Domain Abuse Activity Reporting System

Samaneh Tajalizadehkhoob, PhD
Security Stability & Resiliency Specialist
ICANN Office of CTO (OCTO)
ICANN66 - Montreal

What is DAAR?

- Source: CZDS and other gTLD zone files + third party Reputation Block Lists (RBLs)
    - Takes security threat data from different RBLs and maps that to gTLDs based on information in gTLD zone files.
- DAAR datasets: January 2018 - onwards
    - On avg. data on 1200 gTLDs
    - On avg. data on 195 million domain names
    - Security threat data on Phishing, Malware, Spam and Botnet C&C
- Data collection frequency
    - gTLD Zone file data: Once a day
    - RBL data: Avg. once an hour (RBL specific)

Why Different vs. Existing Work?

- Different source of security threat data
- Historical data collection and analytics
- Monthly stats
- Comprehensive overview of the gTLD space
- Reproducible and replicable methodology
- Continuous system/data improvements

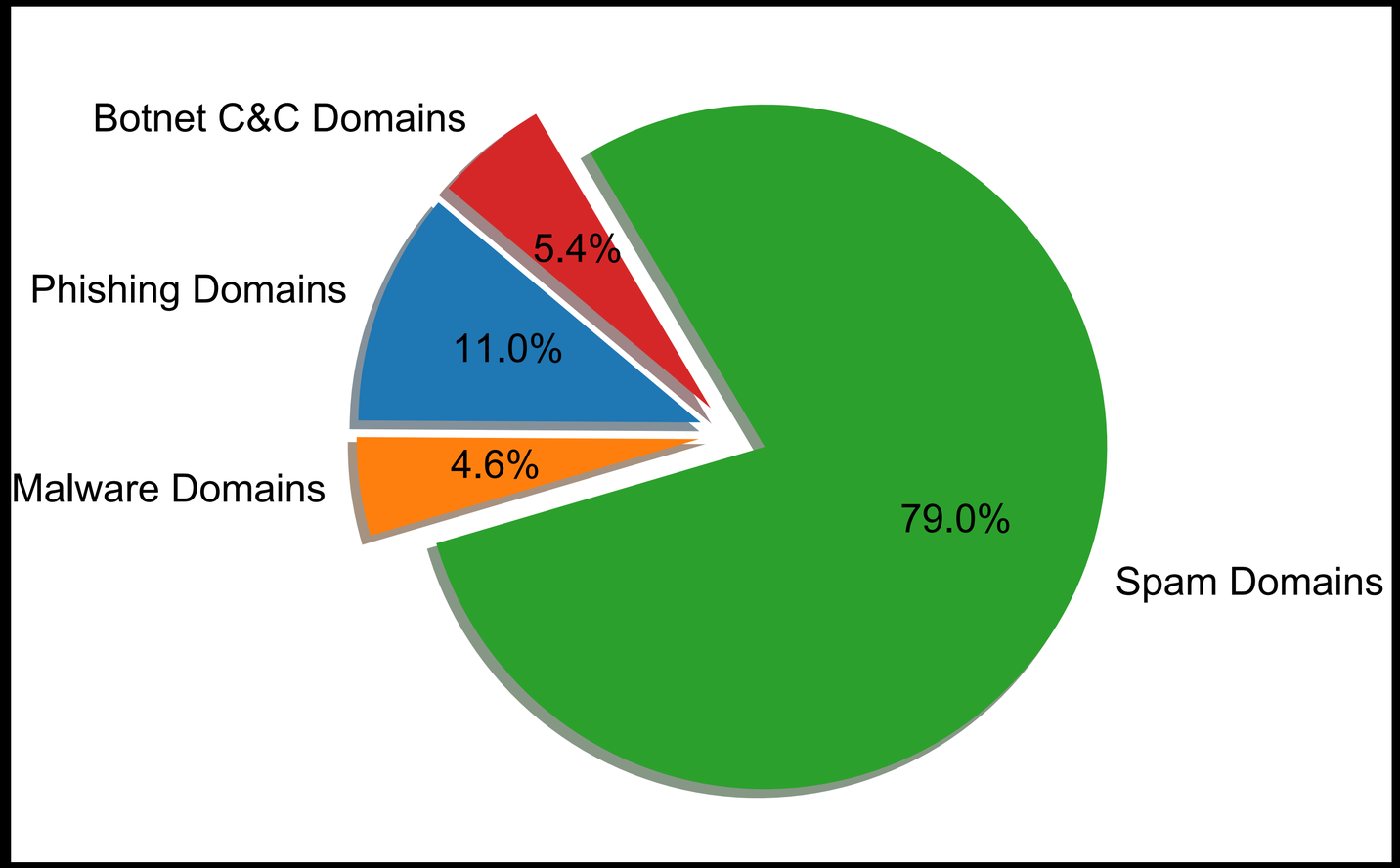
Primary DAAR Example Metrics

- Count of domains in the zone file per gTLD per day
- Count of domains listed per gTLD per day
- Count of newly listed domains per gTLD
- Count of domains listed for the first time per gTLD (during the observation period)

