Proactively hunting for low-reputed infrastructure used by large cybercrimes and APTs

Nabeel Mohamed, Keerthiraj Nagaraj, Janos Szurdi, Alex Starov

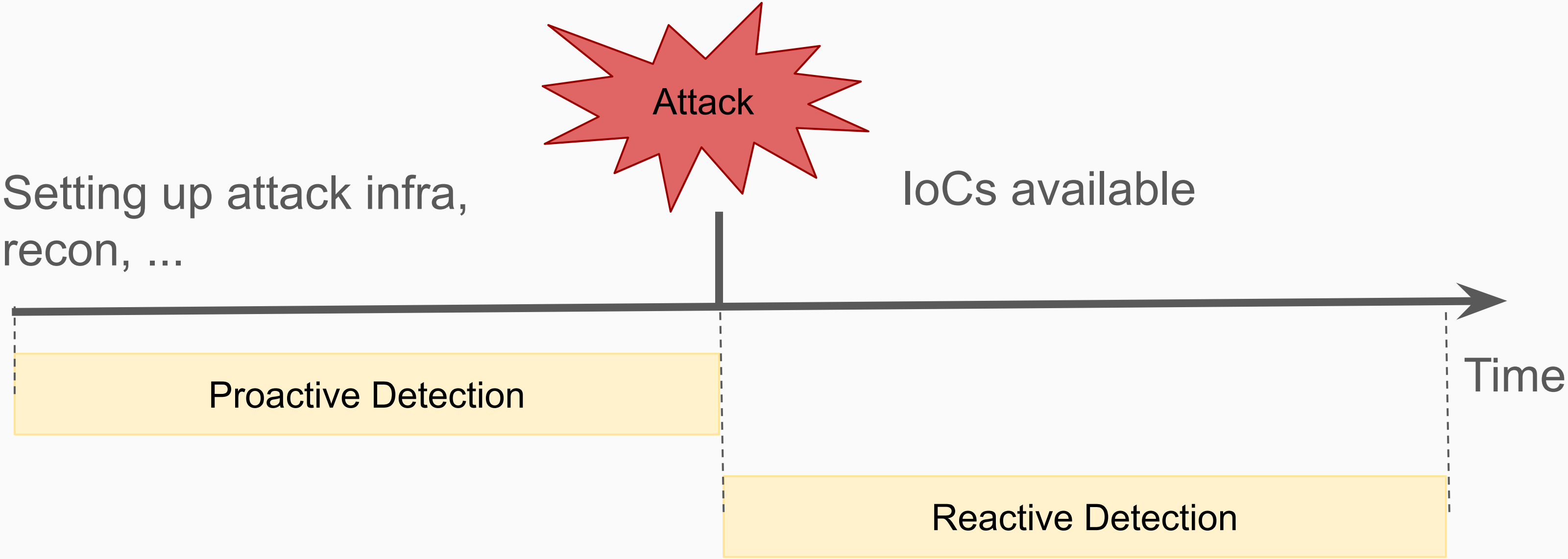
10/05/2024

Agenda

- Motivation with examples
- Methodology
    - Knowledge graph construction
    - Graph AI learner
- Case studies

Introduction

- Reactive: Currently, a lot of attacks are detected ***after*** they are launched
- Proactive: Can we detect attacks ***before*** *they are launched* or ***early*** during the attack?



Observations

Attackers often

- **Rotate** their attack infrastructure (domains, IPs, file hashes, certificates)
- **Automate** hosting related activities
- **Reuse or share** the same attack infrastructure

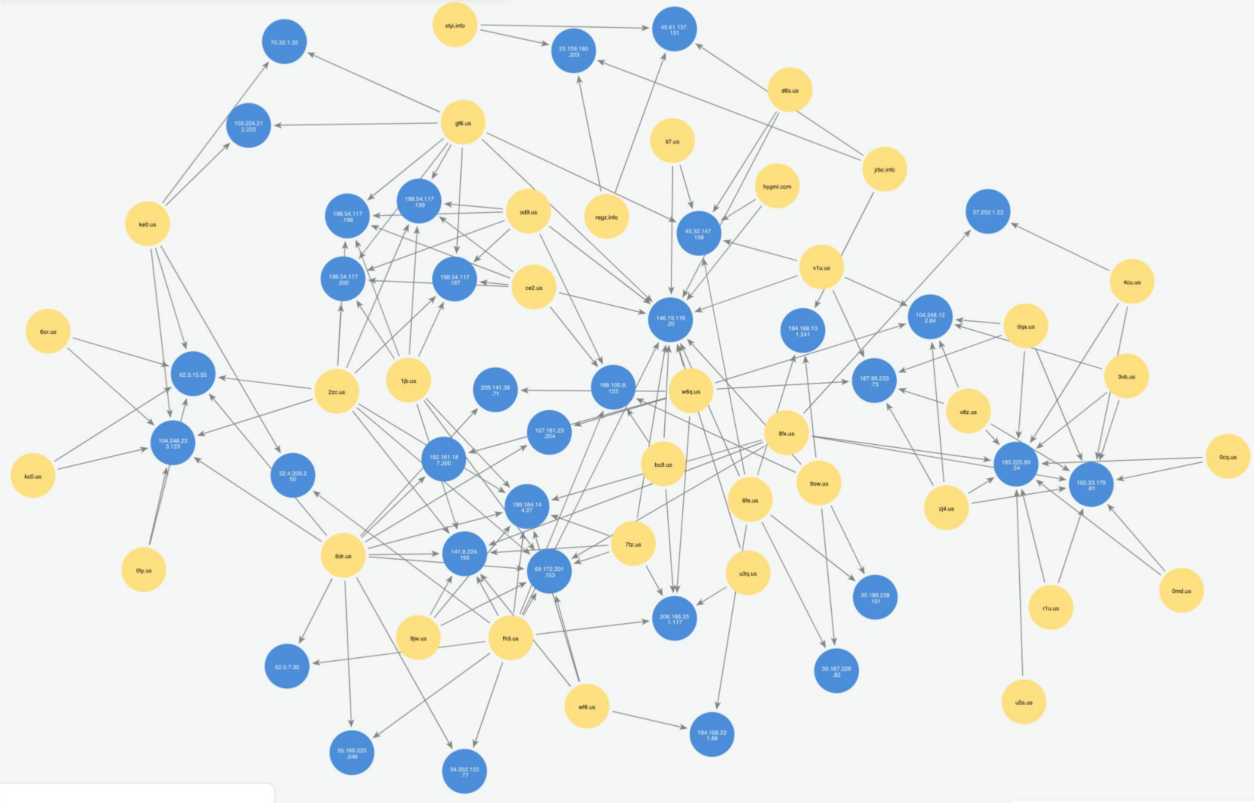
Attackers set up their infrastructure **before** they launch the attack.

Existing analyzers often **detect only parts of** active attack infrastructures.

Pivot on these observations to proactively protect **patient zero** victims.

Example Resource Sharing in the Web

Malicious Domains Share/Rotate Hosting Infrastructure



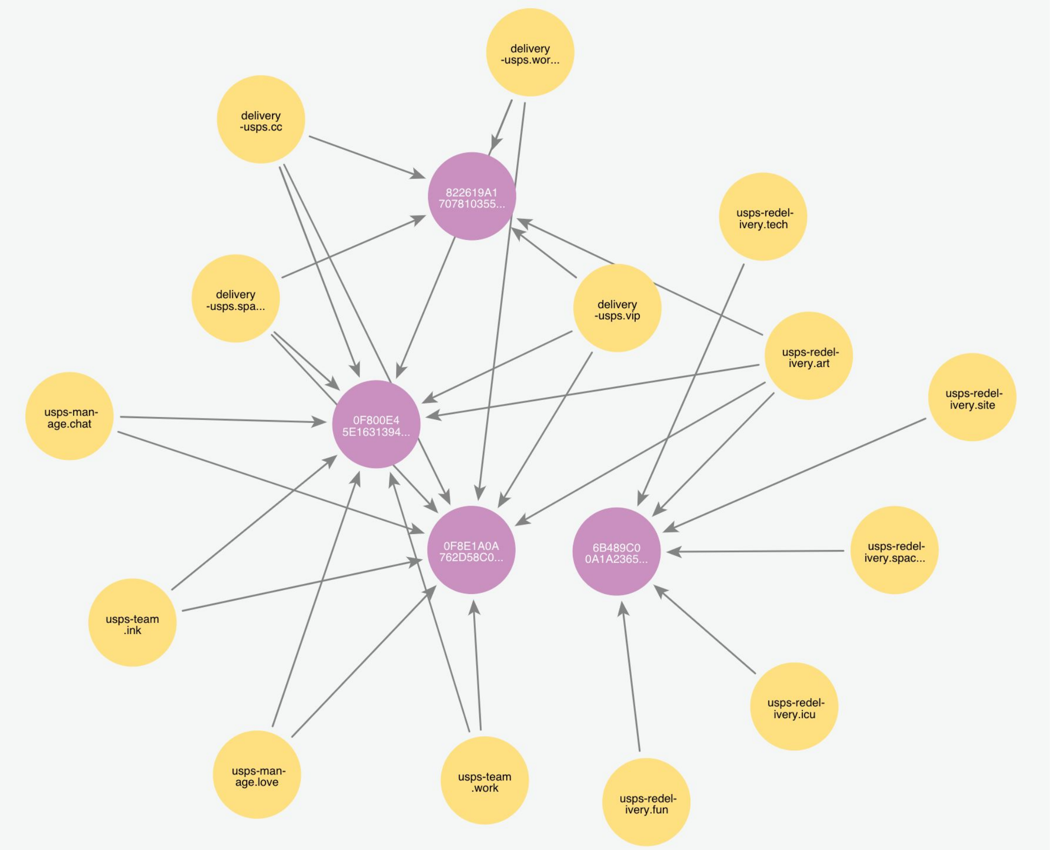
Prolific Puma malicious link shortening service

