CRIME PREVENTION ON SOCIAL NETWORKS FEATURING LOCATION BASED SERVICES

James Maude Majdi Owda
School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University
Manchester, M1 5GD, UK
Telephone: +44 (0)161 247 1520
Email: james.maude@stu.mmu.ac.uk m.owda@mmu.ac.uk

Abstract

In the age of austerity crime is on the increase. The large online presence of the populous is fueling criminals with large amounts of data capable of turning and individual into a victim. The public awareness of the dangers of social networks is low and online crime analysis is in its infancy. This paper presents a novel system for the prevention of crime on social networks. The system will identify risks within users Geo-location information, status updates and online profile. The system analyses location based information as well as using Information Extraction templates and Natural Language Processing to identify threats. The system can successfully identify threats on a graded scale and provide feedback and advice to the user. The work highlights the importance of closely monitoring a digital footprint.

Keywords: Social Networks Analysis, Geo-location, Geo-tagging, NLP, Information Extraction, Crime Prevention, Computer Forensics, Cyber-crime.

1 Introduction

Social networking websites allow individuals to create a virtual presence in the form of a profile and then map their social interactions from the real world to the virtual world by communicating with other users via profiles. This allows users to "form a social network, which provides a powerful means of sharing, organizing, and finding content and contacts" [1]. Facebook, founded in 2004 has the goal of being "a social utility that helps people communicate more efficiently with their friends, family and coworkers". With over 800million users active (users logging in at least every 30 days) the website has rapidly become something of a modern phenomena [2]. As Facebook has evolved it has gained more and more personal information from users. As well as collecting the information you supply when you sign up for the service smartphone integration now makes geo-location possible. Users can 'Check in' at locations to alert contacts where they are, this information can also be

linked into status updates. Criminals are increasingly targeting users of Social Networks by exploiting the information contained within the networks. The purpose of this paper is to create a novel solution to identify potential victims within a social network and warn them of the dangers. There are currently a number of solutions for monitoring and extracting information from social networks. Most of the solutions however are focused on market research and commercial gain from the network. As such these tools would never promote users disclosing less information. In fact promotions companies often offer users rewards for checking in at a location. The starting point for this paper is 'E-Officer for Crime Prevention on Social Networks' [3]. This system can be used to identify threats contained within some basic profile information and status updates. It effectively uses information extraction templates to identify threats. The limitations of the system are that it is not integrated in the online environment. Also it does not take into account any geo-location information. The research into content-based data management [4] suggests that screening the data before placing it online is a good solution. By leveraging the power of privacy controls with content analysis the system recommends whom to share the status updates with. The researchers claim an "F-measure value of 0.831 in correctly recommending safe recipients". Assumed networks of trust, which may be infiltrated, however limit the system. Commercial extraction systems are capable of capturing and analysing large amounts of data from social networks [5]. However privacy features are increasingly limiting the amount of information available. Also these are not solutions for the use of end users due to licensing fees and complex interfaces. Long-term solutions to the problem involve the education of end users. The e-safety courses offered to children and parents [6] can help raise awareness of online safety. By educating end users you can prevent future dangers and develop a online safety oriented mindset. Crime and social networking threats are discussed in section 2 this is concluded by the proposal of a novel system. Current information extraction techniques are reviewed in section 3. Section 4 covers the prototype architecture and Section 5 explains deals with the analysis of data. Section 6 concludes the paper and section 7 contains references.

2 Crime over Social Networks

Social networks are playing an ever-increasing role in peoples lives, with Facebook alone there are over 800 million active users[7]. Many of these users

are sharing large amounts of their personal information. As a direct result of this more people are falling victim to cyber crimes and identity theft. This increased in cyber crime currently costs the UK economy £27bn per annum[8]. Identity theft is a £1,800M problem with personal information being the key to an identity, the more information you can gain about a person the more you are able to steal their identity. Social networks provide a rich array of personal data and users are often ignorant to the threats.
HMRC defines money laundering as "Money laundering means exchanging money or assets that were obtained criminally for money or other assets that are 'clean'."[9]. Cybercriminals exploit trust within social networks to build a money-laundering network. Criminals are increasingly using social networks to track victims locations. The Association of British Insurers warns "home insurance premiums may rise up to 10% this year, due in part to an increase in home invasions resulting from people revealing their whereabouts on social networks."[10] The proposed novel solution to this is the creation of an online system integrated within the social network environment. The system will be capable of analysing data from profile data, status updates and multiple geolocation services. The system will be able to identify threats and warn users. The testing of possible future status updates will be integrated as well as recommendations and advice.

