

Developing a Forensics Tool for Social Media

Theodore Casser
Applied Information Technology
University of Baltimore
Baltimore, MD 21201
tjcasser@gmail.com

Mohammed Ketel
Applied Information Technology
University of Baltimore
Baltimore, MD 21201
mketel@ubalt.edu

ABSTRACT

Millions of users around the world utilize social media sites on any given day, spreading information about their activities, whereabouts and thoughts to friends and interested readers. These same messages can be used to construct a digital and physical path that can be extracted for forensic analysis through application programming interfaces provided by each of the social media outlets. While there has been recent work discussing the spread of social media as a means of tracking news and trends in the world at large, little has been done to study a means to analyze the data available through social media using forensic methods. To fill this gap, an application has been created that can retrieve data created by users via social media applications and allow analysis of the same.

Categories and Subject Descriptors

D.4.6 [Operating Systems]: Security and Protection

K.6.5 [Management of Computing and Information Systems]: Security and Protection

General Terms

Design, Security

Keywords

Forensics, Social media, Security

1. INTRODUCTION

With the increasing usage of social networks, digital forensics faces novel problems and challenges. The number of users of these services increases steadily [3, 4, 8, and 9]. While traditional forensics relies on the physical acquisition of hardware [2] to ensure evidence, this approach does not scale to social networks.

Social media has become the “killer app” of the Internet. Not a day seems to go by without hearing mention of one social media site or another on the news, in the newspaper, or as part of advertising for some product. According to the latest figures from Alexa.com, a leading online traffic-monitoring site, five social media sites rank among the twelve top sites in terms of user traffic. Facebook ranks second in total unique visitors, with only Google’s search site ranking above it in terms of traffic [1].

As referenced by the title of a recent popular film, these sites are used for the expressed purpose of spreading one’s social network.

ACM SE’14, Mar 28-29, 2014, Kennesaw, GA, USA

ACM 978-1-4503-2923-1/14/03.

<http://dx.doi.org/10.1145/2638404.2638491>

Users post their minutiae in ways that had once been left to the province of letters, phone calls and face-to-face conversation. The difference between these styles is the availability of this information and its permanence in the public domain.

Security and data protection in social networks has recently become an active research area. Many researchers have outlined the potential threats and risks associated with using social networking services [5, 7].

2. BACKGROUND

Most uses of social media in the court room tend to be limited to single incriminating messages. Many of these incidents could be attributed to users utilizing social media as if it were email. However, there are times when evidence is not limited to a single message, but rather is a pattern of behavior, best confirmed by discovery of a larger data-set.

A complicating issue with any forensics work against social media is that the user does not have true possession of the information. Deleting a message on social media is still as easy as deleting a file on a local file system. While forensic tools can deal with the deletion on a hard drive by searching sectors that have not yet been overwritten, social media sites are remote systems that collect data in multiple locations, making finding deleted messages impossible.

In the first case, it is often difficult to use the built-in tools for each of the services to find evidence. In the second case, a savvy user might lie to a third party and hide having an account. Further, some states have passed laws making it illegal to require access to a student or employee’s social media accounts, which removes a tool from the school or employer’s arsenal to monitor conduct.

The solution to the first two cases is a tool that can search and cross-reference accounts between services. A forensic application would allow data on searches to be stored and hashed in a fashion that ensures non-manipulation of retrieved information. The tool would go further in presenting an interface to the networks’ search mechanisms to ensure a complete search that allows for discovery of information without alerting the investigated individual.

The third case – deletion of messages – is harder to deal with, and cannot be addressed by this project. Such is the limit of a tool that operates against a network data source – no application not constantly monitoring an account will guarantee capture of all posts made by a given user.

2.1 Prior Research in the Field

There are few papers published thus far in the field of performing forensics against social media networks. Most of the papers that

cover the issue of security in social media concern themselves with the manner in which the users leave themselves vulnerable.

There has been some work by Markus Huber and his colleagues regarding creating a forensic tool for Facebook in specific [6]. Huber's team concentrated on finding methods to retrieve information in a manner that did not trip Facebook's security. Their research is a helpful guide in the creation of a tool that can mine multiple networks, but differs in significant ways:

- The tool created for the paper only accessed Facebook.
- The tool was developed to mine an account based on information retrieved from a local machine.
- The 'social snapshot' tool that they developed required the account credentials of the target account.

3. DESIGN

This research is limited to three of the most popular social media networks – Twitter, foursquare and Facebook. Each of these sites occupies a different space, yet are often used in conjunction with one another. Only publicly documented APIs were used. There were no coding workarounds created to avoid any security measures to avoid running afoul of the terms of service of the sites. The program was configured to be read-only. Finally, the application requires valid login credentials to the underlying services, though not the investigated user's credentials.