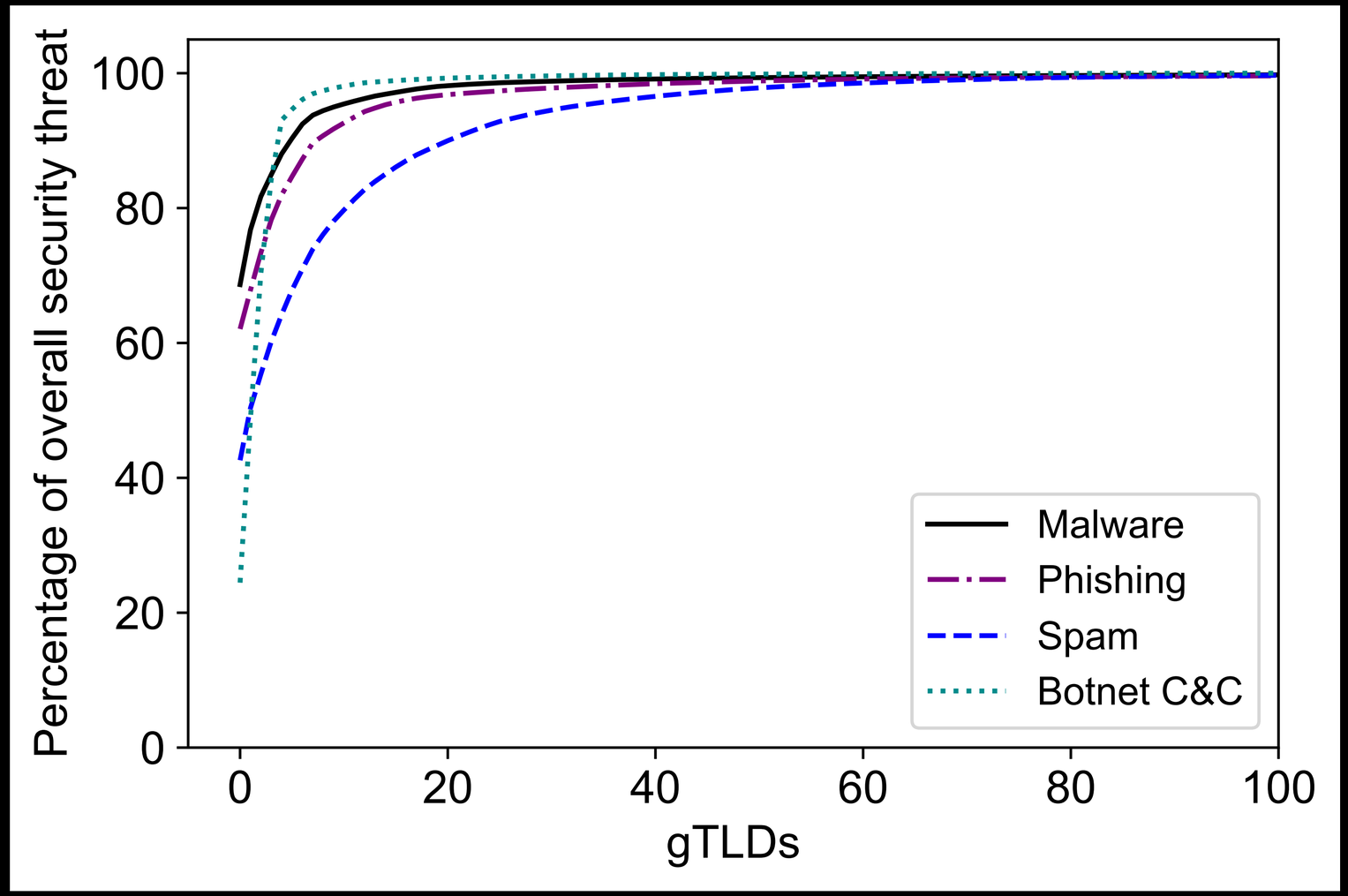
- Percentage of gTLD domain listed
- Percentage of gTLD Phishing domains
- Percentage of gTLD Malware domains
- Percentage of gTLD Spam domains
- Percentage of gTLD Botnet C&C domains
- Time-series changes of the above metrics
- …

