Friend-in-the-middle Attacks: Exploiting Social Networking Sites for Spam

Markus Huber, Martin Mulazzani, Gerhard Kitzler, Sigrun Goluch,
and Edgar Weippl

Abstract

In this work we present our *friend-in-the-middle attacks* on SNSs and how it 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 exemplary on Facebook and estimate the impact based upon a simulation on a regional network of Facebook. Alarmingly, all major SNSs are vulnerable to our attack as they fail to secure the network layer appropriately.

I. Introduction

Criminals, as well as direct marketers, continue to clog mailboxes with unsolicited bulk e-mails such as 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. 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. 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 [1] as well as Jones and Soltren

[2] 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 [3] or cross-platform profile cloning attacks [4]. The leakage of personal information from these platforms creates a remarkable dilemma as this information forms the ideal base for further attacks. 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.

II. Friend-in-the-middle (FITM) Attacks

We define friend-in-the-middle attacks as active eavesdropping attacks against social networking sites. 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 active eavesdropping attacks against web services are well studied and known for decades, these attacks have a severe impact in combination with social networking services. SNSs session hijacking attacks enable more sophisticated attacks on SNSs, which we outline in the following. Moreover, SNSs providers are responsible for a major share of today's world-wide-web traffic and almost all providers fail to protect the communication layer (see Table I). Thus social networking traffic can be intercepted virtually everywhere.

HTTP Session Hijacking Attacks on SNSs. As a precondition the attacker needs to have access to the communication between the SNS and the user. This can be achieved either passively (e.g., by monitoring unencrypted wireless networks) or actively (e.g. by ARP-spoofing on a LAN). 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. Table I shows today's biggest social networking sites and their support for HTTPS and third-party applications.

Social Networking Site				
Name	Claimed users	HTTPS	App. Registration	API
Facebook	500×10^6	Login only	open	Graph API
Friendster	110×10^6	No	open	OpenSocial API
Orkut	100×10^6	Login only	open	OpenSocial API
hi5	80×10^6	No	open	OpenSocial API
LinkedIn	60×10^6	Login only	closed	OpenSocial API

TABLE I: Top five social networking sites and their support for HTTPS and third-party applications.

One can observe that if HTTPS is used at all, today's biggest SNSs provider use it solely to protect the credentials during login. As with traditional eavesdropping attacks, the attacker is able to use the web service to its full extent from the victim's point of view. However in the case of our FITM attacks, further scenarios become available, which are specific to SNSs:

- *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. 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*, written and under the control of the attacker, it is possible to access the data in an automated fashion. Among other things, an application has access to sensitive information (birthday, demographic information, pictures, interests, etc.) and in case of most SNSs also 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. An attacker might install the application, take all the data needed in an automated fashion and remove the appli-

cation 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, which we describe in the following.

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.

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 [6] 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.

With automated data extraction from social networks via FITM attacks, a vast amount of further usable data becomes available to attackers. Prior conversations within the social network like private messages, comments or wall posts could be used to deduce the language normally used for message exchange between the victim and the spam target. For example, a phishing target might find it very suspicious if the victim sends a message in English if they normally communicate in French. Another example are extracted pictures that could be included in the spam and phishing emails to increase their authenticity. Extracted pictures could for example be used to send invitations to shared "photo albums", including a link which promises more pictures given that a user enters his social networking credentials.

III. Social Spam Attacks

Spam and phishing messages via FITM attacks can be delivered using one of various approaches. First, 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. However, if used on a large scale this approach is most likely to get detected by SNSs providers who already implemented a number of anti spam strategies to protect their networks [7]. A more promising approach (from an attacker's standpoint) are out-of-bound spam messages. Out-of-bound messages mean that traditional emails or other forms of sending messages besides the SNS are used to deliver the spam and phishing messages. The traditional email spam is enabled through the availability of real email addresses users make available to their friends. Hence, if the spam attack is carried out over email instead of the SNS platform, these malicious messages cannot be detected by the SNSs providers. In the following we describe a large scale spam attack on basis of FITM attacks. In our scenario, social networking sessions are temporarily hijacked and serve as "attack seeds". Information harvested through these attack seeds is then used to generate both context-aware spam as well as social phishing emails. The attacker poses as a friend of spam attack victims by using extracted pictures and personal information and thus becomes the "Friend-In-The-Middle".

