Collective Models of Disaster: Making a Case for Using Collective Mobile Phone Location Data in Disaster Science

Eric Best

University of Delaware – Disaster Research Center

Disaster Science and Management Ph.D. Student

Collection and analysis of mobile phone location data are exciting new tools for government agencies and researchers. Until recently, mobile phone location data was not a topic of interest to most phone users who are generally unaware that mobile phone carriers, device makers, and other third parties track user locations. Academics proposing research with mobile location data sets required unique access to proprietary data and consistently focused on tracking individuals and individual relationships in ways that are similar to existing corporate efforts (Altshuler, et al., 2011, Bengtsson, et al., 2011, Eagle, et al., 2009, Pentland, 2009).

Because most previous research is modeled after corporate efforts, privacy of monitored users is not important. Referencing Bengtsson et al.'s Haiti migration study, Dr. Matt Duckham was quoted saying, "Emergencies trump privacy, there's the expectation that in an emergency you'll give up your privacy so that people know where to go to help" (Back, 2011).

Not content with ignoring privacy implications, I propose a different approach to using mobile phone location data in disaster scenarios that preserves the integrity of privacy of subjects. Using location data to respond to disasters is not a new concept, but most bestpractice studies in the field focus on individual data, which is used to track or contact users in a manner that specifically identifies users (Fritsch and Scherner, 2005, Strawderman, et al., 2012).

With current technology, it is no longer enough to remove names from data sets and announce that information is no longer individually identifiable, for there is a growing industry merging anonymous collections of information generated by phones to monetize data. Carriers specifically brag about applications for individual location tracking (AT&T, 2012). Because sampled users do not give specific informed consent about how their data is used, any use of individual location data likely does not meet academic IRB standards. My intention is to create useful analysis opportunities for this type of data without competing with these businesses or violating academic standards of individual privacy. I hope to show that useful analysis is possible while maintaining the anonymity of sample respondents. By focusing on community-level data instead of individual tracking data I propose to create analysis methods that satisfy the desires of scientists wanting to use location data while removing the barriers of exclusivity to data and maintaining the requirements imposed by IRB policies governing the use of data for scientific research.

Disaster science and mobile phone location literature that examines collectivities limits focus to real-time data collection without baseline data (Candia, et al., 2008, Liu, et al., 2008). Conscious of the numerous difficulties with analyzing and sharing data after disasters occur (Bharosa, et al., 2010, Laefer, et al., 2006, Ratti, et al., 2006), I propose a system of models that compare aggregate baseline data to developing disaster scenarios. By creating systems where “normal” populations are known quantities, collective disaster models would be able to provide disaster responders with real-time data that has contextual meaning.

Analysis of real-time location data could be faster and cheaper than typical survey targeting methods currently used to figure out ex post facto how community members and the community as a whole react to a disaster. Existing mobile location data linked to phones can be used to identify which users are in a disaster area, where they went in response to the disaster, and a number of other valuable pieces of information that are valuable to emergency managers such as whether they had returned home or not and how long they were gone. Without the need to contact anyone, these data would allow researchers and government agencies to see, in real time, how many people were likely trapped in collapsed buildings or other man-made structures, how many were displaced, and how many returned to their original destination or moved elsewhere.

Until the last few years, it has not been possible to gather or analyze detailed location information for individuals or populations. However, recent changes in GPS technology, mobile phones, batteries, and data storage have made it possible to gather, store, and send detailed real-time location information. In the United States there are now over three hundred million individual location tracking devices in use every day in the form of mobile phones (CTIA, 2011). These devices are constantly transmitting location data to carriers and device makers, resulting in massive data sets collected almost in real time. Data analysis of populations that was impossible a few years ago is now becoming routine.

Mobile phones work by tracking approximate user location, and then selectively sending incoming radio signals to towers close to the last known device location (FCC, 2011). This is an efficiency measure making sure that signals are efficiently routed to the towers near a phone's last location. Handset location data have been used this way for most of the life of mobile phones, but until recently they were only available in real-time to law enforcement with court permission. This is changing. Access to this type of data is becoming much more widespread, which shifts the focus from accessing data to protecting subjects.

