Is My Information Private?
Geo-Privacy in the World of Social Media

B. Kar[1], R. Ghose[2]

[1]Department of Geography & Geology, University of Southern Mississippi, 118 College Drive # 5051, Hattiesburg, MS - 39406
Email: bandana.kar@usm.edu

[2]Department of Geography, University of Wisconsin-Milwaukee, P.O.Box 413, Milwaukee, WI - 53201
Email: rghose@uwm.edu

1. Abstract

Privacy has always been a public concern. It can be dated back to "media privacy" of 1361 that was enacted to protect people from peeping toms in England. In 19th century, privacy became a hot issue with the arrival of modern photography and printing press that enabled easy reporting of personal information and pictures, automated data processing to catalog citizens during World War II, and record keeping Systems in the U.S. These actions led to the enactment of a number of privacy policies, specifically, focusing on data protection - mis-use of data and/or malicious access of data by unauthorized persons. In this era of the Internet, ubiquitous computing combined with the growth of the geo-spatial and the Information and Communication Technology industries has made it possible for anyone to access personal and location information of another person anytime and anywhere. For instance, GeoAPI of Twitter (a social media service) can help track the movement of a person over space and time. Despite having regulatory policies, it is possible to extract location information of a person by using VGI (Volunteered Geographic Information) and CGI (Contributed Geographic Information) available from social media sites. This study explores the extent to which location information obtained from social media sites are reliable and useful and are influenced by an individual's concern and knowledge of privacy.

Keywords: Social media, Geo-Privacy, VGI, CGI, Accuracy, Location Information

2. Introduction

The revolutionary changes in geo-spatial technologies (e.g. Global Positioning System (GPS), ultra-wide-band radio) have enabled the collection and generation of a large amount of geospatial data at different levels of accuracy, coverage and cost. The growth of the Information and Communication Technologies (ICT) (e.g. mobile devices, cell towers, the Internet) has led to the emergence of location-based services (LBS) (Google Latitude and Google Street View) and social media sites (e.g. Twitter and Facebook). Using spatial data, these services and sites provide information about an individual's or a vehicle's location accurately and precisely. Given the uncertainty about how this information can violate a person's privacy, this study explores the extent to which people's knowledge and concern of privacy influences their decision to share location information through VGI (Volunteered Geographic Information) and CGI (Contributed Geographic Information).

Location-based services and social media sites singularly do not violate personal information. However, by coordinating location information with other types of information, such as an individual's address, these services can provide personal information to a third party, thereby leading to location privacy violation. Location privacy though is an unclear concept, it is defined as "the ability to prevent other parties from learning one's current or past location" or "the ability of an individual to move in public space with the expectation that under normal circumstances their location will not be systematically and secretly recorded for later use" (Beresford and Stajano 2003; Blumberg and Eckersley 2009).

Personal privacy - "the quality or state of being apart from company or observation" or "freedom from unauthorized intrusion" (Webster 2010) - has received extensive legal attention with the publication of the 1890 *Harvard Law Review* article "the right to be let alone" (Warren and Brandeis 1890). It is also protected by the U.S. Constitution's Fourth Amendment that prohibits government observation or intrusion of one's home. In contrast, despite numerous legal battles (e.g.

U.S. vs Skinner 2012, Boring vs Google 2010), location privacy did not receive equivalent attention until 2013 and 2014 when a number of bills were introduced in the U.S. Congress for the Location Privacy Protection Act, Geolocation Privacy and Surveillance Act, Electronic Communications Privacy Act Amendments Act, and Online Communications and Geolocation Protection Act (Pomfret 2013). These acts in essence prohibit (1) unlawful attainment and disclosing of geo-location information to government agencies, and (2) companies from obtaining and disclosing geo-location information from electronic communication media without users' consent (Pomfret 2013, GPS.gov 2014).

Though the legal underpinning exists, it is possible to harvest spatial information from crowd-sourced geographic information (VGI and CGI) (Harvey 2013). As Harvey (2013) pointed out, from ethical and legal standpoint, VGI requires users to opt-in or volunteer to share their information (e.g. OpenStreetMap) based on their explicit understanding of how the information will be used as opposed to CGI in which users may opt-out from sharing their information, but it can still be collected (e.g. cell phone tracking). Thus, the users' attitude towards volunteering information and protecting privacy can influence the accuracy of the spatial information obtained from VGI and CGI. Given the wide spread usage of crowd-sourced data and growing concern of public about privacy, this study attempts to explore the following questions identified by other researchers (1) how accurate is the location information obtained from VGI vs CGI and what percentage of VGI and CGI data are fudged and obliterated? (2) how much an individual's concern and knowledge of privacy influences their sharing of location information on social media? (Harvey 2013, Krumm 2008).

