ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

Detecting Fake Social Network Profiles based on Image-Watermarking

Roshne Vasanthamohan[1], N.P.Revathi[2], Dr.V.Anjana Devi[3]

B.E. Student, Department of Computer Science and Engineering, St.Joseph's College of Engineering, Chennai, Tamil Nadu, India[1]

B.E. Student, Department of Computer Science and Engineering. St.Joseph's College of Engineering, Chennai, Tamil Nadu, India[2]

Associate Professor, Department of Computer Science and Engineering. St.Joseph's College of Engineering, Chennai, Tamil Nadu, India[3]

ABSTRACT: On-line Social Networks (OSNs) are increasingly influencing the way people communicate with each other and share personal, professional and political information. Well known sites such as Facebook, LinkedIn, Twitter, and Google+ have millions of users across the globe. With the wide popularity there are lot of security and privacy threats to the users of Online Social Networks (OSN) such as breach of privacy, viral marketing, structural attacks, malware attacks and Profile Cloning. Social Networks have permitted people have their own virtual identities which they use to interact with other online users. It is also completely possible and not uncommon for a user to have more than one online profile or even a completely different anonymous online identity. Sometimes it is needed to unmask the anonymity of certain profiles, or to identify two difference profiles as belonging to the same user. Entity Resolution (ER) is the task of matching two different online profiles potentially from social networks. Solving ER has a identification of fake profiles. Our solution compares profiles based similar attributes. The system was tasked with matching two profiles that were in a pool of extremely similar profiles.

KEYWORDS: image-watermarking; Steganography; spatial domain; least significant bit algorithm; cloning

I. INTRODUCTION

Online social networks, such as facebook and twitter, have become one of the main media to stay in touch with the rest of the world. Celebrities use them to communicate with their fan base, corporations take advantage of them to promote their brands and have a direct connection to their customers, while news agencies leverage social networks to distribute breaking news. Regular users make pervasive use of social networks too, to stay in touch with their friends or colleagues and share content that they find interesting.

Over time, social network users build trust relationships with the accounts they follow. This trust can develop for a variety of reasons. For example, the user might know the owner of the trusted account in person or the account might be operated by an entity commonly considered as trustworthy, such as a popular news agency. Unfortunately, should the control over an account fall into the hands of a cyber criminal, he can easily exploit this trust to further his own malicious agenda. Social networks have permitted people have their own virtual identities which they use to interact with other online users. Social networks such as facebook, twitter and google+ have attracted millions of users.

One of the most widely used social networks, facebook, recently had an initial public offering, which was among the biggest in internet technology. These social networks allow real world people to create online profiles based on the information they give. The profiles are online identities that are capable of being totally independent of their real life identity. The interaction between these profiles happens through direct communication with other users, publishing posts and pictures, expressing opinions on other people's content, etc. Each profile can be seen as a node on a graph and the friendship relations between profiles are the vertices, hence the term social network. Such profiles are created

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

during the registration process. Since the registration process for the average social network requires the user to manually enter their information it is very easy and not an uncommon occurrence to create a profile with fake or erroneous information. It could be to the interest of multiple parties to acquire the public information of these profiles from different social networks to correlate and match data in order to identify a single entity with different profiles. This process of matching profiles into a single entity representing one real world entity is known as entity resolution. Er also has real world uses such as the construction of a more detailed source of information on people, searching for people across different social networks, employers being able to know their employee candidates more before hiring them, improving marketing strategies, detecting fake profiles, etc.

We present an alternative form of comparing profiles that takes advantage of other information that is available, without using training phase. To solve ER, we went farther than just comparing image based features between profiles; we also compared other types of information if it was publically available. Image based features such as the profile's images and posted images were compared with string comparison methods that obtain best results.

II. RELATED WORK

The popularity of social networks inspired many scientific studies in both, networking and security. Early detection systems for malicious activity on social networks focused on identifying fake accounts and spam messages [8], [9], [10] by leveraging features that are geared towards recognizing characteristics of spam accounts (e.g., the presence of urls in messages or message similarity in user posts).

