March 19, 2014

To: The Federal Trade Commision
Re: Mobile Device Tracking

From: Micah Altman, Director of Research, MIT Libraries;
Non Resident Senior Fellow, Brookings Institution

I appreciate the opportunity to contribute to the FTC's considerations of Mobile Device Tracking. These comments address selected privacy risks and mitigation methods. Our perspective is informed by substantial advances in privacy science that have been made in the computer science literature and by recent research conducted by the members of the Privacy Tools for Sharing Research Data project at Harvard University.[1]

Scope of information

The speakers focussed primarily on businesses use of mobile devices to track consumers' movements through retail stores and nearby environments. However, as noted in the comments made by the Center for Digital Democracy [CDD 2014] and in the workshop discussion (as documented in the transcript), the the general scope of mobile information tracking businesses and third parties extends far beyond this scenario.

Based on the current ability for third parties to collect location information from mobile phones alone, third parties have the potential to collect extensive, fine grained, continuous and identifiable records of a persons location and movement history, accompanied with a partial record of other devices (potentially linked to people) encountered over that history.

Information sensitivity

Generally, information policy should treat information as sensitive when that information, if linked to a person, even partially or probabilistically, is likely to cause substantial harm. There is a broad range of informational harms that are recognized by regulation and by researchers and

[1] The Privacy Tools for Sharing Research Data project is a National Science Foundation funded collaboration at Harvard University involving the Center for Research on Computation and Society, the Institute for Quantitative Social Science, the Berkman Center for Internet & Society, and the Data Privacy Lab. More information about the project can be found at http://privacytools.seas.harvard.edu/.
1

[1]The author takes full responsibility for these comments. However, the author wishes to thank the members of the project for their comments and insights and acknowledge that these comments build upon joint work conducted with all member of this project and upon previous analysis developed with David O'Brien and Alexandra Wood (http://informatics.mit.edu/blog/how-provide-public-transparency-individual-privacy-%E2%80%94-comments-osha).

practitioners in the behavioral, medical, and social science fields. [Bankurt & Ander 2006; Lee 1993] Types of potential harms associated with information include loss of insurability, loss of employability, criminal liability, psychological harm, social harm to a vulnerable group (e.g. stereotyping), loss of reputation, emotional harm, and dignitary harm.

Knowledge of an individual's location history and associations with others has the potential to be used in a wide variety of harmful ways. For example, businesses may use this data for discriminatory pricing ; employers could use location history to discriminate in hiring; insurers may red-line specific areas; and third parties could use this information to target users for physical harm. [NAS 2006] Moreover, since all physical activity has a unique spatial and temporal context, location history provides a linchpin for integrating multiple sources of data that may describe an individual.

Re-identification risks

The workshop presenters explained that the MAC addresses associated with identifiers were routinely 'hashed' in an attempt to prevent the data from being identifiable. As discussed in the workshop, this is a weak protection mechanism. And more genrally, there are many examples of datasets that were believed to have been anonymized but were later re-identified. For instance, Latanya Sweeney recently demonstrated how news stories and public records can be used to identify patients in hospital records that have been anonymized by removing patient names and addresses. [Sweeney 2013] Using information such as name, age, residence, gender, hospital, incident date, and details about the incident from news stories, as well as date of birth and ZIP code from online public records, she was able to successfully identify a significant number of individual hospitalization records in state hospital data. As another example, Kobbi Nissim and Eran Tromer recovered records of over one thousand individuals in an anonymized survey by querying an internet access mechanism hosted by the Israel Central Bureau of Statistics web site and, furthermore, demonstrated that it is possible to link these records to individuals. [Ziv 2013]

Furthermor, research shows that human mobility patterns are highly predictable [Gonzalez, et. al 2008] and that these patterns have unique signatures, making them highly identifiiable [de Montjoye, et al 2013] -- *even in the absence of associated identifiers or hashes.* For example, in a recent study [de Montjoye, et. al 2013], over 95% of individuals could be identified using only four location points -- even where location was measured only every hour, and with relatively low precision.

Moreover, locational traces are difficult or impossible to render non-identifiable using traditional masking methods. [NAS 2006] Masking through aggregration and perturbation of individual points is generally insufficient, short of rendering the data useless. [Zimmerman & Pavlik 2008; de Montjoye et al 2013] More sophisticated approaches have been developed that aggregate entire location traces rather than individual points [Fung 2010, ch 14] -- but these methods are

new, not in common practice, and may still leak sensitive information about individuals.

A stronger new mathematical privacy guarantee known as differential privacy has been developed which can provide provable privacy. [Dwork 2011] Protecting geospatial using differential privacy is possible in practice [Mir, et al 2013] -- but thiscurrently requires that private data be stored with a trusted party, and only queries on the data be made publicly available. No method currently exists that allows detailed location data to be and then safely published.