● Malicious domains

● IP addresses

Top hosting services:
- BL Networks
- AS-CHOOPA
- NameCheap
- Amazon
- Digital Ocean

Malicious Domains Share TLS Fingerprints

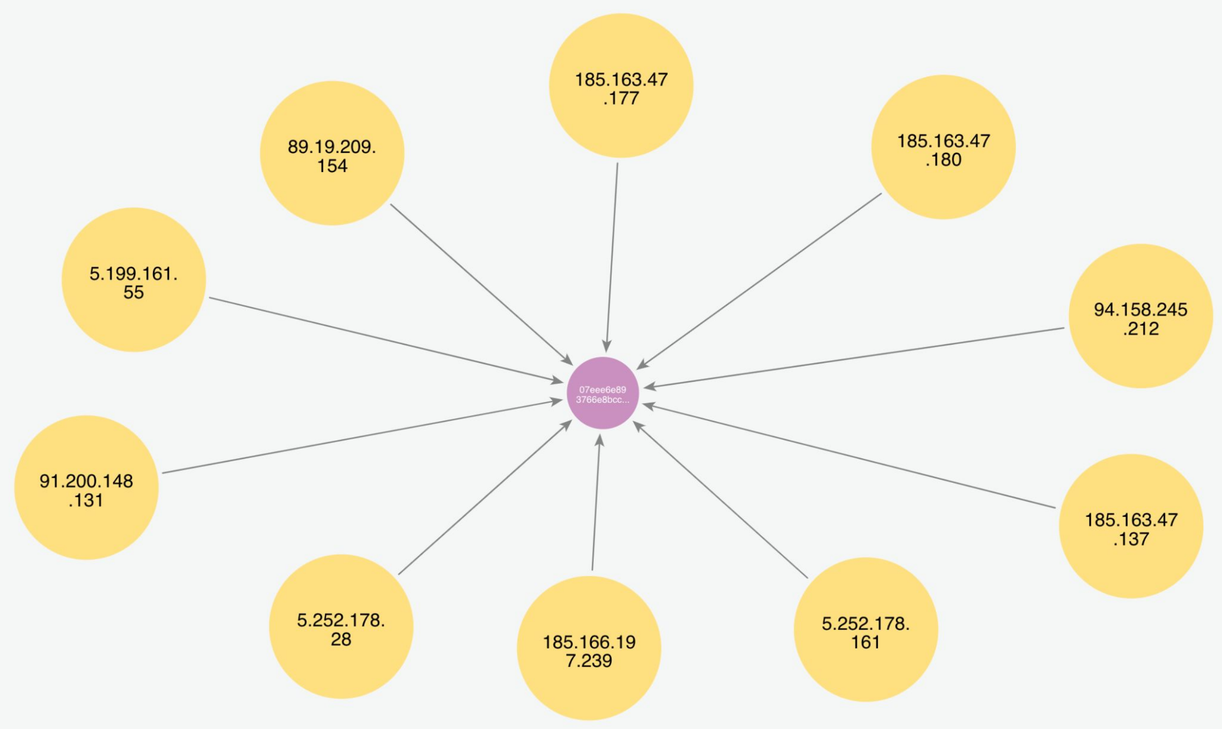


● Malicious domains

● TLS certificate fingerprints

USPS phishing campaign

Multiple IP Addresses Share Same SSH Fingerprint

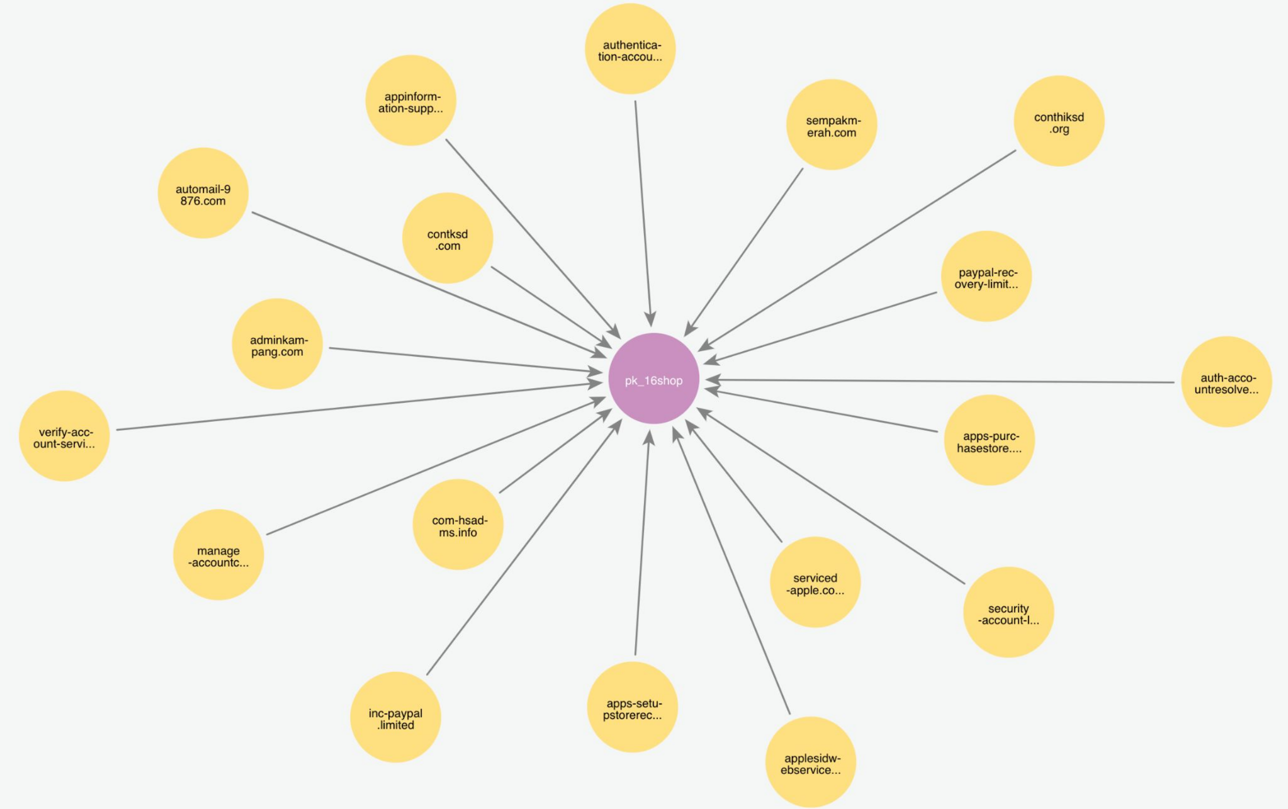


● Malicious IPs

● SSH fingerprint

An active self-signed certificate used by Gamaredon

Multiple Phishing Sites Use the Same Phishing Kit

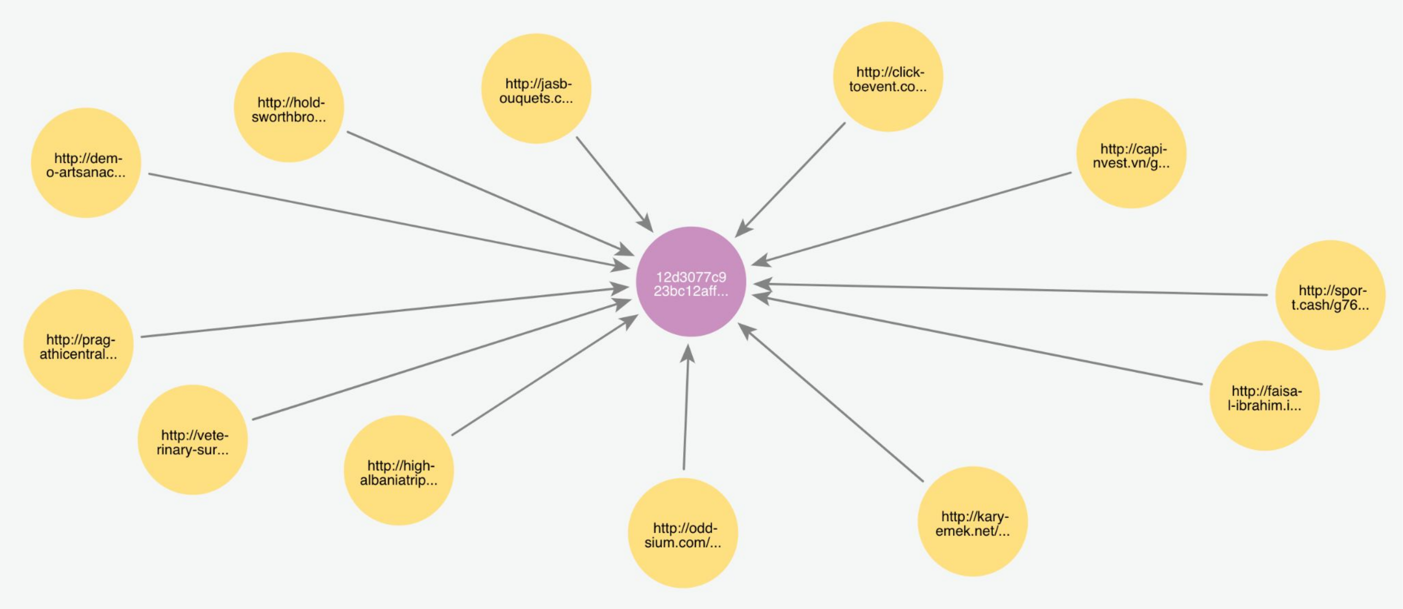


● Malicious domains

● Phishing kit

Phishing sites using 16shop phishing kit