3 Information Extraction

Information Extraction will form part to the system as it is stated that a rule based approach is the best solution when there are lexicons and rule writers [11]. As the kind of information displayed on social networks follows theses rules this should prove to be the best technique. The Facebook site has a dynamic and complex structure. Automatic information extraction often struggles with this type of input [12]. In order to overcome this issue domain knowledge can be applied to create wrappers for information sources. This technique allows the information to be compartmentalized and dealt with in a specific manner. The template model can be used to classify and to a certain extent extract information. However in most systems the data must be interpreted and processed in some way. This makes for a two-stage system, template and domain dependent process or decision engine [13]. The purpose of the decision engine is to determine what can be done with the gathered information.

4 Prototype Architecture

The novel system architecture that has been created can extract information from several aspects of a users profile. Information extraction templates are used to capture data from status updates. The fixed structured data from the profile is also captured to discover what personal information is stored there. The users location based updates are also processed in order to identify threats based upon location information.

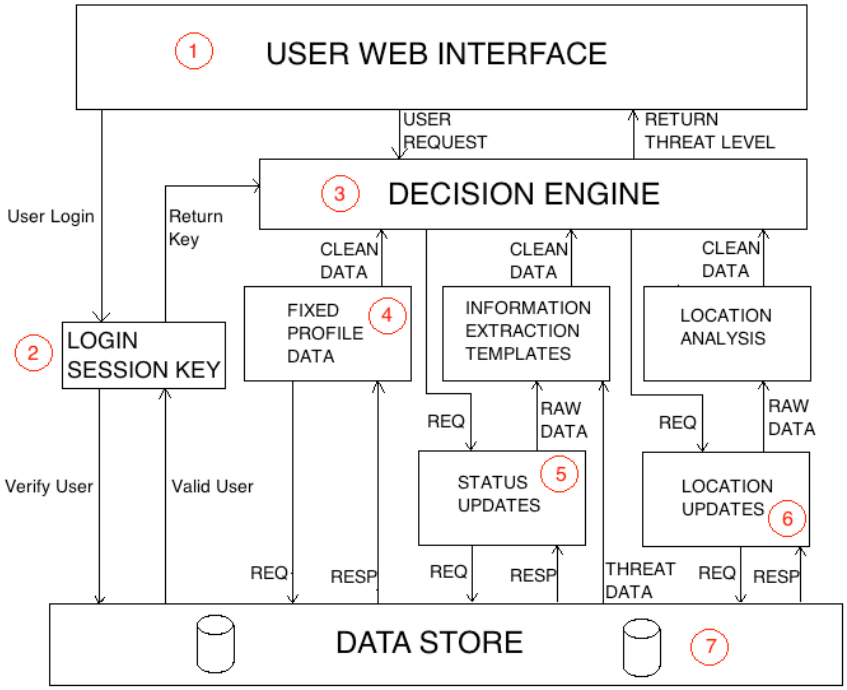

Fig. 1. Prototype System Architecture

The system architecture figure 1 analyses the overall profile, status updates both existing and future and review the users geo-location information. The system also offers advice and allows for recommending the system to friends. *(1) Web Interface* - The web interface has been developed in a combination of PHP, HTML and Java Script with CSS. It is designed to emulate the feel of the Facebook site and present users with a familiar environment. In order to

not overwhelm the user with a large quantity of information at once the system is divided into sections using a Java Script powered tabbed interface. The users log into the system using their Facebook ID this enables them to view their profile information and threats.