Cai et al. [25] proposed a system that detects fake profiles on social networks by examining densely interconnected groups of profiles. These techniques work reasonably well, and both twitter and face- book rely on similar heuristics to detect fake accounts [26], [27]. In response to defense efforts by social network providers, the focus of the attackers has shifted, and a majority of the accounts carrying out malicious activities were not created for this purpose, but started as legitimate accounts that were compromised [12], [2]. Since these accounts do not show a consistent behavior, previous systems will fail to recognize them as malicious.

Grier et al. [2] studied the behavior of compromised accounts on twitter by entering the credentials of an account they controlled on a phishing campaign site. This approach does not scale as it requires identifying and joining each new phishing campaign. Also, this approach is limited to phishing campaigns.

Gao et al. [12] developed a clustering approach to detect spam wall posts on facebook. They also attempted to determine whether an account that sent a spam post was compromised. To this end, the authors look at the wall post history of spam accounts. However, the classification is very simple. When an account received a benign wall post from one of their connections (friends), they automatically considered that account as being legitimate but compromised. The problem with this technique is that previous work showed that spam victims occasionally send messages to these spam accounts [10].

Warningbird [13] is a system that detects spam links posted on twitter by analyzing the characteristics of http redirection chains that lead to a final spam page.

Xu et al. [28] present a system that, by monitoring a small number of nodes, detects worms propagating on social networks. This paper does not directly address the problem of compromised accounts, but could detect large-scale infections such as *koobface* [29].

Yang et al. [30] studied new twitter spammers that act in a stealthy way to avoid detection. In their system, they use advanced features such as the topology of the network that surrounds the spammer. They do not try to distinguish compromised from spam accounts.thomas et al. [14] built monarch to detect malicious messages on social networks based on urls that link to malicious sites. By relying only on urls, monarch misses other types of malicious messages. For example, our previous work [15] illustrates that compa detects scams based on phone numbers and XSS worms spreading without linking to a malicious URL.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

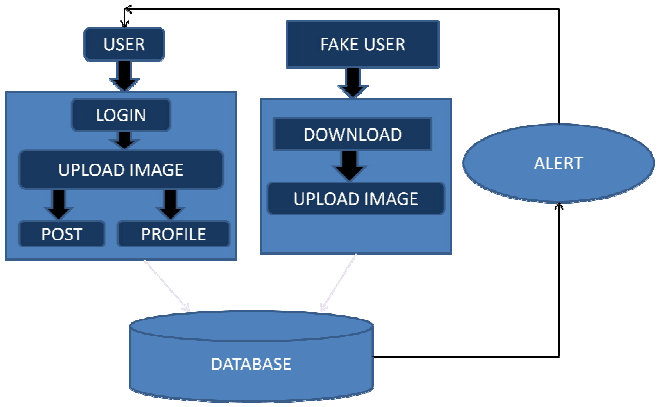
Vol. 6, Issue 3, March 2018

III. PROPOSED SYSTEM

We propose a technique using steganography in which we add an id to the profile and posted pictures which the id will be an email id of the user which is added to the image while uploading. The images downloaded from fake profile users and uploaded it when the notification alert sends to the original users. If the original profile user gives the permission when the picture was uploaded otherwise it was blocked.

A.SYSTEM MODEL

The following illustration describes how the proposed method works. It consists of the login module, data hiding module, profile matching module and alerting user module. The functions of both the user and the fake user are depicted clearly.



B.LOGIN-USER INTERFACE

The Login Form module presents site visitors with a form with username and password fields. If the user enters a valid username/password combination they will be granted access to additional resources on your website. Which additional resources they will have access to can be configured separately. Once logged in, the Login Form module presents the user with a Logout button. Logged in users who are inactive for a predetermined period of time will be automatically logged out. The Login Form module will appear in whatever module position it is assigned to in the current template. It is also possible to have a Login Form that will appear in place of regular content when a Menu Item is clicked.