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 then serves 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).

Finding an optimal attack strategy. In order to find an optimal spam to phishing ratio as

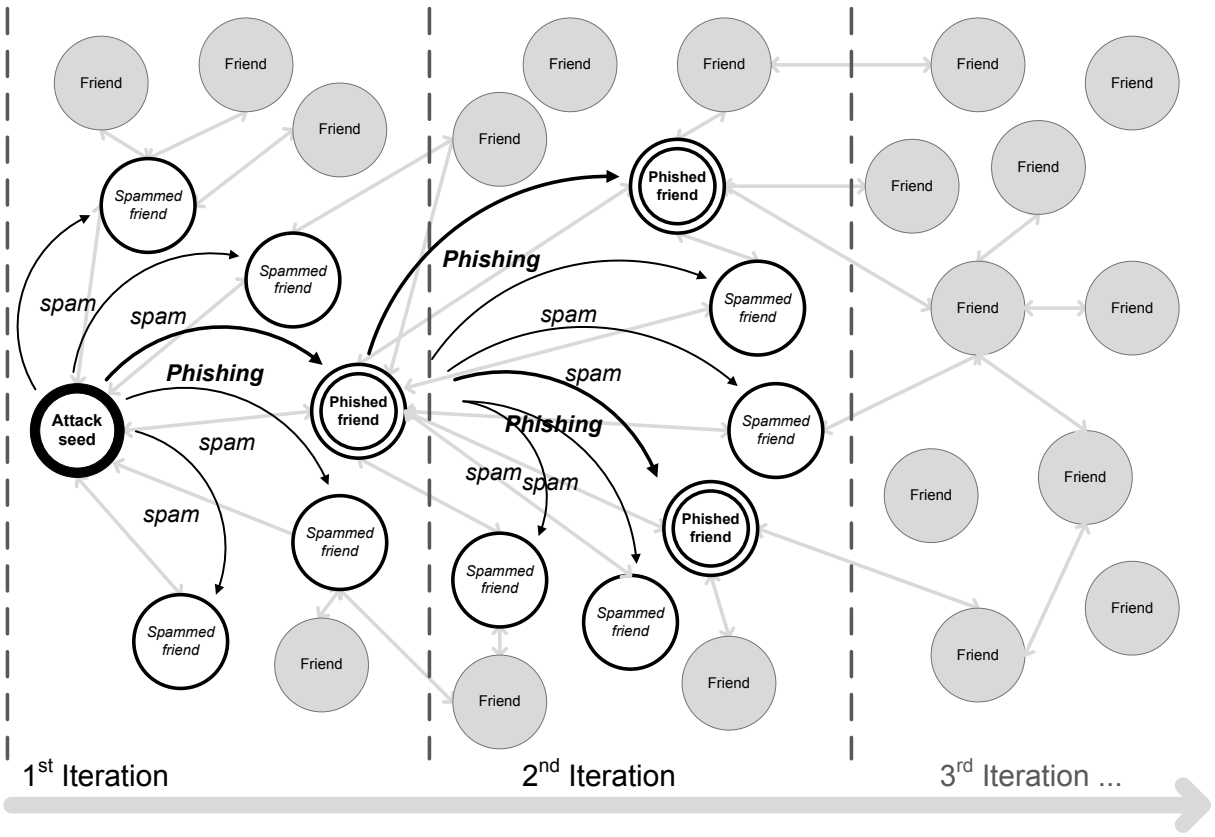

Fig. 1: Outline of a large-scale spam campaign via the friend-in-the-middle attack: A social networking session is hijacked to fetch personal information from a victim's profile. The extracted information is then used for spam and phishing emails targeted at the victim's friends.

well as to determine if increasing the number of iterations or attack seeds yields more spam targets, we modeled a Facebook graph as a configuration model and simulated two different attack strategies. A node in the graph is called a user and its degree is his number of friends.