Example Analytics

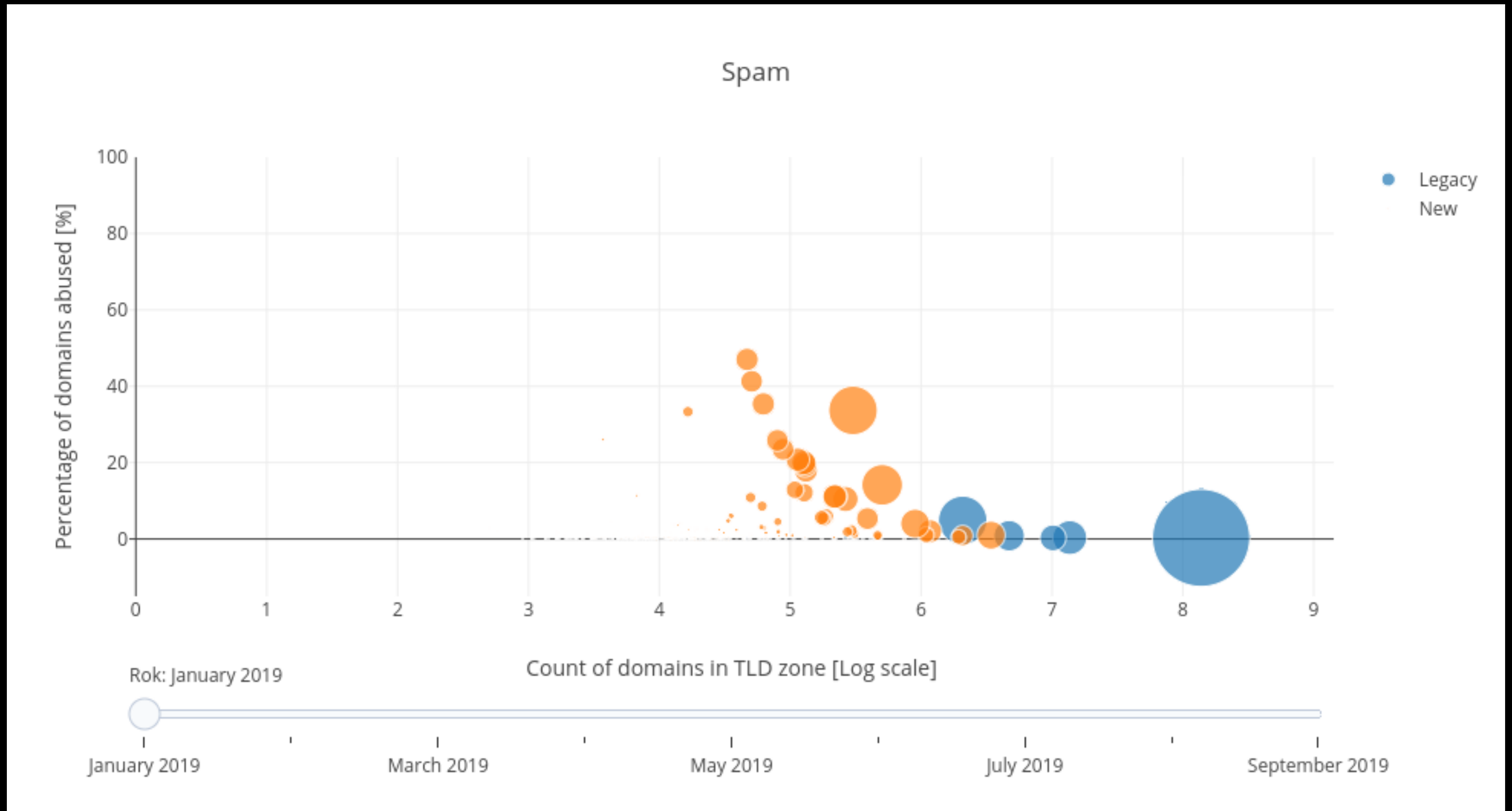
Security Threat Type Breakdown



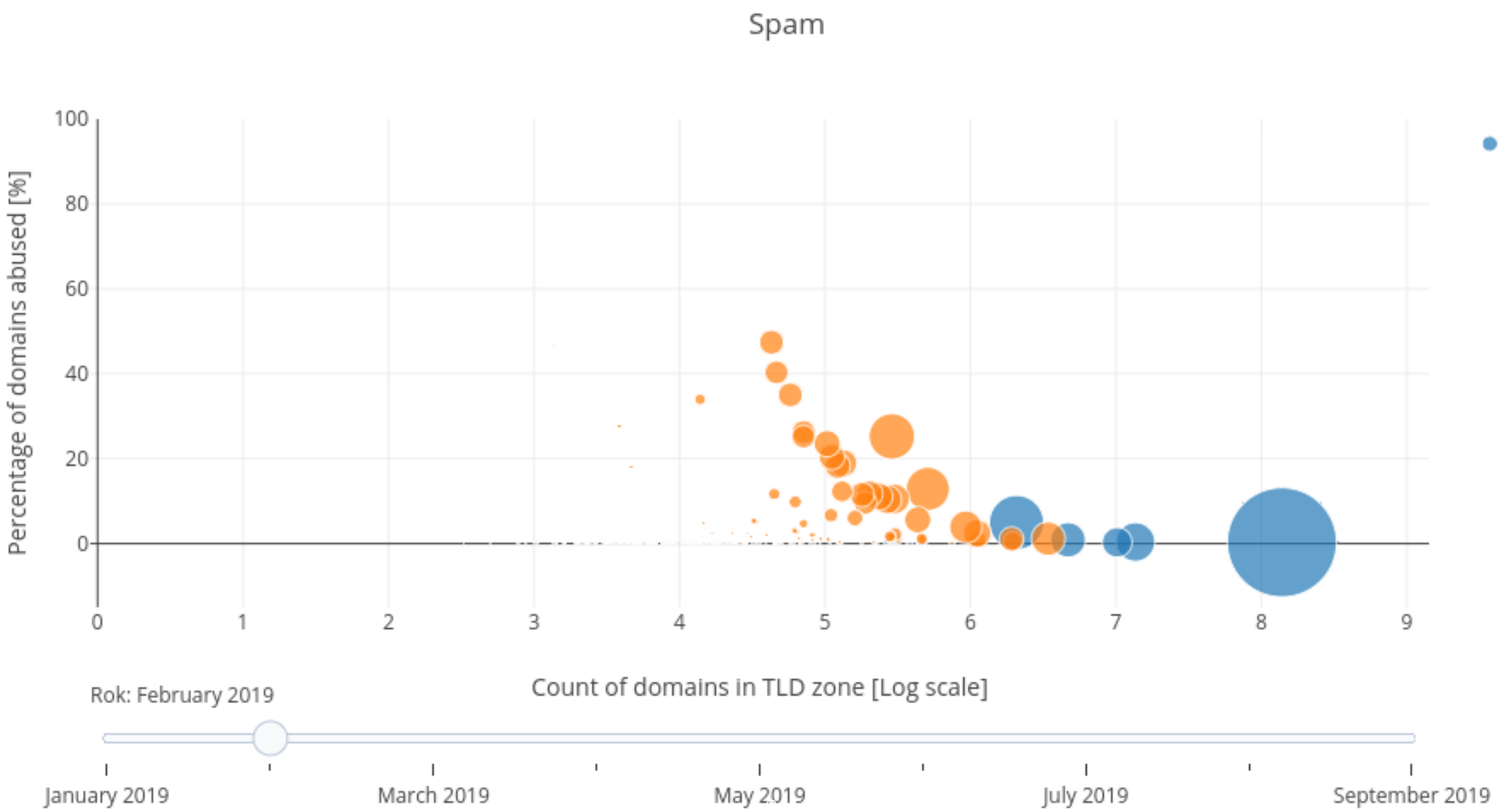
Distribution of Security Threats for top-100 Most Abused gTLDs