Multiple Malicious URLs Distribute Same Malware

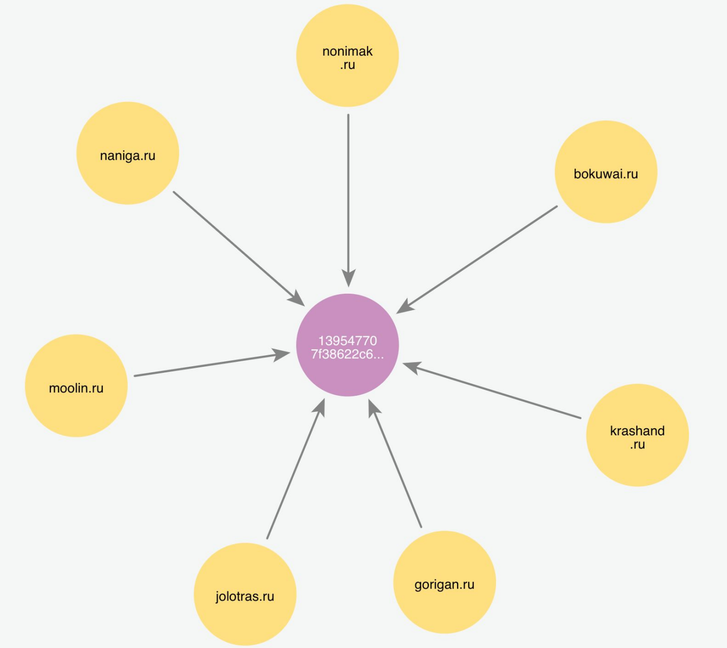

http://click-toevent.co...
http://jasb-ouquets.c...
http://hold-sworthbro...
http://dem-o-artsanac...
http://capi-nvest.vn/g...
12d3077c9 23bc12aff...
http://spor-t.cash/g76...
http://prag-athicentral...
http://faisa-l-ibrahim.i...
http://vete-rinary-sur...
http://high-albaniatrip...
http://odd-sium.com/...
http://kary-emek.net/...

● Malicious URLs

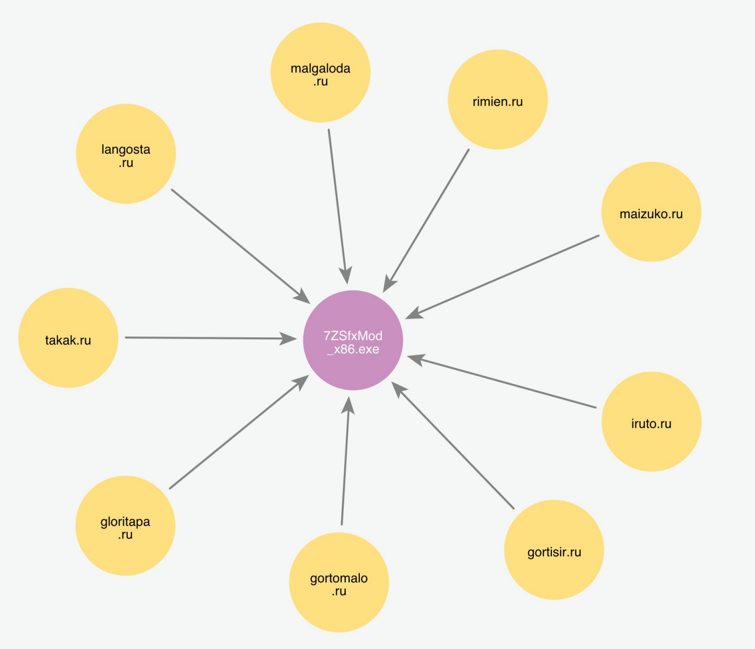
● Malware hash

TeslaCrypt delivery URLs

Same Malware Connects to Multiple C2 Domains



Gamaredon stealer



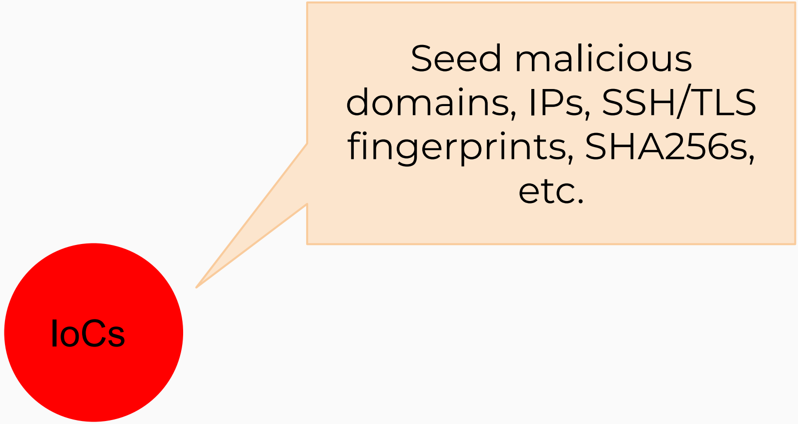
Gamaredon remote admin tool (Pteranodon)

