

Security Issues in Online Social Networks

This article surveys the current state of security issues and available defense mechanisms regarding popular online social networks. It covers a wide variety of attacks and the corresponding defense mechanisms, if available. The authors organize these attacks into four categories — privacy breaches, viral marketing, network structural attacks, and malware attacks — and focus primarily on privacy concerns. They offer an in-depth discussion of each category and analyze the connections among the different security issues involved.

Hongyu Gao Northwestern University

Jun Hu

Huazhong University of Science and Technology

Tuo Huang Yale Law School

Jingnan Wang and Yan Chen Northwestern University ver the past few years, the popularity of online social networks (OSNs) such as Facebook, Twitter, and Orkut has grown tremendously. OSNs are built on real-world social relationships and provide their users with a wide variety of virtual-interaction mechanisms. As OSNs have become critical online communication platforms integrated into society's daily life, the security risks accompanying such developments have raised concerns in industry, academia, and government.

This survey provides a comprehensive view of the security issues in OSNs today. We cover a wide variety of attacks and the corresponding defense mechanisms. We organize these attacks into four broad categories: privacy breaches, viral marketing, network structural attacks, and malware attacks. Privacy breach attacks are either unique to OSNs or become feasible for large-scale studies because of OSNs' emergence. They are the main focus of this survey; the other three categories aren't new by themselves, but they have a new context in OSNs and are worth reexamining. In particular, they heavily exploit the credulity and carelessness of people tricked by social-engineering techniques. These various types of attacks aren't completely separate from one another, but rather are closely intertwined and are sometimes combined.

Privacy Breach Attacks

Users provide an astonishing amount of personal information voluntarily, and OSN service providers store this information. Ralph Gross and Alessandro Acquisti studied the Facebook users in the Carnegie Mellon University network and discovered that 90.8 percent of users uploaded their images, 87.8 percent revealed their birth dates, 39.9 percent shared their phone numbers, and 50.8 percent listed their current addresses.¹ Such an abundance of readily available personal information makes privacy breach a unique angle of attack in OSNs.

Three primary parties interact with one another in an OSN: the service provider, the users, and third-party applications.

Breaches from Service Providers

OSNs' current client-server architecture inherently dictates that users must trust service providers to protect all the personal information they've uploaded. However, service providers can obviously benefit from examining and sharing this information — for advertising purposes, for example. Because service providers have the power to use such information however they wish, researchers have raised serious concerns and have attempted to redress this power imbalance.

Researchers have proposed various alternative OSN architectures as defenses. These proposals suggest that users should dictate the fine-grained policies regarding who may view their information. To enforce this userdefined policy, the OSN stores the information with encryption, so that no entity - not even the OSN service provider - can see the information unless the owner has somehow granted access to it. For example, Persona uses decentralized storage so that users can choose where in the network to store their information.² Persona supports both public-key cryptography (to share information with any single entity in the network) and attribute-based encryption (to share content with entire groups).

Similarly, Lockr separates social network content from OSN functionalities.³ This approach lets users decide where to store their information without interrupting the OSN functionalities. In Lockr, the recipient of digitally signed social relationships can provide these signed social relationships to the OSN as proof to fetch social data. Lockr then ensures that the OSN can't reuse the signed social relationships for unintended purposes.

Jonathan Anderson and his colleagues have proposed an OSN architecture consisting of smart clients and an untrusted central server.⁴ The server stores encrypted data so that it's available only for those who have been granted access to it. Thus, the client can access user information only if the owner's client mediates the access.

Breaches from Other Users

OSNs facilitate communication among friends. While fulfilling this purpose, service providers protect users' privacy from unconfirmed access. As a trade-off, all major OSNs let a user's friends access the personal information the user has uploaded to his or her profile by default, while blocking others from doing so. However, the notion of "friends" in an OSN is merely a social link that the two users have agreed to establish in that OSN, regardless of the actual offline relationship. This discrepancy provides a potential channel for stealing personal information by befriending users in OSNs.

Even the simplest forms of such attacks are successful. For example, 75,000 out of 250,000 random Facebook users contacted using an automatic script accepted the script's request to become a Facebook friend.⁵

Leyla Bilge and her colleagues have presented two more-sophisticated attacks.⁶ The first attack is called same-site profile cloning. An attacker duplicates a user's profile in the same OSN and uses the duplication to send out friend requests to the user's friends. Believing the request has come from a familiar person, the unalerted friends can accept it and thereby expose their personal information to the attacker.

