Redes Neurais Aplicadas à Classificação de Tráfego de Redes de Computadores Utilizando os Protocolos TCP e UDP

Diego José Oliveira, Angela Abreu Rosa de Sá

Instituto de Informática – Centro Universitário do Triângulo
Uberlândia, MG – Brasil

oliveiraj.diego@gmail.com, angelaabreu@gmail.com

Abstract. *This article presents the use of the Artificial Neural Network (ANN) on measurement and classification of traffic data using TCP and UDP transport protocols. I has been used the Adaline and Perceptron ANN to classify a data network traffic in the following categories: Source IP, destination IP, source port, destination port, type protocol (TCP or UDP and if the(connection started from the internal or external network. The implemented system allowed the traffic classification process can be done in a more dynamic way. Thus, the use of Artificial Neural Networks for classification and measurement of traffic on computer networks, using TCP and UDP transport protocols, was considered a productive way and applicable to the problem of classification of traffic data.*

Resumo. *Este artigo apresenta o uso da Rede Neural Artificial (RNA) na mensuração e classificação de dados trafegados utilizando os protocolos de transporte TCP e UDP. Foram utilizadas as RNAs Adaline e Perceptron para classificar o tráfego de uma rede de dados nas seguintes categorias: IP de origem, IP de destino, porta de origem, porta de destino, tipo do protocolo (TCP ou UDP) e se a conexão partiu da rede interna ou externa. O sistema implementado permitiu que o processo de classificação do tráfego seja feito de forma mais dinâmica. Assim, o uso de Redes Neurais Artificiais para a classificação e mensuração de tráfego de redes de computadores, utilizando protocolos de transporte TCP e UDP, foi considerado uma maneira produtiva e aplicável ao problema de classificação de dados trafegados.*

1. Introdução

Com o crescente e diversificado uso da Internet em empresas de todos os portes, um fator que demanda bastante atenção dos responsáveis por redes corporativas é conhecer/definir o perfil de utilização desta rede. Esta tarefa é realizada através da coleta e monitoramento dos mais diversos tipos de tráfego, sejam eles da rede interna para a externa (consumo web considerado mais comum em parte das empresas), da externa para a interna (geralmente quando a empresa em questão provê serviços para acesso externo) ou uma comunicação apenas interna, que por mais comum e numerosa que seja, também deve ser analisada de forma detalhada e atenta [Brownlee et al. 1999].

Desta forma, a análise de tráfego se torna algo simples a ser feito pelo responsável por uma rede de computadores, principalmente sendo ela corporativa. E

ainda, existem muitas ferramentas que dispõe das funcionalidades necessárias para que a execução desta tarefa seja possível, como, por exemplo,Wireshark [Wireshark 2018] e Tcpdump [Tcdump 2018], as quais são capazes de efetuar uma coleta detalhada e completa de todo o tráfego de uma rede, para que possa posteriormente ser analisado. Porém, esta análise depende de conhecimento aprofundado de quem a executa e, principalmente, do tempo dedicado; tornando inviável caso esta coleta seja efetuada diariamente, mesmo em redes de pequeno porte [Lima et al. 2011].

Algumas pesquisas visam melhorar a forma com que a análise de tráfego de redes é realizada, e utilizam Redes Neurais Artificiais para conseguir melhores resultados, conforme foi apresentado em [Barros 2012], [Silva and Júnior 2014] e [Lima et al. 2011]. Através desta abordagem, se torna possível uma análise contínua em uma faixa temporal capaz de definir o perfil de uso de uma determinada rede de computadores, classificar o tráfego, mensurar o seu uso e identificar padrões desconhecidos utilizando a Rede Neural, o que pode significar um acesso não permitido ou anomalia na rede. E ainda, utilizando esta análise é possível efetuar tomadas de decisões com mais eficácia e mais assertividade.

Neste contexto, a proposta deste trabalho é a utilização de dois tipos de Redes Neurais capazes de suprir esta necessidade básica inicial de uma análise de grande volume de dados, sem que haja a necessidade de uma análise completamente manual ou que dependa de profissionais com grande conhecimento técnico para tal finalidade. Foram implementadas as Redes Neurais Adaline e Perceptron com a finalidade de demonstrar o poder de classificação de padrões destes modelos.

Nesse contexto, para apresentar o trabalho desenvolvido, o presente artigo está organizado na seguinte estrutura: na seção 2 foi realizada a revisão da literatura, descrevendo os principais conceitos sobre mensuração e classificação de tráfico de redes, bem como os principais trabalhos que já foram publicados neste tema. Em seguida, na seção 3, é apresentada a teoria e a arquitetura de redes neurais artificiais e, também, os trabalhos publicados na área de classificação de tráfego de redes utilizando a técnica de redes neurais. Já na seção 4, é descrita, em detalhes, a metodologia utilizada para a definição e implementação do sistema de classificação de dados que é proposto neste artigo. Os resultados que foram obtidos com a implementação do sistema desenvolvido, são apresentados na seção 5. E, por fim, na seção 6, são apresentadas as conclusões que foram obtidas a partir dos resultados do sistema proposto.

2. Mensuração e Classificação de Tráfego de Redes

Diferente do uso doméstico das redes de computadores, o uso de redes em ambientes corporativos exige cuidados diferenciados, como priorização de tráfego, limitações de banda para determinadas saídas, armazenamento dos logs de acessos a Internet e outros requisitos que garantem o bom funcionamento dessa rede. A coleta de dados sobre esse tráfego se faz necessária pois através desta é possível classificá-lo de diferentes maneiras, de acordo com as necessidades: protocolos, aplicações, horário, usuários, sessões estabelecidas etc. Estas classificações podem ser utilizadas posteriormente para efetuar a correlação destes dados, que antes estavam em uma forma bruta (ainda sem

tratamento) e utilizar para comparações com padrões já armazenados, em busca de anomalias ou mudanças que possam prejudicar ou afetar de alguma forma o funcionamento da rede [Brownlee et al. 1999].

Através da análise do tráfego de uma rede é possível, com a utilização das técnicas corretas, obter dados relevantes que podem auxiliar a gerência da mesma. A necessidade de ambientes de redes íntegros, de alta disponibilidade e escalabilidade, faz com que empresas e profissionais da área direcionem mais esforços na coleta deste tráfego, para que este possa ser mensurado e classificado, de forma a agregar significado aos dados extraídos e utilizar os resultados para diversos fins, como expansão ou redução no cenário atual da rede, correlação de padrões, identificação de problemas e predição de uso do tráfego [Júnior 2012].

2.1 Aplicabilidade da Classificação de Tráfego

A Internet é um dos principais meios de comunicação utilizado por empresas de pequeno, médio e grande porte, fazendo parte desde a infraestrutura principal da empresa, que mantém suas informações, até os produtos oferecidos por ela, como marketing, serviços e vários outros. Para manter um ambiente como este com o mínimo de perdas, indisponibilidades e ocorrências inesperadas, é buscado cada vez mais efetuar a classificação do tráfego gerado em uma rede.

