Jestr

Journal of Engineering Science and Technology Review 11 (2) (2018) 126 - 134

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Impact of AnonStalk (Anonymous Stalking) on users of Social Media: a Case Study

V. Kanakaris*, K. Tzovelekis and D. V. Bandekas

Department of Electrical Engineering, Eastern Macedonia and Thrace Institute of Technology, Kavala, Greece

Received 1 December 2017; Accepted 22 April 2018

Abstract

In the period that we are living, there is a massive exchange of information, while the communication has drastically changed as more and more people using the social networks. Using this kind of technology undoubtedly leads to exposure of personal information on the part of users, enabling the use of Open-Source Intelligence (OSINT) techniques in order to obtained sensitive data from not authorised users. Separating the medium itself from how it is used by users, we consider how possible it is for a malicious user (intruder) to extract personal data in an automated method, and use them against an unsuspicious user of social media. By this way users can be deceived as well as the user's followers by watching different data from the real one as the attacker has the chance to modify them. Moreover, the exposed users lose any sense of privacy as they become victims, penetrating to their personal information.
The paper describes the architecture and implementation of a method that collects data anonymously from Twitter and Instagram using Twitter REST API and "geoJSON" along with "Instagram Real-Time API" and "Genymotion". After the collection and the processing of data, the output is representing in a graphical map that can be easily accessible and provide all the information to the attacker. In terms of anonymity we followed various procedures that a malicious user could follow in order to hide his tracks and create accounts in both platforms that can be used in the implemented application.

The results show that the intruder can retrieve sensitive personal data (user location via tweets / instas from smartphone). In addition, the ability to map the activity could allow a malicious user to track an unsuspecting person's activity and predict their future locations. This research also revealed that anyone with basic knowledge of computer becomes anonymous intruder of our personal life, causing fear and horror to those that are unsheltered.

Keywords: Twitter; Instagram; Social Media; OSINT techniques; Geo-Location

1 Introduction

Through the years, intelligence is a science that has been well recognized because it always plays an important role in military and business wars. On the military purposes, intelligence is focusing on predicting the actions of enemies or evaluation of situation, through a constantly accumulating information procedure. Hence, intelligence should act with credibility and increase the analysis process. A continuously gathering of information, with reliability and a precise analysis, there will be an amazing improvement on whatever field might be military or business. Consequently, authorities be aware of compilation and breakdown of intelligence.
Generally, intelligence is gathered by professionals, which also undertake the analysis of the collected data 0. The process of analysis acquires good knowledge of analysis techniques, and such knowledge can be found from those who worked in a government organization or commercial industry. Experienced analysts would be great asset to organizations, but due to high cost of recruitment makes them potent only in large scale of organizations and enterprises. The solution for the above-mentioned problem, is the open source intelligence (OSINT) approach for intelligence management, because gathered intelligence easy update and in vast amount. OSINT differs from the traditional intelligence since concentrates data via public accessible corporations without limitation on the access and usage. OSINT gathers information from sources such as social networks, media and web communities [2]. Open source intelligence has the following specifications:

- The sources are publicly available.
- The main purpose is to accomplish specific intelligence needs.
- General functions include gathering, usage, and dissemination.

Open-source intelligence (OSINT), gathers process information processing from open sources for intelligence intentions. Data from social media are monitored for law enforcement purposes in order to prevent and detect terrorist activity. Albeit, the data from social media may be accessed from everyone, it still may constitute copyright violations. A potential way to confront these challenges is to use OSINT tools that integrate legal requirements. Most of OSINT tools that is used have a legally compliant design and meet a variety of requirements on different end-user groups.

Social Networks impel users to promptly deal with information on largely scattered networks. Users can post

*E-mail address: vkanak@teiemt.gr
ISSN: 1791-2377
doi:10.25103/jestr.112.18

content in a diversity of formats, which can be instantly made online to social network [3][4]. Hence, social networks have become significant way for promulgation of information, web content discovery, discussion, opinion sharing. The vast amount of public data that run through social network has the eventuality to release precious new insight to the academic community, marketing agencies, concerned in comprehension of online behaviour and following social tendencies [3][5].

All the above lead to the outcome that everyone can access important personal information or links through them. There are risks and responsibilities that may leave the users in difficult situations with regard to the protection of information on social network platforms and the privacy of personal data. The rise of security risks because of spread of information exchange services on the internet, increasing amount of information, and the fast growth in information and communication technologies has created the protection of privacy to be one of the most debatable and alarming issues [6][7]. Nowadays, the biggest amount of information in the internet derived from social networking platforms. The primary reason for the need to be aware about the use of social networks and linking the importance of privacy is the abuse of personal information by social network platforms or the bad use of the viewable content by other users. Cognizance by the content owner regarding the management of digital information has an essential role in the protection of personal data. It is very difficult to anticipate and track around the world the information that disseminated via internet would be used within a few minutes and how many copies of the information would be produced [8]. Although social media platforms provide to their user's some privacy setting tools, on Twitter these tools the only protection that provide is by making visible the user's information to his followers and not all Twitter users. On Instagram there is no specific protection regarding the user's information and if you are using the platform on E.U. the user has some kind of privacy policy instead of U.S. where in some states unavailable[9][10]. Additional, many social networks ameliorate their advertising policies using the personal data they have already received and they put ads on the site according to the personal interests of the users. Use of personal information for such reasons is clearly stated in the user agreement accepted by the user when signing up to the social platform. Furthermore, some service providers can change the agreement without having the confirmation of the end user [11].

