A Demonstration of DNS^3: a Semantic-Aware DNS Service

Philippe Cudré-Mauroux[1], Gianluca Demartini[1], Djellel Eddine Difallah[1], Ahmed Elsayed Mostafa[1], Vincenzo Russo[2], and Matthew Thomas[2⋆]

[1] eXascale Infolab, University of Fribourg, Switzerland
{firstname.lastname}@unifr.ch

[2] VeriSign Inc., Fribourg, Switzerland
{vrusso,mthomas}@verisign.com

Abstract The Domain Name System (DNS) is a hierarchical and distributed database used to resolve domain names into IP addresses. The current Web infrastructure heavily relies on the DNS service to allow end-users to access Web pages and Web data using meaningful names (like "www.verisign.com") rather than cryptic sequences of numbers (e.g., "69.58.181.89"). The main functionalities of the DNS have been specified more than 25 years ago and have not fundamentally evolved since then. In this paper, we propose to demonstrate DNS^3, an extension of the current DNS service based on security mechanisms and semantic metadata. Specifically, we show how one can embed authoritative RDF triples using the current DNS protocol, and how the naming service can take advantage of the embedded semantic metadata to publish authoritative information about the domains, to improve the performance of domain resolution through prefetching, and to alert end-users of probable threats when visiting potentially harmful domains.

1 Introduction

The Domain Name System (DNS) is a key component of today's Internet ecology: it basically functions as the phone book of the Internet, enabling end-users to access Web pages using easy to remember and meaningful names (such as "www.verisign.com") rather than IP addresses (like "69.58.181.89"). The DNS also provides an important independence layer on the Internet, which allows to regroup various resources under a given name, or to change the physical configuration (e.g., the address) of servers without having to change their logical names.

While there has been a lot of discussion on which Top Level Domains to offer and on the policies used to administer or register domain names, there has been less effort to extend the functionalities of the DNS protocol itself. RFC 2671[3] provides however some interesting extension mechanisms, which are not widely used today beyond toy or proof-of-concept applications (e.g., applications[4] to

⋆ Authors are listed in alphabetical order.

[3] http://tools.ietf.org/html/rfc2671

[4] http://any.io/#twitter-dns

check Twitter statuses using the DNS). In this paper, we describe an extension of the current DNS called DNS^3 (standing for *Domain Name System with Security and Semantics*). DNS^3 takes advantage of the extensions mechanisms of the DNS and of Semantic Web technologies to minimize traffic and offer additional services to the client. More specifically, we use both the DNS Text Record ("DNS TXT") and new cryptographic features offered by the DNS ("DNSSEC") to provide:

DNS prefetching: when querying for one domain, DNS^3 also returns IP addresses for similar domain names that are often co-accessed (e.g., it might include the IP of "www.gmail.com" and of "images.google.com" and "calendar.google.com" when querying for "www.google.com").

Domain Suggestion: when querying for a non-existing domain, DNS^3 suggests potential domain names that are syntactically close to the domain name queried (e.g., it might suggest "www.unifr.ch" when querying for "www.unif.ch").

Malware Alerts: DNS^3 issues a warning when querying for a domain name that has been identified as a malware.

Authoritative Metadata: Finally, DNS^3 allows authoritative name servers to add metadata of their own choosing to their DNS responses.

Beyond the description of DNS^3, we also describe below a proof-of-concept end-to-end system that we plan to demonstrate.

2 System Architecture

Figure 1 gives a simplified graphical representation of our novel DNS architecture, which includes an RDF repository responsible for storing domain-related metadata as well as retrieving relevant triples at the time a DNS request is submitted. As an example of how one could take advantage of this architecture, we describe below the case of malware detection (other use cases are briefly described in Section 3).

There exist various methods to automatically detect malware domains[1]. One can for example identify malware domains by performing a Bayesian analysis on DNS activity logs [2]. In DNS^3, we take advantage of NXD log records (i.e., log records for non-existing domains) and latent semantic indexing to automatically detect and flag malware domains. On the DNS^3 server side, we load NXD log records in a high-dimensional space, and cluster domain names dynamically based on their request patterns using latent semantic indexing. We take advantage of latent semantic indexing to discover new malware domains from an existing list of dangerous domains [5]. The rationale here is that distributed attacks (such as botnet attacks) often use long lists of domain names to coordinate their attacks, and that we can thus discover new malware domains by searching for domains having request logs sharing some commonalities with those of known malware domains. Our experience shows that this is indeed the case in practice (we reach a precision close to 100% on malware detection using this technique with a recall of 68% in our experiments).

[5] taken from instance from `http://www.malwaredomains.com/`

As soon as malware domains are detected, we populate an RDF repository with new triples, storing the domain names along with a value reflecting the confidence of the detection results. Whenever a DNS request targets a malware domain, the server-side of DNS^3 automatically embeds semantic metadata in the TXT field of the DNS response to alert the client that the site he/she requests might be potentially harmful. In addition, we leverage the cryptographic capabilities of DNSSEC to ensure that only authoritative information is received by clients of the DNS^3 infrastructure[6]. With respect to scalability, preliminary experiments have shown no correlation between the TXT field size and DNS response time. Moreover, scalable approaches to latent semantic indexing exists, such as, e.g., Random Indexing.

