turing
by NOMINET

Decoding DNS data

Using DNS traffic analysis to identify cyber security threats, server misconfigurations and software bugs

The Domain Name System (DNS) is a core component of the Internet infrastructure, however there are currently no tools available for real-time analysis of DNS traffic. This means that significant users of DNS data (including registries, ISPs and large enterprises) do not have a clear picture of what is happening on their networks at any one time. The sheer volume of DNS data makes the job of real-time analysis especially challenging, which means that even advanced database systems such as Hadoop and Casandra are unable to collect and process data fast enough for real time analysis.

In order to better understand DNS traffic, Nominet has developed a real-time monitoring and analytics tool. *turing by Nominet* uses a patent pending metadata storage architecture that enables the analysis of terabytes of DNS traffic in seconds. *turing* can identify and locate a wide range of threats to the stability and security of the Internet, including abuse, misconfigurations and software bugs. In the future machine learning capabilities will be added to enable *turing* to automatically detect abnormal traffic.

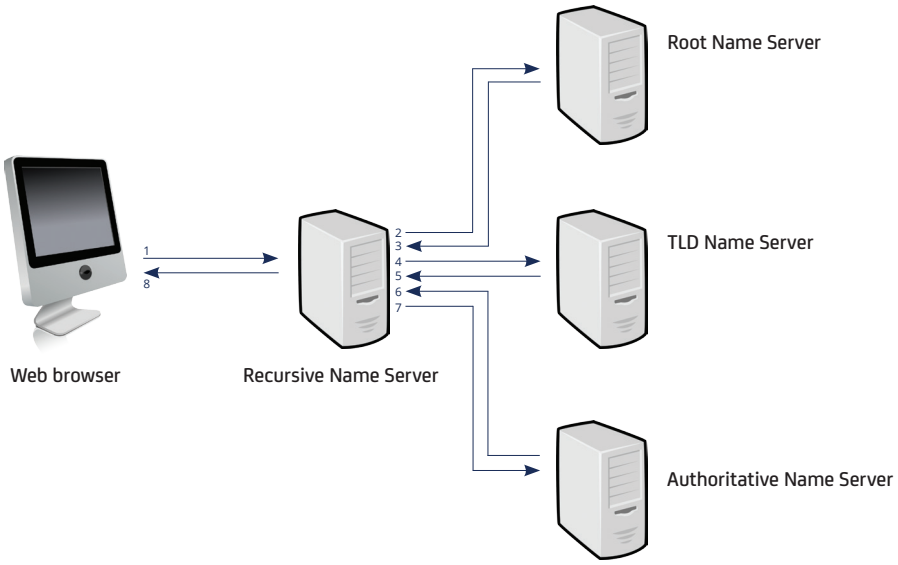
CONTENT

Background

The Domain Name System (DNS) is a core component of the infrastructure of the Internet. DNS is the Internet's equivalent of a telephone directory, storing the names of computers and their IP addresses. Every time someone uses a browser to look at a website or send email, their computer will use the DNS also. DNS also decouples the URL of a website from the physical location of the server that holds its content, meaning that web content can be moved between servers without impacting end users. DNS uses a globally distributed hierarchical tree structure, as illustrated in Figure 1 below. Responsibility for maintaining the DNS records for each level are delegated to various authorities. It is the job of the name server at each tier to accept and respond to DNS requests.

1. Web browser sends DNS lookup request for www.example.com to recursive name server
2. Recursive name server sends request for IP address of name server for .com to root name server
3. Root name server returns name server records of .com name server to recursive name server
4. Recursive name server sends request for IP address of name server for example.com to .com name server
5. .com name server returns name server records of authoritative name server for example.com to recursive name server
6. Recursive name server sends request for IP address of www.example.com to authoritative name server
7. Authoritative name server returns IP address of www.example.com to recursive name server
8. Recursive name server returns IP address of www.example.com (93.184.216.34) to web browser

Figure 1 - DNS architecture



The need for DNS monitoring

As a core element of the Internet's underlying infrastructure, the continuing and efficient operation of the DNS is crucial to any service that runs on the Internet. It is therefore prudent for companies who are responsible for managing Internet infrastructure, such as Internet registries, Internet service providers (ISPs) and large enterprises, to closely monitor the behaviour of DNS traffic. The information can be used to monitor problems, identify threats and configure systems for optimal performance.