2.1 Background

Privacy has been a public concern since 1361 when "media privacy" was enacted to protect people from peeping toms in England (Langheinrich 2013). However, it was not a hot issue until the arrival of modern photography and printing press in the 19th century, which enabled easy reporting of personal information and pictures, the automated data processing to catalog citizens during World War II, and the record keeping Systems in the U.S. (Langheinrich 2013). These actions motivated people to protect their privacy, especially, access to personal information, and led to the enactment of a number of privacy policies focusing on data protection, i.e., mis-use of data and/or malicious access of data by unauthorized persons.

Like privacy, surveillance is not a new concept. The first surveillance instrument - Bentham's Panopticon - a building within which all occupants could be tracked from one vista point - was designed by Samuel Bentham in 1785 (Dobson and Fisher 2007). This was followed by the second surveillance instrument a.k.a. "Big-Brother" - a closed-circuit television (CCTV) to observe public spaces (Dobson and Fisher 2007). The third panopticon, built on geo-spatial technologies and ICT, is more intrusive and enables sharing of personal information with a larger community (Dobson and Fisher 2007). Because of its ability to track an individual's location information, Dobson and Fisher (2003) termed this practice as "geoslavery". In this era of the Internet, ubiquitous computing combined with the growth of the geo-spatial and the ICT industries has made it possible for the pervasive presence of the third panopticon. For instance, GeoAPI of Twitter (a social media service) can help track the movement of a person over space and time.

Location privacy will not be a problem if people were not tempted to share their location information on social media sites and LBS. Given our dependency on ICT and the recent popularity of location-based services (e.g., smart phones, Twitter's location API, Google Latitude, etc.), in addition to the legal communities, actions must be taken by users, developers and providers of these technologies and services to protect privacy and location privacy. Therefore, additional to legal protection, a number of computational countermeasures have been taken to protect personal and location information of a user.

The users of location aware services may use self-regulatory techniques, such as, providing limited personal information, adjusting privacy settings, limiting access to certain people, disabling or refusing to use applications that may lead to losing privacy or provide limited privacy protection, and not opting to share certain information, to protect their privacy (Cottrill 2011). Likewise, a number of approaches are currently available for users to protect their privacy (Tsai et al. 2010): 1) Blacklist: blocks specific individuals from accessing a user's personal information; 2) Friends Only: allows access to only individuals listed as friends; 3) Granularity: restricts the spatial resolution at which a user's location information will be available to others thereby reducing accuracy of location information; 4) Group: allows a user to create a group of members to have access to his/her location information; 5) Invisible: prohibits anyone from accessing a user's location information; 6) Location-based rules: allows a user to identify locations belonging to a specific category (e.g. work or entertainment) for which location information will be available; 7) Network: allows users to release their location information to all members belonging to the users' network; 8) Per-request permissions: allows users to release location data to certain parties who have requested permission to access these data; and 9) Time-based rules: allows users to decide the duration within which their location data may be available to others.

To protect users' privacy, developers use a number of computation approaches: 1) Anonymity: a user's location information is same as other users; 2) Obfuscation: the accuracy of location information is reduced by introducing noise; 3) Aggregation: a user's location information is aggregated with other users' location data; 4) Encryption: the true location information is fudged and/or encrypted (Beresford and Stajano 2003, Leitner and Curtis 2006, Popa et al. 2011, Saponas et al. 2006, Krumm 2008).

For these approaches to be successful users must be aware of their presence, and users must know that their privacy is/can be lost because third parties can access their personal and location information. A user must also take appropriate actions to prevent information sharing. In any case, there is still the possibility of extracting location information from social media sites that have significant potential for real life applications (e.g. crisis mapping). This raises the question, how much crowd-sourced geographic information is reliable and useful.

3. Methodology

The **study-site** will be the University of Southern Mississippi, which is located in the City of Hattiesburg, Mississippi. System architecture similar to the one presented in Stefanidis et al. (2011) will be used to extract information from only one social media site: Twitter. Using twitter's native API and Python, both textual and graphic data will be extracted for two weeks duration for specific hash tags (e.g., USM, Hattiesburg, and University). In addition to extracting real time data, data for the same duration will be purchased from the third party vendor (Gnip) that provides access to twitter feeds to increase data set size. By parsing the extracted data, the spatial information will be obtained (this will include coordinates of the locations where the tweets have been posted). Likewise, from textual and graphic data, geo-tagged information will be extracted for the study site and duration. By geo-coding the location information, a point map will be created, which will be overlaid on the University's location information layer. Finally, the spatial distribution of error will be computed, and total error and error variance based on spatial and temporal distribution will be computed to understand the extent of data fudging done to protect location information. Finally, the survey used in the study exploring knowledge of location privacy (Kar et al. 2013) will be introduced to on-campus students and their user names will be used (after receiving their consent) to extract tweets from these students for the same duration to explore the last question. The spatial distribution of error will also be visualized to explore the impact of location where tweets are posted on accuracy.

Findings and Contributions

We expect that the quality of the data is influenced by the situation rather than public's concern of privacy. For instance, during a disaster, people are motivated to share accurate location information that is relevant to the disaster instead of their own personal space. We also expect that population of specific age group will be more concerned about their own privacy and protecting personal information. For instance, older generation (35+ years) will be more concerned about taking steps to protect their own privacy at any given time.

