

A Threat Model Approach to Attacks and Countermeasures in On-line Social Networks

Borja Sanz*, Carlos Laorden*, Gonzalo Alvarez† and Pablo G. Bringas*

*Laboratory for Smartness, Semantics and Security (S³Lab), University of Deusto
Bilbao, Spain

Email: {borja.sanz, claorden, pablo.garcia.bringas}@deusto.es

†Instituto de Física Aplicada, Consejo Superior de Investigaciones Científicas (CSIC)
Madrid, Spain

Email: gonzalo@iec.csic.es

Abstract—On-line Social Networks (OSN) have become the most popular Internet service today. OSN are being embraced by companies and organisations to help connecting people, assist dealing with cooperative tasks, and develop marketing and public relations campaigns. Despite all their benefits and advantages, as happens with every new technology, they are prone to several security issues. In addition to privacy concerns, there are many other dangerous vulnerabilities that affect security. In this paper, we present our Threat Modelling in OSN, which focuses on identifying attacks against users of OSN and possible countermeasures to mitigate the risks.

I. INTRODUCTION

On-line Social Networks (OSN) have become the most visited sites surpassing information gatherers like Google, MSN, or Yahoo!, consuming most of the time that users spend connected to the Internet, both via desktop and mobile devices.

Although there is no accepted and universal definition for OSN, this paper will use the working definition provided by INTECO and the Agencia Española de Protección de Datos [1]:

‘Services that let their users to create a public profile where they can introduce personal data and information. The users have different tools to interact with each other.’

Therefore, for our purposes, the main features of an OSN and their tools are the popular three C’s: i) Communication, allow sharing knowledge; ii) Community, help finding and integrating communities; and iii) Cooperation, provide tools to develop activities together.

Many enterprises are embracing OSN and integrating them within their strategic plans: viral marketing campaigns; collaborative working environments within the enterprise to allow a free knowledge flow in the new paradigm known as *Enterprise Social Networking* (ESN) [2]; image and reputation promotion of enterprises and people within the enterprises; collaborative content creation via wikis, blogging or microblogging; information exchange with faithful and potential clients, partners, or competitors; search for candidates; etc.

Unfortunately, along with the aforementioned personal and corporative benefits come several web-platform-dependant threats. As expected, with the expansion of OSN, both in and out the enterprise, they are becoming the favourite target for cybercriminals. Actually, in 2009, OSN were one of the main

significant channels to identity theft and information leaking [3], [4], [5], [6]. Furthermore, spam sending and malware distribution through OSN are increasing at an incredible pace [7], [8].

The remainder of this paper is organised as follows: Sec. II provides a short introduction to Threat Modelling (TM); Sec. III presents the assets at risk by OSN; Sec. IV details the attacks that are appearing against those assets through OSN; Sec. V discusses some of the countermeasures to be implemented against the previous attacks; finally, Sec. VI concludes and outlines the avenues of future work.

II. THREAT MODELLING

Threat Modelling is a description of a collection of security aspects, a set of plausible attacks which are able to affect the performance of any computer system. This methodology allows security experts to identify security risks, verify an application’s security architecture, and develop countermeasures in the design, coding, and testing phases [9]. Therefore, analysing and modelling the potential threats that an application faces is an important step in the process of designing a secure application [10]. Some of these threats may be related to the application itself, whilst others are related directly or indirectly to the underlying infrastructures, technologies or programming languages, allowing an easier identification and documentation of the corresponding threats.

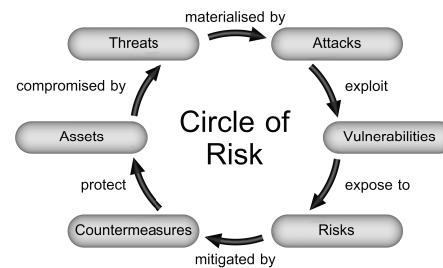


Fig. 1. Threat Modelling’s Circle of Risk.

