Bad Neighborhoods -Learning malicious infrastructure at internet scale

Tamás Vörös – Sophos AI, Data Scientist Rich Harang – Duo Security, Data Science Senior Tech Lead Josh Saxe – Sophos AI, Chief Scientist Konstantin Berlin – Sophos AI, Director



We will talk about

A data driven ML approach to assign 'risk scores' to various blocks of the IPv4 space

Agenda

- Motivation
- Introduce/confirm structural bias
 - Review the hierarchical structure of the internet
 - Demonstrate non-uniform distribution of maliciousness over the IP space on two separate datasets
- Propose 3 architectures that support the structural bias
 - Random forest (RF)
 - Convolutional Neural Network (CNN)
 - Transformers
- o Identify ISP IP space non contiguity as a weak spot of IP-input-only models
 - Generalize knowledge over non-contiguous IP spaces, by utilizing pretrained representation learning

High level threat landscape



Malicious websites



Phishing emails



Command and Control servers



Data exfiltration



IP addresses' structure lends itself to statistical detection modeling





Datasets

- Malicious and benign websites' IP addresses
 - Benign:
 - social network infrastructures
 - search engines
 - online stores
 - video hosting providers, and so on.
 - Malicious:
 - malware repositories
 - phishing sites
 - callhomes
 - Total size: 455248

- Spam associated and non-spam associated IP addresses
 - Benign:
 - static reputation list containing IPs provided by organizations who have had prompt reactions to any kind of abuse of their mail servers. (DNSWL)
 - Malicious:
 - hijacked PCs worms/viruses with builtin spam engines. (Spamhaus)
 - Total size: 432792

Hilbert curves - concept

- 1D and 2D space mapping
- Preserves some notion of closeness or locality
- Two data points that are close to each other in one-dimensional space are usually close to each other after the transformation.

https://blog.benjojo.co.uk/post/scan-ping-the-internet-hilbert-curve



Hilbert curves



(a) Web ip addresses

| Reserved | PNIC | APNIC | wise lighting | | Cara Mater Lispany 11.1.1.05 | wase liquiton a Line | DDN-RVN | | | | | | | |
|--------------------|----------------------|---------------------------|--|--------------------------------------|--|-------------------------|--------------------------|-------------|------------|--------------------------|----------------------------|--------|------------------------------|--------------------|
| Netsia Sagartsar R | IPE | negas fataroin | A <u>T&</u> T | Apple | 2 mesas ingirtini n reas * | ARIN | DISA | Multicast | | | Posorvod | | | |
| Level3 | ne inperton 18935 | Level3 | ne hot krandes term in 1189 | DISA | DISA | ARIN | Where d stress 0.1455 | | | | Reserveu | | | |
| | Army | mejaa sagattijat 13335 | Reserved | RIPE | DSI-North | APNIC | DISA | | | | | | | |
| | () 31116 | where the manual | Daimler AG | mener ingertran | | APNIC | RIPE | APNIC | | | | | Water Sector | AFRINIC |
| APNIC | USPS | Do <u>D</u> NIC | anton belation | No kolas obstat ar tota B, 1 1 MB | , and a second s | APNIC | PSIMet, Inc. | RIPEARIN | APNIC | APNIC | LACNIC | ARIN | tergen engisteler Terrige | |
| APN | IC | ARIŅ | ality and a second seco | RIPE | wine region | 901300 Registrati | AFRINIC | US-DOD | APNIC ARIN | | | TAL | RI | PE |
| ARINR | IPE | APNIC | Prudential | WING REPORT | 8.1145 | antes settimes | APNEC | RIPE | | | AR | ΠN | RIPE | Ware approach |
| Reserved | | | | | | | | LAC | NIC | N. | ncasa nacional Scientes | | | |
| ARIN | | | APNIC APN | | APNIC | APNIC | 19 - P | | | RIPE ARIN LACIT | | | | |
| RIPE ARIN ARIN | | | ADNITC | | | | | đí. | APN | IIC | RIPE | LACNIC | | |
| | | | APNIC | | | | | | | LACNIC APNIC LACNIC RIPE | | | | |
| | 40 | PE. | | | TN | AP | VIC | • Various R | egisti | ries | 1 6 | | ARIN | APNIC |
| | D.T | | | AKIN | | RIPE ARIN | | | | 1003 1003 | | | | Wilson Begalatetet |
| | КŢ | | | ARIN APNIC ARIN ARIN | | | | | | Various Registries | | | | |
| | | 113.94 | | APNIC | AFRINIC | AFRINIC | APNIC | | | - 14 - 15 | • • | | Various R | egistries |

(b) Spam ip addresses

https://github.com/measurement-factory/ipv4-heatmap

SOPHOS

IP Address Modeling

integer

binary

1751389497 01101000.01100100.00010101.00111001 104.100.21.



CNN

...

. . .

 104.
 100
 21
 57
 Prefix

 01101000.01100100.00010101.00111001
 104.100.21.57/32

 01101000.01100100.00010101.00111000
 104.100.21.57/31

 01101000.01100100.00010101.00111000
 104.100.21.57/30

| 01101000.01100100.00010101.00111000 | 104.100.21.57/ <mark>20</mark> |
|-------------------------------------|--------------------------------|
| 01101000.01100100.00010101.00111000 | 104.100.21.57/ <mark>19</mark> |

...

...

0100000.0000000.0000000.0000000 104.100.21.57/2 0000000.0000000.0000000.0000000 104.100.21.57/1 104.100.21.57 Transformer

standard decimal dotted





Results





Utilizing ISP

Fixed ISP: Cloudflare

| Prefix | # Training data | # Test data |
|-----------------|-----------------|-------------|
| 103.21.244.0/22 | 5000 | 0 |
| 23.15.11.0/24 | 3 | 5000 |



Let's add pretrained a component





Final results





Conclusion



