Exploiting Social Networking Sites for Spam

Markus Huber, Martin Mulazzani, Edgar Weippl, Gerhard Kitzler, Sigrun Goluch
SBA Research
Favoritenstrasse 16
AT-1040 Vienna, Austria
{mhuber,mmulazzani,eweippl,gkitzler,sgoluch}@sba-research.org

ABSTRACT

In the ongoing arms race between spammers and the multi-million dollar anti-spam industry, the number of unsolicited e-mail messages (better known as "spam") and phishing has increased heavily in the last decade. In this paper, we show that our *friend-in-the-middle attacks* on social networking sites (SNSs) can be used to harvest social data in an automated fashion. This social data can then be exploited for large-scale attacks such as context-aware spam and social-phishing. We prove the feasibility of our attack exemplarily on Facebook and identify possible consequences based on a mathematical model and simulations. Alarmingly, all major SNSs are vulnerable to our attack as they fail to secure the network layer appropriately.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection; E.1 [**Data Structures**]: Graphs and networks

General Terms

Security, experimentation, theory

Keywords

Social network security, spam, phishing

1. INTRODUCTION

Criminals, as well as direct marketers, continue to clog mailboxes with unsolicited bulk e-mails (e.g., spam and phishing) in the hope of financial gain. So far, their strategy is straightforward, namely to send out a vast numbers of unsolicited e-mails in order to maximize profit on the tiny fraction that falls for their scams. Their pool of target e-mail addresses is normally based upon data harvested with web crawlers or trojans, sometimes even including plain dictionary-based guessing of valid targets. Previous research indicates that social networking sites (SNSs) might change the playing field of spam attacks in the near future. SNSs contain a pool of sensitive information which can be misused for spam messages, namely contact information (email addresses, instant messaging accounts, etc.) and personal information which can be used to improve the believability of spam messages.

Copyright is held by the author/owner(s).
CCS'10, October 4–8, 2010, Chicago, Illinois, USA.
ACM 978-1-4503-0244-9/10/10.

A successful extraction of sensitive information from SNSs would result in spam attacks that are based upon a pool of verified e-mail addresses. Thus messages may have higher conversion rates, increasing the success rate of spam.
Gaining access to the pool of personal information stored in SNSs and impersonating a social network user poses a non-trivial challenge. Gross and Acquisti [6] as well as Jones and Soltren [10] were among the first researchers to raise awareness for information extraction vulnerabilities of SNSs. While their techniques were rather straightforward (automated scripts which retrieve web pages), their results eventually led to security improvements of SNSs. Existing attempts to extract information from SNSs focus on the application layer and can thus be mitigated by adapting a specific social network's application logic. Recent publications devoted to information extraction from SNSs introduced elaborate methods such as the inference of a user's social graph from their public listings [4] or cross-platform profile cloning attacks [3]. The leakage of personal information from these platforms creates a remarkable dilemma as this information forms the ideal base for further attacks. Jagatic et al. [9] showed that they could increase the success rate of phishing attacks from 16 to 72 % using "social data". In social engineering, additional available information on targets could lead to automated social engineering attacks [7]. The main obstacle for large-scale spam attacks on basis of SNSs are the various access protection measures providers offer to keep sensitive information private or at least limit access to a closed circle of friends. Our friend-in-the-middle attack overcomes this obstacle by hijacking HTTP sessions on the network layer, which the majority of SNSs providers fail to secure.

The main contributions of our work are:

- Our friend-in-the-middle attacks on social networks and how they can be used for context-aware spam and social phishing on a large scale.
- An evaluation of the feasibility of our attack on basis of Facebook.
- A simulation to estimate the impact a friend-in-the-middle spam campaign would have.
- A discussion on protection measures and mitigation strategies.

2. FITM ATTACKS

We define friend-in-the-middle (FITM) attacks as active eavesdropping attacks against social networking sites. While

