Towards a Passive DNS Monitoring System

Luca Deri, Lorenzo Luconi Trombacchi, Maurizio Martinelli, Daniele Vannozzi
IIT-CNR
Via Giuseppe Moruzzi 1, 56124 Pisa, Italy.
{ luca.deri, lorenzo.luconi, maurizio.martinelli, daniele.vannozzi}@iit.cnr.it

ABSTRACT
The domain name system (DNS) is a complex distributed database on which several Internet services rely on. As its monitoring is critical, researchers and internet service providers continuously monitor DNS traffic for identifying anomalies, measuring performance, and generating usage statistics.

This paper looks at DNS traffic from a different perspective; it covers the design and implementation of a passive DNS monitoring system whose goal is to understand trends, characterize economical relationships, and also track suspicious activities. The system described on this paper manages the .it country code Top Level Domain (ccTLD). Deployed on .it authoritative name servers, it is currently permanently monitoring all the .it DNS traffic.

Categories and Subject Descriptors
C.2.2 [**Computer-Communication Networks**]: Network Protocols—DNS; C.2.3 [**Network Operations**]: Network monitoring.

General Terms
Measurement, Economics.

Keywords
Domain name system, traffic measurement.

1. INTRODUCTION AND MOTIVATION
IIT-CNR manages the .it ccTLD [12] and the .it Registry since 1987, when IANA (Internet Assigned Number Authority) delegated it. In the past few years IIT-CNR has started a research project [3, 4] focusing on the design and implementation of a passive DNS [8, 9] monitoring system, whose aim is to analyze DNS traffic in order to understand Internet users trends and interests, and also track anomalous traffic pattern behaviors (e.g. DoS attempts and DNS attacks). Analysis of DNS traffic is a widespread activity, as this is one of the core protocols on which the Internet is relying. Nevertheless an area on which both the research community and TLD's seems not to have focused yet [10, 11], is the analysis of DNS traffic for understanding the evolution and trends of Internet users, similar to web search and traffic reports such as Google Zeitgeist [6] and Akamai State of the Internet [1]. Furthermore monitoring DNS activities allowed use to analyze relatively little traffic (the .it DNSs serve in total about 7 million requests/hour) when compared to complex application-protocols probes that instead need to decode a much larger traffic volume (not to mention that they are unable to analyze encrypted traffic) that needs to be diverted to probes by using network taps or span ports. It is worth to notice that the local law forbids techniques that divert unsolicited traffic towards specific monitoring systems; this restriction does not apply to DNS traffic as resolvers contact our servers where the monitoring system is deployed.

This has been the motivation for this work. Namely, we wanted to create a simple yet effective country-wide distributed monitoring platform able to passively monitor DNS traffic for the purpose of understanding trends and interests of a country, geo-locate Internet users, areas of digital-divide, as well suspicious activities. The used methodology required both the analysis of DNS packet payload and displacement of various software monitoring probes at the .it name servers. This has been done in order to have a comprehensive view of all queries performed for the .it domain. Results have been matched against the information records stored in the .it domain database. The outcome of this project, is a novel approach at DNS traffic analysis that is not limiting its scope to traffic volume and query type as most tool do, but also tries to obtain more detailed information about the evolution of a country, interesting domains and its trends.

SAC'12, March 25-29, 2012, Riva del Garda, Italy.

2. MEASUREMENT METHODOLOGY
As stated earlier on this paper, we decided to use a novel approach that is not just focusing on popular metrics (e.g. volume of queries, distribution of query types, response time) but rather tries to understand trends and changes in domain names that reflect the interests of Internet users with respect to specific ccTLD resources. As the authors are working at the .it ccTLD Registry, we have access to the complete .it domain list thus making our observation point unique. For this reason, we have decided to intersect the metrics obtained from DNS traffic analysis with data present in the domain registry database. This allowed us to better interpret some information (e.g. peak of queries for domains about to expire) and also generate unique statistics (e.g. the number of registered domains for which we observe queries and thus activity).

DNS traffic has been measured in three different locations, by analyzing traffic generated by some of .it DNS servers. The traffic has been mirrored towards a server that was running an enhanced version of nProbe [2], an open-source network traffic probe we developed. The probe generated traffic logs that have been analyzed using some home-grown applications. Measured data is then stored on a relational database to which the web console fetches information to display.

3. MEASUREMENT RESULTS
For privacy reasons and also because the collected information is focusing just on .it ccTLD, we decided not to report detailed

numbers on this section, but just to highlight the main results we have obtained. As stated before, this does not limit the scope of our work being the methodology we used pretty general. One of the goals behind this work is not just to understand the state of the Internet in Italy, but also to highlight trends and interests.