Twitter is a platform that allows mobile users to embed their precise geo-location into short textual updates known as tweets. The ability to collect and mine location aware tweets opens new research directions by making it possible to study the spatial as well as textual characteristics of online content. It also provides a means for monitoring social trends and online activity on a regional basis. Before the recent smart-phone boom when mobile access to social networks was limited, Twitter enabled anyone with access to a cell-phone to communicate rapidly with widely dispersed networks of people through SMS [12][13].

As the volume of content shared publicly on social networks continues to grow, the demand for technology that can assist with the collection and mining of this content. Twitter's potential as a tool for research and analysis is underlined by its rapid growth and emergence as a mainstream channel of communication on the Internet. Most existing research into Twitter has focused on social network analysis based on the analysis of usage patterns and the textual content contained within tweets [14][15].

Due to the public nature of most user generated content that flows through its network, Twitter is a particularly useful source for intelligence gathering and large scale data analysis. This is in sharp contrast to Facebook, where user generated content is only made available to pre-selected lists of users [16][17].

Instagram started in 2010, based on previous social network platform called Burbn, that its main function was to allow users to check-in any available smartphone that has enable the GPS or share their location with friends. On the first period of its function added features, such as photo manipulation by offering to the users a variety of filters that could be applied to photographs, caption adding ability, comments and tags. Today, Instagram offers users to get a photo, edit and share by publishing location information. Moreover, as opposed to the other platforms such as Picassa and Flickr which are web- based,

The Instagram Application Programming Interface (API) allow developers to create desktop, web-based and mobile applications that give users the opportunity to have access on features that primarily not provided by the official mobile applications. Particularly, third party mobile applications provide to users a variety set of features of processing photos, which sometimes differ from main function of the official Instagram mobile applications.

This research focuses on the implementation of a method that gathers data anonymously for Twitter and Instagram. These two social platforms and their use is more often on a modern smartphone. Users of Instagram can send a photo or a video of 15 seconds and add a comment with limitation of 2200 characters. The specific toolkit was planned to supply researchers with access to respective information of Twitter and Instagram, in a format convenient for analysis and data mining. The system encloses modules for data collection, spatial offline data storage and retrieval, full-text search, geo-location data, data mapping and export [4][18].

Generally, information regarding location extracted:

- from user's profile data
- from user's location that messages written

On Twitter and Facebook location data can be extracted from user's profile and the visualization of the user's location on the map can be made by adding at coordinates or geo-location data. However, on Instagram, the location from the massage provides longitude and latitude information, avoiding of adding extra data [19][20]. The extraction procedure on Instagram, is done by using the information of the current user's location or from the location that the messages was written. A famous place like Volos (Greece) is interpreted different by the Twitter and Facebook as it means that is very famous among all the locations of the user's friends, while on Instagram means that user is usually located in Volos.

Firstly, the paper focuses on modules for concentrating data of Twitter using the REST API and "geoJSON" in order to extract information about the location of user and concurrent alteration of this data to a spatial format. While real-time Twitter data gathering features have been ameliorated into a number of applications, remarkably, there is a deficiency of tools regarding saving data in a format that trigger users to execute advanced spatial queries. Second, the paper presents modules such as "Genymotion" and

"Instagram Real-Time API" which together can extract data regarding the location of the user. These methods are described as operating at the time of this research took place from July to October 2017.the process of working without leaving traces on the Internet. The third section of the paper describes the process of working without leaving traces on the Internet. Finally, the paper outlines methodology used for data visualization and export. It summarizes, providing limitations and test usage of the method [14][15].

2 Related Work

Many researchers interest mainly on the development of an automatic scheme for OSINT processing. Since OSINT concerns about discovering useful intelligence from large amount of data, data mining techniques are naturally involved in this area. However, there are still few works until now. An Italian cooperation, developed an OSINT platform named SPYWatch [16][17][21] which applied K-means algorithm in order to cluster, classify, and process multilingual documents. Pfeiffer et. al. [22] developed a system called Media Mining System based on MPEG-7, which combines sources such as satellite images, TV images, Web pages, and RSS feeds in order to produce outcomes for early alert, data sharing, and risk assessment. Vincen et al. [23] proposed a system that merge information from various sources, by using probabilistic enhanced scheme in order to get knowledge condition and conflict estimation. Badia et al. [24] analyzed the sentences of documents with the aim to obtain the information about the space and time for OSINT. Dawoud et al. [25] incorporated many data sources from social network analysis to survey the correlations of terrorist networks. Neri et al. [26] analyzed the scandal of Italian Prime Minister in order to present the tactics in analyzing, tagging, and clustering news articles to detect the correlations of them. Kotzias et al. [27] introduced an optimized work with minimum amount of queries and focused on three categories:

- Location identification of a given user
- Location identification of a personal tweet
- Creating model of the spatial density of users.

Eisensten et al. [28] try to solve the user geo-location extraction through geographical topic models by capturing the difference in the language use for a particular theme among user from distant regions. Ahmed et al. [29] proposed a scheme of hierarchical structure of the topics by categorizing them and extracting location-specific topics and place users. In the same way, Cheng et al. [30] used a probabilistic framework for estimating a Twitters user's city-level location based on phrases and not the topics, thus the proposed approach no need to use IP information or external knowledge. Mahmud et al. [31] improve the previous method by using Naives Bayes classifier in order to predict the country, state and city or time zone of a user. Ren et al. [32] uses an approach that estimates a user's location associate with the most of his friend in an accuracy of 56,6% within 100 miles in city-level and 45.2% within 25 miles in town level using ILF and RW filtering. All the above mentioned researches are based on new developed platforms while our approach uses tools and applications with few modifications in order to get full anonymity. Moreover, all these approaches are based on collecting big amount of data and then use several technics such as algorithms [16][17][21] or probabilistic schemes Vincen et al. [23] to extract the information regarding location.

