CRAWDAD NEWSLETTER

Île Sans Fil - "Island Without Wires"

Interview with the Île Sans Fil/WifiDog team who contributed a large hotspot data set to the CRAWDAD archive.

ISF volunteers meet at one of the ISF hotspots.

Michael Lenczner and Benoit Grégoire, together with other colleagues, have collected network data from the Île Sans Fil (ISF) hotspots in Montréal, Québec, Canada, and contributed the data to CRAWDAD. We talked to them about the data collection and the Île Sans Fil hotspots.

CRAWDAD: Could you briefly describe the ISF network? Where is the network located? How popular are your hotspots?

Michael: Île Sans Fil (French for "Island Without Wires") is a non-profit organization that operates a network of free Wi-Fi hotspots in Montréal, Québec, Canada. Although it has been run solely by volunteers since its inception in 2003, it now counts 140 hotspots and over 45,000 users. Part of our mission is to develop technology to promote social cohesion within a geographical community, so all of our hotspots are located in publicly accessible spaces. We work with cafes, restaurants and bars, but also libraries, funeral homes, doctors' offices, and BIAs (Business Improvement Agencies) to cover parks and sections of popular commercial streets.

CRAWDAD: Could you tell us more about the history of the ISF network? How did you come up with the idea?

Michael: The community wireless networking movement phenomenon started in 1999. Open source geeks and amateur radio hams in cities across North America and Europe started hacking on the newly available inexpensive Wi-Fi equipment. One thing that ISF added to this movement was a focus on using the geographic limitations of the technology to support social interaction in those spaces. Besides encouraging and enabling people to spend more time in cafes and parks simply through making internet access available, ISF also added functionality to its networks to foster interaction and local engagement. Part of the inspiration for this was the work of new media artist Julian Bleeker like Wi-Fi Bedouin and Wi-Fi Art Cache.

CRAWDAD: How did you develop the hotspot network and software to support your group's vision?

Benoit: The WifiDog project was started by ISF to solve its own problems with existing solutions. While it has grown into its own project, many design elements (such as content-centered nature of the auth server, the exceptionally extensive multilingual support, and the very thin client) are the direct results of the initial ISF requirements.

CRAWDAD: WifiDog is a software solution for managing hotspots. What challenges are there in administering a hotspot network, and how does WifiDog help?

Benoit: Normally, it is very expensive (in hardware and human resources) to deploy and support a network where every location, and every upstream Internet connection, is controlled by a different organization. WifiDog is extremely small (occupying as little as 30KB of flash) allowing the equipment cost for a typical site to remain well under $100. Furthermore, with WifiDog and FirmwareISF, configuring the equipment at a new site is generally limited to plugging it in the existing network and entering the new router password. Also, DSL and cable connections have proven rather unreliable when taken as a group. WifiDog makes monitoring easier by graphically

displaying the status and location of each hotspot (see http://carte.ilesansfil.org/), sending mail to local support resources assigned to that hotspot in case of fault, and being completely unaffected by network address translation or firewalls upstream, and finally doing all this without any extra software (VPN, SNMP, etc). The auth server also allows increasingly decentralized administration and tech support.

CRAWDAD: Why do you collect network data from hotspots, i.e, what motivates you to collect the data? What is in the data?

Benoit: Most of the data we collect is necessary for the operation of some features (like location-aware content delivery, prior-view aware content delivery, viewing users online at the same hotspot, etc.) Other motivations are detecting and controlling bandwidth hogging, and knowing how, where, and how much our network is used (which is necessary to maintain motivation in a volunteer effort) The actual data collected is extremely limited, for privacy reasons: the account id, MAC and IP address, login and logout time, hotspot and amount of data transferred. We also log content display and click-through, but only for content we display on the auth server.

CRAWDAD: How do you set up and run data collection in each hotspot, and how do you merge the data collected from different hotspots?

Benoit: One must understand that the WifiDog architecture is that of a very thin client (the WifiDog gateway) that only does what cannot possibly (or practically) be done on the central server. When a client connects, the server opens a session on the auth server (that knows what hotspot the user is connecting from), the auth server assigns a session number and gives it to the gateway, and the gateway then counts the in and out traffic

transferred during that session, and periodically sends it to the auth server for each connected client. So no merge is necessary, the only place data is logged is on the auth server.