Review, reporting, and information accountability

Because of the inherent identifiability and sensitivity of the data, individuals would clearly benefit from a regulation that provided increased transparency and accountability for data collection and subsequent use. The current regulatory framework largely lacks mechanisms that would provide accountability for harm arising from misuse of disclosed data.

Consent to data collection, while it should be required is not sufficient protection -- privacy policies are dauntingly complex for individuals to understand, and many summaries provided by data collectors are inaccurate. [Cranor 2012] Furthermore, data collectors will generally have be better informed of potential and actual data use, and in the face of ubiquitous data collection practices, consumers find it difficult to effectively withhold consent -- the playing field is uneven. Thus, transparency, restrictions on disclosure, and accountability for misuse are all essential to achieving an optimal balance of social benefit and individual privacy protection. [Weitzner, et al 2008] Accountability mechanisms should enable individuals to find out where data describing them has been distributed and used, set forth penalties for misuse, and provide harmed individuals with a right of action

A Modern Approach to Sharing Location Information

Addressing privacy risks requires a sophisticated approach, and the privacy protections currently used for location data not take advantage of advances in data privacy research or the nuances they provide in terms of dealing with different kinds of data and finely matching sensitivity to risk. Like treatment of other risks to subjects, treatment of privacy risks should be based on a scientifically informed analysis that includes the likelihood of such risks being realized, the extent and type of the harms that would result from realization of those risks, the efficacy of computational or statistical methods to mitigate risks and monitor access, and the availability of legal remedies to those harmed. [Vadhan, et al 2010]

A modern approach to privacy protection recognizes the following three principles:

- *The risks of informational harm are generally not a simple function of the presence or absence of specific fields, attributes, or keywords in the released set of data.* Instead, much of the potential for individual harm stems from what one can learn about individuals from the data release as a whole when linked with existing data.

- *Redaction, pseudonymization and hashing, are often neither an adequate nor appropriate practice, and releasing less information is not always a better approach to privacy.* As noted above, simple redaction of information that has been identified as sensitive is often not a guarantee of privacy protection. Moreover, in the case of location traces, even substantial aggregation of data often fails to provide adequate privacy protection.

- *Thoughtful analysis with expert consultation is necessary in order to evaluate the sensitivity of the data collected and their associated re-identification risks and to design useful and safe release mechanisms.* Naïve use of any data sharing model, including those we describe below, is unlikely to provide adequate protection.

References

de Montjoye, Yves-Alexandre, et al. "Unique in the Crowd: The privacy bounds of human mobility." *Scientific reports* 3 (2013).

Elizabeth A. Bankert and Robert J. Andur, Institutional Review Board: Management and Function (Boston: Jones and Bartlett, 2006);

Cranor, Lorrie Faith. "Necessary but not sufficient: Standardized mechanisms for privacy notice and choice." *J. on Telecomm. & High Tech. L.*10 (2012): 273.

Cynthia Dwork, "A Firm Foundation for Private Data Analysis," *Communications of the ACM* (2011): 1, 86-95

Fung, Benjamin CM, et al. *Introduction to privacy-preserving data publishing: concepts and techniques*. CRC Press, 2010.

Gonzalez, Marta C., Cesar A. Hidalgo, and Albert-Laszlo Barabasi. "Understanding individual human mobility patterns." *Nature* 453.7196 (2008): 779-782.

Krishnaram Kenthapadi, Nina Mishra, and Kobbi Nissim, "Denials Leak Information: Simulatable Auditing," Journal of Computer and System Sciences 79.8 (2013): 1322-1340.

Raymond M. Lee, Doing Research on Sensitive Topics (London: SAGE, 1993).

NAS 2006, Putting People on the Map: Protecting
Confidentiality with Linked Social-Spatial Data

Latanya Sweeney, "Matching Known Patients to Health Records in Washington State Data" (July 2013), available at http://dataprivacylab.org/projects/wa/1089-1.pdf.

Salil Vadhan, et al., Re: Advance Notice of Proposed Rulemaking: Human Subjects Research Protections (2010), available at http://dataprivacylab.org/projects/irb/Vadhan.pdf.

Daniel J. Weitzner, et al., "Information Accountability," Communications of the ACM 51.6 (2008): 82-87.

Amitai Ziv, "Israel's 'Anonymous' Statistics Surveys Aren't So Anonymous," Haaretz (Jan. 7, 2013), http://www.haaretz.com/news/national/israel-s-anonymous-statistics-surveys-aren-t-so-anonymous-1.492256.

Zimmerman, D. L., Pavlik, C. (2008). Quantifying the effects of mask metadata disclosure and multiple releases on the confidentiality of geographically masked health data. Geographical Analysis 40.1, 52 (25).