The second category of relevant research focuses on the geolocation of the tweets. Ikawa et al. [34] tried to estimate the location of a tweet by consorting location and relevant keyword for previous messages in order to predict the new message location. Kinsela et al. [35] constructed language models of locations using coordinates extracted from geotagged Twitter data and model locations at varying levels of granularity, from zip code to the country level in order to find the location of the user and the tweet. In order to use language model with Terrier they extended Terrier and implemented a query likelihood Language Model with Dirichlet smoothing. This category focuses on location data that accumulating from the association of older messages and its relative information that have regarding location. Our study uses free software that extracts automatically the geo-location information without creating any new model to predict this information.

The third group of relevant approaches focuses on modeling the spatial density of users. Cho et al. [36] developed a model of human mobility that combined periodic short range movements with travel. By modeling user's locations with a mixture of Gaussians centered at "home" and "work" locations they managed to get the diversity of user's behavioral pattern. Lichman et al. [37] proposed and investigated a systematic framework for modeling human location data at individual level using kernel density estimation (KDE) methods. By this way, they improved the previous approach of Cho et al. [36] and avoided the data sparsity. All the above approaches use models and algorithms that predict or show the behaviour of the users regarding the location, but without providing the exact location of the users as our research provides.

The basic difference between our approach and all the above mentioned researches is that they focus:

- on the geo-location of Instagram users and tweets that are associated with the provided area,
- they implemented new algorithms in order to predict as much as many information regarding the location of the messages written on Twitter and the photos that posted on Instagram
- they estimate the location by modeling the behavior of Instagram and Twitter users

while our approach uses free apps, that everyone easily finds them and without need of developing a new one, in order to get information about the geo-location of the user that is used as input, avoid tracking our information (reverse tracking) and keep our anonymity.

The main question on this research is to evaluate the possibility of a malicious user to monitor, without the permission of the target user, in a complete stealth and untraceable way the activity, of the two famous social networks Twitter and Instagram and find out how vulnerable are the data in regard to location.

3. Background Theory

3.1. Instagram Real-Time API

On 2014 Instagram, had more than 200 million users per month, which were sharing more than 20 billion photos, almost 1.6 billion likes and 60 million photos posted each day [40]. In a daily basis Instagram used from smartphones,

deputizing a variety of interests and practices, like shared experiences, instant publication of images from the scene of the experience, tagging friends present, commenting on others' content; promoting photos, 15 seconds videos and tweets.

The Instagram API [41][42] supply a search hook concentrated on tags, offering an immediate comparative opportunity. Using the Instagram API for specific tags then offers results close to Twitter projects, albeit with different metadata, concluding in further methodological questions.

An Instagram API, (Figure 3) query offers a numerous of metadata for respective media shared on the platform. Every media object matching the tag query, meaning that the API returns not just its unique identifier (id) which links to the low and standard-resolution versions of the content (whether image or video), but also metadata that contains usernames, time and date of creation, caption, comments (and user and time information for comments), tags, likes, and location information when a user has geotagged their media. This kind of data permits quantitative and qualitative analyses, whether numbering the quantity of content over time, users, or tags, plotting media based on location data, or searching at the content of the media and their captions. The Instagram information contains more dynamic data points than tweets. Every image or video has its own data point, but, if a user replies to an image by writing a comment, that ends up an additional information to the original data point.

Classifying comments and reviewing the media posted on Instagram, despite how many comments the media might attract, there is no cohesion between data points: while the results can be saved in a database, studying variable comment threads which may alter during the route of the data collection is a novel methodological concern which does not affect Twitter research.

Searching for a specific tag will regain information regarding media published with the relevant tag. However, running the request alike will also provide results regarding the media which has the tag in comments even if the original caption does not relate to the tag, and if the comment and tag were published by the original user. Including the tag can offer to the media a publicity that didn't have formerly, incorporating the differences required for the future uses and intentions around tagging.

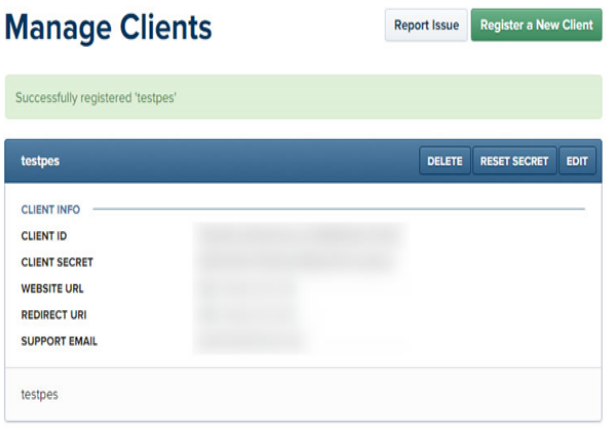

Fig. 1. API Keys from Instagram

With Instagram Realtime API [43][44], someone can control in real time the activity of users, tags and locations of media. In that case, when there is a seeking of a user via Instagram Real-Time API then information is extracted about the new publications. When there is a request about labels the exported information according the label then published in the respective media. Through the inquiring of locations there are new notifications regarding photos or videos uploaded and tagged with a specific location. Also, data can be extracted in relation to new photos posted to an arbitrary location.

