Poster: Social Spear-Phishing

Robin Gonzalez, Michael E. Locasto
University of Calgary

I. Introduction

One of the threats in online communication is its easy analysis by so-called social engineers. In this study, social engineers are *technologically savvy*[1] users who purposely misuse the data in social networks' profiles to further obtain sensitive information from their targeted victims. Note that in our definition the only requirements to become a social engineer are technological savvyness and a targeted victim.

There are many techniques or types of attacks social engineers can use to achieve their goals. In this investigation we focus on one type of attack called *phishing* where a social engineer communicates a deceitful message to their victims. This type of attack commonly seeks to obtain some confidential information from the victim with a communication that hides its true purpose. If the attack is designed to target a specific user with the knowledge of his or her information it is called *spear-phishing*. We study instead the types of attacks we can automatically construct using public information or data that users consciously agree to provide. We also discuss the power of social network conglomeration (i.e., social groups) to expand the scope of these attacks.

We design an Android app we call *Phish or Fish* that asks the user for basic permissions apps commonly ask for and, after the app gathers data that users allowed access to, it automatically constructs a targeted communication (comparable to that of *spear-phishing* attacks and/or *targeted advertisement* and asks the user to rate the credibility of the message in a scale from 1 to 10. The goal of our investigation is to study the feasibility of automatically generating targeted-attacks. For the purposes of this investigation we use phishing and spear-phishing interchangeably when it comes to the messages generated by our framework. Furthermore, we propose a set of experiments we plan to use for the evaluation of targeted attacks on social groups.

We understand some readers will find our approach controversial but the process we follow is no different from the one marketing companies use to target advertise users. We compare these two processes in further sections. For the purposes of this project, however, we do not have any online interaction with users but we rather use our app locally and with our own dataset. The results show that it is feasible to construct a framework that automatically generates messages that use a Facebook profile's public information as their context. Our work could be further extended to work with other social networks (e.g., Linkedin, Twitter, Google Plus) and social sites (e.g., online sites, blogs, forums). We also argue that other authors have obtained a high percentage of success because their attacks were targeted and not broad as phishing attacks commonly are.

[1] *Users who understand how basic online communication (e.g., wall posts, direct messages, likes, photos) in social networks works.*

II. Experiments

For this investigation we are going to measure the time and effectiveness of our framework's process of generating spear-phishing messages. Unfortunately, it will be hard to obtain all ethical permissions to run our experiments on multiple users in this term project. We will, therefore, use our own Facebook accounts for the purposes of collecting data and generating spear-phishing messages.

We are planning to evaluate the structure of the messages, the time it takes to generate them, and the different ways to qualify the messages. We also want to map users to social groups to see the difference in messages generated to several users in the same social group. We are thinking of creating fake Facebook profiles with random likes in all the themes we study to generate more cases in our study.

We have several ideas for different experiments in our study. We want to study the influence of *congeniality*, in the communication of a message between users, on trust or belief. Before we explain our experimental design we need to define what congeniality and trust mean in this context. We need to remember that, in our work, the communication sender-receiver hides its true purpose. In social engineering, the sender looks to obtain trust from the receiver in different ways. Our method consists of the collection and analysis of data from the user for the automatic construction of a message that shares the tastes and interests of the receiver. In other words, the sender aims to relate to the receiver by using publicly available information.

We measure trust at the end of each experiment as a boolean value that indicates whether a receiver believed the message to be benign or not. We send users a non-malicious URL in each message and we say trust is *true* whenever a user visits the URL. The URL links to a web site that thanks users for the participation in our studies and does not collect any sensitive data whatsoever. It only keeps tracks of whether a receiver trusts the communication or not.

A. Phishing social groups

A social group (sg) is a collection of users (a_i) who share one or more common attributes (t_i).

$$sg = a_0 ... a_n$$

$$a = t_0 + ... + t_n$$

A social group is determined based on any possible combination of the following attributes (t_i):

- **Basic info:** age, location, and languages they speak.
- **Popularity:** number of friends, number of events attended, and number of wall posts.
- **Community:** groups they belong to, most visited places (e.g., bars, coffee shops), and most visited locations (e.g., cities, towns).
- **Interests and likes:** music, movies, tv shows, and books the user likes in Facebook.

