Detecção precoce e combate ao registro fraudulento de domínios

Vinicios Barretos
vinicios@acmesecurity.org

LinkedIn

UNESP
ACME! Cybersecurity Research

COOPERAÇÃO EM P&D

Projeto: Detecção de Registro Fraudulento de Nomes de Domínio

IMPACTO NA COMUNIDADE

REPUTAÇÃO DO TLD

O excesso de registros fraudulentos pode levar o usuário a associar o TLD a uma imagem negativa.

PERDAS MONETÁRIAS

Prejuízo causado por negativa de serviço, falsificação de identidade e golpes financeiros.

VAZAMENTO DE DADOS

Phishing focado em obter informações pessoais.

MEDIDAS ESTABELECIDAS

Pedidos recebidos via email

Feeds gratuitos: *OpenPhish*

Fraudes no pagamento

Filtros com REGEX

RUBENS KUHL

GTS 30

TIPOS DE AMEAÇAS

- **Botnets**

  Controlar as máquinas zumbis através de um Command and Control: fraudes, DDoS, malwares.

- **Phishing**

  Obter informações sensíveis como senhas e cartões de crédito se passando por domínios verídicos.

- **Spam**

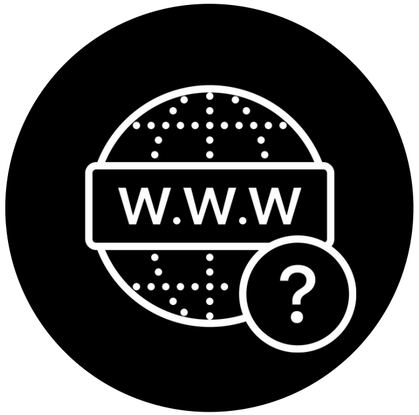
  Domínios utilizados para envio de mensagens indesejadas.

- **Fast-Flux Domain**

  Esconder malwares através de máquinas comprometidas atuando como proxy. IPs alterados frequentemente.

Como identificar fraude?

Análise Textual do nome de domínio



Análise de Coleta Ativa

Análise de Coleta Passiva

Análise Textual

IDENTIFICAÇÃO PREDITIVA

Avaliação do risco de fraude antes mesmo do registro ser concluído.

ESTATÍSTICA + LINGUÍSTICA

A junção de diferentes técnicas permite uma classificação mais precisa.