A demanda pela classificação e mensuração de tráfego tem evoluído bastante em relação às ferramentas e metodologias aplicadas a tal, já que em uma rede de médio e grande porte, o tráfego gerado parte de várias aplicações diferentes e com formas distintas de utilização dos recursos. Considerando que em um ambiente corporativo, onde o foco é destinado a produção, o impacto de um mau uso dos recursos da rede pode ser desastroso; é muito importante a coleta e classificação dos dados trafegados, afim de utilizá-los proativamente na gerência da rede [Júnior 2012].

2.2 Abordagens de Medição Ativas e Passivas

A medição ativa tem como forma de análise a injeção de pacotes na rede, sendo possível avaliar a qualidade do tráfego (QoS) da rede de acordo com um tráfego “comum”, ou com um tráfego específico, como voz, vídeo, etc. Esta abordagem permite que junto com a especificação do tráfego gerado seja feita uma análise da rede, possibilitando modelar o fluxo desse tráfego, gerando um estresse na rede, para identificar o quanto de tráfego a configuração atual da rede pode comportar ou quando, de acordo com a vazão do tráfego, começará a aparecer anomalias, caso existam [Silva and Júnior 2014].

A técnica de medição passiva, a qual será adotada neste trabalho, é utilizada para entender o comportamento da rede em seu estado atual, ou seja, ela é baseada na observação do tráfego de pacotes, sem que haja interferência em seu comportamento. Os dados coletados são salvos em uma base/repositório, para que possam ser classificados e analisados posteriormente. É importante ressaltar que esta captura é geralmente feita no centro da comunicação, ou seja, em um switch, roteador, ou outro equipamento que esteja interligado aos dois nós que se comunicam, sendo capaz de ter a

visão do tráfego de forma completa, desde o início até o fim do mesmo, conforme apresentado na Figura 1 [Júnior 2012].

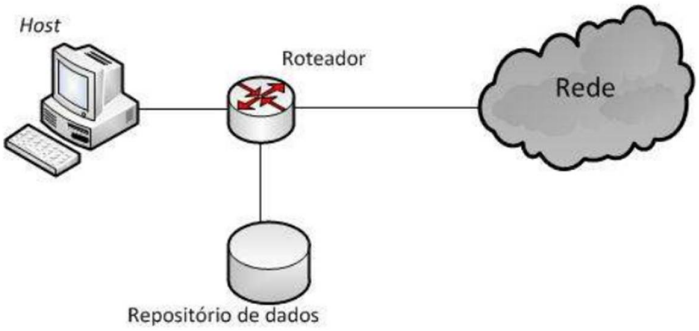

Figura 1. Coleta passiva entre um host e uma determinada rede [Júnior 2012]**.**

2.3 Técnicas de Classificação de Tráfego

Várias técnicas de classificação de tráfego são utilizadas com a finalidade de serem mais assertivas de acordo com cada necessidade. Um desses métodos é com base em portas e consiste em mapeá-las por aplicação, utilizando o cabeçalho do protocolo. Essa técnica, além de ser simples para implementação, tem uma certa vantagem no processamento das informações coletadas, por se tratar de uma separação utilizando classificadores estáticos (portas), que podem ser adicionados de acordo com as necessidades de classificação. Porém, a utilização desta técnica, se torna ineficiente quando se trata de aplicações que utilizam portas dinâmicas para efetuar sua conexão [Zander et al. 2016].

Diferente da classificação com base em portas, a classificação por carga utiliza uma técnica de inspeção profunda do pacote (DPI - Deep Packet Inspection), que identifica qual protocolo da camada de aplicação está sendo usado através de valores hexadecimais específicos de cada um. Porém, esta técnica apesar de bastante acurada na classificação, tem um maior custo computacional e perde sua eficiência quando se trata de tráfego criptografado[Szabó et al. 2008].

Outra classificação, diferente das citadas acima, é a classificação com base em fluxo. Esta não utiliza informações do cabeçalho de um pacote para classificá-lo, pois trabalha com estatísticas geradas através das características do fluxo. Esta técnica utiliza algoritmos de aprendizagem de máquina para gerar um modelo de classificador, para que os dados que são coletados e salvos possam ser classificados[Barros 2012].

3. Redes Neurais Artificiais

O trabalho de pesquisa e evolução sobre Redes Neurais Artificiais (RNA) é baseado na busca pelo entendimento dos neurônios biológicos e no funcionamento do cérebro humano, em relação ao aprendizado. O cérebro humano tem a capacidade de efetuar o processamento de informações bem como controlar de entrada e saída das mesmas, de

forma paralela e evolutiva, em que aprende e armazena informações enquanto efetua outras tarefas. Este aprendizado é crescente e modificado de acordo com as experiências ao qual o cérebro obtém como entrada. Nesse contexto, o principal foco e objetivo das RNAs é conseguir reproduzir a capacidade de aprendizado do cérebro humano [Haykin 2006].

3.1 Arquitetura

As Redes Neurais Artificiais podem ser divididas em arquiteturas, de acordo com a quantidade de camadas que contém. Em uma rede neural de única camada, os únicos elementos são os nós de origem em que se encontram os dados que serão inseridos na rede diretamente sobre a camada de processamento, que é a responsável por entregar diretamente os resultados de saída. Esta rede é considerada de camada única, sendo que a camada referida é a de saída, já que a camada de entrada (nós de origem) não efetuam processamento algum sobre os dados [Haykin 2006], conforme exemplificado na Figura 2.

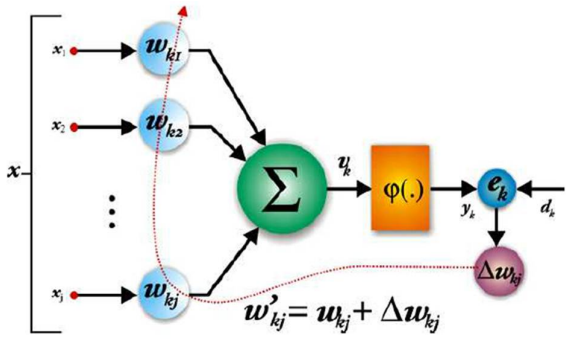

Figura 2. Rede Neural Artificial única camada [Lima 2005]**.**

A Rede Neural multicamadas se diferencia da anterior, pois além dos nós de entrada, esta contém *n* camadas ocultas e, posteriormente, uma camada de saída. Essa configuração permite que quanto maior a quantidade de camadas ocultas, maior a possibilidade de extração de estatísticas de ordem elevada. Seu funcionamento ocorre da seguinte maneira: dados os valores para entrada, estes são direcionados para a próxima camada, que será a camada oculta, que irá processar o valor para treinamento dos neurônios e, ao término do processamento, informará uma saída com os valores que serão utilizados como a entrada da próxima camada, seja ela a última ou não, até alcançar a camada de saída [Haykin 2006].

