Blind Men and the DNS (abstract)

Giovane C. M. Moura Moritz Müller Marco Davids
Maarten Wullink Cristian Hesselman
SIDN Labs
firstname.lastname@sidn.nl

Domain names have been used to provide a simple identification label for hosts, services, applications, and networks on the Internet [12]. They have also been long misused for types of abuse: phishing, malware distribution, spamming, and botnet command-and-control (C&C), among others. Underlying these abuses, we find profitable *business models* that provide the *incentives* for these abusers to continue with such activities.

To curb such practices, the research community has been active in proposing various solutions, such as [6, 7, 8, 3, 2, 5]. While these works advance the state-of-the art and have a clear contribution, they are faced with two main shortcomings: (i) they are constrained by *type* and/or *duration* of their respectively available datasets (due to the difficulty in obtaining such datasets) and (ii) while these solutions cover different sorts of abuse, we lack a survey on domain-related abuses, which leaves the question of how much ground has *not* been covered yet unanswered.

We propose to address both issues by focus on historical and complete datastes from a TLD (historical registration database (`RegDB`), historical DNS traffic to its authoritative servers (`AuthDNS` [10]), and historical data on the DNS records used by each domain on its zone (`Records` [15]).We address the second issue by presenting a survey on domain abuses and discussing their underlying business models and respective monetization methods, and how they create patterns on our datasets.

Table 1 shows a non-comprehensive list of business models and the "strength" of the signal in each of those datasets. Phishing, for example, comes in forms: 0-day (a domain is registered with the sole purpose of abuse) and compromised (websites that are hacked and wind up unknowingly hosting phishing content). Currently, most phishing attacks are compromised ones – so the registration database (`RegDB`) is virtually useless to detect these, since no information is changed for these attacks. However, both types of phishing are followed by a large number of DNS queries to those domains, which can be measured at the authoritative traffic

Business	`RegDB`	`AuthDNS`	`Records`	**Lit**
Phishing(0-day)*	Weak	Strong	Weak	[8, 5]
Phishing(comp.)*	None	Strong	Weak	[14]
Parking (Ads)	Strong-bulk	Weak	Strong	[1, 18]
Parking (Mal.)	Strong-bulk	Medium	Strong	[1, 18]
Fake Goods*	Weak	Weak	Medium	[5, 11]
Drop-Catch	Medium	Medium	Weak	[9]
Botnet C&C	Medium	Strong	?	[16]
Blackhat SEO	Medium	Medium	Strong	[13, 4]

Table 1: Business Models and Datasets/signal "strength". *business that rely on spam

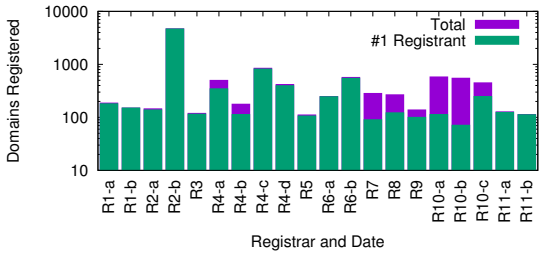

Figure 1: Registrations and registrations by #1 registrant on anomalous days for registrars. One registrant is responsible for most registrations (exception is R9, who uses random data, so we correct for it).

(`AuthDNS`) [8, 5], since spam campaigns are typically used to advertise the phishing URLs.

Another example is domainers, i.e., users that register large number of domains to monetize using ads or malware. Since both perform registrations in bulk, it can be easily be detected using `RegDB`. By performing anomaly detection [17] per individual registrar on the `.nl` zone for June 2016, we found 11 registrars with anomalous bulk registrations, as shown in Figure 1. Most of these were done by one registrant only. Upon closer inspection, we observed that these websites ran regular ads, thus no incentive to hide their ID. Some of them, however, ran adult/malware related advertises, and tend to use randomly generated data, but could detected by analyzing the registration timestamps.

These two examples illustrate the need to understand the business models used by domain abusers and their implications on our datasets, and to develop tailored solutions, taking into account the agility needs of each model. Otherwise, we will remain limited by the lack of complete datasets, winding up like the folks described at the blind and an elephant metaphor[1].

[1] https://en.wikipedia.org/wiki/Blind_men_and_an_elephant

Acknowledgements

We thank the anonymous reviewers of the DINR 2016 workshop for their valuable comments.