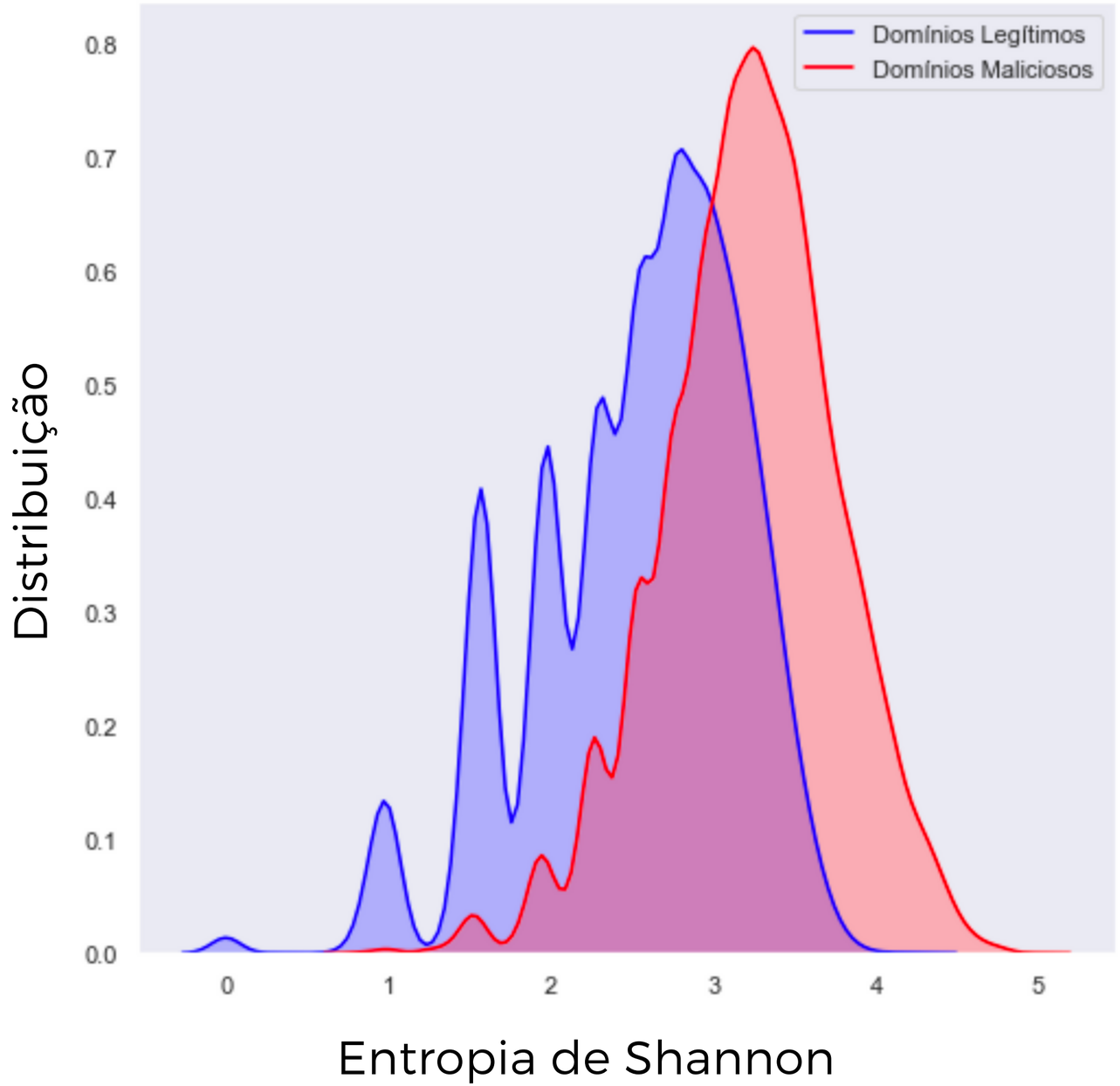
FOCO NO .BR

Desafio: Algoritmos devem ser treinados com datasets em português e em inglês

Entropia de Shannon

CARACTERÍSTICAS

Extração de mais de
10 características
textuais



kuyuhsdfnotr.tld

DGA

amerlcanas.com.br
americanas.com.br

Phishing

promo-ifood.com.br
promo.ifood.com.br

Phishing

Análise Textual

Blacklists

MALWARE DOMAINS
OPENPHISH
...

Whitelists

ALEXA TOP 1 MILLION
CISCO UMBRELLA
MAJESTIC MILLION

Análise de Coleta Ativa

IDENTIFICAÇÃO IMEDIATA

Após a publicação do domínio já é possível realizar sua classificação.