Timeseries Metrics



Timeseries Metrics



Improvement Process

What Feedback We Have Received

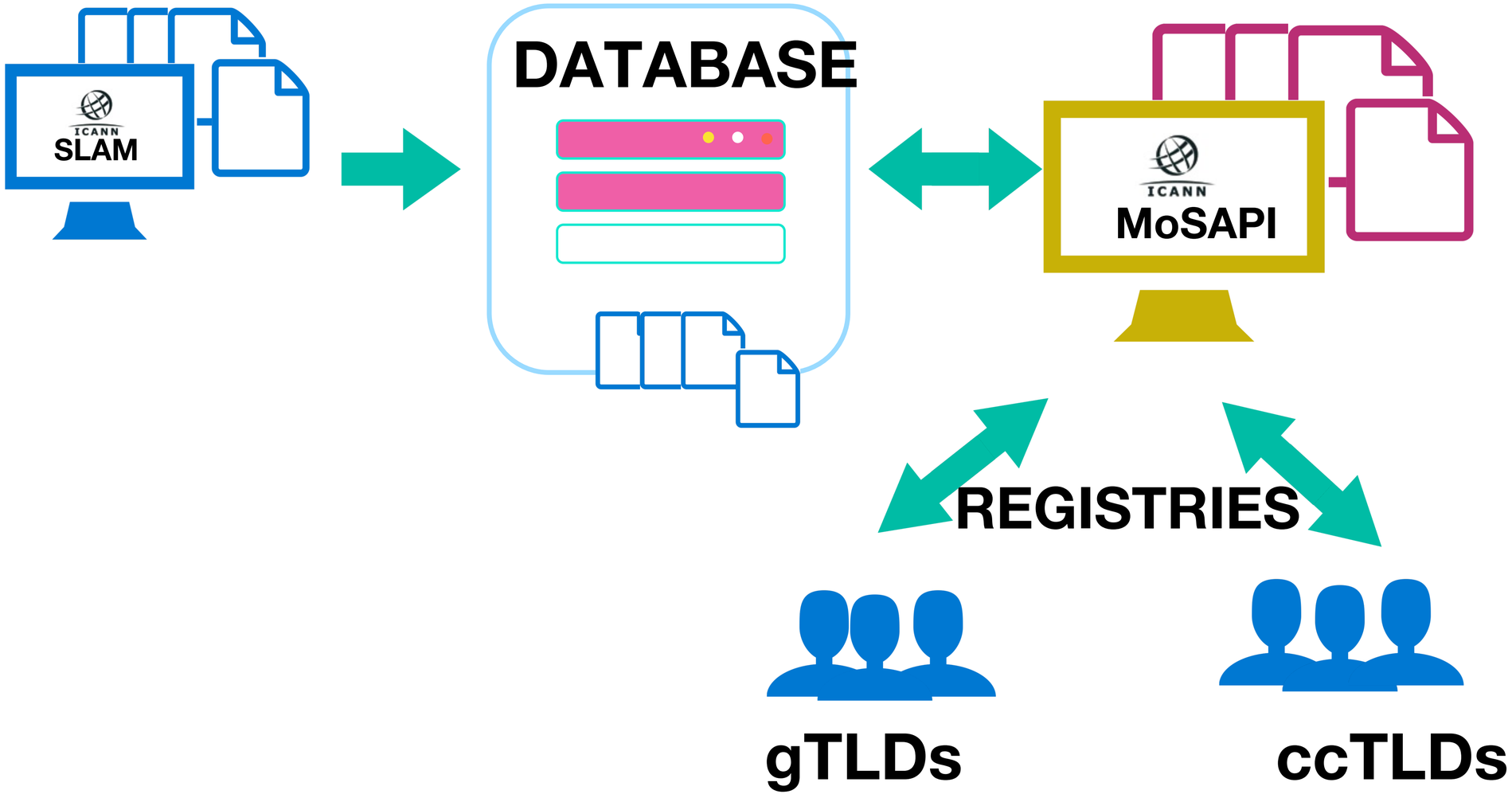
- Requests for more transparency on DAAR progress
- Re-aggregating the DAAR data
- Consistency in using the terms “Abuse” vs “Security Threat”
- Adding threat domain time-to-live data
- Adding ccTLDs to DAAR
- Adding registrar metrics to DAAR
- Publishing DAAR detailed data
- Distinguishing between maliciously registered domains and compromised one
- Better articulation of DAAR’s goal in monthly reports and documentation

What We Have Considered

Sharing DAAR data with Registries

What is MoSAPI?

REST API that allows Registries to retrieve information collected by the SLAM.



