em: https://dl.acm.org/doi/10.1145/3419394.3423658. Acesso em: 22 de maio de 2024.

CHOO, M., **The FreeBSD Project**, 2023. “About FreeBSD”. Disponível em: https://www.freebsd.org/about/. Acesso em: 21 de maio de 2024.

CISCO SYSTEMS, **“Introduction to Cisco IOS® NetFlow—A Technical Overview”**. pp. 1-6, 2006. (Whitepaper). Disponível em: http://www.service-desk.co/white_papers/cisco_netflow.pdf. Acesso em: 21 de maio de 2024.

CLEMM, A., ZHANI, M. F. and BOUTABA, R., “Network Management 2030: Operations and Control of Network 2030 Services”. **Journal of Network Systems Management**, Netherlands, v. 28, p. 726, out./2020. Disponível em: https://link.springer.com/article/10.1007/s10922-020-09517-0. Acesso em: 17 de outubro de 2024.

CONTI, et al., “The Dark Side(-Channel) of Mobile Devices: A Survey on Network Traffic Analysis”. **IEEE Communications Surveys & Tutorials**, USA, v. 20, n4, p. 2658, jun./2018. Disponível em: https://ieeexplore.ieee.org/document/8371242. Acesso em: 16 de outubro de 2024.

D’ALCONZO, et al., “A Survey on Big Data for Network Traffic Monitoring and Analysis”. **IEEE Transactions on Network and Service Management**, USA, v. 16, n3, pp. 800-807, set./2019. Disponível em: https://ieeexplore.ieee.org/document/8789667. Acesso em: 17 de outubro de 2024.

DO CARMO, M., “Coleta e Análise de Tráfego em Redes de Computadores”. **Revista Especialize On-line IPOG**, Goiânia, v. 17, n1, p.10, jul./2019. Disponível em: https://ipog.edu.br/institucional/academico/revista-especialize. Acesso em: 21 de maio de 2024.

EHRLICH, et al., **“Passive Flow Monitoring of Hybrid Network Connections regarding Quality of Service Parameters for the Industrial Automation”**. In: Kommunikation in der Automation (KommA) 2017, Magdeburg, Deutschland, pp. 1-10 (2017). Disponível em: https://www.researchgate.net/publication/321098055_Passive_Flow_Monitoring_of_Hybrid_Network_Connections_regarding_Quality_of_Service_Parameters_for_the_Industrial_Automation. Acesso em: 17 de outubro de 2024.

FELDMANN, et al., “A Year in Lockdown: How the Waves of COVID-19 Impact Internet Traffic”. **COMMUNICATIONS OF THE ACM**, New York, NY, USA, v. 64, n7, pp. 101-107, jul./2021. Disponível em: https://dl.acm.org/toc/cacm/2021/64/7. Acesso em: 21 de maio de 2024.

FOWDUR, T.P and BABOORAM, L. “Network Traffic Monitoring and Analysis”. In: **Machine Learning for Network For Network Traffic and Video Quality Analysis.** 1 ed. Berkeley, CA, USA: Apress, 2024. pp. 51-96.

HAFFEY, M., ARLITT, M. and WILLIAMSON, C., **"Modeling, Analysis, and Characterization of Periodic Traffic on a Campus Edge Network"**. In: 2018 IEEE 26th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Milwaukee, WI, USA, pp. 170-172 (2018). Disponível em: https://cspages.ucalgary.ca/~cwill/papers/2018/Mack-MASCOTS2018.pdf. Acesso em: 18 de outubro de 2024.

HOFSTEDE, et al., “Flow Monitoring Explained: From Packet Capture to Data Analysis With NetFlow and IPFIX". **IEEE Communications Surveys & Tutorials**, Nanyang

Technological University, Singapore, vol. 16, n4, pp. 2037-2040, maio/2014. Disponível em: https://is.muni.cz/publication/1181098/flow-monitoring-explained-paper.pdf. Acesso em: 21 de maio de 2024.

IANA, **"Service Name and Transport Protocol Port Number Registry"**. 2024. (RFC6335). Disponível em: https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.txt. Acesso em: 21 de maio de 2024.

IX.BR, **IX.BR**, 2020. "IX.br alcança marca de 10 Tb/s de pico de tráfego Internet". Disponível em: https://ix.br/noticia/releases/ix-br-alcanca-marca-de-10-tb-s-de-pico-de-trafego-internet. Acesso em: 21 de maio de 2024.

IX.BR, **IX.BR**, 2021. "IX.br bate recorde histórico ao atingir 16 Tbit/s de pico de tráfego Internet". Disponível em: https://ix.br/noticia/releases/ix-br-bate-recorde-historico-ao-atingir-16-tbit-s-de-pico-de-trafego-internet. Acesso em: 21 de maio de 2024.

