Abstract: DINR2016

Title: **Walking The Line: Challenges, Benefits, and Scope of Data Sharing in a Privacy-Sensitive Public Recursive Resolver System**

Presenters: John Todd (Packet Clearing House) jtodd@pch.net
Andre Ludwig (Global Cyberalliance) aludwig@globalcyberalliance.org

Global Cyberalliance (GCA) & Packet Clearing House (PCH) are currently in the process of collaborating in the construction of a global anycast open recursive DNS infrastructure that will block malicious domains for the users of the platform. This highly vetted data on malicious domains will be contributed by industry partners. The project will then take that information and leverage it to block (via NXDomain) responses requesting those malicious domains.

The project plans to provide strict privacy protection capabilities for end users, as well as provide generic telemetry to contributing organizations on the data that is provided for blocks. There are interesting technical issues to address around maintaining privacy while retaining enough value to partner organizations providing the DNSBL data or researchers trying to draw conclusions from the data. We have established basic controls and techniques to attempt to balance our contradictory concerns, but wish to understand any “best practices” currently observed by the research and security community for enhancing the effectiveness of any data we may report while maintaining a high level of privacy/integrity in respect to any PII data.

Our talk with primarily be an introduction to privacy issues we have considered, in the hopes that our comments can generate discussion and participation by the research community in our decision process. Additionally, we have included many operational items on this abstract in the hopes that our project may be noticed by researchers who have specific areas of interest.

Research privacy consideration topics:

- How to adhere to EU PII requirements on a multi-national configuration collecting DNS data?
- What is a best-practices set of data for such DNS or DNSBL metric feeds?(ASN? Timestamps? Observed locations?)
- One-way encryption or stripping of IP data - at what level are results still useful to researchers?
- Full-knowledge vs. “minimal effort” privacy compromises - what is best practice?
- Malicious use of research data - how to protect against backwards-engineering if research data is made public or non-confidential?

Operational Privacy considerations:

- What is the current research on use of ECS for malicious purposes?
- Is ECS whitelisting used by recursive operators? How are lists built?
- What is the growth curve of ECS production and consumption by other recursive/auth resolvers?

- Unique spear-phishing host data - implies leakage if FQDN is provided to researcher and mapping is understood (or embedded)
- What is the prevalence of such spear-phished host usage vs. path unique ("to the right of /") methods, and how will it influence information sharing given PII requirements?

Prediction and validation research for DNSBL:

- Based on historical data, is it possible to create synthetic suspicion values when new or continuing DNS RR's are observed that match criteria? What criteria?
- Ensure false positives are quickly resolved is a significant research item. What are the comparisons that can be done to validate domains provided by DNSBL sources?

Anycast/Global Recursors: Special Interest Items

- Novel "Marco Polo" method of reverse-path validation may provide significant protection in anycast arrays; research needed as to viability in production environment.
- What are the differences in DNS recursion characteristics, DNSBL usage, and other observable behaviors based on geographic distribution?
- How could a widely-peered network provide more interesting DNS topology data vs. a recursive resolver array with few peers?
- Is there a research need for real-time BGP data in conjunction with DNS authoritative destination or query origin? (poisoning sophistication analysis)

Operationally considered questions requiring evaluation not related to privacy/policy:

- Are there existing methods for reducing amplification attack vectors other than the common ones in most recursive system software?
- What are the indicators that we might use to understand the adoption barriers to more complex client settings? (i.e.: alternate configurations that provide more or less information leakage to the operator or third-parties.)
- What are the best methods by which to detect and mitigate "slow bleed" amplification or poisoning attacks in an anycast or unicast environment?
- What is the observed "infection" rate of DNSBL hosts across an area/network/nation?
- Are there significant differences in resolver libraries that can be modeled and compensated for to improve performance/security/privacy?
- What is the prevalence of TCP-capable DNS stub libraries?
- What are the benefits of end-of-TTL pre-fetching for popular RR's?
- How is adoption stalled as complexity rises for recursive configurations?

-- end