As an applicable degree distribution we used the much studied *power-law* distribution [8]. Distributions of the form $p(x) = Cx^{-\alpha}$ are said to follow a power law, where C functions as a *normalization constant*. C is given by the normalization requirement

$$1 = C \int_{x_{min}}^{\infty} x^{-\alpha} dx = \frac{C}{1-\alpha} \left[x^{-\alpha+1} \right]_{x_{min}}^{\infty} \tag{1}$$

Formula 1 shows that: 1) $\alpha > 1$ and 2) for a given $\alpha > 1$ and known limit x_{min} it is easy to compute the normalization constant C. Gjoka et al. [9] presented a new degree distribution for Facebook which does not follow one power-law but two. They found two regimes $1 \leq k < 300$

and $300 \leq k \leq 5000$, each following a power law with exponent $\alpha_{k<300} = 1.32$ and $\alpha_{k\geq 300} = 3.38$. With this specific information it was possible to generate an accurate power law degree sequence for the two intervals $[1; 300[$ and $[300; 5000]$. One attack process behaves as follows: We choose a random node v_i (**user**), the user has a predetermined degree ($d(v_i) = d_i = k$) which is the amount of friends. We *spam* a fixed percentage p of the users friends and *propagate*[1] the remaining ones. This cycle then repeats itself for a given amount of iterations. Our underlying assumption is that an attacker would try to spam as many targets as possible with a minimum number of sent messages. A visual diagram of this cycle can be seen in Fig. 1.

We tested two strategies within our implemented model. **Strategy 1:** randomly choose a user, spam and propagate as mentioned above for $it = 1, ..., l$ times. Stop after l iterations. **Strategy 2:** randomly choose a user, fix the number of iterations ($it = m$), after repeating the cycle m times select another randomly chosen user (attack seed) and repeat the cycle another m times. This repeats itself for a fixed number of times, e.g. $as = 8$. The major difference between strategy 1 and 2 would be the dimension of l resp. m: $l \gg m$

We ran our attack model for *strategy 1* for each number of iterations, from $it = 1, ..., 35$, 1000 times. For each iteration step we averaged over the 1000 results and got the mean value for the potential spam targets. The simulation shows that after a very high slope in the beginning the number of spammed targets slowly levels. The slope is significantly higher for values of p between 50 and 80 percent. According to our simulation the value $p = 70\%$ (70% context-aware spam, 30% social phishing) is the best choice for p. The leveling of all curves yields to the assumption that in a highly clustered structure it is not possible, with this strategy, to elude the whole cluster. For strategy 2 we used the optimized value $p = 70\%$ and varied the number of initial attack seeds, from $as = 1, ..., 35$ for a fixed number of iterations ($it = 1, 2, 3, 10, 20, 35$). For each number of attack seeds and iterations we calculated the mean value for the spammed nodes. Our simulation shows that with a rather small amount of iterations ($it = 3$) it takes only a few attack seeds ($as = 7$) to reach the asymptotic limit of possible spammed targets in our simulated graph. For a more detailed description of our simulation we refer the interested reader to [10].

[1]We assume that phishing has a success rate of 72 % as the propagation is based on social-phishing [6].

