Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 551–558

Procedia Computer Science

www.elsevier.com/locate/procedia

The 4th International Symposium on Emerging Information, Communication and Networks (EICN-2017)

I am at home: Spatial Privacy Concerns with Social Media Check-ins

Jayakrishnan Ajayakumar [a], Kambiz Ghazinour [b]*

[a]*Department of Geography, Kent State University, Kent, Ohio, USA*
[b]*Advanced Information Security and Privacy Lab, Department of Computer Science, Kent State University, Kent, Ohio, USA*

Abstract

Inclusion of Location Based Services (LBS) on social media including geo-tagging, and check-ins has been on the rise. Users share their location information, motivated by self-expression and socialization, as well as to improve their understanding about places. However, sharing of location information develops an increased potential for spatial privacy violations. Along with the advances in LBS, the developments in the field of Geographic Information Systems (GIS) including easy to use web-maps and reverse geocoding services, provide researchers as well as others to analyze and visualize spatial footprints generated by LBS. This study will use geo-tagged check-in data from the microblogging platform Twitter, along with easy to use GIS based tools to demonstrate spatial privacy issues created due to online check-ins. Apart from demonstrating the spatial privacy vulnerabilities, we also analyze the spatial privacy issues based on the privacy policy documents. Finally, we provide suggestions to improve spatial privacy based on policies, and algorithmic techniques.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Location Based Services; Social Media; GIS; Spatial Privacy

1. Introduction

With the advent of Web 2.0 and other technological advancements, social media became highly popular during the last decade[20]. Inclusion of Location Based Services (LBS)[1] on social media including geo-tagging, and check-ins gave the opportunity to people to ask the "where" question about photos, events, places, and even other persons. The

* Corresponding author. Tel.: +-330-672-9061.
E-mail address: kghazino@kent.edu

1877-0509
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.278

ubiquitous use of social media along with location-based services can facilitate easy sharing of a considerable amount of geo-personal information, which are collected with or without awareness of location disclosure[2]. Based on previous studies[3], users tend to share their locational information through LBS to improve their social connectivity and to enhance their virtual experience about places. Apart from users, the locational data is widely used for place-based studies[4] by social scientists and human geographers. Apart from researchers, business firms also use user-generated locational data to improve their customer experience[3]. Along with numerous advantages, locational data from LBS, like any other form of spatial data, increases potential privacy issues. Recent advances in the field of mobile-based technologies enable users of social networking platforms to share locations in the form of GPS coordinates, which could be mapped using any software that supports spatial data visualizations. Friedland and Sommer[5] in their work on cybercasing used geo-tagged tweets to demonstrate real-world attacks. They used publicly available data from other sources to improve contextual information of geo-tagged tweets to expose the vulnerability of shared locations. *Pleaserobme.com*[6], a website developed to raise awareness on spatial privacy issues in social media platforms, demonstrated the malicious potential of location-based data, by identifying users who are not in their homes. They analysed Foursquare check-ins to demonstrate the spatial vulnerability. In their work on privacy in geo-social networks, Vicente et al.[3] elaborate on two major threat categories for location-based social media networks including release of sensitive location information and re-identification through location. They show that apart from revealing sensitive locational information through check-ins and geo-tagging, LBS-based social media platforms reveal highly contextual information such as timestamps, which makes re-identification easier. Li et al. [2] in their work on location disclosure through georeferenced tweets were able to provide confident estimates about home or work location of users, using geo-referenced tweets and land cover datasets. Jin, Joshi and Anwar [7] in their work on access control mechanisms for users, used check-in data from four popular LBS sources to compare and contrast privacy policies related to different location based services.

1.1. Our contribution

This study aims to demonstrate the spatial privacy vulnerabilities created due to posting LBS check-ins in Twitter by using readily available open-source tools for mapping and reverse geo-coding. Apart from demonstrating the vulnerabilities, we also analyze the privacy policies for Twitter and Swarm app (the LBS), and provide suggestions for improving spatial privacy based on policies. The second section of this paper provides details about the data, the filtering methods used to extract relevant information and the pre-processing steps. The third section provides implementation details of the demonstration and the fourth section elaborates on the spatial privacy issues as well as the policy-level analysis for social media sites. Fifth section provides details about some of the algorithms that could be used to improve spatial privacy. The last section will have the conclusion and future works for this study.

2. Data Characteristics

2.1. Data Acquisition

As a part of the NSF-supported SESYNC project "The Socio-Environmental Data Explorer"[8], we have collected around 1.5 billion geo-tagged tweets ranging from 09/08/15 to 10/15/16. The tweets were collected using the Twitter streaming API. From the corpus of tweets, English tweets were filtered out using the “lang” attribute from the metadata available with the tweets.

2.2. Data Filtering and Preprocessing

The locational details from the tweets are extracted from two different tags available with the tweet metadata. Tweeting from mobile devices having GPS capabilities with Twitter location services enabled creates an attribute