Getting DAAR data

Additional methods to get DAAR data:

- DAAR data for the specified date in the URL <base_url>/daar/report/<YYYY>-<MM>-<DD>

- List of dates for which DAAR data is accessible <base_url>/daar/reports?startDate=<startDate>&endDate=<endDate>

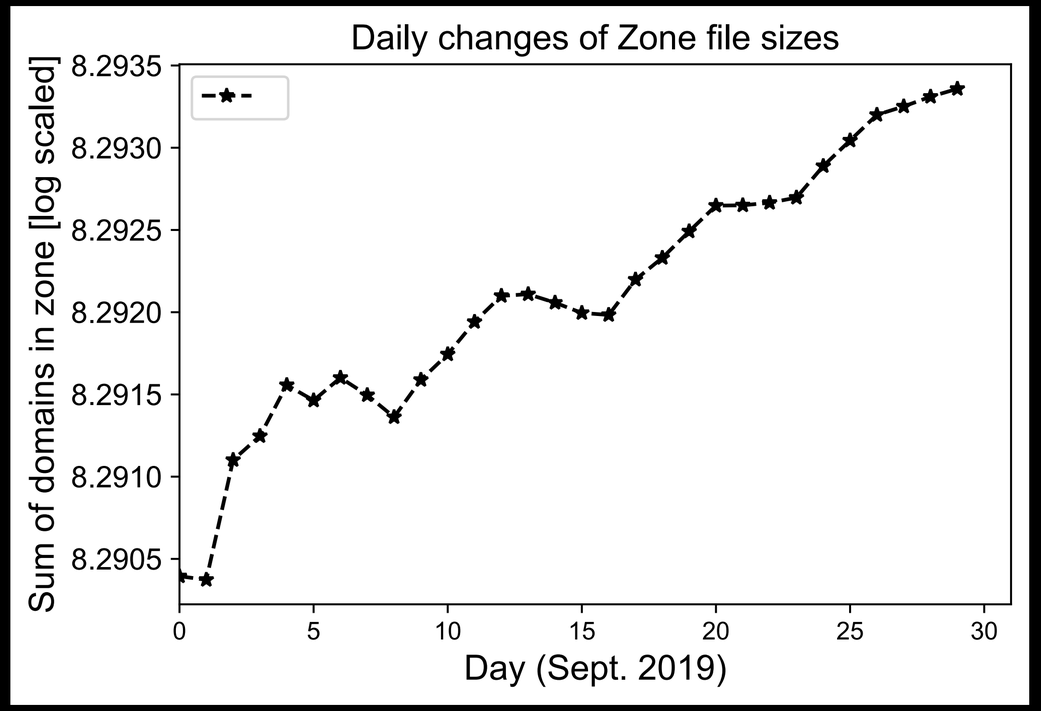
Sharing DAAR Data with Registries

Up to September 2019, only 21 registry operators were pulling this data.

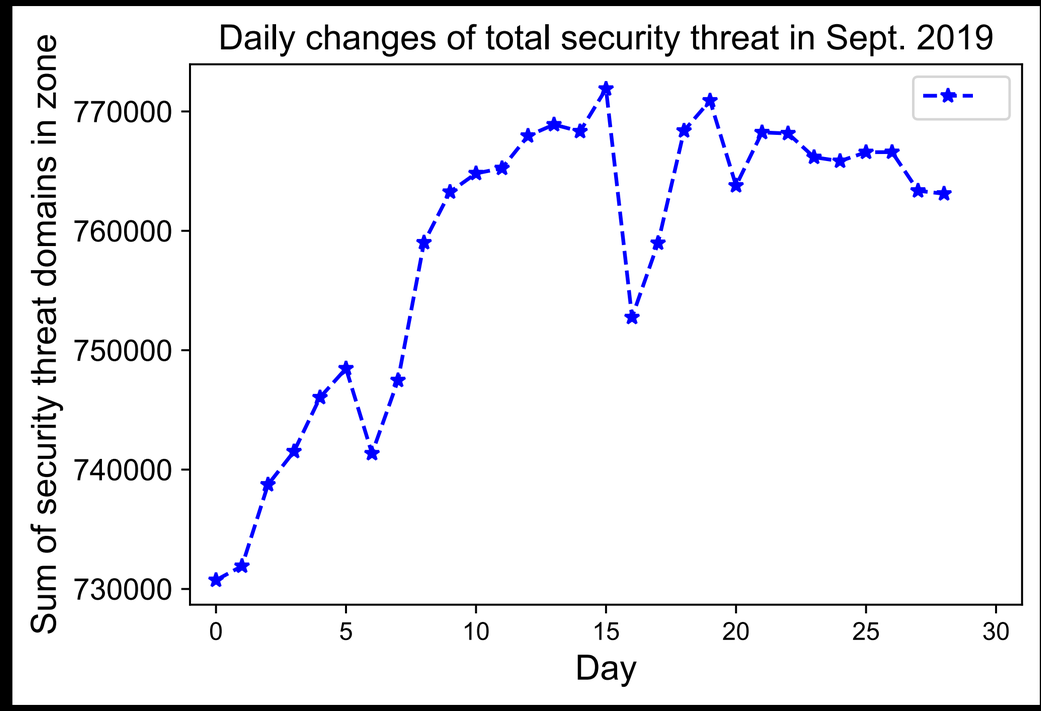
For more info contact

globalSupport@icann.org

Snapshot Data vs Monthly Aggregates



Mean= 8.29
Median= 8.29
Std Dev. = 5.55



Mean = 758522.44
Median = 763736.0
Std Dev. = 12247.01

Newer Metrics & Analytics

Inferential Analysis of Potential Relationship With Abuse Drivers

Example Drivers

- Size of a TLD can be used as an
  - **explanatory factor** for the concentrations of security threat domains
  - **“attack surface”** size for cybercriminals.

What other factors can drive security threat concentrations?

