Análise de Trafego de Rede com Machine Learning para Identificação de Ameaças IoT

Matheus Magalhães de Paula Paiva

www.inatel.br/

Inatel

Resumo

- Com o número crescente de dispositivos IoT, aumenta a preocupação com a segurança desses dispositivos.
- Proposta do trabalho: Detectar ameaças baseada na análise de tráfego de rede, utilizando aprendizado de máquina.

1.0 Introdução

1.1 Soluções para segurança

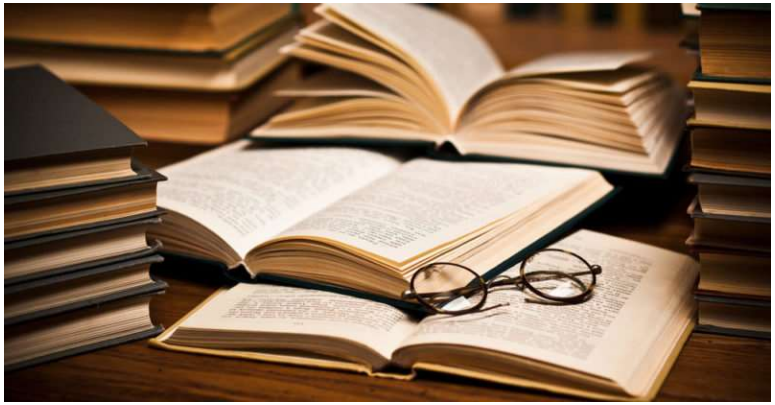
Soluções de segurança para os dispositivos:

- Podem ser tanto a nível de hardware quanto software, como Firewalls.
- Modelos de aprendizado de máquina analisando o tráfego de rede: **abordagem inovadora**.

1.2 Abordagens literárias

Em 2020, Stoain comparou alguns modelos de aprendizado de máquinas, conseguindo obter resultados com acurácia de 99,5%.

No mesmo ano, os autores de Shafiq et al propuseram uma nova métrica para seleção de características denominada CorrAUC, obtendo uma precisão de 96% na detecção de tráfego malicioso.

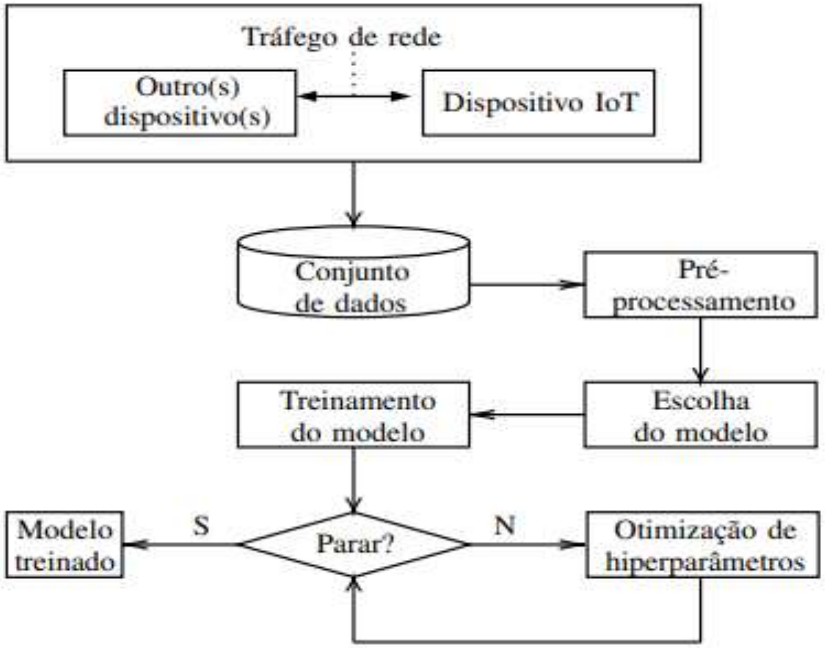


1.3 Objetivo

Este trabalho tem como objetivo escolher, treinar, otimizar e avaliar modelos de aprendizado de máquina aplicados à identificação de ameaças a dispositivos IoT.