CRAWDAD: After collecting the data, how do you store and retrieve them? Do you make any post-processing and/or analysis on the data?

Benoit: Data are stored in a relational database in normalized form and retrieved in realtime as needed. There is no persistent post-processing done on the data.

CRAWDAD: We think that the ISF hotspot data can be used in many ways, e.g., for wireless network research, for developing better hotspot administration tools, etc. What possible uses of the ISF hotspot data do you envision?

Michael: We can definitely make technical improvements if we had a better analysis of our problem points. However, since ISF is a supposed to be a social intervention in our city, our main hope is to have researchers help us by developing a better idea of our users' behavior. We want to know the impact of ISF on our users. Do they principally use one or two hotspots or do they travel all over the city? Can we see if there are groupings of hotspots that are used by different pockets of users? Do new users migrate from less popular hotspots to more popular hotspots or vice-versa? That kind of thing will help us change the way we design and deploy our network as well as be useful information for other cities and businesses using WifiDog.

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

Top 10 Datasets and Tools by number of download users (March 2006 - August 2007)

Rank 1. dartmouth/campus (234 users) - Syslog, SNMP, and tcpdump data for 5 years or more from the wireless network at Dartmouth College.

Rank 2. cambridge/haggle (65 users) - Traces of Bluetooth sightings by groups of users carrying small devices (iMotes) for a number of days.

Rank 3. mit/reality (62 users) - Traces of communication, proximity, location, and activity information from 100 subjects at MIT over the course of the 2004-2005 academic year.

Rank 4. stanford/gates (49 users) - Traces of the Stanford CS department's wireless network.

Rank 5. tools/process/snmp/extract.pl (38 users) - A tool for generating data matching the pattern given by the users.

Rank 6. uw/sigcomm2004 (37 users) - Dataset of wireless network measurement at the SIGCOMM 2004 conference.

Rank 7. ucsd/sigcomm2001 (35 users) - SNMP and tcpdump records from 4 access points at the SIGCOMM 2001 conference.

Rank 8. umass/diesel (35 users) - The bus-based DTN (Disruption-tolerent networks) traces from UMass Amherst campus.

Rank 9. tools/process/pcap/wifi_parser (35 users) - A tool for selectively printing out header fields/flags from either live or recorded 802.11 traffic.

Rank 10. uw/places (32 users) - Location-aware dataset for extracting significant places.

Michael: We learned about CRAWDAD through a CAIDA (Cooperative Association for Internet Data Analysis) conference in San Diego. There is a project being developed to develop a common internet backbone for community networks (http://www.caida.org/projects/commons/).

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

Benoit: The WiFidog auth server already has a fairly sophisticated reporting subsystem. The anonymized dataset provided to CRAWDAD has been formatted in such a way that any SQL-based analysis developed by CRAWDAD researchers would be very easy to adapt to the WifiDog codebase as new reports. We hope that CRAWDAD does contribute them, so the many other groups using WifiDog (https://dev.wifidog.org/wiki/Community) and other researchers can benefit and build on it. We also hope to actively cooperate with CRAWDAD on the scientific aspects of data interpretation.

CRAWDAD: Thank you for the interview!

𝕮   𝕮   𝕮   𝕮
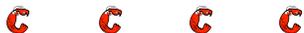
Who is using CRAWDAD?

We analyzed CRAWDAD users who were registered as of September 1, 2007. The analysis shows that about 70% of the users are from North America and Europe and more than 75% of the users are affiliated with academia. The users' interests are distributed over most areas of wireless network research. About 70% of the users intend to use CRAWDAD data for simulation and modeling.

There are 733 CRAWDAD users as of September 1, 2007. The users are distributed over 51 countries but 67% of the users are from 5 countries which are either in North America or Europe (Figure 1). For a world map, visit the CRAWDAD website.

