

Internship proposal 2016/2017

Topic: Learning user context from the observation of network and systems data

Duration: 4 to 6 months

Hosting team: Muse, Inria Paris (<https://team.inria.fr/muse/>)

Apply at: <https://team.inria.fr/muse/join-us-2/>

Mentor:

Renata Teixeira, Senior Researcher, Inria (head of Muse team)

Keywords: Internet measurements, network traffic, HCI, data analysis, data visualization

Description:

Online services and applications are collecting more and more personal data about users to adapt and optimize their services to each individual user. Examples of service optimization range from your personalized Facebook feed to Google maps displaying the area around you. This personalization can take a wide range of inputs from the Web pages a user visits and the films she watches to the places she visits. We are particularly interested in services that adapt to a user's context (e.g., the geographical location in the Google maps case). User context is fairly broad; it can encompass, e.g., the location, the time of day, who the user is with, what the user is doing. Our goal is to be able to infer user context only by observing technical data that we can collect automatically by instrumenting the user's device or the network that carries the user's traffic. To better understand the issues involved in inferring user context, in the European project User-Centric Networking, we have conducted a user study that simultaneously collected both technical and ethnographic data from a population of 25 users across the UK and France. The technical data includes full packet traces, location information, and the set of applications used on each participant's devices. The user observations adopted an ethnographic approach to detail each participant's online activities, routines, and exceptions, as well as their reasoning for conducting certain activities at certain times, locations, or with certain devices. Our initial analysis of this data highlights the gap between the user's reasoning and what we can capture with network and application level statistics alone.

The goal of this internship is two-fold: (i) demonstrate and properly evidence the gap between user reasoning and technical data; and (ii) develop methods to bridge this gap by analyzing the technical data to infer user context. To demonstrate the gap, the student will analyze the transcripts made by the ethnographers after the discussions with users to identify the vocabulary and ways in which users talk about their activities. This analysis will allow us to compare the user reasoning with the technical data we can gather from users. The student will then develop methods to process the technical data to extract the user context. Our manual analysis of the data identified a few data types that help capture context: location data, visited websites, and time of day. To make sense out of the data, however, we had to conduct a fair amount of manual search and inference. The goal is that the student will identify methods to automate this process.

The student should develop scientific skills on network measurements and data analysis. In particular, we use and evaluate analysis methods from natural language processing and time series analysis. If the student is interested, there is a possibility of staying for the doctoral studies after the internship.

Desirable skills:

- Comfortable communicating in English
- Knowledge of data analysis techniques
- Knowledge of network traffic measurements
- Knowledge of matlab or gnu R

References:

P. Brundell, A. Crabtree, R. Mortier, T. Rodden, P. Tennent, P. Tolmie. **The Network from Above and Below**. In Proc. of the ACM SIGCOMM W-MUST workshop, 2011. <http://mor1.github.io/publications/pdf/wmust11-homework.pdf>

A. Brown, R. Mortier, T. Rodden. **An exploration of user recognition on domestic networks using NetFlow records**. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, 2014. <http://mor1.github.io/publications/pdf/homesys14-netflow.pdf> / <http://dx.doi.org/10.1145/2638728.2641560>