Revisiting the Self-Similar Nature of Web Traffic with Modern Webpages and Browsers

Nida Tanveer — **University of Wisconsin**
Sarah Tanveer — **University of Wisconsin**

Motivation and Objectives

- Understanding the **behaviour of network traffic** leads us to making **design decisions** about the Internet
- The **complexity of web pages and web browsers are different** now than when studies about self-similarity were conducted
- analyze how factors like **prefetching, and local caching** impact the **self-similarity of web traffic** in the context of modern browsers and webpages

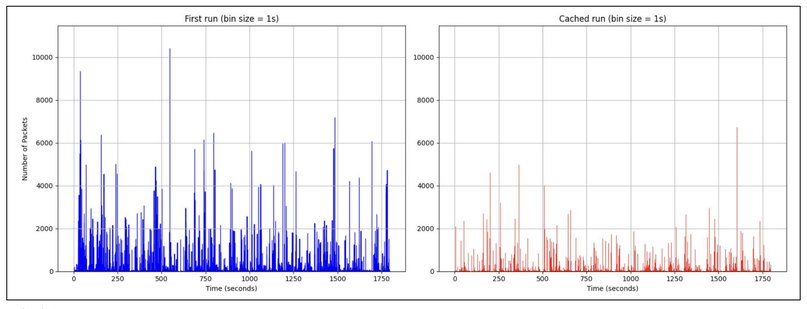
Related Work

- **On the Self-Similar Nature of Ethernet Traffic:**
  - Using large-scale measurements, this study shows the self-similar nature of Ethernet Traffic through graphical and mathematical analysis
- **Generating Representative Web Workloads for Network and Server Performance Evaluation:**
  - This work describes the characteristics of "user equivalents" and how the On/Off Model can be used to mimic real web traffic
- **Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes** and **The Network Effects of Prefetching:**
  - These papers identify prefetching and local caching to be factors that impact self-similarity

Methodology

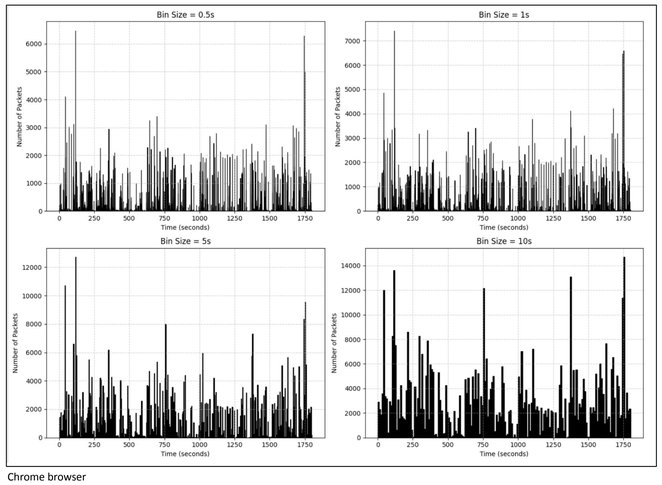
- Using **Selenium**, we **simulate a user** based on the **On/Off model**
- We run **three concurrent users** browsing websites on **three different browsers**
- We capture network traffic using **tcpdump**
- **Caching:** We randomly browse our pages, and **save cache in user profiles**. We repeat the experiment but using only the pages in the user's history
- **Prefetching:** We **turn off prefetching on Firefox** by setting user preferences on Selenium
- We use graphical **R/S analysis** to estimate the Hurst parameter of our collected packet traces.

First Run Vs. Cached



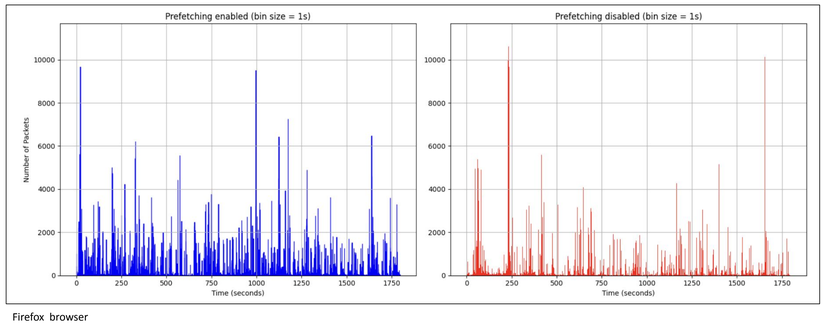
Edge browser

Pictorial "proof" of self-similarity



Chrome browser

Prefetching Enabled Vs. Disabled



Firefox browser

Data

- We compare data on these parameters:
  - Three different browsers using their default settings
  - Day vs night to capture the diurnal pattern of traffic
  - First visit vs Cached
  - Prefetching enabled vs disabled

Key Results

- For all browser types, and runs, the value of the **Hurst parameter > 0.5**, which indicates that our time series data **exhibits long-range dependencies** and therefore is self-similar.
- For Chrome, Edge and Firefox the average value of the estimated Hurst parameter is 0.783, 0.816 and 0.817 respectively.
- The file sizes of the packet traces of the second run (cached) are on average smaller than that of the fresh run (11-42% smaller).
- Burstiness persists, but on the second/cached run, we see that the packet frequencies are much lower.
- We did not observe a noticeable difference between the packet traces for prefetched enabled vs disabled

Conclusion and Future Work

- From our analysis, we conclude that web traffic has remained self-similar
- At this point our study is relatively small-scale so in the future we would like to experiment with more users and a larger variety of websites.
- Since our data is collected from a set of simulated users, we would like to extend our work to using data gathered from real users.
- We would also like to explore more browser configurations such as prefetching and pre-rendering settings.