3.2 Tipos e Características

A Rede Neural Artificial do tipo Perceptron simples, é considerada a forma mais clássica de representação de uma rede neural linearmente separada. Ela contém, basicamente, apenas neurônios com seus pesos que podem ser ajustados. Este algoritmo foi apresentado em 1958 e foi demonstrado que o modelo Perceptron recebendo entradas retiradas de duas classes linearmente separáveis, é possível para o algoritmo convergir e posicionar um hiperplano entre as duas classes, tornando possível a

separação destas. Esta convergência, em relação aos valores inseridos para seu treinamento, é conhecida como o teorema de convergência do perceptron [Haykin 2006].

Já a RNA Perceptron multicamadas (MLP Multi-layer Perceptron), consiste de uma camada de neurônios e entrada, uma ou mais camadas ocultas, que não são parte de entrada ou saída da rede e capacitam a rede a aprender tarefas complexas pois consegue extrair características significativas dos valores de entrada. E em seguida, tem-se a camada de saída que recebe como entrada a saída da camada anterior. Uma das formas de treinamento desta rede é através do algoritmo de retro propagação de erro, conhecido como Backpropagation, que utiliza regra de aprendizagem por correção do erro e permite ajustar os erros encontrados durante o treinamento de forma mais precisa, gerando uma saída mais apurada da rede [Fausett 1993].

Praticamente na mesma época em que a Rede Neural Perceptron simples era estudada, Bernard Widrow desenvolvia o modelo conhecido como Adaline, na Universidade de Stanford. O nome Adaline é um acrônimo para **ADA**ptive **LIN**ear **E**lement [Kovács 2002]. Esta rede neural pode ser treinada utilizando o modelo de treinamento da regra delta, também desenvolvido por Bernard Widrow. Este algoritmo de aprendizagem minimiza o erro quadrático médio e o valor de destino, permitindo um maior espaço de aprendizagem para a rede neural. [Fausett 1993].

3.3 Trabalhos Relacionados

O uso da RNA Multi-layer Perceptron para classificação de QoS (Quality of Service) em conteúdo multimídia para rede virtual privada (VPN) foi apresentado por [Lima et al. 2011] com o intuito de classificar o desempenho de QoS em uma rede VPN, utilizando para testes, o tráfego sem multimídia e com multimídia, e como parâmetros, jitter, throughput, perda de pacotes e dados transferidos. Esta classificação foi efetuada em 3 níveis, alto, médio e baixo em relação a distribuição de conteúdo, após a coleta, análise e devidas classificações, os autores concluíram que o uso da RNA Multi-layer Perceptron apresentou bons resultados para o problema proposto em relação aos testes efetuados.

Outro trabalho relacionado, foi apresentado por [Oliveira 2014] sobre predição de tráfego, usando redes neurais artificiais, para gerenciamento adaptativo de largura de Banda em Roteadores, com o intuito de comparar o desempenho de três modelos de redes neurais (MPL, RNN, SAE) na predição de tráfego para gerenciamento de largura de banda em roteadores. Com o uso de três diferentes técnicas de redes neurais, foi concluído que todas conseguiram efetuar a predição e ajuste de tráfego conforme proposto, porém, foram obtidas diferentes demandas de tempo e poder computacional na fase de treinamento, sendo que a SAE consumiu mais tempo nos casos de teste da solução proposta, pois contém uma etapa extra com o treinamento não supervisionado.

Em [Lima 2014], foi apresentada a utilização de sistemas inteligentes para classificação de tráfego malicioso. Nesse trabalho foram descritos o uso de algoritmos

genéticos AG e redes neurais multicamadas MLP, com a finalidade de classificar tráfegos considerados maliciosos, utilizando sistemas inteligentes. Foram demonstradas várias técnicas possíveis para tal finalidade e comparado o resultado de cada uma em determinado cenário, com diferentes quantidades de dados e campos para treinamento e, posteriormente, classificação do tráfego coletado. Com os resultados de testes e comparações das técnicas utilizadas, foi concluído que o uso de mais de uma técnica de classificação para esta finalidade obtém resultados promissores, apontando assim para a contribuição do uso de redes neurais artificiais na classificação de tráfego malicioso em um cenário real de redes de computadores.

E também, o uso da inteligência artificial em conjunto com a mineração de dados foi apresentado em detalhes em [Santos 2011], onde é demonstrado os benefícios da aplicação de redes neurais e estatística, para a implementação de uma metodologia TRAFCIN (network Traffic Characterization on Computational INteligence), para análise, processamento e extração de informações no tráfego de redes de computadores, com o objetivo de identificar anomalias, falhas, características ruídos derivados deste tráfego gerado diariamente, e também definir um comportamento considerado padrão da rede, através da classificação destes dados, sendo possível traçar um desempenho aceitável dentro destes padrões. Concluiu-se que, de fato houve uma contribuição notável das técnicas aplicadas, obtendo resultados satisfatórios na detecção de anomalias, caracterização do tráfego e reconhecimento de seu comportamento, como pretendido.

Em [Lima 2005] foi apresentado o uso também do modelo de rede neural MPL (Multi-layer Perceptron) utilizando o algoritmo de treinamento *Backpropagation* para a apresentação de uma abordagem simplificada de detecção de intrusão em redes de computadores. Nesta pesquisa, foi proposto que após efetuar coletas de tráfego de rede por determinados espaços de tempo e com filtros que selecionaram o tráfego para análise, é feito um primeiro processamento utilizando o auxílio de um analisador semântico, e posteriormente é feito um pós-processamento que irá preparar os dados para a análise da rede neural, onde é feita a classificação do dado recebido e, assim, gera uma saída para que seja validada uma possível ocorrência de intrusão. Com os casos executados naquel trabalho, foram obtidos bons resultados no que diz respeito a capacidade da rede neural em identificar possíveis intrusões na rede, sendo constatado que mesmo com algumas diferentes variações de ataques utilizados nos dados para reconhecimento, ainda foi possível efetuar a classificação, algo que não seria possível utilizando alguma técnica estática para o mesmo cenário.

4. Metodologia

Esta sessão descreve de maneira detalhada como foi efetuada a implementação de todo o algoritmo responsável pela classificação de serviços derivados dos protocolos TCP e UDP, bem como as ferramentas externas utilizadas na captura, padronização e persistência de dados. E também, são apresentadas a lista de ferramentas e tecnologias utilizadas para que a toda implementação do algoritmo e o resultado final fosse obtido.