In the web 2.0 era, citizens a.k.a. "human sensors" provide valuable and timely information about the post-disaster landscape that can be used in emergency response and crisis mapping (Goodchild 2007, Zook et al., 2010, Bengtsson et al., 2011). Given the crowd-sourced data are available at real time and at fine resolution than other data sets, they can be used as reference data to assess accuracy of other data sets, e.g. remote sensing data. However, the usage of crowd-sourced data can be influenced by the quality of the data. The answers to the research questions identified in this study will make three contributions. First, it will show if public's concern and knowledge of privacy influences the quality of crowd-sourced information, which is one of the future research questions identified by Goodchild (2010) in the field of Geographic Information Science (GIScience). Second, it will contribute to the research on data quality of crowd-sourced information and implications for real-life applications (Lease 2011). Finally, by investigating the privacy concerns of public, it will contribute to the research in location privacy, especially, what triggers people to care about location privacy (Krumm 2008).

References

Authority F, 1973, Stating the obvious: An interdisciplinary approach. *Journal of Entirely Predictable Results*, 63(2):1037–1068.

Bengtsson, L., X. Lu, A. Thorson, R. Garfield, J. von Schreeb, 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): DOI: 10.1371/journal.pmed.1001083.

Beresford, A., and F. Stajano, 2003, Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1): 46-55.

Blumberg, A. J., and P. Eckersley, 2009, On locational privacy, and how to avoid losing it forever. *Electronic Frontier Foundation.* http://www.eff.org/files/eff-locational-privacy.pdf (last accessed 15 May 2014).

Cottrill, C. D., 2011, Location privacy: who protects? *URISA Journal*, 23(2): 49 - 59.

Dobson, Jerome E., and P. Fisher, 2003, Geoslavery. *IEEE Technology and Society Magazine*, 47 - 52. https://msu.edu/~kg/874/geoslavery.pdf (last accessed 5 June 2014).

Dobson, Jerome E., and P. Fisher, 2007, The panopticon's changing geography. *Geographical Review*, 97(3): 27 – 323.

Goodchild, M. F, 2007, Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69 (4): 211-221.

Goodchild, M. F, 2010, Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science*, 1: 3 - 20.

GPS.gov, 2014, *Geolocation Privacy Legislation.* http://www.gps.gov/policy/legislation/gps-act/ (last accessed 15 May 2014).

Harvey, F., 2013, To volunteer or to contribute locational information? Towards truth in labeling for Crowdsourced Geographic Information. In: D. Sui, S. Elwood, M. Goodchild (eds), *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice,* NY, USA, 31 - 43.

Kar, B., and R. C. Crowsey, J. J. Zale, 2013, The myth of location privacy in the U.S.: surveyed attitude vs. current practices. *The Professional Geographer*, 65(1): 47-64.

Krumm, J., 2008, A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6): 391–399.

Langheinrich, M., 2013, Privacy by design - principles of privacy-aware ubiquitous systems. http://cs.gmu.edu/~jpsousa/classes/699/papers/privacy%20Langheinrich.pdf (last accessed 5 June 2014).

Leitner, M., and A. Curtis, 2006, A First step towards a framework for presenting the location of confidential point data on maps-results of an empirical perceptual study. *International Journal of Geographical Information Science*, 20(7): 813-22.

M. Lease, 2011, On quality control and machine learning in crowdsourcing. *Association for the Advancement of Artificial Intelligence.* https://www.ischool.utexas.edu/~ml/papers/lease-hcomp11.pdf (last accessed 15 July 2014).

Pomfret, K. D., 2013, Latitudes and attitudes: Zooming in on geospatial data, privacy and the law in the digital age. *Centre for Spatial Law and Policy,* http://www.spatiallaw.com/Uploads/Latitudes_and_Attitudes.pdf (last accessed 15 May 2014).

Popa, R. A., A. J. Blumberg, H. Balakrishnan, and F. H. Li, 2011, Privacy and accountability for location-based aggregate statistics. *CCS'11.*

Saponas, S., J. Lester, C. Hartung, and T. Kohno, 2006, Devices that tell on you: The Nike+iPod sport kit. *Technical Report 2006-12-06*, University of Washington.

Tsai, J. Y., P. G. Kelley, L. F. Cranor, N. Sadeh, 2010, Location-sharing technologies: Privacy risks and controls. http://cups.cs.cmu.edu/LBSprivacy/files/TsaiKelleyCranorSadeh_2009.pdf (last accessed 5 June 2014).

Warren, S. D. and L. D. Brandeis, 1890. The Right to Privacy. *Harvard Law Review,* 4(5): 193-220.

Zook, M., M. Graham, T. Shelton and S. Gorman, 2010, Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian Earthquake. *World Medical & Health Policy,* 2: 7–33.