3.2. Twitter REST API

Twitter is a famous online application in social media that enables its users to send and receive text-based posts of up to 280 characters, known as "tweets". The service was launched on July 2006. It quickly earned worldwide popularity, with over 180 million active users as of 2014 generating over 390 million tweets daily and managing over 1.6 billion inquires per day. Twitter REST API[3][14][45][46] follows a RESTful API design, following standard HTTP process to extract and handle Twitter resources. Many API calls demand that the user of the application is allowed permission to access their data. Twitter uses the OAuth 2.0 protocol to permit authorized applications to access user data. Twitter REST API uses JSON, XML, RSS, or ATOM data formats to depict the resources, as well as supports pagination. The API offers HTTP GET and POST processes to handle the resources (read, create, update, destroy). The authors of the documentation of Twitter REST API have grouped the methods provided by the API into 22 main categories:

Table 1. Twitter Rest API main categories

Timelines	Direct Messages	Lists	Places & Geo
Tweet	Friends & Followers	Accounts	Trends
Search	Users	Notification	Block
Streaming	Saved Searches	Favourites	Spam
Reporting	OAuth	Help	Legal
Deprecated	Suggested Users		

Moreover, Twitter provides free client libraries for various programming languages including Python, PHP, Ruby, Javascript and Java [47][48][49].

The general workflow of our tool is shown in following lines:

- Social Flow Information
- Cluster Flow Information into events
- Analyze Information to find the event location
- Visualize event location

3.3 Geo-Location

The Geo-Location approximation difficulty has been studied completely by researchers who recommend many solutions to extract user coordinates from Internet social media platforms. These social media platforms include web pages and blogs [28][49][50] etc. All works based on external resources such as gazetteers and databases, to recognize the associated geographical information. In our approach, we do not use any external resource to find the coordinates of the user. Also, the work from Jurgens and Ghahremanlou [46][51] studied the variation of language usage on Twitter. This can also be used to increase our work to ameliorate the accuracy of predicting user geographic location.

There have been also researches on: relations between geotags, geo-location assesment in search engine query logs [50], user privacy of geotags, anticepating geographic

location on proximity [46], and a study of private information trials using correlations between different publicly available pieces of information to extract private information about a person. An additional work contains location prediction of Twitter users based on his/her social network [52]. The most relevant associated work is the content-based approach recommended by Liu et al. [53] to estimate the geo-location of a Twitter user. In our work, we use the Rest API and anonymity on Internet through the process that already described.

3.4. Google maps JavaScript API

The Google Maps API [54] offers the usableness of Google Maps with fast and direct way. It is a web service that is provided by Google and offers street maps and navigation services to websites or mobile applications.

The services offered are:

- Create and display the map.
- Import markers (pinpoints), polygons, popups, polylines (Figure 8), Info Window.
- Event handlers.
- Geocoder: Coordinates / address translation service.
- Direction: Route and package design. route navigation (by car, by public transport or on foot).
- Recognizing business in countries around the world.

The Google Maps API is free for commercial use, and can be accessible to the public without charging the user for each access, and do not produce more than 25,000 accesses map per day [54]. There are also premium packages for a fee covering increased needs in applications and use. For the case of SocialMap python script, the specifications of the free package is more than enough.

The query in the Google Maps API starts by sending an HTTP GET request to the Web application and returns the results in XML or JSON format. The application uses JSON messages as we have mentioned. To use the Google Maps API from an application is required to obtain a "key" [55] (Google Maps API Key) from the creator of the application that uses it and the introduction of the application code.
The capabilities of Google Maps JavaScript API used for SocialMap script are:

- Pinpoints (Markers).
- InfoWindow using activation of the marker touch event (On Click).
- Polylines.
- Changing the map center.

4 Methodology

The first and the main goal of our work was to keep our anonymity on each step anytime. Starting our methodology, we use a point where Internet access cannot connect back to our identity (e.g. home or relatives home, our work etc.). We are located in the popular international chain of coffee in the city center, which offers free wireless Internet access. We will use our laptop computer by booting from USBlive distribution Tails Linux. This distribution offers anonymity and privacy. It uses the Tor network to route all the traffic on the Internet, in order to ensure anonymity. The full name is Tor: The Onion Router (Tor) [38]. The word onion (= onion) indicates multiple "layers" used during the operation. The objectives most used are many and not all necessarily ethical or legal. The Tor network is a network with multiple computers online (Figure 2).

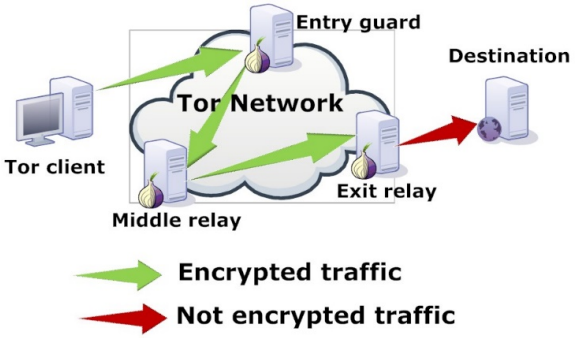

Fig. 2 How TOR network provide anonymity.

An important factor that increases the anonymity of the user connected via Tor is the number of users, because bigger the number, the better anonymity that can be succeeded. Information that exchanged between two computers will transfer encrypted via multiple paths. The connection to a PC is randomly selected each time, while the only available information, is the ending and the starting point in the IP package, providing anonymity. Basically, several relay nodes that are in different points all over the world, are used for scrabbling and encrypting the Internet traffic in multiple layers (onion layers), in such way that at the end of the method can be difficult for someone start searching the information package. In case that a user of Tor tries to access a web site, an encrypted request sent by the browser through the Tor network. The first server that accepts the request is a «keeper» server, which "peel" of the encryption and the request is came across to another randomly selected server. This process run until you reveal all encryption layers together with the last server, the output node, it forwards the user's browser request to the real server hosting the chosen site.