IX.BR, **IX.BR**, 2021. "IX.br chega a 20 Tbit/s de pico de tráfego, nova marca histórica". Disponível em: https://ix.br/noticia/releases/ix-br-chega-a-20-tbit-s-de-pico-de-trafego-nova-marca-historica. Acesso em: 21 de maio de 2024.

IX.BR, **IX.BR**, 2023. "Em nova marca recorde, IX.br ultrapassa os 31 Tbit/s de pico de troca de tráfego Internet". Disponível em: https://ix.br/noticia/releases/em-nova-marca-recorde-ix-br-ultrapassa-os-31-tbit-s-de-pico-de-troca-de-trafego-internet. Acesso em: 21 de maio de 2024.

JAIN, V., **"Wireshark Fundamentals: A Network Engineer's Handbook to Analyzing Network Traffic".** 1ª ed. Berkeley: Appress, 2022. pp. 5-6.

JI, SY., JEONG, BK. and JEONG, D.H. "Evaluating Visualization Approaches to Detect Abnormal Activities in Network Traffic Data". **International Journal of Information Security**, Heidelberg, Deutschland, v. 20, pp. 331–334, jun./2021. Disponível em: https://www.researchgate.net/publication/341582937_Evaluating_visualization_approaches_to_detect_abnormal_activities_in_network_traffic_data. Acesso em: 18 de outubro de 2024.

JOSHI, et al., "Network Traffic Analysis Measurement and Classification Using Hadoop". **International Journal of Advanced Research in Computer and Communication Engineering**, Pune, India, v. 5, n3, pp. 246-247, mar./2016. Disponível em: https://www.ijarcce.com/upload/2016/march-16/IJARCCE%2060.pdf. Acesso em: 16 de outubro de 2024.

KOUMAR, J. and ČEJKA, T., **"Network Traffic Classification Based on Periodic Behavior Detection".** In: 2022 18th International Conference on Network and Service Management (CNSM), Thessaloniki, Greece, p. 374-375 (2022). Disponível em: https://dl.ifip.org/db/conf/cnsm/cnsm2022/52.pdf. Acesso em: 17 de outubro de 2024.

KOUMAR, J., HYNEK, K. and ČEJKA, T., **"Network Traffic Classification Based on Single Flow Time Series Analysis".** In: 2023 19th International Conference on Network and Service Management (CNSM), Niagara Falls, ON, Canada, p. 1 (2023). Disponível em: https://ieeexplore.ieee.org/document/10327876. Acesso em: 18 de outubro de 2024.

KUROSE, J. F. and ROSS, K. W., **“Redes de computadores e a internet: uma abordagem top-down”**. 8ª ed. São Paulo: Pearson; Porto Alegre: Bookman, 2021. pp. 2-4.

LEE, Y. and LEE, Y., “Toward Scalable Internet Traffic Measurement and Analysis with Hadoop”. **ACM SIGCOMM Computer Communication Review**, New York, NY, USA, v. 43, n1, p. 7, jan./2013. Disponível em: https://dl.acm.org/doi/10.1145/2427036.2427038. Acesso em: 18 de outubro de 2024.

LIU, J., LIU, F. and ANSARI, N., "Monitoring and Analyzing Big Traffic Data of a Large-scale Cellular Network with Hadoop". **IEEE Network,** USA, v. 28, n4, pp. 32-34, jul./ago.2014. Disponível em: https://web.njit.edu/~ansari/papers/14Network.pdf. Acesso em: 17 de outubro de 2014.

LUCAS, M. W., **“NETWORK FLOW ANALYSIS”**. San Francisco: no starch press, 2010. pp. 11-151.

MANOHAR, V., “Comparative Study on Network Monitoring Tools”. **International Research Journal of Engineering and Technology (IRJET)**, India, v. 7, n4, p. 299, abr./2020. Disponível em: https://www.irjet.net/archives/V7/i4/IRJET-V7I464.pdf. Acesso em: 18 de outubro de 2024.

MISTRY, et al., **“Network Traffic Measurement and Analysis”.** In: IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, NY, USA, p. 1 (2016). Disponível em: https://ieeexplore.ieee.org/abstract/document/7494141. Acesso em: 17 de outubro de 2024.

NIC.BR, **NIC.BR**, 2021. “Cresce o uso de Internet durante a pandemia e número de usuários no Brasil chega a 152 milhões, é o que aponta pesquisa do Cetic.br”. Disponível em: https://www.nic.br/noticia/releases/cresce-o-uso-de-internet-durante-a-pandemia-e-numero-de-usuarios-no-brasil-chega-a-152-milhoes-e-o-que-aponta-pesquisa-do-cetic-br/. Acesso em: 22 de maio de 2024.

NIC.BR, **“TIC Domicílios 2023”**. São Paulo, p. 12, 2023. (Coletiva de Imprensa) Disponível em: https://cetic.br/media/analises/tic_domicilios_2023_coletiva_imprensa.pdf. Acesso em: 20 de maio de 2024.

