SIDN Labs

https://sidnlabs.nl

December 19, 2016

Peer-reviewed Publication

Title: Increasing DNS Security and Stability through a Control Plane for Top-level Domain Operators

Authors: Cristian Hesselman, Giovane C. M. Moura, Ricardo de O. Schmidt, Cees Toet

Journal: IEEE Communications Magazine, January 2017 (post-print).

DOI: 10.1109/MCOM.2017.1600521CM

Citation:

- Hesselman, C., Moura, G.C. M., de O. Schmidt, R., Toet, C.:Increasing DNS Security and Stability through a Control Plane for Top-level Domain Operators. In: IEEE Communications Magazine
- Bibtex:

```
@article{Hesselman17,
  title={{Increasing DNS Security and Stability
  through a Control Plane for Top-level Domain Operators
  (to appear)}},
  author={Cristian Hesselman and Giovane C. M. Moura
  and Ricardo de O. Schmidt, and Cees Toet},
  journal={IEEE Communications Magazine},
  pages={2--8},
  year={2017},
  publisher={IEEE}
  doi={MCOM.2017.1600521CM}
}
```

Increasing DNS Security and Stability through a Control Plane for Top-level Domain Operators (paper postprint)

Date
19 December 2016

Page
1/11

Classification
Public

Authors
Cristian Hesselman, SIDN
Giovane C. M. Moura, SIDN
Ricardo de O. Schmidt, University of Twente
Cees Toet, SIDN

Contact
cristian.hesselman@sidn.nl

Contact
T +31 (0)26 352 5500
support@sidn.nl
www.sidn.nl

Offices
Meander 501
6825 MD Arnhem
The Netherlands

Mailing address
PO Box 5022
6802 EA Arnhem
The Netherlands

Citation: C. Hesselman, G. Moura, R. de O. Schmidt, and C. Toet, "Increasing DNS Security and Stability through a Control Plane for Top-level Domain Operators", IEEE Communications Magazine, Network and Service Management Series, January 2017, pp. 2-8

Abstract

We present a control plane for operators of Top-level Domains (TLDs) in the Domain Name System (DNS), such as ".org" and ".nl", that enables them to increase the security and stability of their TLD by taking on the role of a threat intelligence provider. Our control plane is a novel system that extends a TLD operator's traditional services and detects potential threats in the TLD by continuously analyzing the TLD operator's two key datasets: the typically large amounts of DNS traffic that it handles and its database of registered domain names. The control plane shares information on discovered threats with other players in the TLD's ecosystem and can also use it to dynamically scale the TLD operator's DNS infrastructure. The control plane builds on a set of open source modules that we have developed on top of a Hadoop-based data storage cluster. They for instance enable TLD operators to run and develop threat detectors and to easily import their DNS traffic into the control plane. Our control plane uses policies to protect the privacy of TLD users and is based on our operational experience of running the .nl TLD (The Netherlands), which we are also using as the use case for our implementation.

Keywords: DNS security and stability, threat detection, automatic management, privacy protection

1 Introduction

Since their inception, domain names have been used as a simple identification label for hosts, services, applications, and networks on the Internet [RFC1034]. Until the mid 1980s, the mappings from domain names to IP addresses were distributed as a text file (HOSTS.TXT) via ftp to the relatively small number of hosts that were connected to the Internet at that time. The Domain Name System (DNS) [RFC1034] replaced this mechanism to provide domain name to IP address mappings in a scalable way and has become a critical part of the Internet infrastructure.

The DNS uses a hierarchical namespace and a tree-like structure in which each level uses so-called

SIDN Labs is the research team of SIDN, the company that manages the Netherlands' Internet extension, .nl. SIDN Labs develops, prototypes and evaluates new technologies and systems that enhance the security and stability of .nl, the DNS and the wider Internet. Visit us at www.sidn.nl and www.sidnlabs.nl.

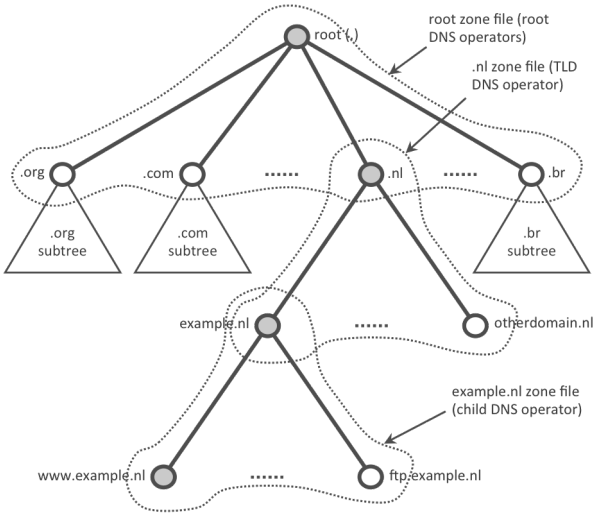

Figure 1. DNS naming hierarchy and DNS operators.