Being the main objective of threat modelling to provide useful guidelines on how to mitigate the associated risks, we

must be able to distinguish the elements corresponding to what we have called the *Circle of Risk* (CoR) (shown in Fig. 1). The CoR is composed of *assets*, which are compromised by *threats*; threats that exploit *vulnerabilities*, which when misused result in *exposure*, which represents a serious *risk*. Finally, the *countermeasures* mitigate the dangers caused by those risks; countermeasures which have as goal protecting the assets. Definitions for the aforementioned terms can be found within the technical dictionaries [11] and [12].

Although the threat modelling process requires the study in detail of every above-mentioned element, in this paper we introduce a first approach to the CoR, focussing on the assets, attacks, and countermeasures.

III. ASSETS AT RISK BY OSN

Every organisation has at disposal several assets that must be protected to guarantee the proper course of its business. The loss, theft, destruction, reduction, or damage of any of these assets could prevent the organisation from achieving its objectives. Therefore, among the assets specially threatened by OSN we can identify [13]:

- 1) Private information: it can be stolen or used against its legitimate owner in order to harass, extort, or send hyper-contextual advertising.
- 2) Financial assets: they can be stolen through on-line banking fraud, telephone fraud, or lost by decreased productivity.
- 3) Intellectual property: it can be stolen, plagiarised, or illegally distributed free of charge, causing economic losses.
- 4) Corporate secrets: their leakage or theft can cause economic losses, reputation damage, or decreased competitiveness.
- 5) Physical security: it can be compromised by stalkers, harassers, criminals, or thieves.
- 6) Computing and network resources: they can be consumed leading to denial of service or decreased Quality of Service (QoS).
- 7) Corporate and personal reputation: it can be irreversibly damaged.
- 8) Digital identity: it can be spoofed or stolen.

In conclusion, the misuse of OSN affects the aforementioned assets, which are compromised by the attacks described in the next section.

IV. ATTACKS IN OSN

OSN have concocted a dangerous cocktail of user-supplied content, open APIs, and web pages heavily loaded with JavaScript and embedded media of all descriptions. And it is an environment that is largely devoid of security standards and practices [14]. Since attacks are aimed at the aforementioned assets, this work introduces the potential attacks that affect OSN organised in categories corresponding to the objective they are oriented to.

A. Private Information

- **Sensitive data retrieval:** Attackers are able to collect users' personal data due to their negligence when publishing private information [15], [16], [17].
- **Sensitive attribute inference models:** The attributes of users who are connected in social networks are often correlated. Zheleva et al. [3] introduced different attacks to infer the hidden sensitive values:
 - **Friend-Aggregate model (AGG):** AGG looks at the sensitive attribute distribution amongst the friends of the person under question.
 - **Collective classification model (CC):** Unlike more traditional methods, in which each instance is classified independently of the rest, collective classification aims at learning and inferring class labels of linked objects together.
 - **Flat-link model (LINK):** Another approach to dealing with links by 'flattening' the data by considering the adjacency matrix of the graph.
 - **Blockmodelling attack (BLOCK):** The basic idea behind stochastic blockmodelling is that users form natural clusters or blocks, and their interactions can be explained by the blocks they belong to.
 - **Groupmate-link model (CLIQUE):** One can think of groupmates as friends to whom users are implicitly linked. In this model, they assume that each group is a clique of friends, thus creating a friendship link between users who belong to at least one group together.
 - **Group-based classification model (GROUP):** Another approach to dealing with groups is to consider each group as a feature in a classifier, inferring sensitive information according the groups a user belongs to.
 - **BASIC** In the absence of relationship and group information, the only available information is the overall marginal distribution for the sensitive attribute in the public profiles. So, the simplest model is to use this as the basis for predicting the sensitive attributes of the private profiles.
- **Data Mining for demographic information:** Using data mining techniques to retrieve public demographic data [18], one could infer unpublished personal data about other users.
- **Automated User Profiling:** Retrieval of users sensitive data by querying social networks for registered e-mail addresses and crawling every profile found to collect personal information [19].
- **De-anonymise OSN users:** It exploits group membership information that is available on social networking sites, which is often sufficient to uniquely identify users, or, at least, to significantly reduce the set of possible candidates [20].
- **OSN Mash-ups:** Link data between independently provided web services to obtain previously unforeseen infer-

ences including highly personal information [21].

