DNS RESOLVER BEHAVIORS

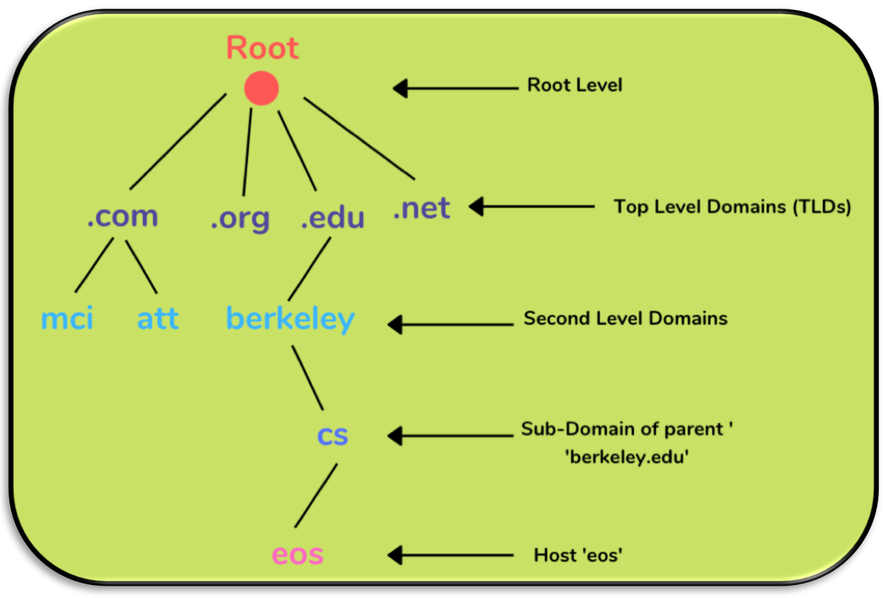
Alexandra Fernandez, Jelena Mirkovic

Overview:

- **The Problem**: DNS resolvers sometimes behave in unexpected ways – sending lots of queries and sending malformed queries. There is a lot of research on the various aspects of the Domain Name System (DNS) but little research on understanding the range of weird resolver behaviors.
- **The Challenge:** The complexity of the DNS ecosystem makes it difficult to enumerate behaviors and to establish the root cause of anomalous behaviors.
- **Our Approach:** We focused on analyzing behaviors of large senders and looking for dominant malformed query patterns, looking to quantify and dissect these behaviors.

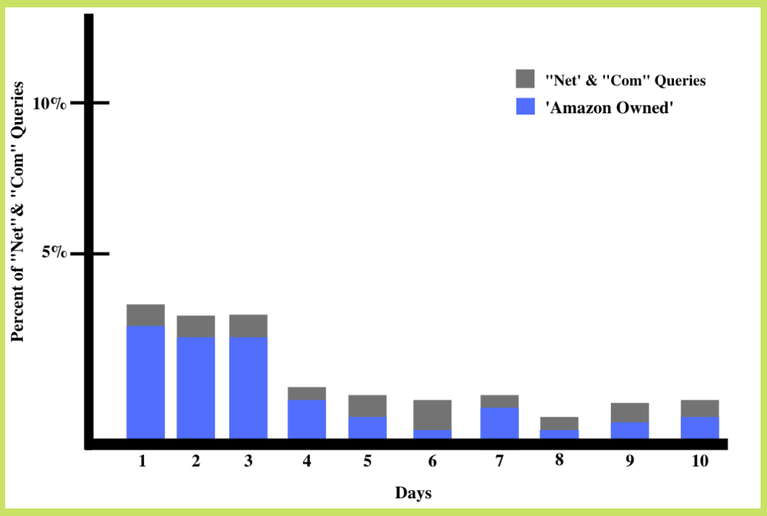
What is DNS?

