Técnicas para Análise de Tráfego de Rede

António Mendes[a84675], Maria Cunha[a93264], and Vicente Moreira[a93296]

Universidade do Minho, Campus de Gualtar, Braga

Abstract. Com o aumento exponencial da dimensão das redes, assim como o aumento do número dos seus utilizadores, surgiu a necessidade de analisar, prever e modelar o comportamento das redes, de forma a aumentar a sua eficiência e segurança, ao facilitar a decisão de alocação de recursos, na previsão antecipada de potenciais ataques na rede e descobrir as necessidades da rede, podendo ser a sua expansão planeada cuidadosamente. Este artigo contém uma breve exploração das várias técnicas de análise de tráfego, contendo também a metodologia mais comum ao desenvolvimento desta análise, assim como alguns dos dados, algoritmos e técnicas mais utilizadas em cada fase desta metodologia.

Keywords: Análise de Tráfego · Machine Learning

1 Introdução

Em 2021, 58,9% da população global tem acesso à internet. Esta encontra-se presente em quase todos os dispositivos que usamos no quotidiano: os *smartphones*, computadores, e vários eletrodomésticos. A navegação e utilização que fazemos desta varia de utilizador para utilizador mas, principalmente quando se fala em telemóveis pessoais, há partilha de dados privados ou confidenciais que devem ser reservados.

Adicionalmente, o aumento do número de serviços essenciais que recorrem a meios *online* resulta, também, numa crescente necessidade de salvaguardar a informação transmitida pela rede.

Assim, criou-se um novo ramo de estudo inserido na àrea da informática dedicado à análise de tráfego na rede. Este estuda métodos de inferência que recebem dados do tráfego *online* de vários dispositivos como *input* e devolve informação com detalhes sobre estes dispositivos e a rede como *output*, criando assim um modelo da rede.

Dado esta capacidade de modelar redes, a análise de trafégo mostrou ter um papel crucial na gestão das redes e da sua segurança. Esta tem uma função proativa na monitorização da rede, evitando a ocorrência de congestionamentos e quebras de segurança através de esquemas preventivos.

Para analisar o tráfego na rede, criaram-se dois tipos de esquemas preventivos: previsão de longo prazo, que fornece um bom modelo do funcionamento da rede, permitindo assim avaliar eventuais problemas futuros e planear de forma cuidada as necessidade de expansão da rede. A outra categoria é a previsão de

curto prazo (mili-segundos a minutos), que permite o controlo de congestionamento ativo e auxilia na gestão de alocação de recursos, assim como melhorar a qualidade de serviço.

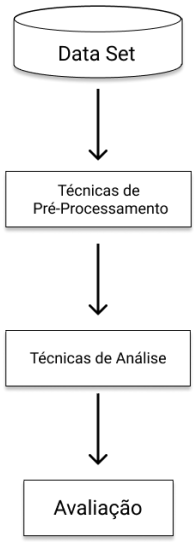
2 Metodologia de Análise de Tráfego

No passado, os administradores eram responsáveis por monitorizar um pequeno número de dispositivos de redes, sendo a capacidade de débito da rede muito reduzida (menos de 100 *Mbps*). Comparando com os dias de hoje, os administradores têm que ser capazes de monitorizar e gerir milhares de dispositivos a alta velocidade (1 *Gpbs* ou mais), sendo necessário recorrer a ferramentas de análise de tráfego de forma a auxiliar na gestão desta.

No entanto, esta análise é complexa dado que pode ser efetuada nos vários níveis de rede, quer ao nível dos pacotes, ao nível do fluxo ou ao nível da rede. Não só, a dimensão dos dados recolhidos e trabalhados gerou a necessidade de criar uma metodologia genérica de análise de tráfego.

Esta é dividida em 4 fases: A recolha dos dados da rede a ser analisada, o pré-processamento destes dados, a análise destes e, por último, a avaliação dos resultados.

Fig. 1. Metodologia Genérica para Análise de Tráfego de Redes



De seguida será apresentada cada uma destas fases, justificando o seu papel e importância no processo, assim como expor algumas das técnicas mais comuns utilizadas.

2.1 Data Sets

Data Sets são definidos como a coleção de dados relacionados de informação, compostos por elementos distintos mas manipulados por computadores como uma unidade só.

Para que haja consistência no formato dos dados recolhidos, foram criados ao longo dos anos vários *standards* de *Data Sets*. Estes contêm vantagens e desvantagens, sendo alguns destes *Data Sets* mais especificos a certas análises. Um dos *standards* mais utilizados frequentemente é o "DARPA *Data Set*", que se especializa na deteção de intrusões.

2.2 Técnicas de Pré-Processamento

Depois dos dados serem devidamente recolhidos e formatados, estes precisam de ser pré-processados de forma a simplificar e facilitar a posterior análise dos mesmos. Este processamento ocorre em duas fases distintas:

Segmentação: Responsável por distribuir os valores contínuos de um atributo em segmentos distintos, reduzindo significativamente o tempo de análise de tráfego.

Seleção: Este passo reduz a quantidade de dados a ser analisada ao selecionar que atributos de valores são relevantes para a análise.

