

View From Above: Exploring the Malware Ecosystem From the Upper DNS Hierarchy

Aaron Faulkenberry, Athanasios Avgetidis, Zane Ma, Omar Alrawi, Charles Lever, Panagiotis Kintis, Fabian Monroe, Angelos D. Keromytis, Manos Antonakakis

Motivation

- Malware is a pervasive and growing threat.
- Malware network behavior is being utilized to combat abuse.
 - Malware network artifacts enable practical defenses.
 - Network behavior is critical to understand where, when and how fast it spreads.
- Unfortunately, previous work on malware communications has been largely limited by partial visibility of the global malware threat.

Contemporary Malware Communication Vantage Points

Host level:

- AV client host logs
- Sandboxes
- Honeypots

Network level:

- Sinkholes
- Malicious Infrastructure Takeover
- Local recursive DNS and network data
- TLD DNS data

- Limitations:
 - Partial global malware threat visibility.
 - Partial malware lifecycle visibility.

Is there a dataset that provides high guarantees in these two qualitative aspects?

Our Vantage Point - Authoritative DNS

Upper DNS

Root



TLD



Authoritative



Recursive DNS

Open Recursive



Local Recursive

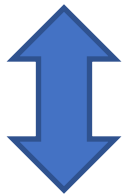


Clients

Users



VPNs/
Proxies



Motivation – Revisiting Common Wisdom

- How is malware distributed among different geographies and *industries*?
- Does malware exhibit different network behaviors on each phase of its lifecycle?
- How effective are different vantage points when studying malware communications?

Enabling Global Longitudinal Analysis

- Popular Registrar Authoritative DNS records (2017-02 to 2021-06)
- Malware Dynamic Executions (30M executions 2018-01 to 2021-04)
- Network Prefix Industry Labeling Dataset
- Supplementary Datasets
 - IP Whois
 - Virus Total
 - Malpedia

Methodology – Malware Dataset



30M
Malware
Sandbox
Executions



- Identify registrar domains & malware
- Remove benign and popular domains

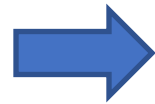


39.7B DNS requests
12,212 domains
245,010 samples

Methodology – Malware Sample Labeling

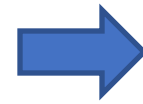


245,010
Malware
Samples



VIRUSTOTAL

Get AV
Malware
Labels



malpedia Malware
Aliases



AVCLASS2




161,322
Labeled
Malware
Samples

Enabling Global Longitudinal Analysis

- Diverse malware:
 - 202 malware families of multiple types
 - Hosted in 151 Countries
- Epidemiological view of connections:
 - 40,937 querying ASes
 - Global visibility
- Lifecycle Analysis:
 - 4-year Time Period
 - All phases of malware domain lifecycle

Malware	Type	Domains	Samples	Server CCs	Client CCs
darkkomet	RAT	3,578	16,441	140	232
njrat	RAT	1,924	10,596	129	229
cybergate	RAT	1,181	2,546	100	219
xtrat	RAT	946	2,801	89	222
bifrose	RAT	700	1,432	62	211
razy	Stealer	667	1,139	110	225
remcos	RAT	563	39,279	103	221
nanocore	RAT	501	2,112	116	227
ponystealer	Stealer	450	4,891	93	222