4.1 Tcpdump

O Tcpdump é uma programa executável que roda em linha de comando de sistemas com base em Unix, utilizado para captura de pacotes em tráfego de redes TCP/IP. O Tcpdump foi utilizado neste trabalho para fazer a captura do tráfego que foi analisado para treinamento/validação da rede neural. O Tcpdump foi executado por um período de aproximadamente 1 hora, onde o computador que contém o executável foi conectado em uma determinada porta de um switch em operação na rede, de preferência o switch principal, onde se concentra todo tráfego. Após essa etapa, todas as portas necessárias deste switch foram espelhadas para a porta onde o tcpdump estava "ouvindo". Este fluxo está ilustrado na Figura 3.

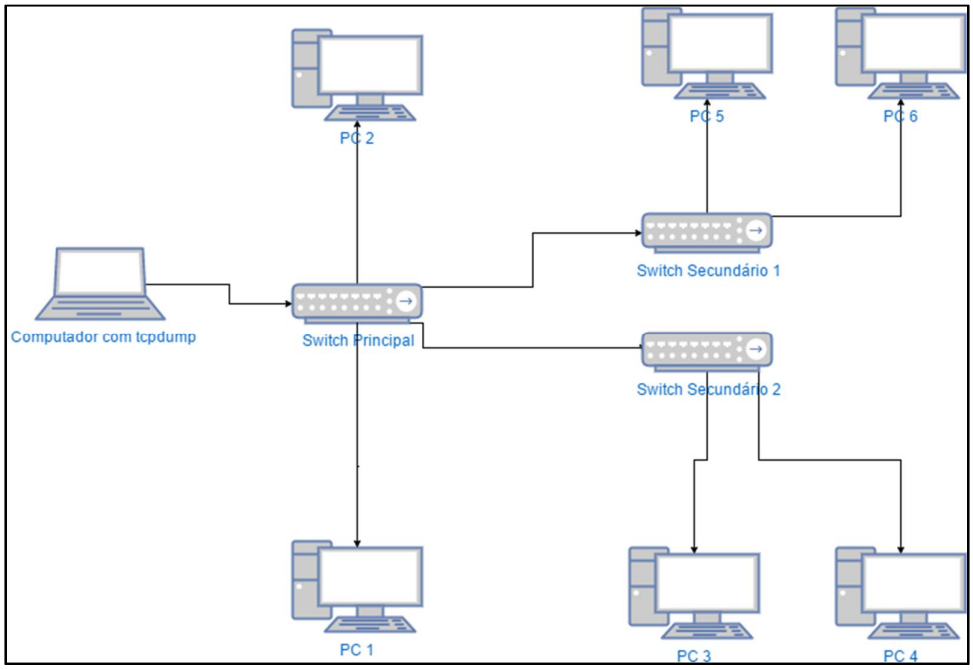

Figura 3. Demonstração de captura de tráfego de rede pelo tcpdump [Autoria própria 2017].

Quando o Tcpdump é executado com o mínimo de parâmetros necessários para seu funcionamento, a placa de rede do computador em questão é colocada em modo promiscuo, onde passa a capturar todo o tráfego que passa por ela. Então, o tcpdump faz a "tradução" de todos os dados que transitam por esta porta. A Figura 4 apresenta a linha de comando e os parâmetros utilizados para esta captura de tráfego.

```
# tcpdump -i ens33 -c 200 -nn tcp or udp > trafegoDeRede
```

Figura 3. Demonstração de captura de tráfego de rede pelo tcpdump [Autoria própria 2017].

O tráfego capturado é salvo em um arquivo para que possa ser padronizado e direcionado a um banco de dados relacional, o que facilitará a busca pelas informações necessárias para uso em posterior análise dos dados.

4.2 Linguagem C#

Foi utilizada a linguagem C# e o framework .NET devido à vasta documentação referente ao desenvolvimento de redes neurais em que a mesma é utilizada. E também, a disponibilidade de drivers de conexão com o SGBD (Sistema de Gerenciamento de Bando de Dados) PostgreSQL e a facilidade de criação de interface gráfica através do uso de Windows Form para facilitar a interação com usuário final, foram fatores que influenciaram na escolha destas tecnologias.

4.3 Extração e Padronização de Informações

Este módulo é responsável pela extração dos dados localizados no banco de dados (tabela tb_trafego_original) e conversão dos mesmos para o padrão que será recebido pela rede neural. Os campos utilizados e extraídos para conversão são: ip de origem, ip de destino, tipo do protocolo de transporte utilizado e porta de destino. Todos os dados são convertidos para um valor binário. Por exemplo, a porta 80 é representada pela sequência correspondente em binário, **01010000**, que passa por uma segunda substituição, em que os **0´s** são substituídos por **-1,** tornando possível o aprendizado por parte da rede neural. Estes dados são salvos na tabela tb_trafego_convertido, que contém os campos a serem utilizados e um id de referência para o tráfego original salvo na tabela tb_trafego_original. Este id de referência é utilizado posteriormente para identificação de cada conexão efetuada. Este fluxo está exemplificado na Figura 5.

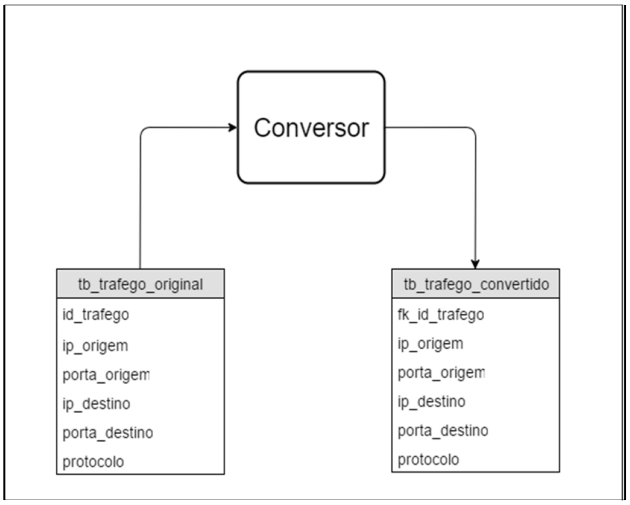

Figura 5. Fluxo de extração, conversão e preenchimento da tabela com o tráfego já convertido [Autoria própria, 2017].

4.4 PostgreSQL

O armazenamento da captura de tráfego e o tráfego já convertido é feito em um banco de dados relacional, utilizando o SGBD PostgreSQL. A escolha deste SGBD se justifica por se tratar de uma ferramenta livre, estável e que comporta uma grande quantidade de dados; sendo possível obter um desempenho aceitável sem exigir muita capacidade do hardware.

4.5 Estrutura da Rede Neural

A rede neural utilizada para efetuar a classificação dos dados de tráfego de rede é a Adaline. Todo o tráfego coletado é convertido para o padrão determinado no tópico 4.3 e é inserido para treinamento da rede neural. Após o treinamento da rede, o dados são inseridos para avaliação e classificação. A Figura 6 ilustra a estrutura da rede neural projetada e utilizada neste trabalho.