AT&T and Apple iPhone Data Collection – An Example of What is Possible Today

In 2011, Apple Corporation created a mobile phone upgrade for their iOS platform that accidentally backed up private mobile device data on computers synced to users' iPhones in plain text (Apple, 2011). These files contained troves of data about device users. Some of the most interesting files included separate location files called CellLocation and WifiLocation, containing location and time information based on cellular radio triangulation and wifi access points. There were other files containing detailed GPS location information. This allowed civilian research communities to see exactly what kind of data mobile carriers store, and it suggested new possibilities for studying disaster preparation, response, and recovery. Real-time data is promising for it gives disaster responders and researchers quick potential snapshots of communities. Nevertheless, the real power is in the data sets that measure multiple points over time for then it is possible to measure movement and make comparisons over time. Using these data for a community of users, emergency managers could create pictorial representations of populations before and after disaster strikes. This comparison would then show the impact of disasters on population entrapment and people in danger, population displacement, evacuees and traffic patterns. Using these data to their fullest potential would make communities safer, allow for better distribution of aid, and enhance disaster recovery by quickly showing which areas of a municipality were functioning as normal. From an academic standpoint, these data also have the potential to change the nature of the disaster survey, allowing for targeted surveying and response verification.

An Individual-Level Case Study

The analysis below is an effort to understand what type of data is collected on smartphones and what could be determined from it without cooperation from the carrier or device maker (Apple, 2011). Following instructions published online (Warden, 2011), I accessed my personal iPhone location tracking files and after a few hours of searching for other files backed up in my phone, had a wealth of data about my whereabouts. There were about 46,000 data points from a ten-month period. About 9,500 observations were from cellular radio locations, and about 35,500 were from Wifi locations. Data from the phone's GPS unit were not included in the study because I knowingly collected and stored those data points. The CellLocation and WifiLocation tables in the backup contained various network values, but the important ones for this analysis are latitude, longitude, and a timestamp. These values are stored in a way that makes seconds distinguishable, but unfortunately they are truncated to a few time points per day. Wifi location data is stored in decimal degrees to two decimal places, which distinguishes points to about 1.11km or less at common United States latitudes. Cell location data is stored in decimal degrees to eight decimal places, which distinguishes points to 1.11mm or less. Looking at these data using current technology it is not possible to obtain accuracy levels beyond about four to five decimal degrees. The data are not accurate to the millimeter, indicating hopes for technological improvements in the near future since Apple chooses to allow space for this level of accuracy.

A mobile phone logs location when a user switches towers or contacts or is contacted by a tower after his or her location has moved (EventHelix, 2004). Of the 299-day initial observation period from 6/22/2010 to 4/16/2011, the database had 46,353 data points, about 155 per day. When results are isolated by day of the week, tracking on Fridays through Sundays occurs statistically significantly more often than on Mondays through Thursdays, supporting the idea that carriers track location more when the device moves, mirroring my travel patterns. There is no indication that tracking frequency changes by month. Looking at the data by location another interesting pattern emerges. Data points truncated to two decimal degrees represent about a square kilometer at latitudes in the United States. Over the study period, my device was in 2,658 different square kilometers, an area smaller than Delaware, where I reside. For my sample location data, the Pareto principle serves as a rough guideline of location predictability. The device recorded about 80% of observations in about 16% of locations.

When I examine data by time and by location, they become much more powerful and predictive. Looking at time and location together, it was relatively easy to determine which combinations of times and locations showed consistent patterns. In the early morning hours the phone was almost constantly in a residence; during weekdays and work hours the phone was usually in the same place (my office). Overall, it was possible to confidently predict the location of my phone about two thirds of the time. Based on a preliminary analysis weekday evenings and weekends were difficult to predict, but during nighttime hours and the workweek, the phone frequently returned to the same locations. Once this type of analysis occurs, it is possible to determine if my phone is “off pattern”.

On an aggregate scale, by repeating this analysis for many users with different predictive times, it would be possible to determine whether an area or time had a larger than expected number of users "off pattern", possibly indicating some type of anomaly. This is possible without contacting users, looking at specific locations, or singling out any one user.

