Caracterização e Classificação do Tráfego da Darknet com Modelos Baseados em Árvores de Decisão

Gustavo Lopes Rodrigues[1], **Lucas Santiago de Oliveira**[1], **Rafael Amauri Diniz Augusto**[1], **Rafael Pinto Mesquita**[1]

[1]Instituto de Ciências Exatas e Informática –
Pontifícia Universidade Católica de Minas Gerais(PUC-MG)

1. Motivação

O artigo apresenta a categorização classificatória do tráfego da *Darknet* como relevante para atividades como a detecção de intrusões e disseminação de malwares, o gerenciamento de *Quality of Service* no contexto de escalabilidade e a prevenção de ataques de *DDoS*.

2. Objetivos

O objetivo do trabalho é a detecção e categorização do tráfego da *Darknet* utilizando uma base de dados representativa: a *CIC-DarkNet2020* e os modelos *Decision Tree* e *Random Forest*. Os resultados também serão comparados com os resultados obtidos por uma classificação feita somente com a origem do tráfego e com outra classificação que verifica a aplicação dos dados provenientes da *Darknet*.

3. Modelo

Para a interpretação do conjunto de características agregados, os modelos de *Decision Trees (DT)* e *Random Forest (RF)* foram escolhidos, devido a sua simplicidade e facilidade na interpretação, tendo apenas como desvantagem o fato que eles não podem ser treinados novamente com novos dados, precisando retreinar o modelo com os dados antigos somados aos novos.

4. Resultados de simulação

A utilização de um modelo mais simples, demonstrou ser mais adequada para a classificação dos dados, quando comparado ao modelo anterior. A acurácia do modelo de Gurdip Kaur usando deep learning foi de 86% enquanto que o método proposto nesse artigo foi de 99.03%.

5. Conclusões

Neste trabalho foi abordado os problemas da detecção e caracterização do tráfego proveniente da *Darknet* através da utilização de modelos de aprendizagem baseados em árvores de decisão, sendo eles a *DT* e a *RF*, que se mostraram capazes de classificar novos registros de tráfego com uma acurácia superior a 98% para cada uma das tarefas de classificação. A proposta citada pelos criadores no artigo para trabalhos futuros é reutilizar esses mesmos modelos de predição para modelos de aprendizado online.