Exposing Impersonation Attacks in Online Social Networks

Oana Goga
MPI-SWS

Giridhari Venkatadri
MPI-SWS

Krishna P. Gummadi
MPI-SWS

1. MOTIVATION

Today, users sign on to most online social networking sites like Facebook, Twitter, and Google+ via *weak identities*, i.e., unverified identities that do not require users to prove that their online identities match their offline (real world) personalities. Weak identities leave the sites vulnerable to a variety of fake identity or *Sybil* attacks. In this paper, we focus on *identity impersonation attacks*, a special class of Sybil (fake identity) attacks where the attacker spoofs (assumes) the identity of another real-world user. As more and more personal data about users becomes publicly available on the Web, impersonation attacks become easier to carry out.

Identity impersonation attacks can be particularly damaging for the victim's online reputation. As people's online data is increasingly aggregated by search engines [1] and used for a variety of purposes including evaluating their suitability for employment [2], impersonation attacks, particularly those that go undetected, can have serious adverse consequences for the victims, even in the offline world. Further, social engineering attacks launched using impersonated identities can result in material and financial losses for the victim as well as the victim's friends [3].

Despite the serious threat posed by impersonation attacks [4, 5] and the woeful state of defenses against the attacks, few research studies, to date, have systematically studied impersonation attacks in online social networks. We argue that the current state of affairs is largely due to the difficulty in gathering extensive ground truth data about impersonation attacks in the real-world. Beyond a few anecdotal examples that are reported in the popular press, we lack large datasets about impersonation attacks in online social networks that can be used to characterize and detect identities participating in such attacks.

In this paper, we take the fist step toward understanding and detecting impersonation attacks by presenting a novel method that allows to gather extensive ground truth of impersonation attacks in current social networks. §2 gives the key idea behind our method and §3 presents some preliminary results on characterizing impersonation attacks.

2. KEY CHALLENGE: GATHERING GROUND TRUTH DATA

We approach the problem of gathering ground truth data of impersonating accounts in two steps:

1. Identify the accounts in a social network that *portray* the same person/entity. The key that enables this step is the recent findings by Goga [6] in ways to match accounts belonging to the same user across different social networks. We leverage these findings to build a rule-based matching scheme that detects whether two accounts portray the same person by emulating human judgment. Our scheme has a 65% true positive rate for a 2% false positive rate.

2. Identify which accounts are impersonating and which accounts are legitimate out of the accounts that portray the same entity. The hardest challenge of this step is that some social networks allow users to maintain multiple legitimate accounts. For example, users on Twitter are allowed to maintain different personal and professional accounts. To meet this challenge, we opportunistically use account suspension signals from Twitter and we exploit the observation that accounts managed by the same person frequently interact with each other, while impersonating accounts never interact with their victims as they do not want to be discovered. We use this method to gather ground truth about 4,054 cases of impersonation attacks on Twitter.

3. PRELIMINARY RESULTS

We use the ground truth data to do a preliminary characterization of impersonation attacks. Our analysis leads to some novel findings:

1. We observe that the higher the social neighborhood overlap between two accounts is (e.g., the accounts have common friends or followers, they retweet or mention the same users), the more likely it is that the accounts are managed by the same person. However, the more similar two accounts are in visual features the more likely it is that one account impersonates the other.

2. Most of the victims of impersonation attacks are ordinary Twitter users and not celebrities or important public figures. This finding reveals a new class of attacks that can impact negatively the online image of any user not just celebrities. We plan to extend our characterization and investigate why attackers target such wide range of users.

4. REFERENCES

[1] http://www.spokeo.com/.
[2] http://www.socialintel.com/.
[3] http://nairobiwire.com/2014/07/mike-sonko-arrested-swindling-public.html?utm_source=rss&utm_medium=rss&utm_campaign=mike-sonko-arrested-swindling-public.
[4] http://www.seattlepi.com/local/sound/article/Racism-and-Twitter-impersonation-prompt-lawsuit-893555.php.
[5] https://www.mediabistro.com/alltwitter/was-twitter-right-to-suspend-christopher-walken_b5021.
[6] O. Goga. *Matching User Accounts Across Online Social Networks: Methods and Applications*. PhD thesis, UPMC, 2014.