- **OSN Aggregators:** Services that integrate several OSN which multiply vulnerabilities by giving read/write access to several social network accounts using a single weak authentication [21].

B. Financial Assets

- **Cross-Site Scripting (XSS):** A type of computer security vulnerability typically found in web applications that enables malicious attackers to inject client-side script into web pages viewed by other users [22], [21], [23].
- **Cross-Site Request Forgery (CSRF):** Unlike XSS, which exploits the trust a user has for a particular site, CSRF exploits the trust that a site has in a user's browser [22], [23].
- **Bank-customer oriented Malware:** In order to maximise their monetary benefits, malware creators target bank customers credentials [24]. The appearance of these attacks has increased [25], due to the use of social networks as a distribution channel. A recent example is *Koobface*¹, which upon successful infection, gathers sensitive information from the victims such as credit card numbers.

C. Intellectual Property

- **Contents publication property of third parties:** It occurs when a user publishes the contents not being the legitimate holder of the intellectual property rights of such material [1].
- **Search engines indexation of protected contents** entails a greater diffusion and therefore an exponential increased number of reproductions [1].
- **Loss of control over contents when users unsubscribe from the on-line service:** OSN based on profiles eliminate, or at least block, all the contents associated to the profile of the user leaving the service, while in platforms based on contents, where members can get to publish works without being associated directly to their profile, material may remain publicly accessible [1].

D. Corporate Secrets

- **Social Engineering:** Manipulating people into performing actions or divulging confidential information using information found in OSN profiles [26].
- **Spear Phishing:** Spear phishing appears genuine to all the employees or members within a certain company, government agency, organisation, or group, using information found in OSN profiles [27].

E. Physical Security

- **Location Inferring** from recognisable places in the image [28].
- **Location Inferring** from IP connection [29].
- **Facial Recognition:** Sophisticated facial recognition algorithms used to identify unknown users [30].

- **Harassment between Adults** Bullying via electronic communication tools [31], [32].
- **Cyber-bullying** Harassment via electronic communication tools from child to child [31], [32].
- **Cyber-grooming (harassment from adult to child)** Sexual exploitation of children on-line [33].

F. Computing and Network Resources

- **Spam and Hyper-contextualised Advertising:** Spam is becoming a major issue for OSN, and the use of hyper-contextualised advertising (i.e. adapt advertising to users preferences) increases the possibility of the junk messages being read [21].
- **Botnets:** Attacks designed solely to disable infrastructure to those that also target people and organisations.[34].

G. Corporate and Personal Reputation

- **Sybil Attacks:** Given a reputation system, a peer may attempt to falsely raise its reputation by creating fake identities – or sybils – and using them to its benefits [35].
- **Classes of attacks against reputations systems:** Hoffman et al. [36] classify attacks against reputation systems based on the goals of the reputation systems.
 - **Self-promoting:** Attackers manipulate their own reputation by falsely increasing it.
 - **Self-Serving or Whitewashing:** Attackers escape the consequence of abusing the system by using some system vulnerability to repair their reputation. Once they restore their reputation, the attackers can continue the malicious behaviour.
 - **Slandering:** Attackers manipulate the reputation of other nodes by reporting false data to lower their reputation.
 - **Orchestrated:** Attackers orchestrate their efforts and employ several of the above strategies.
 - **Denial of Service (DoS):** Attackers may cause denial of service by either lowering the reputation of victim nodes so they cannot use the system or by preventing the calculation and dissemination of reputation values.

H. Digital Identity

- **Credentials Theft** using technical hacking techniques [37].
- **Profile Cloning** consists of identifying a victim and creating a new account with his real name and photograph inside the same social network [38].
- **Cross-site Profile Cloning** identifies victims who are registered in one social network, but not in another and steals their identities creating accounts for them in the network where they are not registered [38].

Finally, it is important to take into account that the danger level of one attack is directly proportional to how dangerous is the vulnerability that being exploited and inversely proportional to the effectiveness of the deployed countermeasures.

