RAISING AWARENESS: AN EXAMINATION OF EMBEDDED GPS DATA IN IMAGES POSTED TO THE SOCIAL NETWORKING SITE TWITTER

Michael B. Flinn, Frostburg State University, mflinn@frostburg.edu
Christopher J. Teodorski, Robert Morris University, cjst4@mail.rmu.edu
Karen L. Paullet, American Public University, kp1803@online.apus.edu

ABSTRACT

We live in a society that depends heavily on the use of technology, specifically the Internet, to complete our daily activities and to stay in touch with friends and family. Perhaps one of the most potentially dangerous activities on the Internet is keeping in touch with others via social media, or social networking sites. This exploratory study examined personally identifiable information (PII), of photos uploaded to TwitPic. A total of 417,056 images were downloaded and processed for the study. The data collected from the TwitPic photos will uncover the GPS location information embedded in the photo's header, as well as, the possible impacts of that information on the user.

Keywords: Twitter, TwitPic, EXIF data, metadata, social network sites, privacy

INTRODUCTION

We live in a society that depends heavily on the use of technology, specifically the Internet, to complete our daily activities and to stay in touch with friends and family. On a daily basis, the Internet is used for personal finance, productivity, time management, and even consuming products. The Internet is also used to watch movies, catch up on missed television shows, online gaming, and networking with others. However, there are potential dangers when keeping in touch with others via social media, or social networking sites.

The use of social network sites such as Twitter, Facebook, and Reddit have transformed the way people keep in touch with friends and family, as well as aggregate information. Sites such as Twitter and Facebook allow people to keep in touch with friends and family, and even make new friends. These same sites also enable individuals to post an enormous amount of personally identifiable information about themselves and their families [6].

Twitter describes itself as "a real-time information network powered by people all around the world that lets you share and discover what's happening now" [16]. Twitter is a social network-driven microblogging service that is primarily made up of 140 character posts, or "tweets" as they are known in the Twitter terminology. TwitPic is a separate service that integrates with Twitter and allows users to share photos, as well as the information in the photos, with their followers. A possible drawback to users sharing their photos is the unintentional sharing of personally identifiable information along with a photo via the EXIF (Exchangeable Image File Format) header. Personally identifiable information has been defined in the 2007 OMB Memorandum [1] on *Safeguarding Against and Responding to the Breach of Personally Identifiable Information* as information which can be used to distinguish or trace an individual's identity such as their name, social security number, biometric records, etc., alone or combined with other personally identifiable information. For the purpose of this paper, we examined personally identifiable information (PII), specifically the location field located in the EXIF header of photos uploaded to TwitPic. The researchers explored the data collected from TwitPic and discuss the impacts that the embedded GPS location may have on the user. The background section discusses the growth of social media networks and the possible PII information that these sites can supply to a would-be cyber-criminal.

BACKGROUND INFORMATION

In August 2009, there were more than half a billion people using online social network sites worldwide [7]. Twitter alone attracted over 44.5 million people. In June 2009, Twitter experienced rapid growth with a single month increase of 7 million people [13]. Before the Internet, and the acceptance of social media, if a person wanted to rob a bank they had to physically be present in the bank, risking the chance of being caught or identified. Now with the use of the Internet, money can be stolen from individual banking accounts without a person ever stepping foot into a bank. In August 2009, [14] the Justice Department exposed the largest credit and debit card breach in U.S. history. Albert Gonzalez and two Russian conspirators were charged with stealing more than 130 million cards. The group hacked the firewalls of several companies' computer systems in order to obtain credit card and debit card numbers. Other ways in which individuals can have their information stolen is with the use of keyloggers. Keyloggers can be installed by cyber-criminals to capture a person's pin number and bank account information making it very easy to transfer or steal that individual's money. With the rise of the Internet, it has also become easy for sexual predators to stalk children.

Before the Internet, for a child abduction to occur, a person would hang out in a local park or near a school playground searching for the right child. Now a predator can search and follow their prey from the comforts of their own home. Much of this can be accomplished with the use of social media, such as Facebook and Twitter, to assist with building a profile on the child. They can follow children in chat rooms and social networking sites waiting and hoping to become their friend. People in general list names, the school they are currently attending, the school that they graduated from, their hometown, their occupation, and pictures of family and friends on social network sites. All of the information that people readily make available online can give a cyber-criminal everything they need to stalk, abduct, track a child, or even steal a person's identity. Cyber-criminals can use the personal information to harass, or even harm, a person and their family. A cyber-criminal "can choose someone they know or a complete stranger with the use of a personal computer and the Internet. The information that is available about people on the Internet makes it easy for a cybercriminal to target a victim" [9].