C.DATA HIDING

In this module, it consists of a new stenographic algorithm for hiding data in images. Here we have also used a Steganography algorithm. Steganography is the practice of hiding secret message within any media. Most data hiding systems take advantage of human perceptual weaknesses. Steganography is often confused with cryptography because the two are similar in the way that they both are used to protect secret information. Here we have tested few images with different sizes of data to be hidden and concluded that the resulting steno images do not have any noticeable changes.In this module,the concern user who uploads the image will have an id that will be hidden within the image.Once another user who downloads the image cannot see the image as it is hidden.We have also used water mark techniques that will not be visible even for the users. Steganography technique finds its main application in the field of secret communication.The main advantage of this algorithm is to keep the size of the cover image constant while the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

secret message increased in size. It can be used by intelligence agencies across the world .Hence this new stenographic approach is robust and very efficient for hiding data in images.

D.PROFILE MATCHING

In this module, If the user who uploads the entire image can be viewed by another user. Another user can downloaded the image but they cannot upload the same image this can be checked by the hidden id. The profile will be checked if the third party who upload the same image, this will be checked by the database. If the profile matches with another profile, another user cannot upload the same it consists of a new stenographic algorithm for hiding data in images. Another user can, Use the Image or else can upload the Image internal entry criteria matching system that checks for a primary match based on hard-coded, Already some data inside is there are not check. This profile matching module will check if another user who uploads the image which is in exists with another user. There by this can avoids the fake user.

E.ALERTING USER

If the profiles match, then the concern user will be alerted by the alert message.The user will be notified as their profile image has be tried to upload by the another user and the user can block the person or else allow its user wish. User will also be notified with the fake users name, mail id, uploaded image, uploading time and system MAC Address. criteria match fails, no further weighing point match is attempted and the profile is either created newly or rejected based on parameter settings for this interface ID in fake profile.So finally give a some Alert Message to the original User.

EVALUATION METRIC

In order to evaluate the performance of the watermarked images, there are some quality measures such as SNR, PSNR, MSE, and BER.

The MSE (mean square error) is defined as average squared difference between a reference image and a distorted image. It is calculated by the formula given below

$$MSE = 1/XY\left[\sum_{i=1}^{\hat{}}\sum_{j=1}^{'}(c(i,j)-e(i,j))\text{^}2\right]$$

X and Y are height and width respectively of the image. The c (i, j) is the pixel value of the cover image and e (i, j) is the pixel value of the embed image. [18]

SNR (Signal to Noise ratio) measures the sensitivity of the imaging. It measures the signal strength relative to the background noise. It is calculated by the formula given below, [23]

$$SNR_{dB} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right)$$

The PSNR (peak signal to noise ratio) is used to determine the degradation in the embedded image with respect to the host image. It is calculated by the formula as

PSNR = 10 log10 (L*L/MSE)

L is the peak signal value of the cover image which is equal to 255 for 8 bit images. [18]

The BER (bit error ratio) is the ratio that describes how many bits received in error over the number of the total bits received. It is calculated by comparing bit values of embed and cover image.

BER = P/ (H*W)

H and W are height and width of the watermarked image. P is the count number initialized to zero and it increments by one if there is any bit difference between cover and embed image.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

IV.PSEUDO CODE

SPATIAL DOMAIN LEAST SIGNIFICANT BIT ALGORITHM

Step1-Open Image

This step will open the file and save header in a file and save the palette value of body in another file.

Step2- Split the body of the image file

This step will split the body image in equal blocks to use these blocks to hide text.

Step3-Convert text watermarking to ASCII code and then convert to Binary code.

Step4- Divide the stream binary code to parts of 24 bits. Every part represents three characters of text watermarking, and compare with pixels in palette of image.

V.RESULTS

The following database results show the records of all profiles available and the information regarding the pictures posted by the users. The relative ids of users are also include in the database. After blocking, the status of the action is also updated in the database.

DATABASE RESULTS

The pictures uploaded by the original users.