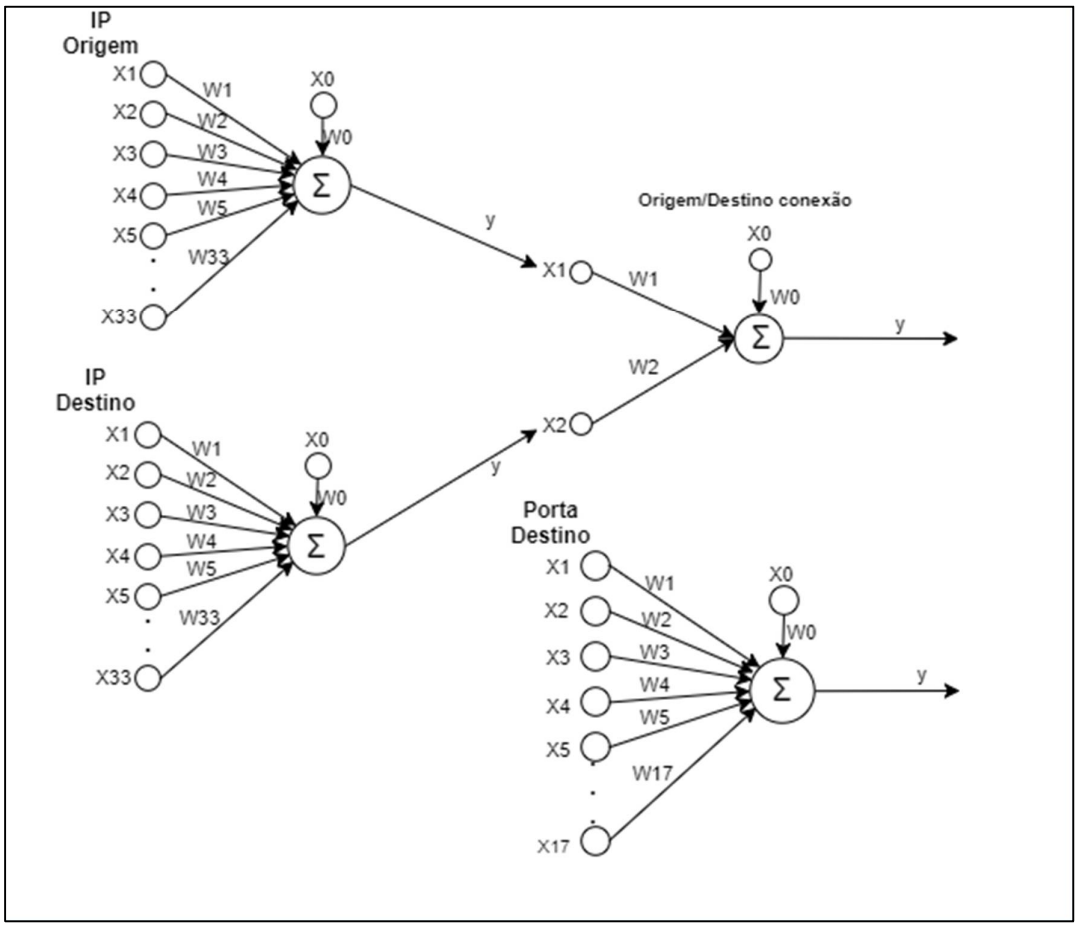

Figura 6. Fluxo de entrada, classificação e saída da rede neural para o tráfego de teste, inserido após a rede neural estar completamente treinada [Autoria própria, 2017].

4.6 Coleta de Dados

Para os testes efetuados neste trabalho, foi utilizado o tráfego coletado de uma rede empresarial de pequeno porte, onde nela contém os seguintes equipamentos: a) Três switches, sendo um principal responsável por receber todo tráfego de entrada e saída, e efetuar o repasse para o gateway e dois switches secundários responsáveis pelo tráfego de cada setor onde está alocado; b) Um firewall com a funcionalidade de filtro de pacotes, controle de acesso e gateway; c) Um PABX responsável por todo tráfego de telefonia; d) Um servidor de arquivos. Além desses equipamentos, a rede contém aproximadamente 30 computadores de acesso comum, utilizados nas atividades diárias e responsáveis pela maior parte do tráfego gerado. Os principais serviços desta rede são de desenvolvimento de software e suporte ao usuário final, onde existe um alto consumo principalmente de conteúdo HTTP e SIP, devido ao serviço de atendimento telefônico.

O treinamento das redes neurais utilizadas para a classificação do tráfego foi feito com base na faixa de IP´s utilizada na rede local, para que fosse possível identificar a origem e destino da conexão. Com relação às portas, foram utilizadas para aprendizagem um número total de 525 portas entre TCP e UDP, dentre estas as mais comumente utilizadas em comunicações de uma rede desse porte.

Para os testes realizados neste projeto foram utilizados dados fixos e conhecidos para o treinamento das redes neurais, como endereços IP da rede local analisada, portas consideradas mais utilizadas comumente nesta rede, e outras conhecidas. Os dados foram utilizados para validar o funcionamento da rede neural e sua capacidade de classificar os padrões definidos. E ainda, os dados foram coletados de uma rede com funcionamento real e adaptados ao formato necessário para que fossem utilizados.

4.7 Passos do Treinamento da Rede Neural

O processo de treinamento das redes neurais se divide nas seguintes etapas (Figura 7): a) Definição da faixa de IP's da rede local e geração de todos os endereços IP válidos para hosts disponíveis na rede em um arquivo; b) Definição de quais portas serão utilizadas como referência para treinamento, o que ajuda a diminuir o treinamento desnecessário na rede neural, porém pode ser utilizada a quantidade de portas que se julgue necessário. Os protocolos de transporte utilizados para esta análise foram TCP e UDP.