As more people begin to use social networking sites they are becoming more vulnerable to cyber-criminals. Cyber-criminals can easily search for a victim without ever leaving their house. Many users of social network sites are aware of the possible pitfalls of failing to secure their personally identifiable information using the privacy settings of the site. However, what about the personally identifiable information placed in the photos that these individuals place online? With the circulation of cheaper GPS enabled cameras, the risks of unknowingly exposing personally identifiable information increases.

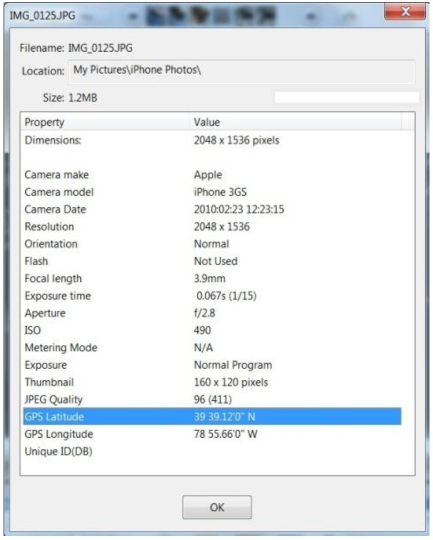
Social network sites "not only allow users to create personal information spaces which are easily accessible from anywhere on the Web, but also gives them the tools to share their personal artifacts with others and take advantage of others shared artifacts"[11]. As people use Twitter to follow their friends and family throughout the day they also post pictures to TwitPic. This allows the individual who posts that tweet to visually represent the text of the tweet by including a photo. With GPS receivers being built into many of the lower cost cell phones and other devices, the consumer may be unaware of the full capabilities of the device and unintentionally post sensitive location information.

With geo-locating hardware (GPS receivers) being added to cell phones and cameras, some of the pictures that are available on TwitPic contain embedded information (longitude and latitude coordinates) about where the picture was taken. Cyber-criminals can take advantage of the information that is available in these photos. They can then use a program such as Picasa, or other tools, to analyze the information in the EXIF header. When posting a picture to TwitPic, the poster has the option to include additional information, with regard to the picture, for their friends and followers. This additional information is then passed to Twitter. For example, suppose a family is taking pictures while they are celebrating the 8th birthday of one of their children. They post to Twitter the following, "Celebrating Adrian's 8th birthday at our favorite restaurant down the street. You just can't get burgers like this any place else." In addition they upload a picture to TwitPic to accompany the tweet. The picture and tweet were posted from an iPhone or Blackberry. In the picture, you see a family huddled around a young boy blowing out the candles on a birthday cake. Often times when predators want to lure a child they find a conversation that the child is familiar with to gain their trust. Imagine this tweet in the hands of a child predator. The predator can check the date of the post, and discover when the photo was taken. He has already gathered the name and age of the boy, if he analyzes the EXIF data in the image, he will discover the GPS coordinates of the restaurant, which he already knows is "down the street". Directions to the restaurant can be obtained with a quick check of Google maps. He now knows that he is in very close proximity to their house. The predator could use this information to start looking for the child, or he could wait until an additional photo is posted. If the additional photo includes a picture of their house or is taken inside their home, he can look at the EXIF data for that image and determine the exact location of the child based on the embedded GPS coordinates.

In a recent study conducted by Ullrich [17], his team tested prevalence of EXIF headers in photos from TwitPic. His team wrote a script to capture 15,291 images. Scripts are miniprograms, or a series of commands that are issued to carry out a specific repetitive function [2]. One might think of a script as a macro, which is often found in office productivity software. A second script was written to analyze the information obtained from the pictures. Of the 15,291 images that were analyzed, 399 images included location of the camera when the picture was taken and 102 pictures included the name of the photographer. Most of the pictures were being uploaded from cell phones with the majority of pictures coming from the iPhone [17]. In addition to the visible information in a digital photo, it contains an EXIF header.