"authoritative name servers" to provide pointers to the next lower level. As an example, consider a user tying to reach the website www.example.nl (see Figure 1). The user's computer first connects to a resolver, which is a recursive name server that interacts with authoritative name servers on behalf of the user and that is usually located in the network of the user's Internet access provider. The resolver obtains a reference to the ".nl" namespace from the root name servers, then a reference to "example.nl" from the .nl name servers, and finally the reference to "www.example.nl" from the name server of example.nl. This last name server knows the requested IP address, which the resolver returns to the user, allowing its browser to reach www.example.nl.

The second level of the DNS namespace currently contains over 1,300 Top-level Domains (TLDs), classified into country code TLDs (such as ".nl" and ".br"), generic TLDs (such as ".com" and ".org"), and new generic TLDs such as ".amsterdam" and ".shop". The operators of these TLDs manage the TLD's authoritative name servers and the database of all registered second-level domain names (usually of the form [domain].[tld]). They regularly export the database contents to a so-called zone file, which is the input for the TLD's authoritative DNS servers. The other levels in the DNS tree follow this same principle, as Figure 1 illustrates.

A recent development is that some TLD operators have extended their traditional role as DNS operator to also take on the role of threat intelligence provider. They leverage the updates of their domain name database and the DNS traffic they handle on their name servers to detect potential threats in their TLD, such as phishing sites [1], DDoS attacks on the DNS [2][3], and sites that distribute malware. The underlying rationale is to protect the TLD's users by making this threat information available to other players in the TLD, such as hosting and access providers, thus helping them to better fight these threats (collaborative security).

The contribution of our work is that we have developed and implemented a so-called "control plane" that enables TLD operators to become a threat intelligence provider. The control plane is a novel system that extends a TLD operator's traditional services (registration and DNS) to automatically derive potential threats from DNS traffic, database updates, and potentially other sources. Our control plane makes this threat information available to other players in the TLD and can also use it to dynamically scale the TLD operator's DNS services. Together, these two functions increase the level of automation of operating a TLD because threat detection and DNS reconfiguration are mostly manual and ad-hoc tasks today.

Our control plane builds on several open source modules we have developed on top of a Hadoop-based data storage cluster. They for instance enable TLD operators to detect phishing sites and to easily import their DNS traffic into the control plane. Our modules are currently being used by at least six TLD operators, including those of .ca (Canada) and .at (Austria). Our control plane uses policies to protect the privacy of TLD users and is based on our operational experience of running the .nl TLD (7th largest TLD, 5.6 million domain names). We are also using .nl as the use case for our implementation.

In this paper, we focus on the design and principles of our control plane and refer the interested reader to our previous work for more technical details and extensive analysis.

We first provide an overview of infrastructure that a TLD operator typically manages (Section 2). Next, we discuss the threats TLDs are exposed to (Section 3), the functions our control plane needs to mitigate them (Section 4), and how we realized the control plane (Section 5). We end with a discussion on related work (Section 6) and conclusions and future work (Section 7).

2 TLD Operator Infrastructure

A TLD operator traditionally manages the set of authoritative name servers for the TLD (Section 2.1) and the TLD's registration database (Section 2.2).

2.1 Authoritative Name Servers

Because TLD operators form the second highest level in the DNS naming hierarchy (see Figure 1), they typically use multiple layers of redundancy to provide their DNS services in a fault-tolerant way. For example, they replicate their name servers across multiple DNS services (e.g., ns1.dns.nl and ns2.dns.nl for the .nl TLD), use multiple types of name server software, and use IP anycast [3] to replicate their DNS services across sites. The advantage of IP anycast is that it also enables TLD operators to scale their DNS capacity to deal with an increasing DNS load and to reduce response times by placing machines closer to end-users. IP anycast relies on the Internet's inter-domain routing protocol (BGP) [RFC4271] to route clients to the closest name server and is heavily used by the DNS root (eleven of its thirteen "letters" use anycast across more than 500 different locations [3]).

As an example, the DNS infrastructure for the .nl TLD consists of six unicast name servers and two anycast services. The anycast service is distributed across several dozens of sites, with one anycast service mostly co-located with large Dutch access providers ("local anycast") and the other worldwide (through third parties). We use several different types of name server software for reasons of diversity and changes to our infrastructure go through a tightly controlled change management process.