3.1 Software and Hardware Design

Each of these sites also makes available an application-programming interface (API) that is meant to interface with user information on their database to allow for integration with third party sites. We have chosen to use the PHP interfaces to the services to allow for creating a web application to conduct forensic investigations. There is little difference in the functionality provided across the different language sets, but the ease of editing a web application and its subsequent ability to be offered online using that language led to its selection for this project. A commercial-grade application meant for desktop use might better utilize Java or .NET depending on the environment in a forensic laboratory. As this research is done with the intent of creating a prototype application, and not indeed a production-ready version, we felt that using a programming language that permits rapid development and adjustment through un-compiled, scripted code would be an acceptable prototyping method.

The resulting web application is hosted on an Apple Macintosh Mini running Apache 2, available over a cable modem line to the Internet. Since this application is intended as a prototype for what a real social media forensic package might look like, choice is made to pursue a course that would resemble the real installation as closely as possible. MacOS is descended from the BSD Unix variant and as such, is similar to the majority of servers on the Internet today. Likewise, Apache's httpd package is far and away the most popular web server software in use across all sites and is easily configured to allow execution of PHP files. The connectivity to the Internet is admittedly less than optimal; while weighing the design decisions for this project, however, it occurred that not all potential (legal) clients of such a package might have a large-bandwidth connection to the Internet, so a slower connection may be deemed acceptable.

3.2 Targeted Social Media Networks

As previously mentioned, the application targets three of the most popular social media applications – Twitter, Foursquare and Facebook. These three networks were chosen due to their unique places in the social media patchwork and the overlap in usage by users – users who have an account on foursquare generally also have one on Facebook or Twitter as well, and users on Twitter and Facebook often cross-post their information between the services.

3.3 Application Ground Rules

Several rules have been established before proceeding with this project.

Only publicly documented APIs were used. There were no coding workarounds created to avoid any security measures to avoid running afoul of the terms of service of the sites. The goal of the application was to allow searches against the information stored on each network in a manner that would be admissible in court, and data that is obtained through undocumented breaches in security likely would not fall into that category.

Second, the programs were configured to be read-only. The three web applications in question all permit their APIs to be used for posting to the services, given proper permissions.

Third, in places where required, the application requires valid login credentials to the underlying service that is being browsed, though not the investigated user's credentials. For all three networks, no login is required to inspect some publicly accessible data. However, much more data is available once a user has connected to the system with valid credentials.

4. RESULTS

The application was broken into three parts, each concentrating on one of the three social media frameworks. Results found with each of the three social media frameworks will be presented in the following sections in order of ascending popularity.

4.1 Foursquare


Foursquare's security regarding data is well built-out – it is impossible to ask for the comprehensive list of venues visited by any other than the authenticated user.

There is a major failing to the design of foursquare's user information API – one can search for a user based on membership in other social media sites or personal information such as a name, phone number or email address. Once a user is found, all of their personal information that has been entered – including phone and other social network IDs - is returned.

A user search also returns information regarding where the user is 'mayor' (most frequent check-ins in the last sixty days). This information does not reveal when they were last at a location or any pattern to the visits. Given the nature of foursquare and its encouragement for gathering 'badges' based on locales visited, there is ample incentive for users to claim mayorships as frequently as they are able, which can be retrieved by anyone through the public API.

Information available through foursquare helps to further investigations in other ways. Many users tend to claim mayorship of their home and workplace. If the addresses of their workplace and home were not known at the start of an investigation, the

On the other side of the equation, the location information API can be used to further refine information on a user once those locales frequented by a user are identified. Foursquare lists all users currently at a location to anyone who views the site, except for those who explicitly opt out of being listed. Foursquare also lists the total number of users who have visited a location and the number of times those users have checked-in. And, separate from these figures, foursquare displays a record of who the mayor is, and how frequently the mayor has visited during the prior sixty days.

Home City:	Thoudos Cuser
Phone:	Baltimore, MD
Twitter:	
Email:	tcuser@gmail.com
Photo:	
Badges:	63
Mayships:	16
<h2>Mayships</h2>	
Starbucks	2520 Quarry Lake Dr, Baltimore, MD 21209
Quarry Wine and Spirits	2516 Quarry Lake Drive, Baltimore, MD 21209
The Fresh Market	2510 Quarry Lake Dr, Baltimore, MD 21209
Our House	
McLanAC	, Owings Mills, MD
University of Baltimore - Langsdale Library	1420 Maryland Avenue, Baltimore, Maryland 21201
Starbucks @ UB Barnes & Noble	62 West Oliver St, Baltimore, MD 21217
School Of Information Arts and Technologies Labs	1420 N. Charles St, Baltimore, Maryland 21201
Gordon Plaza @ UB	Maryland Ave, Baltimore, Maryland 21201
Academic Center (AC)	1420 N. Charles St, Baltimore, Maryland 21201
UB Parking @ The Fitzgerald	62 W Oliver st, Baltimore, MD
Student Center (SC)	21 W Mount Royal Ave., Baltimore, MD 21201
The Associated	101 W Mt Royal Ave, Baltimore, MD 21201
Garco's At Quarry Lake	Greenspring Ave., Baltimore, MD
Valley Pediatrics	5 Park Center Ct, 300, Maryland 21117
American Jewish Congress - Maryland Chapter	3723 old court of, Pikesville, MD