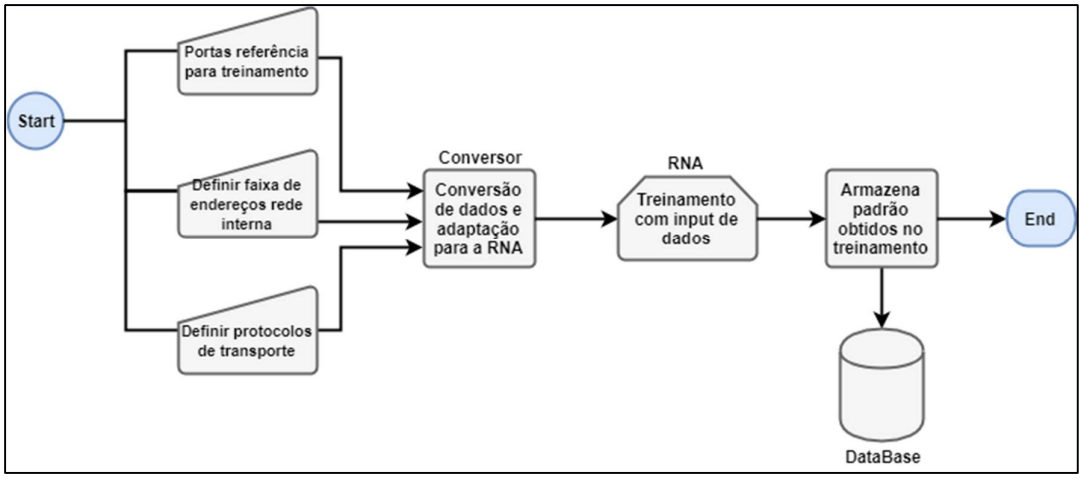

Figura 7. Fluxo de entrada e conversão de dados para o treinamento da rede neural [Autoria própria, 2017].

4.8 Fluxo de classificação de dados

Após a rede neural estar treinada e ter convergido para a melhor classificação possível, é efetuada a conversão do tráfego de rede coletado e esses dados são avaliados na rede neural para classificação. O fluxo desse procedimento é apresentado na Figura 8.

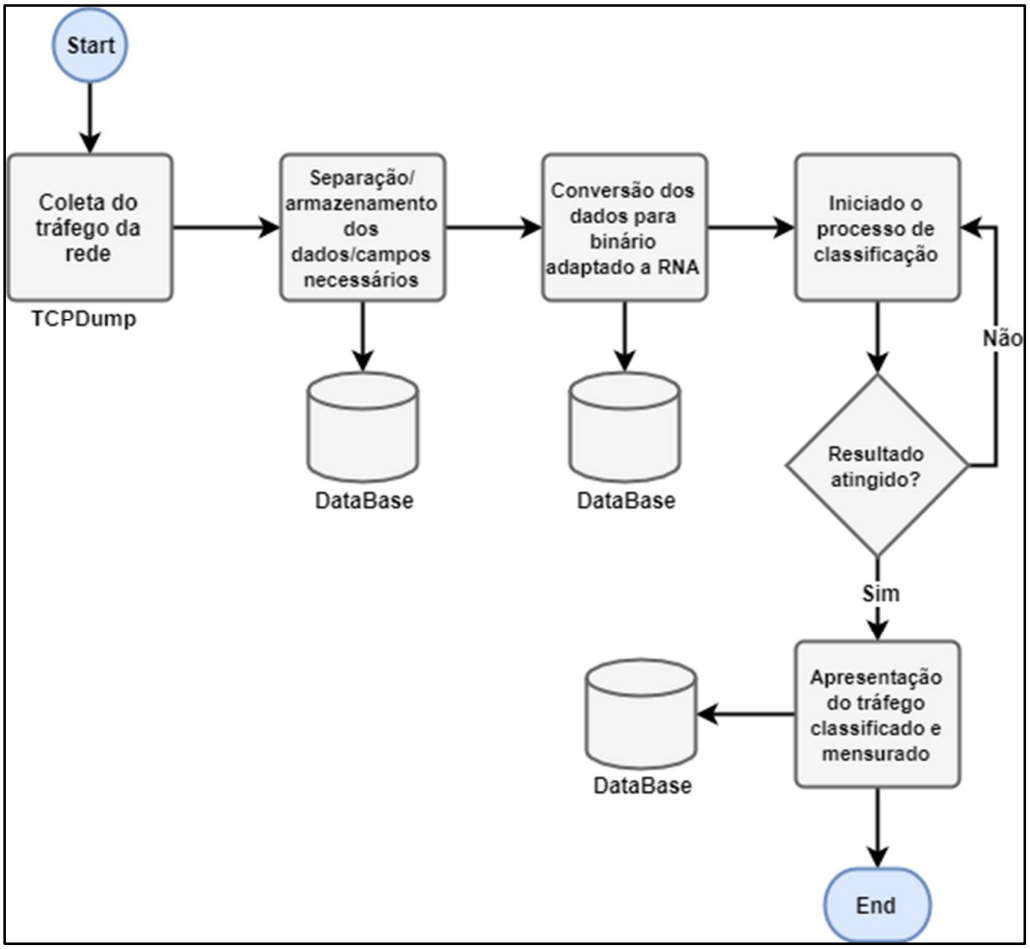

Figura 8. Fluxo de coleta, conversão do tráfego de rede utilizado para classificação e mensuração do tráfego TCP/UDP [Autoria própria, 2017].

Após completar esta fase, que é decisiva para definir a eficiência do treinamento efetuado que foi efetuado, o processo de classificação está concluído. Desta forma, é possível através do conjunto de dados de saída, definir a quantidade de conexões partindo da rede local para externa, qual o tipo de conexão mais predominante dentre o tráfego coletado, qual o protocolo de transporte mais utilizado, se existem muitas conexões partindo da rede externa para a interna e quais IP´s/serviços estão sendo procurados.

4.9 Sistema Desenvolvido

A aplicação foi desenvolvida com intuito de ser de fácil utilização, tanto pra profissionais experientes quanto para profissionais com menos conhecimento específico do tema. A tela inicial contém as opções necessárias para o usuário treinar a Rede Neural (Figura 9), que tem em destaque as funcionalidades iniciais, bem como um número que descreve a função do objeto.