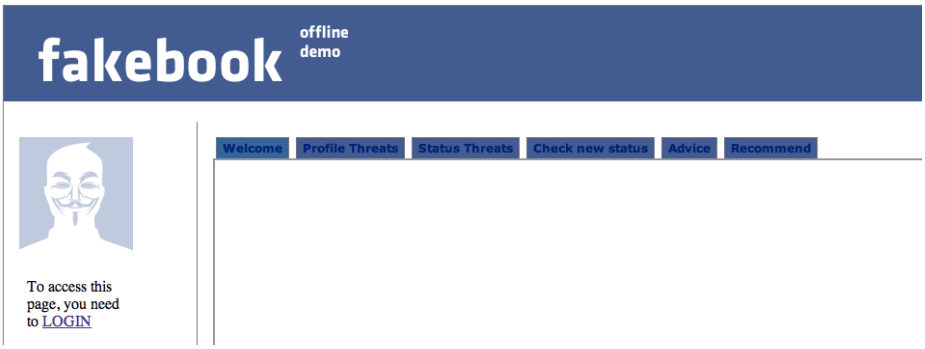

Fig. 2. Web Interface

(2) Login Session Key - This key is generated upon a successful user login and grants access to the users information. In the offline testing system it grants access to the database in the live version it allows the system to request data via the Facebook API. The key is only valid while the user is logged into the system and is destroyed upon logout. *(3) Decision Engine* - The decision engine receives requests from the web interface for data. The data requested depends on the function the user is wishing to perform. The engine can then request a variety of information from the data stores. As the data varies a lot the system uses three separate processing engines to handle the data request. *(4) Fixed Profile Data* - This handles the request for fixed information stored within the profile such as name, date of birth and address. The data is checked to see which elements are present. As users may have different privacy settings and information the data is cleaned and standardized. *(5) Status updates* - The recent updates are collected and processed by the Information Extraction engine. The engine uses templates fed from a database of key threat words. This also ranks the threat level of individual pieces of information. The engine passes an array of possible threats along with the threat levels back to the decision engine. *(6) Location Information* - Location information is gathered from the database this can be in the form of a Geo-tagged status update or a Facebook 'check-in' this information is processed to determine the users most recent locations. The users location is passed to the decision engine and compared to the address information from the fixed profile data. *(7) Data*

Store - The data store in the systems is a MySQL database in the current system the data store contains two main areas. The first is the threat data for the information extraction templates. This stores all the keywords that the system uses to identify threats from statuses. The second part of the system contains the user profiles. In a real world implementation this part of the database would be replaced with developer API calls to the Facebook database.

5 Analysis

In order to test the system the profiles of 30 volunteers were taken for analysis. In order to preserve the test subjects identities the names and certain identifying characteristics have been changed. However no extra data was added or removed and status updates and geo-location information remain true. All volunteers gave explicit consent for their data to be used. The data was manually entered into the database in the exact format presented by the Facebook developer API. This way when the system goes live there are only a few lines of code to change from the offline to online system. The system ranks the threat at three different levels. Low Threat - For the low threat profile the user only shows limited personal information on their profile. Medium Threat - The medium threat profile is where the user is displaying some personal information and may give away some personal or location information through status updates or 'check-ins'. High Threat - The high threat profile is where a user has personal information displayed on their profile to friends. They also give out details of location and plans. The system successfully identifies threats in the profiles tested figure 3. The majority of the threats are medium. This indicates that users are aware of privacy settings but possibly not aware of leaking information in status updates.

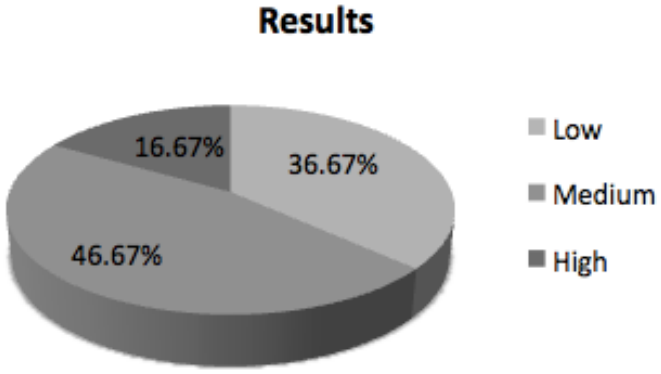

Fig. 3. Profile Analysis Results

The system successfully identifies a High Threat profile in figure 4 and feeds this back to the user in an easy to understand format. The user is clearly shown how a potential attacker can pull together a large amount of personal information including the location of the user.