● Malicious domains

● File hashes

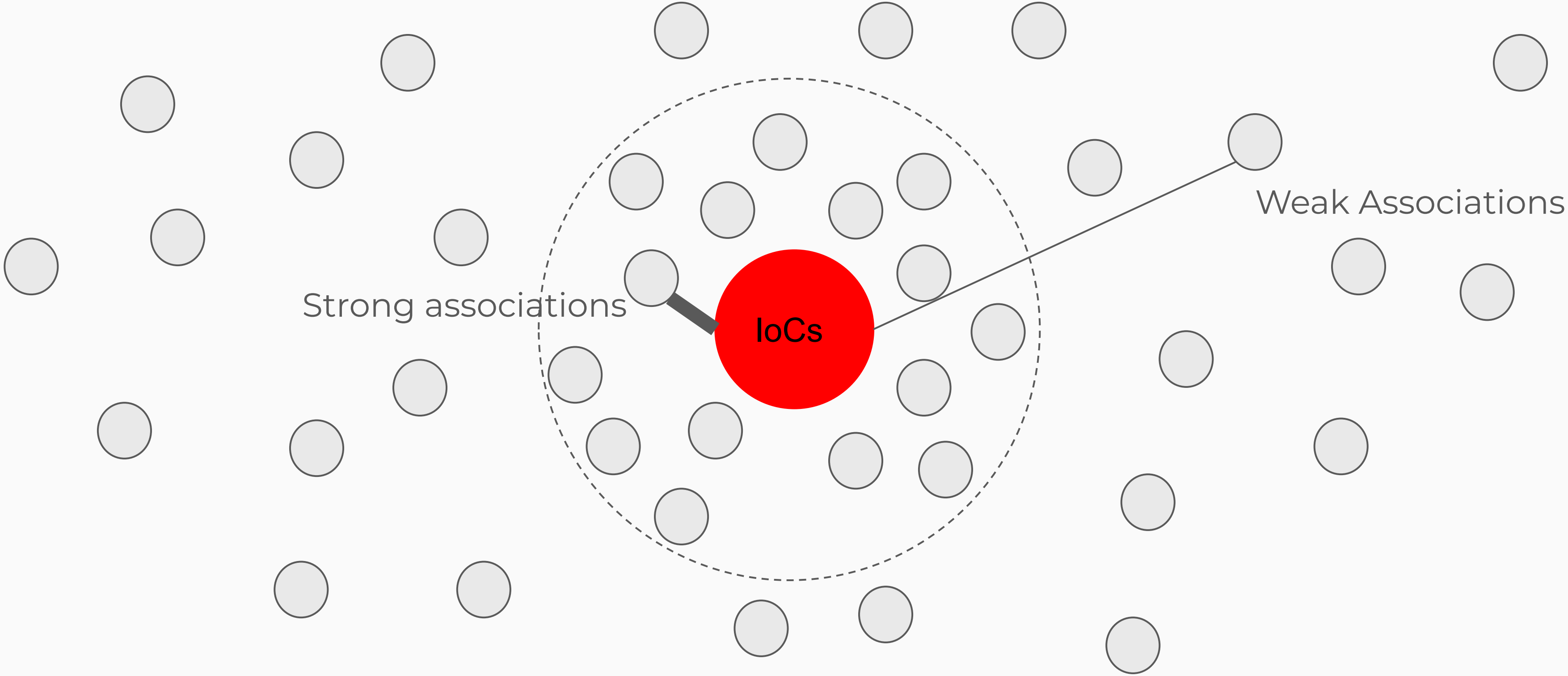
Our Approach

Key Idea: Automated Pivoting + Feature Similarity

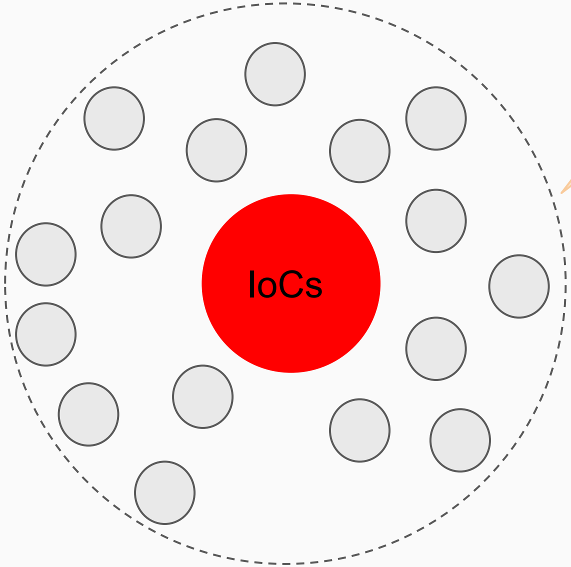


Seed malicious domains, IPs, SSH/TLS fingerprints, SHA256s, etc.

Key Idea: Automated Pivoting + Feature Similarity

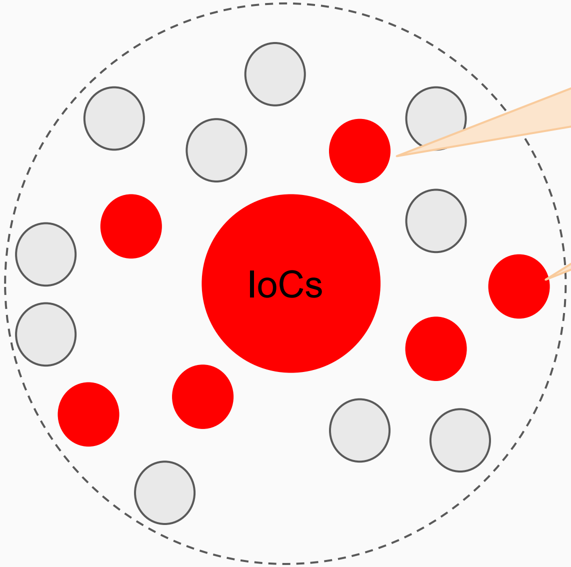


Key Idea: Automated Pivoting + Feature Similarity



High toxicity neighborhood

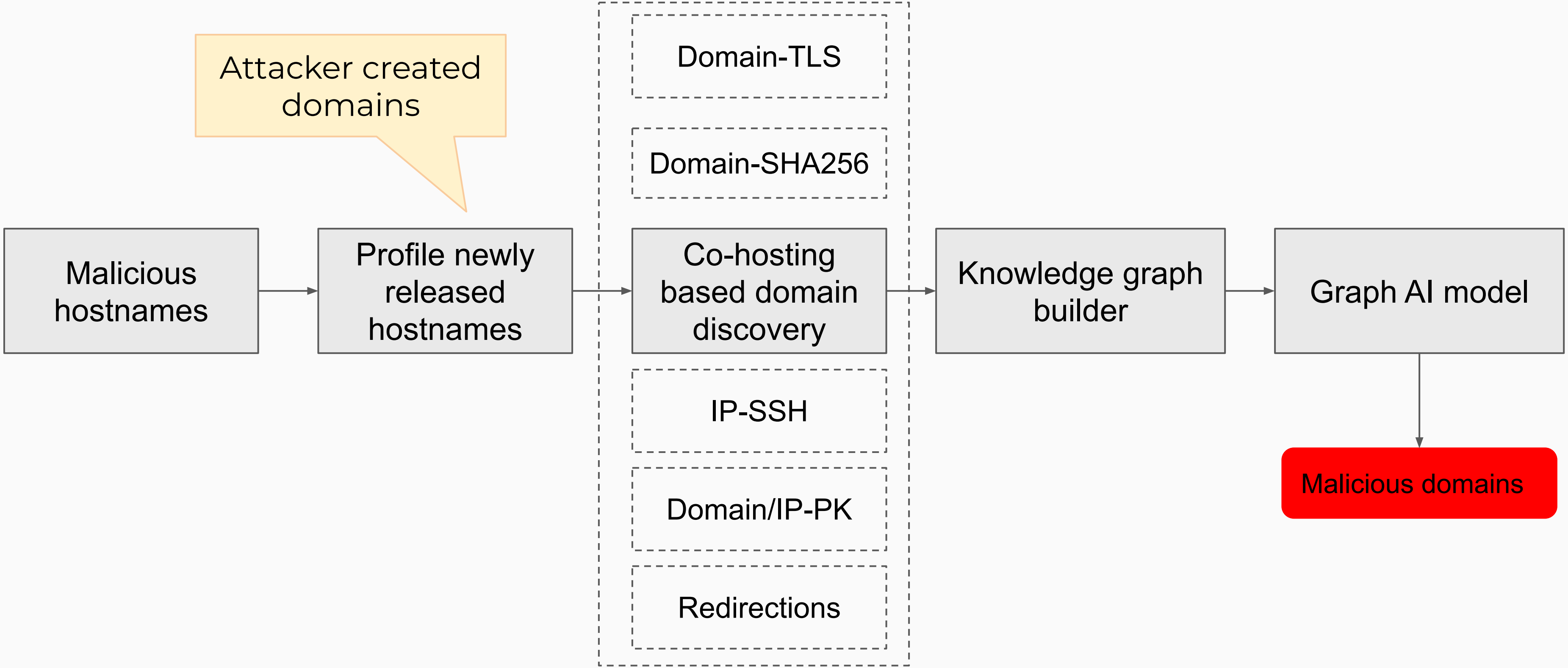
Key Idea: Automated Pivoting + Feature Similarity



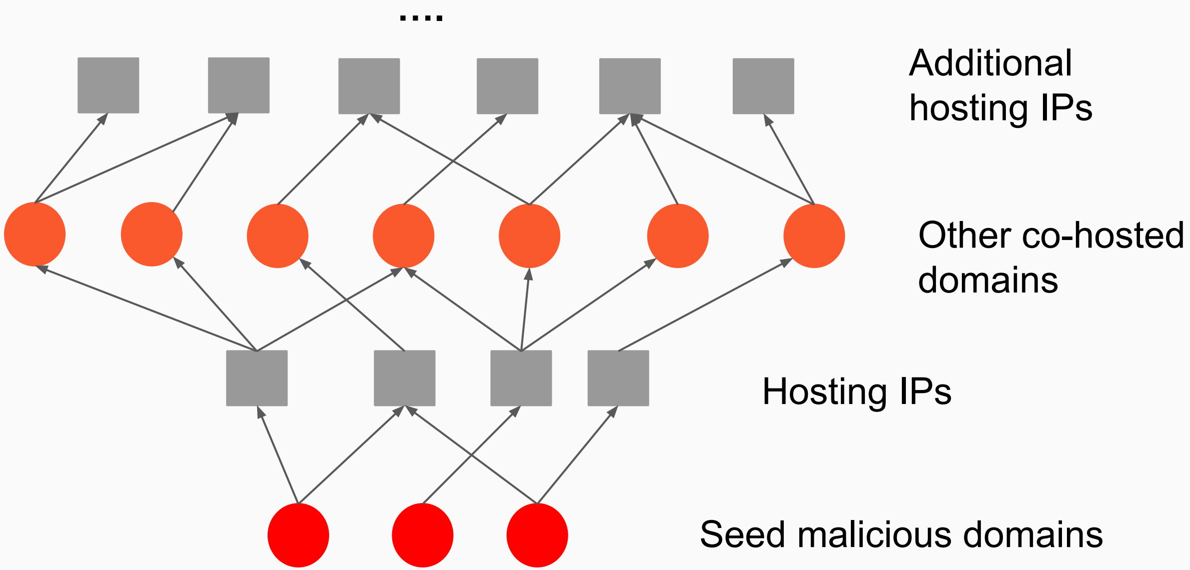
Domains similar in features to known malicious domains*

* Same applies to IPs

Overall Pipeline



Guided Discovery of Domains (Co-Hosting Relationship)