Four of our six unicast name servers together handle around 850 million DNS queries a day coming from approximately 1.3 million resolvers[1]. This is a subset of the total amount of queries because resolvers use local caches to avoid having to completely walk the DNS tree for every lookup. This increases performance and DNS scalability, but implies that authoritative servers only receive part of the queries that a resolver receives from clients.

2.2 Registration Database

A TLD operator's registration database usually contains all the second-level domain names in a TLD, which are of the form "[domain].[tld]" (some TLD operators also allow for third level registrations, such as under .com.br). The TLD operator typically enables so-called "registrars" to register a domain name (or update or delete it) in the database on behalf of Internet users, which are called "registrants". A registration corresponds to adding a leaf under a TLD in the DNS tree (see Figure 1).

Different registrars provide different registration interfaces, but the registrar-registry interface is often based on the Extensible Provisioning Protocol (EPP) [RFC 5730]. Registrars typically sell domain names in combination with hosting services.

As an example, the .nl registration database is synchronized across multiple sites, contains 5.6 million domains, and serves around 1,500 domestic and international registrars. We offer both an EPP and a web-based interface and generate and export the .nl zone file to our name servers every hour.

3 Threats

The DNS and the domain names in a TLD are exposed to various threats. Some affect the services of a TLD operator, others those of other players within the TLD. We distinguish four types of threats in this paper and refer to [RFC3833] for a more detailed description of DNS-related threats.

Zone file integrity violation: these threats involve compromising the TLD zone file (cf. [16]), for instance by stealing users or registrar credentials, allowing the attacker to change certain records in the zone file. This

[1] http://stats.sidnlabs.nl

will lead the authoritative server to respond to queries with fraudulent answers, ultimately pointing the user to a malicious domain name.

Name server unavailability: this type of threat purposely reduces the availability of name servers in the DNS, for instance through a DDoS attack [2][3][4]. This results in name servers becoming unavailable or instable (partial availability), which means that clients do not receive a response to their DNS request (in time) and are unable to reach the intended server.

DNS response integrity violation: bad actors tamper with DNS responses, for instance through man-in-the-middle attacks, DNS hijacking, or cache poisoning [RFC3833]. This results in a user being redirected to a malicious or unsolicited server. DNSSEC [RFC4035] detects this type of attack at the resolver.

Abuse: the DNS is being exposed to various sorts of abuse, such as phishing, malware distribution, and command-and-control botnet channels. While the malicious content is hosted outside the DNS, the DNS is misused to direct victims to such sites.

4 Data and Functions

The goal of our control plane is to leverage the data that a TLD operator handles (see Section 4.1) to detect potential threats in the TLD (Section 4.2) and to automatically reconfigure the TLD operator's name servers (Section 4.3). The analysis of the TLD operator's data requires a third function, which is privacy protection (Section 4.4).

4.1 TLD Operator Data

A TLD operator has two key datasets that it can use to detect threats: DNS authoritative traffic (incoming DNS queries for domains in the TLD's zone) and the TLD's domain registration database. The latter furthermore gives a TLD operator a real-time view on domain registration changes (creates, deletes, updates) across different registrars.

TLD operators can use these data sets to automatically detect patterns and suspicious behaviors in their zone. For example, the TLD operator would be able to detect spam campaigns based on bulk registrations, which has been reported on in [14]. It would also be able to detect phishing attacks based on unusual DNS traffic patterns for a domain that has just been registered (see Section 5.2). TLD operators could furthermore cautiously carry out active measurements on all domain names in their zones and use this information to augment the threat detection logic.

While resolvers and DNS operators at lower levels in the DNS hierarchy would be able to carry out a similar analysis, they miss the real-time centralized view that a TLD operator has as a result of its position at the second-highest level in the DNS (see Figure 1). This makes it difficult for them to detect and correlate malicious domain names created through different registrars, such as the automatically generated domain names that botnets use.

The limitation of a TLD operator's data is that it provides a "sampled" view on the DNS because resolvers cache queries [15]. Also, TLD operators will likely gradually receive less DNS information because of QNAME minimization [RFC7816], which is a recent DNS extension that reduces the amount of data in DNS queries to protect the privacy of users. With QNAME minimization resolvers for instance only put "example.nl" in the queries they send to TLD operators instead of "www.example.nl", which is the Fully Qualified Domain Name (FQDN). The uptake of QNAME minimization is currently limited.

