

Student Statement: Jessie W.H. Zou Memorial Award for Excellence in Undergraduate Research
Zhucheng (Michael) Tu

Over my last three academic terms, I worked with Professor Jimmy Lin on a project, named Prizm, that enables a Raspberry Pi to be used as a device for monitoring and analyzing personal web activities. While citizens of today spend a significant amount of time using the web, we have little insights into questions such as how we distribute our time and network resources among the web sites and applications we use, where we get answers from for each of the queries we search for, and how often we visit the same web resource. While there has been an emergence of devices and apps for tracking personal activities, from monitoring fitness levels to managing diet, there currently does not exist an agent to analyze all of our web activities comprehensively – one that is capable of monitoring web activities initiated from across multiple devices and multiple clients. Prizm is a prototype that aims to fill this gap by getting a Raspberry Pi to behave as a substitute for the home wireless router, so that when devices use Prizm as their personal wireless access point the software can log the activities and provide subsequent analyses that answer the questions we posed. Although the initial goal of the software is limited to the analysis of web activities, the infrastructure for this task can be reused to support even more ambitious goals of personal web archiving, offline web usage, and offline search.

Prizm works by placing an HTTP proxy between two network interfaces that can log the HTTP requests passing through. One network interface is for the subnet that contains devices connected to the Prizm wireless access point while the other network interface ultimately connects to the internet. A hook in the proxy can send metadata about the request and response to a message queue, which gets consumed by workers that persist the data to a database. Getting this kind of architecture running on a Raspberry Pi together with the wireless access point is quite novel, with many software engineering challenges involving reliability and performance. Users can then query the data using SQL or visualize it on a web portal. We were able to use Prizm under this set-up as a complete substitute for the normal home router for using the web, with the added benefit of collecting data for personal web analytics.

This research project started out as an undergraduate research project and then evolved into a Software Engineering Capstone Design project. During the 2017 Software Engineering Design Symposium, judges Professor Joanne Atlee and Professor Mahesh Tripunitara were impressed by the ambition of the project. The project also generated enthusiasm from the chair of the Lifelogging Tools and Application Workshop, Dr. Cathal Gurrin. I will be pursuing a Masters of Mathematics at the University of Waterloo to continue working on Prizm to realize more of its ambitions alongside other research in information retrieval.