ICANN 80 POLICY FORUM

SSAC Panel on Artificial Intelligence & Machine Learning in DNS Security:

Advanced DNS Abuse Threat Detection & Mitigation

Agenda

- **Introduction**
- **Current Landscape of using Machine Learning algorithms in detecting DNS Abuse Research**
- **Areas for Further Research and Focus**
- **The Role of SSAC in Supporting Research Efforts**
- **Closing Remarks**

Moderator and Panelists

Jeff Bedser	Graeme Bunton	Carlos Gañan	Chris Lewis-Evans	Greg Aaron
				
CEO CleanDNS	Executive Director NetBeacon Institute	Principal SSR Specialist ICANN	Director of Governmental Engagement and Internet Abuse Mitigation Clean DNS	President Illumintel Inc.

Introduction to Artificial Intelligence (AI)

Broad Concept

- ➜ AI refers to the broader concept of machines being able to carry out tasks in a way that we would consider “smart” or exhibiting human-like intelligence.

Scope

- ➜ Encompasses a range of technologies that simulate human intelligence processes, including reasoning, learning, problem-solving, perception, and language understanding.

Types of AI

- ➜ Can be classified into narrow AI (specialized in a single task) and general AI (which can perform any intellectual task that a human can do)

AI Deployment Techniques

Rule-Based Systems

➜ Early AI systems used pre-defined rules for decision-making

Search and Optimization

➜ Algorithms to solve problems by searching through possible solutions

Expert Systems

➜ AI systems that use databases of expert knowledge to offer advice or make decisions

Natural Language Processing (NLP)

➜ AI techniques to understand and generate human language

Introduction to Machine Learning (ML)

Subset of AI

➜ ML is a subset of AI that focuses on the development of systems that can learn and make decisions from data without being explicitly programmed.

Functionality

➜ Uses algorithms to parse data, learn from it, and make informed decisions based on what it has learned.

Types of ML

➜ Includes supervised learning (learning from labeled data), unsupervised learning (finding patterns in unlabeled data), and reinforcement learning (learning based on rewards from actions)

ML Techniques

Statistical Models

➜ Uses statistical techniques to make predictions or decisions based on data.

Neural Networks

➜ A type of ML model inspired by the human brain, used particularly in deep learning.

Decision Trees, Random Forests

➜ Algorithms used for classification and regression tasks.

Support Vector Machines

➜ Used for classification and regression tasks in high-dimensional spaces.

Current Landscape of using Machine Learning Algorithms in Detecting DNS Abuse Research

