

# POSTER: CRAWDAD: A Wireless Network Data Archive for WiSec Researchers

Tristan Henderson  
School of Computer Science  
University of St Andrews  
St Andrews, Fife, UK  
tnhh@st-andrews.ac.uk

David Kotz  
Department of Computer Science  
Dartmouth College  
Hanover, NH, USA  
kotz@cs.dartmouth.edu

## ABSTRACT

We present the CRAWDAD wireless network data archive, the largest archive of its type. CRAWDAD hosts 115 wireless network datasets and tools, many of which are of use, and have been used, for wireless network research. CRAWDAD datasets should be of interest to the WiSec community, and similarly that the community might be interested in contributing data to the archive.

## Categories and Subject Descriptors

C.2.3 [COMPUTER-COMMUNICATION NETWORKS]: Network Operations—*Network monitoring*

## General Terms

Measurement

## Keywords

Wireless network measurement, data archiving

## 1. INTRODUCTION

Sharing and reusing research data is increasingly discussed and recognised as a crucial part of the scientific process [1, 3, 8], and indeed is being encouraged or mandated by research funding bodies. [2, 6] We run the CRAWDAD network data archive as a resource for wireless network researchers to both deposit and share their data, and for other researchers to download and use data in their research. We believe that many of the datasets in the CRAWDAD archive are of interest to the WiSec community. Indeed some past WiSec papers have used CRAWDAD datasets [5]. By advertising the CRAWDAD archive to the community, we hope to help further wireless security research and also encourage further contributions to the archive.

## 2. THE CRAWDAD DATA ARCHIVE

The CRAWDAD<sup>1</sup> (Community Resource for Archiving Wireless Data at Dartmouth) data archive (<http://crawdad.org/>) was founded

<sup>1</sup>A “crawdad” is a crustacean, also known as a crayfish, crawfish or yabby. See <https://en.wikipedia.org/wiki/Crawdada>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).  
WiSec’14, Jul 23-25 2014, Oxford, United Kingdom  
ACM 978-1-4503-2972-9/14/07.

in 2005 [4], and initially funded for the first three years through an NSF Community Resource award. The original NSF proposal summarised the archive as follows:

The investigators propose to develop an archive of wireless network data and associated tools for collecting and processing the data, as a community resource for those involved in wireless network research and education. Today, this community is seriously starved for real data about real users on real networks. Most current research is based on analytical or simulation models; due to the complexity of radio propagation in the real world and a lack of understanding about the behavior of real wireless applications and users, these models are severely limited. On the other hand, the difficult logistical challenges involved in collecting detailed traces of wireless network activity preclude most people from working with experimental data.

The investigators have years of experience collecting data from wireless networks, and have released this data to the research community. The community’s hunger for this data clarified the need for a community-run facility with a larger capacity and staff to develop the necessary tools.

Starting with a single wireless network dataset collected by the investigators, CRAWDAD has grown to become what we believe is the largest data archive of its type.<sup>2</sup> As of June 2014 we host 115 datasets and tools, used by over 6,200 users from 101 countries. CRAWDAD datasets have been used in over 1,450 papers, as well as for teaching and standards development.

## 3. SOME RELEVANT DATASETS

Some CRAWDAD datasets that might be of interest to the WiSec community include:

- **hope/nh\_amd** — Packets generated by the badges at The Next HOPE (Hackers On Planet Earth) conference held July 16-18, 2010.
- **sapienza/probe-requests** —  $\approx 11$  million 802.11 probe requests sent by  $\approx 160,000$  different devices.

<sup>2</sup>This is perhaps not as impressive as it might sound, as in general, there are few network data archives and even fewer dedicated to wireless networks!

- **st\_andrews/locshare** — Traceset of a privacy study, including encounters, sharing preferences, and accelerometer readings, conducted at University of St Andrews.
- **uportorwthaachen/vanetjamming2012** — This dataset contains traces of 802.11p packets, collected in an anechoic chamber and outdoors in Porto (Portugal) in 2011, with and without the presence of an RF jamming signal with constant, reactive, and pilot jamming patterns.
- **uportorwthaachen/vanetjamming2014** — This dataset contains traces of 802.11p packets, collected in a rural area located in the periphery of Aachen (Germany) in 2012, with the presence of an RF jamming signal with constant, periodic, and reactive jamming patterns.

#### 4. CONTRIBUTE YOUR DATA

We are also interested in hosting any wireless data that you may have collected in the course of your research. Sharing your data will help the research community, and also help you gain visibility for your work. Sharing data has been shown to increase citations [7] and is a good way of advertising your research. The CRAWDAD archive can provide storage, persistent citable links, and your data will be advertised to our 6,000 users.

Note that any sensitive data should be sanitised before release, and we require that contributors sign an agreement to indicate that appropriate IRB/ethics approval was obtained for any shared research data.

If you would like to contribute some data or know more about the process, then please e-mail [crawdad@crawdad.org](mailto:crawdad@crawdad.org).

#### 5. CONCLUSIONS

We are thankful to our funders, including the NSF, Intel, Aruba Networks, ACM SIGCOMM and ACM SIGMOBILE.

#### 6. REFERENCES

- [1] G. Andrienko, A. Gkoulalas-Divanis, M. Gruteser, C. Kopp, T. Liebig, and K. Rechert. Report from Dagstuhl: the liberation of mobile location data and its implications for privacy research. *ACM SIGMOBILE Mobile Computing and Communications Review*, 17(2):7–18, 19 July 2013. doi:10.1145/2505395.2505398.
- [2] EPSRC policy framework on research data. Online at <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>.
- [3] B. A. Fischer and M. J. Zigmund. The essential nature of sharing in science. *Science and Engineering Ethics*, 16(4):783–799, Dec. 2010. doi:10.1007/s11948-010-9239-x.
- [4] D. Kotz and T. Henderson. CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth. *IEEE Pervasive Computing*, 4(4):12–14, Oct. 2005. doi:10.1109/mprv.2005.75.
- [5] D. C. Loh, C. Y. Cho, C. P. Tan, and R. S. Lee. Identifying unique devices through wireless fingerprinting. In *Proceedings of the first ACM conference on Wireless network security (WiSec)*, pages 46–55, Alexandria, VA, USA, Mar. 2008. doi:10.1145/1352533.1352542.
- [6] NSF policy on dissemination and sharing of research results. Online at [http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag\\_6.jsp#VID4](http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag_6.jsp#VID4).
- [7] H. A. Piwowar, R. S. Day, and D. B. Fridsma. Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2(3):e308+, 21 Mar. 2007. doi:10.1371/journal.pone.0000308.
- [8] J. C. Wallis, E. Rolando, and C. L. Borgman. If we share data, will anyone use them? data sharing and reuse in the long tail of science and technology. *PLoS ONE*, 8(7):e67332+, 23 July 2013. doi:10.1371/journal.pone.0067332.



**CRAWDAD**

Community Resource for Archiving Wireless Data At Dartmouth  
<http://crawdad.org>