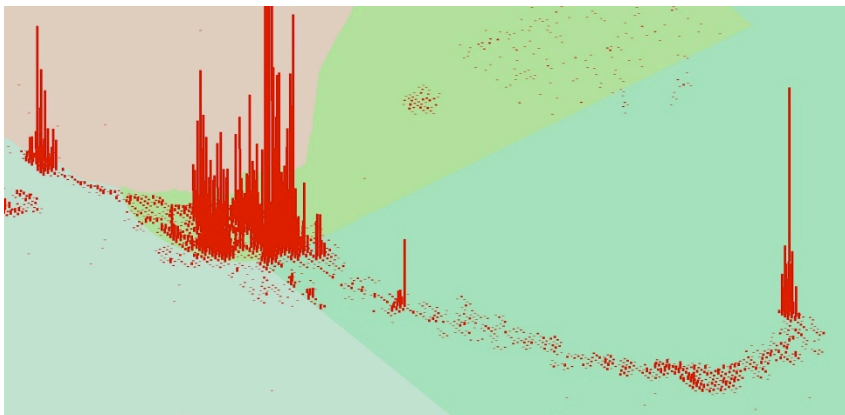

Figure 1: Aggregate Location Data for Eric Best's Mobile Phone – Mid Atlantic Region

Figure 1 shows location data along the I-95 corridor for my mobile phone under the sample observation period. The map shows Pennsylvania, Maryland, Delaware, and New Jersey with the Philadelphia airport in the upper left and Annapolis, Maryland in the lower right. X and Y points are location, and Z values (height) are the frequency of tracking for each location over a year. Presenting the data about the detailed tracking of my own whereabouts shows how useful this tracking can be, and how safe it can be with a few simple privacy filters. However, before data like these can be used to benefit emergency management and planning organizations, it is important for those collecting it or using it to acknowledge they are tracking users in detail, and to keep these data safe. While I am personally glad to have the opportunity to look at my own data, I am equally concerned that often carriers, device makers, and third parties are storing and using these data sets in inappropriate manners.

Although there has not been an issue with Apple/AT&T data breaches at this time, user location data is being stored in a way that indicates data security is clearly not a priority. Apple has not been very forthcoming about why these data are collected and stored for a year, why they are stored as unencrypted files, and why the storage format implies much more detail than the data presents (Apple, 2011). Companies that store this type of data owe it to consumers to be honest about their data collection practices, and the priority they give to data protection. The potential of misuse of data sets like these is huge, but there is also so much positive potential for analysis and disaster planning and response that this type of personally identifiable information has uses in the planning community. Once privacy issues are resolved, we could learn a great deal about population movement in disasters from location and timestamp data taken together, without identifying or tracking any individual users.

Monitoring Collectivities

While these data are invasive to track for one user, they can be useful in the aggregate without similar privacy violations. In the field of disaster science, data like these from a multitude of users may help determine things like who responds to evacuation warnings, where people congregate in disasters and other emergencies, and for many other purposes where space and time are important dimensions. With informed consent and appropriate anonymizing techniques, these data could be used in very productive ways without violating the privacy of individual smart phone users. These data, even in crippled forms and for a single user, are useful, but their greatest promise is at the level of the collectivity. Timestamps to the second and location data to a few feet for a large number of people would be useful for everyone from advertisers and retailers to urban planners and hospitals. These aggregate data are stored by private corporations but for now are unavailable to the emergency management organizations and emergency planners that could benefit from them the most.

Conceptual Models to Improve the Field of Disaster Science and Management Using Collective Mobile Phone Location Data

Aggregate data allows for additional analysis of location data without many of the conventional drawbacks. By focusing on decentralized data and established population baselines, it is possible to increase utility to the disaster management community by improving the tools available to identify disasters and recovery. Using the wealth of data collected on individuals as explained above, I propose the following models using non-identifiable collective data to improve the tools available to governments and the disaster science and management community.

Conceptual Model One: Using Mobile Phone Location Data to Improve Warning Times Through the Creation of Predictive Mass Behavior Models

The first conceptual model is a proposal to use existing mobile phone location data to improve disaster-warning times for emergency responders and populations. I propose to develop a model that would learn normal aggregate behavior for a community over time, and create a model of "normalcy" for that community. With a large enough number of phones and an observational period of many months to learn about normal location behavior, it would be possible to create a system that could identify usage in and out of their normal patterns. These usages could then be individually assigned time ranges considered predictable or unpredictable. Using this simple and anonymous “predicable/unpredictable” system, profiles could be developed for communities and a sense of location predictability and normalcy could be quickly displayed in real time.

This type of system would be different for every community, and would change based on day of the week, time of day, holidays, seasons, and other factors. With enough data, the model would be able to give users a predictability ratio compared to the norm, and an indication of how abnormal a situation is statistically. After a long enough observation period for refining data, the model would be able to output what percentage of a population is unexpectedly doing something unpredictable, and how far from past norms that percentage is. Ideally, a measured ratio that is very far from expected norms would be part of a set of metrics to determine early warning signals for disaster response units, 911 centers, law enforcement, health care organizations, and community leaders.