AI/ML and Proactive Abuse Approaches

NETBEACON INSTITUTE

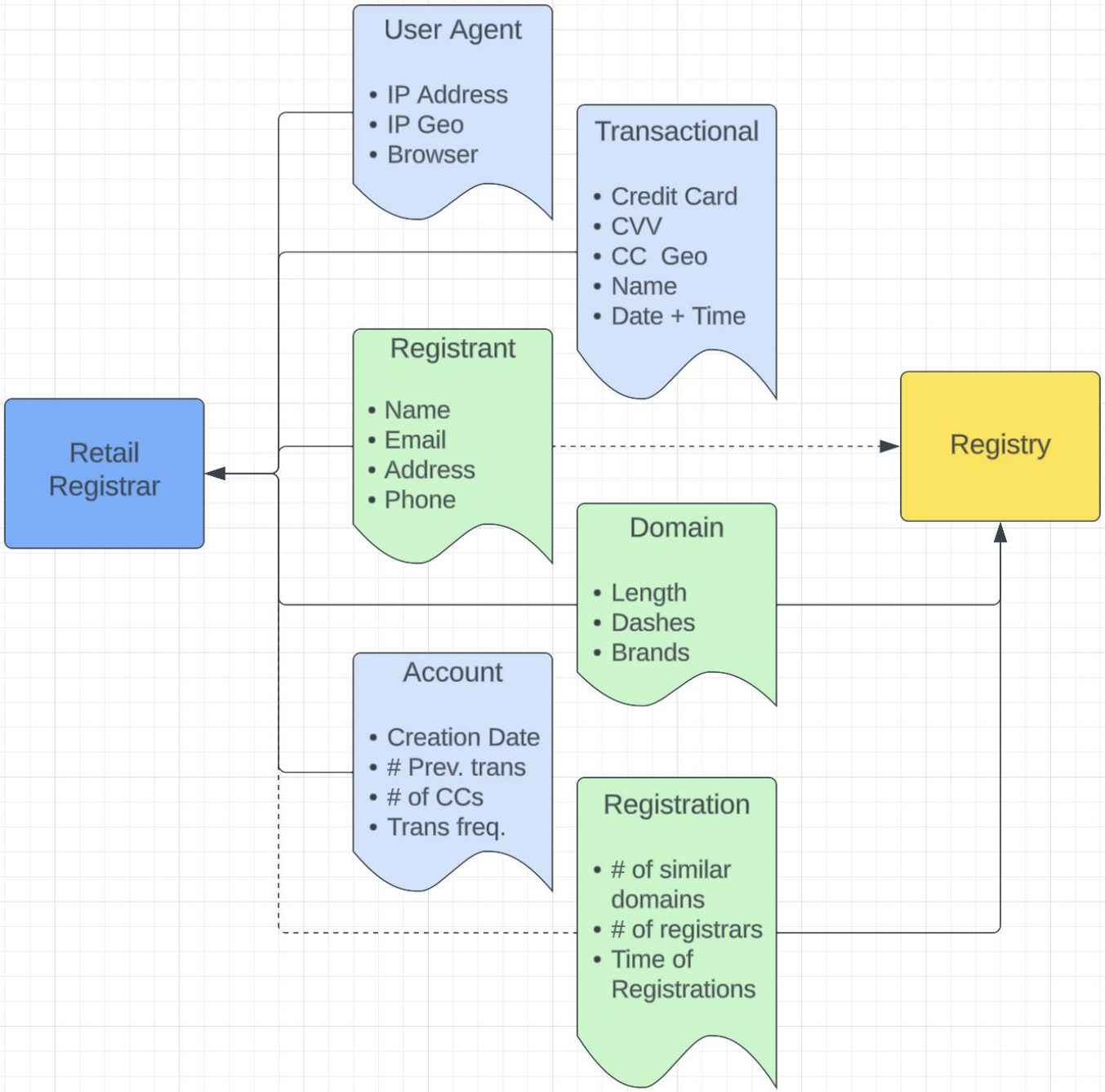
Can we predict potentially abusive domain names *before* the abuse has happened?

(as in, before an abusive website resolves or email is sent)

Key Findings

- AFAIK, AI detection of malicious names only implemented at ccTLDs
- Results are not an unambiguous success
- 3 approaches: domain attributes, registrant attributes, registration attributes
- (Retail) Registrars have much MUCH more useful data
- New Customer + string in list = review, can get you surprisingly far

Using what data?



