CRAWDAD NEWSLETTER

Welcome to the first issue of the CRAWDAD newsletter! We will be announcing updates and publishing articles that highlight our data and tools, our web site, and researches related with our community resources.

The Devil Always Hides in the Details

Interview with a CRAWDAD user

Wei-Jen Hsu working with the CRAWDAD data

We talked with Wei-Jen Hsu, a CRAWDAD user, about how to make good use of the CRAWDAD data. Wei-Jen Hsu is currently a graduate student at University of Florida. He has been using several CRAWDAD data sets to analyze user behavior and encounter patterns in WLANs on different university campuses. He has published several papers on WLAN user characterization and modeling.

CRAWDAD: Hi Wei-Jen. Could you please introduce yourself to our newsletter readers?

Wei-Jen: I got my bachelors and master degree on 1999 and 2001, respectively, both from National Taiwan University, Taiwan. After I got my master degree, I worked for Chung-Hwa telecom, the largest telecom company in Taiwan, for 2 years. After some working experience I felt like pursuing higher study and heading into academia, so I joined USC for the Ph.D. program on 2003. I got my Engineer degree from USC on 2006 and moved with my advisor, Prof. Ahmed Helmy, to UFL to continue my pursuit for Ph.D.

CRAWDAD: We would like to know more about your "MobiLib" project and your research interests.

Wei-Jen: Our research interest, at high level, focuses on finding realistic mobility models to evaluate the performance of mobile wireless networks, and understanding its impact on network performance, with the help of real-world data. This has been one of the major themes in our group. We have tried surveys and real-world observation to understand human mobility before, and now move on to collecting such data electronically, with WLAN or other handheld devices. MobiLib is a project initiated by my advisor, Dr. Ahmed Helmy. Our objective is to collect pointers to people/projects who are collecting data that helps to evaluate wireless networks in with more realistic assumptions, so the community as a whole can be more aware of these efforts. It was started around January 2005 and the website was up and running in May 2005. For people who are interested to know more, you are very welcomed to visit our website at http://nile.cise.ufl.edu/MobiLib/

CRAWDAD: Why did you need network data for your research? In other words, what research problem motivated you to look for the data?

CRAWDAD Supplies Researchers with Wireless Data

The Community Resource for Archiving Wireless Data at Dartmouth (CRAWDAD), funded by the National Science Foundation, continues to help researchers and educators worldwide who are interested in using wireless network data and associated tools for collecting and processing the data.

The CRAWDAD project stores large data sets collected from real wireless networks, and develops tools for collecting, sanitizing, and analyzing this data. Used at over 320 universities and research labs by more than 520 users, the CRAWDAD archive has contributed to the wireless research community in a broad range of areas including workload characterization, location-aware services, network management, and protocol development.

CRAWDAD currently supplies 23 data sets and 12 tools to its users.

Wei-Jen: As of early 2005, I worked on a project to analyze and compare mobility patterns and structures of nodal encounters in different campuses. I read several papers published with different data sets and they focused on slightly different stuff. That time, my advisor told me to look into the problem of comparing these data sets. That was before CRAWDAD was available so I ended up getting data sets from multiple sources. If CRAWDAD were available back then, it could have helped me to find and get the data sets faster. As a side comment, "realistic mobility" is actually an orthogonal issue to wireless networks although it has major impacts on network performance. I ended up analyzing the problem using network data for two reasons: (1). These are the largest data sets available with relation to human mobility. (2). In general, people in our community lack trust in data obtained through other means (e.g., surveys, observations, etc.).

CRAWDAD: Were there any challenges in getting the data?

Wei-Jen: I assume that you mean the challenges in "collecting" data. Getting existing data, especially if they are published, would be easy these days. I must thank those data authors for their generosity in sharing the data. In collecting data, on the other hand, the most common problem is that campus network operators are in general very careful about releasing any data from their networks. Privacy of their users is always a major and legitimate concern. After they are willing to release data, it still took time to find out what exactly we need and how to get it. This went on for rounds and took some time.

CRAWDAD: After obtaining the data, how did you process and analyze them? Did you use existing tools, or develop your own? Why?