Fig. 4. Web Interface Identifying High Threat Level

The user in maintains a minimum amount of personal data and does not give away any extra details in status updates. The profile is correctly identified as a low risk level. When viewing past status updates or testing potential new ones the system can display words and phrases that might create a threat. These words are displayed in a word cloud shown in figure 5. The words are sized and ordered in terms of threat level and by clicking on a word the user is presented with an explanation of the possible threat.

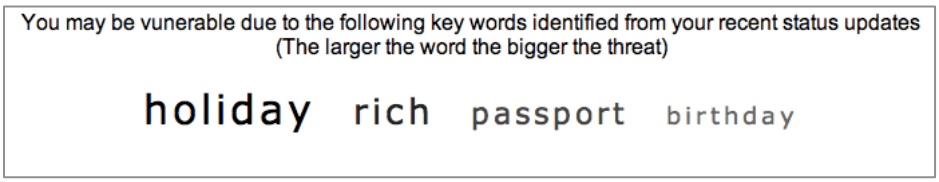

Fig. 5. Word cloud to display possible threats in order of threat ranking

When the volunteers were made aware of the results 70% removed information from their profile. This is a good result as it has not only highlighted past information leaks and dangers but also provided feedback that will modify future behavior. This should lead to an overall safer social networking experience.

6 Conclusions

In conclusion the system can effectively identify threats contained within a users Facebook profile. The system communicates the threats effectively to the end user. This communication is critical, as the system on its own cannot protect the user. The system creates an effective feedback loop as users begin to change their behavior based upon the results of the system. The system greatly raises the awareness of the user to the dangers of social networking. As social networks continue to expand in user numbers and features, technology such as this is becoming ever more critical in the fight against crime.

7 References

[1] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," 2007, p. 29.

[2] Facebook.com, "Factsheet," 2011. [Online]. Available: http://www.facebook.com/press/info.php?factsheet. [Accessed: 03-Nov-2011].

[3] Hussain,Owda, "E-Officer for Crime Prevention on Social Networks." 2011. in Cyber Forensics 2011

[4] M. Jakob, Z. Moler, M. Pechoucek, and R. Vaculin, "Content-Based Privacy Management on the Social Web," in *2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2011, vol. 3, pp. 277-280.

[5] mozenda.com, "Data Extraction, Web Screen Scraping Tool, Mozenda Scraper," 2011. [Online]. Available: http://www.mozenda.com/default. [Accessed: 10-Nov-2011].

[6] roareducate.com, "Roar Educate - Interactive education about cyber-safety, online security and digital citizenship (AU)," 2011. [Online]. Available: http://www.roareducate.com/au/. [Accessed: 10-Nov-2011].

[7] Facebook.com, "Statistics," 10-Oct-2011. [Online]. Available: http://www.facebook.com/press/info.php?statistics. [Accessed: 20-Oct-2011].

[8] Home Office, "The Cost of Cyber Crime Report," Oct. 2011.

[9] 100 P. S. HM Revenue and Customs, "Introduction to Money Laundering Regulations," 17-Aug-2010. [Online]. Available: http://www.hmrc.gov.uk/MLR/getstarted/intro.htm#1. [Accessed: 03-Nov-2011].

[10] ABI, "Association of British Insurers - We know where you are this summer - think twice about disclosing your holiday plans online," 2011. [Online]. Available: http://www.abi.org.uk/Media/Releases/2011/07/We_know_where_you_are_this_summer__think_twice_about_disclosing_your_holiday_plans_online.aspx. [Accessed: 30-Jan-2012].

[11] V. M. Filho, R. B. . Prudencio, F. A. . de Carvalho, L. R. Torres, L. Rodrigues, and M. G. Lima, "Automatic Information Extraction in Semi-structured Official Journals," in *10th Brazilian Symposium on Neural Networks, 2008. SBRN '08*, 2008, pp. 51-56.

[12] Keekyoung Seo, Jaeyoung Yang, and Joongmin Choi, "Building intelligent systems for mining information extractionrules from web pages by using domain knowledge," in *IEEE International Symposium on Industrial Electronics, 2001. Proceedings. ISIE 2001*, 2001, vol. 1, pp. 322-327 vol.1.

[13] M. Bouzeghoub, Z. Kedad, and E. Métais, *Natural language processing and information systems: 5th International Conference on Applications of Natural Language to Information Systems, NLDB 2000, Versailles, France, June 2000 : revised papers*. Springer, 2001.