International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Cybercrime Analytics on Social Media

Swati Kumari[1], Sanjay S Pawar[2], Zia Saquib[3]

[1, 2]UMIT, SNDT Women's University, Mumbai, India

[3]CDAC, Gulmohar Cross Road no.9, Juhu, Mumbai, India

Abstract: ***Online Social Networking (OSN) sites, such as Facebook and Twitter, process large amount of data on regular basis. By carefully studying the underlying patterns of these data sets, we can find out the disturbances that are relevant for cyber-criminal investigations. By using the Facebook API and Twitter API, extraction of data is performed based on the word related to the disturbance. The raw data collection set will be stored in huge volume, containing many from Facebook, like posts and information such as counts of like, shares and comments, etc., and from twitter, like tweets/status, retweets' count, follower count, friends count, etc. When the user enters a search text, the relevant post will be fetched by the machine learning based algorithms. Main goal of the project is to analyse such information which may be useful to support cybercrime investigations.***

Keywords: Facebook, Twitter, OSN, OSN API

1. Introduction

Cybercrime are crimes that involves a computer and a network [5]. It has many types like scams, cyber terrorism, cyber extortion, cyber warfare, etc. The frequently discussed typeis cyber terrorism, which in general, can be defined as an act of terrorism committed through the use of cyberspace for computer and resources.

Social Media is an open platform where people are inter-connected and everyone can share their views directly without any intermediaries in the flow of communication. Popular online social networks are like Facebook, twitter, YouTube, etc, which are easily accessible, free/affordable, broad, and a convenient way to spread a message/thought about everything, thus making it a popularly used medium to broadcast the message and aims by any terror organization. As their tool are cheap, accessible, broad dissemination of message and allows unfiltered communication with everyone.

Social Media is changing the pace of how people receive news about terrorist attacks and the reactions. The first information to the public about an incident is more likely to come through social media platforms such as twitter rather than through traditional news outlet. Social media activity increases significantly during events like sports, natural calamities, etc. Terrorists use the social media for disinformation, fund raising, recruitment, communication and networking, especially, whenever an event takes place. Increase in terrorism is more through social media.

Primary objective of this research is to identify the cyber terrorism through online social networking sites, especially through Facebook and twitter. These two online social networking sites are popular among wide range of people across the globe. The terrorist organizations use it to directly spread their audio-visual messages via 'posts' and 'tweets' on Facebook and Twitter respectively. Their objective is to spread the malicious entities like rumours, protest, violence, hoaxes, malware attacks, phishing attacks, chaos, luring victims into scams and misinformation. So, in this research, the malicious entities are collected from Facebook & Twitter along with various parameters and stored to a database. Various challenges are faced during the process of collection of the data.

FACEBOOK

Recently in Jan 2017, the Facebook community update showed that 1.86 billion users are actively using it. It is used by people to post the status, videos, pictures, audios and share the same to their connections which may be friends, peoples, community, pages, groups and organization.

Anyone can search for their friends by either 'name' or 'email address', any pages, community and organization like celebrities, political parties, government pages/group and terror organizations pages/group. Any user 'follows', 'likes' the pages & community to get the notification updates and may 'comment' to give any sentiment emotion/view towards any post.

Apart from keeping in touch with friends and family and resorting to Facebook for daily news and updates about what is going on around the world, the newly introduced features of hashtag support and graph search for posts has increased the level of visibility of public content, either directly or indirectly.

TWITTER

In Oct 2016, Twitter had 317 million monthly active users. The authorized registered user can post a 'Tweet' consisting maximum 140 characters, 'retweet' the tweet of others, user can follow/un-follow anyone and may tweet to anyone. Even if the user is not registered, then they can read the posted tweets only.

This public nature of Twitter, and reach has made it first choice of almost everyone including law and order agencies, journalists and celebrities.