Despite the importance of the DNS, historically it has been not been possible to actually monitor and analyse DNS traffic in real-time due to a lack of suitable tools. Without a current and detailed view of DNS traffic, companies have had a huge blind-spot that poses a significant risk to their business and the Internet.

There are around 4 billion queries to the .UK DNS servers each day, each consisting of both a request and response pair.

Existing DNS monitoring tools

A small number of DNS traffic visualisation tools do exist, however they are very basic, offering little more than monitoring of the performance of DNS servers, rather than analysing the traffic that passes through them. An alternative is to record raw traffic on an ad-hoc basic and then manually analyse it, but this approach is also limited as it is only able to look at a tiny snapshot of data at a time.

The key challenge in monitoring DNS traffic is the enormous volume of data that is generated.

For example, there are around 4 billion queries to the .uk DNS servers each day, each consisting of both a request and response pair. The scale and speed of the data poses challenges not only with data collection, but also with storage and analysis. Generic "off the shelf" database systems are simply not built to deal with this very specialist use-case. Even the latest NoSQL distributed database systems such as Hadoop and Cassandra are unable to process the data at the speeds required for real time collection and analysis.

turing overview

Introduction

In order to address the lack of suitable tool for monitoring DNS traffic in real time, Nominet has developed turing, a state of the art DNS analysis tool. The objective of turing was to create an easy to use, interactive, near real time analysis tool using off the shelf hardware and de-facto software standards.

To meet these criteria Nominet brought together a wide range of functionality, combining compressed storage and multidimensional time series with big data techniques, advanced statistics and an intuitive GUI. The result is a patent pending system that is years ahead of anything else on the market. turing is already helping the UK Internet to become more stable by identifying abuse, misconfigurations and software bugs.

Key features

Data collection

turing collects a wide range of DNS messages, including UDP, TCP, IPv4, IPv6 and EDNS packets. It stores all DNS messages in detail, correlating request and response messages at the point of capture. It also captures and stores advanced metadata such as the resolver's IP address, port number and latency so that specific queries can be accessed and analysed immediately or stored for later analysis.
turing's modular design enables the system to easily scale from monitoring a single server up to multiple servers. Additionally it can be configured for use on both recursive or authoritative DNS servers and runs on standard x86 hardware.

Data analysis

turing has a intuitive touch and gesture-based user interface that can display details of billions of DNS queries and then in milliseconds zoom down to one packet. The visualisation app is built using HTML5, meaning that it can be viewed in a web browser. Visualisations and reports can be filtered on 23 different attributes of the DNS messages and combinations of these, and also on query name and source IP address. An API is provided so that developers can integrate other systems including third party monitoring, visualisation and analysis tools.