	Fileid	Filename	Storedname	Comment
	kumaresan	download.jpg	kumaresan_download.jpg	hai
	kumar	download.jpg	kumar_download.jpg	hai
	kumaresan	bommu_prosss.jpg	kumaresan_bommu_prosss.jpg	hai
	kumar	bommu_prosss.jpg	kumar_bommu_prosss.jpg	hai
	kumar	download.jpg	kumar_download.jpg	hai
*	(NULL)	(NULL)	(NULL)	(NULL)

The profile names along with their passwords, date of birth, gender and mail id that were entered by users.

	username	password	dob	gender	mail
	kumar	12345	2017-07-26	Male	kumaresan@uniqtechnologies.co.in
	kumar	12345	2017-07-27	Male	kumaresan458@gmail.com
	Roshne	harrypotter	2018-01-05	Female	roshnemohan@gmail.com
	Roshne	harrypotter	1996-11-09	Female	roshnemohan@gmail.com
	Roshne	harrypotter	1996-11-09	Female	roshnemohan@gmail.com
	Roshne	harrypotter	1996-11-09	Female	roshnemohan@gmail.com
	Roshne	harrypotter	1996-11-09	Female	roshnemohan@gmail.com
	Roshne	harrypotter	1996-11-09	Female	roshnemohan@gmail.com
	Roshne	harrypotter	1996-11-09	Female	roshnemohan@gmail.com
*	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

The status of uploading by the fake users is updated according to the original users' actions.

phone	SecurityQuestion	Answer	Macaddress	Permit	otp	id
8883638181	What is your petâ??s name?	123	40-8D-5C-F5-86-2A	Allow	38081757	a3VDvsxJ
8883638181	What is your petâ??s name?	zzz	40-8D-5C-F5-86-2A	Allow	93219771	ysA163yK
9791419565	What is your petâ??s name?	Bitto	28-50-33-2D-56-15-00-00-00-00	Allow	36010273	3ZqoZ8BF
9791419565	What is your petâ??s name?	Bitto	28-50-33-2D-56-15-00-00-00-00	Allow	36010273	nt3gc8m7
9791419565	What is your petâ??s name?	Bitto	28-50-33-2D-56-15-00-00-00-00	Allow	36010273	77VW2YK1
9791419565	What is your petâ??s name?	Bitto	28-50-33-2D-56-15-00-00-00-00	Allow	36010273	wbgLm3Fr
9791419565	What is your petâ??s name?	Bitto	28-50-33-2D-56-15-00-00-00-00	Allow	36010273	XuDV4omO
9791419565	What is your petâ??s name?	Bitto	28-50-33-2D-56-15-00-00-00-00	Allow	36010273	19A0U6U4
9791419565	What is your petâ??s name?	Bitto	28-50-33-2D-56-15-00-00-00-00	Allow	36010273	8a4k1gg2
(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)

After blocking the fake user from uploading

Fileid	Filename	OldFileid	Comment	Imagetype	Macaddress	status
kumar	download.jpg	kumaresan	hai	Posts	40-8D-5C-F5-86-2A	Block
kumar	bommu_prosss.jpg		hai	Posts	40-8D-5C-F5-86-2A	Block
(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)

VI. CONCLUSION AND FUTURE WORK

We solved Entity Resolution with our system and used it to compare online user profiles from social networks in order to identify matches. Our systems are comparing the two images and identify that fake or not. We are using Steganography Algorithm and that algorithm hides the information inside the image. In this way new images upload in our profile and that image compare to existing user profile. If the image is fake when send notification to original user. The original user allows the uploading notification that images was uploaded otherwise blocked.

- The above mentioned limitations can be solved in the future enhancements of this project.
- This system can be easily extended with additional and more comprehensive similarity measures.
- Removal attacks intend to remove the watermark data from the watermarked object. Such attacks exploit the fact that the watermark is usually an additive noise signal present in the host signal. Thus, techniques to avoid this can be implemented.
- We will recommend a modified SVD dependent watermarking to enhance the results further. And also we will utilize embedding plus to improve the security.