On Oct 21, 2015, Twitter began to roll out the ability to attach "# Poll Q" to tweets. Poll was open for up-to 7 days & voters were not personally identified.

A Twitter user gets quick latest updates regarding anything happening around the world. This makes the terrorist organizations to use Twitter social media to interact with the people and to know the count of the number of people it

reached, the sentiment response of the people. They use it to attract the audience towards them to manipulate their views, spread rumours, etc.

The terrorists group account is there in Twitter and having many followers. They continuously use it to tweet about attacks plan, credit of an attack, recruit members for their group, and gather knowledge of the action taken against them by the law and real time conversation with audience as it's the fastest means to get real time updates of any major activity in the world.

Being a member, you can post the tweet and receive messages of a network of contacts. You can build your own network of contacts and invite others to receive your tweet, follow other members' posts.

You can follow accounts of anyone from your interested fields like sports, politics, music, celebrities, or everyday moments of your favourite person, so that you can see all their tweets. You can search and see the whole conversation involving the person in the result of that search. Twitter provide a hash-tag (#) feature which helps to connect a tweet to the tweets that talk about the same thing. It provides option to add a geo-location or geo-tag, to let others know the location from where the tweet is posted.

2. Literature Review

Our examination will be focussed on how to extract the data from the OSN site like Facebook and twitter through their API; how this API fetch and response data in format to store the data in database&how a programmable and non-programmable developer can extract the data.

2.1 Online Social Networking API

Our project focuses on very famous online social networking sites which are very popular for real time update of news and many more. A social networking service is a platform to build social networks or social relations among people who, for example, share interests, activities, backgrounds or real-life connections and a variety of additional services. But they are also used by terrorist to spread their message, rumour and disturbance. Almost all OSN including Facebook and twitter, have an API for user to interact service programmatically and exchange data. Some Free and not free Software tools are available which uses these API as a third party for exchanging the data without any programming knowledge.

An Application Programming Interface (API) is a particular set of rules and specifications that a software program can follow to access and make use of the services and resources provided by another particular software program that implements that API. It serves as an interface between different software programs and facilitates their interaction, similar to the way the user interface facilitates interaction between humans and computers [6]. Whenever a developer or tool requests information from an API, they need to call (i.e., more technically speaking, create a request to) that API. Many open APIs have strict limits on how many times people can make a call in order to limit traffic and not overwhelm the API with requests [7].

Facebook API is Facebook graph API andtwitters' API are REST API and Streaming API.

2.2 Facebook Graph API

The Graph API is the primary way to get data in and out of Facebook's platform. It's a low-level HTTP-based API that you can use to query data, post new stories, manage ads, upload photos and a variety of other tasks that an app might need to do. The Graph API is HTTP based, so it works with any language that has an HTTP library, such as cURL, rule [2].

Information on Facebook composed of nodes, edges and fields. "Things" such as a User, a Photo, a Page, a Comment are represented as node and connection between them is called as edges. Information about those "things" is represented as fields.

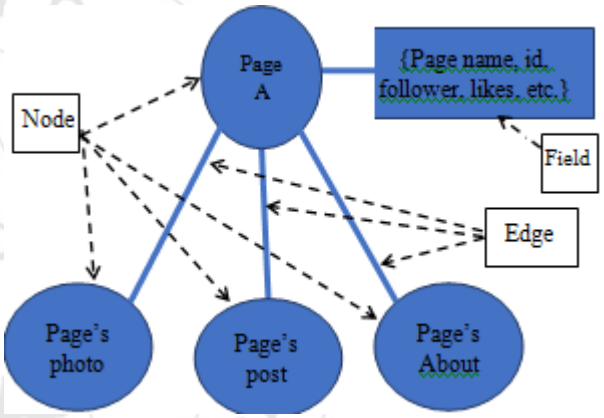

Figure 2: Example of nodes

Example – Page A, pages' photo, posts, about are the nodes and these nodes are connected by edges and information about page A like page name, id, followers, like, etc are the fields.

