Visualizing DNS Datasets for Alert-driven Threat Analysis

Rosa Romero-Gomez, Yacin Nadji, Panagiotis Kintis, and Manos Antonakakis
School of Electrical Engineering and Computing
Georgia Institute of Technology
{rgomez30,yacin,kintis,manos}@gatech.edu

Index Terms: H.5.2 [Information Interfaces & Presentations]: User Interfaces - Graphical User Interfaces (GUI); C.2.0 [Computer-Communication Networks]: GeneralSecurity and Protection; I.3.8 [Computer Graphics]: Applications

1 Introduction

Domain names are frequently used by attackers when constructing malicious infrastructure due to their ubiquity and low cost. This also makes them useful, both: (1) when investigating or detecting malicious activity and (2) when network detectors alert on suspicious domain names. While the security community understands the behavioral properties of malicious domain names [5, 8, 7], little work has been done to enable network operators in the investigation of Domain Name System (DNS) datasets — especially in relation to malicious DNS indicators. No detector is perfect, and techniques to explore features associated with malicious use of the DNS would enable operators to quickly separate false positives and true positives from their security appliances' output. When detectors are correct, these relationships from historic resolutions would also allow network operators to identify additional infections in their networks, or even prevent future malicious communications from occurring.

In an effort to address these issues, this poster presents an interactive, multi-view visualization prototype system for DNS datasets focused on assisting network operators performing *alert-driven threat analysis*. We accomplish this by leveraging open DNS datasets [10], although passive DNS datasets [13] could be collected and used as well. The visualization features of this prototype are derived from state-of-the-art security research used to identify related infrastructure and detect malicious uses of the DNS. To the authors' knowledge, this is the first proposed approach to visualize large, historic DNS datasets with the specific focus of expanding knowledge and highlighting potentially malicious network infrastructure. The code and datasets used in this poster will be made publicly available at the Active DNS project website[1] upon publication.

2 Related Work

In spite of the huge amount of data collected by network measurement projects, currently few free and commercially available DNS visualization tools exist.

DNSViz [1] is a free visualization tool which, given a domain name, visualizes the chain of name servers and certificates involved in the DNSSEC chain of trust. Relationships between DNSSEC components are represented as a non-interactive directed graph. Flying Term [11] is a visualization system to help operators identify and understand querying behavior due to anomalies such as misconfigurations and security events. DNSentinel [2] is a commercially available visualization tool that operators can use to issue custom IP address or domain name queries returning a node-graph depiction of IP address/domain name relationships. More recently, DNSMON [3] has been proposed to allow actively comparing the availability and responsiveness of DNS servers at the root and TLD level. Our visualization approach differs primarily from that of the visualization tools described above in its more interactive and multi-dimensional nature, providing different perspectives on DNS traffic. We advocate different DNS views with the specific focus of expanding network operators' knowledge and highlighting potentially malicious network infrastructure.

