Finding Malicious Websites Using Twitter Streams

Muhammad Ashhad Sheikh

ashhadsheikh@hotmail.com

Abstract

With the increase in trend of social media there's also an increase in fraudsters. So it's important that we monitor such frauds and blacklist them before they produce any potential harm to the people. I developed a program that searches for all the malicious links appearing in the public twitter stream.

Result Summary: I ran my program for two consecutive days searching for tweets containing a URL and at least one of the keywords muscle, weight, diet, acai, cambogia, lose fast, and miracle pill, and during this period my program downloaded more than 65,000 tweets. For the 50 most tweeted URLs I ran my program to collect different *host-based* features of the URL and then tagged each feature to further use them in classification. My classification functions assigned a specific fitness value to each tagged feature and then the URL with the fitness score exceeding the certain threshold was classified as malicious. After the experiment I manually checked each URL as well as ran different online tests to check the authenticity of URL and found that 61% of the URLs that were being marked as malicious by my program were accurate.

1 Introduction

According to FTC survey report for both online and offline fraud published in 2011 it was being observed that almost 10.8% of American adults have faced the fraud. Among them around 5.1 million were the victims of fraud weight-loss products [2].

With the ever increasing use of internet most of the frauds in today's world are being carried out online since for fraudsters it's much easier to grab a larger population online. According to FBI Complaint cell they have received around 0.3 million complaints for Internet crime in 2014 but the true incidents rate is even higher and is estimated to be three times this figure. [3]

In general the fraudster's main aim is to gain as much attention as he can so that people fall prey to his fraud. So if they are spreading their fraud through online medium their main aim will be to redirect as much traffic as they can to their website. The more hits to the website increases the chances for anyone to have caught into this network of fraud. So in today's world with everyone having a handheld device and it's been observed that the teens spend around 9 hours daily on social media [8]. With this much people using social media with this intensity its best for anyone to use social media for spreading their fraud.

Similarly there are now techniques available for law enforcement agencies to catch these online criminal activities and to prevent people from falling prey to them. So we can proactively discover and monitor these frauds which will help protection groups to take down these websites in smaller timescale which will ultimately help consumer to be saved from fraud.

2 Background

Among different social networking sites available, for study purposes twitter is ranked best among researchers as it provides a medium to study through its API that allows anyone to write custom programs to read the public twitter stream. Anyone through its automated program can fetch twitter stream with different kind of filters applied to it.

One Limitation to twitter API is that it not sends complete stream or the total tweets that are being tweeted but it only send 1% of the total tweets being posted at any given time [4].

However if the script left to be running for a longer duration, this one percent can be enough for performing certain kind of results and experiments on the stream.

3 Methods

I have developed a program [9] which searches for different keywords in the twitter streams. The terms I chose to filter the stream are based on FTC report [2] and the previous research [1] which may be used in fraudulent messages and tweets. Following are the terms I used "muscle, weight, diet, acai, cambogia, lose fast, and miracle pill". As the previous research [1] concludes that only looking for some words in tweets doesn't actually means that the tweet contains fraud so there's a need of some advance technique to look for such frauds in the stream.

Whenever a new tweet was being posted with any word in the above list my program was capturing that tweet and storing that in database with the twitter handle of the person tweeting that and the timestamp at which the tweet was being posted. This way I had gathered more than 65,000 tweets in my database.

The second and most important part of the experiment was to classify the tweets based on certain technique that whether the tweet contains the URL pointing to malicious website or not. So I chose *Host-Based* methods [5] and ran different tests on the URL to detect whether the URL is pointing towards the benign website or the malicious Website.

According to the research [5] one can accurately detect that whether the website is malicious or not by its URL since, there are multiple tests that can be run on the URL to detect its authenticity. I used different host based models to classify whether a URL is redirecting towards an authentic web or the website contains malicious contents.

I used the following hostname properties to classify the website

3.1 IP Address Properties

These are the properties associated with IP Address and there are different tests through which I passed the IP address of the host. i.e. I have gathered around 0.25 Million blacklist URLs from different online sources[6] and saved all those URLs locally in my database, now each time a link was being received from the stream. It was first being passed through the function which determines that whether this link is blacklisted or not.

The second URL property which I examined was that whether A, MX and NS records of a given IP are in the same autonomous system or not, because malicious websites may have them placed in different autonomous systems.

3.2 WHOIS Properties

WHOIS protocol which is documented in RFC 3912 is a query and response protocol which is widely used to query internet resources, and a lot of information can be fetched through this protocol like Registrars information, Created and updated date of domain. As the previous research [7] states that malicious URLs tries to be registered for the minimum duration of time and they change their URLs very frequently. So a lifetime of a domain is very important in this aspect, that most of the trusted URLs uses the domain that are there for years but on the other hand the malicious URLs are created for a very short span and they change their domain names very frequently. I tested this property using the WHOIS protocol and determined the created and expiry date of URL which made me calculate the lifetime of a URL, and the shorter the lifetime the more are the chances that the URL is pointing towards a malicious website.

3.3 Geographic Properties

Through geographic properties one can determine the region, continent, country or even the exact coordinates about where the website is being hosted. So with these properties I queried the location of a server and then checked in the list of countries where most of malicious URLs are being hosted [7]. If the URL is pointing to that country it doesn't necessarily means that the website is malicious but it certainly raises the level of to which one can believe that the URL is malicious.