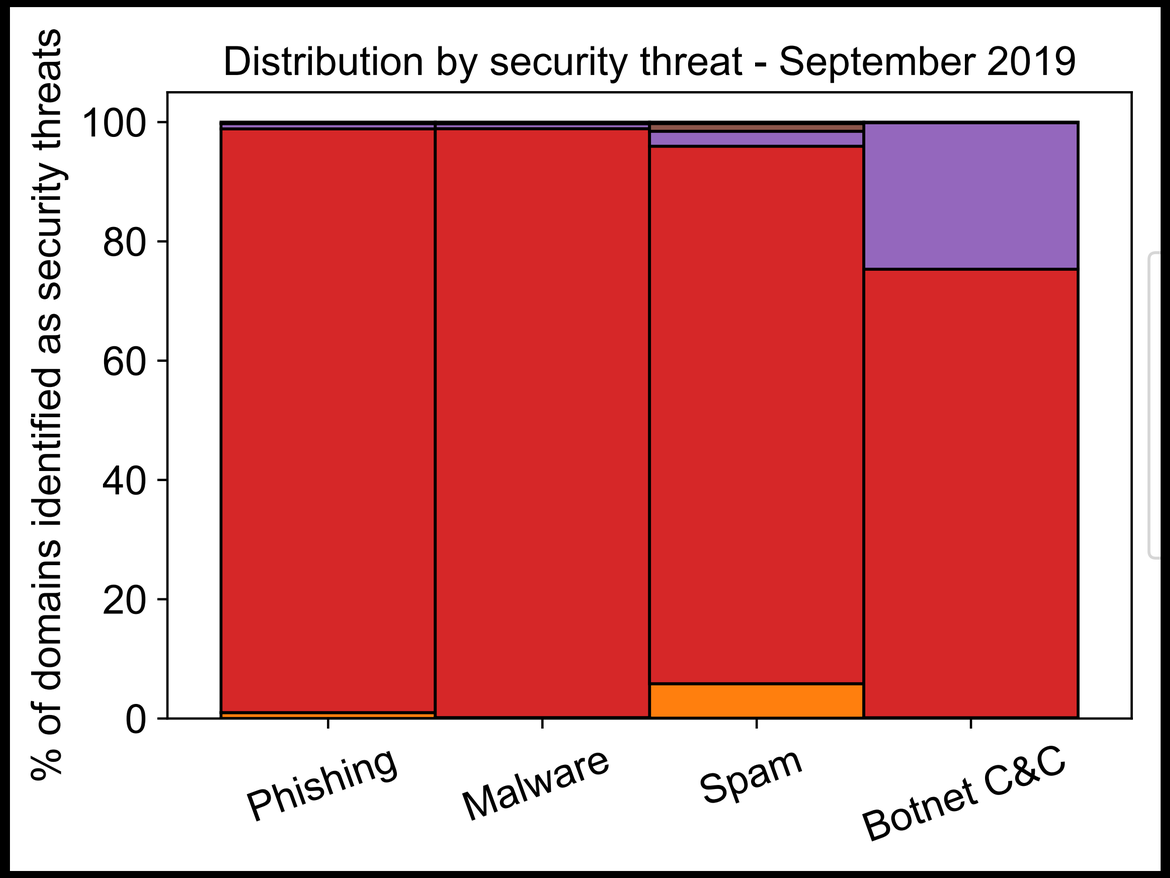
Real World Variables

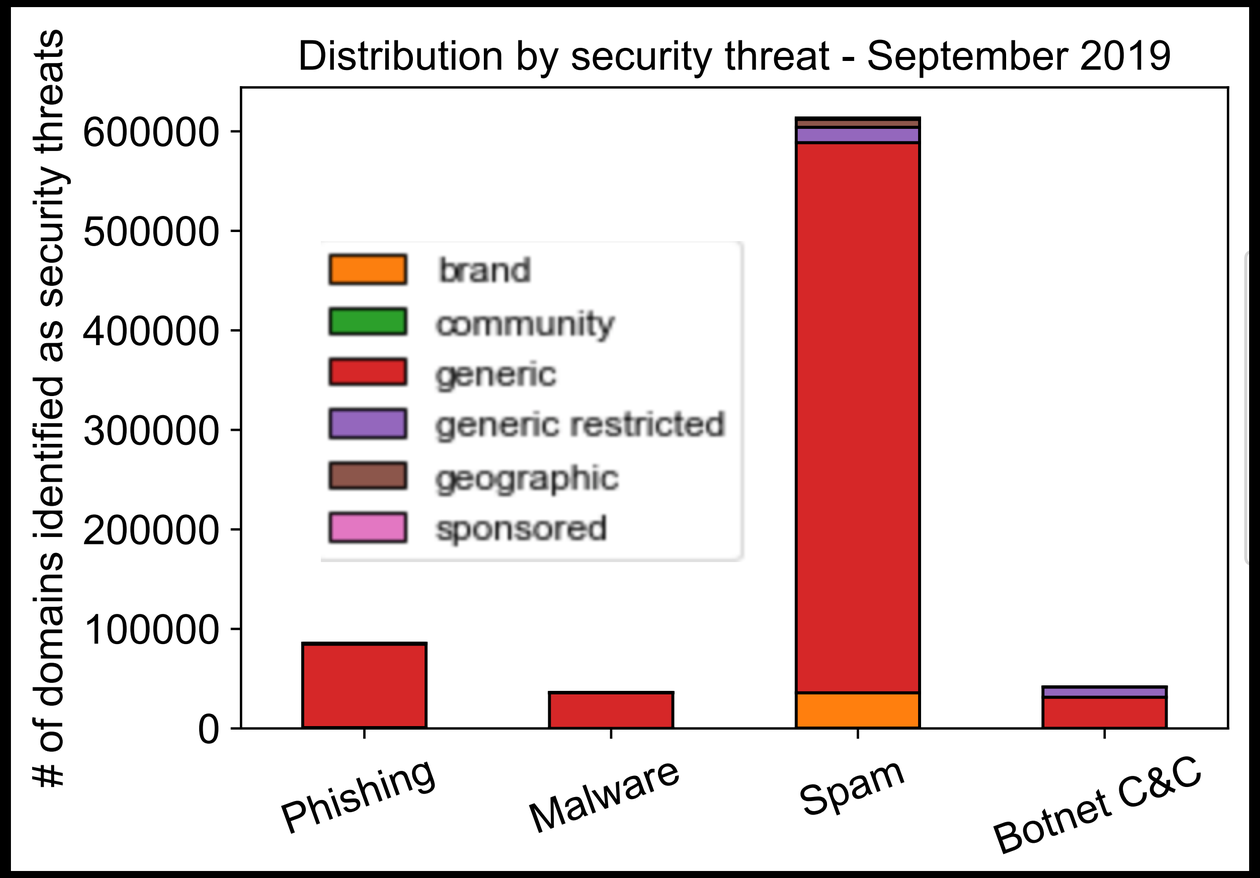
- More than one factor might drive abuse within TLDs
- Therefore factors should be taken into account at the same time