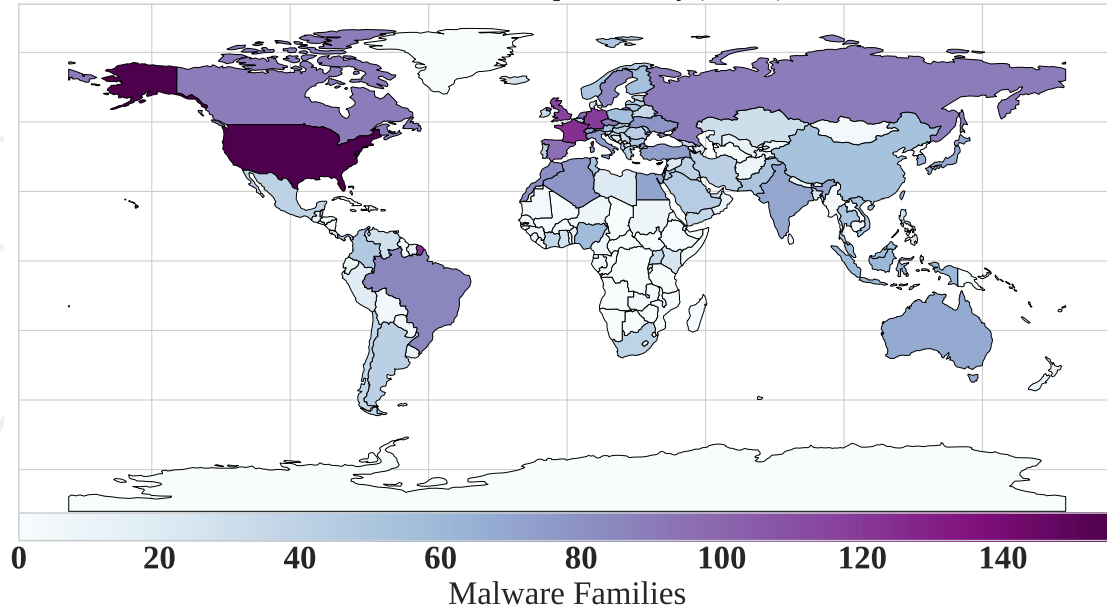


How is malware distributed among different geographies and industries?

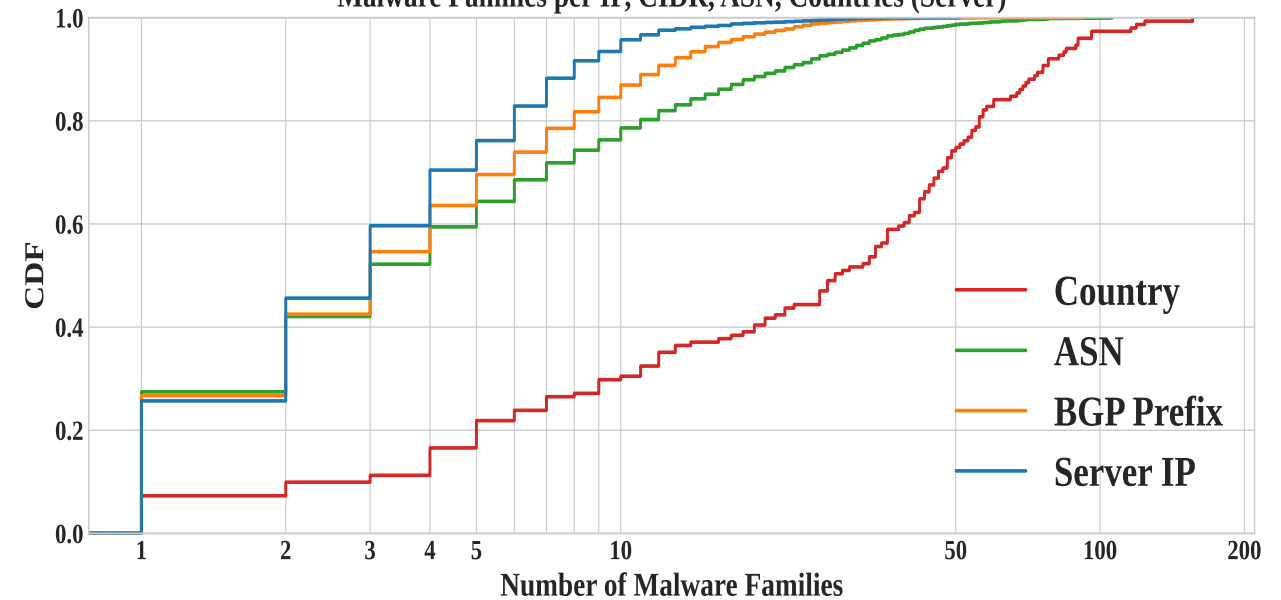


Hosting Infrastructure

Malware Families per Country (Server)