Wei-Jen: I mostly develop my own tool since most processing I have done is question driven. I asked those questions to myself so I was the right person to write the codes. Also, while the availability of existing tools would be great help, it also means that the exact same problem I look at has been done before. That's one thing we want to avoid as researchers.

CRAWDAD: What was the result? How did the CRAWDAD data help you solve your research problem?

Wei-Jen: Looking into multiple data sets from various campuses (USC, Dartmouth, UCSD, MIT), I found similarities in mobility features albeit these are very different campuses and the user bodies are also different. This points to fundamental features of human mobility, which I was not aware of before looking at the data. I later utilized the findings to create a mobility model. The data sets had played a central part on my research and I think what CRAWDAD is doing now – putting these valuable resources together – can be very helpful in the future.

CRAWDAD: Could you give some advice to other researchers who plan to use CRAWDAD data?

Wei-Jen: The devil always hides in the details. After you start dealing with the data you will find there are plenty of problems to solve, before you start to get some results. Be patient with that. Although researchers tried their best in cleaning up these data, there are always limitations. Be careful about that.

CRAWDAD: What do you think would be a benefit of CRAWDAD to users like you?

Wei-Jen: Provide the community a portal to access data sets and also keep us updated on what's new in the society, and hence a better and easier communication within our community.

CRAWDAD: Could you describe your future research and future use of CRAWDAD data?

Wei-Jen: I will continue to use the WLAN data sets to find out more fundamental properties on user mobility or usage patterns. I may utilize such understanding on protocol design or applications. Although it is further away from the data sets, it still roots from the data. I will also keep watching what's available in the community through CRAWDAD.

CRAWDAD: Do you have any suggestions for the CRAWDAD project so that we can better serve data providers and users?

Wei-Jen: Here is one idea: CRAWDAD can issue a Call For Dataset if some data set is of common interests of the community but not available yet.

CRAWDAD: Thank you for the interview!

CRAWDAD Datasets (as of March 1, 2007)

- dartmouth/campus
- ibm/watson
- cambridge/haggle
- mit/reality
- ucsd/sigcomm2001
- stanford/gates
- umass/diesel
- cambridge/inmotion
- nus/contact
- ucdavis/unitrans
- toronto/bluetooth
- uw/places
- gatech/vehicular
- intel/home
- ucsb/ietf2005
- mannheim/compass
- columbia/ecsma
- intel/placelab
- dartmouth/wardriving
- uw/sigcomm2004
- upmc/content
- princeton/zebranet
- ucsb/meshnet

Sniffing Out Useful Information

Interview with a CRAWDAD contributor

Interview with a CRAWDAD contributor We corresponded with Amit Jardosh, who collected network data at the IETF 2005 meeting and contributed the data to the CRAWDAD archive (as ucsb/ietf2005 dataset). We talked about the data collection and CRAWDAD. Amit is currently a PhD candidate in the MOMENT Lab of the Department of Computer Science at University of California at Santa Barbara. He works in the area of large-scale wireless networks. His research interests include large-scale wireless network measurements, protocol design and development, and network deployment.

Top: *The CONAN team - Amit Jardosh (top left), Krishna Ramachandran (top right), Prof. Elizabeth Belding (bottom left), and Prof. Kevin Almeroth (bottom right)* **Bottom**: *Three sniffers used at the most recent IETF meeting.*

CRAWDAD: Hi Amit. Could you please introduce yourself to our newsletter readers?

Amit: I am a PhD candidate at UCSB and a member of the MOMENT Lab at the Department of Computer Science. The lab is co-led by Professors Elizabeth Belding and Kevin Almeroth. The majority of my research with them has focused on the experimental and empirical understanding of large-scale wireless networks and development of wireless protocols.

CRAWDAD: We would like to know more about your "CONAN" project. Could you describe the project and your research interests?

Amit: My colleague Krishna

Ramachandran and I initiated the CONAN project along with our advisors Elizabeth Belding and Kevin Almeroth. The primary goal of this project is to make future large scale wireless LAN and mesh networks scale well so that they can support hundreds to thousands of simultaneous users. The project involves 3 steps. The first step is to develop novel measurement techniques to understand the impact of hundreds of users using large-scale wireless networks. The second step is to understand performance flaws, detect occurrences of congestion, and diagnose protocol deficiencies that degrade the performance of wireless networks during congestion. The third and final step is to suggest and develop novel algorithms and schemes that allow wireless networks to better handle excessive load and congestion - so that they are better geared for the future.