EXIF HEADER

The EXIF header contains many pieces of information about a photo. The majority of the fields contain benign information such as EXIF version, manufacturer of the camera, model of the camera, and the software on the camera. However, many people may not know that this same header contains fields that can contain personally identifiable information such as the date, time, owner of the device, and GPS location, as seen in figure 1, at which the photo was taken [16].



IMG_0125.JPG

Filename: IMG_0125.JPG

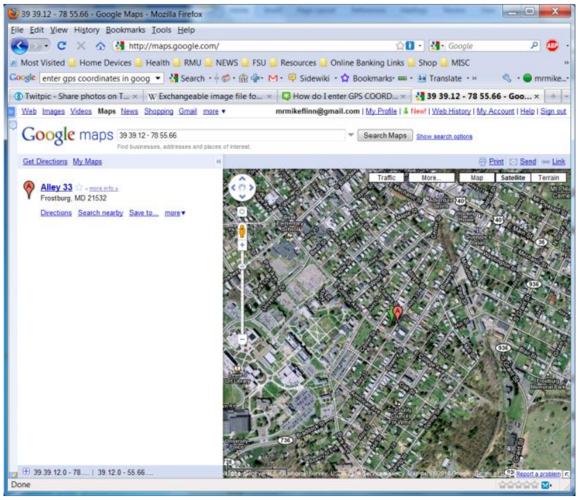
Location: My Pictures\iPhone Photos\

Size: 1.2MB

Property	Value
Dimensions:	2048 x 1536 pixels
Camera make	Apple
Camera model	iPhone 3GS
Camera Date	2010:02:23 12:23:15
Resolution	2048 x 1536
Orientation	Normal
Flash	Not Used
Focal length	3.9mm
Exposure time	0.067s (1/15)
Aperture	f/2.8
ISO	490
Metering Mode	N/A
Exposure	Normal Program
Thumbnail	160 x 120 pixels
JPEG Quality	96 (411)
GPS Latitude	39 39.12'0" N
GPS Longitude	78 55.66'0" W
Unique ID(DB)	

Figure 1- Example of EXIF Data Stored in Photos

Once an individual posts a photo containing this information online, it is easy to use a mapping service such as Google Maps to discover the location, as seen in figure 2, at which the photo was taken. The below map was obtained from the GPS latitude and longitude coordinates from figure 1. The user can type in the coordinates from the EXIF data to find the exact location from where the photo was taken.



39 39.12 - 78 55.66 - Google Maps - Mozilla Firefox

http://maps.google.com/

39 39.12 - 78 55.66

Alley 33
Frostburg, MD 21532

File Edit View History Bookmarks Tools Help
Most Visited
Home Devices
Health
RMU
NEWS
FSU
Resources
Online Banking Links
Shop
MISC
enter gps coordinates in goog
Search
Sidewiki
Bookmarks
Translate
mrmike...
Twitpic - Share photos on T...
Exchangeable image file fo...
How do I enter GPS COORD...
39 39.12 - 78 55.66 - Goo...
Web Images Videos Maps News Shopping Gmail more
mrmikeflinn@gmail.com | My Profile | New! | Web History | My Account | Help | Sign out
Google maps
Search Maps
Show search options
Find businesses, addresses and places of interest.
Get Directions
My Maps
Print
Send
Link
more info »
Directions
Search nearby
Save to...
more
Traffic
More...
Map
Satellite
Terrain
40
36
936
Done

Figure 2 - Google Maps Displaying the Coordinates from an EXIF Header

Using a few simple programming scripts, the authors created a program that collected samples of photos from TwitPic in order to discover the number of users that upload photos containing the GPS location of the photo. This is a more rudimentary form of the way pleaserobme.com works. Pleaserobme.com was a website that illustrated how easy it is to rob a person's home based on the information they post on sites such as Twitter. When posting the information to Twitter the information becomes publicly available allowing viewers to know when a person is not home, hence, knowing when to rob their houses [3]. The intent of pleaserobme.com was to raise awareness for personal security and privacy on the Internet. The service used geo-location information, similar to geotagging in photos, to identify where the user was located at the time of the tweet.