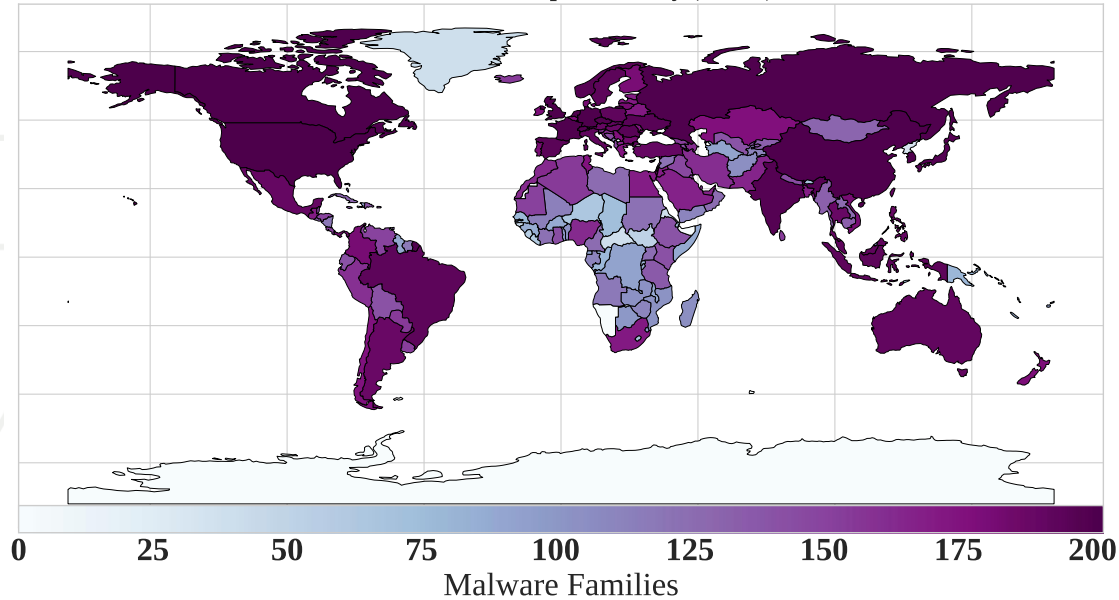
Malware Families per IP, CIDR, ASN, Countries (Server)



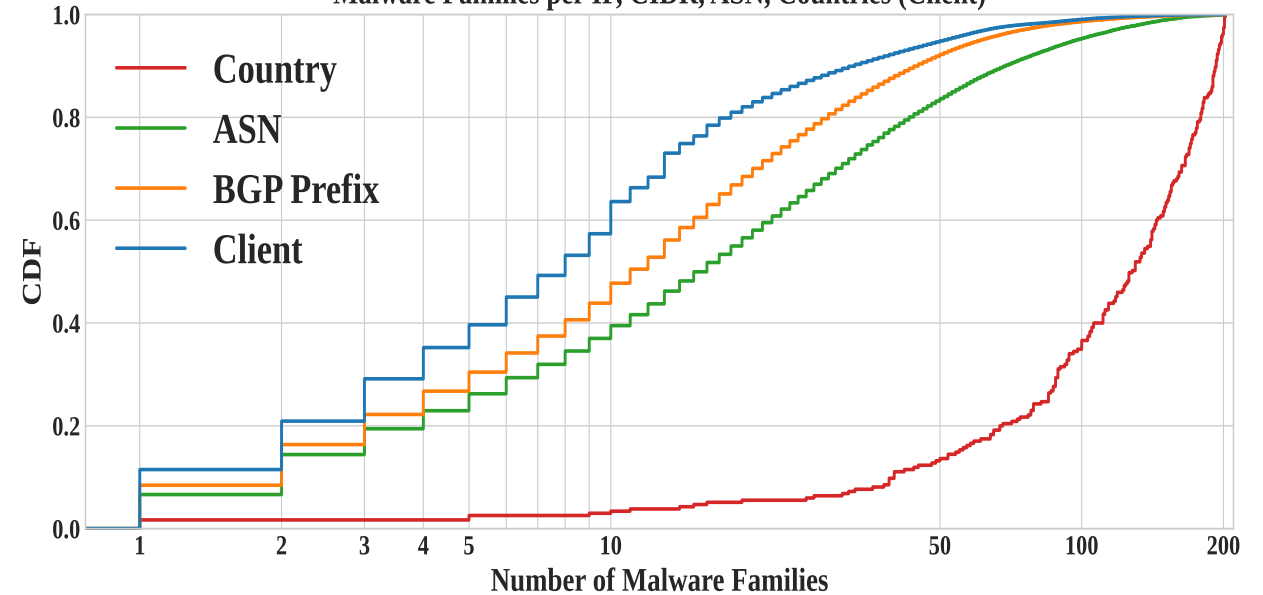
- Malware is largely aggregated in countries with large-hosting providers.
- Most malware families are hosted in multiple countries with 74.3% of malware hosting IPs are associated with 2 or more malware families.
- Our results agree with prior work on different points on the DNS hierarchy.

