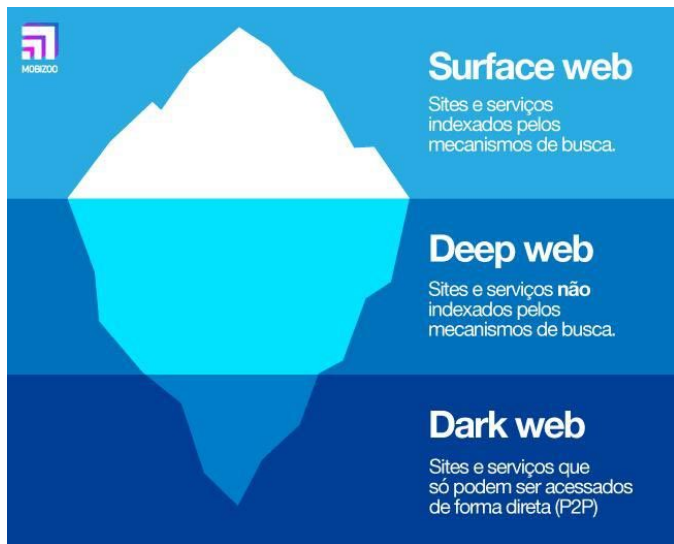


# Caracterização e Classificação do Tráfego da Darknet com Modelos Baseados em Árvores de Decisão

Gustavo Lopes Rodrigues  
Lucas Santiago de Oliveira  
Rafael Amauri Diniz Augusto  
Rafael Pinto Mesquita

# Introdução à Dark Web e sua importância



Acreditamos que todos devem poder explorar a internet com privacidade. Nós somos o “Tor Project”, uma organização sem fins lucrativos dos EUA. Promovemos os direitos humanos e defendemos sua privacidade online por meio de software livre e redes abertas.

- Tor Project (traduzido)

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

Autores: Mateus Coutinho Marim, Paulo Vitor Barbosa Ramos, Roberto Massi de Oliveira, Alex B. Vieira e Edelberto Franco Silva



# Análise da Dark Web pode ajudar a Surface Web

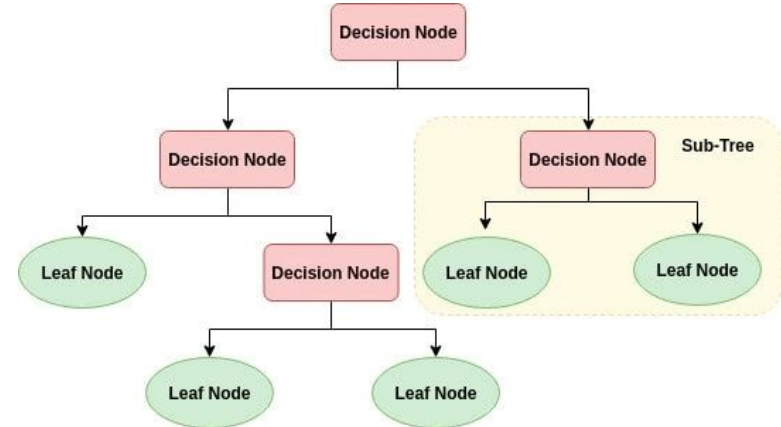
- Avanço do estado-da-arte de inteligência artificial voltada para redes e *P2P*.
- Categorização em tempo real de tráfego na web.
- Otimizações para navegadores em diferentes serviços como VOIP, transferência de arquivos, *streaming* de vídeo, etc.
- Detecção de invasões e *malwares* por análise de conexões.
- Disponibilidade de serviço no contexto de escalabilidade e a prevenção de ataques de *DDoS*.

Mas como é possível classificar o tráfego da web?



# Árvores de Decisão e Random Forest

- Lista de perguntas e respostas rotuladas
- Hierarquicamente arranjadas
- Estruturas bem-definidas.
- Mais resistente a *overfitting* (Random Forest!)



## Dados para treinamento e fitting dos modelos

- CIC-Darknet2020, disponibilizado pelo Canadian Institute for Cybersecurity at the University of New Brunswick.
- *Dataset* com milhares de entradas contendo rótulos para serviços tipicamente utilizados na darknet.
- Portas, IPs tipicamente utilizados, protocolos de conexão, protocolos de transporte, etc.

