Summarising Wireless Network Datasets

Charlotte Knight

September 2019

Contents

1 Introduction

(CURRENTLY JUST DOER DESCRIPTION) With such large quantities of wireless traffic now travelling through networks at ever increasing rates, processing of this data can be challenging. By introducing a summarisation step before any main processing the overall efficiency of information extraction from wireless network datasets may be increased. The aim of this project will be to create a summarised report from large datasets in order to enable more efficient onward processing of the data. This will mean using statistical approaches to maintain some identified information from the dataset while reducing the overall quantity of data that must be stored and processed. The output summaries created may utilise an existing format if my research can identify an appropriate one. My project will produce a command line application involving new approaches to summarisation to run over data collected in the CRAWDAD archives, the approach taken may (in this project or otherwise, dependent on time constraints) also be extended to work on datasets in real-time so as to eliminate the need for storing large datasets before summarisation.

2 CRAWDAD Usage

2.1 Research

When completing this research the focus has been on one particular dataset from the CRAWDAD archive. This dataset was chosen because it is one of the most popular datasets in the archive, having been cited by 374 papers at the time of writing [1]. The most frequently cited dataset however is cambridge/haggle, the reasoning for deciding not to focus on this instead is that the Cambridge dataset is comparatively small in size and would therefore benefit much less from the summarisation which this project hopes to provide.
The papers that have been selected for use in this research were chosen because they all cite the dartmouth/campus dataset. A simple online search was used to retrieve the most recent papers citing the chosen dataset, this was how all papers aside from [8] and [4] were selected. The remaining papers were chosen based on aiming to include all apparent general areas of research that have used the dartmouth/campus dataset in the past 9 years. Table 1 shows a summary of the type of information each paper needed to use from the dartmouth/campus dataset.

2.2 Summary of Results

The usage of the Dartmouth University CRAWDAD dataset is primarily regarding network mobility and DTNs. As such, the most often needed information seems to be identifiers (mainly IP addresses) for both mobile devices and access points, and the times of connections.

There is also research into SDN and wide-band small-cell networking using the dartmouth/campus dataset. These uses seem to require a wider variety of

		Properties Needed		
Paper	Brief Description	Device/AP Identification	Time of Transmission	Transmission Quality/Rate
[5]	Mobility	x	x	
[2]	Software Defined Networking	x		x
[10]	Network Selection	x		x
[9]	Opportunistic Routing	x	x	
[3]	Delay Tolerant Networking (N/A)	x	x	
[4]	Delay Tolerant Networking	x		x
[6]	Wireless Sensor Networks (N/A)			
[11]	Offloading	x	x	
[7]	Machine to Machine Communication	x	x	
[8]	Distribution of Inter-Contact Time	x	x	

Table 1: Table of the properties of CRAWDAD dartmouth/campus data used in various research projects in which it was cited. Papers in which the dataset was not used are marked "(N/A)" in the description field.

information from the data, however these instances are much less frequent than those mentioned above. Bandwidth, quality of connection, and MAC address are only mentioned in these less common cases.

3 Format Selection

4 Design of Code

5 Implementation

6 Outcome/Evaluation

References

[1] *About CRAWDAD*. 2014. URL: http://crawdad.org/about.html.

[2] Ahmed Amokrane et al. "Flow-Based Management For Energy Efficient Campus Networks". In: *IEEE Transactions on Network and Service Management* 12 (Dec. 2015), pp. 1–1. DOI: 10.1109/TNSM.2015.2501398.

[3] G. Baudic, Tanguy Pérennou, and Emmanuel Lochin. "Following the Right Path: Using Traces for the Study of DTNs". In: *Computer Communications* 88 (May 2016). DOI: 10.1016/j.comcom.2016.05.006.

[4] Zhenxin Feng and Kwan-Wu Chin. *State-of-the-Art Routing Protocols for Delay Tolerant Networks*. Oct. 2012. URL: http://arxiv.org/pdf/1210.0965.p.

[5] Yan SHI, Shanzhi Chen, and Xiang XU. "MAGA: A Mobility-Aware Computation Offloading Decision for Distributed Mobile Cloud Computing". In: *IEEE Internet of Things Journal* PP (Nov. 2017), pp. 1–1. DOI: 10.1109/JIOT.2017.2776252.

[6] M. Sun et al. "Efficient Articulation Point Collaborative Exploration for Reliable Communications in Wireless Sensor Networks". In: *IEEE Sensors Journal* 16.23 (Dec. 2016), pp. 8578–8588. DOI: 10.1109/JSEN.2016.2611594.

[7] S. N. Swain, R. Thakur, and S. R. M. Chebiyyam. "Coverage and Rate Analysis for Facilitating Machine-to-Machine Communication in LTE-A Networks Using Device-to-Device Communication". In: *IEEE Transactions on Mobile Computing* 16.11 (Nov. 2017), pp. 3014–3027. DOI: 10.1109/TMC.2017.2684162.

[8] K. Wei et al. "Distribution of inter-contact time: An analysis-based on social relationships". In: *Journal of Communications and Networks* 15.5 (Oct. 2013), pp. 504–513. DOI: 10.1109/JCN.2013.000090.

[9] M. Xiao, J. Wu, and L. Huang. "Community-Aware Opportunistic Routing in Mobile Social Networks". In: *IEEE Transactions on Computers* 63.7 (July 2014), pp. 1682–1695. DOI: 10.1109/TC.2013.55.

[10] Yu-Han Yang and K. J. Ray Liu. *WAVEFORM DESIGN AND NETWORK SELECTION IN WIDEBAND SMALL CELL NETWORKS*. June 2014. URL: http://hdl.handle.net/1903/14834.

[11] Bowen Zhou et al. "An Online Algorithm for Task Offloading in Heterogeneous Mobile Clouds". In: *ACM Transactions on Internet Technology* 18 (Jan. 2018), pp. 1–25. DOI: 10.1145/3122981.