4.2 Threat Detection

The purpose of threat detection is to automatically detect potential threats in a TLD, such as phishing domains and unavailability of DNS name servers (see Section 3). To accomplish this, the control plane needs to be able to quickly analyze large datasets covering a year or more of relatively high-volume DNS data. Speed is crucial to quickly detect and mitigate threats such as the appearance of phishing sites, which will affect fewer victims the sooner they are removed.

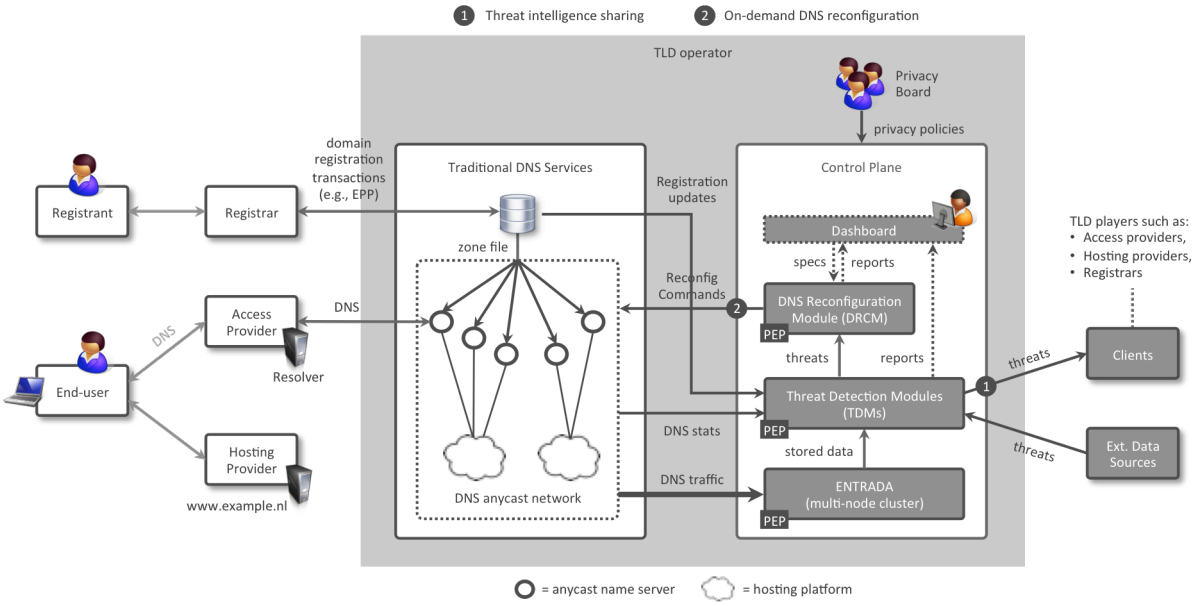

Figure 2. A TLD operator's traditional DNS services (left) and its control plane (right).

To accomplish this, the control plane needs to provide near real-time response times when analyzing a TLD operator's datasets and needs to continuously store large volumes of DNS and other data. Data Streaming Warehouses (DSWs) [5] are designed with this in mind: they continuously digest incoming data and use optimized file formats (columnar storage) and parallel processing to achieve near real-time response times. DSWs can also be easily extended with extra nodes, enabling the control plane to increase its capacity when the TLD operator's datasets grow. DSWs typically also provide an easy interface for data analysis, which eases application development and interaction with a human operator.

Our control plane's DSW needs be able to obtain the transport and IP-level information in DNS packets, which might for instance be relevant to detect reflection attacks based on ICMP messages. The DSW should also introduce limited changes on the TLD operator's name servers. This is essential because TLD-level name servers are high availability resources that are typically tightly managed (see Section 2.1). The format for importing DNS packets from name servers into the control plane should furthermore be widely used so that different TLD operators can easily implement our control plane irrespective of their particular name server setup (see Section 2.1).

We discuss our DSW in Section 5.1 and our threat detection modules and their performance in Section 5.2.

4.3 On-demand DNS Reconfiguration

The purpose of on-demand DNS reconfiguration is to dynamically adapt the DNS anycast infrastructure of a TLD operator, for instance to handle a DDoS threat (name server unavailability) as it occurs. TLD operators frequently use IP anycast because of its ability to handle stress situations [3] and because allows them to easily scale their authoritative name server infrastructure (see Section 2.1)

By adapting we mean starting and stopping anycasted and virtualized DNS name servers at specific (external) hosting platforms (cf. [6]). The result is that our control plane manages a potentially large set of DNS name servers that grows and shrinks dynamically over time, which is unlike today's static and relatively small DNS anycast networks. A precondition is that the control plane is able to interface with the TLD operator's name servers so that it can send reconfiguration commands to them.