1. REFERENCES

[1] Alrwais, S., Yuan, K., Alowaisheq, E., Li, Z., and Wang, X. Understanding the Dark Side of Domain Parking. In *23rd USENIX Security Symposium (USENIX Security 14)* (San Diego, CA, Aug. 2014), USENIX Association, pp. 207–222.

[2] Antonakakis, M., Perdisci, R., Dagon, D., Lee, W., and Feamster, N. Building a Dynamic Reputation System for DNS. In *USENIX security symposium* (2010), pp. 273–290.

[3] Antonakakis, M., Perdisci, R., Lee, W., Vasiloglou II, N., and Dagon, D. Detecting Malware Domains at the Upper DNS Hierarchy. In *USENIX Security Symposium* (2011), pp. 16–32.

[4] Du, K., Yang, H., Li, Z., Duan, H., and Zhang, K. The ever-changing labyrinth: A large-scale analysis of wildcard dns powered blackhat seo. In *25th USENIX Security Symposium (USENIX Security 16)*, USENIX Association.

[5] Giovane C. M. Moura, Moritz Muller, Maarten Wullink, and Cristian Hesselman. nDEWS: a New Domains Early Warning System for TLDs. In *IEEE/IFIP International Workshop on Analytics for Network and Service Management (AnNet 2016), co-located with IEEE/IFIP Network Operations and Management Symposium (NOMS 2016)* (April 2016).

[6] Hao, S., Kantchelian, A., Miller, B., Paxson, V., and Feamster. PREDATOR: Proactive Recognition and Elimination of Domain Abuse at Time-Of-Registration. In *Proceedings of the 2016 ACM CCS* (October 2016).

[7] Hao, S., Thomas, M., Paxson, V., Feamster, N., Kreibich, C., Grier, C., and Hollenbeck, S. Understanding the Domain Registration Behavior of Spammers. In *Proceedings of the 2013 Conference on Internet Measurement Conference* (New York, NY, USA, 2013), IMC '13, ACM, pp. 63–76.

[8] Hao, Shuang and Feamster, Nick and Pandrangi, Ramakant. Monitoring the Initial DNS Behavior of Malicious Domains. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference* (New York, NY, USA, 2011), IMC '11, ACM, pp. 269–278.

[9] Lever, C., Walls, R., Nadji, Y., Dagon, D., McDaniel, P., and Antonakakis, M. Domain-Z: 28 Registrations Later.

[10] Maarten Wullink, Giovane C. M. Moura, Müller, M, and Cristian Hesselman. ENTRADA: a High Performance Network Traffic Data Streaming Warehouse. In *Network Operations and Management Symposium (NOMS), 2016 IEEE* (April 2016).

[11] McCoy, D., Pitsillidis, A., Jordan, G., Weaver, N., Kreibich, C., Krebs, B., Voelker, G. M., Savage, S., and Levchenko, K. PharmaLeaks: Understanding the Business of Online Pharmaceutical Affiliate Programs. In *Proceedings of the 21st USENIX Security Symposium* (Bellevue, Washington, USA, August 2012), USENIX Association.

[12] Mockapetris, P. *RFC 1034 Domain Names - Concepts and Facilities*. Internet Engineering Task Force, 1987.

[13] niche site project. Private Blog Networks http://nichesiteproject.com/private-blog-networks/, Sept. 2016.

[14] Noroozian, A., Korczynski, M., Tajalizadehkhoob, S., and van Eeten, M. Developing security reputation metrics for hosting providers. In *8th Workshop on Cyber Security Experimentation and Test (CSET 15)* (2015).

[15] OpenINTEL. OpenINTEL Open Access – http://openintel.nl/, 2016.

[16] Stone-Gross, B., Cova, M., Cavallaro, L., Gilbert, B., Szydlowski, M., Kemmerer, R., Kruegel, C., and Vigna, G. Your botnet is my botnet: analysis of a botnet takeover. In *Proceedings of the 16th ACM conference on Computer and communications security* (2009), ACM.

[17] Vallis, O., Hochenbaum, J., and Kejariwal, A. A novel technique for long-term anomaly detection in the cloud. In *6th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 14)* (2014).

[18] Vissers, T., Joosen, W., and Nikiforakis, N. Parking Sensors: Analyzing and Detecting Parked Domains. In *Proceedings of the 2015 Network and Distributed System Security Symposium (NDSS 2015), San Diego, California, USA.* (2015).