Modeling Security Threat Concentrations

	Dependent Variable: Median security threat counts			
	Phishing domains	Malware domains	Spam domains	Botnet C&C domains
	(1)	(2)	(3)	(4)
Size of domains in zone	2.320***	2.169***	2.077***	1.101***
	(0.074)	(0.119)	(0.057)	(0.034)
Spec 13 TLD	0.298	−0.866	2.119***	−14.289
	(2.980)	(11.423)	(0.490)	(666.315)
Restriction code	−0.240*	−0.345	0.060	−0.048
	(0.120)	(0.284)	(0.067)	(0.194)
TLD Type: New	1.487***	0.622	1.122***	−3.510***
	(0.336)	(0.700)	(0.236)	(0.439)
Constant	−8.306***	−7.909***	−4.395***	0.642*
	(0.611)	(1.010)	(0.463)	(0.322)
Observations	1,202	1,202	1,202	1,202
Note:				*p<0.05; **p<0.01; ***p<0.001
				Standard errors in brackets

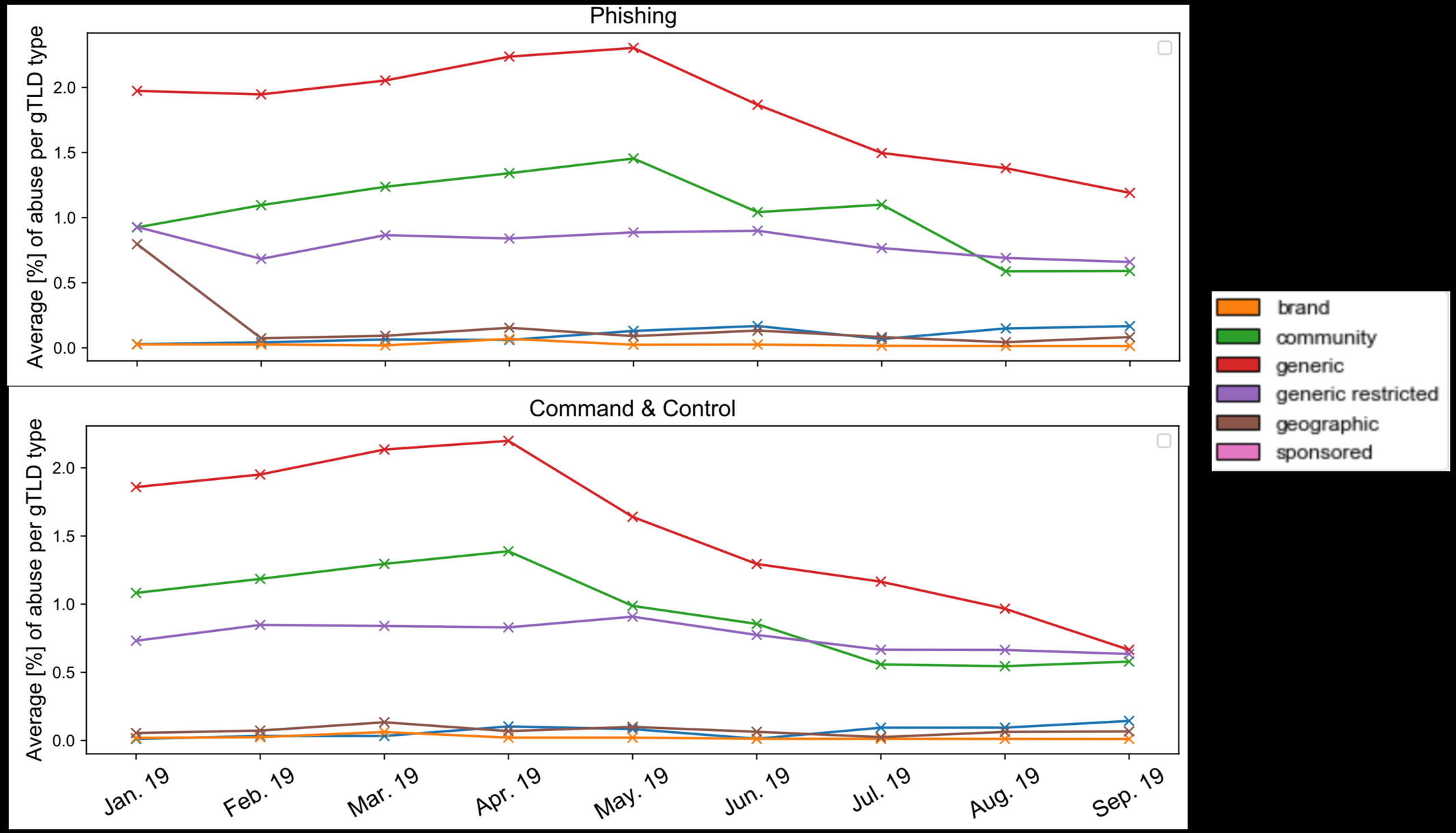
Distribution of Domains by Restriction Types