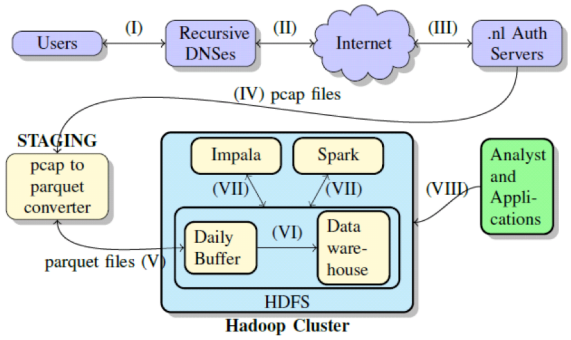

Figure 3. ENTRADA overview.

Automatic reconfiguration requires the control plane to collect a rich set of statistics on every DNS anycast node that it manages. This includes basic statistics such as processing and storage resources usage, which may be collected using tools such as Nagios[2]. More extended statistics include EDNS Client-Subnet (ECS) extensions [RFC7871]. ECS contains crude geographic information on the location of clients, which the control plane may use to map query demands to the geographical location of end-users (i.e., queries' origin) rather than of resolvers

Our ultimate goal is that the control plane raises the abstraction level of operating a DNS name server infrastructure, allowing human operators to focus on handling rare incidents because the control plane handles the "regular" ones automatically. We expect this will require advanced visualizations through a TLD operator-wide dashboard (cf. [7]), but that is outside the scope of this paper.

We discuss our reconfiguration module and its performance in Section 5.3.

4.4 Privacy Protection

Privacy protection is an important function because the DNS traffic that the control plane analyzes for threat detection and DNS reconfiguration contains IP addresses of resolvers and domain names being looked up, which may constitute Personally Identifiable Information (PII), depending on the jurisdiction. For example, under Dutch law this type of information is regarded as PII [8].

[2] http://www.nagios.org

Privacy protection requires a mechanism that allows a TLD operator to systematically balance the privacy of Internet users on the one hand and the targeted increase in the security and stability of the TLD by the control plane on the other. This mechanism needs to be flexible so that it can work with applicable privacy regulations and it needs to be easy to use for engineers who need to develop new software to detect a new type of threat. It also needs to protect privacy through technical means within the control plane. We refer to [8] for more details on privacy requirements.

We discuss our privacy protection mechanism in Section 5.4.

5 Realization

Figure 2 provides an overview of our control plane, which consists of a high-speed data streaming warehouse called "ENTRADA" (Section 5.1), Threat Detection Modules (Section 5.2), a DNS Reconfiguration Module (Section 5.3), and a privacy framework (Section 5.4).

5.1 ENTRADA

ENTRADA[3] (ENhanced Top-level Domain Resilience through Advanced Data Analysis) [7][9] is our open source DSW for the TLD control plane. ENTRADA consists of a set of modules that run on top of Apache Hadoop[4], which is open source as well.

Figure 3 provides an overview of the ENTRADA DSW and how it stores DNS authoritative traffic. Steps I—III refer to domain name resolution. We export the incoming DNS traffic from the .nl authoritative servers to a staging server (step IV), in which the raw PCAP format is converted to an optimized open source column storage format (Parquet, step V), and later imported into the Hadoop File System (HDFS, VI). Impala[5] provides a massively parallel processing query engine with a standard SQL interface (VII). Applications and services use this interface to connect to ENTRADA.

[3] http://entrada.sidnlabs.nl
[4] http://hadoop.apache.org
[5] http://impala.io

We choose PCAP as our format for importing DNS traffic from name servers because it includes transport and IP-level headers in addition to their DNS payloads, because it requires little to no changes on name servers (a mirror port on the network or a PCAP process on the name servers), and because it is widely used (see Section 4.2).

ENTRADA delivers the performance we need to build Threat Detection Modules (Section 5.2) and perform hypothesis tests. For example, we showed in [9] that ENTRADA is able to analyze the equivalent of 52TB of PCAP data in less than 3.5 minutes in a four data node cluster, using Impala and SQL syntax, which would be infeasible using PCAP format.

Our ENTRADA instance for .nl currently receives DNS traffic from four of our six unicast authoritative name servers and has been operational on our research network uninterruptedly as of March 2014. It currently stores more than 320 billion DNS query-response pairs in 15TB of Parquet-compressed format.

5.2 Threat Detection Modules

A Threat Detection Module (TDM) is an ENTRADA application that discovers potential threats, possibly in combination with other data feeds such as domain name database transactions, logs, and feeds from external threat information providers such as ShadowServer[6].

