CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth

Jihwang Yeo[*], David Kotz[†], Tristan Henderson[‡]
Department of Computer Science, Dartmouth College
Hanover, NH 03755, USA
{jyeo,dfk,tristan}@cs.dartmouth.edu

ABSTRACT

Wireless network researchers are seriously starved for data about how real users, applications, and devices use real networks under real network conditions. *CRAWDAD*, a Community Resource for Archiving Wireless Data at Dartmouth, is a new NSF-funded project to build 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 identify and evaluate real and interesting problems in mobile and pervasive computing. This report outlines the *CRAWDAD* project, the kick-off workshop that was held at MobiCom 2005, and the latest news.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Measurement

Keywords

wireless networks, measurement, wireless network data

1. INTRODUCTION

Have you ever been searching for real wireless data, i.e., data captured from live wireless networks, to understand the usage of real networks? Or have you ever needed such data for your research in wireless network or mobile computing, to identify the real problems or to evaluate possible solutions? If so, here is a relevant resource for you: http://crawdad.cs.dartmouth.edu — the website of *CRAWDAD*, the Community Resource for Archiving Wireless Data at Dartmouth. But before you open your web browser, let us take a quick look at *CRAWDAD*: the project's genesis, the first *CRAWDAD* workshop, and the latest news.

2. THE GENESIS OF *CRAWDAD*

Researchers who work with wireless networks or mobile computing are seriously starved for data. Data captured from live wireless networks would help us all to understand how real users, applications, and devices use real networks under real conditions, and how mobile users actually move about. This data helps us to identify and understand the real problems, to evaluate possible solutions, and to evaluate new applications and services.

On the other hand, most research today is based on analytical or simulation models. These models are severely limited by the complexity of real-world radio propagation and the lack of understanding about behavior of wireless applications and users. Experimental studies, however, are extremely difficult to set up. To collect data about real users on real networks requires a considerable amount of equipment, specialized software for collecting and sanitizing data, organizational permission and assistance to collect data, and human-subjects research clearance from the appropriate institutional review board (IRB).

At Dartmouth College we are fortunate. We have a campus-wide wireless infrastructure, with comprehensive data collection mechanisms to gather traces of wireless users and their behavior. We have developed an extensive toolset for collecting, sanitizing, and analyzing the trace data. We have a cooperative network-management organization, and experience with the IRB process. We have a history of sharing our data with the research community. Several other research groups from around the world, in both academia and industry, have used our data. In our experience, the need for this sort of data is great.

To meet this need, the US National Science Foundation is funding an effort to turn this Dartmouth resource into a true community resource: an archive with the capacity to store wireless trace data from many contributing locations, with the staff to develop better tools to handle the data. The *CRAWDAD* project works with community leaders to ensure that the archive meets the research community's needs, the other leading centers that develop network tracing tools and metadata, and key research organizations and corporations. We plan to lay the foundation that will ensure continuing support for the archive after NSF funding ends.

3. *CRAWDAD* WORKSHOP 2005

At MobiCom 2005, in Cologne, Germany, we held a workshop to launch the new *CRAWDAD* project. In the evening of the last day of the main conference, about 30 people gathered to learn more about *CRAWDAD* and share their thoughts on its direction. The *CRAWDAD* workshop consisted of an invited talk by Ravi Jain (then of DoCoMo Labs USA), and a lively group discussion.

[*]Jihwang Yeo is a staff member of the *CRAWDAD* project.

[†]David Kotz is a professor of computer science at Dartmouth College.

[‡]Tristan Henderson is a research assistant professor of computer science at Dartmouth College.

3.1 The importance of measurement

Jain gave an inspiring and educational talk about the importance of measurement in our field. He sees the mobility and networking research communities beginning to mature, as evidenced by the increased interplay between theoretical and experimental research. In his view, the two "dance" in a supportive cycle: experimental data allows analysis and modeling, which enhances and enables new theoretical research, which in turn generates new requirements for data, which inspires new research. He noted that many of the early wireless-network data-collection studies brought realism to the study of wireless-network traffic and user mobility. On the theoretical side, the MANET (mobile ad hoc network) community has long used arbitrary mobility models, such as random waypoint, that have no basis in reality but are theoretically tractable. Now, these communities are beginning to meet in the middle, attempting to build realistic — but usable — mobility models based on real mobility traces.

