DGA Domain Detection and Classification with Passive DNS and Deep Learning

João Rafael Gregório
Kim Morgan de Oliveira Ito Porto
Adriano Mauro Cansian

OARC 44 - 2024

unesp

ACME!
Cybersecurity Research

Domain Generation Algorithms (DGA)

-Algorithms used to generate large amounts of domain names for malicious use

-Examples:

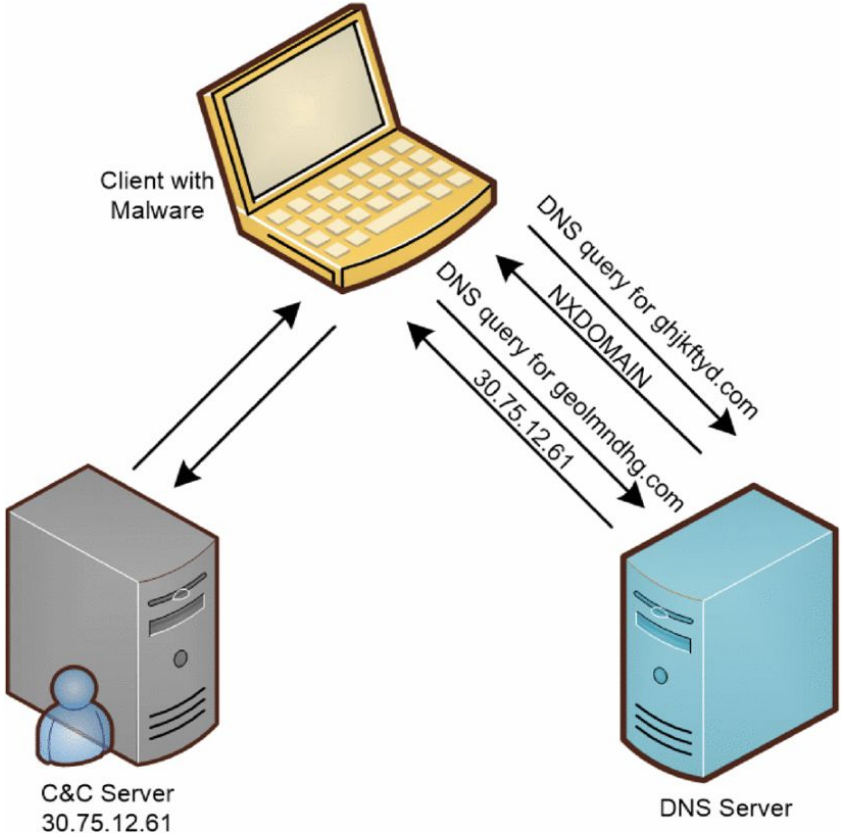
    -Botnets and Advanced Persistent Threats

    -Communication maintenance and obfuscation

-Used by threat actors to avoid detection and prosecution

Ex: growthsupple.net, 1pb98u4egqbcwzes185mpfyvc.com, u035zy.com

Who are Algorithmically Generated Domains?



Source: H. Shahzad; A. R. Sattar and J. Skandaraniyam. DGA Domain Detection using Deep Learning.(2021).

Why Deep Learning?

-No need for Feature Engineering

-Performance for high data volume

-High precision for pattern recognition

-and more…

Ex: growthsupple.net, 1pb98u4egqbcwzes185mpfyvc.com, u035zy.com,

Ahmed, S.F., Alam, M.S.B., Hassan, M. et al. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif Intell Rev 56, 13521–13617 (2023). https://doi.org/10.1007/s10462-023-10466-8