For example, an abnormal degree of unpredictability on a weekday may indicate that a large number of community members are staying home from work to prepare for a natural disaster or because they are sick. Without revealing any user identifications, it would be possible to get a pulse on a city or a community, and verify much of time that the society is functioning as expected. While this class of model would require a great deal of initial data gathering and large amounts of storage space, conceptually it is relatively simple to develop and maintain. Because the model focuses on when people are behaving predictably, and not specifically where they are when they deviate from their pattern, the simple binary system of "predictable/unpredictable" returns a result quickly and easily, and no relationships between monitored devices are required for analysis. This fact makes this type of model significantly less complicated and processor intensive than other proposed models.

Conceptual Model Two: Using Real-Time Location Concentration Data for Advances in Real-Time Warning Systems

The second conceptual model is also a proposal to use existing mobile phone location data to improve disaster-warning times for emergency responders and populations. However, instead of relying on an aggregation of individual predictability patterns, this proposed model will be location-based. The concept of this second model is to catalog data about specific locations within communities without ever tracking users, meeting the goal of not creating datasets of individually identifiable information. Because the model does not track user paths from one location to another, privacy is maintained. By tracking density in locations over time, it is possible to model predictive density numbers and then monitor communities in real-time.

This system would be able to flag locations with unexpected data concentrations. Ideally, this type of system would alert emergency operations centers to accidents (unexpected increases in concentration on highways in the middle of the day), threats (unexpected disbursement from a shopping mall in the middle of the afternoon), large gatherings (even if functioning as intended), and other situations where people are more or less concentrated than normal. With enough refinement, this model would ideally complement 911 and other emergency systems by alerting emergency managers to situations requiring a response faster than potential 911 calls. This system would be particularly useful on highways where mass stoppages outside of rush hours might mean that accidents occurred (Zhao, 2000).

This proposed model could be of great use to governments and emergency management organizations without violating end user privacy. By focusing on collective data by location instead of specific user behavior patterns, the data provides valuable information without violating user privacy in any way.

Conceptual Model Three: Using Mobile Phone Data to Improve Evacuation and Rescue Efforts

The third conceptual model is a proposal to use existing mobile phone location data to improve evacuation and rescue efforts without tracking individual mobile phone users. This proposed model would be the easiest proposal to implement and use because it does not require prior observation for locations or users.

An overwhelming percentage of the population use mobile phones, so during or after a mandatory evacuation or a disaster mobile phones still on in a disaster area likely indicate members of the population requiring emergency assistance (CTIA, 2011). Using maps of the disaster zones with active cell phones sending real-time data to towers, it is possible for emergency managers to quickly visualize areas where rescue teams may be required. The intention of this proposed system is to bolster emergency call centers and improve house-to-house searches by prioritizing locations for rescue workers to visit first after a disaster. By focusing on areas with active devices, emergency responders can first check areas where cellular towers show there are active phones (and likely active users). This time saving measure will allow rescue workers to check areas where people are likely to be before going house to house through an entire community.

This proposed model would allow emergency operations workers to better prioritize their rescues. Since this proposed system does not even require a background dataset or learning model, it would be set up quickly and easily with cooperation between community governments and emergency managers and mobile phone companies. The user data is already collected on an individual basis, but through anonymous aggregation, it could significantly benefit emergency responders without further violating user privacy or requiring additional data collection or storage. It would even be possible to create a feed that emergency responders could analyze without the ability to store the data.

Conceptual Model Four: Using Mobile Phone Location Data to Confirm Population Return During Disaster Recovery

The fourth conceptual model is a proposal to use existing mobile phone location data to improve disaster recovery efforts and surveys without tracking individual mobile phone users. After a disaster, communities are forced to distribute abnormal amounts of resources and manage large numbers of volunteers (Barsky, et al., 2007). The fourth conceptual model is an ideal tool for this task. By periodically tracking normal levels of mobile phone activity in communities during normal times, it is possible to compare post-disaster community levels of mobile phones and determine roughly what percentage of the population remains in an area or has returned after a disaster. Comparing these aggregate concentrations to average concentrations before the disaster, community governments and emergency responders would have the ability to see which areas require reconstruction services or volunteers first. By tracking normal neighborhood conditions and conditions after a disaster, recovery services (e.g., power and water companies) could determine how to prioritize service fixes to restore power or water to the greatest number of people in the smallest amount of time.