Therefore, all members of a social group share a set of attributes $t = \{t_0, ..., t_n\}$ where $n > 1$.

We divide users who use our application in different social groups according to their interests. One of our hypothesis is that we can attack each group with a single message that relates to the group. We posit that the attack will have high percentage of success or *trust* for every individual that belongs to the social group. We are planning to follow a static group comparison design for this experiment where the observations are the levels of trust for each group and the treatment the congeniality of the message. For this, we will generate 10 messages for each social group followed by a 5 question survey about the attack. The survey asks a set of questions about the validity and credibility of the message as well as what made them click or not on the URL. Each survey is linked to a unique user ID and saved in our repository for further studies.

We claim that congeniality influence trust or credibility and that, during the experiment, people who feel more related to the message will trust or believe it more. Since we will have different and numerous social groups we can compare the results of a group with others. The external validity of this experiment relies on the fact that we will send the message to all social groups at the same time. Since we are measuring *trust* we argue that the history of the users does not matter but rather the impact the message has at the point it was sent. In other words, we assume that no other outside elements besides tech savvyness influence in their decision. Moreover, the survey will ask the reasons why the users did not trust the message to eliminate any potential biases not related to the structure of the communication.

B. A phishing game

We are planning to design our application as a game for several reasons. First of all, it incentivizes its use in a world where having a successful application gets harder everyday. Secondly, it increases the interaction and communication between users and thus our population list. Thirdly, people tend to increase their learning capabilities with educational games[2]. One of the motivations in our research is to instruct users about the dangers in online communication and make them aware of the things they post in public social networks. A user who plays the phishing game would have to follow these steps:

1) Nominate one of your Facebook friends as your phishing victim.
2) Choose between sending an automatic phishing message constructed by the app or manually construct a message yourself that talks about your interests.
3) Send the message to your victim of choice.
4) The victim would have to guess if the message was sent by you or was automatically constructed by an algorithm.
5) If the victim guesses right he obtains a point, otherwise the user who sent the message gets a point.

The manually constructed message is saved into our database for further use as a different attack to a social group the user belongs to. Since the sender is also allowed to manually construct a message it will be harder to study the congeniality changes. We posit, however, that receivers will more often trust a message manually constructed by a friend rather than the automatically generated and thus it will have more congeniality. Therefore, our hypothesis is that manually constructed messages by a friend will be more trustworthy and relatable than the app's messages.

C. Pre and post security awareness

Our observation and treatment variables will be the same as before. This experiment will follow a Solomon four-group design where pre-experiment and post-experiment surveys will be used to increase the validity of the results. We assume that from the point when users replies to the first survey until the point they reply to the last survey their history is the same. In other words, the user does not have enough time to think differently about communication attacks in the timeframe between both surveys. We plan to educate users who participate in our studies using the post-experiment surveys. The set of questions we ask the user in the pre-experiment survey is the following:

1) I often reply to messages from strangers in social networks.
2) I often visit links if they interest me.
3) I spend more than two hours per day on social network activity.
4) One of my passions is either music, literature, films, or tv shows.

This experiment will also follow a static group comparison design to compare the results of all the respondents. Since we manipulate congeniliaty by including more or less interests (i.e., relating more or less to the receiver) we can observe how trust changes as a result. We can repeat the same experiment for all social groups and thus it also possesses external validity. On the other hand, the post-experiment (last) survey aims to study the impressions of users towards the communication attacks we send. Some of the questions we plan to include in this survey are the following:

1) The message was written by a human being.
2) The message looked too suspicious.
3) I didn't feel a need to click on the link.
4) I thought the link was an attack/virus.

Similarly to the pre-experiment surveys we link each answer to the respective user ID who answered the question. At the end of the study we calculate pre-knowledge and post-knowledge levels for every user in our study and they represent how likely they are to fall for the attack and it will also help us to automatically construct better and more suitable attacks.

[2]http://news.stanford.edu/news/2013/march/games-education-tool-030113.html