In general, you can read APIs by making HTTP GET requests to nodes or edges on those nodes. Almost all requests are passed to the API at graph.facebook.com. Each node has a unique ID which is used to access it via the Graph API. ID to make a request for a node[2]:

GET graph.facebook.com
/{node-id}
or edge:
GET graph.facebook.com
/{node-id}/{edge-name}
You can generally publish to APIs by making HTTP POST requests with parameter to the node:
POST graph.facebook.com
/{node-id}
or edge:
POST graph.facebook.com
/{node-id}/{edge-name}

Most Graph API requests require the use of access tokens which your app can generate by implementing Facebook

Login. We can get one quickly through the Graph API Explorer. Click on the Get Token button of the Explorer then Choose the option Get User Access Token after that just click the blue Get Access Token button then You'll see a FacebookLogin Dialog, click **OK" here to proceed.

Now press the "Submit" button. It'll take a few seconds to process, but you should now see a whole bunch of information appear in the response panel. What appears here for you is determined by the privacy settings of your profile, but there should at least be some basic fields:

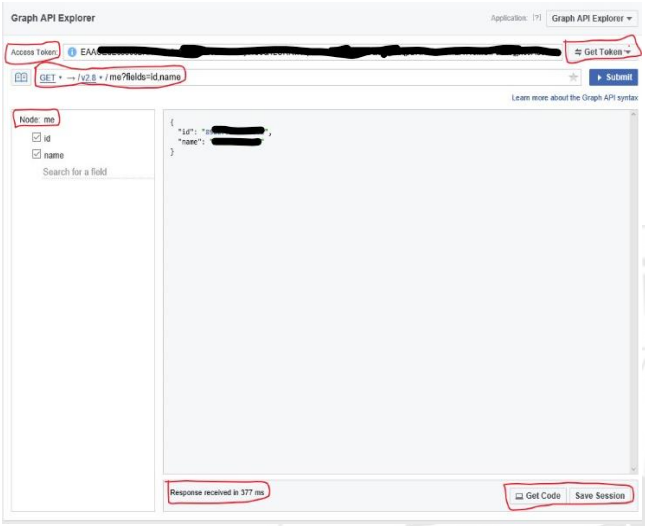

Figure 3: Graph API Explorer

Access Token is not a permanent token, it has a time limit and that information you can see by clicking on that "i" symbol with access token. Once it expires it will ask you to refresh it. When the user creates its new App, it gives the "App Id" and "Secret Key". Here the "App Id" is a permanent Token which never expires. This Token and key should not be shared with others to avoid misuse. This Access Token is required by programmer and non-programmer both while extracting data for authentication.According to your data requirement you can select a permission and version also, you can see in fig-3. Here the Selection is divided into three parts –(a) User Data Permissions, (b) Events, Groups and Pages, (c) Others. Whatever the permission is selected by user only that data will be only extracted. Facebook privacy is highly secured stricted, everyone maintains their Facebook account privacy, so only limited and publicly allowed data will be extracted.Currently the version v2.8 is latest one, every version has some additional updated and something deprecated. In v2.8 we can fetch the post reactions which will show in type – LIKE, ANGRY, HAHA,LOVE. Through the reactions, we will do sentimental analysis.

If the user is familiar with programming the many option are available to extract the data using SDKs. But if the user is non-programmable then using free or paid software tool are available online to extract and analyse analysis the data. Format of extracted data will be JSON, but programmer can also save in .csv form also.