Using TOR in order to have anonymity, we created an encrypted mailbox using tutanota.com service, which doesn't ask the declaration of personal data. Moreover, the anonymity is very strong as the IP addresses of incoming and outgoing email is not recorded in this service. Furthermore, encryption of emails (subject, content, attachment) and contacts provided. The process of encryption and decryption take place only on the local computer and not at the server. The usage of email is for registration on all other Online Services.

We have used a sever cloud with low specifications, in which we have installed a Virtual Private Network (VPN) server on Ubuntu 16.04 LTS 64 bit, increasing the percentage of anonymity. The following figure shows the connection to the Internet.

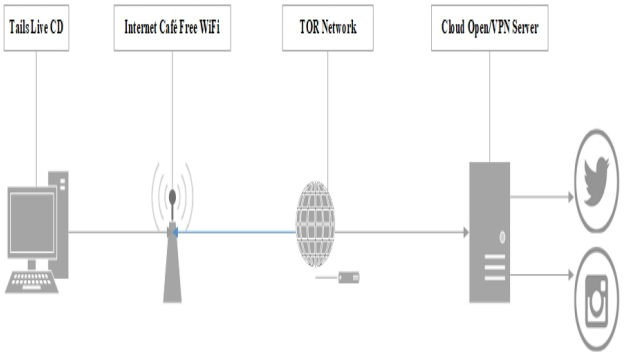

Fig. 3. Connection to the targets using

With a view to enhancing anonymity, we created an encrypted mailbox through tatanota.com service. This service did not record the IP addresses of incoming and outgoing emails, it offered local encryption and decryption of the emails and it did not require the declaration of personal information during the registration process. Furthermore, we used this email account for all the online services which we registered in.

Ensuring the anonymity, without providing our identity during the payment procedure of server cloud was the next goal. We have not used famous electronic payment methods such as Paypal or credit cards because requiring our identification. The transactions can be anonymous by using electronic currency bitcoin and its derivatives [39] . As stated in the website, bitcoin is "a consensual network that enables a new payment system and a completely digital form of money." This is a primary decentralized payment network among peers (peer-to-peer) functioned by users without central authority or intermediaries.

Generally, there are three methods for someone to get bitcoins:

- Mining directly using CPU, GPU, asic mining
- Buy a bitcoin via an ATM with Euros or US Dollars.
- Mining another electronic currency conversion via exchanging service in bitcoin.

The first process is neither profitable nor efficient for someone that has to use the mining project in multiple CPU / cores and graphics cards in the same machine or use multiple computers. At the time was written the survey, as already reported previously, this method was not cost-effective.

The second process is almost new, and uses automated banking machines, using bitcoins and Euros as a medium of exchange, without providing anonymity.

The third process, which also chosen on this survey, is the generation of alternative cryptocoins using CPU and GPU and exchanged them into bitcoins. It should be mentioned that this way it is not economically profitable because the value of current demanded for the finding of cryptocoins is much greater than the value of cryptocoins, just give us the undetectable payment method which will use it to buy online services that we need further in our cause.

The Monero (XMR) digital currency has been chosen to do mining. Besides the product is open source, which offers secure transactions, as each transaction is encrypted. It is very safe because the transactions are not visible to anyone in the global transaction file (blockchain). Exchanging the cryptocurrency Monero to Bitcoin was the next step, as last one was more to online markets.

After obtaining a well-respected amount of bitcoins, we tried to find cloud servers services that could paid in bitcoins instead of other payment methods (Paypal, credit card) in order to preserve our anonymity. The supplier that received the bitcoin payment had the same specification like any other big cloud server providers (amazon, google, digital ocean etc). In our case, we have used the bithost [60], which based on digital ocean infrastructure that is high in the preference of developers. It requires 7 U.S. dollars paid in bitcoins for each month of the cloud server we chose (1 CPU Core, 512MB RAM, 20GB SSD storage, 1TB Data Transfer). On this server we have install Linux Ubuntu 16.04 LTS 64 bit.

The next step, we have subscribed at the API (Application Program Interface) of the networks we will use (Twitter, Instagram, Google) using the email address that we register on tatanota.com service. Moreover, all these APIs need second validation via SMS, and in our case we have used a non-registered prepaid phone number from Cyprus network in order, not get our identity in case they try to find us.

After the registration of the three APIs we have developed a python script that run simultaneously the three APIs in order to gather data from the account and depict the location from the user on the map using the google maps application. This script, install on virtual machine that we have rent on bithost in order to get the data from the social platforms fast and with anonymity. At the following lines gives a detailed description of the methodology.

The system starts using a crawler service that gather feeds according to keywords provided by the user. Those feeds are then used as input to the service that detects, the output returns a list of events that happened during a specific interval, each depicting to a cluster of tweets. The last step, localize the tweet clusters [28].

The task of recognizing locations associated with an event is defying due to rarity of feeds that quite include information regarding the location. Following such a scenario, the feeds depicting physical events usually incline to include spatial landmarks that can be used as unlimited tags in order to define potent locations. Nevertheless, due to the volatilized nature of these feeds the task needs some pre-processing steps:

- Check the keywords regarding location from the tags
- Detects if there is any pattern

before the spatial information can really be extracted. The events are then depicted on a graphical map as shown in Figures 4 and 5 (both depicting areas of mainland Greece), with markers pointing out precise locations. An information pop up box is connected with each event location to provide samples of tweets delineating the event in question.