Graph AI-based Detection of Malicious Domains

Graph Schema

- Nodes
  - Domain
  - Subdomain
  - IP
  - File hash
  - TLS/SSH certificate fingerprint
- Edges
  - Domain-Subdomain
  - Domain-IP
  - Domain-FileHash
  - IP-SSH, Domain-TLS

Labeled Data

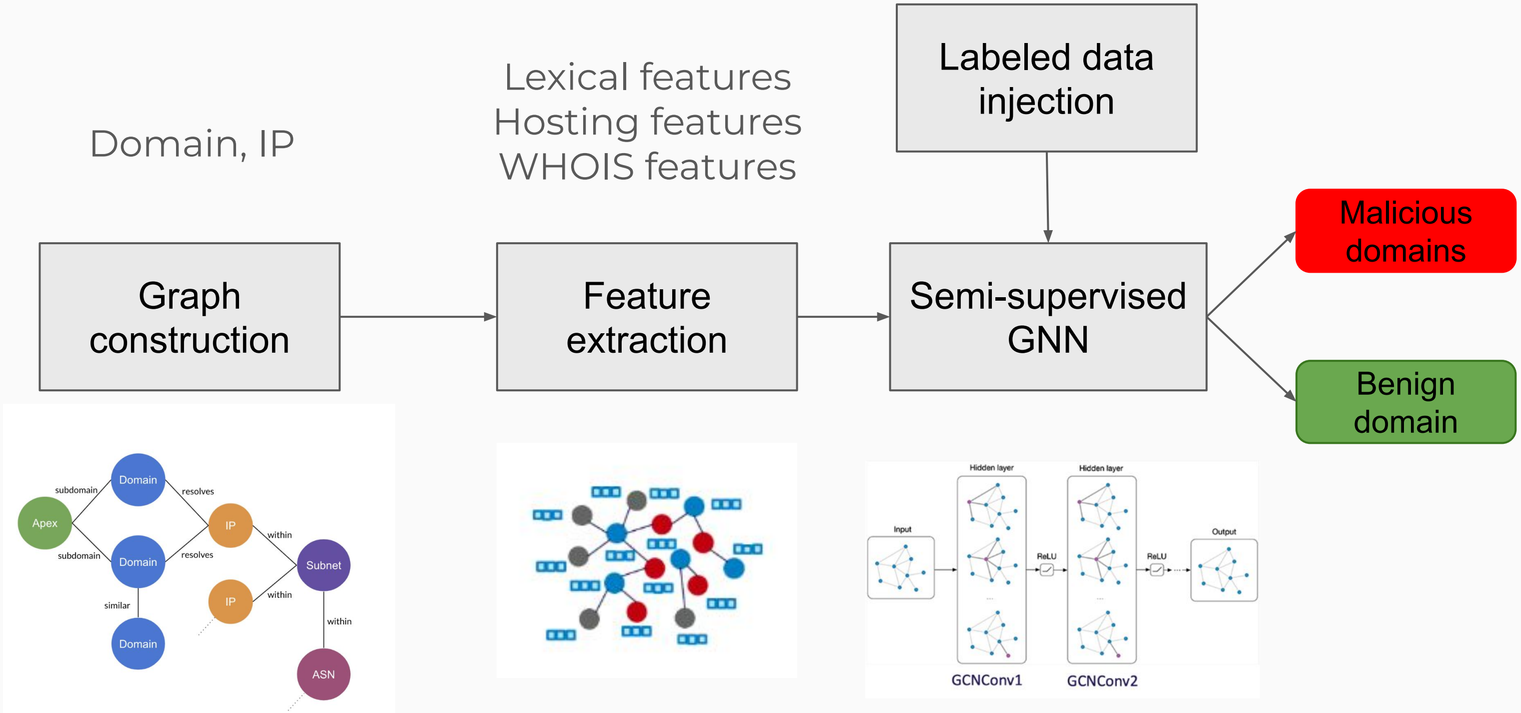
- Malicious
    - In-house malicious domains
- Benign
    - Tranco top 100K domains
    - In-house benign domains

Features

- **Lexical features** (e.g., # brand/suspicious keywords, # hyphens)
- **Hosting features** (e.g., # IPs, hosting duration)
- **WHOIS features** (e.g., age, days to expiration, privacy)
- **Certificate features** (e.g., type, issuer)
- **IP features** (e.g., # domains, ASN, CC)
- **Content-based features** (e.g., # iframes, webform?)

Training the Graph AI (GNN) Model

(2K from each class)



Domain, IP

Graph construction

Lexical features
Hosting features
WHOIS features

Feature extraction

Labeled data injection

Semi-supervised GNN

Malicious domains

Benign domain

GraphSAGE

Preliminary Results

Model	Precision*	Recall*
Local features	81.05	70.10
Shallow embedding (node2vec)	84.07	72.23
Shallow embedding (metapath2vec)	86.22	74.54
Local features + Shallow embedding	89.01	78.32
GNN	**95.20**	**92.30**