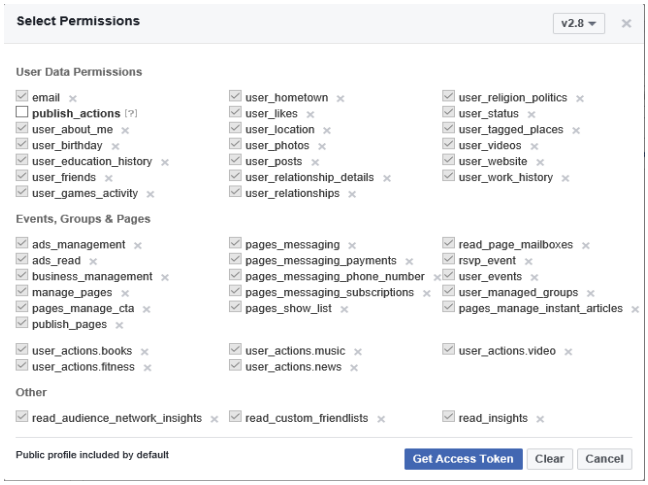

Figure 4: Select Permission

2.3 Twitter API

There are several APIs, public Twitter API consists of a REST API and a Streaming API. The Streaming API provides low-latency high-volume access to Tweets. Twitter APIs use the JSON data format for responses.To make authorized calls to twitter's APIs, your application must first obtain an OAuth access token on behalf of a Twitter user[3]. The way you will obtain such tokens will depend on your use case.

REST APIs provide programmatic access to read and write Twitter data. Create a new Tweet, read user profile and follower data, and more. The REST API identifies Twitter applications and users using OAuth; responses are in JSON format. If you want to monitor or process Tweets in real-time then use Streaming API.

According to our project requirement, we are using application-only authentication, rate limits are determined globally for the entire application. If a method allows for 15 requests per rate limit window, then it allows you to make 15 requests per window — on behalf of your application. This limit is considered completely separate from per-user limits. Rate limits are divided into 15 minute intervals. All endpoints require authentication, so there is no concept of unauthenticated calls and rate limits. There are two initial buckets available for GET requests: 15 calls every 15 minutes and 180 calls every 15 minutes. Features of Authentication-only authenticate is secure of password and standard libraries and code.

The Twitter Search API is part of Twitter's REST API. It allows queries against the indices of recent or popular Tweets. It searches against a sampling of recent Tweets published in the past 7 days.

To get a twitter API key login to dev account with twitter login information then to go "My App" and click "Create New App". Fill all the details; agree the terms & condition and click "Create your Twitter application". In next page, click "Key and Access Token", you will get "API Key (Consumer key)" & "API Secret (Consumer Secret)" and click "Create my access token", you will get "Access token" & "Access token secret". All key and token user will need while fetching the data and it should be kept safe from

misuse of others. In Figure 4, you can the information how its look.

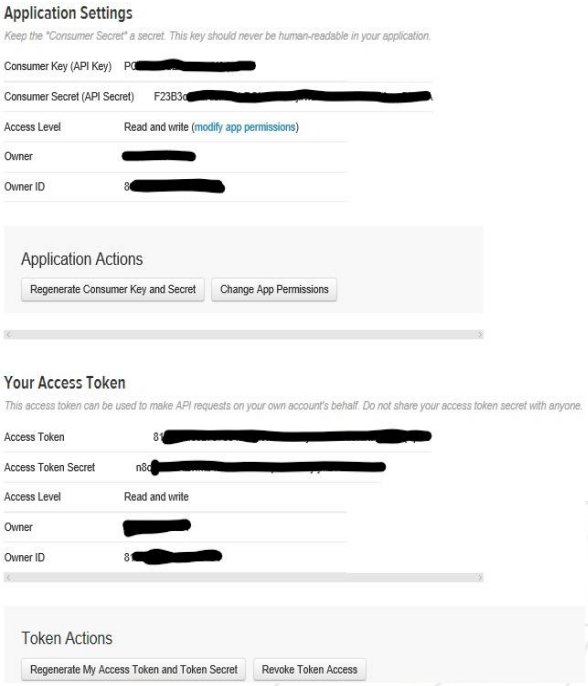

Figure 4: Twitter API Key and Secret and Access Token

3. Methodology

The data required and needs to be extracted from Facebook & twitter is based on the effects of the events caused by terrorism/protests to spread fear and chaos everywhere or to brain-drain people to join/support the hideous act.The extraction of data related to rumours, violence, riots, terror attacks, bomb blasts, etc., by using the Software Development Kit (SDK) orby free/paid online software tools is a simple process as the data will be extracted from present date to the available old data.