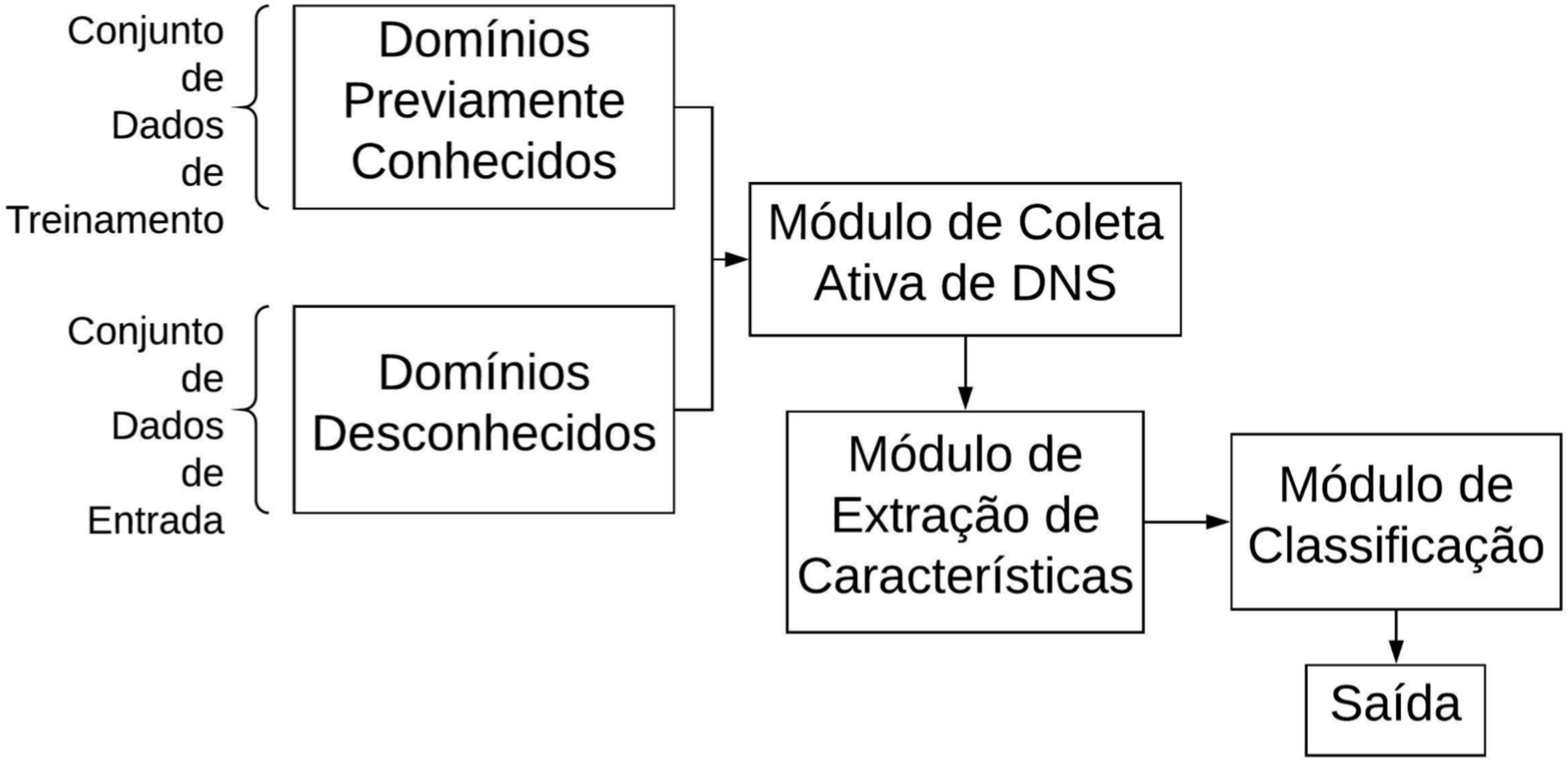
ANÁLISE DOS RRs

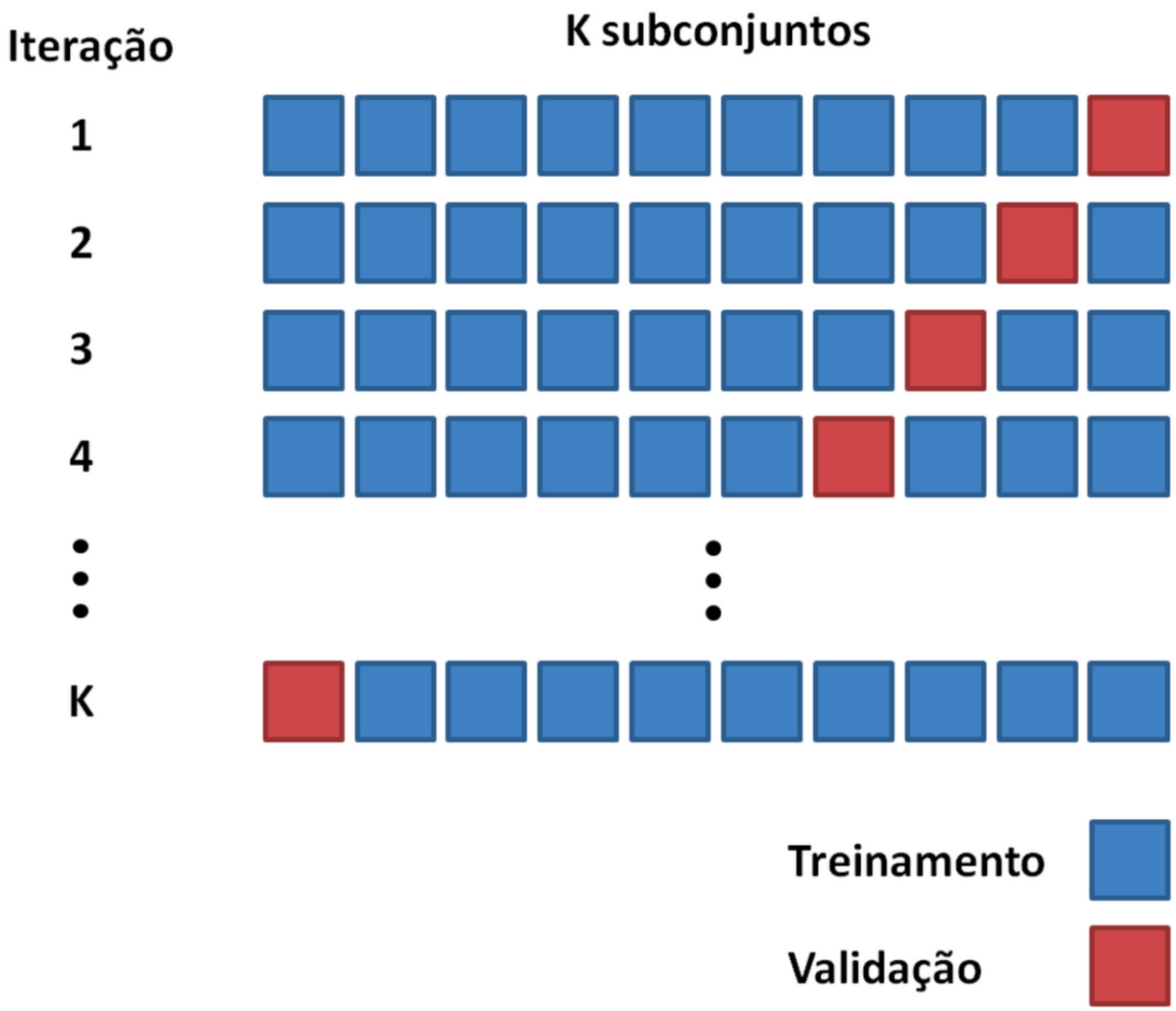
Utilização das respostas obtidas com queries DNS.

MACHINE LEARNING

Utilização de algoritmos supervisionados com base em dados rotulados.

Arquitetura





K-Fold Cross Validation

CARACTERÍSTICAS

TTL
Número de RRs MX
Campo Expire
+9 Features

Análise de Coleta Passiva

MAIOR ABRANGÊNCIA

Através desse módulo é possível detectar ameaças complexas mais rapidamente.

DISPONIBILIDADE

Monitorar e identificar tráfego dos domínios requer alto nível na hierarquia do DNS.

REDES NEURAIS

Detectar similaridade e variações no comportamento das consultas ao domínio ao longo do tempo.