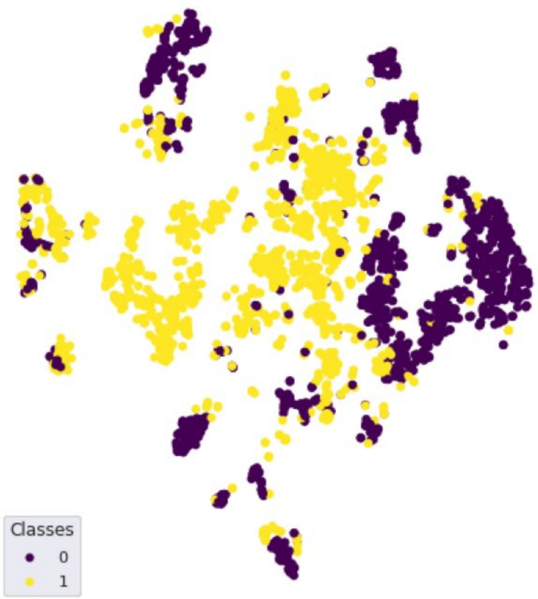
* At 0.5 default cut-off threshold

Metric\Thresh.	0.50	0.98
Precision	95.2%	99.9%
Recall	92.3%	53.1%

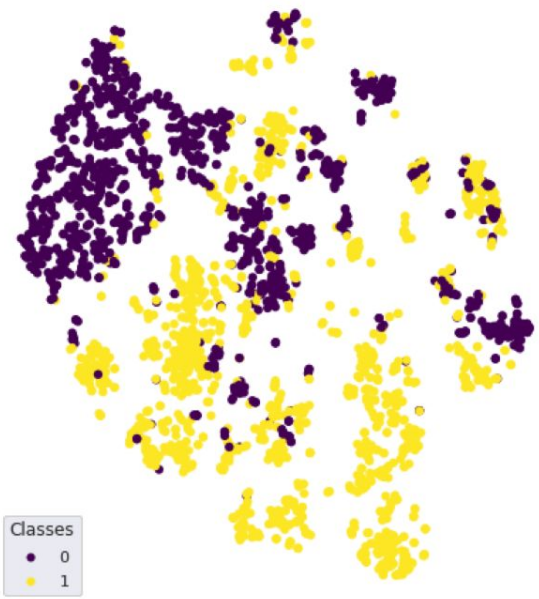
Results - Why it works

● Benign

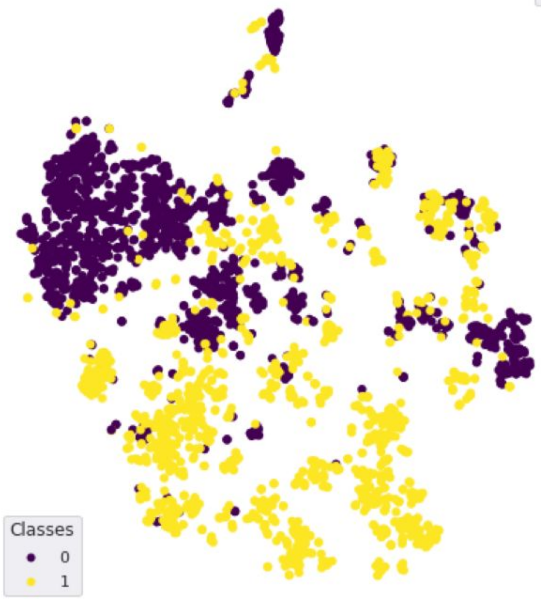
● Malicious



Week 1



Week 2

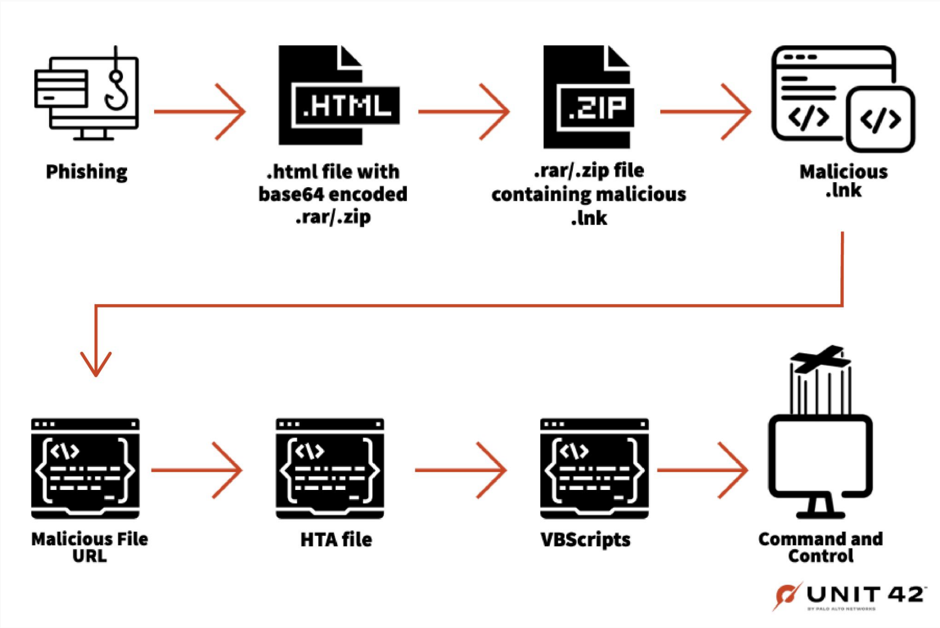


Week 3

Case Studies

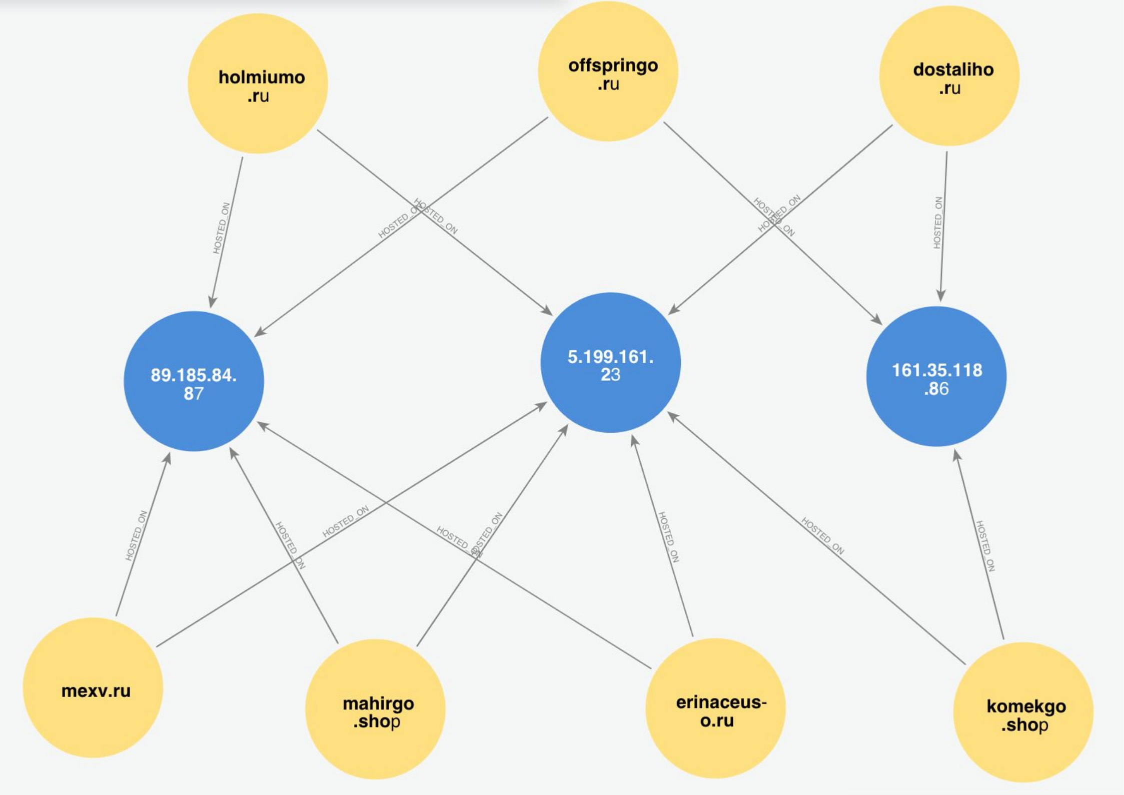
Case Study 1: Gamaredon APT

- A prominent Russian APT group targeting mainly Ukraine
- Operational since 2014



Gamaredon - Seed Domains

- offspringo.ru
- dostaliho.ru
- komekgo.shop
- mexv.ru
- erinaceuso.ru
- mahirgo.shop
- holmiumo.ru