Client Analysis

Malware Families per Country (Client)



Malware Families per IP, CIDR, ASN, Countries (Client)



- A significant amount of different malware families plague nearly every country.
- Most networks and countries are not strongly correlated with a single malware family.

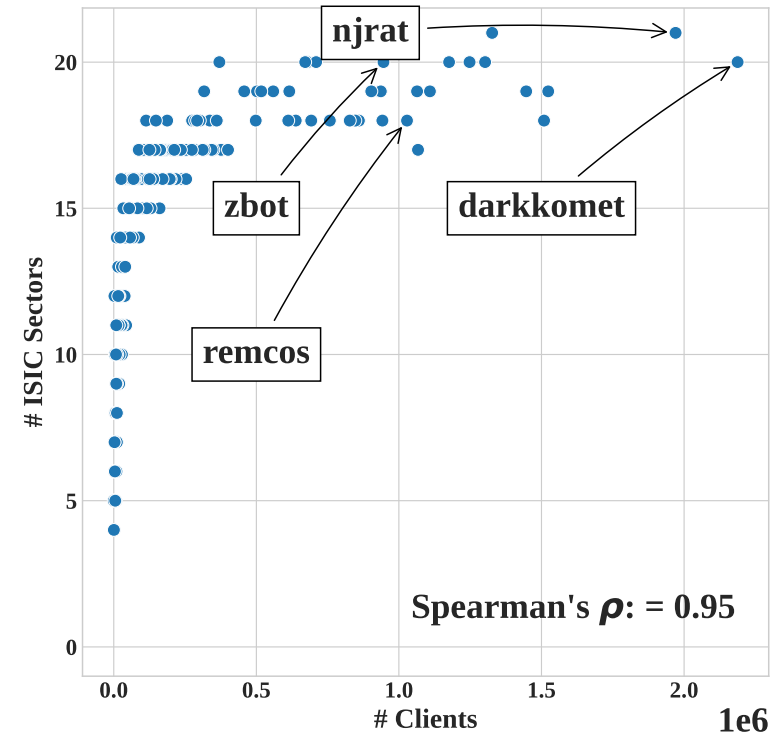
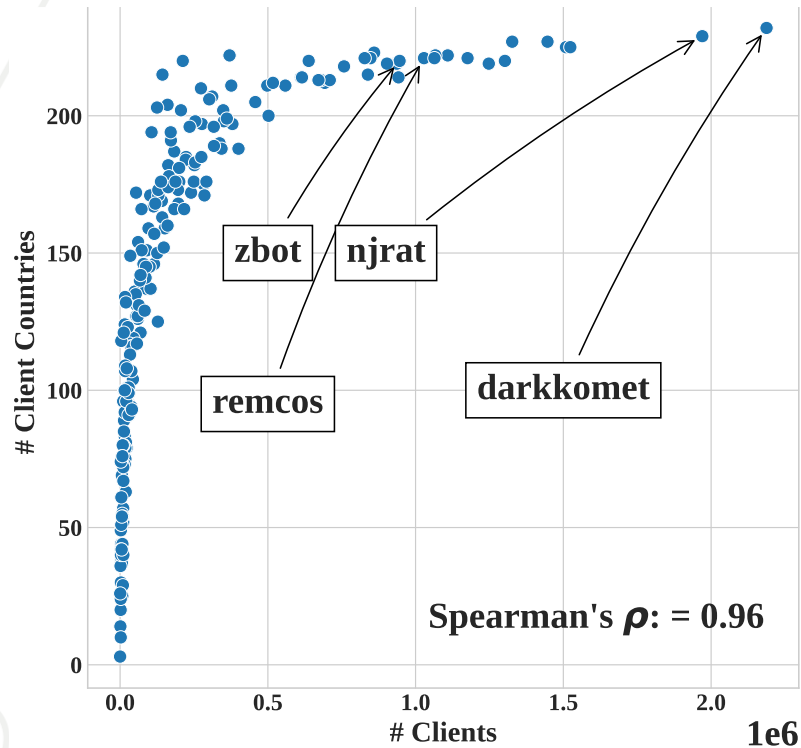
Industry Analysis

