

Master Thesis - Problem Description

Johan Alexander Nordstrand Rusvik

1 Description

1.1 Background

OpenCellID [1] is the worlds largest collaborative community project for collecting GPS positions of cell towers. The GPS positions of the cell towers are calculated by looking at data collected with crowdsourcing. Everyone with a smartphone can download an application free of charge which registers, relative to the connected cell tower, signal strength, mobile country code, mobile network code, along with other parameters. It also registers the current GPS position of the smartphone. All of these parameters makes a measurement, and all measurements are collected by a server and stored in a database. As of August 2014, the database contained more than 6 million unique Cell IDs and 1 Billion measurements.

OpenCellID provides a visual service that consists of a world map where you can zoom in and out, and move around the globe. The service plots every measurement along with the calculated position of cell towers within the map's current scale. The algorithm for calculating a cell tower position is very simple. It looks at every measurement connected to a given cell tower and calculates the mean of the longitudes and latitudes of the GPS position of the smartphones, and gives this mean as the cell tower position.

1.2 Dataset

1.3 Scope and Goal

1.3.1 Implementing existing methods and improving them

Yang, Varshavsky, Liu, Chen and Gruteser [6] describes two existing methods (Weighted Centroid and Strongest Received Signal Strength) for cell tower localization and a process to improve the accuracy of these methods. First, I want to implement the two existing methods and test them on the dataset provided by opencellid.org. Then I want to implement the suggested improvements and test them on the same dataset. This will give me a basis for my own research to improve the methods further, or to come up with my own method.

1.3.2 My own improvement of existing methods

I want to improve the algorithm mentioned to calculate the position of cell towers. The dataset provides several parameters that are not being taken advantage of. Including some of these parameters in the calculation might significantly improve the exactness of cell tower positions. The tools I want to use to accomplish this is signal detection theory and geometry.

1.3.3 Blindspots

If I manage to create an algorithm that gives close to the exact location of cell towers I might also be able to search for and locate any smaller or larger blindspots, if they exist. That is, areas that are not covered by any cell. Looking at the entire world would be too time consuming. The areas of most interest to me would be the city I currently live in, Bergen, or my birth city, Oslo.

1.3.4 Position of mobile phone

If I'm able to create an algorithm that gives close to the exact location of cell towers I want to investigate the possibility to find the exact location of the mobile phone, by not using the GPS functionality. By knowing which cell towers the mobile phone is connected to during a specific time interval, and by using the phones own equipment for measuring speed, time and direction (without using the GPS), it might be possible to calculate the exact position of the mobile phone by using triangulation.

1.3.5 Health issues

I would also like to investigate the health issues associated with mobile phones and base stations. For now it is not a priority as it is a bit of a sidestep from the main goal of improving the algorithm for calculating the location of cell towers. It would be a rough overview of the discussion surrounding this issue.

1.3.6 Crowdsourcing

1.4 Schedule

18.09.14 - 18.11.14: Implementing and testing the methods described in [6], both the original and the improved versions.

18.11.14 - 20.12.14: Documenting the results previously obtained and how I got there.

20.12.14 - 15.03.15: Researching for and implementing my own method for calculating the localization of cell towers.

15.03.15 - 01.06.15: Testing my method, documenting the results obtained and how I got there, and completing the written part of my master thesis.

2 Terms and Terminology

2.1 Cells

A *base station* is responsible for sending and receiving data to and from multiple wireless hosts [4]. The hosts use the base station to connect to the larger network. *Cell towers* in *cellular networks* and access points in 802.11 wireless Local Area Networks, also known as WiFi networks, are examples of base stations.

A cellular network is partitioned into a number of geographic coverage areas, known as *cells*. In a *Global System for Mobile Communications* (GSM) cellular network, the correct term for a cell tower is *base transceiver station* (BTS). The BTS transmits signals to and receives signals from the mobile stations in its cell, typically mobile cell phones. The coverage area of a cell depends on many factors, including the transmitting power of the BTS, the transmitting power of the user devices, obstructing buildings in the cell, and the height of base station antennas. Many systems today place the BTS at corners where three cells intersect, so that a single BTS with directional antennas can service three cells.

2.2 Signal Detection Theory

The work done by Peterson, Birdsall and Fox in 1954 [5] is considered to be the first fully developed work on signal detection theory. They describe it as follows: Suppose an observer is given a voltage varying with time during a prescribed observation interval and is asked to decide whether its source is noise or signal plus noise. What method should the observer use to make this decision, and what receiver is a realization of that method? The receiver is typically a mobile phone or some other device that can receive such a voltage. The voltage is also called the receiver input.

Heeger gives a more introductory version [3]. When the receiver receives input, two kinds of noise factors contribute to the uncertainty; external noise and internal noise. The evaluation of the input can result in four different outcomes:

- Hit: There is some information (signal) contained in the voltage and we found it.
- Miss: There is some information (signal) contained in the voltage but we did not find it.
- False Alarm: There is no information (signal) contained in the voltage but we still think we found some.
- Correct Reject: There is no information (signal) contained in the voltage and we did not find any.

Another factor that plays a part is the criterion for how often we decide there is some information contained in the voltage. If we set a low criterion we will find the information every time it is there but we will also get a lot of false alarms. If we set a high criterion we will not get many false alarms, but may also miss a lot.

2.3 Computational geometry

2.3.1 Triangulation

2.4 Crowdsourcing

The word crowdsourcing is used for a wide group of activities that take on different forms. The flexibility and adaptability of crowdsourcing makes it an effective and powerful practice, but difficult to define and categorize. Jeff Howe says that crowdsourcing "is a business practice that means literally to outsource an activity to the crowd". Enrique Estell-Arolas propose a more general, but specific definition [2]:

"Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should

participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage that what the user has brought to the venture, whose form will depend on the type of activity undertaken.”

With respect to my project, the crowdsourcing has already been carried out. The initiators of the opencellid project has already gathered loads of data from the public, which you can download for free. The process described in the background section above, which was used to gather this data, matches both of the definitions mentioned here.

References

- [1] Opencellid. www.opencellid.org.
- [2] Estellés-Arolas and Gonzáles-Ladrón de Guevara. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 2012.
- [3] David Heeger. Signal detection theory. 1997.
- [4] Kurose and Ross. *Computer Networking: A Top-Down Approach*. Pearson Education, 6th edition, 2013.
- [5] Birdsall Peterson and Fox. The theory of signal detectability. 1954.
- [6] Liu-Chen Yang, Varshavsky and Gruteser. Accuracy characterization of cell tower localization. 2010.