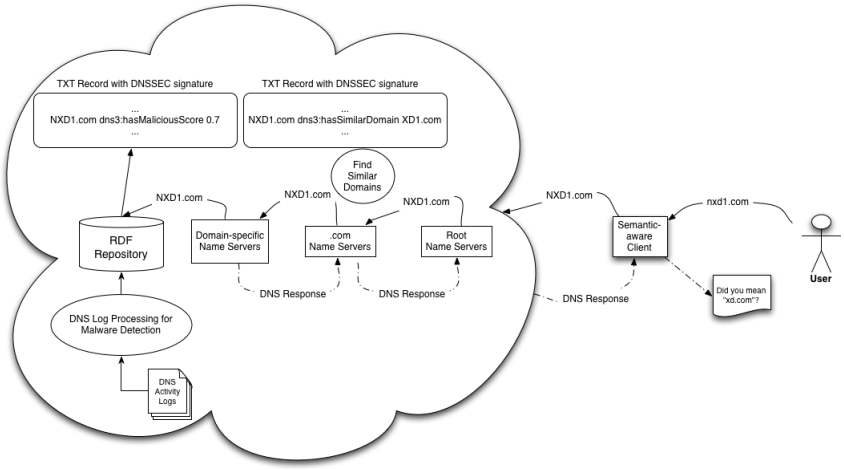

Figure 1. The process flow of DNS^3. First, a DNS resolution request is submitted to the infrastructure. Then, recursively, the request is forwarded to the responsible Name Server which queries the RDF Repository to retrieve relevant triples. The resulting metadata is embedded in the DNSSEC signed TXT record. Additional metadata can then be added to the response by Top Level Domain Name Servers. Finally, the response is returned and parsed by semantic-aware clients to enhance the user experience.

2.1 An Ontology for DNS Metadata

In order to be able to embed metadata in the TXT record of DNS responses, we need a standard language to express statements about internet domains, their IP addresses, and their relationships. To that end, we have defined a simple ontology that complies with the needs of the various use cases mentioned above[7].

[6] We use the method described by the IETF "DANE" working group: `https://datatracker.ietf.org/wg/dane/charter/`

[7] The ontology is available for download at: `http://diuf.unifr.ch/xi/dns3`

3 Demonstration

We have developed a fully functional end-to-end prototype of DNS^3. The server-side comprises a modified DNS deamon, a Java-based package to cluster domain names using latent semantic indexing, and a semantic repository running our dipLODocus[RDF] database system [3] to efficiently store and process triples. The server side continually clusters domain names, and dynamically embeds semantic metadata into DNS responses whenever necessary.

The client-side consists of semantic-aware clients. We have developed a Firefox extension to catch triples embedded by DNS^3 in DNS responses in order to enhance the browsing experience (in addition, we are also currently developing a smartphone client implementing the same functionalities). Whenever the client queries for a malware domain, the system automatically displays an alerts about the danger associated to the domain. A similar functionality has been implemented to suggest domain names when querying for non-existing domains. The semantic-aware client also takes advantage of domain prefetching: when querying for a domain, DNS^3 returns in addition a list of IP addresses for similar domain names (as suggested by our clustering method), thus drastically limiting the number of DNS requests that have to be performed. Our demonstration includes a small panel showing which domain names were prefetched.

Finally, we also demonstrate how arbitrary authoritative metadata can be securely served using DNS^3: when accessing special sites that have registered triples in addition to their IP addresses for their domain, the triples are supplied to the client using the DNS^3 infrastructure. In our demonstration, the triples are then displayed in a special panel and presented to the user.

4 Conclusions

DNS^3 is to the best of our knowledge the first DNS system to take advantage of semantic Web technologies to enhance the browsing experience. We have briefly described in this paper four use-cases built using semantic DNS functionalities. We are very excited by the first DNS^3 end-to-end prototype that we have built and hope that it will constitute a solid basis for further advances in the nascent domain of semantic DNS infrastructures.

5 Acknowledgment

This work is supported (in part) by the Swiss National Science Foundation under grant number PP00P2_128459.

References

1. M. Feily, A. Shahrestani, and S. Ramadass. A survey of botnet and botnet detection. In *Third International Conference on Emerging Security Information, Systems and Technologies, 2009. SECURWARE'09.*, pages 268–273. Ieee, 2009.
2. R. Villamarín-Salomón and J.C. Brustoloni. Bayesian bot detection based on DNS traffic similarity. In *ACM SAC*, pages 2035–2041. ACM, 2009.
3. Marcin Wylot, Jigé Pont, Mariusz Wisniewski, and Philippe Cudré-Mauroux. dipLODocus[RDF]–Short and Long-Tail RDF Analytics for Massive Webs of Data. In *international Semantic Web Conference (ISWC)*, 2011.