GEOTAGGING

Geo-location hardware is now embedded in many cell phones such as the iPhone and Blackberry and high-end cameras such as the Cannon Rebel and Nikon 5000 DSLR. The GPS receivers enable the phone or camera to record the GPS location in the EXIF header automatically. The use of this technology "denotes the marking of digital resources with geographical coordinates mostly used for images" [11]. With the increased use of mobile devices and cameras, geotagging has increased in popularity. Nations [8] defines geotagging as marking a video, photo or other media with a location. A digital photo is geotagged when the device can capture the location of where the photo was taken and adds the GPS coordinates to the EXIF header of the photo. Besides GPS coordinates, the date and time in which the photo was taken, the kind of camera and the camera settings to include shutter speed, image stabilization and image format can be recorded in the EXIF header.

DISCUSSION

In addition to the posting of text, Twitter is often utilized to post links to various content including multimedia such as pictures, videos, and songs. A significant driver of microblogging is the mobile phone and its ability to send SMS messages or text messages. Twitter describes themselves as "the evolution of SMS and Instant Messaging." Approximately 82% of cell phones in North America and 95% of cell phones in Western Europe have a camera built directly into the device [5]. According to a 2009 report published by ABI Research [13], the number of GPS enabled devices was expected to grow by 240 million units which is a 6.4 percent increase from 2008. As a result, developers have created applications to take advantage of a camera enabled cell phone.

Third party applications, such asTwitPic.com, have been developed to facilitate the posting of multimedia content to Twitter. TwitPic specifically enables the user to post photos to their TwitPic account. A message is then passed to their Twitter account that a photograph has been posted. Several TwitPic applications have been designed for the various mobile phone platforms such as the iPhone, Blackberry, and Android. As seen in figure 3, the user and the TwitPic service perform several steps to share their photograph.

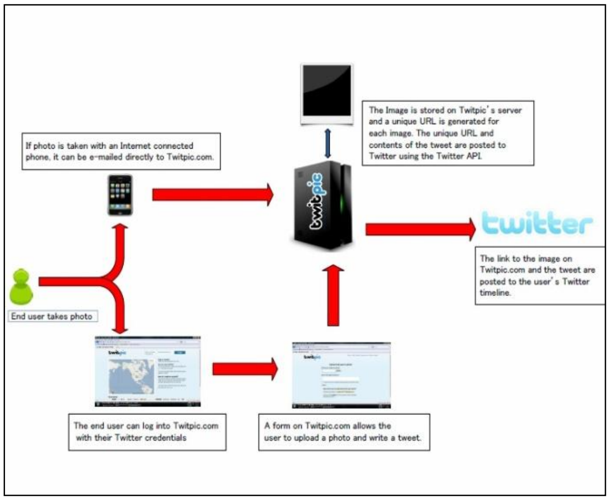
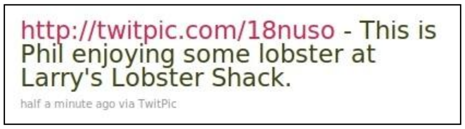

If photo is taken with an Internet connected phone, it can be e-mailed directly to Twitpic.com.
The Image is stored on Twitpic's server and a unique URL is generated for each image. The unique URL and contents of the tweet are posted to Twitter using the Twitter API.
twitpic
twitter
The link to the image on Twitpic.com and the tweet are posted to the user's Twitter timeline.
End user takes photo
The end user can log into Twitpic.com with their Twitter credentials
A form on Twitpic.com allows the user to upload a photo and write a tweet.

Figure 3- Steps performed to upload a photograph to TwitPic

When a user takes a photograph, they have two options when they want to post the photo to their TwitPic account. The first option allows them to upload the photograph from their computer by logging into their TwitPic account via the Internet. The second option enables a user, with an Internet enabled smart phone to upload the photograph directly from their phone to their Twitpic account. Once the photo is uploaded to the TwitPic account, the TwitPic server passes a message to the user's Twitter account. This message contains several items: the user's Twitter authentication information, a URL to the photograph on TwitPic's servers, and a personal message. Finally, Twitter's servers post a tweet containing all of the information submitted by the user.