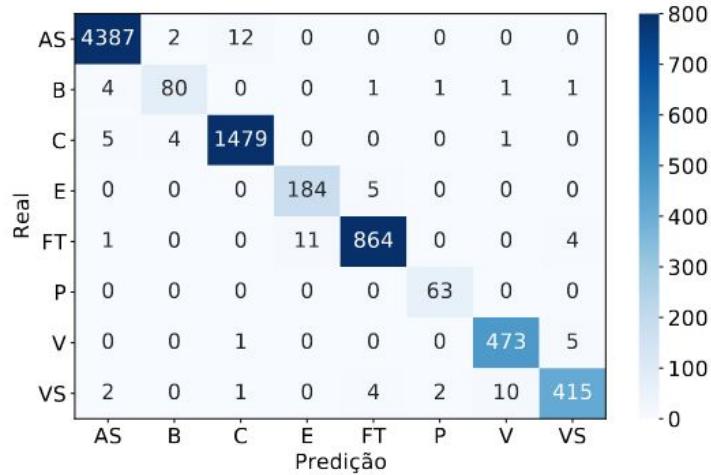
# Dados para treinamento e fitting dos modelos

Traffic Category	Applications used
Audio-Stream	Vimeo and Youtube
Browsing	Firefox and Chrome
Chat	ICQ, AIM, Skype, Facebook and Hangouts
Email	SMTPS, POP3S and IMAPS
P2P	uTorrent and Transmission (BitTorrent)
Transfer	Skype, FTP over SSH (SFTP) and FTP over SSL (FTPS) using Filezilla and an external service
Video-Stream	Vimeo and Youtube
VOIP	Facebook, Skype and Hangouts voice calls

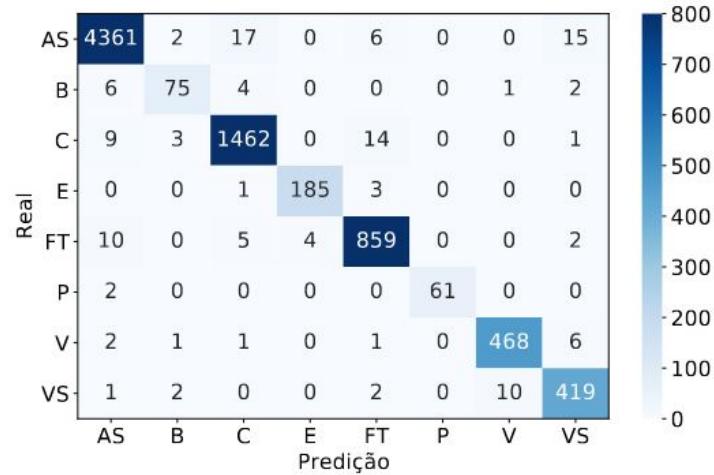


# Resultados

- 99% e 98% de acerto respectivamente



(a) Decision tree



(b) Random forest

# Um sucesso? Não! Mas por que?

- Foram agrupadas as portas padrões com o serviço que se refere a essa porta. O nó-raiz das árvores foram as portas.
- Se um usuário mudar uma porta padrão, o tráfego será classificado erroneamente!
- Não houve separação entre dados de treino, teste e validação. Os modelos decoram as respostas. *Overfitting* se mostrou um problema sério para esse modelo!
- Ponto especialmente fraco para árvores de decisão e random forests, que têm uma tendência alta a se “viciarem” e sofrerem por overfitting.
- Os modelos escolhidos pelo autor exigem uma atenção especial do pesquisador.

## Um sucesso? Não! Mas por que?

- Uso incorreto da expressão “internet benigna” para referir-se à “surface web”. Esse é um termo inexistente na literatura e não reflete em nada real.
- Persiste no artigo uma visão infantilizada da Dark Web. Os autores insistem em citar malware e hacking como se fossem uma parte integral da Dark Web.

# Aplicações do artigo e considerações finais

- É um problema não-determinístico, logo, modelagens probabilísticas como CNNs e estratégias de Aprendizado Profundo são uma alternativa melhor ao Aprendizado de Máquina.
- Maior resistência a ruídos e *outliers*.
- Estado da arte: *“DIDarknet: A Contemporary Approach to Detect and Characterize the Darknet Traffic using Deep Image Learning”*.

# Referências

- Arash Habibi Lashkari, Gurdip Kaur, and Abir Rahali. 2020. DIDarknet: A Contemporary Approach to Detect and Characterize the Darknet Traffic using Deep Image Learning. In *2020 the 10th International Conference on Communication and Network Security (ICCNS 2020)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3442520.3442521>
- Tor Project - Acessado em maio de 2022
- Repositório do artigo: Github - Acessado em maio de 2022
- Trocando portas do FTP: IBM - Acessado em maio de 2022

Obrigado!!