We used various techniques for tagging the content of .it internet domain. At the moment we have divided web sites in several macro-categories and for each category we daily report about domain queries. One of the results we would like to achieve in the long run, is to find out whether trends on specific topics we observe on the Internet also match the statistics which are published every year by official statistical companies and institutions.
Among supported metrics, we are measuring the (potential) economical value of domain registrars. In particular based on domain responses and not on domain registration records, we aggregate the number of domains that are resolved by specific DNS. This information is obtained from DNS responses, where the authoritative DNSs for the queried domains are returned. The HHI (Herfindahl-Hirschman Index) index [5], a widespread indicator of competition among companies, has been used to evaluate the concentration ratio of registrar. The result has confirmed that even if in Italy the first 10 registrars do register more than 60% of Italian domains (HHI national index of 1389 making this a not very competitive market), most of those registrars do not host the active domains, thus making the hosting market much more competitive than the registry market (the HHI index based on name servers is less than 200).
We have created a correlation between the searched domain type and the country/AS (Autonomous System) that has originated such query. This is because we want to answer questions such as "What are Germans mostly interested when searching for .it web sites?" and "How domain interests change for specific ISPs?". In addition to domains being queried, one of the questions we wanted to address is which AS path [7] is used to reach .it TLD DNS servers. This information is necessary for both understanding which are the best IXPs (Internet Exchange Point) for deploying anycast DNS servers, and also which ISPs and transit ASs are mostly used for reaching our DNS servers.

Other than measuring valid domain name requests, the monitoring system we developed also keeps track of requests for non-existent domains. This information is used for several purposes:

- Identify IP address that are scanning the domain registry for the purpose of creating a domain database.
- Potential sources of spam messages or computers affected by viruses.
- Misspelled domain names.

As we keep raw records of domain queries, we decided to classify them in two families: domain name scanners, and unknown domains that might have been misspelled by requestors. This classification is necessary as we have learnt that domain scanning is often a slow and distributed activity (e.g. on all ccTLD domain servers), that might be easily hidden when queries are originated from DNSs of large ISPs that issue million queries/day. For this purpose we have developed a component that exploits the Levenshtein distance for implementing this classification, and decide whether a certain non-existing domain is similar to an existing one (thus we assume that its name was misspelled) or not similar to any registered domain. In the latter case we assume this was a scanner, unless such domain will be registered within a short amount of time (less than a week) as the query might have been used to check for domain existence. A side result of this research is that we have realized that there are domains such as yahoo.it that get about 1/3 of requests of its abbreviation yhoo.it that is used as shortcut for Yahoo!'s Facebook page.

4. OPEN ISSUES AND FUTURE WORK

As we are monitoring only .it domain names, our work does not take into account all italian domains that have been registered under another TLD such as .com and .net. Although we acknowledge that this can be a limitation of our methodology, the local law does not allow us to place probes across the country to collect DNS traffic statistics on requests that are not sent to our servers.
We are planning to refine the mechanism used to tag a site with respect to its content. In particular we are considering to extend our system with a web crawler that could attempt to visit www.<registered domain>.it (note that not all registered domains have a web site) and based on the content of the site, tag it automatically according to the categories we defined. Doing this, we will greatly complement the information currently gathered by our system.

5. FINAL REMARKS

This paper presented a novel approach to DNS traffic analysis, whose goal is to understand the trends and interests of a country by analyzing queries to ccTLD domain servers. This work has demonstrated that although the DNS is not an ideal protocol when compared to HTTP, we have created a law-compliant solid monitoring system that is permanently monitoring .it DNS traffic and assisting daily activities at the italian DNS registry.

REFERENCES

[1] Akamai Technologies, State of the Internet: Q4 2010 Report, http://www.akamai.com/stateoftheinternet/, 2010.

[2] L. Deri, nProbe: an Open Source NetFlow Probe for Gigabit Networks, Proc. of Terena TNC Conference, 2003.

[3] A. Bonaccorsi et al., Internet Diffusion and Internet DomainsL Looking for a New Metric: the Case of Registrations by Italian Individuals, Proc. of ICWI Workshop, 2002.

[4] M. Martinelli et al., Measuring Internet Diffusion in Italy, Proc. of IFIP TC6/WG6.4 Workshop on Internet Technologies, 2002.

[5] A. Hirschman, The Paternity of an Index, The American Economic Review Vol. 54, No. 5, 1964.

[6] Google Inc., Google Zeitgeist 2010, http://www.google.com/zeitgeist, 2010

[7] R. Govindan and H. Tangmunarunkit, Heuristics for Internet Map Discovery, Proc. of ACM SIGCOMM, 2000.

[8] J. Postel, Domain Name System Structure and Delegation, RFC 1591, 1994.

[9] P. Mockapetris, Domain Names - Implementation and Specification, RFC 1035, 1987.

[10] The Measurement Factory, dnstop and dsc tools, http://dns.measurement-factory.com/tools/, 2006.

[11] P. Ren et al., Visualizing DNS Traffic, Proc. of VizSEC'06, 2006.

[12] N. Brownlee, DNS Root/gTLD Performance Measurements, Proc. of PAM Conference, 2001.