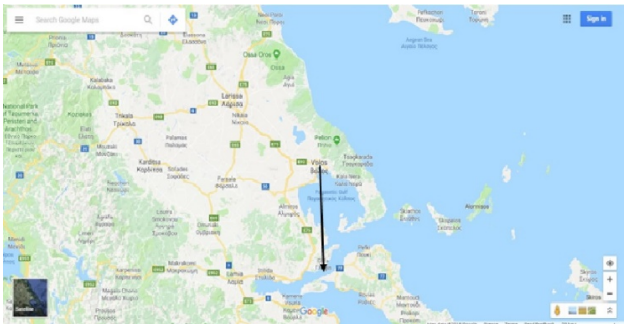

Fig. 4 Example of Google Maps Javascript polyline arrow

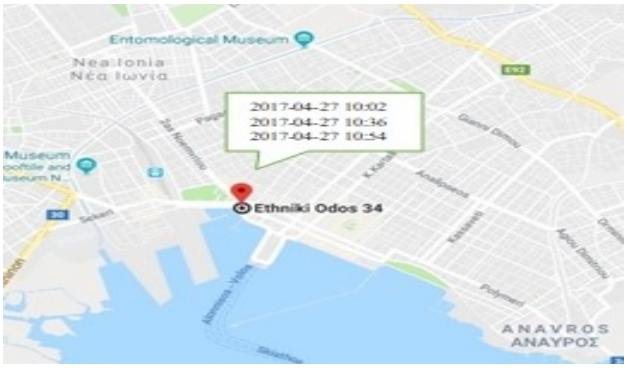

Fig. 5 Collected tracks depicted by Google pinpoint Info balloon

At the two previous Figures (4,5) there is an example of how a malicious users monitor target users in Twitter and Instagram. In our approach, we monitor the location of a Greek celebrity, on Twitter and Instagram using the Rest API and Instagram API respectively. Using this integrated library in our python script, isolate the fields of interest of the social media APIs and store the results then in csv files (Fig. 6).

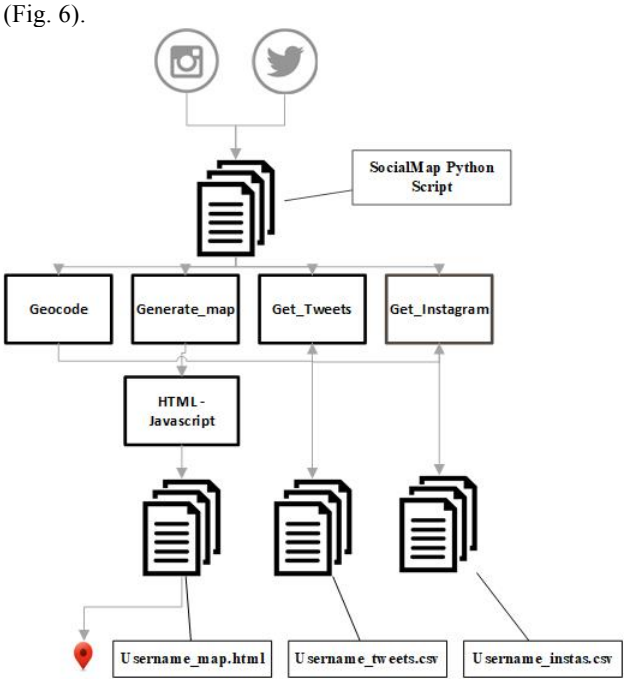

Fig. 6. Diagram of SocialMap python script.

Fig. 7 JSON part of code for Twitter

Fig. 8 Revealing Location using Polyline method

5 .Conclusions and Future Work

In this research, we have presented the possibility of an intruder to get personal data in an automated, undetectable way and use them against an unsuspicious person that uses social media. Specifically, we have monitored (without consent) user activity on two major social networks (Twitter and Instagram), and draw conclusions relating to user locations at each point in time. The results showed that:

- Users of Twitter and Instagram are vulnerable to location disclosure without their consent
- The attacker can remain anonymous during all the monitoring process and cannot be traced back
- A malicious user can predict the future location of the target user using the Twitter and Instagram Rest APIs as well as the approach of Jurgens and Ghahremanlou that use the variation of language usage [46][51].

5.1 Implications for research and practice

Open-source intelligence (OSINT) is gathered public information from available sources. The term "open" means obvious, in public available sources. The OSINT techniques and methods presented in the research demonstrated that anyone could gain access to Twitter, Instagram and Google APIs without disclosing an identity and purpose, and without leaving evidence that can be related back to us easily.

The very nature of social media itself means that we cannot fully protect end users from such malicious actions, as the services require the creation of content by the users themselves, which unfortunately tends to convey the personality of the medium. The use of common sense, the separation and the preservation of sensitive personal information, and the raising of awareness on security issues amongst the user community, can help to significantly reduce the exposure to such attacks.

Morosi et al and their research found that mining the privacy of social media may predict future users' movements as well as attitudes. Moreover, respective surveys from Zheng et. al. [58] have manage with their study, not only to find geographic topics but also allow depiction of users' hidden interests about the location. All these studies together with this one predispose that not only

there is no longer privacy on social media but now it is possible activities that we would like to do to be envisaged.

5.2.Limitations and future research directions.

In terms of limitations, it should be acknowledged that the current work focused only on two social networks, Twitter and Instagram. With respect to that, if the target user does not have an account on these two networks then they cannot be monitored. Another current limitation is the absence of real-time reporting, as the program needs to collect data, analyse them, and then export the results on a map. If the user changes location during this very short time, then it is necessary to repeat the process and collect new tweets in order to relocate them. Improvements of this work, it could be the support of more social networking platforms, such as Facebook, and the production of a graphical web interface / application with user-friendly environment and real-time reporting efficiency. Moreover, there could be integrated operations to provide automated reports per user view, for a given specific time period, and the correlating percentage of exposure. An interesting extension that could be added, is foretelling the future position of the target user based on their history and the frequency of visits to a particular site. Moreover, with the addition of Machine Learning [59], the techniques could potentially be adapted for the benefit of companies that would like to determine the degree of threat of leakage of important information from their own employees (insider threats), or to determine the psychological state/social profile of employees that may be in critical positions, as inferred from their presence in social media. Finally, worth to mention that this tool with the adaptation of Machine Learning and the future improvement stated above could make a significant contribution to law enforcement as well.