The second attack is cross-site profile cloning. The attacker identifies a user from OSN A, along with this user's friend list. The attacker then duplicates the profile to OSN B, where the user hasn't yet registered, and sends out friend requests on OSN B to the target's friends who have also registered on OSN B. Cross-site profile cloning is potentially more dangerous than same-site cloning because it's less likely to arouse suspicion.

Currently, no definitive defense can protect against such attacks. However, Bilge and her colleagues suggest increasing users' alertness concerning their acceptance of friend requests.⁶ Also, improving the strength of Captcha can help prevent large-scale profile-cloning attacks using automated scripts.

Breaches from Third-Party Applications

As OSNs expand their services, third-party applications are flourishing because of user demands for additional functionalities. Although these applications reside on the OSN platform, a third party develops them, so they're essentially untrusted. In addition, users must grant the application access to their personal data before they can install those applications, because such access is necessary for some applications to perform their functionality. For example, a horoscope application must know the user's birthday.

Unfortunately, neither the service provider nor the users know exactly which piece of information is truly necessary for the applications. As a result, they must trust the applications to correctly declare the information they need. In addition, the mechanism to monitor how the applications manipulate the personal information is missing. This leaves the door open for the applications to misuse that information. For example, a popular Facebook application, Compare Friends, promised users privacy when they expressed opinions about their friends, and then later offered to sell that information.⁷

Kapil Singh, Sumeer Bhola, and Wenke Lee have proposed XBook to counter such attacks.⁷ They use information flow models to control what untrusted applications can do with the information they receive. In the XBook design, applications have a set of components. Any communication between two components, or between a component and an external entity, can occur only via XBook APIs. When adding a particular application, the user receives a list of the personal information that the application requests to access and share with external entities. Then, XBook ensures that the application can access and share the information only according to what the user has explicitly agreed. However, because XBook relies on the list of personal information provided by the application in the first place, it only solves the problem of monitoring how the application manipulates this personal information. How to determine which information the application actually needs is still an open question.

Recently, Facebook updated its privacy policy so that applications must obtain specific approval from users before gaining access to any personal information that isn't available to "everyone."

Reidentification and De-anonymization

OSN operators and researchers are increasingly sharing anonymized (that is, all personally identifiable information has been removed) social network structure with other researchers, application developers, and advertisers. Unfortunately, de-anonymizing attacks might be able to reidentify a particular user in the anonymized social network, thus defeating anonymization and breaching that user's privacy.

Lars Backstrom, Cynthia Dwork, and Jon Kleinberg have presented an active attack and a passive attack.⁸ In the active attack, the adversary registers a few accounts in the social network, creates a link pattern among those accounts, and connects them to the target users. After the anonymization, the adversary can efficiently reidentify the created nodes as well as the target users. In the passive attack, the adversary doesn't create new nodes or edges for de-anonymization, but rather exchanges structural information with a small coalition of friends and uniquely identifies this coalition's subgraph, which enables those colluding friends to locate themselves.

Arvind Narayanan and Vitaly Shmatikov have presented another large-scale passive attack.⁹ In this attack, the adversary identifies individual users by building a mapping between the target anonymized graph and the auxiliary graph on the basis of the intuition that the network topology for the same user in different social networks is still similar.

Gilbert Wondracek and his colleagues show that group membership in the OSN is sufficient to uniquely de-anonymize the user.¹⁰ In addition, they exploit the attack technique called history stealing to identify the groups to which the user behind a browser belongs. Consequently, a malicious website could uncover user information (for instance, their full names).

The implication of de-anonymizing attacks is far-reaching. For example, in recent years, identity theft has become a widespread concern. To an identify-theft criminal, the most critical, valuable piece of information is a person's social security number (SSN). Knowing the structure of the nine-digit SSN, an attacker can use information revealed from OSNs to uncover the first five digits.¹

Defense against de-anonymization attacks often involves increasing the difficulty of data acquisition for the attacker because such attacks usually require a relatively large volume of data. Little can be done from the client side, whereas server-side defense is more promising. For example, Wondracek and his colleagues suggest using dynamic hyperlinks to effectively hinder the automatic collection of data.¹⁰ Even adding a simple, alphanumeric string two characters long as the token for each URL can increase the attacker's search space by a factor of 3,844.

