ICANN 71
VIRTUAL POLICY FORUM
14–17 June 2021

Understanding Reputation Block Lists

Thursday, 17 June 2021
10:30-12:00 CEST

Program

Part 1	Part 2
Introduction	Interviews
Part 3	**Part 4**
Presentation and Panel Discussion	Conclusion

Introduction

Part 1

Participants

Moderator

⊙ LG Forsberg	iQ Global AS

Interviews

⊙ Carel Bitter	Spamhaus
⊙ Ben Coon	WMC Global
⊙ Roman Hüssy	abuse.ch

Presentation

⊙ Samaneh Tajalizadehkhoob	ICANN organization

Panel Discussion

⊙ Joanna Kulesza	At-Large Advisory Committee
⊙ Reg Levy	Registrar Stakeholder Group
⊙ Matt Thomas	Registries Stakeholder Group

Interviews

Part 2

Presentation and Panel Discussion

Part 3

Understanding Reputation Blocklists

ICANN's Point of View

Dr. Samaneh Tajalizadehkhoob
ICANN Office of CTO

17 June 2021

What are Reputation Blocklists (RBLs)?

- IP blocklists or domain (hostname) blocklists
- Regarded as malicious, untrustworthy, or simply bad reputed
  - to feed DNS firewalls to prevent malicious traffic from coming into one's network or connecting to malicious domains or IP addresses
  - to filter out spam or phishing email
  - used by large content delivery networks to prevent delivery of malicious content to their customers
  - as part of incident response or law enforcement purposes, to identify malicious infrastructure involved in attacks
- Sharing mechanisms
  - Commercial: available through rate-limited, license-based, or pay-per-use mechanisms and are maintained by for-profit companies specialized in threat intelligence
  - Open source: openly and freely available for anyone to collect and use, provided by diverse set of companies
- Threat specific (e.g., PhishTank) as opposed to more general lists (e.g., SURBL)

General Characteristics & Draw backs

- Overspecialization: Each list geared towards specific purpose [1]
- Limited coverage & overlap, limited vantage points: datafeed maintainers may have honeypots in certain geolocations, therefore they may miss malicious sources [2,4]
- Limited transparency/documentation on internal methods: a general lack of documentation of data collection and curation processes
- Absence of unified methodology: substantial methodological differences in data collection, curation, maintaining, and labeling blocklists which can lead to different effects on coverage, reliability, effectiveness, and speed of reporting (aka update cycle) [2]

Why is it Important to Know the Drawbacks?

- To inform users such as network operators, researchers, security companies relying on these security resources
- To design more effective defenses and curation methods that account for the complementary strengths and limitations of individual blocklists when used in isolation or in combination

ICANN SSR's use of RBLs

- Domain Abuse Activity Reporting (DAAR)
  I. Takes domain names from TLD registry zone files
  II. Takes domain names from a preselected set of reputation feeds for phishing, malware, botnet command & control and spam as a delivery vector ***
  III. Overlaps domains from the first and second step
  IV. Processes and calculates daily rate of domains in zone that appear in the RBLs
  V. Generates daily, monthly and time series statistics, analytics and visuals to see
    - Where DNS security threats are concentrated
    - How this concentration changes over time

- *** This step contains extensive preprocessing, cleaning, unifying the RBL data feeds

ICANN SSR's use of RBLs

- ICANN Compliance Support (SSR's research)
  - Takes domain names from TLD zone files
  - Maps domain names to their corresponding registrar IDs and registrar families using the BRDA** data
  - Takes domain names from a preselected set of RBLs for phishing and malware for a specific period of time ***
  - Collapses domains from the first and second step
  - Calculates metrics showing which registrars have a higher degree of security threat concentrations in one point of time and over time

** Important to note that so far we only can use BRDA for compliance purposes

*** This step contains extensive preprocessing, cleaning, unifying the RBL data feeds

ICANN SSR's use of RBLs

- Other research projects
  - Predicting DNS threats
    Historical analysis of the RBLs can be used to extract patterns that characterize malicious domains
  - Distinguishing maliciously registered vs. compromised domains using a similar technique to COMAR [5]
    - Only a subset of domain-based RBLs make this distinction

ICANN's Current Evaluation Criteria

- We monitor reputation feeds for a period of time before including any as part of our research work. We use:
  - Reputed lists within academia and industry based on publications
  - Lists with better documented data sanitization and record removal processes & compliment the existing set, in terms of coverage

ICANN's Future Evaluation Criteria

- We are working on developing a more comprehensive method to evaluate an RBL in terms of
  - Purity
    - Manual False Positives/False Negatives analysis based on a ground truth
  - Coverage
    - The percentage of overall threat domains that are listed
  - Responsiveness
    - Indication of responsiveness of one reputation feed in comparison to the others in a set
  - Accuracy
    - How detailed the information of a domain is in a reputation feed
  - Agility / Stability
    - The consistency of domain names / ranking in lists
  - Liveliness
    - How much of listed domain names are TPs and *active* when they appear in a feed

Among others ….

References on Block List Evaluations

1) Ramanathan, Sivaramakrishnan, Jelena Mirkovic, and Minlan Yu. "BLAG: Improving the Accuracy of Blacklists." *NDSS*. 2020 https://par.nsf.gov/servlets/purl/10205652

2) Feal, Álvaro, et al. "Blocklist babel: On the transparency and dynamics of open source blocklisting." *IEEE Transactions on Network and Service Management* (2021).

3) S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang, “An empirical analysis of phishing blacklists,” 2009.

4) L. Metcalf and J. M. Spring, “Blacklist ecosystem analysis: Spanning jan 2012 to jun 2014,” in *Proceedings of the 2Nd ACM Workshop on Information Sharing and Collaborative Security*. New York, NY, USA: ACM, 2015.

5) Maroofi, Sourena, et al. "COMAR: Classification of Compromised versus Maliciously Registered Domains." *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2020.

Conclusion

Part 4

ICANN

One World, One Internet

Visit us at **icann.org**

@icann

facebook.com/icannorg

youtube.com/icannnews

flickr.com/icann

linkedin/company/icann

soundcloud/icann

instagram.com/icannorg