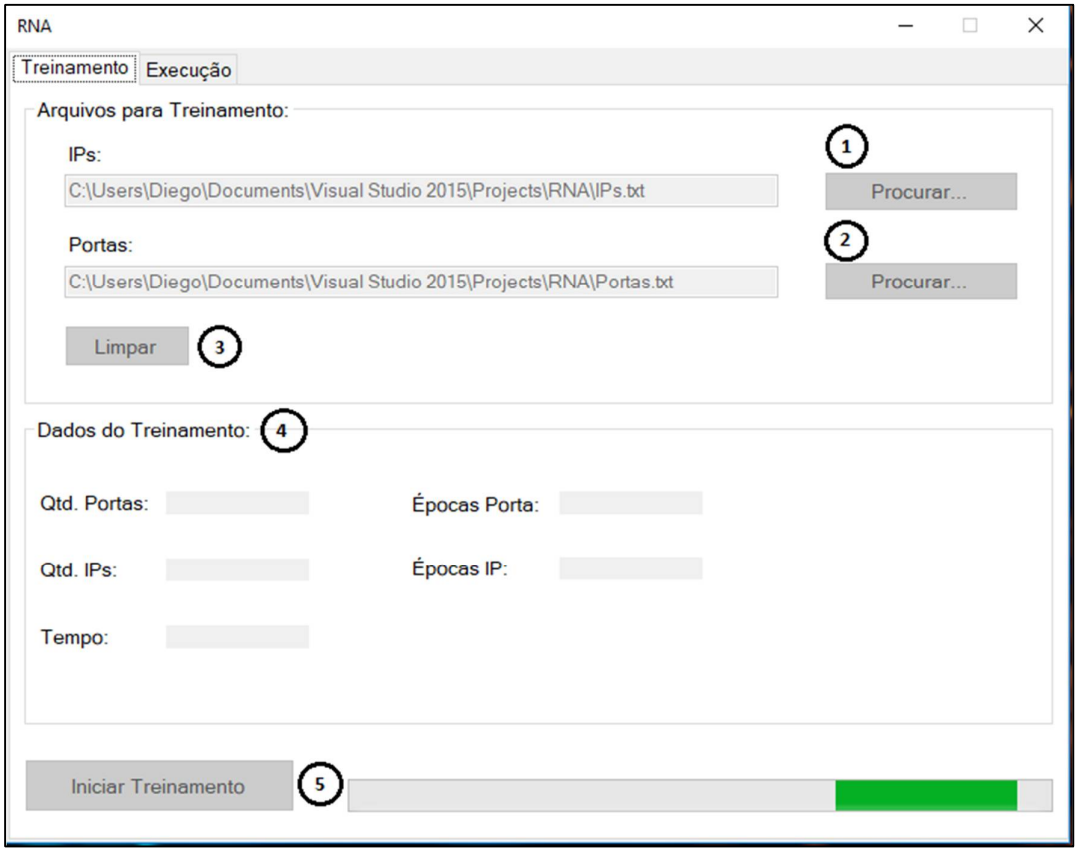

Figura 9. Tela inicial da aplicação [Autoria própria, 2017].

A interface apresentada na Figura 9 é a responsável pela interação com o usuário que será responsável pela inserção dos dados iniciais para treinamento. Nesta tela serão inseridos os dados referente à faixa de IP´s da rede interna e as portas utilizadas para treinar a rede neural. Após esta inserção, será possível iniciar o treinamento ou limpar as entradas de texto que serão preenchidas com o caminho do arquivo utilizado. As funcionalidades mais importantes dessa tela são:

1. Botão “Procurar”: responsável por selecionar o arquivo de texto contendo a faixa de IP´s para treinamento.

2. Botão “Procurar”: obtém o arquivo de texto contendo as portas para treinamento.

3. Botão “Limpar”: Anula a seleção dos arquivos que foram selecionados usando os botões 1 e 2.

4. Dados obtidos após a conclusão do treinamento. Eles mostram a quantidade de entradas para portas e IP´s utilizada, bem como o tempo gasto e a quantidade de ciclos que cada tipo de entrada gastou no treinamento.

5. Botão que inicia o treinamento das redes neurais. Após iniciado, uma barra de progresso mostra que o processo está em andamento.

Já a segunda tela (Figura 10), é responsável pela entrada dos dados que serão classificados após as redes neurais estarem completamente treinadas. Após clicar no botão "Iniciar classificação", é apresentada de forma legível a mensuração do tráfego classificado, onde são mostradas os 10 prováveis tipos de conexão da rede, e a quantidade em porcentagem de conexões de entrada e saída.

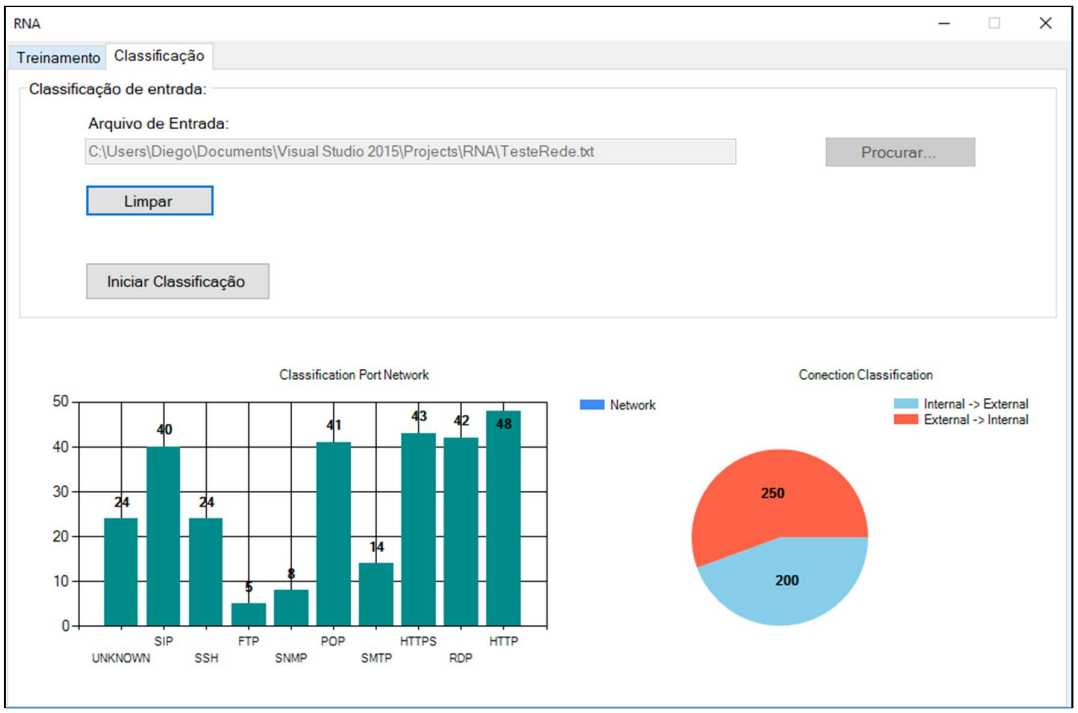

Figura 10. Tela de classificação e acompanhamento da mensuração de tráfego pela aplicação [Autoria própria, 2017].

5. Resultados

As métricas capturadas com a aplicação em execução, bem como a quantidade de dados utilizada para treinamento, estão apresentadas na Tabela 1. Estes valores podem variar de acordo com a capacidade de processamento do computador que os recebe para execução.

Tabela 1. Métricas coletadas da fase de treinamento [Autoria própria, 2017].

Tipo	Quantidade	Tempo	Ciclos	Taxa de Aprendizagem	Precisão
Porta	100	00:10:14s	232309	0.0000001	0.000000001
IP	354	00:03:10s	1989	0.0005	0.00001
Origem/destino	50	00:01:14s	1000	0.001	0.001

Os dados utilizados para treinamento são baseados na faixa IP´s da rede local, por exemplo: 192.168.1.0/24, neste caso, os endereços validados utilizados serão de 192.168.1.1 a 192.168.1.254. Estes serão convertidos para o formato equivalente em binário e por fim utilizados para treinamento. As portas foram escolhidas com base em uma análise de tráfego que considerou o tipo de trabalho executado na empresa Sendo assim, foi feita a seleção de 525 portas UDP/TCP (http, ssh, ftp, entre outras) que foram também convertidas para o formato necessário antes do treinamento. Foi também efetuado o treinamento da Rede Neural com tipo "Origem/destino", que representa a origem e destino da conexão em se tratando do IP local ou externo, que permite definir de onde foi iniciada a maior parte das conexões estabelecidas no tráfego utilizado para classificação.