References

[1] Koops, B. J., Hoepman, J. H., and Leenes, R. (2013). Open-source intelligence and privacy by design. Computer Law & Security Review, 29(6), 676-688.

[2] Heuer, R. J. (1999). Psychology of intelligence analysis. Lulu. com

[3] Giordano, A., Spezzano, G., Sunarsa, H., and Vinci, A. (2015). Twitter to integrate human and Smart Objects by a Web of Things architecture. In Computer Supported Cooperative Work in Design (CSCWD), 2015 IEEE 19th International Conference on (pp. 355-361). IEEE

[4] Cho, H. R., and Choi, M. (2014). Replacing Socket Communication by REST Open API for Acquisition Tax Analyzer Development. In Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on (pp. 462-468). IEEE.

[5] Puzis, R., Yagil, D., Elovici, Y., & Braha, D. (2009). Collaborative attack on Internet users' anonymity. Internet Research, 19(1), 60-77.

[6] Hugl, U. (2011). Reviewing person's value of privacy of online social networking. Internet Research, 21(4), 384-407.

[7] Tan, X., Qin, L., Kim, Y., & Hsu, J. (2012). Impact of privacy concern in social networking web sites. Internet Research, 22(2), 211-233.

[8] Schoen, H., Gayo-Avello, D., Takis Metaxas, P., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. Internet Research, 23(5), 528-543.

[9] Twitter (2017). Availble from https://support.twitter.com/articles/20169886 [Accessed 1 September 2017]

[10] Instagram (2017). Available from https://www.instagram.com/about/legal/privacy/ [Accessed 1 September 2017]

[11] Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Understanding the predictive power of social media. Internet Research, 23(5), 544-559.

[12] McCreadie, R., Soboroff, I., Lin, J., Macdonald, C., Ounis, I., and McCullough, D. (2012). On building a reusable Twitter corpus. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 1113-1114). ACM.

[13] Rusk, D., and Coady, Y. (2014). Location-based analysis of developers and technologies on github. In Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on (pp. 681-685). IEEE.

[14] Cieszko, M., and Legierski, J. (2014). Graph Based Messaging APIs-concept and implementation. In FedCSIS (pp. 925-932).

[15] Giridhar, P., Abdelzaher, T., George, J., and Kaplan, L. (2015). Event localization and visualization in social networks. In Computer Communications Workshops (INFOCOM WKSHPS), 2015 IEEE Conference on (pp. 35-36). IEEE.

[16] Baldini, N., Neri, F., and Pettoni, M. (2007). A multilanguage platform for open source intelligence. Data Mining and Information Engineering, 18-20.

[17] Neri, F., and Priamo, A. (2008). SPYWatch, Overcoming Linguistic Barriers in Information Management. In Intelligence and Security Informatics (pp. 51-60). Springer Berlin Heidelberg.

[18] Lim, B. H., Lu, D., Chen, T., and Kan, M. Y. (2015). # mytweet via Instagram: Exploring User Behaviour across Multiple Social Networks. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 (pp. 113-120). ACM

[19] Perwitasari, A., Akbar, S., & Saptawati, G. P. (2015, November). Software architecture for social media data analytics. In 2015 International Conference on Data and Software Engineering (ICoDSE) (pp. 208-213). IEEE.

[20] Xie, Y., Chen, Z., Cheng, Y., Zhang, K., Agrawal, A., Liao, W. K., & Choudhary, A. (2013). Detecting and tracking disease outbreaks by mining social media data. Dimensions, 17(16), 16-70.

[21] Neri, F., and Geraci, P. (2009). Mining Textual Data to boost Information Access in OSINT. In Information Visualisation, 2009 13th International Conference (pp. 427-432). IEEE.

[22] Pfeiffer, M., Avila, M., Backfried, G., Pfannerer, N., and Riedler, J. (2008). Next Generation Data Fusion Open Source Intelligence (OSINT) System Based on MPEG7. In Technologies for Homeland Security, 2008 IEEE Conference on (pp. 41-46). IEEE.

[23] Vincen, D., Stampouli, D., and Powell, G. (2009). Foundations for system implementation for a centralised intelligence fusion framework for emergency services. In Information Fusion, 2009. FUSION'09. 12th International Conference on (pp. 1401-1408). IEEE.

[24] Badia, A., Ravishankar, J., and Muezzinoglu, T. (2007). Text Extraction of Spatial and Temporal Information. In Intelligence and Security Informatics, 2007 IEEE (pp. 381-381). IEEE.

[25] Dawoud, K., Alhajj, R., and Rokne, J. (2010, August). A global measure for estimating the degree of organization of terrorist networks. In Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on (pp. 421-427). IEEE.

[26] Neri, F., Geraci, P., and Camillo, F. (2010). Monitoring the Web Sentiment, The Italian Prime Minister's Case. In Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on (pp. 432-434). IEEE.

[27] Kotzias, D., Lappas, T., and Gunopulos, D. (2015). Home is where your friends are: Utilizing the social graph to locate twitter users in a city. Information Systems.

[28] Eisenstein, J., O'Connor, B., Smith, N. A., and Xing, E. P. (2010). A latent variable model for geographic lexical variation. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (pp. 1277-1287). Association for Computational Linguistics.