- The Domain Name System (DNS) is the Internet's system for converting alphabetical, human-readable, names into numeric IP Addresses that a computer can understand.
- This system is crucial and is invoked whenever a user wants to access a remote server (e.g, visit a URL, SSH into a server, run a phone app that reports to a server, play a multiplayer video game, etc.)
- Sometimes this system is also used to direct Internet traffic to the nearest server to minimize user-experienced latency.



Current Results:

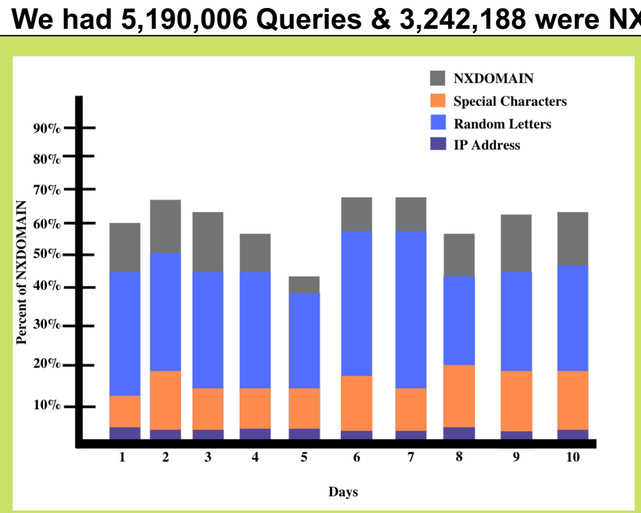
- **Large Sender Analysis:** We investigated 10 days of B root data from 2017-2019. One interesting pattern we found was that the top 6 senders for “.net” and “.com" were from Amazon cloud. They sent almost exclusively the same two queries: “.net” and “.com” without any other domain name info.
  - **Implications:** These queries are likely coming from rented cloud machines that may be misconfigured or used for malicious purpose.



- **Malformed Query Analysis:** We identified TLDs present at the B-root server and identified queries that do not contain a valid TLD and thus will result in NXDOMAIN (no such domain) replies from the server. We then investigated patterns in these malformed queries.
- **Discovery:** Different query patterns such as:
  a) Queries with Special Characters
  b) Queries made up of just Random Letters
  c) Queries made up of IP Addresses

  Example: 2019 Morning Data

  We had 5,190,006 Queries & 3,242,188 were NXDOMAIN (62%)



- **Machine Learning Clustering of Queries:** We used an Sklearn Clustering model on a DITL data file. The model clustered all IP Sources that queried above a certain threshold into different clusters. Our model then gave us 44 different clusters, where each clusters queries were similar.

Cluster Number	Cluster Size (IP)	Cluster Size (Queries)	Cluster Rate	Owners of IPs	Sample Query
zero	681	887,072	40.1%	Asia Pacific Network Information Centre	“fxkxclnvbx”
twenty nine	117	34,841	1.6%	Google	“nonexist-9d9c360e.node.consul”
two	36	5,455	.25%	Asia Pacific Network Information Centre	“.”
seven	25	16,047	.74%	Latin American and Caribbean IP address Regional Registry	“cyghdsgxch.local”
three	20	3,608	.17%	RIPE Network Coordination Centre	“dr._dns-sd._udp.31.”
thirty nine	12	3,205	.15%	Namecheap, Inc.	“localhost”
one	12	5,222	.24%	RIPE Network Coordination Centre	“kvqqjiztatdhymi”
twenty	9	827	.04%	Charter Communications, Comcast Cable Communications	“com”
nine	8	104,599	4.8%	Amazon Data Services Ireland Limited	“net”
eight	7	2,846	.13%	RIPE Network Coordination Centre	“icyvuvjmmny”

Table 1: 10 Largest Clusters (Largest to Smallest)

Future Works:

- Distinguish when a resolver is being used with malicious intent versus being misconfigured.

If interested contact: Alexandra Fernandez
amfernandez@loyola.edu
Work performed under REU Site program supported by NSF grant #1659886

USC Viterbi School of Engineering
Information Sciences Institute