Two Deep Learning Models

- **Detector**: Binary Classification

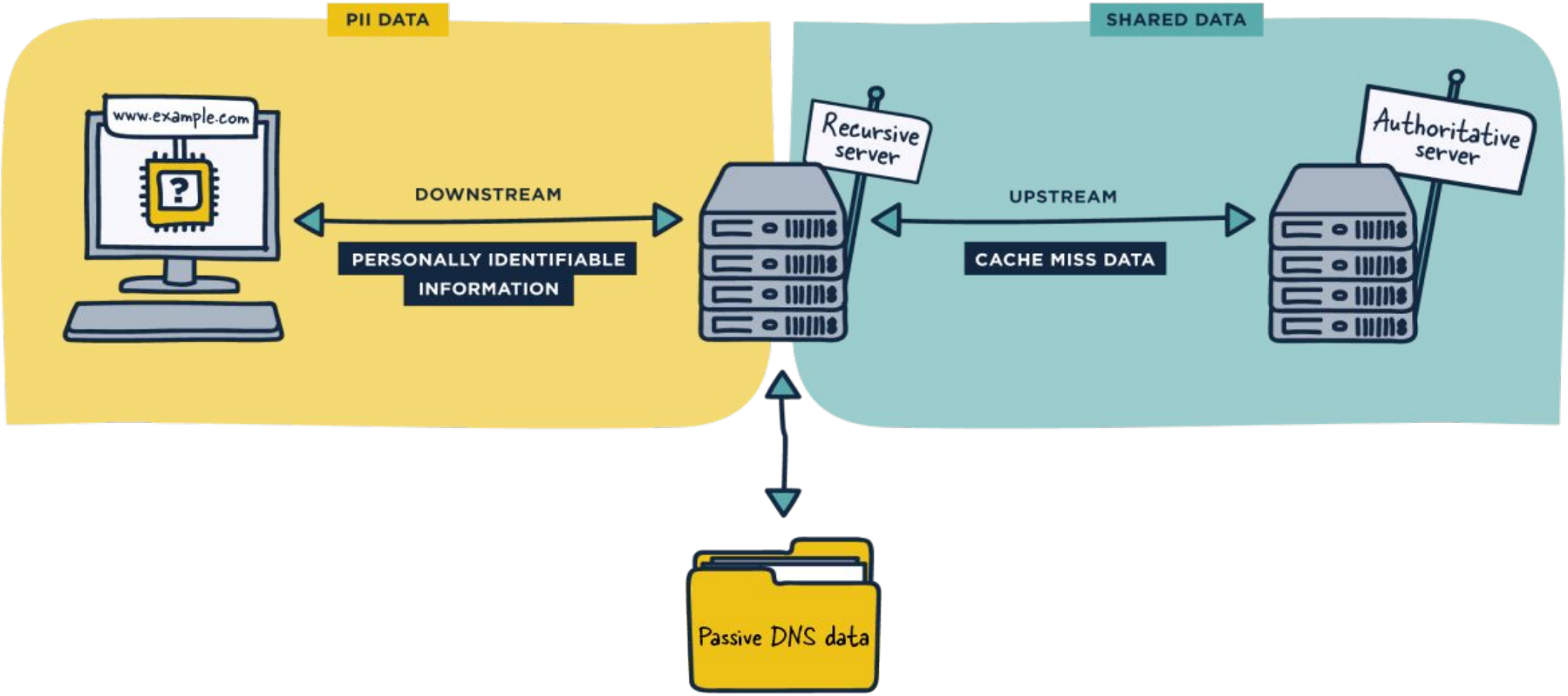
  Classify domains between Legitimate and DGA