NETBEACON INSTITUTE

Bibliography Link

A non-exhaustive list of the academic papers on the subject

Leveraging Machine Learning in DNS Abuse Research

An Overview of SSR Research

Carlos H. Gañán, PhD
Office of the CTO

ICANN 80
June 2024

Security, Stability & Resiliency (SSR) Research

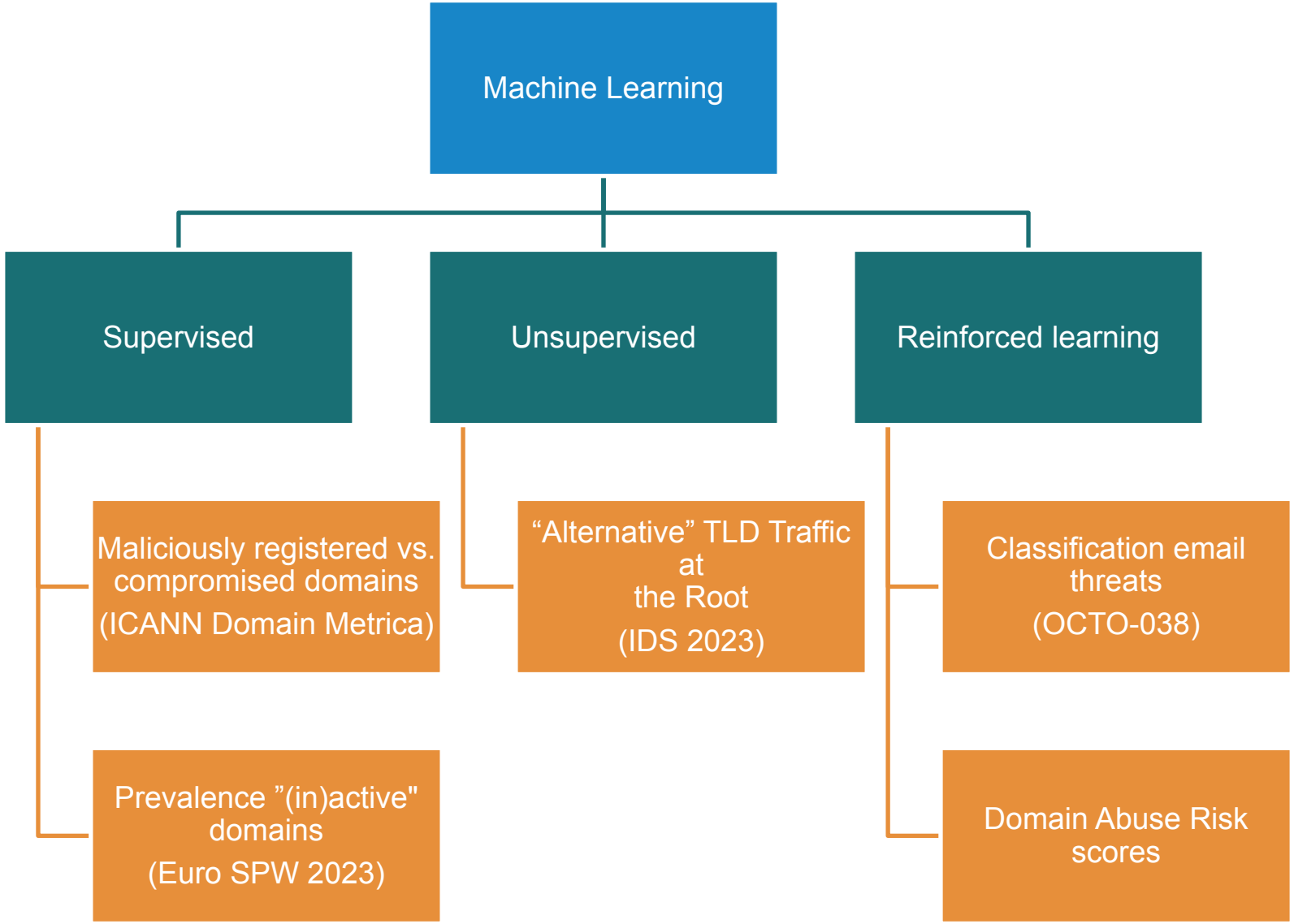
Samaneh Tajalizadehkhoob
Director, Security, Stability & Resiliency Research

Security, Stability, & Resiliency Research

Siôn Lloyd
Principal SSR Research Specialist

Carlos H. Gañán
Principal SSR Research Specialist

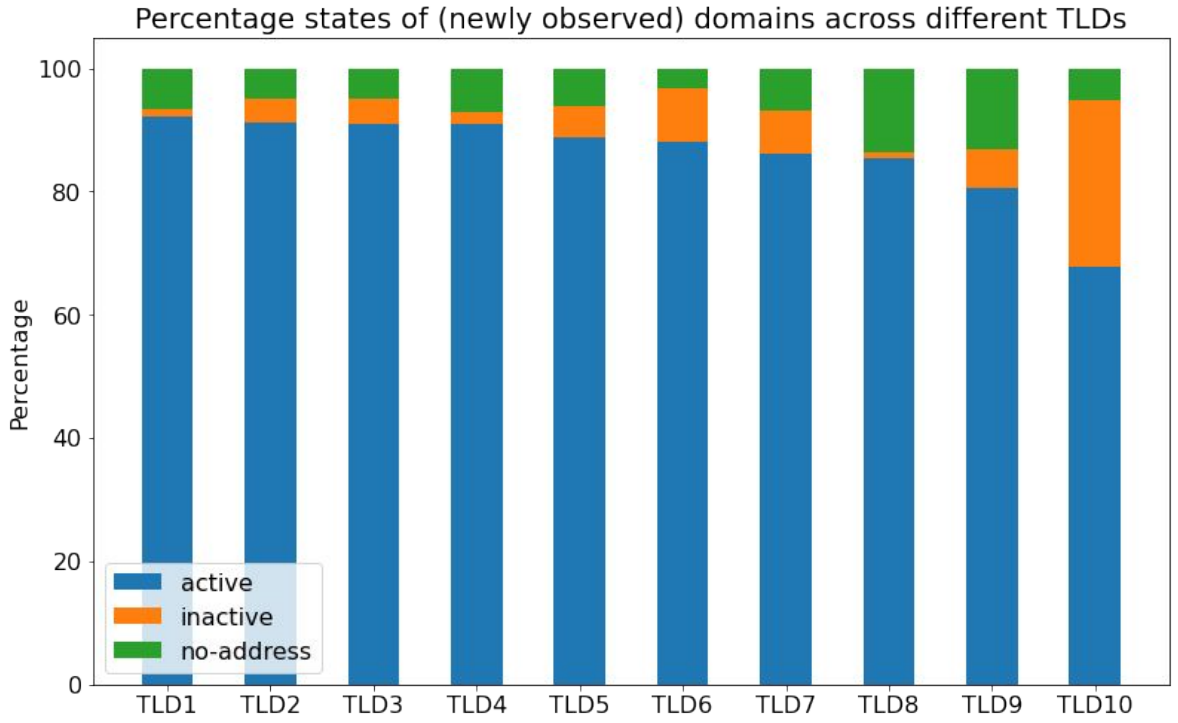
SSR Research Leveraging Machine Learning