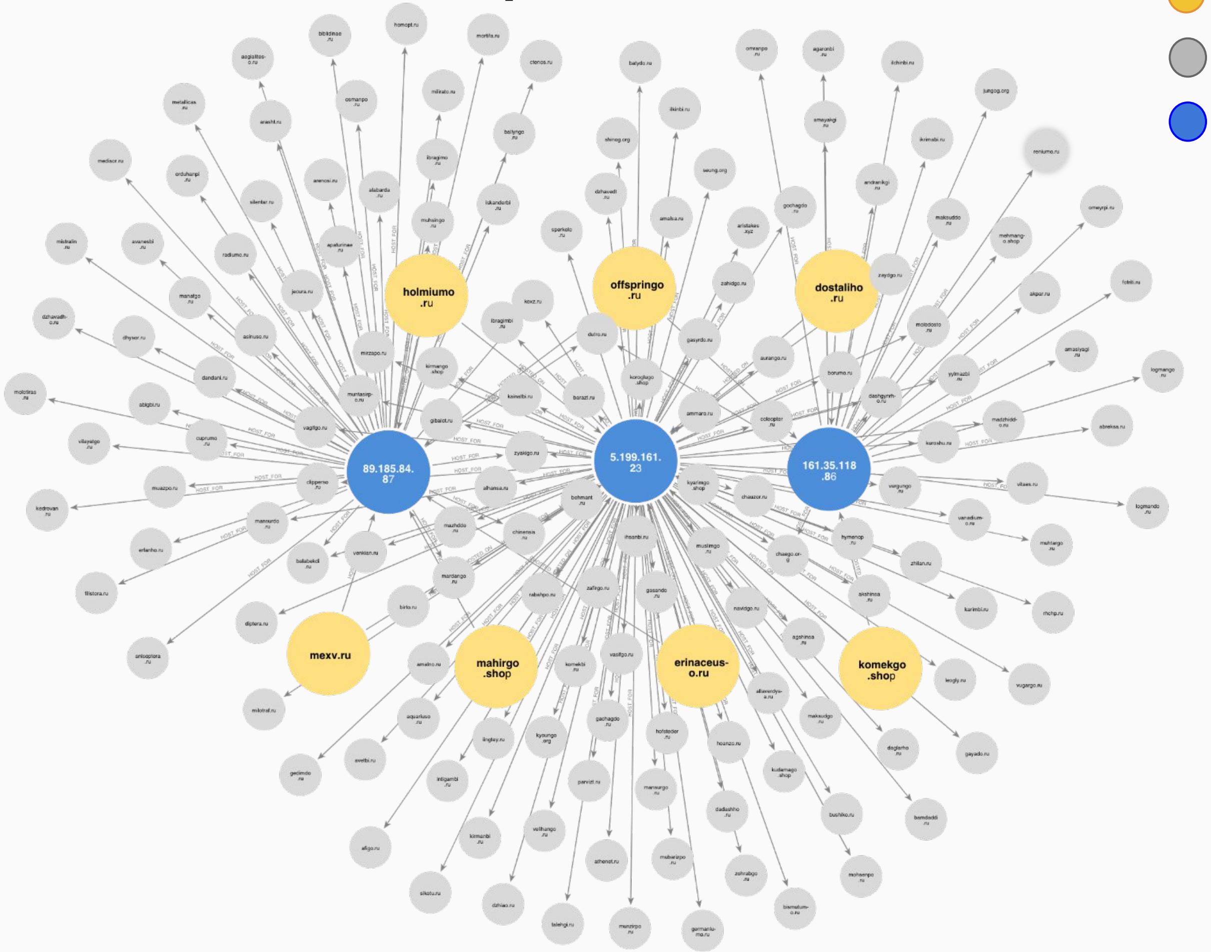
Hosting Infrastructure

Gamaredon - Guided Expansion

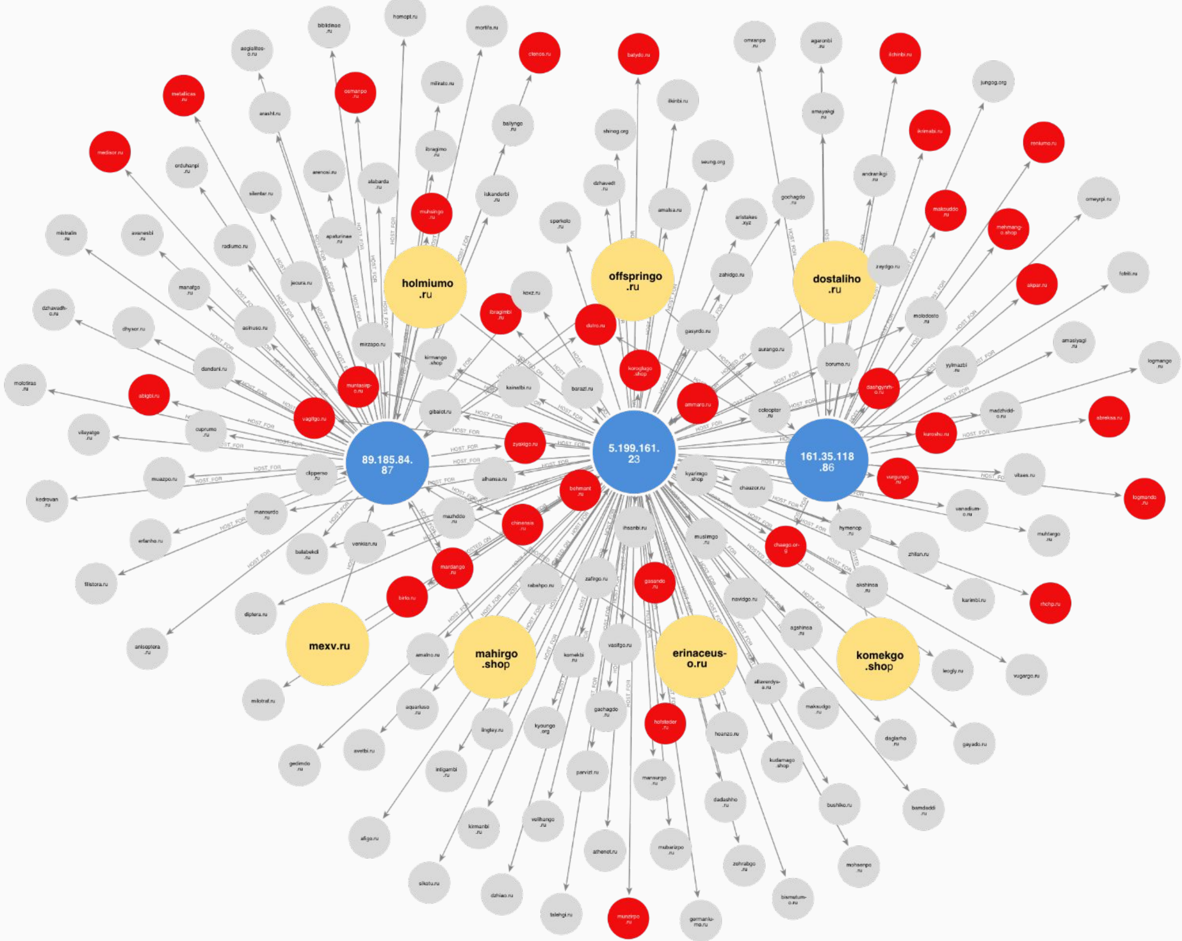
● Seed malicious domains

● Expanded unknown domains

● IP addresses



Gamaredon - Flagged Malicious Domains



🟡 Seed malicious domains

⚪ Expanded unknown domains

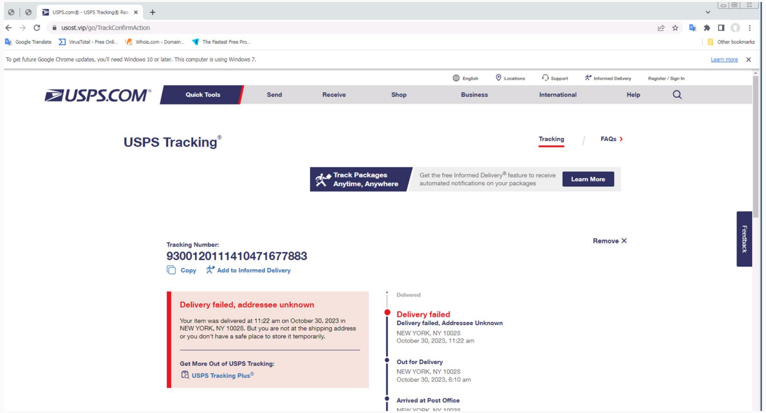
🔵 IP addresses

🔴 Flagged malicious domains

Later 34 domains were flagged later as Malware by other vendors.

Case Study 2: Postal Phishing Campaign

- A recent campaign targeting USPS and 12 other national postal services around the world.
- Attack vector: Smishing
- Collected ~450 seed domains from this campaign
  - Hosted on ~400 unique IP addresses
- Identified ~5000 additional domains hosted on these ~400 IP addresses in the last 3 months.
  - ~30% of them later flagged malicious by other vendors



usost.vip/go/TrackConfirmAction

To get future Google Chrome updates, you'll need Windows 10 or later. This computer is using Windows 7.

USPS Tracking®

Track Packages Anytime, Anywhere

Get the free Informed Delivery® feature to receive automated notifications on your packages

Tracking Number:
9300120111410471677883

Delivery failed, addressee unknown

Your item was delivered at 11:22 am on October 30, 2023 in NEW YORK, NY 10028. But you are not at the shipping address or you don't have a safe place to store it temporarily.

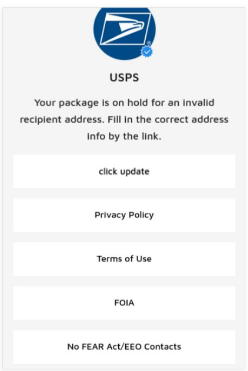
Get More Out of USPS Tracking:
USPS Tracking Plus®

Delivered

Delivery failed
Delivery failed, Addressee Unknown
NEW YORK, NY 10028
October 30, 2023, 11:22 am

Out for Delivery
NEW YORK, NY 10028
October 30, 2023, 6:10 am

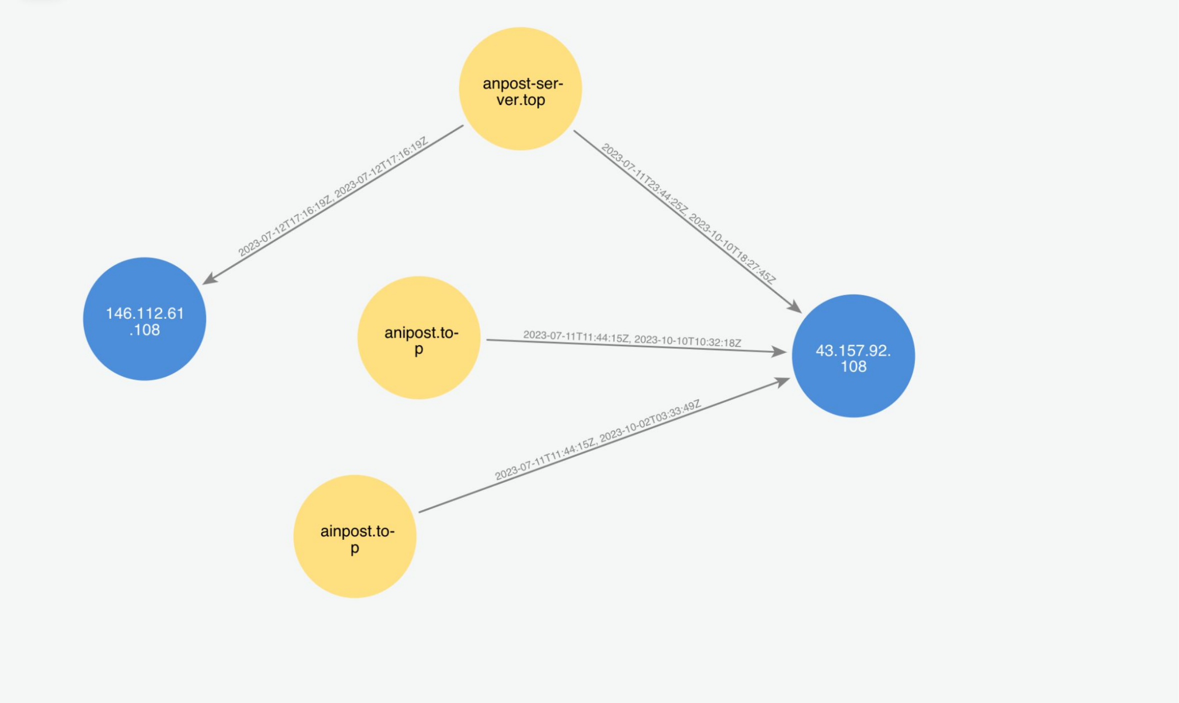
Arrived at Post Office
NEW YORK, NY 10028



USPS

Your package is on hold for an invalid recipient address. Fill in the correct address info by the link.