ISIC Section	Clients	Malware Families
Information & Communication	3,108,546	202
Wholesale & Retail Trade	567,729	202
Education	29,741	201
Professional, Scientific & Technical Activities	11,576	196
Manufacturing	4,837	192
Government, Defence	4,697	178
Financial & Insurance Activities	3,670	183
Human Health & Social Work Activities	3,785	172
Accommodation & Food Service Activities	2,785	148
Transportation and Storage	624	155
Arts, Entertainment & Recreation	421	140
Electricity, Gas, Steam & A/C Supply	333	127
Administrative and Support Service Activities	199	141
Extraterritorial Organizations and Bodies	164	120
Other Service Activities	149	149
Real Estate Activities	96	86
Construction	74	38
Mining and Quarrying	17	23
Agriculture, Forestry and Fishing	5	18
Water Supply, Sewerage.	5	8


- Most industry sectors are impacted by over the half of the malware families.
- 72.7% of the malware families were found in more than 10 different industry sectors.
- AV client host log previous work [1] : 37% of malware families only seen in 1 sector.

[1] Kotzias, P., Bilge, L., Vervier, P. A., & Caballero, J. (2019, February). Mind Your Own Business: A Longitudinal Study of Threats and Vulnerabilities in Enterprises. In *NDSS*.

Industry Analysis



- Higher number of clients querying a malware family correlate strongly with more countries and industries.
- As malware families grow, so does the diversity of their victims.

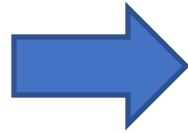


Does malware exhibit different network behaviors on each phase of its lifecycle?

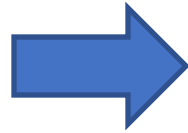


Malicious Domain Lifecycle

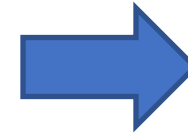
Registration



Detection



Takedown/
Expiration



Post
Takedown/
Expiration

Malware Lifecycle: Registration to Detection

Domains	Registration to Detection						
	ASNs	(%)	ASCCs	(%)	Sectors	(%)	Days
10%	0	(0.00)	0	(0.00)	0	(0.00)	0
25%	0	(0.00)	0	(0.00)	0	(0.00)	1
50%	6	(2.63)	3	(7.69)	1	(20.0)	4
75%	22	(10.0)	9	(21.8)	2	(37.5)	19
90%	54	(23.4)	18	(40.0)	3	(57.1)	79
max	2,243	(96.6)	136	(100)	14	(100)	1,154

Registration to Detection	
ASNAME	Domains
AMAZON-AES	772
CORBINA-AS PJSC "Vimpelcom"	756
GOOGLE	645
LEVEL3	611
AMAZON-02	602

- Shortest window with < 20 days for 75% of the domains.
- Only a small fraction of all observed ASNs and countries are observed in this window.
- Big recursives and DNS networks first appear in this phase.

Malware Lifecycle: Detection to Expiration/Takedown

Domains	Detection to Expiration/Takedown						
	ASNs	(%)	ASCCs	(%)	Sectors	(%)	Days
10%	19	(11.6)	5	(11.5)	0	(00.0)	1
25%	59	(36.2)	15	(42.9)	2	(37.5)	23
50%	101	(54.0)	26	(62.5)	4	(62.5)	30
75%	164	(70.4)	37	(79.0)	5	(80.0)	100
90%	369	(86.5)	54	(90.4)	7	(100)	419
max	11,650	(100)	187	(100)	15	(100)	1,661

Detection to Expiration/Takedown	
ASNAME	Domains
WINTEK-CORP	1,745
GEORGIA-TECH	1,738
OVH OVH SAS	1,662
MFENET	1,649
PAN0001	1,641

- Most ASNs and industries observed in this window compared to the others.
- We first observe many scanners and AV related networks that can artificially inflate the infected population counts.

