TIDE – Threat Identification Using Active DNS Measurements

Anna Sperotto
University of Twente
The Netherlands
a.sperotto@utwente.nl

Olivier van der Toorn
University of Twente
The Netherlands
o.i.vandertoorn@student.utwente.nl

Roland van Rijswijk-Deij
University of Twente and SURFnet bv
The Netherlands
r.m.vanrijswijk@utwente.nl

ABSTRACT

The Domain Name System contains a wealth of information about the security, stability and health of the Internet. Most research that leverages the DNS for detection of malicious activities does so by using passive measurements. The limitation of this approach, however, is that it is effective only once an attack is ongoing. In this paper, we explore a different approach. We advocate the use of active DNS measurements for pro-active (i.e., before the actual attack) identification of domains set up for malicious use. Our research makes uses of data from the OpenINTEL large-scale active DNS measurement platform, which, since February 2015, collects daily snapshots of currently more than 60% of the DNS namespace. We illustrate the potential of our approach by showing preliminary results in three case studies, namely snowshoe spam, denial of service attacks and a case of targeted phishing known as CEO fraud.

CCS CONCEPTS

• **Networks → Network measurement**; **Network security**;

KEYWORDS

DNS, active measurements, network security

ACM Reference format:
Anna Sperotto, Olivier van der Toorn, and Roland van Rijswijk-Deij. 2017. TIDE – Threat Identification Using Active DNS Measurements. In *Proceedings of SIGCOMM Posters and Demos '17, Los Angeles, CA, USA, August 22–24, 2017 (SIGCOMM 2017),* 3 pages.
https://doi.org/10.1145/3123878.3131988

1 INTRODUCTION

The Domain Name System (DNS) is one of the core Internet infrastructures. It is also an extremely rich source of information about the security, stability, and in general, health of the Internet. Monitoring of the DNS – especially when performed at a large scale – can yield important information about the use and security of the Internet. A notable approach is *passive DNS* (pDNS) [6], a system that monitors DNS queries and responses issued from a recursive resolver towards authoritative name servers. From a security perspective, pDNS is used to investigate DNS anomalies [1, 2]. Several types of anomalies can be identified, for example domains used for spam campaigns, fast flux and malware.

All these approaches are, however, reactive, thus becoming effective only once an attack is already taking place. Less studied in literature, with the exception of [3], is instead the possibility of using DNS measurements for *pro-actively* identifying domains set up for malicious use.

We propose to use active DNS measurements for gaining insight into malicious activity in the making. Our approach leverages the uniqueness, both in duration and coverage, of the OpenINTEL[1] large-scale active DNS measurement, which takes daily snapshots of the resource records for 60% of the global DNS namespace. Such a data set is therefore a novel vantage point for observing malicious activities. The rationale for this research is that sophisticated attacks require careful preparation, relying on different attack phases and additional infrastructure. In preparing an attack, attackers are forced to expose information about their infrastructure, several aspects of which we will be able to observe in the DNS. Our intuition, backed-up by preliminary results, tells us that a structural, automatic analysis of the content of the DNS can give us an advantage on the attacker. The research we will conduct in the upcoming TIDE project aims, firstly, at devising an alternative way of detecting attacks compared to passive measurements; and secondly, at investigating the possible time advantage active measurements give us on the attacker, enabling us to pro-actively identify threats *before* the actual attack takes place.

2 DATASET

Since February 2015, the OpenINTEL[1] large-scale active DNS measurement platform collects daily snapshots of the data in the DNS [5]. The measurement currently queries 60% of the global DNS namespace. At the time of writing, the measurement covers the zones `.com`, `.net`, `.org`, `.info`, `.mobi`, the new gTLDs defined by ICANN, and a set of ccTLDs such as `.nl`, `.se`, `.ca`, `.fi`, `.at`, `.dk` and `.nu`. For each domain name, the measurement performs, for the apex and `www` label, a set of 11 queries including `A`, `AAAA`, `MX`, `NS`, `TXT`, `SOA` records, and for signed domains also the `DS` and `DNSKEY` records.

3 CASE STUDIES

We illustrate the potential of our approach by presenting anecdotal evidence for three case studies: snowshoe spam, denial of service attacks, and a targeted form of phishing known as "CEO fraud".

Snowshoe spam. In snowshoe spam, attackers distribute the load of spam among a large set of sources, aiming to evade detection based on reputation systems (e.g. blacklists). Snowshoe spam is therefore notoriously difficult to detect. We compare a dataset of 15k domains known to be related to snowshoe campaigns with a reference dataset of 15k regular domains (from the Alexa Top

[1] https://openintel.nl/

SIGCOMM 2017, August 22–24, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5057-0/17/08.
https://doi.org/10.1145/3123878.3131988

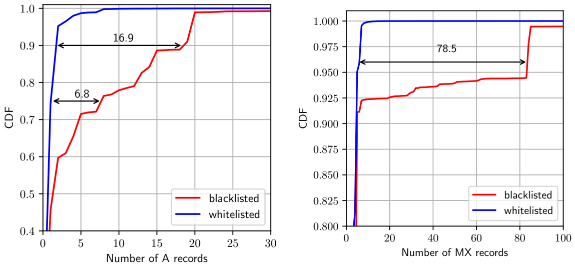

Figure 1: CDF of the number of A records (left) and MX records (right) for snowshoe spam and regular domains

1M list). Fig. 1 shows the CDF of the number of A and MX records for spam and regular domains. This analysis indicates that at the 90^{th} percentile for the A record distribution, spam domains have on average 16.9 records more than regular domains. Similarly, at the 95^{th} percentile of the MX record distribution, spam domains have 78.5 records more than regular domains. This result shows that characteristics of spam related domains significantly deviate from the ones for regular domains, thus enabling detection in DNS data.