Postal Phishing Campaign: Seed Domains and Hosting Infrastructure

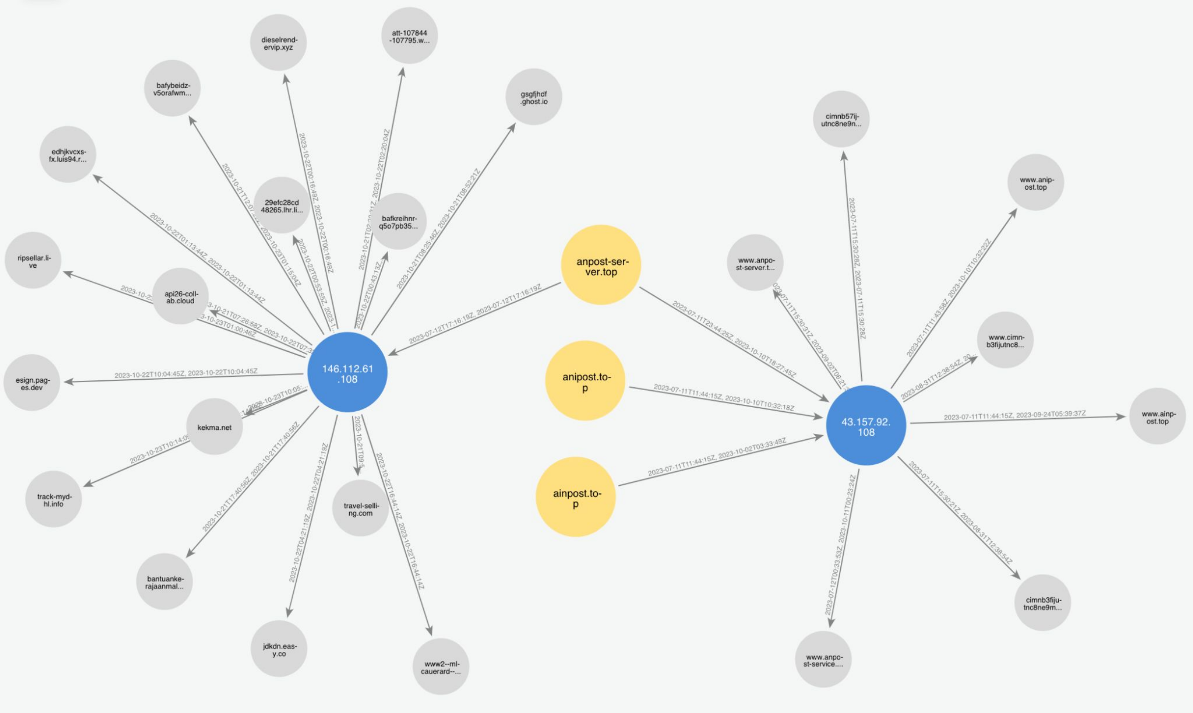

anpost-ser-
ver.top
2023-07-12T17:16:19Z, 2023-07-12T17:16:19Z
2023-07-11T23:44:25Z, 2023-10-10T18:27:45Z
146.112.61
.108
anipost.to-
p
2023-07-11T11:44:15Z, 2023-10-10T10:32:18Z
43.157.92.
108
2023-07-11T11:44:15Z, 2023-10-02T03:33:49Z
ainpost.to-
p

● Seed malicious domains

● IP addresses

Hosting infrastructure shared by phishing domains targeting anpost[.]com (Ireland’s national postal service).

Postal Phishing Campaign - Graph Expansion



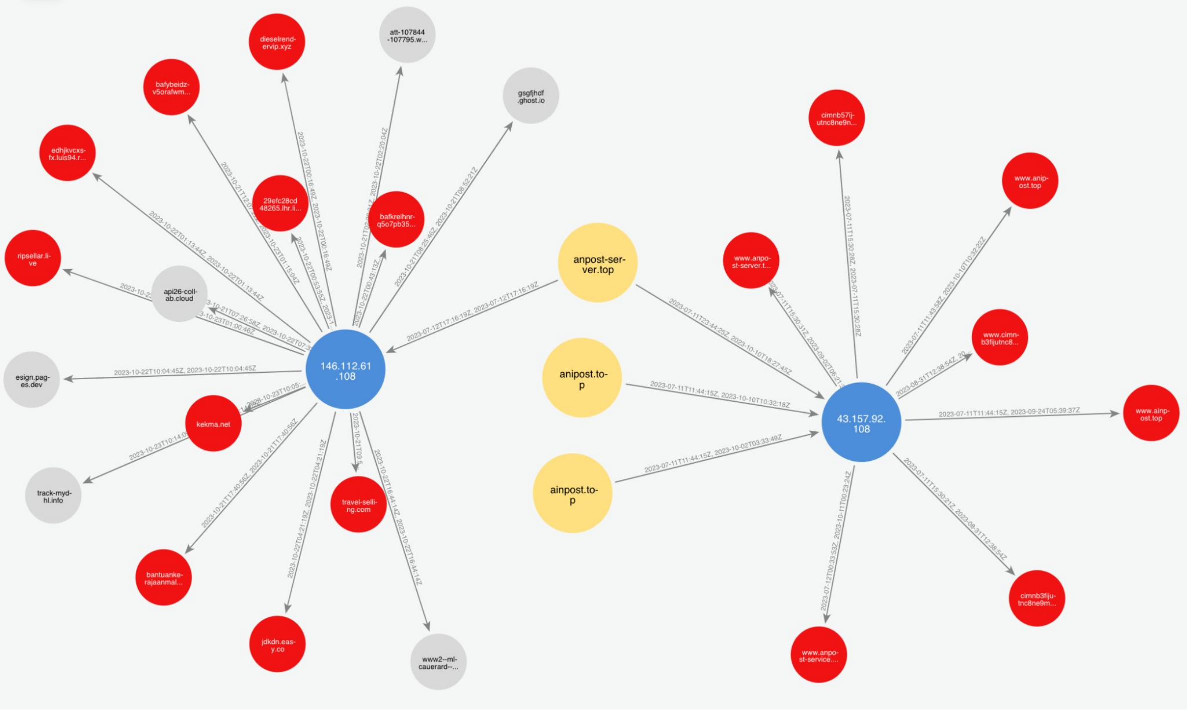
● Seed malicious domains

● Expanded unknown domains

● IP addresses

Graph expansion for the phishing pages targeting An Post (anpost[.]com)

Postal Phishing Campaign - Flagged Malicious Domains


dieselrend-ervip.xyz
att-107844-107795.w...
bafybeidz-v5orafwm...
gsgfjhdf.ghost.io
edhjkvcxs-fx.luis94.r...
29efc28cd48265.lhr.li...
bafkreihnr-q5o7pb35...
ripsellar.li-ve
api26-coll-ab.cloud
146.112.61.108
esign.pag-es.dev
kekma.net
track-myd-hl.info
travel-selli-ng.com
bantuanke-rajaanmal...
jdkdn.eas-y.co
www2--ml-cauerard--...
anpost-ser-ver.top
anipost.to-p
ainpost.to-p
cimnb57ij-utnc8ne9n...
www.anpo-st-server.t...
www.anip-ost.top
www.cimn-b3fijutnc8...
43.157.92.108
www.ainp-ost.top
cimnb3fiju-tnc8ne9m...
www.anpo-st-service....
2023-10-22T00:16:49Z, 2023-10-22T00:16:49Z
2023-10-22T02:20:04Z
2023-10-21T08:25:46Z, 2023-10-21T08:52:21Z
2023-10-22T01:13:44Z, 2023-10-22T01:13:44Z
2023-10-23T01:15:04Z
2023-10-22T00:53:55Z
2023-10-22T00:43:13Z
2023-10-23T01:00:46Z
2023-10-22T10:04:45Z, 2023-10-22T10:04:45Z
2023-07-12T17:16:19Z, 2023-07-12T17:16:19Z
2023-10-21T17:40:56Z, 2023-10-21T17:40:56Z
2023-10-22T04:21:19Z, 2023-10-22T04:21:19Z
2023-10-22T16:44:14Z, 2023-10-22T16:44:14Z
2023-07-11T15:30:28Z, 2023-07-11T15:30:28Z
2023-07-11T11:43:58Z, 2023-10-10T10:32:22Z
2023-07-11T23:44:25Z, 2023-10-10T18:27:45Z
2023-07-11T11:44:15Z, 2023-10-10T10:32:18Z
2023-07-11T11:44:15Z, 2023-09-24T05:39:37Z
2023-07-11T11:44:15Z, 2023-10-02T03:33:49Z
2023-07-12T00:33:53Z, 2023-10-11T00:23:24Z
2023-07-11T15:30:21Z, 2023-08-31T12:38:54Z

- 🟡 Seed malicious domains
- ⚪ Expanded unknown domains
- 🔵 IP addresses
- 🔴 Flagged malicious domains

Summary

- Threat actors unintentionally leave behind traces of information
  - Domains, IPs, Certificates, File Hashes, Phishing Kits
- How we can pivot on these traces to find malicious domains before they are weaponized
  - Building a knowledge graph
  - Training a GNN over the knowledge graph
- Two examples showing that our detector can proactively uncover criminal infrastructure
- Uncovered tens of thousands of high-confidence malicious domains in the last two months

Q&A

Nabeel Mohamed - mmohamednabe@paloaltonetworks.com
in linkedin.com/in/**myoosuf**

Janos Szurdi - jszurdi@paloaltonetworks.com
in linkedin.com/in/**szurdi**