Viral Marketing

Because OSNs are formed by real people, they are tempting targets for viral marketing, which has therefore easily invaded them. The public perception that OSNs encompass friends, family, and acquaintances makes users prone to trust messages they receive in OSNs. In addition, aided with the information extracted from user profiles, spammers can often exploit social-engineering tricks to enhance viral marketing's effectiveness.

Spam in OSNs

Two types of spamming activities in OSNs are worth noting. The first is context-aware spamming,¹¹ which is likely to have a high click-through rate because of the increased authenticity recipients perceive. The spammer achieves the high click-through rate by taking advantage of the shared context among friends on social networks. Moreover, OSNs provide search functionality to help locate users with certain properties (location, school, workplace, and so on), and this functionality enhances the spammer's ability to discover a well-defined target set.

The second type is broadcast spamming, which doesn't have specific targets, but rather abuses public interaction mechanisms to disseminate information. For example, spammers might pollute a collaborative tagging system such as www.delicious.com. In such a system, any user could annotate any resource in the system using free-form tags. Spammers exploit this freedom to associate the resource with misleading tags that direct users to the adversary's links.¹²

Phishing and Account Attacks

A phishing attack targets OSN users' confidential information (OSN account credentials, email address, online banking, and so on). If it targets OSN account credentials, such an attack is usually combined with spamming to complete the viral-marketing process. Phishing attacks in OSNs don't differ essentially from those traditionally executed through email, although they have higher success rates. Tom Jagatic and his colleagues have conducted experiments that involve real phishing attacks on real users. They show that, aided with information obtained from OSNs, phishing is four times more effective than "blind" attempts.¹³ With the obtained personal information, an attacker can better impersonate the victims' friends and acquaintances and spoof email messages to direct the victims to sites where they're prompted for their usernames and passwords.

Defense Mechanisms

The centralized administration of OSNs facilitates the enforcement of spam control. The proposed countermeasures form three categories: detection, demotion, and prevention. Paul Heymann, Georgia Koutrika, and Hector Garcia-Molina offer detailed descriptions on all three forms of countermeasures.¹⁴

Plausible defense mechanisms for phishing attacks are mostly on the client side, given the voluntary nature of users' sharing their confidential information. Digitally signed emails can prevent spoofed email messages from fooling recipients; browser toolbars can also alert users when they're prompted for usernames and passwords at suspicious sites.¹⁵

Network Structural Attacks

One popular type of attack is the Sybil attack, which an adversary launches by exploiting the network structure.¹⁶

Sybil Attacks and Social Networks

In a Sybil attack, an individual entity masquerades as multiple simultaneous identities. Researchers have extensively studied Sybil attacks in the other areas of computer networks such as peer-to-peer (P2P) systems. However, such attacks also pose serious threats to OSN security because social networks contain many users interacting spontaneously as peers interacting in a P2P network. The fundamental problem is that one entity can control multiple identities in the system. By manipulating these identities, the adversary can render the result of the applications running on the system questionable, if not incorrect. A concrete example is that by controlling many identities, the adversary can promote the popularity and reputation of an account in e-commerce settings by voting the target account as "good."

In a more sophisticated scenario, a deanonymization attack can leverage a Sybil attack.⁸ The adversary strategically creates new accounts and links them in the network, so that when the anonymized network is released, he or she can recover information using the particular topological feature introduced by the Sybil accounts.

To launch a Sybil attack, the adversary needs only the OSN accounts. Depending on the different attack goal, the number of accounts required can vary from as small as seven (as in the attack demonstrated by Backstrom, Dwork, and Kleinberg⁸) to thousands (for manipulating voting results).

Defense Mechanisms

Currently, three main categories of defense mechanisms can help prevent the entrance of Sybil nodes or identify those nodes. Although researchers have proposed all these defense mechanisms for P2P systems, they apply to OSNs as well.