An example of a TDM we developed is the New Domains Early-Warning System (nDEWS) [10], which leverages the known fact that newly registered malicious domains receive a much higher number of DNS queries immediately after their registration than normal domains.

Figure 4 shows this based on the daily number of queries for 20,000 randomly chosen normal domains (purple line) and phishing domains (green line). nDEWS thus enables a TLD operator to monitor all new domains added to its zone on a daily basis. It uses the k-means clustering algorithm to classify them based on

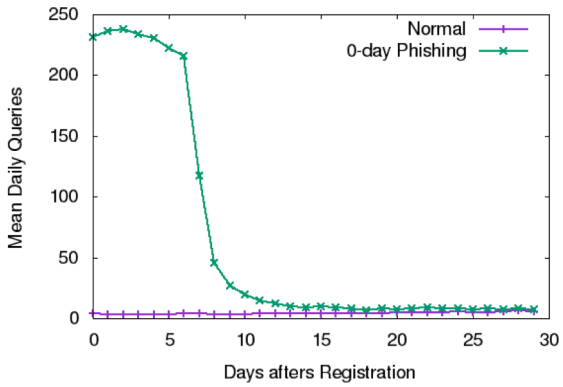

Figure 4. Queries to normal and phishing domains.

their DNS query patterns.

We evaluated nDEWS using historical 8 month-data collected from one of the .nl authoritative servers. nDEWS yielded almost 3,000 suspicious domains, which we had to validate using several techniques because we did not have a ground truth for them, since the contents of their websites might have changed during this period.

We are also evaluating nDEWS using current data and performing web content analysis if a domain is classified as suspicious. Besides phishing, nDEWS is able to detect other type of suspicious sites, such as allegedly counterfeit drugs and shoes. We automatically share the information coming out of nDEWS with 32 .nl registrars as part of a pilot.

Another TDM we have developed detects the DNS traffic pattern of a specific botnet. We identified what this pattern looks like and our TDM uses ENTRADA to continuously scan for it. When our TDM detects a resolver that exhibits this behavior, it sends the resolver's IP address to the Abuse Information Exchange (AbuseHUB)[7]. Members of this platform include large Dutch access providers, who use the information to cleanup the botnet infections located within their network. With this TDM we are thus able to actively disrupt the distribution of spam-mail and other malicious activity.

[6] http://shadowserver.org

[7] http://www.abuseinformationexchange.nl (in Dutch)

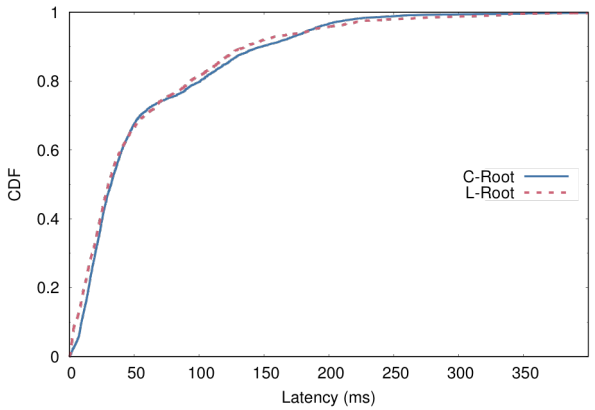

Figure 5. Distribution of latency for C- and L-Root.

We refer to [7] for other TDMs we have developed.

5.3 DNS Reconfiguration Module

The DNS Reconfiguration Module (DRCM) dynamically decides which name servers to start or stop at which locations. The DRCM is a logical entity that may be fully distributed across the name servers of a DNS anycast service (cf. [6]).

Our current DRCM focuses on minimizing the latency between resolvers and the TLD operator's authoritative name servers. To develop the DRCM, we studied the impact of the number of anycast instances and their physical locations on the latency of the anycast service and reported on this study in [11]. By measuring real-world anycast deployments from C-, F-, K- and L-Root DNS name servers using the RIPE Atlas framework, we were able to show that a handful of well-placed anycast instances provide a better and more stable latency than a large-scale infrastructure consisting of several dozens of nodes. For example, C-Root with 8 anycast sites (4 in Europe and 4 in North America) achieved a worldwide median latency of 32 ms, while L-Root with 144 sites (18x more than C) all over the globe achieved a median latency of 30 ms.

Figure 5 shows the distribution of latency for C-Root and L-Root as seen from around 7,900 vantage points around the globe. Note that the larger deployment of L-Root did not result in a shorter distribution tail as well: the 75th-percentile of the latency distribution is 76 ms for C-Root and 73 ms for L-Root. These results suggest that connectivity of the anycast site is far more important for the performance of the anycast service than the number of deployed sites, which is an important finding for our DRCM.