The URL placed in the Tweet links to a scaled-down image of the original photo. For example, if the original's resolution was 3264 x 2448, the Tweet, seen in figure 4, would contain a link to a scaled-down image of 600 x 450. The scaled-down image does not contain an EXIF header. However, it is possible to download the original image, which can contain the EXIF header, uploaded to TwitPic by appending "/full" to the end of the URL. The API (Application programming interface), provided by Twitter, enabled the researchers to extract the necessary information from the photos found at the "full" URL.



http://twitpic.com/18nuso - This is Phil enjoying some lobster at Larry's Lobster Shack.

Figure 4 - Tweet containing a URL to a scaled down image

Twitter and TwitPic provide APIs that allow developers of various third-party applications to integrate with both applications. The APIs supplied by the companies have accelerated the adoption of both of these platforms by enabling developers to create applications for multiple platforms, such as iPhone, Blackberry, and Android, which integrate seamlessly with Twitter and TwitPic. The TwitterApp website lists 275 applications for Twitter, examples include applications like Tweetdeck, Twitterfeed and Twitterholic [15].

Two tools were utilized by the researchers to download and extract the EXIF data located in geotagged photographs. First, a Python script was developed which leveraged the Twitter API. The script performed a simple set of tasks: it located, isolated, and downloaded the images from TwitPic. Next, an application called ExifTool was utilized. ExifTool is a library written in the Perl programming language that allows for the programmatic reading, writing, and editing of metadata in a wide variety of formats [4]. The ExifTool application was used to extract the EXIF data contained in the image. This data was then stored in a file for later analysis. The process of obtaining the GPS data was completed in two stages.

The first stage of the overall process for obtaining the images was to obtain original photos from TwitPic. It is necessary to construct the URL of the original image due to the EXIF data being stripped from the downsized image. The pieces of the process functioned as followed:

- Step 1 - The script would search for the appearance of twitpic.com in a tweet.
- Step 2 – Fifty results were returned.
- Step 3 – Step 1 and Step 2 were placed in a loop, which executed one hundred times.
- Step 4 – The list of returned results were parsed.
- Step 5 – The URL of the down-sized image (lower resolution) was extracted.
- Step 6- URL was augmented to produce the location of the original image.
- Step 7 – The image found at the original image URL was then downloaded directly from TwitPic and stored in a directory that corresponded with the date it was visited.

During stage 2 of the process, the original images were passed to the ExifTool utility. The program extracted the EXIF data from the original image. For back-up purposes, all extracted EXIF data was stored to two locations. A text file was created locally for each image that contained the data found in the EXIF header. In addition, the data was stored in a MySQL database that enabled searching and indexing of the data.

The process was housed on an Ubuntu 9.10 Linux server. Utilizing the UNIX scheduling tool called cron, the process was set to run repeatedly throughout the day from November 29, 2009 to January 12, 2010. A total of 417,056 images were downloaded and processed. An average of 9,698 images downloaded and processed per day. The Twitter API does allow for limiting of search results based upon geography, however, the geographic area is self-reported by the user. As a result, there is a possibility of the geographic area being incomplete or incorrect due to user error. For the purpose of this study, no geographic limitation was placed upon the search results; the collected data includes images from around the globe.

Each time the cron job was executed, the first stage would gather 5000 TwitPic links. The result was between 4800 and 4900 images for each execution of the first stage. There are several identifying factors for these results. There were times that the URL of an image location was broken. In addition, only individual photos were downloaded; thus, links to groups of photos were excluded. As seen in figure 5, of the 417,056 images collected and analyzed, 14,074 contained the GPS position EXIF header, representing 3.3% of the images examined. Of the 14,074 images that were analyzed, 75% of the images uploaded to TwitPic containing GPS results came from the Apple iPhone. The Palm Pre, the LG, and the HTC RAPH800 all represented 3% of the GPS containing results. These results are consistent with the 2010 study conducted by Ullrich.