Trusted certification. In trusted certification, only verified users can enter the network. This has proven to be the only technique that can potentially completely eliminate Sybil attacks.¹⁶ The approaches proposed by Miguel Castro and his colleagues,¹⁷ as well as numerous others, fall into this category. However, this method requires a centralized authority verifying each individual identity, and this requirement is considered a shortcoming for the actual deployment of such techniques in P2P networks. Fortunately, the OSN naturally has a centralized authority - the site itself - which makes such approaches somewhat feasible. However, such verification eventually must be done manually, so there are still scalability issues for larger networks.

Resource testing. The second category of defense involves resource testing, which investigates computing ability, storage ability, network structure, network bandwidth, and the number of IP addresses associated with the nodes representing actual users. Haifeng Yu and his colleagues have proposed SybilGuard,¹⁸ a decentralized approach that identifies Sybil nodes having extremely small "quotient cuts"

between them and the honest nodes. Established trust relationships – represented as friendship links – among legitimate users in a social network form an honest region of the network and are "fast mixing" (that is, the legitimate nodes have good connectivity to the rest of the social graph). Meanwhile, attackers could create many Sybil nodes, but relatively few of them will establish trust relationships in the honest region. On the basis of this insight, SybilGuard relies on a special random walk in the graph, as well as the intersections between different walks, to identify small quotient cuts between the honest region and the Sybil region.

Recurring costs. Sybil attacks can't be launched until a significant number of Sybil nodes are created. Therefore, some approaches try to impose an additional cost during node creation. Besides using Turing tests such as Captcha, using recurring validation mechanisms can also significantly increase the cost of creating many Sybil nodes.¹⁹

Malware Attacks

In addition to file sharing and email, attackers are also exploiting OSNs to spread malicious software. Attackers can spread worms and establish botnets more easily because of the rich and frequent interactions in the OSN. Malware can propagate over social networks via profile, interaction, and third-party applications.

The Koobface worm is one of the most notorious worms in OSNs.²⁰ It's the first malware to have a successful and continuous run propagating through social networks. It executes an attack by automating Internet browsers to perform the following tasks:

- 1. Register and activate a Facebook account using a Gmail address.
- 2. Join random Facebook groups, adding Facebook friends.
- 3. Post messages on the friends' walls that contain links to the Koobface loader component.

Another worm also targeting Facebook uses cross-site request forgery (CSRF or XSRF) to spread itself.²¹ This worm lures unsuspecting users to click on a link. If the user does so, he or she opens an exploited page that calls a sequence of other pages and scripts. Eventually, the worm presents a form submission to Facebook

Table 1. Comparison of the different types of attacks.						
Measure	Information leak	De-anonymizing	Phishing	Sybil	Malware	Spamming
Attack difficulty	Easy to medium	Medium	Easy	Hard	Hard	Easy
Server defense effectiveness	Yes, but with limited effectiveness	No	No	Yes	Yes, but with limited effectiveness	Yes
User defense effectiveness	Yes	No	Yes	No	Yes, but with limited effectiveness	No
Threat to user	High	Medium to high	High	Medium	High	Low

as if the victim himself submitted a URL for a wall post and clicked on the "Share" button to confirm the posting. Next, all his friends will see this message as well as the link.

Almost all major OSNs are targeted by one type of malware or another, such as worms, information stealers, and password stealers – for example, Grey Goo targeting Second Life, JS/SpaceFlash targeting MySpace, Kut Wormer and Scrapkut targeting Orkut, and Secret Crush targeting Facebook.²²

On the defense side, Wei Xu, Fangfang Zhang, and Sencun Zhu have proposed a maximumcoverage algorithm that selects a subset of normal OSN users to whom the defense system attaches "decoy friends" to monitor the entire social graph.²⁰ Once the decoys receive suspicious worm propagation evidence, the system performs local and network correlations to distinguish actual worm evidence from normal user communication.

Comprehensive View of OSN Attacks

The categories of attacks we've discussed are representative of the most common threats reported today. Table 1 summarizes the different types of attacks.

In Table 1, "attack difficulty" refers to the technical difficulty of launching the attack. Certain types of attacks, such as Sybil attacks, require advanced technical knowledge to create the necessary components and exploit the propagation vectors. Other types, such as spamming, require very little knowledge of computer network systems. "Server defense effectiveness" indicates whether a server-side defense mechanism exists and, if so, whether this defense is effective; similarly, the table compares userside defense mechanisms.