2.0 Metodologia

2.1 Etapas do modelo



2.2 Conjunto de dados

Utilizado o dataset IoT-23, disponível no laboratório *Statrosphere*.

Possui 23 capturas no total, sendo:

- 20 capturas de malware.
- 3 capturas de tráfego benígno.

2.3 Pré-processamento

- Concatenação dos 23 conjuntos de dados em um único arquivo;
- Padronização dos rótulos das classes;
- Remoção de amostras que não correspondem a cenários mais frequentes no *dataset*;
- Preenchimento de valores ausentes;
- Remoção de amostras duplicadas;
- Separação dos dados em subconjunto de treinamento e validação.

2.4 Identificação dos modelos de aprendizado

- Após o processamento de dados identificou-se os modelos a serem utilizados na tarefa de classificação:
  - Árvores de decisão;
  - Adaboost;
  - Random Forest;
  - Perceptron;
  - Nearest Centroid;

2.5 Treinamento dos modelos e otimização de hiperparâmetros

- O sucesso ou fracasso do processo de treinamento se deve a configuração dos hiperparâmetros.
- Deve-se avaliar diversas combinações para que o modelo tenha seu desempenho otimizado (máxima acurácia)

- Cada modelo foi submetido a uma rodada de otimização realizada pela biblioteca optuna.

3.0 Experimentos e discussão

3.1 Acurácia dos classificadores

Posição	Classificador	Acurácia (%)
1	Decision Tree	99,86
	Random Forest	99,86
2	Adaboost	93,46
3	Perceptron	73,40
4	Nearest Centroid	59,51

3.0 Experimentos e discussão

3.1 Treinamento dos classificadores

Posição	Classificador	Treinamentos
1	Decision Tree	100
	Nearest Centroid	100
2	Perceptron	83
3	Random Forest	18
4	Adaboost	10

4.0 Conclusão

4.1 Conclusão

- Modelos simples são treinados mias rapidamente;
- Árvores de decisão e ensembles produziram os melhores resultados;
- Metodologia eficaz, pois produziu modelo com capacidade de detecção próxima a 99,9%.

4.0 Conclusão

4.2 Sugestão de trabalhos futuros

- Empregos de técnicas de pré-processamento mais sofisticadas;
- Utilização de outros modelos de aprendizado de máquina;
- Utilização de outros conjuntos de dados;
- Utilização de método de aprendizado não supervisionado capaz de armazenar novos dados e fazer um treinamento para um possível novo ataque que venha a ser criado no futuro

Perguntas

Referências

Sebastian Garcia, Agustin Parmisano, & Maria Jose Erquiaga. (2020). IoT-23: A labeled dataset with malicious and benign IoT network traffic (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4743746

N.A. Stoian. Machine Learning for Anomaly Detection in IoT networks: Malware analysis on the IoT-23 Data set. 2020.