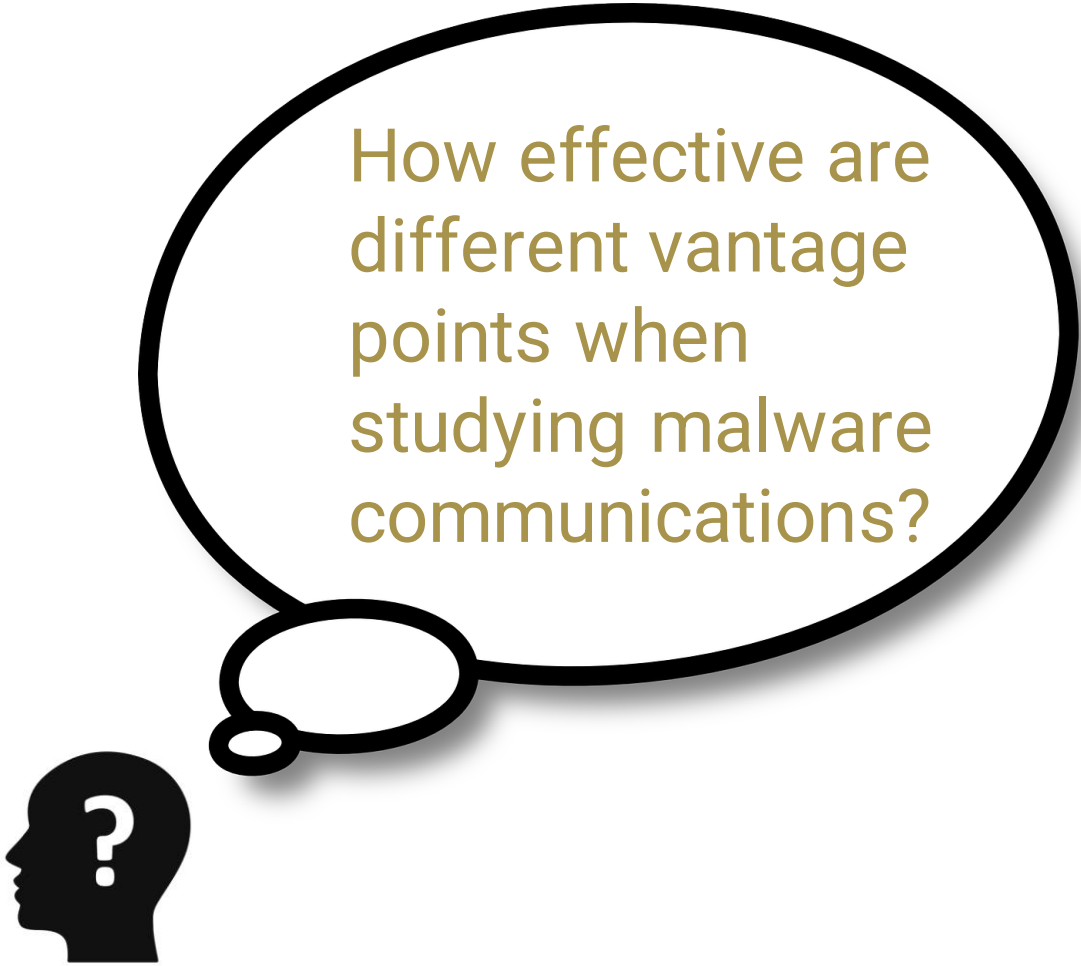
Malware Lifecycle: Post Expiration/Takedown

Domains	Post Expiration/Takedown						
	ASNs	(%)	ASCCs	(%)	Sectors	(%)	Days
10%	10	(6.80)	1	(1.02)	0	(0.00)	238
25%	36	(21.9)	3	(9.09)	0	(0.00)	526
50%	76	(38.6)	10	(22.7)	1	(16.7)	963
75%	132	(55.3)	18	(40.0)	2	(30.0)	1,180
90%	229	(71.3)	28	(58.6)	3	(50.0)	1,256
max	4,644	(100)	95	(100)	10	(100)	1,558

Post Expiration/Takedown	
ASNAME	Domains
CNNIC-ALIBABA-US-NET-AP Alibaba (US) Technology Co., Ltd.	1,387
CNIX-AP China Networks Inter-Exchange	1,363
CHINATELECOM-TIANJIN Tianjij,300000	1,226
InterConnect ML Consultancy	1,114
FSOL-AS F-Solutions Oy	1,084

- We observe a long tail of new ASNs that will first query a malicious domain only post expiration or takedown.
 - This could be attributed to network mobility of infected clients, new infections and/or new scanner networks.
- Previous work [2] has suggested that 99% of the traffic in this window is not real victim related.