Technical overview

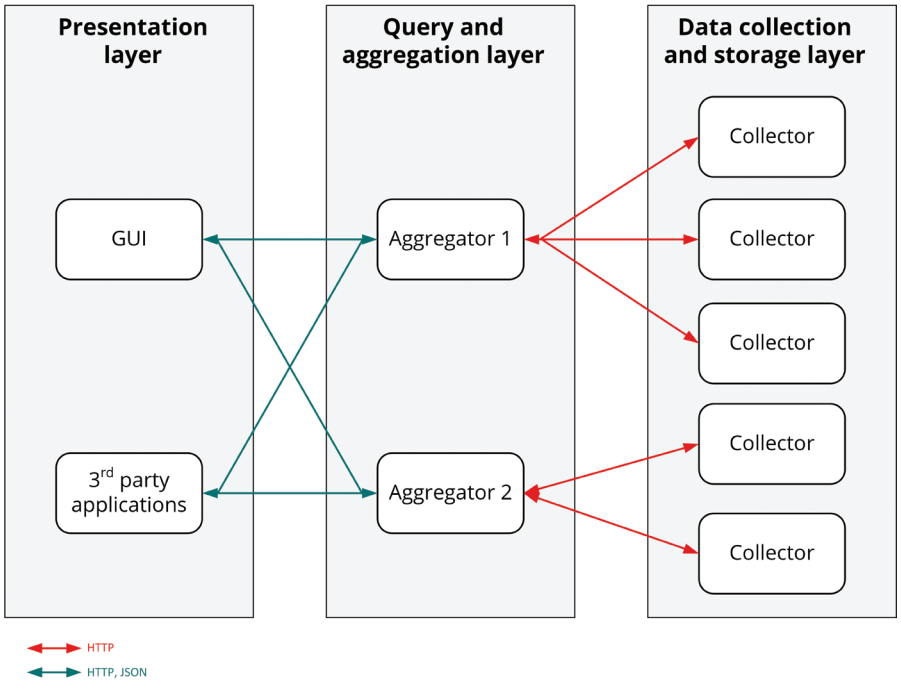
System Architecture / Overview

turing enables advanced DNS analysis by harvesting, collating and storing all DNS traffic for multiple selected name servers. This data is then presented in a front-end visualisation that is built using HTML5. This application communicates with a centralised data processing system, called the Aggregator, that in turn communicates with one or more remote collector nodes that collect the network traffic passing to and from the DNS servers being monitored, as illustrated in figure 2 below.

Data collection and storage

One of the major innovations in *turing* is its patent-pending data collection and storage system. *turing* stores all DNS messages in detail, correlating both request and response messages at the point of capture. It also captures and stores advanced metadata such as the sender's IP address, port number and latency so that specific queries can be accessed and analysed immediately or stored for later analysis. The packet data is stored together with a timestamp, which allows for extremely fast retrieval of packet data related to a specific time series. *turing*'s data collection and storage system consists of two main components: the Collectors, which collect DNS information from the network, and the Aggregator, which collates DNS metadata from each of the Collectors.

Figure 2 - System architecture



Data retrieval and analysis

Thanks to the unique method that *turing* uses to store data, through its distributed data collection and storage architecture, creation of a multi-dimensional signature and time stamped request / response pairs, it is able to retrieve and analyse terabytes of DNS traffic in seconds.

User interface

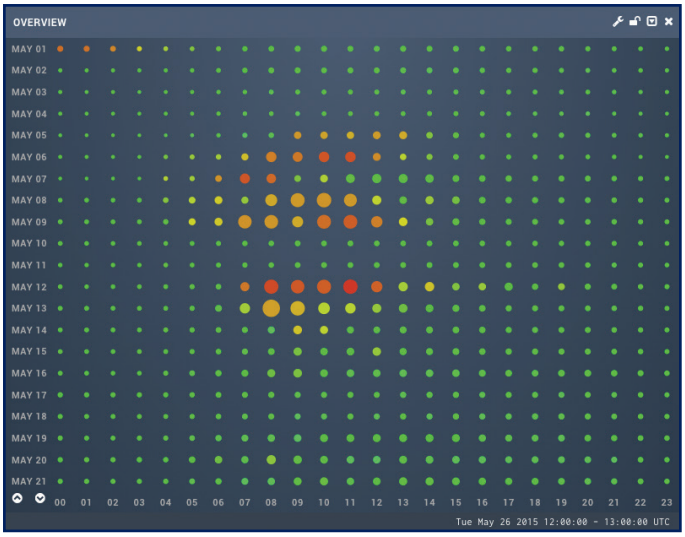
The user experience of *turing* is designed for touch control, allowing for filtering of huge datasets with single gestures. The *turing* front-end visualisation tool is an entirely browser-based HTML5 web application, which has the benefit of not requiring the user to install any specialist client software.

turing uses a RESTful API to enable the GUI to communicate with the database, as shown in figure 2.
This API is also available to third party monitoring, visualisation and analysis tools, enabling them to interact directly with the database using the same JSON over HTTP query interface as used by the GUI.

