

Workshop Report — CRAWDAD Workshop 2007

Jihwang Yeo^a

jyeo@cs.dartmouth.edu

Tristan Henderson^b

tristan@cs.st-andrews.ac.uk

David Kotz^a

dfk@cs.dartmouth.edu

^aDepartment of Computer Science, Dartmouth College, Hanover, NH 03755, USA

^bSchool of Computer Science, University of St Andrews, St Andrews, Fife, UK

Wireless network researchers are hungry for data about how real users, applications, and devices use real networks under real network conditions. CRAWDAD, the Community Resource for Archiving Wireless Data at Dartmouth, is an NSF-funded project that is building a wireless network data archive for the research community. We host wireless data, and provide tools and documents to make it easy to collect and use wireless network data. We hope that this resource will help researchers to identify and evaluate real and interesting problems in mobile and pervasive computing. This report outlines the CRAWDAD project and summarizes the third CRAWDAD workshop, held at MobiCom 2007.

I. Introduction

The third CRAWDAD workshop was held in conjunction with the MobiCom conference on September 14, 2007, in Montréal, Québec, Canada. Thanks to our NSF support, the workshop was free.

II. About CRAWDAD

If you are not familiar with the CRAWDAD project, we encourage you to visit the website¹ to learn more. Briefly, the NSF is funding an effort to build a true community resource: an archive with the capacity to store wireless trace data from many contributing locations, and with the staff to develop better tools for collecting, sanitizing, and analyzing the data. This Community Resource for Archiving Wireless Data At Dartmouth, CRAWDAD, works with community leaders to ensure that the archive meets the needs of the research community, works with the other leading centers that develop network tracing tools and metadata, and works with research organizations and corporations to ensure continuing support for the archive after NSF's funding ends.

Our goal is to serve you, the researchers and educators, in collecting and using wireless data.

Jihwang Yeo is a staff member of the CRAWDAD project. Tristan Henderson is a Lecturer of School of Computer Science, University of St Andrews, Scotland. David Kotz is a Professor of Computer Science at Dartmouth College. CRAWDAD is sponsored by the National Science Foundation under award number 0454062.

¹<http://CRAWDAD.cs.dartmouth.edu>

III. CRAWDAD Workshop 2007

The workshop began with a brief introduction of the CRAWDAD project. Following this were five invited talks by two CRAWDAD data contributors, one CRAWDAD user, and two of our research collaborators. This year the workshop ended with a poster session featuring five posters.

The following summarizes the invited talks and posters. Interested readers may visit the workshop website² for a detailed agenda, slides and posters.

III.A. Welcome and Introduction

David Kotz, CRAWDAD Principal Investigator, welcomed the attendees and talked about the recent growth of the archive. He highlighted the growing numbers of users, data, and tools. As of December 2007, the archive includes 37 data sets, 17 tools, and 110 papers have used CRAWDAD data. At the time of the workshop, there were 746 users from 425 institutions, and this has since risen to 917 users from 520 institutions at the time of writing.

III.B. Invited Talks by the Community

We invited three speakers from the CRAWDAD community — two data contributors and one data user.

In the first invited talk, Benoit Grégoire from Île Sans Fil, one of the CRAWDAD data contributors, discussed the Île Sans Fil (Island Without Wires) hotspot network.³ Île Sans Fil is a non-profit organization that

²<http://CRAWDAD.cs.dartmouth.edu/workshop2007>

³<http://www.ilesansfil.org>

operates a network of free Wi-Fi hotspots in Montréal, Québec.

Benoit and his colleagues recently contributed their data set,⁴ which was collected from 206 public Wi-Fi hotspots in Montréal from April 2004 to August 2007.

He then described the challenges they encountered in deploying and operating the hotspots: e.g., high equipment cost, equipment failure, low network speed, and difficulty in deploying and controlling networks in different organizations. To deal with these challenges, they have developed a software solution, called WifiDog, for easy deployment and management of each hotspot. He demonstrated the features of WifiDog and showed several interesting data analysis results. When he concluded the talk, he mentioned that they hope that the *CRAWDAD* users can analyze their data `ilesansfil/wifidog` to find groups of hotspot users using some clustering techniques.

Stefan Saroiu, from the University of Toronto, talked about his research on incorporating social information in data collection. He has contributed the `toronto/bluetooth` data set, which is a set of traces of Bluetooth activity in different urban environments.

In the talk, Stefan first described examples of social information, such as the relationships between people like whether they are friends, strangers, and familiar strangers. He then argued that wireless trace collection in future should incorporate gathering social data to build better mobile systems. He showed how we can build better mobile systems by exploiting social information with two research examples: improving DTN (Delay Tolerant Network) routing algorithms and slowing down the spread of mobile worms. He showed that exploiting social structure leads to substantial performance gains for DTN routing protocols, and social networking firewalls offer an attractive solution against worm infections. As a final remark, he stressed that we can improve mobile systems by using social information.

Ahmed Helmy, from the University of Florida, talked about his wireless measurement research, especially his experiences with the MobiLib trace library. He and his research group members are active users of the *CRAWDAD* archive and have published over six papers using the *CRAWDAD* data sets.

Ahmed argued that gaining insight into deployment context is important for network protocol design and we can gain such insight by analyzing extensive measurements and traces. He called this approach “trace-driven context-aware protocol design.” He also gave an overview of existing libraries of wireless traces in-

cluding *CRAWDAD* and MobiLib. The MobiLib library, built by Ahmed’s research group, consists of over 15 traces collected from various organizations. In particular, it contains the USC WLAN traces that were collected by his research team. (we are working to mirror these data sets in *CRAWDAD*.) He described several cases of his research using MobiLib and discussed the lessons learnt from the research. In conclusion, he suggested some research directions and issues, such as privacy, ambiguity, and security questions.