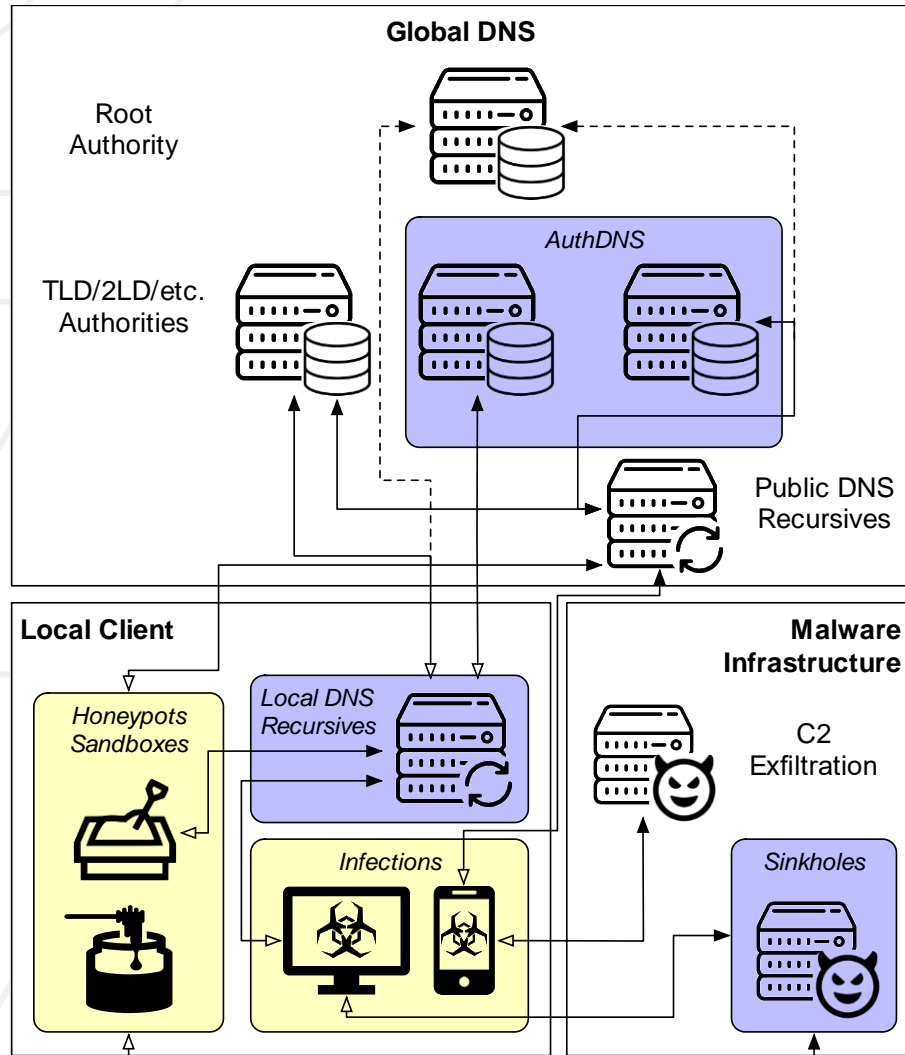
[2] Rezaeirad, M., Farinholt, B., Dharmdasani, H., Pearce, P., Levchenko, K., & McCoy, D. (2018). {Schrödinger's}{RAT}: Profiling the Stakeholders in the Remote Access Trojan Ecosystem. In *27th USENIX Security Symposium (USENIX Security 18)*



How effective are
different vantage
points when
studying malware
communications?



Vantage Point Comparison



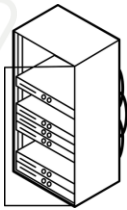
- Three Measurement Planes:
 - Global DNS
 - Local Client
 - Malware Infrastructure
- Four Qualitative Perspectives:
 - Infection Visibility
 - Infection Precision
 - Client Granularity
 - Infrastructure Visibility

Vantage Point Comparison Highlights



Infection Visibility

- AuthDNS provides high guarantees both geographically and temporally compared to host and local network datasets.



Infrastructure Visibility

- AuthDNS is the richest dataset at the domain level, however recursive DNS also provides good visibility.



Infection Precision

- DNS datasets and sinkholes are noisier compared to AV client host logs.



Client Granularity

- Big recursives, NAT-ing, VPNs and proxies limit the ability of global DNS and network datasets compared to local.

Conclusions

- While malware is largely hosted in specific geographies it affects many different countries and industries.
- Malicious domains DO exhibit different network behavior in each phase of their lifecycle.
- Client infection estimation need to account for non-victim traffic on every phase of the malware lifecycle.
- We provide independent verification of previous studies on malware epidemiology.

Questions