We have also setup a worldwide anycast testbed[8], which we are using to further investigate the relationship between number of anycast sites and their respective connectivity to service latency, in particular to understand the efficiency and impact of traffic engineering through anycast for the mitigation of DDoS attacks. We are actively probing the anycast infrastructure to understand the effects of runtime reconfigurations. We also evaluated the use of the ECS extension (see Section 4.3) based on real data measured at two authoritative name servers that are authoritative for popular second-level domains such as apache.com and we have modified them to receive and process queries with ECS extension.

5.4 Privacy Framework

Our Privacy Framework protects the privacy of the users of a TLD [8]. Its key concept is a privacy policy, which is a structured document in natural language that defines what data ENTRADA and its applications process for a particular purpose using which data filters. A filter is an operation that ENTRADA or an application applies on personal data. Examples are pseudonymization and aggregation. Filters form an essential element in the Privacy Framework, because they ensure that the privacy policies are verifiably enforced by technical means.

ENTRADA application developers and researchers formally submit privacy policies to the privacy board, which is a body within the TLD operator's organization that reviews policies. The privacy board approves or rejects the policy and informs the author through a policy evaluation report.

After policy approval, the author implements it as part of a Policy Enforcement Point (PEP), which is the technical component within ENTRADA or one of its applications that realizes an approved privacy policy and actually applies the policy's filters at run-time.

[8] http://www.anycast-testbed.nl

Our implementation of the framework for .nl conforms to both EU and Dutch laws and we reported it to the Netherlands Data Protection Authority. Our privacy board consists of a technical expert, a legal expert, and a member of our management team. They approved several privacy policies as of mid 2015, which we activated through PEPs.

6 Related Work

To the best of our knowledge, we are the first to propose a system that enables TLD operators to become a threat intelligence provider and increase the robustness of their DNS services. Scattered prior work on individual components does however exist.

The operator of the .uk TLD (United Kingdom) developed Turing[9], a system that appears to be similar to ENTRADA. Turing is however a commercial closed-source solution and there exists little publicly available information about its technical implementation. As far as we know, they did not extend their platform with functions to dynamically reconfigure name servers nor did they include privacy protection mechanisms. We are also unaware of deployments of Turing at TLD operators other than at the .uk operator.

There have been several research works that use DNS TLD data for detection of malicious domains, but not as part of a larger modular system such as our control plane. Hao et al. [12] analyzed the initial lookup behavior of malicious domains under .com and .org using a spam trap. Also, different methods exist to classify malicious websites. For example, Abbasi and Chen [13] present a comparison of tools to detect fake websites and perform content analysis to classify the websites.

The dynamic reconfiguration of DNS anycast networks is a technique that has been used to guide clients to the best node of a Content Distribution Network (CDN) in terms of network latency [6]. The topic has however not been explored before in the context of TLD operators, which need to support all networked applications that use the DNS and cannot assume that the roles of DNS operator and content provider are collocated as in the case of [6].

[9] http://nominet.uk/turing

7 Conclusions and Future Work

We presented a control plane for operators of Top-level Domains (TLDs) in the Domain Name System (DNS) that enables them to increase the security and stability of their TLD by becoming a threat intelligence provider. Our control plane is a system that extends a TLD operator's traditional services and leverages the DNS traffic and the domain registration transactions that a TLD operator handles. The control plane continuously stores and analyzes this data to automatically detect potential threats in the TLD and shares this information with other players in the TLD, such as hosting and access providers. It can also use the information to dynamically scale the TLD operator's DNS infrastructure, which increases the robustness of the TLD operator's DNS services.

Our control plane builds on the "ENTRADA" open source software, which we have developed on top of a Hadoop-based data storage cluster. ENTRADA enables TLD operators to easily feed their authoritative DNS traffic into the control plane, to run our threat detection modules, and to add their own. ENTRADA is currently being used by at least 6 operators of country code TLDs. It comes with a Privacy Framework that enables TLD operators to manage the personally identifiable information of TLD users.

Our future work consists of further refining and implementing the control plane, for instance in terms of modeling the DNS ecosystem using a variety of data sources, extend the control plane to other types of DNS operators, the interfaces a TLD operator needs to provide towards its DNS services, the impact of adding and removing nodes from a DNS anycast network, and new threat detection modules such as for the detection of booter sites.

