DNS and how it works?: Learning DNS by Cricket Liu

What you should learn:

- Why DNS was developed
- What the DNS namespace, domains, and domain names are
- What zones are and how they're related to delegation
- What resolvers and name servers are
- How resolvers and name servers cooperate in name resolution
- Types of name servers, including primaries, secondaries, caching- and authoritative- only and forwarders
- The syntax and purpose of the most common types of resource records
- How to use nslookup and dig, the most common DNS query tools

A brief history of the domain name system

- Before DNS (during ARPANET) there was HOSTS<TXT which was a single file mapping the name of all the ARPANET's hosts to their IP addresses
- HOSTS.TXT was maintained by a single organization and was distributed from a single point
- The hosts appear in no particular order
- Hosts would submit their portion of HOSTS.TXT to the NIC and there it would aggregated into a single file available via FTP

Problems with HOSTS.TXT

- Consistency/latency
- Traffic and load on SRI-NIC
- Collisions (had to search the table for naming collisions before naming the host)
- Searching the file was linear (O(n)) (took a long time since it wasn't sorted, sometimes the search would time out before reaching the file)

Goals for the design of DNS

- Local administration, regarding a partition of a database, but with global availability

- Hierarchical namespace (to allow easy partitioning of the database and ensure the uniqueness of hostnames)

What is the domain name system?

- The Internet's naming service
  - Responsible for mapping those funny dotted names you often see
    - In email addresses: event@infoblox.com (portion after @ is relevant to DNS)
    - In URLs: http://www.infoblox.com/ (portion after // and before / is relevant to DNS)
    - In IP addresses
- A distributed database
  - Partitioned into zones
  - Which are managed in a distributed, decentralized way
  - But which are available globally, across the Internet
  - Organizations can manage their zone almost autonomously yet have that information available globally

- DNS like other databases typically have indexes
- DNS's indexes are those funny dotted names, which are properly called domain names
- Domain names are actually paths in an inverted tree

Inverted Trees

- An inverted tree is made up of nodes and links between nodes
- Each node is connected to its parents by a single link
- One node can be parent to arbitrarily many children
- Each node has a label, between 0 and 63 bytes in length
- The root node (at the top) has a special, reserved label, written ""
- Otherwise, the main restriction is that all of the children of a node have different labels

Domain Names

- Clearly a label isn't sufficient to uniquely identify a node in the tree
- To do that, you need the node's domain name: the series of labels from the node to the root, read with dots separating the labels

How to read a domain name

Domains, Delegation, and Zones

- Domains are groups of related nodes
    - Technically, they're a subtree of the namespace
- Domains are identified by a domain name- the domain name of the node at the apex of the subtree
- You can think of the com domain as all domain names in the name space that end in com

Subdomains

- Subdomains are domains contained within another domain
    - The apex of the subdomain is within the parent domain
    - In terms of their domain names, the subdomain ends in the name of its parent (ex: hp.com hp as a subdomain ends in com the domain)
    - Infoblox.com is a subdomain of com because it ends in com

Domain Management

- An organization typically manages a domain
    - For example, ICANN controls the root domain
    - Verisign controls the com domain
    - But both ICAAN and Verisign do not control all the nodes below com

Delegation

- Delegation allows an organization to assign control of a subdomain to another organization
- The parent now contains pointers to the sources of data in the subdomain
- The parent and child are now separate administrative containers, separate zones

The Internet's Namespace: The Early Years

- Initially, the Internet's namespace was simple:
- 7 top-level domains
    - com, for commercial organizations (hp.com)

  - edu, for educational organizations (berkeley.edu)
  - gov, for US governmental organization (nasa.gov)
  - int, for international treaty organizations (nato.int)
  - mil, for US military organizations, such as the US air force (af.mil)
  - net, for networking organizations, such as NSFNET (nsf.net)
  - org, for non-commercial organizations, such as the Internet System Consortium
- These are now known as Generic Top Level Domains (GTLD)

The Namespace Expands

- Later, top-level domains were reserved for every country in the world, called country code top-level domains, or ccTLDs
- For example, Australia's, au, China's cn, Frace's fr, Germany's de
- ICANN later allowed the addition of more top-level domains
- And quite recently, opened up the market for top-level domains to all comers

DNS 101 by verisign (https://www.verisign.com/assets/DNS101.pdf)

- The Domain Name System (DNS) ensures the availability Web sites, email and Web systems by mapping domain names to Internet Protocol (IP) addresses, represented as a series of numbers and letter
- The DNS uses specialized servers to translate names such as www.verisign.com into IP addresses that allow data and information to reach its destination
- This process, called DNS resolution, allows people to type more memorable domain names into a browser to reach Web sites and send email messages
- Today there are nearly 300 million registered domains. Domain names allow people to organize, navigate, and understand the Web. They provide a literal address that directs Internet users to the area of the Web to which they want to go.
- There are a few different parts of what we consider a "domain name" described below:
  - www
    - Third-level domain: Also called the subdomain, this is the portion of the domain name that appears before the second-level domain name. The most common third-level domain name is www, but can take many other forms, for example, blogs.verisign.com

        - Second-level domain: This is the unique part of the domain name that appears to the immediate left of the TLD. People register for the second level domain to differentiate themselves or their offering from other sites
        - Top-level domain: Top Level Domains (TLDs) are the highest level of organization on the Web. There are typically two kinds: Generic TLDs such as .com, .net, and .org; and Country Code TLDs, two-letter codes approved for use by specific regions, such as .uk, .au, and, .de
- How the DNS works: The DNS uses the following steps to map domain names to IP addresses, allowing people to search for websites and send email using familiar names instead of strings of numbers and letters. The process of translating a domain name into an IP address is called DNS resolution
    - Step 1: You type a domain or web address into a browser. What your browser does is send a message to the network asking for help (this is called a query)
    - Step 2: Your computer queries (contacts) one of the machines that your ISP gave to your computer, called recursive resolvers, which should either have the IP address cached, or be able to go out and "recursively" find it
    - Step 3: If your ISP's recursive resolvers don't have the address, they query the DNS root name servers for the IP address
    - Step 4: The root name servers direct (or "refer") your ISP's recursive resolver to appropriate TLD name servers by examining the top level domain
    - Step 5: Each TLD has its own set of name servers, and after the resolver asks them for the IP address, they refer it to another (more appropriate) set of authoritative DNS servers by reviewing the second level domain of the query
    - Step 6: Your ISP's recursive resolver then queries the referred authoritatively DNS name servers for the IP address. Each domain has an assigned set of authoritative DNS name servers that are responsible for knowing everything about the domain, including the IP addresses
    - Step 7: Your ISP's recursive resolver retrieves the A record(which is the DNS record for mapping IP addresses) for the web site you are looking for from the authoritative name server and stores the record in its local cache in case anyone else queries it
    - Step 8: Finally, your ISP's recursive server returns the A record to your computer, which reads and passes the IP address to your browser. The browser then opens a connection to

the web site. The entire process generally happens in a few tenths of a second and is transparent to the end user

Resources:

https://learning.oreilly.com/videos/learning-dns/9781771373692/

https://learning.oreilly.com/library/view/managing-mission-/9781789135077/

https://learning.oreilly.com/library/view/learning-coredns/9781492047957/

https://learning.oreilly.com/library/view/pro-dns-and/9781430230489/

Textbook used for Internet section