DISSERTATION DEFENSE

Ayush Goel

Overcoming Barriers to Information Exchange on the Web

Monday, September 25, 2023
1:00pm – 3:00pm
3725 Beyster
Hybrid – Zoom

ABSTRACT: The advancement of the human species relies on its ability to communicate complex information and ideas with each other. In the modern era, such exchange of information has been expedited using digital mediums, with the world wide web (WWW) being one of the most used platforms. Despite its ubiquitous use, the web today suffers from two main problems: 1) Poor performance, in terms of page load latency perceived by users and the rate at which pages can be crawled at scale. 2) High ephemerality resulting in link rot and necessitating massive web archiving initiatives.

In this dissertation, I have built a number of systems to combat both the performance and ephemerality issues of the modern web by leveraging trends in web pages and exploiting them with the help of fine-grained program analysis of web computations. First, I proposed two separate systems that seek to reduce the total client-side computation for end-user page loads: memoizing JavaScript execution and automatically parallelizing JavaScript execution across different cores of mobile devices. Second, I built Sprinter, a general-purpose web crawler that carefully combines browser-based and browserless crawling to significantly increase the crawling throughput while ensuring near-perfect fidelity of resource fetches. Third, I built Jawa, a web crawler that downloads pages specifically for web archives by exploiting the differences between live and archived pages in order to reduce their storage overhead and eliminate any fidelity issues incurred while loading the archived pages.

CHAIR: Prof. Atul Prakash and Prof. Harsha V. Madhyastha