PADILLA, J. J., “Análisis del Comportamiento del Tráfico en Internet Durante la Pandemia del Covid-19: el Caso de Colombia”. **Entre Ciencia e Ingenieria**, Pereira, Colombia, v. 14, n28, pp. 26-30, dez./2020. Disponível em: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1909-83672020000200026&lng=en&nrm=iso. Acesso em: 18 de outubro de 2024.

PONCE, D., TIPANTUÑA, C. and ESPINOSA, C., “Analysis of Internet Traffic in Ecuador”. **IEEE Access**, USA, v. 11, pp. 126370-126371, out./2023. Disponível em: https://ieeexplore.ieee.org/abstract/document/10313312. Acesso em: 17 de outubro de 2024.

SANTOS, et al., **“An Efficient Approach for Device Identification and Traffic Classification in IoT Ecosystems”**. In: 2018 IEEE Symposium on Computers and Communications (ISCC), Natal, Brazil, p. 00308 (2018). Disponível em:

https://www.computer.org/csdl/proceedings-article/iscc/2018/08538630/17D45We0UCS. Acesso em: 18 de outubro de 2024.

SANTOS, O., **"Network Security with NetFlow and IPFIX: Big Data Analytics for Information Security"**. Indianapolis: Cisco Press, 2015. p. 1-81.

SHAHRAKI, et al., "Active Learning for Network Traffic Classification: A Technical Study". **IEEE Transactions on Cognitive Communications and Networking**, USA, v. 8, n1, p. 422, mar./2022. Disponível em: https://ieeexplore.ieee.org/document/9566310. Acesso em: 18 de outubro de 2024.

SHEN, et al., **"**Machine Learning-Powered Encrypted Network Traffic Analysi: A Comprehensive Survey". **IEEE Communications Surveys & Tutorials**, USA, v. 25, n1, pp. 791-798, set./2022. Disponível em: http://www.thucsnet.com/wp-content/papers/meng_comst2022.pdf. Acesso em: 17 de outubro de 2024.

SPLUNDER, J. V., **"Periodicity Detection in Network Traffic"**. 2015. 90 f. Tese (Master in Mathematics (MSc)) – Universiteit Leiden, Leiden, p. 18, 2015. Disponível em: https://studenttheses.universiteitleiden.nl/handle/1887/3597183. Acesso em: 18 de outubro de 2024.

TANENBAUM, A., FEAMSTER, N. and WETHERALL D., **"REDES DE COMPUTADORES"**. 6ª ed. São Paulo: Pearson; Porto Alegre: Bookman, 2021. pp. 1-17.

TREMEL, et al., **"VITALflow: Visual Interactive Traffic Analysis with NetFlow".** In: NOMS 2022 - 2022 IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, p. 1 (2022). Disponível em: https://sappan-project.eu/wp-content/uploads/2022/04/VITALflow_2022.pdf. Acesso em: 18 de outubro de 2024.

TREVISAN, et al., "Five Years at the Edge: Watching Internet From the ISP Network". **IEEE/ACM Transactions on Networking**, USA, v. 28, n2, pp. 561-568, abr./2020. Disponível em: https://ieeexplore.ieee.org/abstract/document/8976293. Acesso em: 17 de outubro de 2024.

VELAN, P. and JIRSIK, T., **"On the Impact of Flow Monitoring Configuration".** In: NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, pp. 1-6 (2020). Disponível em: https://ieeexplore.ieee.org/document/9110361. Acesso em: 18 de outubro de 2024.

WASSERMAN, S. A., **"Machine Learning for Network Traffic Monitoring and Analysis"**: Application to Internet QoE Assessment and Network Security. 2022. 167 f. Tese (Doctor of Technical Sciences) – Technische Universität Wien, Vienna, p. 12, 2022. Disponível em: https://repositum.tuwien.at/handle/20.500.12708/20297. Acesso em: 17 de outubro de 2024.

YAHYAOUI, H. and ZHANI, M. F., **"On Providing Low-cost Flow Monitoring for SDN Networks"**. In: 2020 IEEE 9th International Conference on Cloud Networking (CloudNet), Piscataway, NJ, USA, p. 1 (2020). Disponível em: https://www.researchgate.net/publication/346298205_On_Providing_Low-cost_Flow_Monitoring_for_SDN_Networks. Acesso em: 18 de outubro de 2024.

YAMANSAVASCILAR, et al., **“Application Identification via Network Traffic Classification”**. In: 2017 International Conference on Computing, Networking and Communications (ICNC), Silicon Valley, CA, USA, p. 843 (2017). Disponível em: http://www.barisyamansavascilar.com/media/pdfs/2017_-_Application_Identification_via_Network_Traffic_Classification.pdf. Acesso em: 18 de outubro de 2024.