Prevalence and implications of “Active” Domains – EuroSPW 2023

Objective: quantify the impact of "(in)active" domains through rule-based classifiers in DNS abuse metrics.

- **Data**
  - Zone file, active DNS measurements, registration data.
- **Methodology**
  - ML models trained on DNS lookup results to differentiate between active and inactive domains.
  - Rule-Based Classifier:
    - Development of rules to categorize domains based on DNS configurations (e.g., IP, CNAME, NS records).
    - Rule-based systems applied to identify specific domain states (e.g., parking services, verification pending).



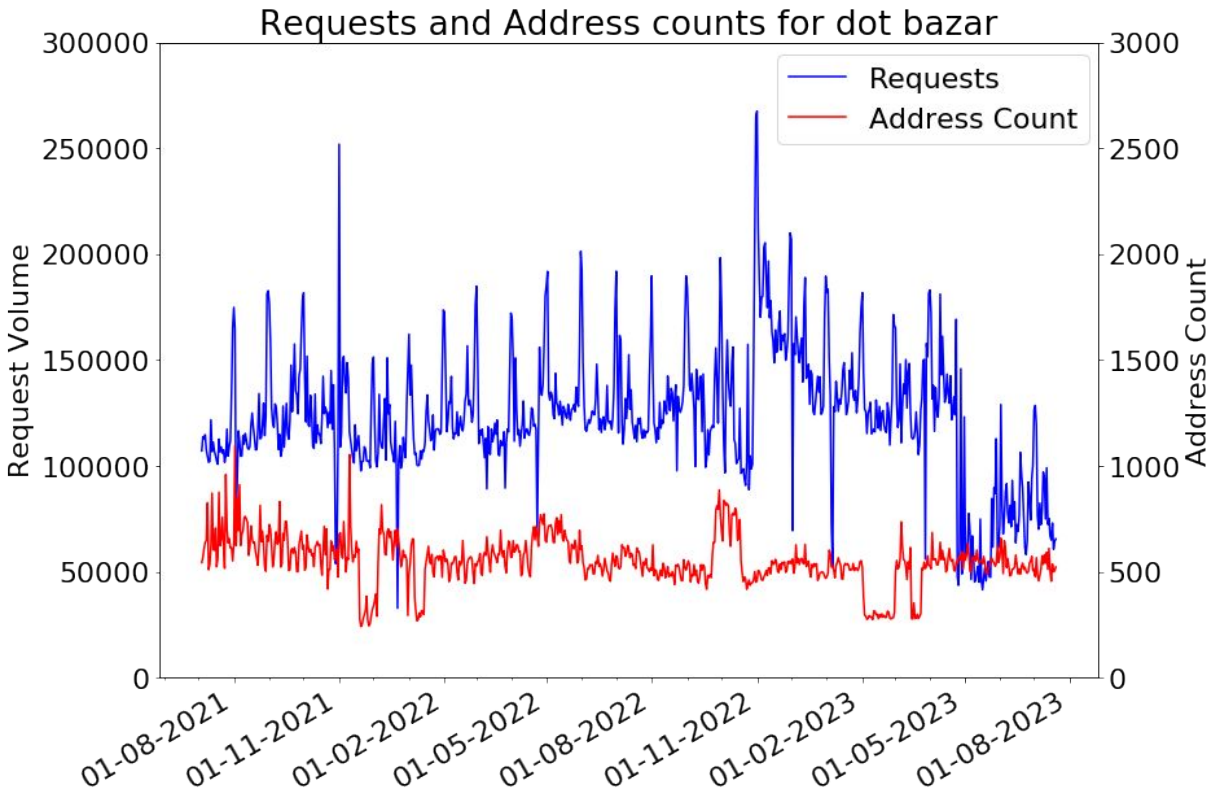
“Alternative” TLD Traffic at the Root – IDS 2023