CRAWDAD: Why did you collect the network data at the IETF? In other words, what motivated you to collect the data and why did you choose the IETF as a collection venue?

Amit: The wireless networks set up at IETF meetings have historically been one of the most dynamic and heavily utilized networks that handle one of the largest set of simultaneous users of a single wireless network. Our lab's prior involvement with the IETF community helped us identify these networks as ideal examples of

what networks of the future will experience. Moreover, little is known about a 'typical' construction of a large-scale wireless network and therefore simulators cannot be used easily to simulate their operation. As a result, studying the IETF networks was an obvious choice to forward the progress of the CONAN project. The other advantage of collecting data at the IETF meetings is that they occur 3 times a year - which gives us an opportunity to test different measurement techniques in successive meetings – improving and learning with each collection exercise.

CRAWDAD: How did you set up and conduct the experiment? Did you use existing measurement tools, or develop your own? Why?

Amit: The experimental setup for data collection at the IETF is primarily divided in to two phases. The first phase is the survey phase where we gather information about the network, and use that information to qualitatively and quantitatively decide where to place the sniffers for data collection. The placement of the sniffers depends on the expected network activity close to the location and/or logistical constraints. Since the objective of CONAN is to understand congestion in wireless networks, we typically choose sniffer locations which are likely to experience a flash crowd of users and/or heavy traffic loads. The second phase is the collection phase, in which we configure the wireless network cards

in the RFMon mode (for prism2 chipset cards) or monitor mode (for atheros cards). We select these modes of operation for the cards because they allow the card drivers to tag each captured frame with information on the received signal strength, the data rate at which the frame was sent, and the receive power of the frame. We use the ethereal capture tool to collect the frames transmitted from the vicinity of the sniffers; this process is called 'vicinity sniffing'.

CRAWDAD: Were there any challenges in collecting data at IETF? How did you handle them?

Amit: We face several challenges in collecting data at IETF meetings. First, the topology of the network is likely known only a day before the commencement of the main meeting. As a result, the setting up of the data collection tools and sniffers needs to be worked upon in that single day. Second, the wireless network administrators and the meeting organizers are likely to change the topology of the network, the AP channels, the transmit power and data-rates, and/or also reschedule meetings - without any sufficient notice. As a result, we were required to adapt the data collection setup on-the-fly, to ensure that necessary and sufficient data is being collected all the times. Third, the data collection exercise does not offer the flexibility to alter and correct parameter settings because the IETF network is a transient network and every network operates differently. Once a meeting is over, any changes to the data collection effort has to occur after a hiatus of at least 3 months. Data collec-

tion exercises on university and corporate campus networks do not face this problem because in such scenarios researchers have the opportunity to modify and rerun experiments if the experiments fail to produce desirable results at any time. Fourth, the data collected at the IETF meeting requires backing up once in a day at the very least. This is required to avoid loss of data due to mishaps such as lost or damaged equipment. The fifth and biggest challenge is data privacy. The data that we collect at the IETF needs to be processed and stored securely so that no private information about the users of the network is revealed. Also, when making the data available to other researchers, we need to make sure that MAC and IP addresses are anonymized and the data payloads are removed.

CRAWDAD: After collecting the data, what tools did you use to process and analyze the data?

Amit: We first used the Dartmouth data anonymization tools to anonymize MAC and IP addresses and remove the frame payloads. We use these tools to avoid any privacy issues that may arise. We then use a slicer tool written in C to parse through the capture files to extract relevant per-frame fields or aggregate metrics based on the objective of the study.

CRAWDAD: What was the result? How did the experiment or data help you solve your research problem?

Amit: The data we have collected so far has helped us in three

areas. First, it has helped us define and understand congestion in wireless networks - the main objective of the CONAN project. Second, it has helped us devise tools and metrics that allow users and APs to evaluate wireless network conditions online - so that smart decisions may be made to improve their performance. Third, the data has helped us identify specific problem areas of large-scale wireless networks and thereby design strategies that mitigate the ill-effects of those problem areas.

CRAWDAD: Could you give some advice to other researchers who plan data collection?