2.3 Técnicas de Análise (*Mining*)

Esta terceira fase, a fase de análise, é a fase mais importante do processo para a avaliação de tráfego de rede dado a grande dimensão e complexidade dos dados presentes a serem analisados. Para este processo, é frequente e necessário recorrer a técnicas de *Machine Learning*.

Machine Learning (ML) é uma solução bem conhecida e explorada, não só para a classificação de tráfego na *internet*, mas também na previsão e geração de conhecimento. Esta estratégia também é bem capaz de analisar uma rede com tráfego encriptado visto que ML é capaz de efetuar uma análise da rede sem ter acesso ao conteúdo dos pacotes. Tráfego encriptado tem se tornado relevante para que se possa evitar intrusões no meio de informação a ser transmitida. Adicionalmente, é capaz de detetar tráfego anormal, sendo este relevante para descobrir e caracterizar anomalias provenientes de fontes maliciosas, ou mesmo de fontes que, sem saber, prejudicam a infrastrutura da rede, afetando a privacidade, quer esta seja de um utilizador privado ou de uma empresa.

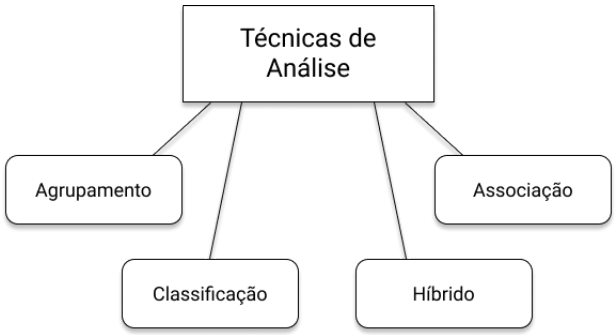
A análise de tráfego com ML começou a ser usado em 2005, no entanto, alguns problemas persistem devido à rápida evolução e escabilidade desta, entre outras mais razões, sendo algumas como:

- Informação disponível limitada

- Escabilidade de classificação de tráfego é complicada
- Falta de soluções adaptativas devido ao dinamismo e evolução da internet
- Soluções necessitam de uma validação correta

Apesar da ajuda que *Machine Learning* fornece para a análise de tráfego na rede, existem várias categorias de como esta análise é efetuada, sendo que cada uma tem os seus objetivos e especializações, assim como as suas vantagens e desvantagens nos resultados obtidos. Os métodos de análise podem ser divididos em 4 categorias: Agrupamento, Classificação, Híbrido e Associação.

Fig. 2. Categorias de Técnicas de Análise de Tráfego



Agrupamento - Esta categoria é focada em agrupar dados semelhantes em grupos, de forma a que cada grupo contenha dados semelhantes acerca de uma rede. Um dos algoritmos mais populares nesta categoria é conhecido por "k-*means*".

Classificação - Classificação é uma forma de análise caracterizada por atribuir classificações aos dados. Esta tem como objetivo classificar todo o tráfego da rede como tráfego normal ou malicioso. Esta estratégia apresenta o desafio de reduzir o número de falsos positivos e falsos negativos. Algoritmos como *Suport Vector Machine* (SVM) são utilizados para efetuar esta classificação, apesar desta ser computacionalmente pesada.

Híbrido - Este método de análise é constituido pela combinação de dois ou mais métodos de análise e conseguem alcançar bons resultados, como por exemplo, a utilização do algoritmo SVM com "*Decision Tree*" (DT) de forma a detetar intrusões de forma mais eficiente.

Associação - Esta categoria avalia cada par de atributos nos dados com um item, obtendo assim coleções de items. Esta coleção é depois usada para descobrir padrões e relações entre os vários atributos dos dados. Um dos algoritmos usados nesta área é conhecido por "*Fuzzy Apriori*" e auxilia na deteção de pacotes de rede normais aos maliciosos.

2.4 Avaliação

No fim, o modelo gerado é avaliado através de várias métricas obtidas, dependendo da análise escolhida. Várias métricas como: o número de Falsos-Negativos, ou seja, o número de pacotes maliciosos incorretamente classifcados como normais, o número de Falsos-Positivos, a taxa de deteção, a taxa de precisão e a percentagem de previsões bem-sucedidas.

3 Conclusão

Para concluir, salienta-se a importância de um estudo contínuo na área de análise de tráfego de redes, visto que esta está constantemente a evoluir e a expandir, criando novas necessidades como novos *Data Sets*, técnicas de análise e avaliação de dados. Trata-se de um tópico muito relevante e extenso que merece um estudo aprofundado de cadas as etapas envolventes no processo.

Assim, recomenda-se veemente a leitura e pesquisa cuidada sobre esta área, realçando o valor desta.

References

1. Joshi, M.,Hadi, T.: A Review of Network Traffic Analysis and Prediction Techniques.
2. Pacheco F.,Exposito E.,Gineste M.,Bausoin C.,Aguilar J.: Towards the Deployment of Machine Learning Solutions in Network Traffic Classification: A Systematic Survey. IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 21, NO. 2, 2019
3. Conti M.,Li Q.,Maragno A.,Spolaor R.: The Dark Side(-Channel) of Mobile Devices: A Survey on Network Traffic Analysis. IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL.20, NO. 4, 2018