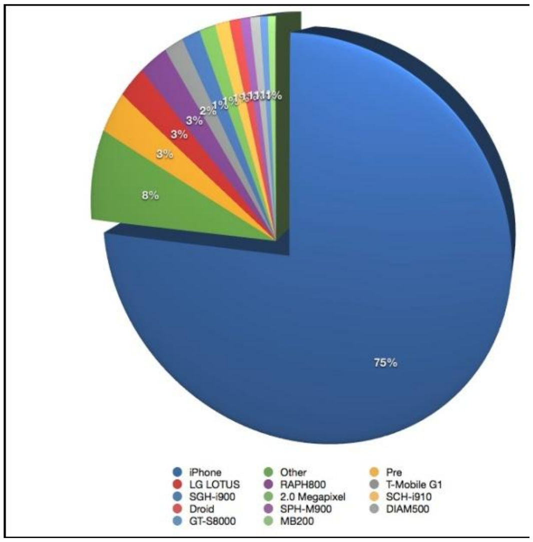


Figure 5 – Images collected and analyzed from TwitPic

CONCLUSION AND FUTURE RESEARCH

The current exploratory study revealed information gathered from examining EXIF headers obtained from images uploaded to TwitPic. Although the number may seem small, a total of 14,074 images, 3.3% contained embedded GPS coordinates in the photo showing where the picture was taken. The current study was limited to examining images uploaded to TwitPic over a six-week period. The researchers plan to continue analyzing the images uploaded to TwitPic over a longer period of time. The researchers would also like to examine the ability to correlate the text of the tweet with the image and the embedded EXIF data in the image to see if it would be possible to further analyze the potential misuse of this information. Additionally, as previously mentioned, the study was not limited in geography, so additional work needs done to understand the prevalence of this information by geographic area.

REFERENCES

1. Clayson, J. (2007). Safeguarding against and responding to the breach of personally identifiable information retrieved on January 6, 2010 from http://www.whitehouse.gov/omb/memoranda/fy2007/m07-16.pdf.

2. Evans, A., Martin, K., & Poatsy, M. A. (2010). *Technology in action* (6th ed.). Upper Saddle River, N.J.: Pearson/Prentice Hall.

3. Fletcher, D. (2010). Please Rob Me: Site shows dangers of foursquare, Twitter. *Time Magazine.* Retrieved on March 5, 2010 from http://news.yahoo.com/s/time/20100218us_time/08599196487300

4. Harvey, Phil. (2010). Exiftool by Phil Harvey. Retrieved on January 20, 2010 from http://www.sno/phy.queensu.ca/-phil/exiftool/

5. Johnson, S. (2009, May 13). Tiny cameras have big market. Retrieved on March 15, 2010 from http://www.physorg.com/news161457142.html

6. Krishnamurthy, B., &Wills, C. (2009). On the Leakage of Personally Identifiable Information via Online Social Networks. WOSN Spain

7. Many Online Social (Anonymous, 2009, August). “Many online social networks leak personal information to tracking sites, new study shows” retrieved on January 19, 2010 from http://www.sciencedaily.com/releases/2009/08/090824151307.htm

8. Nations, D. (2010). What is geotagging? Retrieved on February 17, 2010 from http://webtrends.about.com/od/glossary/i/what-geotagging.htm?p=1

9. Paullet, K. (2009). An exploratory study of cyberstalking; students and law enforcement in Allegheny County. UMI 3376412 retrieved March 22, 2010 from Dissertations and Thesis Database.

10. Please Rob Me. (Anonymous, 2010). Please rob me. Raising awareness about over-sharing. Retrieved on March 15, 2010 from http://pleaserobme.com

11. Razavi, M., & Iverson, L. (2009). Improving personal privacy in social systems with people tagging.

12. Safran, C., GarciaBarrios, V.M., and Ebner, M. (2009). The benefits of Geo-Tagging and Microblogging in m-Learning: A Use Case.

13. Schonfeld, E. (2009). Twitter reaches 44.5 million people worldwide in June retrieved from http://techcrunch.com/2009/08/03/twitter-reaches-445-million-people-Worldwide-in-june-comscore on February 3, 2010.

14. Suarez, R. (2009). Record setting cyber theft stirs questions on security. Retrieved on March 3, 2010 from http://www.pbs.org/newshour/bb/business/july-December09/hacker_0818.html

15. The Only Twitter. (Anonymous, 2010). The only Twitter application list you'll ever need. Retrieved on March 25, 2010 from hhtp://www.squidoo.com/twitterapps

16. Twitter is a Real (Anonymous, 2010). Twitter is a real-time information network powered by people all around the world that lets you share and discover what's happening now. Retrieved on February 16, 2010 from http://twitter.com/about

17. Ullrich, J. (2010). Twitpic, Exif & GPS: I know where you did it last summer. Retrieved on February 10, 2010 from http://isc.sans.org/diary.html?story/id=8203.