Amit: I would like to give some generic, but very important advice to researchers new to the area of data collection in wireless networks. First, plan the exercise, and test your tools thoroughly so that you are sure that they work as they should. Second, ensure that your hardware is good enough. In other words, the sniffers have enough storage space, they are fast enough (memory and CPU) to record the majority of the frames that are transmitted in their vicinity. Third and most important is to make sure that the placement of sniffers and the data collected by them is relevant and useful for your evaluation.

CRAWDAD: How did you learn about CRAWDAD? How did you contribute your data to CRAWDAD?

Amit: We learnt about CRAWDAD via Tristan [Henderson]. We have contributed the ucsb/ietf05 data set to the CRAWDAD repository by making it available online.

CRAWDAD Tools
(as of March 1, 2007)

- tools/collect/snmp/snmputil.exe
- tools/process/snmp/extract.pl
- tools/analyze/snmp/index.pl

- tools/analyze/snmp/mac_addr.pl
- tools/analyze/snmp/pick_mac_addr.pl
- tools/sanitize/generic/AnonTool
- tools/collect/802.11/wrapi_plus
- tools/process/pcap/wifi_parser
- tools/process/pads/snmp_parser
- tools/analyze/802.11/Wit
- tools/process/syslog/syslog_parser
- tools/sanitize/generic/FLAIM

CRAWDAD: What do you think would be a benefit of CRAWDAD to data contributor like you?

Amit: As a contributor, I am looking forward to hearing from users of our data sets about its advantages and disadvantages. Based on such feedback, the CONAN project is likely to provide many more data sets which can be successfully used for further research.

CRAWDAD: Do you have any suggestions for the CRAWDAD project so that we can better serve data providers and users?

Amit: I think CRAWDAD is currently doing a splendid job in bringing together users and contributors of data in the wireless networking research community. This allows a effective and faster dissemination of useful information in the form of data-sets. It also presents users with a common interface for downloading relevant data sets for their use. One suggestion I would like to make is to ask users to log feedback about the usefulness of the specific data sets and also how and when they used it. Users are also requested to suggest metrics and methodologies that they find useful. This feedback will help us and other researchers collect more appropriate data-sets for further interesting research.

CRAWDAD: Thank you for the interview!

··

What's New

CRAWDAD Workshop
[October 5, 2006]

You can find the agenda and the slides at http://crawdad.cs.dartmouth.edu/events.php?show=2006

This year we had 37 people in our sign-up sheets, but the 49 chairs in the room were nearly full during parts of the workshop.

Wiki pages added
[August 1, 2006]

We are pleased to announce that we have added wiki pages to the CRAWDAD site. http://crawdad.cs.dartmouth.edu/wiki

The CRAWDAD wiki provides HOWTO documentation in the following categories:

- Sanitization
- Collection
- IRB (Institutional Review Board) and the requirements for human-subjects research clearance
- Data processing
- CRAWDAD Datasets
- CRAWDAD Tools
- Help on CRAWDAD site
- FAQ (and many more...)

We also set up the following wikis for special-interest "Areas" within our collection:

- MANET (http://crawdad.cs.dartmouth.edu/manet)
- Education (http://crawdad.cs.dartmouth.edu/education)

CRAWDAD Welcomes Tracy Camp as MANET Area Editor
[May 30, 2006]

We are very pleased to announce that Dr. Tracy Camp has been appointed MANET (Mobile Ad hoc NETworking) Area Editor for the CRAWDAD project.

She will seek out interesting MANET data sets. She will also collect software tools that can help researchers use MANET data sets or take MANET measurements.

If you have any MANET data sets (or tools to help process MANET data), we encourage you to contact Tracy to discuss contributing these to CRAWDAD.

We invite articles for the next issue!

Please contact us if you have ideas for articles, interviewees, etc.

Newsletter Editors

JIHWANG YEO

Programmer/Administrator of the CRAWDAD project

jyeo@cs.dartmouth.edu

DAVID KOTZ

Principal Investigator of the CRAWDAD project. Professor of Computer Science, Dartmouth College

dfk@cs.dartmouth.edu

TRISTAN HENDERSON

Co-PI of the CRAWDAD project. Lecturer of School of Computer Science, University of St Andrews, St Andrews, Scotland

tristan@cs.st-andrews.ac.uk

AND