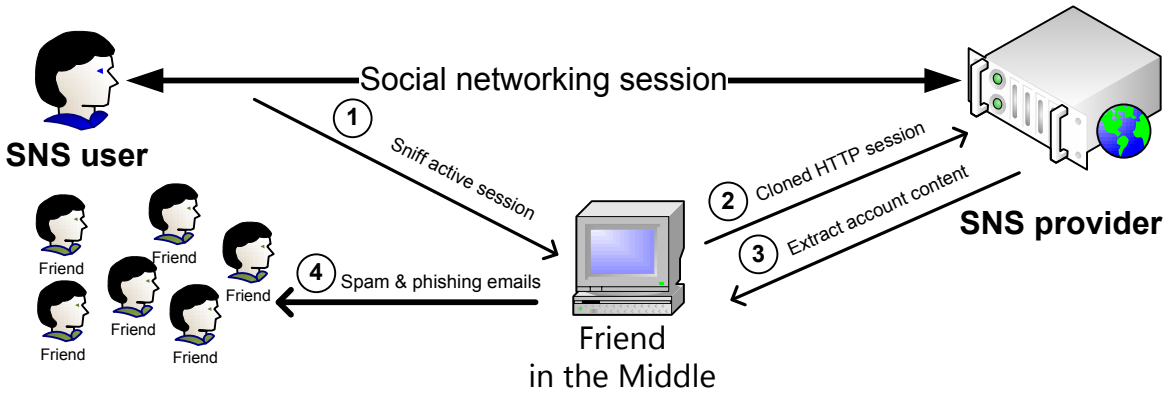

Figure 1: Outline of a large-scale spam campaign via the friend-in-the-middle attack: A social networking session is hijacked to fetch personal information from a victim's profile. The extracted information is then used for spam and phishing emails targeted at the victim's friends.

active eavesdropping attacks against web services are well studied and known for decades [2], we claim that active eavesdropping attacks against SNSs are fundamentally different for two reasons. First of all do SNSs session hijacking attacks allow various sophisticated attacks on the application layer (see below), and second of all could social networking traffic be intercepted virtually anywhere (e.g. according to [1] is Facebook, at the time of writing, responsible for 30 per cent of the world wide web traffic). Our FITM attack is based on the missing protection of the communication link between users and social networking providers. By hijacking session cookies, it becomes possible to impersonate the victim and interact with the social network without proper authorization. While at first glance the risk of hijacking social networking seems like yet another threat to privacy, we outline that FITM attacks enable large-scale spam attacks. Within this section, we first explain various attack scenarios on basis of session hijacking and describe how FITM attacks could be misused for large-scale spam campaigns on basis of Facebook.

HTTP Session Hijacking Attacks on SNSs. As a precondition the attacker needs to have access to the communication between the SNSs and the user. This can be achieved either passively (e.g., by monitoring unencrypted wireless networks) or actively (e.g., by installing malicious software on the victim's computer). The adversary then simply clones the HTTP header containing the authentication cookies and can interact with the social network, unbeknownst to the SNS operator or user. The victim is unable to detect or prevent such attacks and the attacker is able to use the social network to its full extent from the victim's point of view. As with all HTTP session hijacking attacks, it becomes possible to both retrieve information (*data acquisition from the social network*) as well as to insert malicious requests on the behalf of a user (*data publication into the social network*). However in the case of our FITM attack, further scenarios become available to attackers, which are specific to social networking sites:

- *Friend injection* to infiltrate a closed network
- *Application injection* to extract profile content
- *Social engineering* to exploit collected information

The rudimentary security and privacy protection measures of SNSs available to users are based on the notion of "friendship", which means that sensitive information is made available only to a limited set of accounts (friends) specified by the SNS user. Once an attacker is able to hijack a social networking session, she is able to add herself as a friend on behalf of the victim and thus infiltrate the target's closed network [8]. The *injected friend* could then be misused to access profile information or to post messages within the infiltrated network of friends.

By *installing* a custom third-party *application* [11], written and under the control by the attacker, it is possible to access the data in an automated fashion. Among other things, an application has access to sensitive information (birthday, email address, demographic information, pictures, interests) and in case of most SNSs to information of friends of the application user. Third-party applications such as online games have become a popular amusement within SNSs, and hiding a malicious application without any activity visible to the user is possible. Thus, the application is likely to remain undetected within a pool of installed third-party applications. This ultimately enables an attacker to extract profile content in a stealthy way as this retrieval method does not cause as much noise as a burst of separate HTTP requests. Even worse, the attacker might install the application, take all the data needed in an automated fashion and remove the application afterwards. This would be completely undetectable to the user and most likely to the SNSs providers as well.

Whereas *social engineers* traditionally relied upon context information gathered through dumpster diving or quizzing people over the phone, with FITM attacks the context information harvesting process becomes automated. We thus claim that FITM attacks allow sophisticated social engineering attacks. Two such social engineering attacks based on information extraction from social networking sites are context-aware spam and social phishing. These advanced versions of traditionally spam and phishing messages are described bellow as they are ultimately used to show the devastating effect a large-scale FITM attack might cause.