There is one other opportunity to track a user's foursquare information. It is possible, if a user configures it, to have foursquare automatically broadcast check-ins to both Twitter and Facebook, leaving a trail behind in those services.

- A user without a profile image cannot be mayor, even if they are the most frequent visitor to a venue.
- Private check-ins are unavailable for browsing and do not count towards claiming the mayorship.

It should be noted that users are able to delete check-ins from their history using the web interface. There is no way to retrieve

By its nature, Twitter is publicly accessible – one does not need an account in most circumstances to view the messages being sent out by a given user. It is possible to set an account private and require that users request access; however, most users do not take this step. As such, there is an opportunity to analyze the flow of messages, estimated by Neil Savage [8] in his article as being nearly ninety million messages per day.

For any public user, it is easy to get a list of all messages sent out, including the responses to messages sent by others. Each message features useful information from a forensic standpoint – it includes the timestamp, the content of the message, if it was a ‘retweet’ (reposting) and metadata pertaining to the message. Some of the metadata is useful forensically such as the client application used to post the message, which can be used to help with planning further steps in an investigation. Further, the messages themselves may offer up other hints as to what the user is interested in, which may correlate to other activities being investigated.

Search for Users View Timeline			
Username: <input type="text" value="HowardKramer"/>			
<input type="button" value="Search"/>			
Date	Tweet	Client	Location
Mon Apr 29 19:39:38 +0000 2012	Good talk @smonmairing at #firstdata conference. Fr 40Knot score.	Twitter for iPhone	Map It
Mon Apr 30 16:58:07 +0000 2012	I gave @Kramer +K about Carling on @kikout http://t.co/9Y6uqW	Twitter Button	
Mon Apr 30 16:56:35 +0000 2012	I'm huge. According to @kikout, my Kikout score is 40. How influential are you? http://t.co/ZZX8TUD	Twitter Button	
Mon Apr 30 12:27:37 +0000 2012	#tuesday in Camps Square before work... Seriously http://t.co/0Yp6gW	Twitter for iPhone	Map It
Sun Apr 29 12:56:44 +0000 2012	I'm watching Jonelle Monks at New Orleans Jazz Festival 2012	iOS	
Sun Apr 29 10:47:56 +0000 2012	Listening to Dr. John. Must be the right place...	Twitter for iPhone	Map It
Sun Apr 29 18:49:34 +0000 2012	was in the right place... Must of been the wrong time. http://t.co/PW4T5D	FourSquare	Map It
Sun Apr 29 10:12:47 +0000 2012	Too hip for me. We'll see. @ Green Goddess http://t.co/UPML8ic	FourSquare	Map It
Sat Apr 28 17:21:20 +0000 2012	I'm at Desire Oyster Bar New Orleans, LA http://t.co/04W4GCE	FourSquare	Map It
Sat Apr 28 18:46:44 +0000 2012	I'm at Cafe Beignet New Orleans, LA http://t.co/0R0z5P	FourSquare	Map It
Sat Apr 28 12:09:35 +0000 2012	I'm at The Saint Hotel New Orleans, LA http://t.co/5779z6L7Y	FourSquare	Map It
Sat Apr 28 04:12:34 +0000 2012	I'm at Mercedes-Benz Superdome New Orleans, LA http://t.co/7eQ6MSA	FourSquare	Map It
Fri Apr 27 18:37:26 +0000 2012	Do you know someone with super Process Management experience? Check this out (and what an awesome person to work for... http://t.co/0mepHk5)	LinkedIn	
Fri Apr 27 14:12:03 +0000 2012	My @Kikout score is 27. I improved it by 7 points over the past day http://t.co/9v6p5AN	Twitter Button	
Fri Apr 27 12:09:43 +0000 2012	According to @kikout, my Kikout score is 20. Sounds about right. http://t.co/PyHnGy	Twitter Button	
Fri Apr 27 19:26:35 +0000 2012	@HowardKramer has earned 1 achievements on the new @kikout http://t.co/QB5gWlao	Twitter Button	HowardKramer
Thu Apr 26 13:45:48 +0000 2012	Happy MCS Day! #HTTR	web	
Thu Apr 26 13:42:35 +0000 2012	Elay and its PayPal's positioned to grow http://t.co/14PMkY	Twitter Button	
Thu Apr 19 23:44:46 +0000 2012	These people are nuts about LAX. @ Genesee Valley w/ 2 others! http://t.co/NDWw5Cg	FourSquare	Map It