IV. Impact Evaluation

In order to make assertions on the impact of a FITM spam attack, an experiment which mimics a real large-scale attack would provide valuable insights, but raises also serious ethical concerns. Due to the viral nature of our spam attack it is impossible to get ethical approval for all involved experiment subjects beforehand. Hence, we applied the following twofold approach: we made an empirical evaluation on the number of possible sessions that could have been hijacked, without collecting any data or injecting any malicious requests. We furthermore simulated the impact of our FITM attack on basis of well established research in the area of social network security. This prevents us from processing or storing sensitive information like birthdays or real e-mail addresses, also due to legal constraints. We decided to evaluate the impact of this large-scale spam campaign on basis of Facebook. FITM attacks based on Facebook serve in our opinion as a good example because it is by far the biggest SNS at the time of writing, HTTPS is only used to protect login credentials and Facebook supports external applications. Furthermore, injections of third-party applications into Facebook profiles promise access to a plethora of personal information.

Finding attack seeds. To conduct the FITM attack, numerous attack vectors can be used: DNS poisoning, cross-site request forgery (*CSRF*), wireless networks, deep packet inspection from an ISP or other malicious entity that has access to the traffic between the client and the SNS. However, we used our proof-of-concept FITM application to analyze HTTP cookies from Facebook sessions passing through a Tor exit node. The Tor network [11] is a widely deployed anonymization network which hides the user's IP address on the Internet. It is expected to be used by hundreds of thousands of users every day and is believed to be the most heavily used open anonymization network today. The Tor infrastructure relies on servers run by volunteers, hence anyone can support the Tor project by setting up a dedicated Tor server. For our experiment, we have set up a Tor exit node on a minimal GNU/Linux Debian server with a relay bandwidth rate of 5 Mbit. The server was furthermore configured to allow HTTP traffic (TCP port 80) from the Tor network to the Internet, and the Tor daemon was restarted on a daily basis. We then counted the number of Facebook sessions that were observable to our Tor server and could have been used for FITM attacks. To prevent counting the same user multiple times, we saved hash values of the static session information in an encrypted file container.

Simulating a large-scale attack. We used the anonymous regional network collected by Wilson et al. [12], which contains 3×10^6 Facebook nodes, as basis of our attack simulation. The goal of our simulation was to estimate how many Facebook nodes would be affected by a large-scale FITM spam attack given a specific amount of initial attack seeds. We used the optimized spam to phishing ratio of 70 to 30 per cent, which we found earlier through our simulations on the generated Facebook subgraph (attack strategy 1). Furthermore, as our attack optimization results of strategy 2 suggested, we used a small number of iterations ($it = 3$).

A. Results

During a period of 14 days, approximately 6.1×10^6 HTTP requests passed through our Tor exit node. Facebook was the most requested domain and was responsible for 7.68 % of all overall requests. The second most frequent social networking site was Orkut which caused 0.49 % off all HTTP requests. We observed 4267 unique Facebook sessions throughout our experiment which could have been hijacked for friend-in-the-middle attacks. Furthermore, our cookie analysis suggests that the majority (92.81 %) of observed unique Facebook sessions were persistent sessions.

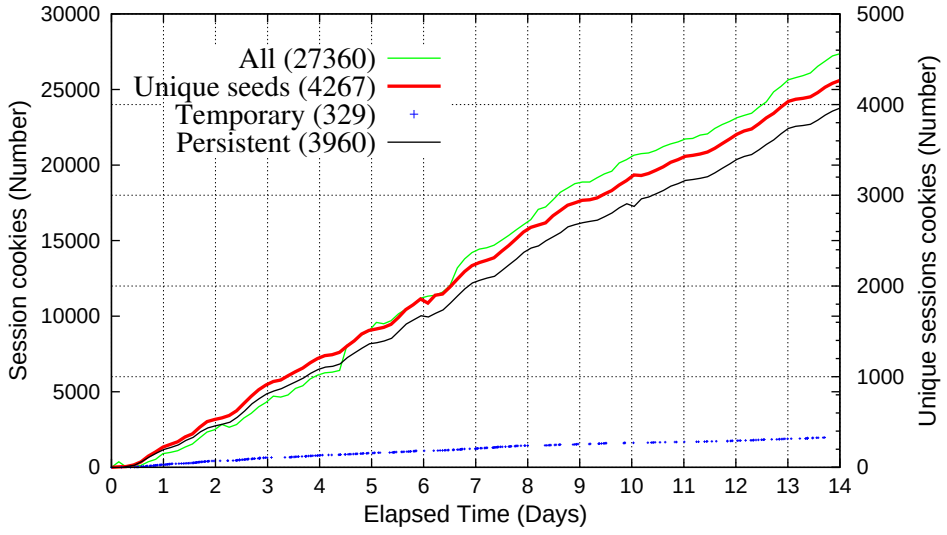

Fig. 2: Number of sessions found through our Tor exit node server within 14 days.