Após efetuado o treinamento, a aplicação foi executada utilizando como entrada dados reais coletados e tratados pra serem encaixados no padrão que a Rede Neural foi treinada. Nesta etapa, será utilizado o aprendizado obtido no treinamento para classificar as entradas. A tabela 2 apresenta detalhes da quantidade de entradas fornecidas, o tempo gasto para classificar as mesmas e a porcentagem de sucesso/acerto que a Rede Neural obteve na classificação das mesmas. É importante ressaltar que o resultado dessa classificação é proveniente da saída da RNA Adaline, a qual compõe a estrutura proposta neste artigo; conforme já apresentado na Figura 6.

Tabela 2. Porcentagem de sucesso na classificação da Rede Neural dos dados coletados [Autoria própria, 2017]

Tipo	Quantidade	Tempo	Porcentagem de sucesso na classificação da Rede Neural
Porta	100	00:01:10s	99%
IP	200	00:01:12s	100%
Origem/destino	100	00:00:58s	100%

As porcentagens de sucesso na classificação dos dados que foram obtidos com os testes realizados no sistema, indicam que a arquitetura proposta no presente trabalho é promissora para ser utilizada como uma ferramenta para a monitoração passiva e classificação do tráfego de redes. O principal diferencial do sistema apresentado é que ele não necessita ser previamente alimentado com um grande volume de dados para que possa funcionar corretamente. E, também, ele dispensa a necessidade de uma análise completamente manual ou que dependa de profissionais com conhecimento técnico aprofundado para tal finalidade. Dessa forma, o sistema proposto tem potencial para ser inserido no cenário das ferramentas já existentes para classificação do tráfego de redes.

6. Conclusão

Tendo como foco a classificação e mensuração do tráfego de redes utilizando os protocolos de transporte TCP e UDP, foi apresentado neste trabalho conceitos iniciais sobre redes de computadores e a importância de classificar e mensurar o tráfego que nesta transita. Foi introduzido, também, a capacidade que as Redes Neurais podem

agregar ao objetivo do trabalho. Para tanto, foram citados alguns modelos utilizados e trabalhos referentes ao tema que foram também aplicados com sucesso com intuito de obter padrões fornecidos e identificar anomalias, falhas e outros fatores.

O presente trabalho apresentou uma dentre muitas maneiras de aumentar a capacidade de classificação de tráfego em redes de computadores. Foi utilizada a Rede Neural Adaline para efetuar a identificação de padrões e, posteriormente classificá-los; levando em consideração padrões de protocolos pré-definidos. Entretanto, estes padrões podem ser ajustados para possibilitar o aumento de padrões que são identificados.

Os resultados obtidos neste trabalho sugerem que a arquitetura do sistema proposto tem potencial para ser utilizado como uma nova técnica para classificação do tráfego de redes, pois, considerando a maneira em que os dados foram tratados, o sistema proposto permite que o processo seja feito de forma mais dinâmica. E também, em comparação aos demais sistemas já existentes, o trabalho apresentado apresenta o diferencial de não necessitar ser previamente alimentado com um grande volume de dados para que possa funcionar corretamente. Assim, o uso de Redes Neurais Artificiais do tipo Adaline para a classificação e mensuração de tráfego de redes de computadores, utilizando protocolos de transporte TCP e UDP foi considerado uma maneira produtiva e aplicável ao cenário.

Como sugestão para trabalhos futuros, pode ser considerado a utilização de Machine Learning para identificação e classificação de padrões, de forma automatizada, relacionados ao tráfego de redes de computadores.

7. Referências

Barros, M. T. (2012). Classificação de Fluxos IP como Ferramenta para Engenharia de Tráfego na Internet. Universidade Federal de Campina Grande.

Brownlee, N., Mills, C. and Ruth, G. (1999). Traffic Flow Measurement: Architecture. https://tools.ietf.org/html/rfc2722, [accessed on Jan 18].

Fausett, L. (1993). *Fundamentals of Neural Networks: Architectures, Algorithms And Applications*. Pearson.

Haykin, S. (2006). *Redes Neurais. Princípios e Prática*. 2. ed. Bookman.

Júnior, P. G. L. (2012). Classificação de tráfego baseado em mineração de fluxos de dados. Universidade Federal de Pernanbuco.

Kovács, Z. (2002). *Redes Neurais Artificiais. Fundamentos e Aplicações*. 4. ed. Livraria da Física.

Lima, I. V. (2005). Uma abordagem simplificada de detecção de instrusão baseada em Redes Neurais Artificiais. Universidade Federal de Santa Cataria.

Lima, P. L., Filho, H. S., Lima, R. R., Oliveira, R. and Neto, A. P. (2011). Classificação de QoS em Conteúdo Multimídia para Rede VPN utilizando Rede Neural Multi-layer Perceptron. http://www.lbd.dcc.ufmg.br/colecoes/ein/2011/Artigo_1.pdf, [accessed on Mar 5].

Lima, R. A. G. (2014). Utilização de Sistemas Inteligentes para Classificação de Tráfego Malicioso. Universidade Federal de Viçosa.

Oliveira, T. P. (2014). Predição de Tráfego, Usando Redes Neurais Artificiais, Para Gerenciamento Adaptativo de Largura de Banda em Roteadores. Univerisidade Federal de Uberlândia.

Santos, A. C. F. (2011). Uma Metodologia Para Caracterização do Tráfego de Redes de Computadores: Uma Aplicação em Detecção de Anomalias. Instituto Nacional de Pesquisas Espaciais.

Silva, P. H. D. and Júnior, N. A. (2014). Ferramenta IPERF: geração e medição de Tráfego TCP e UDP. *Notas Técnicas*, v. 4, n. 2, p. 1–13.

Szabó, G., Orincsay, D., Malomsoky, S. and Szabó, I. (2008). On the Validation of Traffic Classification Algorithms. [M. Claypool & S. Uhlig, Eds.]In *International Conference on Passive and Active Network Measurement*.

Tcdump, . (2018). TCDUMP and LibPcap. https://www.tcpdump.org/, [accessed on Mar 2].

Wireshark (2018). WireShark. https://www.wireshark.org/, [accessed on Feb 13].

Zander, S., Willians, N. and Armitage, G. (2016). Internet Archeology: Estimating Individual Application Trends in Incomplete Historic Traffic Traces. [M. Allman & M. Roughan, Eds.]In *Proceeding of the Passive and Active Measurement Conference*.