REFERENCES

[1] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and Evaluation of a Real-Time URL Spam Filtering Service," in IEEE Symposium on Security and Privacy, 2011.
[2] W. Xu, F. Zhang, and S. Zhu, "Toward worm detection in online social networks," in Annual Computer Security Applications Conference (ACSAC), 2010.
[3] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: the underground on 140 characters or less," in ACM Conference on Computer and Communications Security (CCS), 2010.
[4] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting Spammers on Twitter," in Conference on Email and Anti-Spam (CEAS), 2010 Computer and Communications Security (CCS), 2010.
[5] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting Spammers on Social Networks," in Annual Computer Security Applications Conference (ACSAC), 2010.
[6] T. Jagatic, N. Johnson, M. Jakobsson, and T. Jagatif, "Social Phishing," Comm. ACM, vol. 50, no. 10, pp. 94–100, 2007.
[7] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots + machine learning," in International ACM SIGIR Conference on Research and Development in Information Retrieval, 2010.
[8] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary, "Towards Online Spam Filtering in Social Networks," in Symposium on Network and Distributed System Security (NDSS), 2012.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

[9] S. Lee and J. Kim, "WarningBird: Detecting Suspicious URLs in Twitter Stream," in Symposium on Network and Distributed System Security (NDSS), 2012.

[10] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "Compa: Detecting compromised accounts on social networks," in Proceedings of the Network and Distributed System Security Symposium, San Diego, CA, February 2013.

[11]H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao, "Detecting and Characterizing Social Spam Campaigns," in *Internet Measure- ment Conference (IMC)*, 2010.

[12]M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "Compa: De- tecting compromised accounts on social networks," in *Proceedings of the Network and Distributed System Security Symposium*, San Diego, CA, February 2013.

[13]"Oauth community site," http://oauth.net.

[14]W. B. Cavnar and J. M. Trenkle, "N-gram-based text categoriza- tion," in *In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, 1994, pp. 161–175.

[15]J. C. Platt, "Fast Training of Support Vector Machines Using Sequential Minimal Optimization," in *Advances in Kernel Methods - Support Vector Learning*, 1998.

[19]G. Stringhini, G. Wang, M. Egele, C. Kruegel, G. Vigna, H. Zheng, and B. Y. Zhao, "Follow the Green: Growth and Dynamics in Twitter Follower Markets," in *ACM SIGCOMM Conference on Internet Measurement*, 2013.

[16]G. Stringhini, M. Egele, C. Kruegel, and G. Vigna, "Poultry Markets: On the Underground Economy of Twitter Followers," in *SIGCOMM Workshop on Online Social Networks*, 2012.

[17]"Weka - data mining open source program," http://www.cs.waikato. ac.nz/ml/weka/.

[18]"Alexa top 500 global sites," http://www.alexa.com/topsites.

[19]"Chipotle faked its twitter hack," http://mashable.com/2013/07/24/ chipotle-faked-twitter-hack/, 2013.

[20]"MTV and BET hack their own twitter accounts," http://mashable. com/2013/07/24/chipotle-faked-twitter-hack/, 2013.

[21]Z. Cai and C. Jermaine, "The Latent Community Model for De- tecting Sybils in Social Networks," in *Symposium on Network and Distributed System Security (NDSS)*, 2012.

[22]C.Ghiossi, "Explaining Facebook's Spam Prevention Systems," http://blog.facebook.com/blog.php?post=403200567130, 2010.

[23]Twitter, "The twitter rules," http://support.twitter.com/entries/ 18311-the-twitter-rules, 2010.

[24]J. Baltazar, J. Costoya, and R. Flores, "KOOBFACE: The Largest Web 2.0 Botnet Explained," 2009.

[25]C. Yang, R. Harkreader, and G. Gu, "Die Free or Live Hard? Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers," in *Symposium on Recent Advances in Intrusion Detec- tion (RAID)*, 2011.