Using *turing* to find and investigate DNS issues

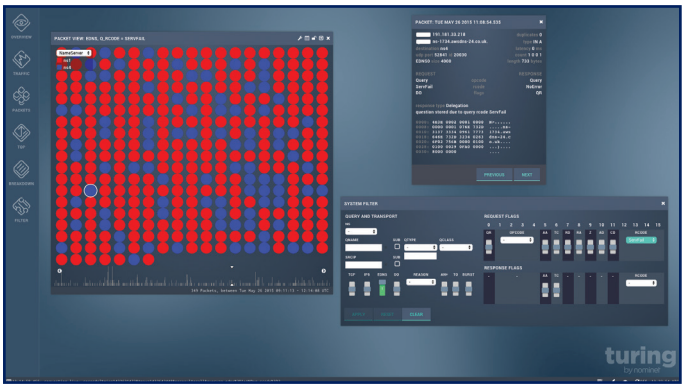
turing can be used both to investigate specific known problems and to identify abnormal patterns in DNS traffic which may signal new problems. The *turing* traffic overview is designed to make it easy for users to spot traffic abnormalities, by displaying the volume of traffic and the proportion of successful requests. The overview can display several weeks of DNS traffic, with each hour of the day represented by a coloured dot of varying size as shown in figure 3. The colour of each dot ranges from green, representing a high proportion of positive returned addresses, to red, representing a low proportion. The size of each dot represents the relative volume of traffic for that specific period, with a larger dot representing a larger amount of traffic.

Figure 3 - Overview of DNS traffic



If any anomalies are identified, the user can zoom in and view details of specific days or hours, enabling them to further pinpoint the issue. It is also possible to view the correlation between request types and responses, and to view any of the more than 20 variables that are recorded. The entire dataset can also be filtered by specific values, allowing them to be highlighted and further investigated. If the user already knows of a particular error they can jump straight to filtering the traffic for specific criteria, enabling them to view only traffic they are interested in, as illustrated
in figure 4.

Figure 4 – *turing* data filtered by a specific variable value



turing successes

As one of the world's largest Internet Registries, Nominet has been using *turing* to improve the health and security of the Internet. The following are just some examples of how it has helped.

Botnet identification

turing has been used to create a unique fingerprint that can quickly identify infection by a prevalent botnet. This was done by analysing spikes in MX records and zeroing in on the characteristics of non-resolving requests, identifying and diagnosing spam patterns. With the ability to identify such infections almost immediately, this discovery can drastically reduce the amount of spam across name servers.

Latency issues identification

turing is used to identify latency issues that can result in crippling transmission and processing delays. By identifying and analysing re-query traffic, Nominet has pinpointed issues with specific servers so that solutions could be found.

Long domain name bugs anderrors

turing has been used to isolate and analyse an increase in SERVFAIL responses, identifying that they were the result of non-protocol compliant, long domain names (255+ bytes). This was causing problems with Google's Public DNS and also highlighted a hidden bug within BIND. Both problems have subsequently been resolved.

Malware Index Case detection

Nominet used the processing power of *turing* to identify the Index Case of an aggressive piece of malware by tracking back from infected machines using the Domain Generation Algorithm. This has enabled the prediction and identification of subsequent infections, severely limiting the spread of this particular malicious software.

Man-in-the-middle attack detection

Through analysis of source ports, *turing* can identify Kaminsky-style blind spoofing/caching attacks, detecting when resolvers are not choosing ports at random. By identifying non-random resolvers, we have detected and prevented Man-in-the-middle attacks.

Future development

turing is already being used to help the Internet to become more stable by identifying abuse, misconfigurations and software bugs. As we continue to develop it, look out for a cloud hosted version (making it easier to start analysing your data), further integration with 3rd party management tools and a wealth of new features like spike detection, additional queries, smart heuristics and BGP monitoring.

turing by NOMINET

About Nominet

Nominet is a public purpose not-for-profit company that is responsible for running the .uk, name registry since 1996 and the .cymru and .wales domains from September 2014. This includes protecting, promoting and supporting the online presence of more than 10 million domain names and handling around 4 billion DNS requests each day. With over 18 years experience in running one of the busiest and most successful Internet registries in the world, Nominet is one of the leading authorities on DNS. Nominet works closely with Internet Corporation of Assigned Names and Numbers (ICANN) and the Internet Engineering Task Force (IETF) and is responsible for authoring the DNSSEC and other DNS standards.

Find out more:

If you would like any more information about turing please contact us:

nominet.uk/turing
turing@nominet.uk
+44 (0) 1865 332255