[29] Ahmed, A., Hong, L., and Smola, A. J. (2013). Hierarchical geographical modeling of user locations from social media posts. In Proceedings of the 22nd international conference on World Wide Web (pp. 25-36). International World Wide Web Conferences Steering Committee.

[30] Cheng, Z., Caverlee, J., and Lee, K. (2010). You are where you tweet: a content-based approach to geo-locating twitter users. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 759-768). ACM.

[31] Mahmud, J., Nichols, J., and Drews, C. (2012). Where Is This Tweet From? Inferring Home Locations of Twitter Users. ICWSM, 12, 511-514.

[32] Ren, K., Zhang, S., and Lin, H. (2012). Where are you settling down: Geo-locating Twitter users based on tweets and social networks. In Information Retrieval Technology (pp. 150-161). Springer Berlin Heidelberg.

[33] Baldini, N., Neri, F., and Pettoni, M. (2007). A multilanguage platform for open source intelligence. Data Mining and Information Engineering, 18-20.

[34] Ikawa, Y., Enoki, M., and Tatsubori, M. (2012). Location inference using microblog messages. In Proceedings of the 21st international conference companion on World Wide Web (pp. 687-690). ACM.

[35] Kinsella, S., Murdock, V. and O'Hare, N., 2011, October. I'm eating a sandwich in Glasgow: modeling locations with tweets. In *Proceedings of the 3rd international workshop on Search and mining user-generated contents* (pp. 61-68). ACM.

[36] Cho, E., Myers, S. A., and Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1082-1090). ACM.

[37] Lichman, M., and Smyth, P. (2014). Modeling human location data with mixtures of kernel densities. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 35-44). ACM

[38] Weinberg, Z., Wang, J., Yegneswaran, V., Briesemeister, L., Cheung, S., Wang, F., & Boneh, D. (2012, October). StegoTorus: a camouflage proxy for the Tor anonymity system. In Proceedings of the 2012 ACM conference on Computer and communications security (pp. 109-120). ACM

[39] Bitcoin. (2017) Available from https://bitcoin.org/el/faq#what-is-bitcoin . [Accessed 25 September 2017]

[40] Instagram Blog. (2017). Available from http://blog.instagram.com/post/8756150468/a-real-time-api-for-next-generation-apps , [Accessed 1 October 2017]

[41] Ferrara, E., Interdonato, R., and Tagarelli, A. (2014). Online popularity and topical interests through the lens of instagram. In Proceedings of the 25th ACM conference on Hypertext and social media (pp. 24-34). ACM.

[42] Highfield, T., and Leaver, T. (2014). A methodology for mapping Instagram hashtags. First Monday, 20(1).

[43] Jaiswal, A., Peng, W., and Sun, T. (2013). Predicting time-sensitive user locations from social media. In Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on (pp. 870-877). IEEE

[44] Reuter, C., & Scholl, S. (2014). Technical Limitations for Designing Applications for Social Media. In Mensch & Computer Workshopband (pp. 131-139).

[45] Huberman, B. A., Romero, D. M., and Wu, F. (2008). Social networks that matter: Twitter under the microscope. Available at SSRN 1313405.

[46] Ghahremanlou, L., Sherchan, W. and Thom, J.A., 2014. Geotagging twitter messages in crisis management. *The Computer Journal*, *58*(9), pp.1937-1954.

[47] Shankar, P., Huang, Y. W., Castro, P., Nath, B., and Iftode, L. (2012). Crowds replace experts: Building better location-based services using mobile social network interactions. In Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on (pp. 20-29). IEEE.

[48] Bhat, F., Oussalah, M., Challis, K., and Schnier, T. (2011). A software system for data mining with twitter. In 2011 IEEE 10th International Conference on Cybernetic Intelligent Systems (CIS).

[49] Vis, F. (2013). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. First Monday, 18(10).

[50] Romsaiyud, W. (2013). Detecting emergency events and geo-location awareness from twitter streams. In The International Conference on E-Technologies and Business on the Web (EBW2013) (pp. 22-27). The Society of Digital Information and Wireless Communication.

[51] Jurgens, D., Finethy, T., McCorriston, J., Xu, Y. T., and Ruths, D. (2015). Geolocation prediction in Twitter using social networks: a critical analysis and review of current practice. In Proceedings of the 9th International AAAI Conference on Weblogs and Social Media (ICWSM).

[52] Rahimi, A., Cohn, T., and Baldwin, T. Twitter User Geolocation Using a Unified Text and Network Prediction Model. Volume 2: Short Papers, 630.

[53] Liu, C. (Ed.). (2014). Principle and application progress in location-based services. Springer.

[54] Bruns, A., and Stieglitz, S. (2013). Towards more systematic Twitter analysis: Metrics for tweeting activities. International Journal of Social Research Methodology, 16(2), 91-108.

[55] Google Maps.. Available from https://developers.google.com/maps/documentation/javascript/ [Accessed 28 August 2017]

[56] Daume, S. (2016). Mining Twitter to monitor invasive alien species—An analytical framework and sample information topologies. Ecological Informatics, 31, 70-82.

[57] Google Geocoding API. Available from https://developers.google.com/maps/documentation/geocoding/intro [Accessed 28 August 2017]

[58] Zheng J, Liu S. and M. Ni L., (2014). User characterization from geographic topic analysis in online social media, In Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on, Beijing, 2014, (pp. 464-471).

[59] Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014, June). Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. InACL (1) (pp. 1555-1565).

[60] Bithost (2017) Available from https://bithost.io/prices/. [Accessed 25 September 2017]