Objective: Analyze and classify traffic patterns of decentralized DNS reaching the L-root.

- TLDs not in DNS Root: .bazar, .coin, .eth, .crypto, etc.

- **Data**
  - Sources: DNS Magnitude measurements, EmerCoin blockchain data.

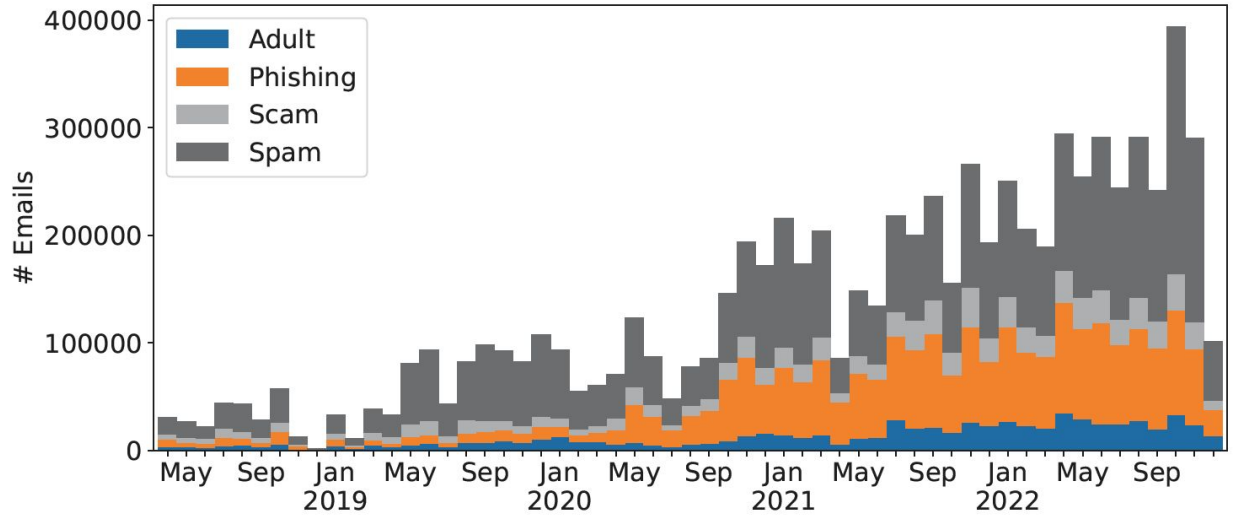
- **Methodology**
  - Clustering Algorithms: Used to group similar traffic patterns.
  - Anomaly Detection: Identify unusual traffic spikes and misdirected queries.
  - Traffic Analysis:
    - .bazar shows high request volumes linked to DGA activity.
    - .coin, .lib, .emc have lower but consistent traffic.



Classification email threats – OCTO 038

Objective: Develop a methodology to accurately classify unsolicited email threats using ML techniques.

- **Data**
  - Source: APWG phishing email feed.
    - 10.85M emails (2018-2022).
  - Categories: Phishing, Scam, Spam, Adult content.
- **Methodology**
  - Feature Extraction: Bitcoin addresses, file hashes, URLs, email/IP addresses, domains.
  - Classification:
    - Models: SVC, NB, LSTM, TF-IDF, LinearSVC, BILSTM
    - High performance (F1-score > 87.4%)



Ongoing research efforts

- **Domain risk scores using GNN**
  - ➜ Using relational graph structures to leverage labeled and unlabeled data (integrates IP addresses, email, and DNS protocol features) estimate domain abusive propensity score.