One alternative source for attack seeds is eavesdropping on a WLAN. Indicative experiments on a university's WLAN access point showed that we could gather 60 seeds within seven hours.

The main drawback of this method is that seeds are not dispersed as evenly over the entire social graph, as many students are friends on Facebook or share at least one common friend.

seeds	seeds [%]	targets	targets [%]
250	0.01	$1.94{\cdot}10^5$	6.28
450	0.01	$2.12{\cdot}10^5$	6.86
750	0.02	$2.27{\cdot}10^5$	7.35
1000	0.03	$2.40{\cdot}10^5$	7.77
1500	0.05	$2.58{\cdot}10^5$	8.35
2000	0.06	$2.70{\cdot}10^5$	8.74
3000	0.10	$2.90{\cdot}10^5$	9.39
4000	0.13	$3.03{\cdot}10^5$	9.80

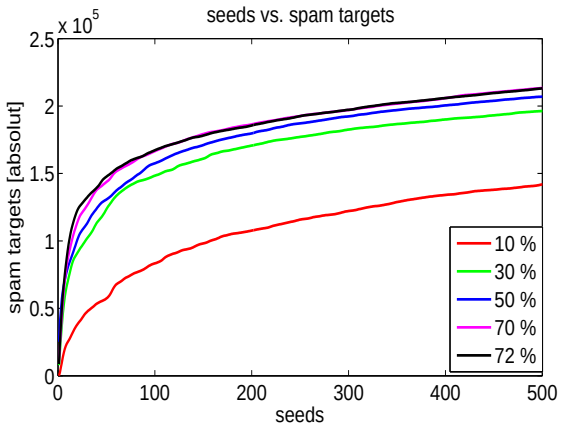

Fig. 3: Shows the results of our large-scale attack simulation. The table on the left outlines how many targets would be reached with our 4000 collected seeds. The figure on the right shows the amount of seeds vs. the amount of potential spam targets with different phishing success probabilities.

The table in Fig. 3 (left) shows our simulation results for 4000 seeds with a propagation success probability of 72%. We observed over 4000 seeds in two weeks which corresponds to 0.13 per cent of the total amount of nodes of the regional network. With only 250 seeds it is possible to find 1.94×10^5 potential spam targets (6.28 % of the total nodes). With 4000 attack seeds within the regional network, the FITM attack finds fewer nodes that have not been already spammed. Thus our result of 3.03×10^5 overall spam targets is a conservative estimation as seeds collected trough a Tor server belong to a disperse set of Facebook clusters. Figure 3 (right) shows the impact the phishing success rate has on the overall spam targets. As the figure illustrates, even with a 10%-success rate, it is possible to spam 3% of the regional network. For higher success probabilities the difference between the curves decreases.

B. Protection Measures

In order to effectively mitigate FITM attacks, SNSs providers have to ultimately ensure that all communication between their users and their platform is done over HTTPS. At the time of