+28 CARACTERÍSTICAS

PRÉ-PROCESSAMENTO

ENRIQUECIMENTO DOS DADOS

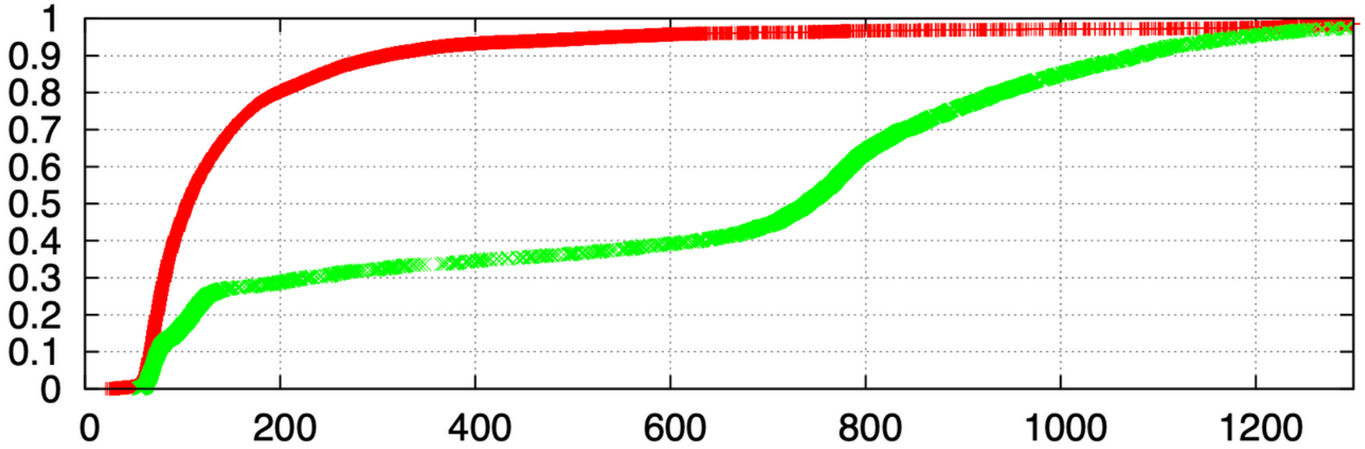
Informações detalhadas providas pelo uso de bases de GeoIP, identificando o ASN e existência em Blacklists.

FEATURES

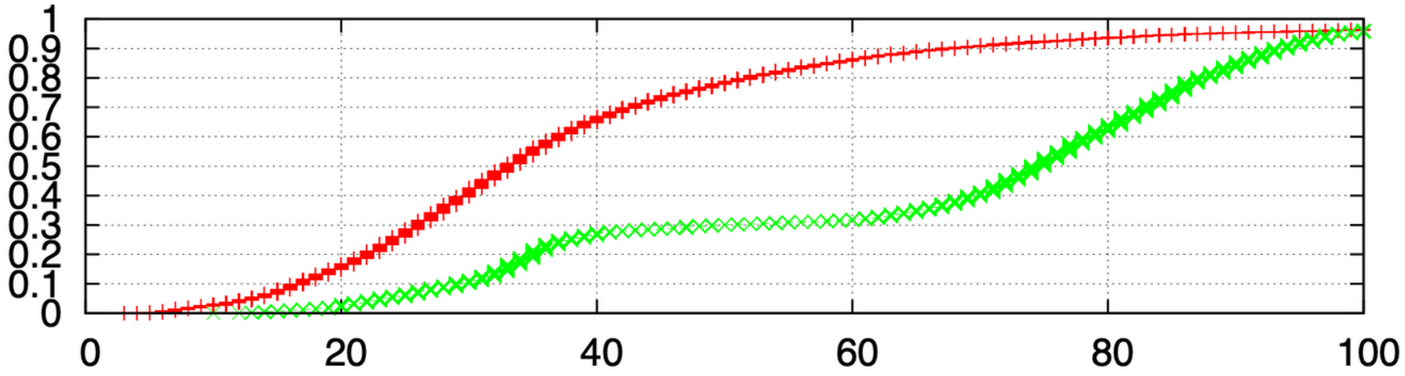
Variação: IPs, Localização e TTL
Atividade do domínio
Padrões de acesso