Jain made an interesting analogy to the human-genome project. The genomics research world is exploding because the availability of a detailed, common data set keeps the entry barrier low for a wide variety of research. In contrast, many aspects of wireless-network research have been very difficult because it is hard or impossible for most researchers to obtain realistic wireless data. Wireless-network providers generally do not want to release their data, and measuring a network yourself takes a tremendous amount of time and resources. So, he encouraged the community to support and engage with the *CRAWDAD* project.

He noted many challenges for the community. User privacy, he emphasized, is crucial. Everyday network users are unwittingly caught in traces of live networks, typically without consent or even being informed. Most careful research groups do obtain the necessary human-subjects research approval and take great care to sanitize the data and respect users' privacy. However, Jain predicted that the data-collection community will need to find more effective ways to obtain informed consent.

3.2 Tackling the important questions

For the group discussion, we posed several questions to the group. These included: what sort of data is needed, what metadata is needed, what tools are needed, how we protect human subjects, and how the data may be used for educational purposes.

Regarding data, many participants were interested in user mobility and thus requested that *CRAWDAD* contain user- or device-mobility traces as well as traffic traces. Some were interested in MANETs, including data from both controlled experiments and test beds. One participant was interested in active measurements, such as an interference map generated from a campus Wi-Fi network by using probes to learn about the interference between access points. Some people focused on security and wanted to find data that exhibited network attacks. Many were interested in data from large network providers, both cellular networks and wireless ISPs, although those in the room who were close to that industry were not optimistic that such traces were obtainable. A few were interested in data from mesh and community wireless networks. In fact, some suggested that mesh network operators would be very interested in collecting data, as many of these networks are just getting off the ground and might find performance data useful in network design and deployment.

The question regarding metadata provoked a long discussion. What information must be supplied along with a network trace to understand that trace's context? It depends on the use of the data, to be sure, but might include: the network topology and geography; the number and type of users and the character of the user population; configuration information about network components including brand, model, and firmware versions; and information about the data-collection methodology and glitches such as power failures.

The discussion on tools focused particularly on those for sanitizing data in ways that protect users' privacy and (where necessary) data providers' anonymity. Many participants requested tools and documentation about how to collect wireless-network and mobility data. Others wanted visualization tools to help examine the data, analysis tools to extract information from the data, or educational tools that could allow use of the data in a classroom.

The discussion on human subjects emphasized the need for obtaining human subjects approval for conducting wireless network measurement studies. By measuring wireless network users, we are potentially invading users' privacy. It is paramount, and indeed a legal requirement, to protect the privacy of these users.

Unfortunately it was getting late in the evening when we reached the final workshop agenda item, on educational uses of data, and this discussion was very short. We would be interested in hearing from any readers who are using wireless data in their current or future classes.

4. THE LATEST NEWS

We have been approaching wireless network operators and researchers to encourage them to contribute their data to our archive. As a result, we recently made major changes to the *CRAWDAD* website, with several new features:

- several new tools and data sets, including data from Bluetooth networks, MANETs, and DTNs (disruption tolerant network);
- a structured metadata description of each data set and tool; and
- basic and advanced searches on Crawdad data, tools, authors, and papers.

We have started compiling "HOWTO" documents on various topics such as collecting data, sanitising data, writing an IRB proposal and so forth, to make it easy for other researchers to conduct measurement studies. We plan to post these documents on the *CRAWDAD* wiki on the website. We also plan to set up working groups to deal with specific issues, for instance particular types of data (MANET, VANET) or topics such as education. Please contact us if you would be interested in helping.

5. SUMMARY

If you would like to learn more about *CRAWDAD*, please visit http://crawdad.cs.dartmouth.edu. You can access our data and tool collection, view their metadata and relevant published papers, and subscribe to a mailing list. We also welcome suggestions and volunteers to help collect and organize data.