The major hurdle while extracting the data from Facebook was due to the user's privacy settings of Facebook, i.e. if the user has configured the privacy settings in such a way that none can access its information, then the tools won't be able to extract all the information of that user. Another hurdle is that the policy of Facebook strictly prohibits posting of disturbing posts, removes it itself, making it unavailable to extract using the tool. In comparison to Twitter, Facebook does not provide much information about the user such as friend list and check-ins. It only provides with the count of friends in a user's account, but displays the User ID and name of those friends in the list who are registered with developers Facebook account.Twitter provideshashtag (#) support and is commonly used by users, Facebook users also use hashtag but they can't fetch data based on hashtag. A survey shows that twitter is having wider range of tweets related to Terrorism/chaos when compared to Facebook.

Extraction of data by using an official SDK provided by Facebook for various platforms like iOS, JavaScript, Android, PHP, etc required programming knowledge. For Non-programmers, user will have to use software tool for the same, which might be a free or paid service.

Facebook and Twitter, both responses the data in JSON format, which makes the data to be read and understand easily. The software tool also allows storing the data in Excel format, .csv format and .db format for Facebook. For Twitter, the data can be extracted with Twitter API by using Python, Tweepy, R Lang, NodeXL, .java, etc. Before initialization of extraction of data, the user has to register in the Facebook and Twitter developer site. In Facebook, the user has to get access token after login, which expires and requires refreshing to get new access token every time. When you create your own Application, an Application ID, API version, Application Secret Key will be generated which won't expire. In Twitter, after login in the developer site, the user will get the access token, access token secret, consumer key and consumer secret. This information is always required to extract the data using SDK or software tool. We have used both, SDK and software tool, i.e. Facepager, to extract the data.

4. Data Collection

Facepager was designed for fetching public available data from Facebook, Twitter and otherJSON-based API[1]. All data is stored in a local SQLite database and may be exported to csv. Even you can copy JSON data to clipboard and save it in .json form. Very common to process to Facepager once you start with any Facebook or twitter data extraction is, first you have to login in your online social networking site through the Facepager and it will automatically fetch the access token and Secret key information. After that you have create new database in your machine and save it with .db file name. Now you can start adding column filed name then add nodes and the select parameter and start fetching data. That .db file you can open in SQLite.

4.1 FACEBOOK

With the Facebook module, you can get data via the so called Graph API. You can access public available data or any data you can see with your account. On Facebook, every object has at least one ID. This identifier is shown in the column "Object ID" in the main window and is the starting point for all queries to the Facebook-API. You need to add some IDs as top nodes to get any data.

Some examples for Object IDs are:

- Me - Refer to yourself
- ISRO – Refer to the page https://www.facebook.com/ISRO/. As you can see identifiers are included as the last part of URLs.
- 1448364408720250 - Also refers to the page https://www.facebook.com/ISRO/.

Every object has such a numerical id. When fetching data for ISRO you will find this numerical id in the detail data.

Some useful parameters are limit, since and until. Request is limited so specify the number of individual object that should return on each page. You can fetch data after the date mention, example date in the form since=" YYYY-MM-DD"; 'Until' is used to grab data before the date.

Normally a page has much information like feeds, photos, posts, videos, notes, links, groups, statuses, etc. In my project the Facebook page information is extracts post message, created time, updated time, reactions, shares counts, likes count of comments, message tags, etc.