3 Visualization Approach

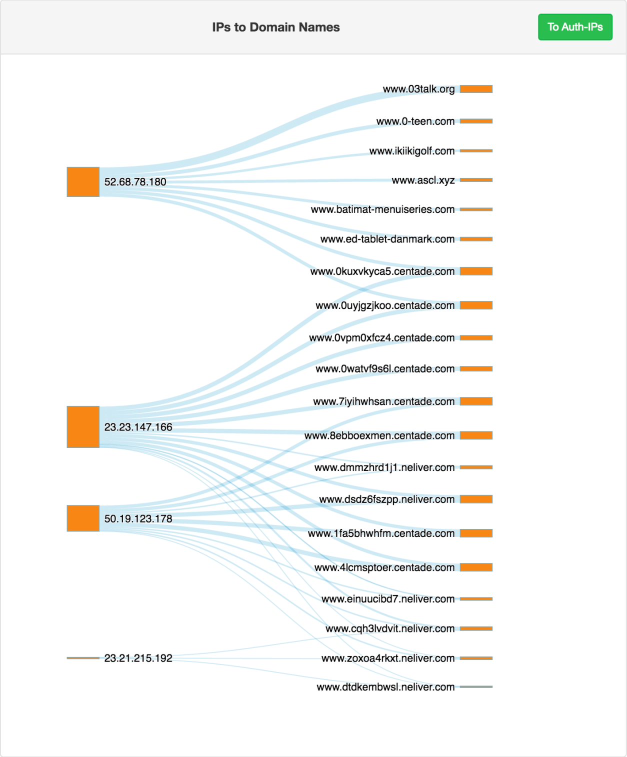
IPs to Domain Names
To Auth-IPs
52.68.78.180
23.23.147.166
50.19.123.178
23.21.215.192
www.03talk.org
www.0-teen.com
www.ikiikigolf.com
www.ascl.xyz
www.batimat-menuiseries.com
www.ed-tablet-danmark.com
www.0kuxvkyca5.centade.com
www.0uyjgzjkoo.centade.com
www.0vpm0xfcz4.centade.com
www.0watvf9s6l.centade.com
www.7iyihwhsan.centade.com
www.8ebboexmen.centade.com
www.dmmzhrd1j1.neliver.com
www.dsdz6fszpp.neliver.com
www.1fa5bhwhfm.centade.com
www.4lcmsptoer.centade.com
www.einuucibd7.neliver.com
www.cqh3lvdvit.neliver.com
www.zoxoa4rkxt.neliver.com
www.dtdkembwsl.neliver.com

Figure 1: Sankey diagram for related domain names to the domain name $www[dot]0kuxvkyca5.centade[dot]com$ (from April 1st to April 8th 2016). This diagram allows us to quickly identify fluxing IP addresses for these related domain names, which are being increasingly used in many illegal practices including malware downloads, and other forms of Internet fraud.

Inspired by the three high-level categories of features proposed by Antonakakis et al. [4], this visualization prototype (based on the use of the D3.js Javascript library[2]) supports the exploration of DNS datasets through multiple distinct visualizations, called views.

[1] http://www.activednsproject.org/

[2] https://d3js.org/

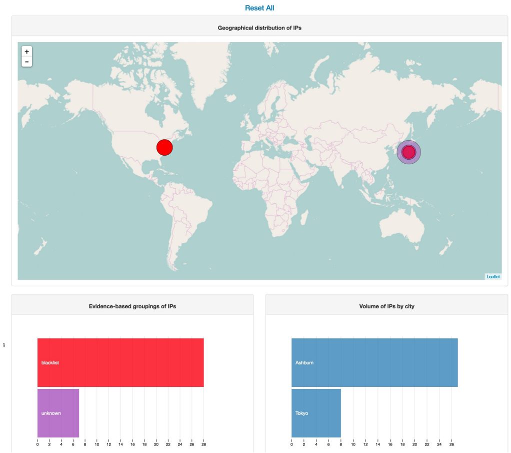

Figure 2: Geographical map showing the distribution of related historical IP addresses to the domain name $www[dot]0kuxvkyca5.centade[dot]com$ (from April 1st to April 8th 2016). The bar charts below show that most related domain names are already blacklisted (using red coloring) and mainly located in Japan and United States.

The use of multiple views in our approach follows the *diversity* and *complementarity* rules proposed by Baldonado et al. [12]. Multiple views can help users understand complex relationships among different data sets with multiple attributes. The views include:

- A *Network-based* view to help network operators understand how attackers allocate their network resources. For a given domain name d in a selected period of time, it displays the connections between domain names, IP addresses, and IP addresses related to Authoritative Name Servers using multiple *sankey diagrams* (see Figure 1).

- An *Evidence-based* view to help network operators to measure the level of *homogeneity* of DNS querying traffic. It displays an overview of the geographical distribution of historical IP addresses and encodes connections to malicious behaviors by using colors and providing multiple grouping options (see Figure 2). Intuitively, a domain name d can be considered suspicious when there is evidence that d or its IP addresses are (or were in previous months) associated with known malicious activities.