Context-Aware Spam. Context-aware spam can be generated from data harvested with FITM attacks, increasing the effectiveness of the spam. Brown et al. [5] identified three context-aware spam attacks which might be misused: relationship-based attacks, unshared-attribute attacks, as well as shared-attribute attacks. While the first attack is based on relationship information, the two remaining variations use content extracted from social networking sites such as geographic information or a user's birthday. The social network itself might be used for sending the spam, e.g. by writing the message to other users' walls, or by sending it via private messages.

Social-Phishing. Phishing is a common threat on the Internet where an attacker tries to lure victims into entering sensitive information like passwords or credit card numbers into a faked website under the control of the attacker. It has been shown [9] that social phishing, which includes some kind of "social" information specific to the victim, can be extremely effective compared to regular phishing. For example such information might be that the message appears to be sent from a person within the social environment of the victim, like a friend or a colleague from work. The social graph is therefore not only for the social network operator of value, but for an attacker too. Especially if it contains additional information like a valid email address or recent communication between the victim and the impersonated friend. With automated data extraction from social networks, a vast amount of further usable data becomes available to the spammers.

Large-scale spam campaigns through FITM attacks. Figure 1 illustrates the outline of a spam campaign exploiting our novel FITM attack. **(1)** In the first step, a network connection is monitored. Once the FITM application detects an active social networking session, it clones the complete HTTP header including the session cookie. **(2)** The cloned HTTP header serves then as a valid authentication token for the SNS provider and is used to temporarily hijack the SNS user's session. **(3)** In order to extract the profile content as well as information on the target's friends, a custom third-party application is added to the target's profile. Once all information has been extracted the application is removed from the profile. Additional queries are used to fetch the email addresses of the target's friends in case they cannot be retrieved through the third-party application. **(4)** The extracted email addresses and account content are used to generate tailored spam and phishing emails. While the spam messages contain the actual payload of the attack, the phishing emails are used to steal credentials of the target's friends for further propagation (the FITM attack starts again from (3) with the phished SNS account credentials).

We decided to evaluate the impact of a large-scale spam campaign on basis of Facebook. FITM attacks based on Facebook serve in our opinion as a good example because it is the biggest SNS at the time of writing, HTTPS is only used to protect login credentials and Facebook supports custom applications. Furthermore, injections of third-party applications into Facebook profiles promise access to a plethora of personal information. Within the Facebook application framework, third-party applications can access the following information:

- *Basic context information*: Full name, geographical location, birthday, affiliations, education, etc.
- *Likes and interests*: Favorite books, movies, tv-series, music, quotations, etc.
- *Private content*: Sent and received messages, photos, videos, etc.

In addition, third-party applications within Facebook are allowed to access the information of a user's friends as well. Thus an application injection in Facebook enables the extraction of a pool of valuable context information from the targeted user as well of his/her friends. Email addresses of users are not accessible through third-party applications and the addresses can be collected by using the hijacked user session. We created a proof-of-concept implementation of our novel FITM attack in the Python scripting language for Facebook.

3. REFERENCES

[1] Alexa. Site info: Facebook, 2010. [Online; accessed 20-January-2010], `http://www.alexa.com/siteinfo/facebook.com/trafficstats`.

[2] S. Bellovin. Security problems in the TCP/IP protocol suite. *ACM SIGCOMM Computer Communication Review*, 19(2):48, 1989.

[3] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All your contacts are belong to us: Automated identity theft attacks on social networks. In *18th International World Wide Web Conference*, April 2009.

[4] J. Bonneau, J. Anderson, R. Anderson, and F. Stajano. Eight friends are enough: social graph approximation via public listings. In *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, pages 13–18. ACM, 2009.

[5] G. Brown, T. Howe, M. Ihbe, A. Prakash, and K. Borders. Social networks and context-aware spam. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, pages 403–412. ACM New York, NY, USA, 2008.

[6] R. Gross and A. Acquisti. Information revelation and privacy in online social networks (the Facebook case). In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, 2005.

[7] M. Huber, S. Kowalski, M. Nohlberg, and S. Tjoa. Towards automating social engineering using social networking sites. *Computational Science and Engineering, IEEE International Conference on*, 3:117–124, 2009.

[8] M. Huber, M. Mulazzani, and E. Weippl. Who on earth is "mr. cypher": Automated friend injection attacks on social networking sites. In *Proceedings of IFIP/SEC 2010*, 2010.

[9] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, 2007.

[10] H. Jones and J. Soltren. Facebook: Threats to Privacy. *Project MAC: MIT Project on Mathematics and Computing*, 2005.

[11] A. Nazir, S. Raza, and C.-N. Chuah. Unveiling facebook: a measurement study of social network based applications. In *IMC '08: Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pages 43–56, New York, NY, USA, 2008. ACM.