¹<http://news.cnet.com/koobface-virus-hits-facebook/>

V. MAJOR COUNTERMEASURES IN OSN

Countermeasures reduce the vulnerabilities in a system. In this section, we present these countermeasures grouped into the following main categories: platform countermeasures and user countermeasures. The former refers to countermeasures which collaborative platforms must implement in order to prevent attacks directed both to platforms and users, while the latter intends to introduce the best practices to improve users' privacy habits.

A. Countermeasures addressed to the Platform

1) *Technological Security of the Platform*: System administrators of collaborative networks should be aware that their users share personal data through their services. Therefore, they should protect their networks against potential attacks, employing tools especially made to combat against phishing and phishing [39] cases, not to mention one of the most annoying threats of the current times: the spam². Regarding network connections, they should make use of secure connections applying technologies (e.g. *Security Socket Layer (SSL)* [40]), to ensure private data transmissions.

On the other hand, social networks provide users with little control over their personal data. As a consequence, identity theft and fake profiles are common issues. These platforms should provide tools to prevent cases of identity theft, to allow legitimate users to get back the control of the account after the theft, or to strengthen user identification before registration. Additionally, it is recommended to implement technological measures to verify the age of the users, in order to protect children against inappropriate contents or behaviours.

2) *User's Data*: OSN need to facilitate access to the Terms of Service and User Conditions displaying all the information in understandable terms. To this end, these documents must employ a perfectly understandable language by any kind of user. After reading the document, the user should know its rights and obligations during the use of the service.

Besides, OSN must guarantee the users a complete control over their published information. Therefore, a social network must implement several procedures in order to satisfy the following:

- Users should know the intended use by the social network of both personal and published data.
- Users should be able to apply the rights to access, rectify, cancel, and oppose to data concerning them published in the OSN.
- User profile configuration should default to maximum privacy, allowing to later changing it according to personal preferences.
- Users should be able to prevent the publication of unauthorised data. The use of tagging mechanisms requesting user's approval is one of the approaches aimed at the achievement of this goal.

Furthermore, OSN must protect users' data against the indexation of search engines by using appropriate codification.

3) *Author's Royalties*: Author's rights must be protected. OSN must provide users with tools that allow reporting the existence of contents protected by author's rights. Additionally, social networks need adequate staff or automatic tools to check all uploaded contents and establish if such contents are subject to intellectual rights.

Besides, OSN users must know the nature of the rights to authorship and the importance to respect them for the correct use of the service, through general conditions when creating new accounts, FAQs, etc.

4) *User Awareness*: It is essential that OSN encourage their users to know the use that social networks make of their personal data, the advertisement systems present in the platform and the potential threats that users face while using on-line services. Similarly, it is necessary to display information related to the security of the platform, including the measures that users should take in case of abuse of their rights.

B. Countermeasures addressed to the Users

1) *User's Behaviour*: The user must read the Terms of Use and Privacy Policies of the OSN, both before the registering process and every time any change occurs. Once the user has registered, it must configure properly the privacy settings, so that only his friends have access to the published contents.

Users have absolute control over the information that they want to publish. They are therefore responsible for the publication of excessive information putting at risk their intimacy or their whereabouts. In this sense, it is recommended not to publish intimate information in personal profiles that could be seen by everybody. Users must also be careful when publishing audiovisual or graphical contents, trying not to put at risk other users' privacy.

Moreover, friendship relations are the core of these networks. Once defined the privacy settings, users must be careful with friend requests. Users should only accept friend requests coming from people already known and avoid accepting compulsively any request for friendship because it could result in privacy issues.

2) *Technological Concerns*: There are security and technological considerations that users must take into account in order to increase the level of security. First, users should use different user-names and passwords to access different social networks. Second, they should use strong passwords to prevent brute force attacks. Finally, they should use updated security software and operating system.

3) *Special Considerations for Children*: Under-age users are specially vulnerable. Thus, they need extra care to ensure that their personal data is not disclosed. Parents or guardians should be consulted for every sensitive action when using social networks (e.g. content uploading and publishing personal information), being able to abort their children's actions.