Figure 2 shows that 78% of the users are affiliated with academia – either as students or faculty, which indicates that academia is the main beneficiary of CRAWDAD.
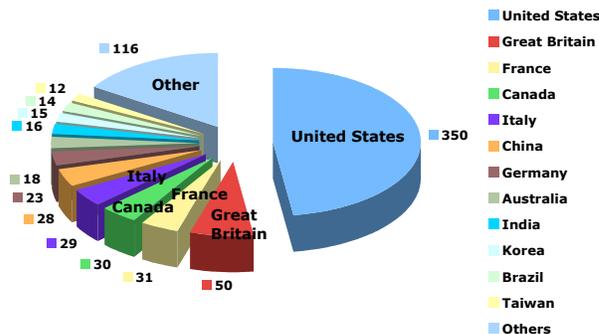

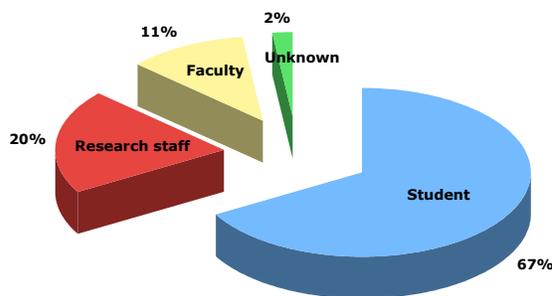*Figure 1. Distribution of CRAWDAD users' countries.*

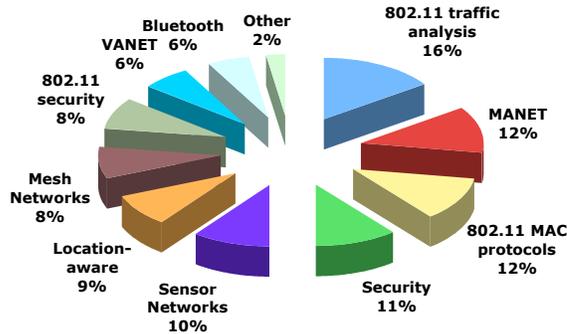*Figure 2. Distribution of CRAWDAD users' occupations.*

Figure 3. Distribution of areas of interest.

For Figure 3 and Figure 4, we count multiple answers from each user. The users' interests are widely distributed over the areas of wireless network research. Most popular research interests include 802.11 (36%) and security (19%).

About 69% of the users intend to use CRAWDAD data for simulation and modeling, which corresponds to the target usages that motivated the CRAWDAD project.
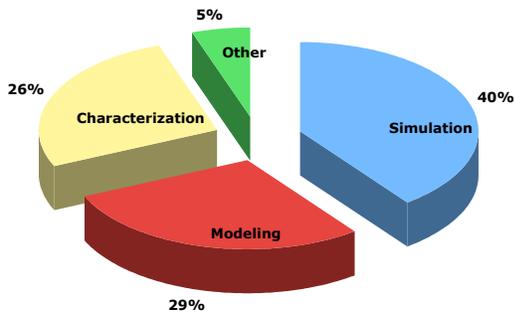
Figure 4. Distribution of intended uses of CRAWDAD.

We invite articles for the next issue!

Please contact us at crawdad@cs.dartmouth.edu if you have ideas for articles, interviewees, etc.

CRAWDAD-DatCat Gateway

The DatCat catalog indexes Internet measurement data (http://www.datcat.org). The goal of DatCat is similar to that of CRAWDAD: to easily share data among researchers and provide them with accurate metadata. Therefore, it would be ideal if both systems can share data and metadata. We set out to have DatCat index all CRAWDAD data sets.

One difficulty, however, is that even though the two systems contain metadata that have many items in common, they use different structure and format to maintain their metadata. To deal with this problem, the CRAWDAD and DatCat teams collaborated to develop a software gateway that can automatically convert the metadata. They identified the difference in metadata structure between the two systems, and then wrote a set of scripts to map one metadata structure to the other.

The collaboration was successful and the CRAWDAD-DatCat gateway has been available since August 2007. We, the CRAWDAD team, appreciate the effort of the DatCat team for implementing the gateway. We especially thank Emile Aben for writing most of metadata conversion scripts. We hope that the gateway will help network researchers easily find data, either wired or wireless.

News

CRAWDAD Workshop and Poster Session

The CRAWDAD 2007 Workshop will be held on September 14 in conjunction with MobiCom in Montréal, Québec, Canada. This year we will have a poster session after the workshop. The following four posters were accepted and will be presented at the poster session:

• 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 WiFi network. T. Ojala; T. Hakanen; O. Salmi; M. Kenttälä; J. Tiensyrjä; P. Närhi (University of Oulu, FI)

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