Table 1 also compares the threat levels these attacks pose to users. We consider an attack as posing a high threat level when it directly threatens the sensitive data that the users wish to protect or causes irreversible damage to the system's privacy and anonymity.

Finally, the different types of attacks are often closely intertwined. An adversary could launch one type of attack in the form of another (such as using a Sybil attack for de-anonymization, as we discussed earlier), or one attack could require the assistance of (or data obtained from) another attack.

For example, account information obtained from phishing can be used for spamming. The efficacy of doing so is twofold: the undermined accounts provide spamming targets (most commonly, their friend lists), as well as enabling access to otherwise protected information that the spammer could use for context-aware spamming. Of course, an adversary could also use such information for de-anonymization and information stealing, or could use these undermined accounts to originate the spread of malware as well.

Similarly, it's possible to use stolen or deanonymized user information for phishing. As we've discussed previously, such information greatly increases a phishing attack's success rate by lowering the recipients' alertness because of messages spoofed with their personal and relationship information. Therefore, an effective defense against any type of security threat in OSNs requires considering other threats at the same time. While addressing a particular issue, we need to keep in mind whether other types of attacks can circumvent the designed methodology, as with a privacy control system and a profilecloning attack: a successful profile-cloning attack can undermine the system, regardless of how advanced or sophisticated that system is.

Designing an effective defense can be difficult because increased security procedures might reduce the enjoyment of the Web service. OSN service providers often face a tradeoff between security and user friendliness. For example, Bilge and her colleagues suggest increasing Captcha tests' strength and frequency.⁶ In practice, however, users might get frustrated by overly difficult Captchas or be deterred from using certain functionalities because of frequent Captcha tests. Moreover, a commercial setting would also require considering business interests during the development of security mechanisms.

Another aspect that the literature has repeatedly emphasized is user awareness. For the majority of security threats, if users don't take the initiative to protect their information, most server-end defenses would fail disastrously.

O SNs are playing an increasingly important role in the Internet community today. Their future development hinges on their ability to deliver enjoyable services without undermining users' information security. We hope this article provides insights and clues that will lead to future improvements in OSN security.

References

- R. Gross and A. Acquisti, "Information Revelation and Privacy in Online Social Networks," *Proc. ACM Workshop Privacy in the Electronic Soc.* (WPES 05), ACM Press, 2005, pp. 71–80.
- R. Baden et al., "Persona: An Online Social Network with User-Defined Privacy," *Proc. ACM SIGCOMM Conf. Data Comm.* (SIGCOMM 09), ACM Press, 2009, pp. 135–146.
- A. Tootoonchian et al., "Lockr: Better Privacy for Social Networks," Proc. 5th Int'l Conf. Emerging Networking Experiments and Technologies (CoNEXT 09), ACM Press, 2009, pp. 169–180.
- J. Anderson et al., "Privacy-Enabling Social Networking over Untrusted Networks," *Proc. 2nd ACM Workshop Online Social Networks* (WOSN 09), ACM Press, 2009, pp. 1–6.
- K. Jump, "A New Kind of Fame," Columbia Missourian, 1 Sept. 2005 (updated 21 July 2008); www. columbiamissourian.com/stories/2005/09/01/a-new-kindof-fame.
- L. Bilge et al., "All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks," *Proc. 18th Int'l Conf. World Wide Web* (WWW 09), ACM Press, 2009, pp. 551–560.
- K. Singh, S. Bhola, and W. Lee, "XBook: Redesigning Privacy Control in Social Networking Platforms," *Proc. 18th Usenix Security Symp.* (SSYM 09), Usenix Assoc., 2009, pp. 249–266.

