CS 494/594 Homework 2 (Winter 2022)

Instructor: Dr. Nirupama Bulusu **Due Date: 02/01/2022**

594 Students:

Besides the homework problems below, please review **only one** of the following two papers. You can choose.

- [Adhikari12] Adhikari, V. K., Guo, Y., Hao, F., Varvello, M., Hilt, V., Steiner, M., & Zhang, Z. L. (2012). **Unreeling netflix: Understanding and improving multi-CDN movie delivery.** In Proceedings of IEEE INFOCOM 2012 (pp. 1620-1628).
- [Halvorson15] **From .academy to .zone: An Analysis of the New TLD Land Rush**, Tristan Halvorson, Matthew F. Der, Ian Foster, Stefan Savage, Lawrence K. Saul, and Geoffrey M. Voelker, *Proceedings of the ACM Internet Measurement Conference*, Tokyo, Japan, October 2015.
  URL: http://www.sysnet.ucsd.edu/~voelker/pubs/newtlds-imc15.pdf

FROM .ACADEMY TO .ZONE: AN ANALYSIS OF THE NEW TLD LAND RUSH (Reviewed by Parth Parashar)

This paper, basically, tries to analyze the different types of domain registrations in the new TLDs (Top Level Domains) with a view to determine the registrant behaviour in this new world of naming abundance and complexity. The authors also try their hand at examining the cost structures and monetization models for the new TLDs, with an aim of identifying which registries are more profitable as compared to others. The financial analysis carried out by the authors suggests that only half of the total registries have earned enough to cover their application fees.

In today's ultra-modern and constantly-evolving world, each and every successful company needs a good Internet presence, and most of these companies rate a memorable domain as an integral part of this presence. This particular paper aims at identifying the overall impact of the New gTLD (Generic Top-Level Domain) Program on the domain name ecosystem. But, with the advent of hundreds of new TLDs, the authors expect many smaller companies to find it infeasible and implausible to defend their name in each TLD. Also, such a sharp surge in simple-word second-level domains could make it quite difficult for speculators to resell even simple, desirable names.

The following contributions are made by the authors in this paper:

a. They classify registration intent on the basis of a methodology derived from their previous work on xxx, and the major contribution includes a timely result of the above methodology applied to the TLD landscape during the current period of swift and sharp expansion.
b. The authors also determine the program's impact on the old TLDs, both on the registration rates as well as on the types of registrations.

c. Finally, the authors examine registry profitability to learn about the places where the registration money goes and also about the different kinds of TLDs that get the most registrations.

The authors' findings suggest that the new gTLDs have yet to provide the required value to the Internet community in the same way as legacy TLDs do. The findings also suggest the fact that the new TLDs have yet to have a significant impact on the old TLDs. Currently, "com" continues to dominate the Internet domain name registration activity.

There are three important actors in the DNS (Domain Name System) ecosystem:

a. Registries: They operate the TLDs and have a contract with ICANN (Internet Corporation for Assigned Names and Numbers) for each TLD.
b. Registrars: They sell domain names in many different TLDs, and also have an ICANN accreditation.
c. Registrants: They are the entities that buy domain names.

The basic goal of the authors, through this paper, is to explore how the above actors have reacted to the rapid expansion of the DNS name space. The emergence of many new TLDs (managed by unique registries), various zone files and numerous reports from ICANN, form the basis of the data in this particular paper.

Each domain is assigned to one of these seven categories: No DNS, HTTP Error, Parked, Unused, Free, Defensive Redirect and Content. Registration intent can be classified into one of these three categories: Defensive (These registrants purchased a new domain to defend an existing Web presence), Primary (These registrants own domains with the intent of establishing a Web presence) and Speculative (These registrants have the intention of profiting off of the name itself and never plan to develop a meaningful Web presence.

There are certain limitations with the model prescribed by the authors in this paper:

a. The pricing model does not include premium domain name sales.
b. For any given TLD / registrar pair, the authors could only record a single price, when domain name prices changed over time. This particular limitation could be a part of future study.
c. The authors estimated wholesale prices as 70% of the lowest price for domains in the TLD.

The authors conclude that new gTLDs, while gathering significant revenue for registrars, have yet to provide a reasonable value to the Internet community when compared to legacy TLDs.

1. (20 points) Peer-to-Peer File Distribution

Ben Bitdiddle and Alyssa Hacker launch rival file distribution companies. Ben chooses a client-server architecture for his file distribution company. Alyssa chooses a peer-to-peer architecture, in which both the server and the peers can participate in the file upload. Alyssa notices that while all the peers choose to download, only 10% of the peers actively upload, with the other 90% being free-riders. Ben's company has 1 million users and Alyssa's company has 10 million users, with each user having an upload rate of 100 kbps and a download rate of 5 Mbps. Both Ben and Alyssa have servers with upload rates of 100 Gbps. Assume all file sizes are 1 GBytes. Unit conversions: 1 Gbps = 10^9 bps, 1 Mbps = 10^6 bps, 1 kbps = 10^3 bps, 1 GByte = $(1024)^3$ bytes

(i) Whose company, Ben or Alyssa's will have the minimum distribution time to distribute a file to all clients? Show your calculations. (ii) What would the minimum distribution time be with no free riders in Alyssa's company?

2. *(10 points) Distributed Hash Table* After a year of running her peer-to-peer file distribution company, Alyssa notices that file lookup through her central server is becoming extremely slow. Alyssa reads a research paper on distributed hash tables and decides to emulate the functionality of a BitTorrent tracker using a distributed hash table. After learning how knowledgeable you are about DHTs, she hires you as a consultant to design and implement a DHT tracker for her. How would you design the DHT? Who would be the peers? What would be the keys, and the value stored for each key? How would you bootstrap the system?

3. *(10 points) DNS*

   Alyssa Hacker thinks that DNS works by having each DNS server forward a query it can't answer to its parent until the query lands at a server that can answer it. Ben Bitdiddle disagrees with her and shows her how DNS actually works. Describe how Alyssa's version of DNS differs from the way DNS actually works. Given DNS_{Alyssa}, a DNS system that works as Alyssa described it, and DNS_{Ben}, an actual DNS system, which is better: DNS_{Ben} or DNS_{Alyssa}? Justify your answer.

4. *(20 points) Web and HTTP*

Alyssa Hacker's company has a development office in Beaverton, Oregon. It consists of a 1 Gbps company local area network (LAN) connected to the Internet via an access link of 100 Mbps. Suppose that the average object size is 1,000,000 bits and that the average request rate from the company's browsers to the origin servers is 90 requests per second. Also suppose that the amount of time it takes from when the router on the Internet side of the access link forwards an HTTP request until it receives the response is three seconds on average. Model *the total average response time* as the sum of the average access delay (that is, the delay from Internet router to company router) and the average Internet delay. For the average access delay, use $\Delta/(1 - \Delta\beta)$, where Δ is the average time required to send an object over the access link and β is the arrival rate of objects to the access link.

(a) (10 points) What is the total average response time?

(b) (10 points) Alyssa decides to install a cache in the company LAN to improve the response time. Suppose the cache hit rate is 1/3. What is the total average response time?

5. *(20 points) Wireshark Labs: HTTP*

6. (20 points) Wireshark Labs: DNS