*Source: https://en.wikipedia.org/wiki/List_of_Internet_top-level_domains

Distribution of Restriction Type over Threats over Time



On More Transparency

- Mailing list for for the ICANN Org and the ICANN community to work together on topics and projects related to DNS abuse measurements, data, tools, and analysis such as the Domain Abuse Activity Reporting (DAAR) as well as any other projects and topics in this area.

- **dns-abuse-measurements@icann.org**

Adding ccTLD Metrics to DAAR

WE ARE READY! Please Apply!

- Many ccTLDs have asked us how they can take part in DAAR
- As of today, ccTLDs will be able to provide their zone files for inclusion in DAAR
- They will be able to pull their own aggregated DAAR data via MOSAPI
- We encourage you to come forward and participate

ICANN

Adding ccTLD Metrics to DAAR

How?

- Send an email to globalSupport@icann.org if they want to offer their zone files …

WE ARE READY! Please Apply!

Moving Forward ….

Moving Forward

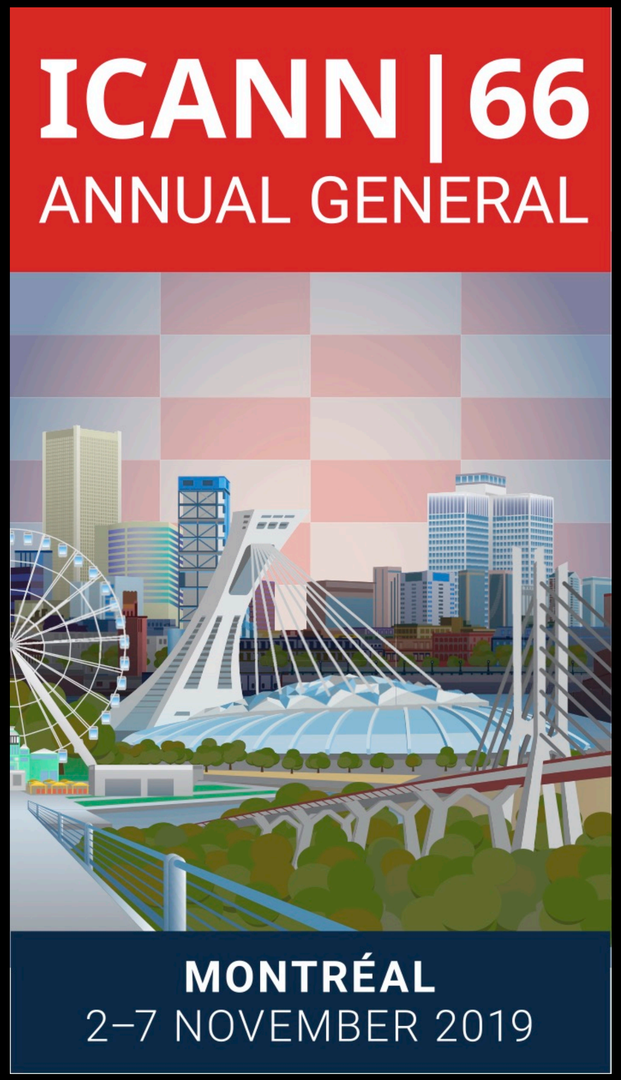
- DAAR v2
- Incorporating more blacklist feeds
- Developing RBL evaluation cycle
- Developing Registrar metrics
- Carrying out research on factors that drive security threat within registrars and registries such as registration policies

Why Read DAAR Monthly Report & Docs?

Did not show

- Detailed of the methodology and data collection
- Details of data feeds
- More detailed analytics
- We invite you registries to access the data
- To replicate the results

Questions?



Get in touch with us

- daar@icann.org
- samaneh.tajali@icann.org
- john.crain@icann.org

Abuse discussions mailing list

- dns-abuse-measurements@icann.org

Access DAAR Monthly Reports

- https://www.icann.org/octo-ssr/daar

Work Breakdown

	Detailed Task	Frequency	Party
Data Collection	Zone files	Daily	iThreat Cyber Group
	WHOIS	Daily	iThreat Cyber Group
	Abuse feeds	Daily	iThreat Cyber Group
Data preprocessing	Prepossessing all the data feeds to remove anomalies, false positives, and others.	Daily	iThreat Cyber Group
Data Aggregation	Aggregate all the 3 data types, merge them and produce abuse metrics	Daily/Monthly	iThreat Cyber Group
Data Analytics	Cleaning the data and producing aggregated statistics and analytics	Monthly	Samaneh (OCTO-SSR)
Monthly Reports	Publishing DAAR white paper including monthly & historical analysis of TLD abuse	Monthly	Samaneh (OCTO-SSR)