Additionally, parents and guardians should take into account several considerations. The computer should be placed in a common area of the house, establishing some rules about the use of Internet. Parental control and content-blocking

²<http://blog.facebook.com/blog.php?post=40218392130>

systems should be installed and effectively working, and, finally, minors should be aware of the dangers that OSN might represent.

VI. CONCLUSION

On-line Social Networks represent one of the last and most important Internet services. Albeit most enterprises hesitate whether to ignore completely the OSN, this new phenomenon can not be ignored, but neither can be integrated into the business model without knowing the risks. In this paper, we presented a first approach to an OSN Threat Modelling that discovers the first elements to take into account when attempting to protect a system. To that end, we identify the assets at risk, the attacks that can compromise them, and we propose some countermeasures to protect against these attacks (the mapping attack-countermeasure is provided in Table I).

The future work of this OSN TM is oriented in three main directions. First, we will complete the aforementioned ‘Circle of Risk’ (see Fig. 1), with the exposures that suffer the assets and the risks that represent them. Second, we plan on developing a taxonomy which organises all the existing OSN threats, attacks, vulnerabilities, and countermeasures. Finally, we will study the feasibility of adding weighted variables to the taxonomy in order to help identifying assets at risk and, hence, supporting the hardening of a system.

ACKNOWLEDGEMENTS

The work described in this paper was supported by the Spanish *Ministerio de Ciencia e Innovación*, project CUCO (MTM2008-02194), and *Ministerio de Industria*, project Cenit SEGUR@, Security and Trust in the Information Society, (BOE 35, 09/02/2007, CDTI).