III.C. Invited Talks by Collaborators

The workshop featured talks by two of our research collaborators, Toni Hakanen from the University of Oulu and Colleen Shannon from CAIDA.

Toni Hakanen from University of Oulu, talked about the panOULU (public access network OULU) network.⁵ We, the *CRAWDAD* team, are collaborating with the panOULU team on hosting their WLAN traces and analyzing some *CRAWDAD* data sets. The panOULU network provides free city-wide wireless broadband Internet access in the City of Oulu, Finland, and is operated in cooperation by City of Oulu, University of Oulu, and many other organizations.

Toni first introduced the panOULU network by describing its history, structure, and service features. He said that the panOULU network covers areas such as downtown Oulu, the City Library and many other public service facilities in Oulu, and the network is already being used by thousands of users. He talked about R&D resources using the panOULU network, such as management, monitoring, and network analysis systems. He showed several interesting network analysis results from these systems.

Colleen Shannon from CAIDA/DatCat talked about the *CRAWDAD*-DatCat gateway, a collaborative effort between *CRAWDAD* and DatCat. The DatCat catalog indexes Internet measurement data.⁶ The goal of DatCat is similar to that of *CRAWDAD*: to easily share data among researchers and provide them with accurate metadata.

Colleen first gave an introduction and demo about the DatCat catalog. She then explained how the DatCat team had collaborated with the *CRAWDAD* team for developing the *CRAWDAD*-DatCat gateway, with which both *CRAWDAD* and DatCat systems can share data and metadata. As a result of this work, 24 *CRAWDAD* data sets are indexed on the DatCat catalog at the time of writing.

⁴<http://crawdada.cs.dartmouth.edu/ilesansfil/hotspot>

⁵<http://www.panoulu.net/>

⁶<http://www.datcat.org>

III.D. Posters

This year's workshop featured a poster session with five posters:

- Comparing Mobility and Predictability of VoIP and WLAN Traces; Jeeyoung Kim, Yi Du, Mingsong Chen, Ahmed Helmy (University of Florida, US).
- Analyzing Gender-gaps in Mobile Student Societies; Udayan Kumar, Nikhil Yadav, Ahmed Helmy (University of Florida, US).
- Synthetic traffic generation based on Measurement-driven modeling of large Wireless Local Area Networks; Manolis Ploumidis, Elias Raftopoulos, Maria Papadopouli (University of Crete, GR).
- Connection Duration Statistics for Wireless Devices; P. Ranjan, J. Li (Intelligent Automation, Inc., US).
- Characterizing the usage of a large municipal Wi-Fi network; T. Ojala, T. Hakanen, O. Salmi, M. Kenttälä, J. Tiensyrjä; P. Närhi (University of Oulu, FI).

The accepted posters include three trace characterization works (Kim et al., Kumar et al. and Ojala et al.) and two modeling works (Ploumidis et al. and Ranjan et al.).

Kim et al. compare VoIP traces with the whole WLAN traces in terms of the performance of mobility predictors. They find that all predictors perform quite poorly with VoIP users compared to general WLAN users.

Kumar et al. analyze the difference in WLAN usage between male users and female users. They observe that there exists a distinct difference for different genders: for instance, male users have longer session duration than female users in most cases.

Ojala et al. characterize the usage of a large municipal Wi-Fi network. They show a set of simple analysis results, e.g., number of devices, kinds of devices, ratio between 'heavy' users and 'one-time' users, number of sessions per client and session length.

Ploumidis et al. present a methodology for modeling wireless workload in terms of flows and sessions. They implemented a synthetic traffic generator based on the methodology and evaluated it by comparing the traces the generator produced with real traces.

Ranjan et al. examine real network usage times and show that network usage times empirically follow either a slowly dying tail or prominent multi-modal distribution.

IV. Conclusion

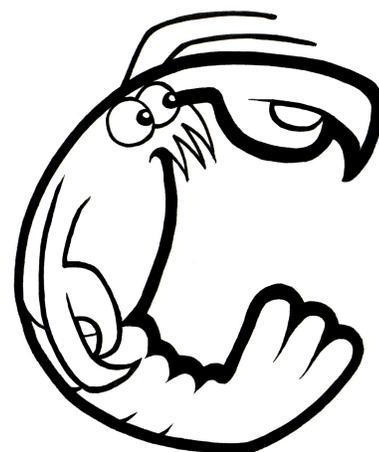
The workshop was a great success and resulted in many action items for us.

One of them is building a real community. While there are many users of the data and many useful papers being published, we have struggled, however, to build a real community. Most feedback and questions are still dealt with in a centralised fashion from the CRAWDAD staff, and there is little user participation on our wiki or mailing list. You can help! We want to build a self-sustaining community for the long term.

We also seek to address some ongoing fundamental challenges regarding our data-archiving process: it is hard to collect wireless data, it is hard to sanitize them for release, and it is sometimes hard to obtain permission to share them with others.

If you would like to learn more about *CRAWDAD*, please visit the website.

You can access our data and tool collection, view metadata and relevant published papers, subscribe to a mailing list, and post your research experiences to the wiki. We also welcome suggestions and volunteers to help collect and organize data.



CRAWDAD.cs.dartmouth.edu