Figure 2. Twitter with geolocation

A less-frequently used feature with Twitter is the ability to embed geo-location information within a given message. For many users, the only location information in a message is the location set in their profile, which may or may not be useful – some users use geographic coordinates, while others use less descriptive terms. Some clients, such as foursquare, post a separate, message-specific set of coordinates. If a user frequently posts from foursquare, one can use Twitter to accurately map a user's routine. When one integrates these results with Google Maps, it is possible to generate a map of their schedule and route.

There are two principle ways information can be hidden on Twitter:

- A user's direct messages are not available through the public API – only those of the logged-in user.
- Private users' message streams are hidden from public view. Access to these feeds requires explicit approval from the user.

A cautious user can also go back and delete prior tweets within a limited period of time. These deleted tweets may, however, linger on through their passage to other social media, such as Facebook, or through archiving on dedicated sites on the Internet that monitor Twitter.

4.3 Facebook

Despite recent issues with privacy, the API offered by Facebook is robust from a security standpoint. Users are provided tools to lock down the information shared with the world, which has been tightened in the wake of a few high profile incidents. Nearly every call requires a user to log in to execute queries against their database. It should be noted that the security settings are not quite as intuitive as they ought be, which leads to many users revealing more information than they intend.

Searching for a user is accomplished via simple string matching. The information available through a search, however, is sparser than that offered by the other two networks profiled – it is limited to the user ID and the user's real name, but not an easily referenced user name.

By default, users are set with their access open to all. This allows an investigator to browse through a user's 'wall' and status messages. Depending on permissions, there is also the ability to view lists of friends, uploaded multimedia, and the list of other users' multimedia in which they have been 'tagged'.

The greatest use of Facebook as a forensic data source is for the running stream of information. The strength of Facebook's wall and feed is the ability to see the conversation that comes from the user's comments. Where Twitter offers an ability to reply to a message, Facebook has a threaded conversation stream involving all participants that is picked up by the methods exposed through their API, allowing the investigator to see the conversation in its entirety.

Some difficulties exist with getting information from Facebook:

- Individual posts to a user's timeline can be narrowed down to limit access to smaller groups than the default, which may prevent an investigator from seeing all information on an otherwise public timeline.
- The direct means of querying Facebook only returns information on the user's timeline.
- It is not easily apparent how to query applications hosted by Facebook for information.

Facebook presents users with the most straightforward and obvious mechanism for deleting content out of the three services. Every post is annotated with a button that allows users to choose to delete a message. Such messages are unable to be retrieved through any means.

Facebook presents an advantage not offered by either of the other services utilized for this research – it does not truly delete accounts. Unlike the other services, requesting deletion of an account simply turns an account 'inactive' rather than deleting it. While this research did not consider this case, retrieval from a 'deleted' account should be the same as with an active account.

5. CONCLUSION

Social media can be a rich source of information in conducting an investigation, if one knows where to search and what data can be captured. The information available from a given social media source varies based on security settings, but all three sites examined yielded at least some useful information.

There are challenges that need to be overcome with regards to the limitations presented by each site, but with some effort, information on a user's activities and interests can be gleaned. Likewise, there are still some points where the application can be further refined to help to develop a robust tool.

6. REFERENCES

- [1] Alexa Top 500 Global Sites, www.alexa.com/topsites/global
- [2] B. Carrier. File System Forensic Analysis. Addison-Wesley Professional, 2005.
- [3] Facebook statistics, www.statisticbrain.com/facebook-statistics/
- [4] Foursquare statistics, www.factbrowser.com/tags/foursquare/
- [5] H. Gao et al., "Security Issues in Online Social Networks," IEEE Internet Computing, vol. 15, no. 4, 2011.
- [6] Markus Huber et al., "Social Snapshots: Digital Forensics for Online Social Networks," ACSAC '11 Proceedings of the 27th Annual Computer Security Applications Conference, 2011.
- [7] Long Jin et al, "Understanding User Behavior in Online Social Networks: A Survey," IEEE Communications Magazine, 2013.
- [8] Neil Savage, "Twitter as Medium and Message," Communications of the ACM, 54, 18- 20, 2011.
- [9] Twitter statistics, www.statisticbrain.com/twitter-statistics/