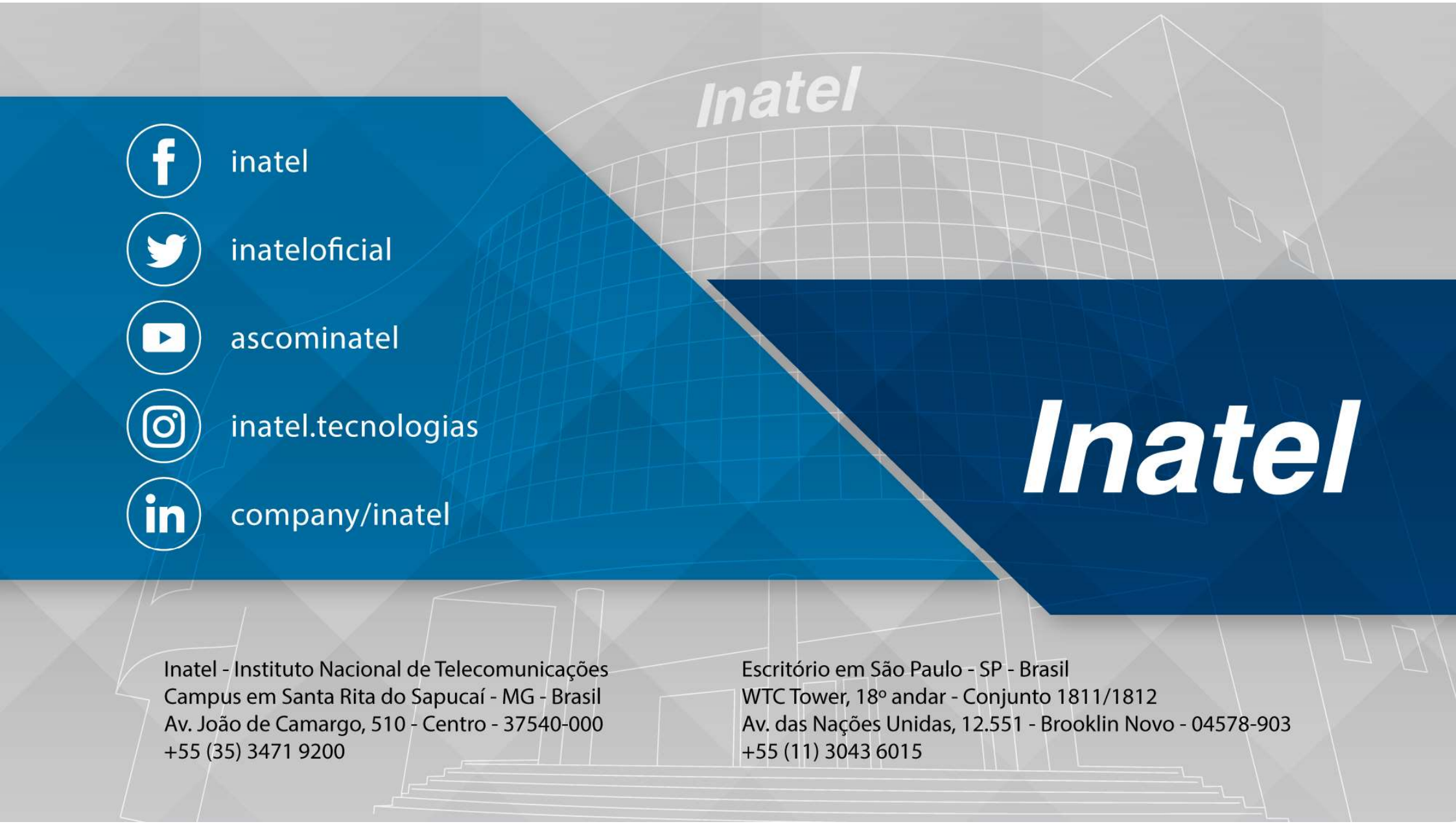
N. Moustafa. The Bot-IoT dataset. 2019. DOI: 10 . 21227/r7v2-x988.

M. Shafiq, Z. Tian, A.K. Bashir, X. Du e M. Guizani. “CorrAUC: A Malicious Bot-IoT Traffic Detection Method in IoT Network Using Machine-Learning Techniques”. Em: IEEE Internet of Things Journal 8.5 (2021), pp. 3242–3254. ISSN: 23274662. DOI: 10 . 1109/JIOT.2020.3002255.

https://scikit-learn.org/stable/user_guide.html. Acesso em 06/12/2022

https://pandas.pydata.org/docs/user_guide/index.html#user-guide. Acesso em 06/12/2022

https://docs.scipy.org/doc/scipy/ Acesso em 06/12/2022



inatel

inateloficial

ascominatel

inatel.tecnologias

company/inatel

Inatel

Inatel - Instituto Nacional de Telecomunicações
Campus em Santa Rita do Sapucaí - MG - Brasil
Av. João de Camargo, 510 - Centro - 37540-000
+55 (35) 3471 9200

Escritório em São Paulo - SP - Brasil
WTC Tower, 18º andar - Conjunto 1811/1812
Av. das Nações Unidas, 12.551 - Brooklin Novo - 04578-903
+55 (11) 3043 6015