Análise de Coleta Passiva

Modelar perfil das requisições



Diversidade de AS



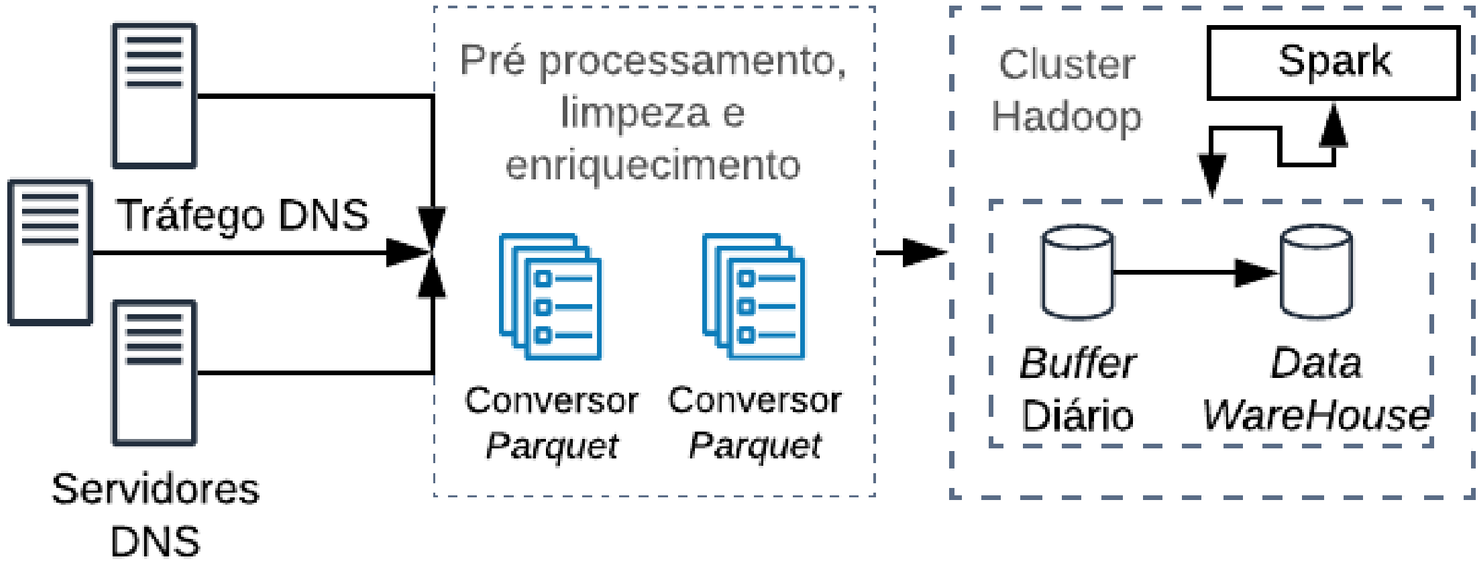
Diversidade de Country Code

Desafio

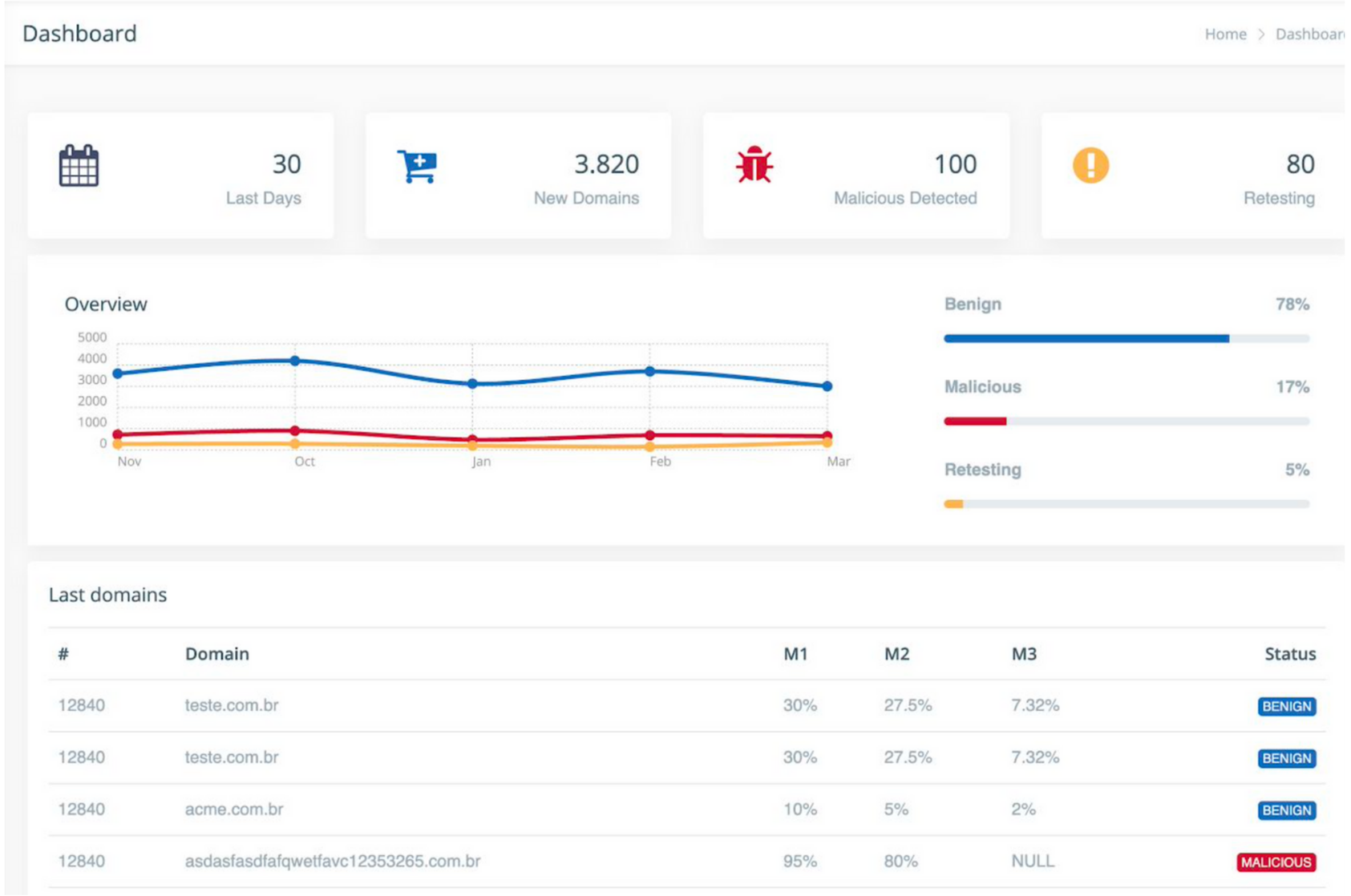
ANÁLISE DE TRÁFEGO DOS

4.105.583

DOMÍNIOS REGISTRADOS NO .BR



Framework



Integração dos módulos

Consultas API
Resultados unificados
Exportar relatórios
Análise de tráfego
Métricas aprimoradas
Re-treinamento
Feed novos domínios

Resultados

PRECISÃO

Módulo 1
77%

Módulo 2
89%

Módulo 3
90%

RÁPIDA DETECÇÃO

Tempo entre registro e detecção da atividade maliciosa cai de forma expressiva.

DIFICULTA ATAQUES

Junção dos 3 tipos de análise enfraquece a ameaça.

TeamDNS

LEOPOLDO FERREIRA

Data Scientist

Mestrado

MARCOS SILVEIRA

Data Scientist

Mestrado

VINICIOS BARRETOS

DevOPS / Data Engineering

Graduação

Agradecimento: Alvaro Pompeu e Rafael Luizete

Obrigado!

Perguntas?

LinkedIn

Vinicios Barretos

vinicios@acmesecurity.org

PGP Key ID: 9B9FE184