- **Domain uptime via deep learning**
  - ➜ Utilizing deep learning survival analysis to estimate the uptime of domains involved in DNS abuse, leveraging advancements in DL techniques tailored for time-to-event analysis.

- **KGs + LLMs for malicious domain extraction**
  - ➜ Automating the extraction of malicious domain information from unstructured text and contextualizing threat vectors using state-of-the-art Large Language Models (LLMs) and Knowledge Graphs (KGs).

Machine Learning in detecting DNS Abuse

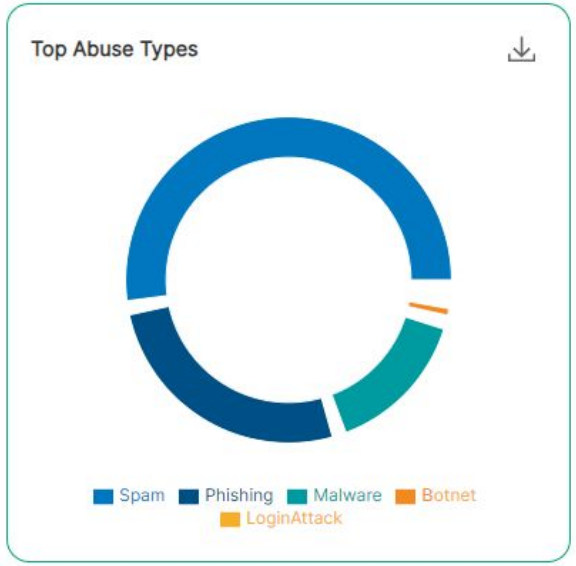
June 9, 2024

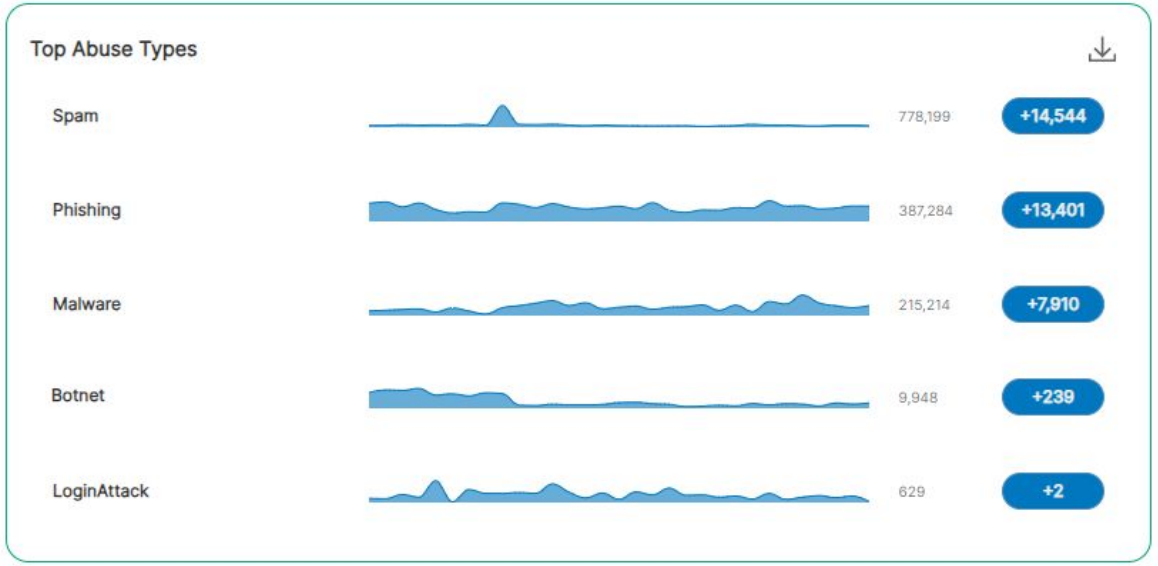
CleanDNS.com

CleanDNS

Current Landscape

- Validate and verify reports at scale:
  - Check for data entry ~1.5million reports per month
  - Validate abuse type
  - Produce recommended courses of action
  - Automate further verification checks
  - Verify or identify further evidential requirements
  - Provide further pivot points
  - DGA detection

