

Workshop Report — CRAWDAD Workshop 2006

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 seriously starved 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 second CRAWDAD workshop, held at MobiCom 2006.

I. Introduction

The second CRAWDAD workshop was held in conjunction with the MobiCom conference on September 25, 2006, in Marina del Rey, California. In all, 38 attendees from 29 institutions signed our attendee list. Thanks to our NSF support, the workshop was free and registration was not strictly policed, so a better gauge of participation might be the number of occupied seats. During parts of the workshop, all the 49 chairs in the room were full.

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

Jihwang Yeo is a staff member of the CRAWDAD project. Tristan Henderson was a Research Assistant Professor of Computer Science at Dartmouth College at the time of the workshop, and moved to the University of St Andrews in December 2006. 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>

NSF's funding ends.

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

III. CRAWDAD Workshop 2006

The workshop began with a brief introduction and a demonstration of the CRAWDAD website. Following this were three invited talks by a CRAWDAD data contributor and two CRAWDAD users, one of who used our data for educational purposes. Finally, we brainstormed on the future of CRAWDAD with three breakout discussion groups covering tools, MANETs (Mobile Ad Hoc NETWORKS), and modeling. After hearing back from each group, the workshop concluded.

The following summarizes the presentations and discussions. Interested readers may visit the workshop website² for a detailed agenda and slides.

III.A. Welcome and CRAWDAD Website Demo

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. The archive now includes 18 data sets, 10 tools, and 44 papers have used CRAWDAD data. At the time of the workshop, there were 346 users from 213 institutions, and this has since risen to 462 users from 280 institutions at the time of writing.

Jihwang Yeo, a CRAWDAD staffer, demonstrated the website, and gave an overview of the currently

²<http://CRAWDAD.cs.dartmouth.edu/workshop2006>

archived data and tools. He also discussed how to correctly cite the *CRAWDAD* resources in papers, and encouraged users to visit the wiki³ for “HOWTO” documentation and online discussion.

III.B. Invited Talks by *CRAWDAD* Community

In the first invited talk, Dr. James Scott from Intel Research Cambridge discussed the Huggle project, for which several traces have been collected. He mentioned the motivation and issues in collecting traces and gave useful tips for those who plan to collect traces. “*CRAWDAD* helps by handling requests for traces, and by generating citations to make it worth it”, said James, who has contributed two data sets to the *CRAWDAD* archive. He also made several excellent suggestions for the *CRAWDAD* archive: for example, adding summary characteristics to the metadata and allowing users to search data by application type, traffic volume, or file size.

The second invited speaker was Professor Mark Hansen from the Department of Statistics at UCLA. He talked about the use of *CRAWDAD* data in his graduate course in statistical computing. He gave an overview of the course, describing the goals of the course, the tools and data used for the course, and the assignments. He then detailed those assignments that used a big data set like the Dartmouth campus wireless LAN traces from in the *CRAWDAD* archive, describing the approach, interactions with the *CRAWDAD* team, and some results. He mentioned that a better documented back-story could help students interpret what they were doing, and that trace sanitization, although necessary, does remove some of the context useful in statistical analysis.

Professor Songwu Lu, from UCLA’s Computer Science Department, talked about his research on the behavior of wireless network traffic. He described his experience of using four *CRAWDAD* traces collected by different institutions, pointing out such problems as varying qualities of traces, difficulty in synthesizing traces, and missing data. He suggested that the community develop better tools, standard data and metadata formats, and standardized benchmarks.

III.C. Break-out Group Discussion

After a short brainstorming session, we selected three break-out topics that concern the future of *CRAWDAD*: Tools, MANETs, and Modeling. Three break-out groups discussed each topic for about 30 minutes.

³<http://CRAWDAD.cs.dartmouth.edu/wiki>

The Tools break-out group first agreed that the research community needs data collection standards and principles, such as common formats and metrics, and guidelines for data collection. Suggested guidelines include desirable collection practices, for instance identifying metrics for evaluating the data, and recording device failures and anomalous behavior. The group argued that such standards and principles are necessary because uniform formats and metrics can be a basis for developing tools, and can also be used for comparing different data. The group also agreed that one of the most needed tools is a tool for visualizing the measured network traffic.

The MANET break-out group members started the discussion on what would be a desirable data set. They agreed that the “dream” data set would be a multi-dimensional, multi-layer data set that includes mobility information. An example dream data set would be in three dimensions of time, spatial, and layers, contain multiple layer information over physical to application layer, and include mobility information such as locations or contact patterns. They discussed the challenges in developing MANET benchmarks, pointing out that the scenarios may depend on both environment and application. They also made a list of MANET test beds, which will shortly appear on the *CRAWDAD* MANET Area wiki.⁴ Furthermore, they suggested several ideas for the *CRAWDAD* website, for example, an online wish list and a forum.

The Modeling break-out group first defined a model as a “more tractable means of understanding data” or a “general method of simulating lots of different environments.” Based on the definitions, they discussed which data are needed for modeling. Various sorts of data were suggested, for example, always-on traces with GPS, multi-resolution traces, WiMax traces, or vehicular traces.

In addition, they discussed other modeling issues like meta-models, validation, and tools. They wondered if it is possible to develop a model that adapts to create the best model for a specific scenario. Another question they raised is how we can validate models, especially in the face of limited trace data. They also agreed on the need for tools that extract common characteristics from many traces.

IV. Conclusion

The workshop was a great success and resulted in many action items for us. We have begun preparing guidelines for data collection and tool development.

⁴<http://CRAWDAD.cs.dartmouth.edu/manet>

Next, we will enhance the metadata on our archive, and improve our search capabilities, by adding useful fields, including summary characteristics, application types, and file sizes.

We also plan to publish an e-newsletter that highlights users' activities, and add more special-interest areas, e.g., a "mobility modeling" area. Please contact us if you are interested in helping.

If you would like to learn more about *CRAWDAD*, please visit the the website. You can access our data and tool collection, view their 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