- **Classifier**: Multiclass Classifier

  Classify a detected domain into a threat class

  Based on CNN

Passive DNS



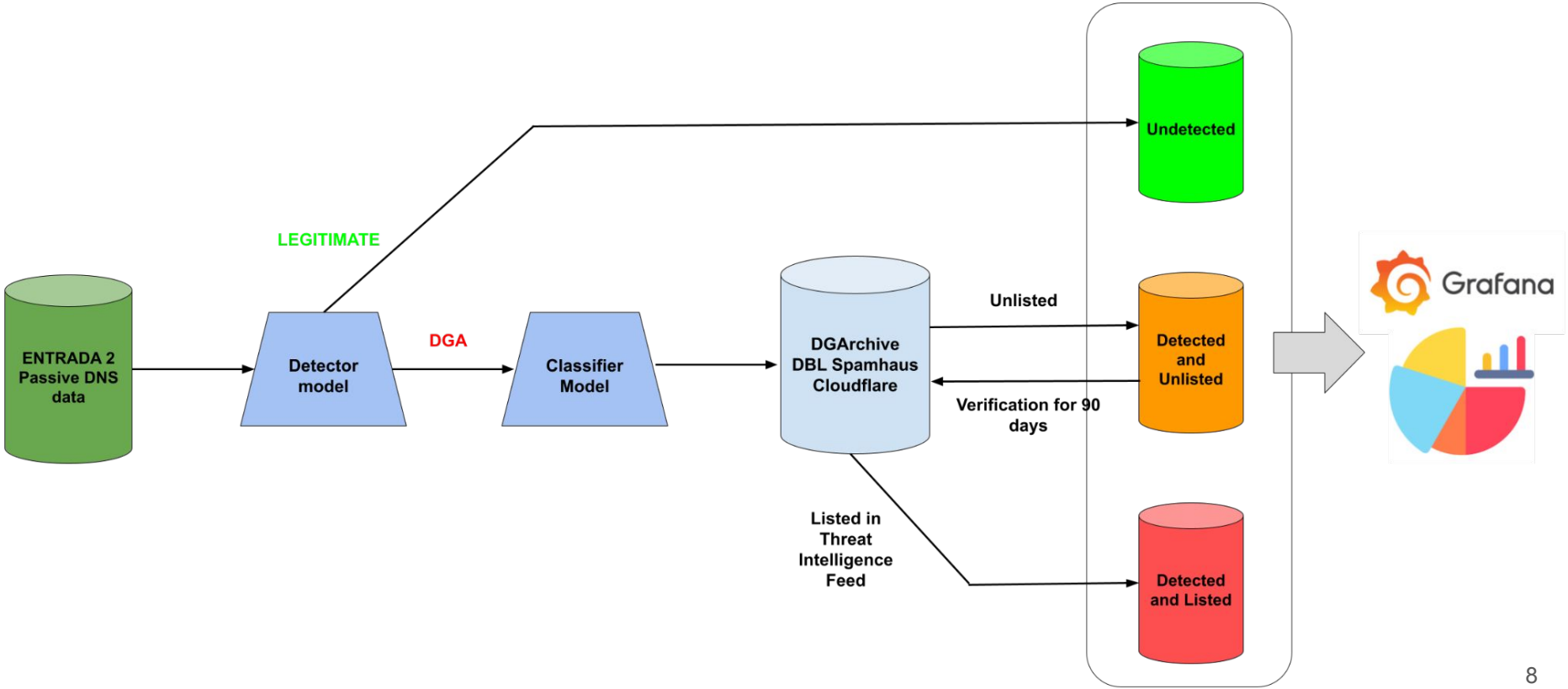
Source: https://www.spamhaus.com/resource-center/what-is-passive-dns-a-beginners-guide/