Denial of service attacks. DNS amplification is a form of distributed DoS attacks in which an attacker will prompt a service to answer fake queries seemingly generated by the target. The attacker will typically send a query for which he knows the response will be very large, to maximize the amplification effect. An effective way for achieving this in the DNS is to use a domain under the control of the attacker himself. In our dataset, we see evidence of this behavior. An example is given in Fig. 2, for the domain *sunrisecx.com*. Fig. 2 shows that, while the number of records was initially modest, starting from March 2015 we observe that the domain has been inflated by adding more than 200 A records. Responses to ANY queries for this domain are estimated to be close to 3500 bytes. During the period when the domain was inflated, there is evidence based on the AmpPot project [4] that the domain was used in amplification attacks. We make this visible in Fig. 2 by indicating the window of time in which attacks were observed. It is important to note that, by using DNS data, the malicious domain is observed two weeks before it was first used in attacks.

CEO fraud. In CEO fraud, attackers send an email to e.g., the financial department of a company impersonating the CEO and requesting support for a transfer of funds. The analysis of a set of domains used in CEO fraud that targeted a Dutch ISP on August 30, 2016, highlighted that the TXT records contained an Office 365 specific token. Since such a token is linked to a specific Office 365 environment, we identified it as key characteristic of this phishing campaign. Using this information, we were unable to uncover a much larger set of malicious domains, allowing us to actively warn potential targets. Table 1 shows that 1) malicious domains were active before the fraud was reported for the first time and 2) that new malicious domains are progressively appearing until September 7, 2016, after which no new domains fitting this pattern appear.

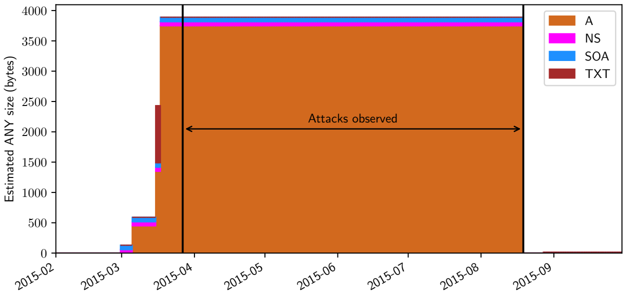

Figure 2: Example of and artificially inflated domain (sunrisesecx.com)

	#Domains sharing a specific Office 365 token												
	August						September						
TLD	26	27	28	29	30	31	1	2	3	4	5	6	7
.com	36	36	77	77	199	259	306	334	334	352	352	394	404
.net	-	2	2	2	17	17	20	38	43	43	44	54	54
.org	-	15	15	15	18	18	23	23	26	26	28	28	28
Total	**36**	**53**	**94**	**94**	**234**	**294**	**349**	**395**	**403**	**421**	**424**	**476**	**486**

Table 1: CEO fraud domains Aug./Sept. 2016

4 CONCLUSIONS

This poster presents work-in-progress results illustrating pro-active identification Internet threats using the DNS. Our analysis is based on an active DNS dataset that is unique in duration (over two years) and coverage (around 200M domains), and which therefore gives us an advantage in detecting suspicious activity. The preliminary results for the three cases of snowshoe spam, DDoS attacks and CEO fraud have shown both the feasibility of this approach and the time advantage we can gain on the attacker. We therefore believe further research in this area is beneficial to the security community. Such an approach is not without challenges, however. First, given the extensiveness of the measurement, the analysis approach needs to be *scalable* and *automatic*. Second, patterns for malicious activities are likely to change over time, and new patterns will emerge, which calls for *adaptability*. Last, since malicious activities can be identified before an attack takes place, ethical considerations on the *reliability* of the results need to be taken into account.

Acknowledgments. This work has been partially supported by the D3 project (NWO 628001018) and by SURF, the Netherlands collaborative organisation for higher education and research aimed at breakthrough innovations in ICT. Special thanks to C. Rossow and J. Krupp for sharing DDoS attack data from the AmpPot measurement.

REFERENCES

[1] M. Antonakakis, R. Perdisci, W. Lee, N. Vasiloglou II, and D. Dagon. 2011. Detecting malware domains at the upper DNS hierarchy. In *Proc. of the 20th USENIX Security Symp.*

[2] L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi. 2011. EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis. In *Proc. of the 2011 Network and Distributed System Security Symp. (NDSS 2011).*

[3] A. Kountouras, P. Kintis, C. Lever, Y. Chen, Y. Nadji, D. Dagon, M. Antonakakis, and R. Joffe. 2016. Enabling Network Security Through Active DNS Datasets. In *Proc. of the 19th Int. Symp. on Research in Attacks, Intrusions, and Defenses: (RAID 2016).*

[4] L. Krämer, J. Krupp, D. Makita, T. Nishizoe, T. Koide, K. Yoshioka, and C. Rossow. 2015. AmpPot: Monitoring and defending against amplification DDos attacks. In *Proc. of the 18th Int. Symp. on Research in Attacks, Intrusions, and Defenses (RAID 2015)*.

[5] R. van Rijswijk-Deij, M. Jonker, A. Sperotto, and A. Pras. 2016. A High-Performance, Scalable Infrastructure for Large-Scale Active DNS Measurements. *IEEE Journal of Selected Areas in Communications* 34, 7 (2016).

[6] F. Weimer. 2005. Passive DNS Replication. In *Proc. of the 17th FIRST Conference (FIRST 2005)*. http://www.enyo.de/fw/software/dnslogger/first2005-paper.pdf