4.2 Twitter

The concepts of fetching data from Twitter are very like the Facebook module described above. Starting nodes for fetching data from Twitter usually are search terms, user_id or user_names or screen_name. For example, if you want to get the most recent tweets of someone, you would add his or her username or screen_name as a node and then do search user_timeline with parameter user_id and screen_name. Once you fetched some tweets, look at the detail data. For searching hashtag (starting with #) and users(starting with @), you have to do search tweets and with parameter "q". Every tweet has a unique numerical id, which subsequently may serve as starting point or otherwise be used in the parameters.

Additional parameter you can add to filter the tweets like count and max_id, until, since_id, result_type, etc. With max_id parameter you can then restrict the results to contain only tweets posted before a given ID (plus the tweet with the given ID). In analogy to max_id you may restrict the result by specifying a since_id. Result_type can be popular, recent and mixed. Count can be maximum 100. Until fetch the data before YYY-MM-DD date of recent 7 days tweets.

In this project, the information extracting for twitters tweets are like text, created at, user screen_name, favourite count, retweet count, entities hashtag text, entities users mention name, entities display url, etc. Twitter allow limited data to extracts, at one time 3,000 tweets only fetched and after that it will through error of rate limit exceed.

5. Limitations of Facebook

In this section, we look at the various limitations and challenges posed by Facebook, which possibly makes extraction and analysing data from this network a difficult task. Facebook social network site is highly privacy maintaining site which doesn't allow access to the details of pages and users, if there is privacy maintained by the account holder. Once some data is deleted by user or removed by Facebook, it is not available to extract. We can't fetch the users friends; Facebook Graph API only give count of friends and name, user id of those friends who are registered with the Facebook developer site.

Figure 5: Fetching Information of Friends

There are four levels privacy setting at Facebook for visibility of information are - Only me, Friends, Friends of friends, Public. Users can also select who can send them friendship requests, and who can search them.

Such tremendous control over contents by the privacy settings configuration is a serious implication from a research standpoint. When it comes to identifying and analyzing malicious content, it is hard to find effective solution using only a part of content.

Privacy rules applied on users' profile and network information makes it harder to analyze the sources of malicious content in Facebook. Apart from gender, name, and username, all other profile information about a user is not available publicly, unless explicitly specified by the user. Vital pieces of information like a user's work, education, description, location, account creation time, birth date, etc. are virtually impossible to extract from Facebook. This implies that even if a user is identified as malicious, it is hard to analyze and extract features, which can be used to differentiate a benign user from a malicious one.

6. Conclusion

Today, anything that happens in the real world is talked about on online social media. From sports to storms, terrorist attacks, bomb blasts, earthquakes, and even elections, users share thoughts and information about literally everything using online social media services. Unfortunately, when this platform is used in wrong way by terror organizations, it spreads fear and chaos among people. The users of OSN are provided freedom of speech and their interpretation & reaction to such harmful events/discussions plays important role in the influence of such acts. In this survey, we have talked about the data collection related to disturbance like rumours, terror attack, bomb blast, terrorism, etc. The extracted data will have all the posts, comments and other such information in huge volume, which will be used in the project to identify the searched word or sentence and creating a link between the users, their connections with other users and their area of interests. From this information, we can sort out the people in support to such negative messages. All this can be achieved by doing analysis of data to avoid such disturbances in future which can prevent harm to OSN users and maintaining the security.

References

[1] https://htmlpreview.github.io/?https://github.com/strohne/Facepager/blob/master/src/help/help.html#_Toc405676327
[2] https://developers.facebook.com/docs/graph-api
[3] https://dev.twitter.com/docs
[4] https://developers.facebook.com/tools-and-support/
[5] https://en.wikipedia.org/wiki/Cybercrime
[6] Arun, K., and M. Gomathy Nayagam. "Building Applications with Social Networking API's." International Journal of Advanced Networking and Applications 5.5 (2014): 2070.
[7] http://pravidhiasia.com/what-is-an-api/