3.4 Domain Name Properties

For a given domain name there's a A-record for that name which is the IP address of that domain name. While the A-record maps domain name to IP there's a reverse DNS lookup too which resolves the IP address to a hostname. So I have ran test to check FCrDNS (Forward-confirmed reverse DNS) in which I had checked that both the forward and reverse DNS maps to a same IP address. Since malicious URLs may try phishing and through which capturing user credentials.

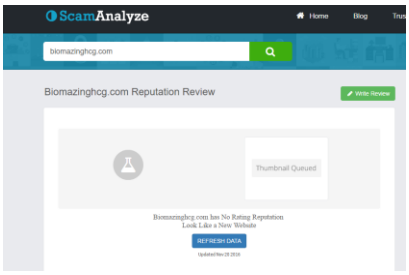
4 Results

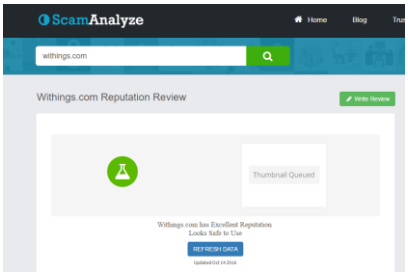
As explained in the above sections I ran script to capture 65,000 tweets and in those tweets I first sorted them and collected the top 50 URLs that were being tweeted mostly. On those top 50 URLs I ran my feature collector program which captured the features for each and every URL. And then tagging was done on the features being collected in the form of 0 or 1, 1 being malicious and 0 being benign.

URL	isBlacklisted	isFullCircle	isASNSame	isCountryMalicious	Lifeline
http://giodet.com/s17hdf-ever_best_americas_...	0	0	0	0	1
http://mecklet.com/s1n23k-weighing_celebrities...	0	0	0	0	1
http://mecklet.com/s17hqd-feeds_dog_a_tumbl...	0	0	0	0	1
http://trendytopic.info/funfacts/here-is-how-m...	0	1	0	0	1
http://www.mymeizitang.com/	0	0	0	0	1
http://www.lidaweightloss.com/	0	1	0	0	1
http://en.voltzro.info/health/christmas-detox-di...	0	1	0	1	1
http://3weekweightlossdiet.wordpress.com	0	0	0	0	1
http://llir.biz/gz?eob	0	0	1	0	1
http://tiewa.ciao.jp/diet/	0	0	0	0	1
http://BiomazingHCG.com	0	1	0	0	0
http://yourweightlosspartner.com/do-diet-pills-...	0	0	1	0	1
http://peachisoda.blogspot.com/	0	1	0	0	1

After the complete features were tagged a fitness function was required to decide whether the URL contains the malicious content or not. So I made a fitness function in which I assigned a different weight to each feature being gathered, the fitness function was making decisions based on the certain threshold which I set after going through the previous research that was being conducted previously [5].

After the URLs were tagged by the system I used an online tool [10] to manually open each URL and check the correctness of my solution that how accurate it's tagging the malicious URLs.





Among those 50 top URLs that were gathered through the tweets, 61% of the URLs were marked correctly as malicious when manually checked and also verified through the online tool.

5 Discussion

My experiment was a follow up to the paper [1] that described how simple key words search in tweets can help identify potential fraud offers, however, I have followed a slightly more technical approach towards the problem solution with using host-based features of an IP address, A main feature of IP address that I haven't added in my experiment is the Lexical analysis which is also used to classify the malicious URLs. The following research [11] conducted at The University of Alabama states how we can identify malicious URLs using lexical features. A future experiment may include these features too which will increase the accuracy to which we detect malicious content in twitter streams.

Another important aspect is addition of an AI powered model to the system through which the system will itself learn continuously and will predict the URLs more accurately by the time. For example the features tagged in this experiment can be acted as genes in *Genetic Algorithm* to make the system Artificially Intelligent.

A future study can also include the improvements to the fitness functions through which the malicious URLs can be accurately detected.

References

1. Rothchild D. Finding Fraudulent Websites Using Twitter Streams. Technology Science. 2015092905. September 29, 2015. http://techscience.org/a/2015092905
2. Federal Trade Commission. Consumer Fraud in the United States, 2011. The Third FTC Survey. April 2013. https://www.ftc.gov/sites/default/files/documents/reports/consumer-fraud-united-states-2011-third-ftc-survey/130419fraudsurvey_0.pdf
3. Federal Bureau of Investigation. 2014 Internet Crime Report. Inter- net Crime Complaint Center. Accessed September 15, 2015. https://www.fbi.gov/news/news blog/2014-ic3-annual- report/
4. Twitter Developers. Is There a Limit to the Amount of Data the Streaming API Will Send Out? Twitter, March 30, 2012. https://twittercommunity.com/t/is-there-a-limit-to-the-amount-of-data-the-streaming-api-will-send-out/8482
5. Justin Ma, Lawrence K. Saul, Stefan Savage, Geoffrey M. Voelker. Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs https://cseweb.ucsd.edu/~voelker/pubs/mal-url-kdd09.pdf
6. squidGuard – Blacklists http://www.squidguard.org/blacklists.html
7. Detecting Malicious URLs- LAVASOFT
8. http://edition.cnn.com/2015/11/03/health/teens-tweens-media-screen-use-report/
9. https://github.com/ashhadsheikh/TwitterFraud
10. http://scamanalyze.com/
11. Aaron Blum, Brad Wardman, Thamar Solorio, Gary Warner. Lexical Feature Based Phishing URL Detection, Using Online Learning http://mmnet.iis.sinica.edu.tw/botnet/file/20110512/20110512_2.pdf