- L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore Art Thou r3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography," *Proc. 16th Int'l Conf. World Wide Web* (WWW 07), ACM Press, 2007, pp. 181–190.
- A. Narayanan and V. Shmatikov, "De-anonymizing Social Networks," *Proc. 20th IEEE Symp. Security and Privacy* (SP 09), IEEE CS Press, 2009, pp. 173–187.
- G. Wondracek et al., "A Practical Attack to De-anonymize Social Network Users," *Proc. IEEE Symp. Security and Privacy* (SP 10), IEEE CS Press, 2010, pp. 223–238.
- G. Brown et al., "Social Networks and Context-Aware Spam," Proc. ACM Conf. Computer Supported Cooperative Work (CSCW 08), ACM Press, 2008, pp. 403–412.
- B. Markines, C. Cattuto, and F. Menczer, "Social Spam Detection," Proc. 5th Int'l Workshop Adversarial Information Retrieval on the Web (AIRWeb 09), ACM Press, 2009, pp. 41–48.
- T.N. Jagatic et al., "Social Phishing," *Comm. ACM*, vol. 50, no. 10, 2007, pp. 94–100.
- P. Heymann, G. Koutrika, and H. Garcia-Molina, "Fighting Spam on Social Web Sites: A Survey of Approaches and Future Challenges," *IEEE Internet Computing*, vol. 11, no. 6, 2007, pp. 36–45.
- 15. N. Chou et al., "Client-Side Defense against Web-Based Identity Theft," Proc. 11th Ann. Network and Distributed System Security Symp. (NDSS 04), Internet Soc., 2004; www.isoc.org/isoc/conferences/ndss/04/proceedings/ Papers/Chou.pdf.
- J.R. Douceur, "The Sybil Attack," Proc. Revised Papers from 1st Int'l Workshop Peer-to-Peer Systems (IPTPS 02), LNCS 2429, Springer, 2002, pp. 251–260.
- M. Castro et al., "Secure Routing for Structured Peerto-Peer Overlay Networks," ACM SIGOPS Operating Systems Rev., Winter 2002, pp. 299–314.
- H. Yu et al., "SybilGuard: Defending against Sybil Attacks via Social Networks," *IEEE/ACM Trans. Networking*, vol. 16, no. 3, 2008, pp. 576–589.
- P. Maniatis et al., "Preserving Peer Replicas by Rate-Limited Sampled Voting," ACM SIGOPS Operating Systems Rev., vol. 37, no. 5, 2003, pp. 44–59.
- W. Xu, F. Zhang, and S. Zhu, "Toward Worm Detection in Online Social Networks," *Proc. 26th Ann. Computer Security Applications Conf.* (ACSAC 10), ACM Press, 2010, pp. 11–20.
- N. FitzGerald, "New Facebook Worm Don't Click Da' Button Baby!" blog, 23 Nov. 2009, http://fitzgerald. blog.avg.com/2009/11/new-facebook-worm-dont-clickda-button-baby.html.
- C. Schmugar, "The Future of Social Networking Sites," McAfee Security J.: Security Vision from McAfee Labs, Fall 2008, pp. 28–30.

- **Hongyu Gao** is pursuing a PhD in computer science at Northwestern University. His research interests include networking and security, with emphasis on security problems in online social networks. Gao has a BS in computer science from Peking University. Contact him at hongyugao2013@u.northwestern.edu.
- Jun Hu is pursuing a PhD in computer science at Huazhong University of Science and Technology in Wuhan, China. She's also a visiting scholar in the Department of Electrical Engineering and Computer Science at Northwestern University. Her research interests include network security and network measurement. Hu has a BS in computer science and technology from Chongqing University of Posts and Telecommunications. Contact her at junehu1210@gmail.com.
- **Tuo Huang** is pursuing a JD at Yale Law School. He completed the work described in this project while he was a graduate student at Northwestern University. His research interests include legal and policy issues in intellectual property, Internet governance, and telecommunications. Huang has an MS in electrical and

computer engineering from Northwestern University. He's a member of Phi Beta Kappa. Contact him at tuo. huang@yale.edu.

- Jingnan Wang is pursuing a PhD in electrical engineering at Northwestern University. Her research interests include statistical image processing, shape analysis, and feature extraction. Wang has an MSc in telecommunications from Technische Universiteit Eindhoven. Contact her at jingnanwang@u.northwestern.edu.
- Yan Chen is an associate professor in the Department of Electrical Engineering and Computer Science at Northwestern University. His research interests include network security, measurement, and diagnosis for large-scale networks and distributed systems. Chen has a PhD in computer science from the University of California, Berkeley. He's a member of IEEE. Contact him at ychen@northwestern.edu.

Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.