writing only XING fully supports HTTPS, which leaves SNS users with browser extensions as the only working mitigation strategy. Browser extensions such as EFF HTTPS Everywhere [13] offer a transitional mitigation strategy to the average user by attempting to force HTTPS for requests that would have been normally transferred over HTTP. A transition from HTTP to HTTPS is a non-trivial tasks as SNSs have become complex web services spawning over multiple domains and HTTPS creates additional performance and network costs. For our described scenario full HTTPS support is the obvious solution, FITM attacks might however also be carried out using e.g. stolen SNSs credentials as initial attack seeds. Thus sophisticated protection schemes are required, so far two possible schemes have emerged. Privacy enhancement extensions aim to limit the amount of information that can be extracted from user profiles. An example for this class of mitigation strategies is the flyByNight application for Facebook [14], which encrypts messages between users with strong cryptography. The second class are privacy-enabling architectures which can be used atop of social networking services. Such an approach was for example chosen by the designers of FaceCloak [15], which intends to hide sensitive user information encrypted on separate servers and by providing fake information. Both approaches have the shortcoming that only text-based content can be protected leaving out e.g. pictures that could be used for context-aware spam and social phishing. Thus in our opinion novel privacy-by-design SNSs would ultimately help to mitigate a number of know security attacks against SNSs including FITM attacks.

V. Conclusion

In this article we have introduced friend-in-the-middle (FITM) attacks which are active eavesdropping attacks against social networking sites. By cloning a user's authentication cookie which is transmitted in an unencrypted way, it becomes possible to completely impersonate the user. This can then be used to collect sensitive information in an automated fashion which ultimately enables large context-aware spam campaigns that propagate via social phishing. FITM attacks are applicable to the great majority of currently deployed SNSs, such as Facebook, Friendster, and Orkut. Based on FITM attacks we described the following subsequent exploits: (1) Friend injection, (2) Application injection, and (3) Social engineering. We furthermore evaluated the impact of a large-scale spam attack on basis of FITM attacks. Our experiments showed that finding possible FITM attack seeds for spam campaigns is cheap regarding time and hardware

resources. Our attack simulation results furthermore suggest that based on the 4000 possible attack seeds we observed, 3.03×10^5 users could have been targeted with context-aware spam. There are a number of limited protection strategies available to social networking users. Hence social networking providers ultimately have to protect their users against FITM attacks by securing the communication channels of their services with HTTPS. As there is no (monetary) incentive for them to do so, we believe that our attack remains applicable for the time being.

REFERENCES

[1] 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*, 2005, pp. 71–80.

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

[3] 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*. ACM, 2009, pp. 13–18.

[4] 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.

[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*. ACM New York, NY, USA, 2008, pp. 403–412.

[6] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Communications of the ACM*, vol. 50, no. 10, pp. 94–100, 2007.

[7] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Fighting spam on social web sites: A survey of approaches and future challenges," *IEEE Internet Computing*, pp. 36–45, 2007.

[8] M. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, September 2005.

[9] M. Gjoka, M. Kurant, C. Butts, and A. Markopoulou, "Walking in facebook: A case study of unbiased sampling of osns," in *Proceedings of IEEE INFOCOM '10*, San Diego, CA, March 2010.

[10] M. Huber, M. Mulazzani, E. Weippl, G. Kitzler, and S. Goluch, "Friend-in-the-middle attacks," SBA Research, Tech. Rep. TR-SBA-Research-0710-01, 2010. [Online]. Available: http://www.sba-research.org/wp-content/uploads/publications/FITM_TR-SBA-Research-0710-01.pdf

[11] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," in *13th USENIX Security Symposium*, San Diego, CA, USA, August 2004.

[12] C. Wilson, B. Boe, A. Sala, K. Puttaswamy, and B. Y. Zhoa, "User interactions in social networks and their implications," in *Proceedings of the 4th ACM European conference on Computer system*. ACM New York, NY, USA, 2009, pp. 205–218.

[13] "EFF HTTPS Everywhere," https://www.eff.org/https-everywhere.

[14] M. Lucas and N. Borisov, "flybynight: Mitigating the privacy risks of social networking," in *Proceedings of the 7th ACM workshop on Privacy in the electronic society*. ACM, 2008, pp. 1–8.

[15] W. Luo, Q. Xie, and U. Hengartner, "FaceCloak: An Architecture for User Privacy on Social Networking Sites," in *Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 3, 2009.