Passive DNS

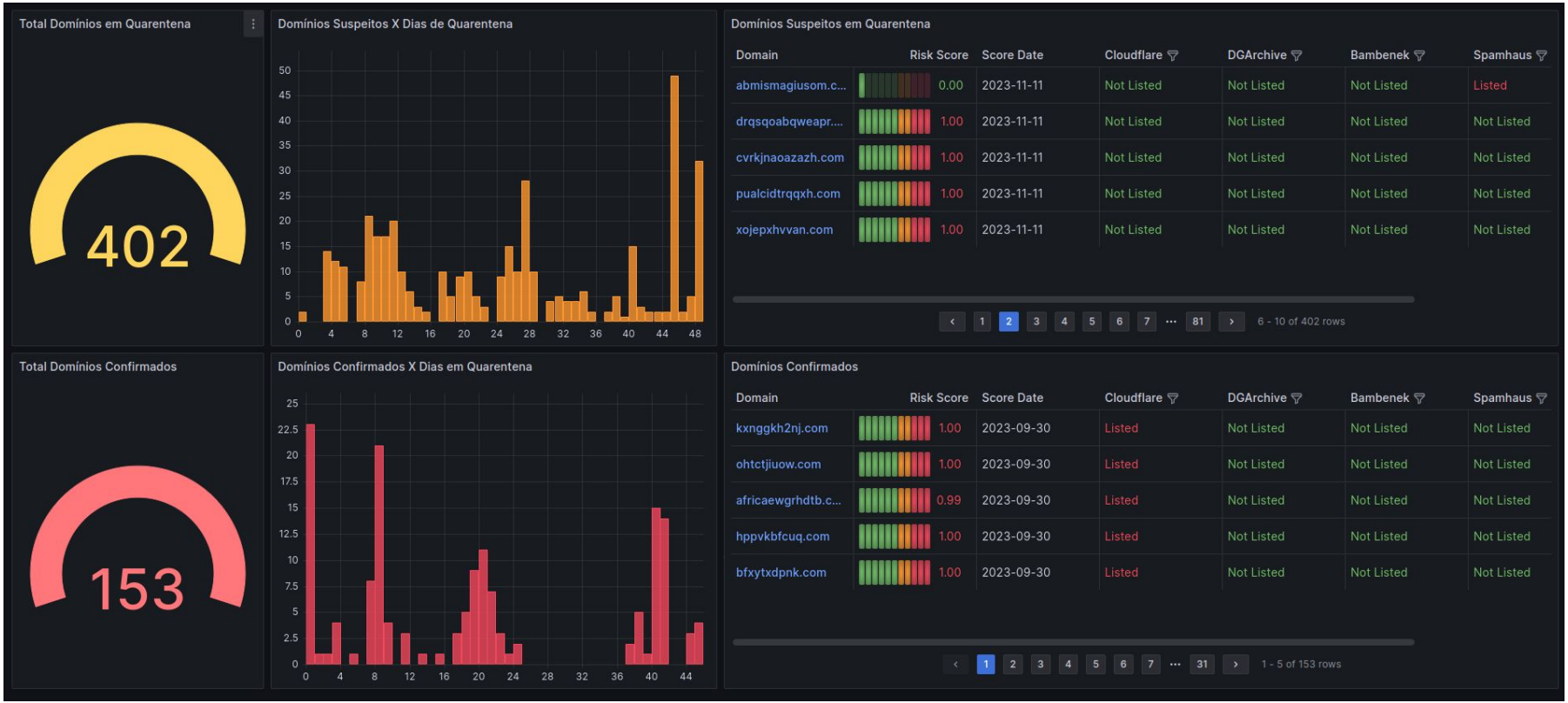
SIDN ENTRADA 2

- 02 Sites (Assis and SJRP) collecting DNS queries across São Paulo State University
- Currently more than 3.5 billion entries and growing

Process Model



pDNS Monitor



Status

- Deep Learning Models (Detector and Classifier) - OK
- Passive DNS Integration - Advanced…
- Monitoring Dashboard - Advanced…

Publications

- 26th ICEIS 2024

“Deep Convolutional Neural Networks and Character-Level Embeddings for DGA Detection”

https://doi.org/10.5220/0012605700003690

- Journal MDPI - Applied Sciences

”Class Incremental Deep Learning: A Computational Scheme to Avoid Catastrophic Forgetting in DGA Multiclass Classification”

https://doi.org/10.3390/app14167244

Objectives achieved

-Detection of DGA domains in real network traffic

-Detection of DGA domains before their listing in threat intelligence feeds

-Identification of previously undetected C2 Servers

-Identification of infected clients, allowing for proactive measures to be taken inside the network

References

- SIDN Labs ENTRADA 2 - https://www.sidnlabs.nl/en/news-and-blogs/entrada-2-0-is-here
- SILVEIRA, M. R. et al. Detection of newly registered malicious domains through passive dns. In: 2021 IEEE International Conference on Big Data (Big Data). [S.l.: s.n.], 2021. p. 3360–3369
- SUN, X.; LIU, Z. Domain generation algorithms detection with feature extraction and domain center construction. PLOS ONE, Public Library of Science, v. 18, p. 1–25, 01 2023 https://doi.org/10.1371/journal.pone.0279866.
- Computer Incident Response Center Luxembourg (CIRCL) Passive DNS 2.0. https://www.circl.lu/services/passive-dns
- Research and Education Networks Information Sharing and Analysis Center (REN-ISAC). https://www.ren-isac.net/member-resources/pDNS.html

Thank you!

rafael.gregorio@unesp.br
kim@acmesecurity.org
adriano.cansian@unesp.br