

Thursday, May 27, 2010

Mobile Data: A Gold Mine for Telcos

A snapshot of our activities, cell phone data attracts both academics and industry researchers.

By Tom Simonite

Cell phone companies are finding that they're sitting on a gold mine--in the form of the call records of their subscribers.

Researchers in academia, and increasingly within the mobile industry, are working with large databases showing where and when calls and texts are made and received to reveal commuting habits, how far people travel for public events, and even significant social trends.

With potential applications ranging from city planning to marketing, such studies could also provide a new source of revenue for the cell phone companies. "Because cell phones have become so ubiquitous, mining the data they generate can really revolutionize the study of human behavior," says [Ramón Cáceres](http://www.research.att.com/people/Caceres_Ramon/index.html) (http://www.research.att.com/people/Caceres_Ramon/index.html), a lead researcher at AT&T's research labs in Florham Park, NJ.

If you were an AT&T subscriber and were near Los Angeles or New York between March 15 and May 15 last year, there's a 5 percent chance that your data was crunched by Cáceres and his colleagues in a study of the travel habits of the company's subscribers. The researchers amassed millions of call records from hundreds of thousands of users in 891 zip codes, covering every New York borough, 10 New Jersey counties, as well as Los Angeles, Orange, and Ventura counties in California.

The data set is a collection of call detail records, or CDRs--the standard feedstock of cell phone data mining. A CDR is generated for every voice or SMS connection. Among other things, it shows the origin and destination number, the type and duration of connection, and, most crucially, the unique ID of the cell tower a handset was connected to when a connection was made.

That let the AT&T team know the location of a phone to within a mile radius at the time each CDR was generated, making it possible to determine the distance traveled from home by each cell phone every day. The group found that, on average, people living in Manhattan travel 2.5 miles most days, compared to five miles in Los Angeles. "But we also found that when you look at the longest trips people make, people that live in New York go significantly further, 69 miles on a weekday compared to 29 in Los Angeles," Cáceres says.

Cáceres hopes to work with city planners, who would usually have to resort to expensive and limited surveys to gather such information. "This kind of data can help them decide how to invest resources, for example if they want to know where to build a new train or subway station," he says. The AT&T work was presented at a [recent workshop \(http://www.inma.ucl.ac.be/~blondel/netmob/\)](http://www.inma.ucl.ac.be/~blondel/netmob/) in Cambridge, MA, earlier this month as part of the [NetSci conference \(http://www.netsci2010.net/\)](http://www.netsci2010.net/) on network science.

For now, Cáceres's group is looking to collaborate rather than commercialize. But cell phone networks are thinking about monetizing their data, says Jean Bolot, a researcher at network operator Sprint. This means a "two-sided" business model where they not only serve end users but also make money through relationships with other businesses. "This is new in the telco space but not in other areas--look at Google, for example," he says.

Since almost everyone has a cell phone, the scale of the data is immense compared to other sources. Mobility patterns might, for example, be used to adjust property or billboard advertising prices. "Just about every operator on the planet is probably thinking about this right now," says Bolot.

Another study, presented by [Francesco Calabrese \(http://web.mit.edu/fcalabre/www/\)](http://web.mit.edu/fcalabre/www/), a research scientist at MIT, and colleagues correlated location traces from roughly a million cell phones in greater Boston with listings of public events such as baseball games and plays, showing how people traveled to attend these events. "We could partly predict where people will come from for future events," the team wrote in a report on their work, suggesting it could be possible to provide accurate traffic forecasts for special events.

The surge of research in this area has been enabled by the development of algorithms that can efficiently handle large networks consisting of millions of links, says [Vincent Blondel \(http://www.inma.ucl.ac.be/~blondel/\)](http://www.inma.ucl.ac.be/~blondel/), a professor of applied mathematics at

Université Catholique de Louvain, near Brussels, who organized the Cambridge workshop.

Blondel's research includes an analysis of connections between two million cell phone users in Belgium. It revealed that the French-speaking and Dutch-speaking populations of the country are barely connected by calls and texts. "This is interesting, since there are already discussions within Belgium about splitting the country in two," says Blondel.

Research in this area is typically focused on aggregate information and not individuals, but questions remain about how to protect user privacy, Blondel says. It is standard to remove the names and numbers from a CDR, but correlating locations and call timings with other databases could help identify individuals, he says. In the MIT study, for example, the team could infer the approximate home location of users by assuming it to be where a handset was most located between 10 p.m. and 7a.m., although they also lumped people together into groups by zip code.

"I feel the scientific community should take responsibility for finding out how to trade off having useful data and protecting privacy," says Blondel. He is investigating the effect of techniques like using approximate rather than exact location information, or blurring the exact time stamps of calls from a data set.

Copyright Technology Review 2010.

Upcoming Events

[Tech Connect World \(http://www.techconnectworld.com\)](http://www.techconnectworld.com)

Anaheim, CA

Monday, June 21, 2010 - Friday, June 25, 2010

<http://www.techconnectworld.com> (<http://www.techconnectworld.com>)

[2010 IEEE Conference on Innovative Technologies for an Efficient and Reliable Electricity Supply \(http://www.ieee-energy.org/\)](http://www.ieee-energy.org/)

Waltham, Massachusetts

Sunday, September 27, 2009 - Tuesday, September 28, 2010

<http://www.ieee-energy.org/> (<http://www.ieee-energy.org/>)