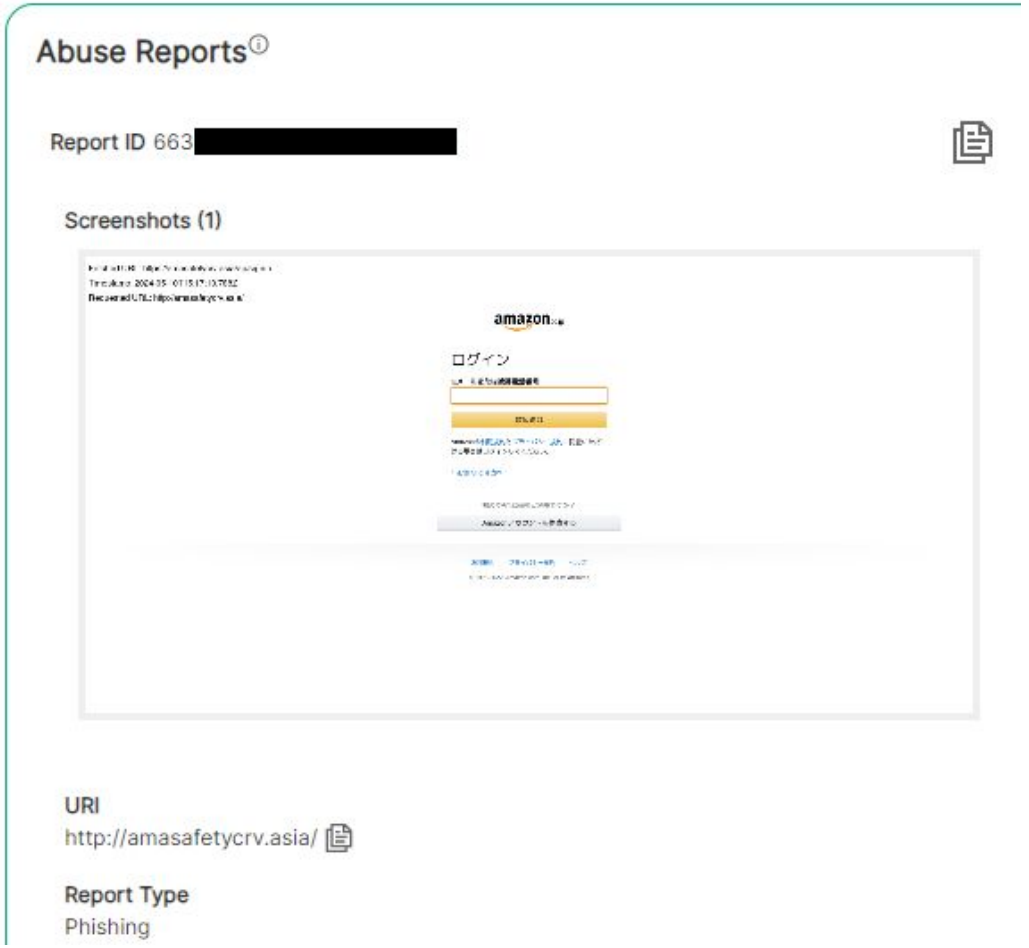




CleanDNS

Current Landscape

- Criminals try to hide phishing pages from investigators:
  - Current use of ML to try different access methods to get a good capture



CleanDNS

Continuing Research and Focus

Utilization of AI to identify victims.

Use of LLM to give a non-technical description of malware activity.

Considerations for automated threat response: Automated response strategies are essential for timely mitigation of DNS abuse.

Impact of data correlation on intelligence sharing: Data correlation enables sharing insights to enhance collaborative DNS abuse prevention.

Continuing Research and Focus

Challenges:

- Over identification - False Positives and Negatives
- High Implementation and Maintenance Costs
- Data Privacy and Security Concerns
- Complexity and Scalability Issues
- Adversarial Attacks / Learning

AI in DNS Security

AI in DNS Security

- Attacks are increasing in number, speed, personalization.
- Defenders using AI to recognize attacks (image recognition)
- New paradigm: predictive, proactive defense, rather than reactive.
- Quandary: how do you send and evaluate abuse complaints when an attack has not happened yet?
- Solution: Trusted reporters.
- Solution: Hosting providers, registrars, and registry operators can use AI to recognize malicious registrations, indicators on hosting, etc.

Interisle ILLUMINTEL

Areas for Further Research and Focus

The Role of SSAC in Supporting Research Efforts

Closing Remarks