REFERENCES

- [1] Study on the privacy of personal data and on the security of information in social networks, Tech. rep., INTECO (2009).
- [2] J. DiMicco, D. Millen, W. Geyer, C. Dugan, B. Brownholtz, M. Muller, Motivations for social networking at work, in: Proceedings of the ACM 2008 conference on Computer supported cooperative work, ACM New York, NY, USA, 2008, pp. 711–720.
- [3] E. Zheleva, L. Getoor, To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles, in: Proceedings of the 18th international conference on World wide web, ACM New York, NY, USA, 2009, pp. 531–540.
- [4] B. Krishnamurthy, C. Wills, On the leakage of personally identifiable information via online social networks, in: Proceedings of the 2nd ACM workshop on Online social networks, ACM, 2009, pp. 7–12.
- [5] J. Lindamood, R. Heatherly, M. Kantarcioglu, B. Thuraisingham, Inferring private information using social network data, in: Proceedings of the 18th international conference on World wide web, ACM, 2009, pp. 1145–1146.
- [6] B. Chen, D. Kifer, K. LeFevre, A. Machanavajjhala, Privacy-Preserving Data Publishing, Foundations and Trends in Databases 2 (1-2) (2009) 1–167.
- [7] Z. Mazur, H. Mazur, T. Mendyk-Krajewska, Security of Internet Transactions, Internet-Technical Development and Applications (2009) 243.
- [8] W. Luo, J. Liu, J. Liu, C. Fan, An analysis of security in social networks, Dependable, Autonomic and Secure Computing, IEEE International Symposium on 0 (2009) 648–651.
- [9] F. Swiderski, W. Snyder, Threat modeling, Microsoft Press Redmond, WA, USA, 2004.
- [10] L. Desmet, B. Jacobs, F. Piessens, W. Joosen, Threat modelling for web services based web applications, in: Eighth IFIP TC-6 TC-11 Conference on Communications and Multimedia Security (CMS 2004), Springer, 2004, pp. 161–174.
- [11] U. E. Gattiker, The Information Security Dictionary: Defining The Terms That Define Security For E-business, Internet, Information And Wireless Technology (KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE), Kluwer Academic Publishers, Norwell, MA, USA, 2004.
- [12] R. Slade, Dictionary of Information Security, Syngress Media Inc, 2006.
- [13] G. Alvarez, Amenazas 2.0: los riesgos de las redes sociales online en la empresa, PC World 274 (2010) 64–70.
- [14] S. Mansfield-Devine, Anti-social networking: exploiting the trusting environment of Web 2.0, Network Security 2008 (11) (2008) 4–7.
- [15] B. Ng, A. Kankanhalli, Y. Xu, Studying users’ computer security behavior: A health belief perspective, Decision Support Systems 46 (4) (2009) 815–825.
- [16] J. Bryce, M. Klang, Young people, disclosure of personal information and online privacy: Control, choice and consequences, Information Security Technical Report 14 (3) (2009) 160–166.
- [17] B. Huberman, E. Adar, L. Fine, Valuating privacy, IEEE security & privacy 3 (5) (2005) 22–25.
- [18] D. Jensen, J. Neville, Data mining in social networks, in: Dynamic Social Network Modeling and Analysis: workshop summary and papers, 2003, pp. 287–302.
- [19] M. Balduzzi, C. Platzer, T. Holz, E. Kirda, D. Balzarotti, C. Kruegel, Abusing social networks for automated user profiling, Tech. Rep. EURECOM+3042, Institut Eurecom, France (03 2010).
- [20] G. Wondracek, T. Holz, E. Kirda, S. Antipolis, C. Kruegel, A Practical Attack to De-Anonymize Social Network Users.
- [21] G. Hogben, Security issues and recommendations for online social networks, Position Paper. ENISA, European Network and Information Security Agency.
- [22] G. Alvarez, S. Petrovic, A new taxonomy of web attacks suitable for efficient encoding, Computers and Security 22 (5) (2003) 435–449.
- [23] Web security threat classification v2.0, Tech. rep., Web Application Security Consortium (2010). URL http://projects.webappsec.org/f/WASC-TC-v2_0.pdf
- [24] M. Pemble, Evolutionary trends in bank customer-targeted malware, Network Security 2005 (10) (2005) 4–7.
- [25] P. Gutmann, The commercial malware industry, in: DEFCON conference, 2007.
- [26] M. Workman, Gaining access with social engineering: An empirical study of the threat, Inf. Sys. Sec. 16 (6) (2007) 315–331.
- [27] T. N. Jagatic, N. A. Johnson, M. Jakobsson, F. Menczer, Social phishing, Commun. ACM 50 (10) (2007) 94–100.
- [28] M. Zhang, Content-based Image retrieval, Artificial Intelligence for Maximizing Content Based Image Retrieval (2009) 115.
- [29] Y. Jiang, B. Fang, M. Hu, X. Cui, Techniques for determining the geographic location of IP addresses in ISP topology measurement, Journal of Computer Science and Technology 20 (5) (2005) 689–701.
- [30] P. Phillips, Support vector machines applied to face recognition, Advances in Neural Information Processing Systems (1999) 803–809.
- [31] T. Beran, Q. Li, Cyber-harassment: A study of a new method for an old behavior, Journal of Educational Computing Research 32 (3) (2005) 265–277.
- [32] Q. Li, Cyberbullying in schools: A research of gender differences, School Psychology International 27 (2) (2006) 157.
- [33] D. Roberts, Cyber-Victimisation in Australia: Extent, Impact on Individuals and Responses (2008).
- [34] E. Cooke, F. Jahanian, D. McPherson, The zombie roundup: Understanding, detecting, and disrupting botnets, in: Proceedings of the USENIX SRUTI Workshop, 2005, pp. 39–44.
- [35] A. Cheng, E. Friedman, Sybilproof reputation mechanisms, in: P2PECON ’05: Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems, ACM, New York, NY, USA, 2005, pp. 128–132.
- [36] K. Hoffman, D. Zage, C. Nita-Rotaru, A survey of attack and defense techniques for reputation systems, ACM Comput. Surv. 42 (1) (2009) 1–31.
- [37] J. Erickson, Hacking: the art of exploitation, 2nd edition, No Starch Press, San Francisco, CA, USA, 2008.
- [38] L. Bilge, T. Strufe, D. Balzarotti, E. Kirda, All your contacts are belong to us: automated identity theft attacks on social networks, in: Proceedings of the 18th international conference on World wide web, ACM New York, NY, USA, 2009, pp. 551–560.
- [39] M. Jakobsson, S. Myers, Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft, Wiley-Interscience, 2006.
- [40] E. Rescorla, SSL and TLS: designing and building secure systems, Addison-Wesley, 2001.