This proposed model also has substantial potential benefits for disaster science researchers who typically must rely on site visits and phone surveys to determine evacuation levels and community damage and recovery. These conventional methods are time consuming, expensive, and frequently inaccurate, and could greatly benefit from complementing surveys with mobile phone data of community populations (Groves, 2004). While this fourth model is similar to the Bengtsson et al. study of Haiti population migration, it will eliminate any individual identifiers that could be used to back into individual user identities such as individual SIM card identification (Bengtsson, et al., 2011).

Weaknesses

The main weakness of the class of models proposed is also one of the main benefits; the proposed conceptual models do not identify or track individuals, and there is no individually identifying information that can be recovered using current analysis techniques. By examining collectivities instead of identifiable individuals, the models sacrifice some precision and behavioral predictability in exchange for the ability to explore models that do not violate privacy. By creating platforms where users would not be identifiable even when proposed data are merged with other datasets, it is possible to create models with a lower degree of precision than would be ultimately possible. However, this weakness is by design to prevent privacy issues, legal issues, and moral issues that accompany tracking users through individually identifiable data in unprecedented ways. The hope is that by avoiding identifying individuals, it will have a positive impact in a novel way without encountering the traditional privacy and proprietary information roadblocks associated with individual location and behavioral information. Conventional mobile phone location research focuses on indentifying individuals, but there are important conclusions that can be drawn by focusing only on collectivities.

Another weakness is that the conceptual models will not immediately revolutionize any industry. The hope is to create a high return on investment by learning things from data already collected by mobile phone carriers. It is not a revolutionary new method of data collection, but an evolutionary method of data analysis. The proposed models are designed as complementary tools; increasing response speed, improving community-level monitoring accuracy for civilian agencies, and decreasing cost, time, and inaccuracies in disaster survey research. They are explicitly not intended to replace current tools used in these roles. The hope is that the proposed conceptual models, if developed and implemented on a wide scale, can provide significant benefit to emergency responders and social science and disaster researchers.

A final weakness of the conceptual models on a development scale is cost. It is simply not possible for one researcher to purchase, analyze, store, and disseminate all mobile phone location that is available at this time. Instead, the hope is to inspire the purchase of enough data to make a prototype model for one community in order to inspire emergency management services and governments to explore relationships with mobile phone carriers to create systems that could improve community warning, response, and recovery.

Improving Privacy of Subjects

The important part of these proposed conceptual models is what they leave out, not what is included. By ignoring the individually identifiable aspects of the available data that make it valuable to marketers and law enforcement, I hope to contribute to a much wider knowledge base that so far is neglected in favor of individual-focused location data. There are already firmly established industries focused on identifying individuals and predicting their behavior for marketing and law enforcement purposes, and I hope to show that these data have applications far beyond these individual uses.

The hope is that these conceptual models will show researchers that the individual is not the only important thing to keep track of when working with location information. With the exception of traffic and human flow engineering, location data analysis focuses almost exclusively on identifying individuals and predicting the future behavior patterns once identified. By proving that individually identifiable data can be modified in a way that truly guards privacy, I also hope to show that these data are ideal candidates for these extended applications. Because these data are already collected and stored, it is important to manage them in a way that can be used to maximize benefit for society even after they are used to benefit the companies that create it.

Contributions to the Field of Disaster Science and Management

By monitoring location data for whole communities, it is possible to create models that can benefit utility industries, community and state governments, charities, and academic researchers without sacrificing the privacy that should be available to law-abiding mobile phone users. By removing exclusivity barriers and IRB violations, it is possible for academic researchers to use location data without fear of ethical issues. The hope is that these conceptual models will foster service provider and government agreements to use data that are already collected and stored in a productive manner to all citizens. By removing this kind of research from the individual privacy debate, the hope is to be able to analyze data in a manner that is not threatening to established business relationships and individuals. If successful, these prototype models could improve emergency responder and service provider efficiency, save lives in disasters, and make better academic study of communities possible during both normal times and during emergencies and disasters. Because data collected in these proposed models do not identify individuals (even after mergers with other data sets), results could be shared without worry of privacy violations, making them widely available to interested parties.

Mobile phone location data are already used to assist marketers and law enforcement agencies. By developing applications that use the data on a mass scale these same benefits are available to the academic and practicing disaster communities without sacrificing the advantages afforded to the firms that benefit from individually identifiable data and the privacy rights of individual citizens.