Acknowledgements

We thank Kees Neggers (SIDN Supervisory Board), Aiko Pras (University of Twente), Marc Groeneweg (SIDN), Moritz Müller (SIDN), and the anonymous reviewers, who provided valuable feedback at various stages of this paper.

Ricardo de O. Schmidt's work is sponsored by the SAND project (http://www.sand-project.nl).

References

[1] "Phishing Activity Trends Report, 1st Quarter 2016", Anti-Phishing Working Group (APWG), May 2016, https://docs.apwg.org/reports/apwg_trends_report_q1_2016.pdf

[2] A. Ozgit, "DDoS attack on .tr", ICANN55 Tech Day, Marrakech, Morocco, March 2016, https://meetings.icann.org/en/marrakech55/schedule/mon-tech/presentation-ddos-07mar16-en.pdf

[3] G. C. M. Moura, R. de O. Schmidt, J. Heidemann, W. B. de Vries, M. Müller, L. Wei and C. Hesselman, "Anycast vs. DDoS: Evaluating the November 2015 Root DNS Event". In: Proceedings of the ACM Internet Measurement Conference (IMC 2016). Santa Monica, CA, USA. November 2016

[4] J. J. Cardoso de Santanna, R. M. van Rijswijk-Deij, R. J. Hofstede, A. Sperotto, M. Wierbosch, L. Zambenedetti Granville, and A. Pras. "Booters – An Analysis of DDoS-as-a-Service Attacks" In IFIP/IEEE International Symposium on Integrated Network Management (IM), 2015, 243-251

[5] A. Bar, A. Finamore, P. Casas, L. Golab, and M. Mellia, "Large-scale Network Traffic Monitoring with DBStream, a System for Rolling Big Data Analysis," in Big Data (Big Data), 2014 IEEE International Conference on, Oct 2014, pp. 165–170

[6] A. Flavel, P. Mani, D. Maltz, N. Holt, J. Liu, Y. Chen, and O. Surmachev, "FastRoute: A Scalable Load-Aware Anycast Routing Architecture for Modern CDNs", 12th USENIX Symposium on Networked Systems Design and Implementation (NSDI'15), Oakland, USA, May 2015

[7] M. Wullink, M. Müller, M. Davids, G. Moura, and C. Hesselman, "ENTRADA: Enabling DNS Big Data Applications", APWG Symposium on Electronic Crime Research (eCRIME 2016), Toronto, ON, Canada, June 2016

[8] C. Hesselman, J. Jansen, M. Wullink, K. Vink, and M. Simon, "A Privacy Framework for DNS Big Data Applications," Technical Report, Nov 2014, https://www.sidnlabs.nl/downloads/whitepapers/SIDN_Labs_Privacyraamwerk_Position_Paper_V1.4_ENG.pdf

[9] M. Wullink, G. Moura, and C. Hesselman, "ENTRADA: a High Performance Network Traffic Data Streaming Warehouse", IEEE/IFIP Network Operations and Management Symposium (NOMS16), Istanbul, Turkey, Apr 2016

[10] G. Moura, M. Müller, M. Wullink, and C. Hesselman, "nDEWS: a New Domains Early Warning System for TLDs", IEEE/IFIP International Workshop on Analytics for Network and Service Management (AnNet 2016), Istanbul, Turkey, Apr 2016

[11] R. Schmidt, J. Heidemann, and J.H. Kuipers, "Anycast Latency: How Many Sites Are Enough?", Tech Report ISI-TR-2016-708, May 2016, http://wwwhome.cs.utwente.nl/~schmidtr/docs/ISI-TR-2016-708.pdf

[12] S. Hao, N. Feamster, and R. Pandrangi, "Monitoring the Initial DNS Behavior of Malicious Domains," in Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, ser. IMC '11. New York, NY, USA: ACM, 2011, pp. 269–278.

[13] A. Abbasi and H. Chen, "A Comparison of Tools for Detecting Fake Websites," Computer, no. 10, pp. 78–86, 2009.

[14] S. Hao, A. Kantchelian, B. Miller, V. Paxson, and Feamster, " PREDATOR: Proac tive Recognition and Elimination of Domain Abuse at Time-Of-Registration," in Proceedings of the 2016 ACM CCS, October 2016.

[15] Y. Yu, D. Wessels, M. Larson, and L. Zhang, "Authority Server Selection in DNS Caching Resolvers," SIGCOMM Computer Communication Review, vol. 42, no. 2, pp. pp. 80–86, Mar. 2012.

[16] M. Korczynski, M. Krol, and M. van Eeten, "Zone Poisoning: The How and Where of Non-Secure DNS Dynamic Updates", ACM Internet Measurement Conference 2016, November 2016