- A *Zone-based* view to show network operators the level of similarity across related domain names in order to develop hypotheses. It displays the top-level domain (TLD) distribution of domain names' strings using a *zoomable node-link tree diagram* (see Figure 3).

The general starting point for displaying these views is a suspicious domain name from an alert or log. Once the different views are displayed, this visualization approach relies on *brushing and linking* and *zooming* mechanisms as principal interaction paradigms. The use of brushing and linking allows to overcome the shortcomings of single visualization techniques [9]. Similarly, zooming mechanisms are applied to address visual scalability issues [6].

Acknowledgements

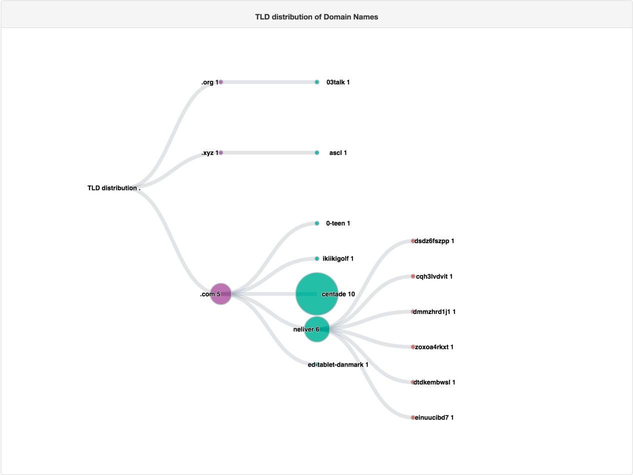

Figure 3: TLD distribution of related domain names to the domain name $www[dot]0kuxvkyca5.centade[dot]com$ (from April 1st to April 8th 2016). The third-level labels for 2LD $.neliver$ are very similar. They appear to be randomly generated and likely part of the same malicious campaign.

This material is based upon work supported in part by the Infoblox, Inc. grant no. 2106DGN. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Infoblox, Inc. We also thank the reviewers for their insightful feedback and comments.

References

[1] Sandia national laboratories, dnsviz. [online]. `http://dnsviz.net`, 2016.

[2] Src cyber, dnsential domain name service visualization tool,. `http://www.srcinc.com/what-we-do/cybersecurity/dnsentinel.html`, 2016.

[3] C. Amin, M. Candela, D. Karrenberg, R. Kisteleki, and A. Strikos. Visualization and monitoring for the identification and analysis of dns issues. In *Proceedings of the Tenth International Conference on Internet Monitoring and Protection*, 2015.

[4] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster. Building a dynamic reputation system for dns. In *USENIX security symposium*, pages 273–290, 2010.

[5] D. Dagon, G. Gu, C. P. Lee, and W. Lee. A taxonomy of botnet structures. In *Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual*, pages 325–339. IEEE, 2007.

[6] S. G. Eick and A. F. Karr. Visual scalability. *Journal of Computational and Graphical Statistics*, 11(1):22–43, 2002.

[7] G. Gu, J. Zhang, and W. Lee. Botsniffer: Detecting botnet command and control channels in network traffic. 2008.

[8] T. Holz, C. Gorecki, K. Rieck, and F. C. Freiling. Measuring and detecting fast-flux service networks. In *NDSS*, 2008.

[9] D. A. Keim. Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics*, 8(1):1–8, 2002.

[10] A. Kountouras, P. Kintis, C. Lever, Y. Chen, Y. Nadji, D. Dagon, M. Antonakakis, and R. Joffe. Enabling network security through active dns datasets. In *19th International Symposium on Research in Attacks, Intrusions and Defenses*. IEEE, 2016.

[11] P. Ren, J. Kristoff, and B. Gooch. Visualizing dns traffic. In *Proceedings of the 3rd international workshop on Visualization for computer security*, pages 23–30. ACM, 2006.

[12] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interfaces*, pages 110–119. ACM, 2000.

[13] F. Weimer. Passive dns replication. In *FIRST conference on computer security incident*, page 98, 2005.