References:

[1] Altshuler, Y., Aharony, N., Pentland, A., Elovici, Y., and Cebrian, M. (2011). "Stealing Reality: When Criminals Become Data Scientists (or Vice Versa)." *Ieee Intelligent Systems*, 26(6), 22-30.
[2] Bengtsson, L., Lu, X., Thorson, A., Garfield, R., and von Schreeb, J. (2011). "Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti." *Plos Medicine*, 8(8).
[3] Eagle, N., Pentland, A., and Lazer, D. (2009). "Inferring friendship network structure by using mobile phone data." *Proceedings of the National Academy of Sciences of the United States of America*, 106(36), 15274-15278.
[4] Pentland, A. (2009). "Reality Mining of Mobile Communications: Toward a New Deal on Data." *The Global Information Technology Report 2008-2009*, S. Dutta, and I. Mia, eds., INSEAD/World Economic Forum, Geneva, 75-80.
[5] Back, A. (2011). "Mobile Phone Tracking Could Revolutionise Disaster Aid Response." *The Converstation*, The Conversation Media Group, Melbourne, Australia.
[6] Fritsch, L., and Scherner, T. (2005). "A multilaterally secure, privacy-friendly location-based service for disaster management and civil protection." *Networking - Icn 2005, Pt 2*, 3421, 1130-1137.
[7] Strawderman, L., Salehi, A., Babski-Reeves, K., Thornton-Neaves, T., and Cosby, A. (2012). "Reverse 911 as a Complementary Evacuation Warning System." *Natural Hazards Review*, 13(1), 65-73.
[8] AT&T (2012). "AT&T Location Information Services." <http://www.wireless.att.com/businesscenter/solutions/industry-solutions/cross-industry/location-based-services.jsp>. (January 4, 2012).
[9] Candia, J., Gonzalez, M. C., Wang, P., Schoenharl, T., Madey, G., and Barabasi, A.-L. (2008). "Uncovering individual and collective human dynamics from mobile phone records." *Journal of Physics a-Mathematical and Theoretical*, 41(22).
[10] Liu, S. B., Palen, L., Sutton, J., Hughes, A. L., and Vieweg, S. (2008). "In Search of the Bigger Picture: The Emergent Role of On-Line Photo Sharing in Times of Disaster." University of Colorado Boulder, Boulder, CO, 10.
[11] Bharosa, N., Lee, J., and Janssen, M. (2010). "Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises." *Information Systems Frontiers*, 12(1), 49-65.
[12] Laefer, D. F., Koss, A., and Pradhan, A. (2006). "The need for baseline data characteristics for GIS-based disaster management systems." *Journal of Urban Planning and Development-Asce*, 132(3), 115-119.
[13] Ratti, C., Frenchman, D., Pulselli, R. M., and Williams, S. (2006). "Mobile landscapes: Using location data from cell phones for urban analysis." *Environment and Planning B-Planning & Design*, 33(5), 727-748.
[14] CTIA (2011). "Wireless Quick Facts." <http://www.ctia.org/advocacy/research/index.cfm/aid/10323>. (December 20, 2011).
[15] FCC (2011). "FAQS - Wireless Phones." <http://www.fcc.gov/encyclopedia/faqs-wireless-phones#how>. (December 21, 2011).
[16] Apple (2011). "iOS 4.3.3 Software Update." <http://support.apple.com/kb/dl1358>. (December 21, 2011).
[17] Apple (2011). "Apple Q&A on Location Data." *Apple Press Info*, <http://www.apple.com/pr/library/2011/04/27Apple-Q-A-on-Location-Data.html>. (April 27, 2011).
[18] Warden, P. (2011). "iPhone Tracker." <http://petewarden.github.com/iPhoneTracker/>. (December 21, 2011).

[19] EventHelix (2004). "Intra MSC Handover Call Flow (GSM Inter BSC - Intra MSC Handover Call Flow)." EventHelix.com, United States, 3.
[20] Zhao, Y. (2000). "Mobile Phone Location Determination and Its Impact on Intelligent Transportation Systems." *IEEE Transactions on Intelligent Transportation Systems*, 1(1), 10.
[21] Barsky, L. E., Trainor, J. E., Torres, M. R., and Aguirre, B. E. (2007). "Managing volunteers: FEMA's Urban Search and Rescue programme and interactions with unaffiliated responders in disaster response." *Disasters*, 31(4), 495-507.
[22] Groves, R. M. (2004). *Survey Errors and Survey Costs*, Wiley-Interscience, Hoboken, NJ.