Poultry Markets: On the Underground Economy of Twitter Followers

Gianluca Stringhini, Manuel Egele, Christopher Kruegel, and Giovanni Vigna
Computer Security Lab, University of California, Santa Barbara
{gianluca, maeg, chris, vigna}@cs.ucsb.edu

ABSTRACT

Since Twitter has emerged as one of the easiest ways of reaching people, companies started using it to advertise their products. However, creating a functional network of followers to whom to promote content is not a straightforward task. On the one side, collecting followers requires time. On the other side, companies need to establish a reputation to lure users into following them.

A number of websites have emerged to help Twitter users create a large network of followers. These websites promise their subscribers to provide followers in exchange for a fee. In addition, they offer to spread their promotional messages in the network. In this paper, we study the phenomenon of these *Twitter Account Markets*, and we show how their services are often linked to abusive behavior and compromised Twitter profiles.

1. INTRODUCTION

Nowadays, Twitter is the most popular microblogging site on the Internet. The rapid growth that Twitter sustained over the recent past resulted in over 300 million subscribed users [3]. People use Twitter to stay in touch with their friends, as well as to get news from people or organizations that they find interesting [19]. Additionally, companies and celebrities started using Twitter to promote themselves and their brands [10, 18].

The same reasons that make Twitter very useful for legitimate users and businesses also attract malicious parties. Recent research showed how miscreants use Twitter to spread spam and malicious content that aims to infect their victims' computers [9, 14, 24, 25]. Attackers on Twitter can increase the impact of their malicious actions by compromising legitimate, influential accounts [13]. By gaining control over a legitimate account, miscreants can leverage the network of trust this account established in the past to spread malicious content more effectively [7]. Intuitively, this makes sense: users are more likely to click on links or re-share content posted by a familiar user they trust, rather than the content posted by a non-related account that contacted them randomly.

A Twitter account can be compromised by either stealing its credentials with a phishing attack [5], luring the user into authorizing a rogue *OAuth* application [2], or by leveraging vulnerabilities on the social networking site, such as Cross-Site Scripting (XSS) [9].

In this paper, we consider an account as compromised if a third party has obtained access to that account (e.g., through a phishing attack or by tricking the user into authorizing a rogue application) and uses this account in a way that violates Twitter's terms of service [4]. This includes, for example, posting unrelated updates on trending topics, or so-called mention spam where a large number of tweets mention users that have no relation with the account that sends the tweets.

Unfortunately, many of these compromised accounts end up under the control of so-called *Twitter Account Markets*. These markets use compromised accounts for two different purposes: *inflating followers* and sending *promoted tweets*. These markets provide services that simplify building up influence in the social network by artificially inflating the number of followers and offering spam-like advertising services so-called *promoted tweets*. However, the way these markets operate directly violates Twitter's terms of service. Twitter Account Markets make use of compromised accounts for two different purposes: *inflating followers* and sending *promoted tweets*.

Inflating followers. Attackers can use the victim accounts that they control to artificially increase the number of followers of any target Twitter account. By doing this, the victim will start seeing the tweets sent from the target account on her timeline. The timeline is the stream of tweets generated by all accounts that a user follows, combined with her own tweets. In addition, the increased number of followers makes the target account look more trustworthy and influential to other users, who will then be more likely to follow it.

Promoted tweets. The attacker can use the compromised account to send arbitrary tweets. This, of course, includes tweets containing malicious content or spam messages. Such tweets can be used to lure more users into getting their accounts compromised, or can be used to promote questionable websites and products.

The reason for the success of these markets is the following: one of the main challenges companies and individuals face to successfully promote their products is to reach their target audience with advertisement. Creating a legitimate

WOSN'12, August 17, 2012, Helsinki, Finland.

and functional network of followers for a Twitter profile requires substantial investments in terms of both time and effort. Therefore, many Twitter accounts decide to buy their followers, instead of growing their social circles over time.

Commonly, these Twitter Account Markets advertise their services as being legitimate and in line with all applicable rules. This tricks the victims into voluntarily handing over their account credentials (users might see this as a form of payment). However, the way in which these markets use these accounts after they obtained the credentials directly violates the Twitter terms of service [4].

In this paper, we study the threats posed by these *Twitter Account Markets*, and their effect on Twitter users. Our experiments focus on two campaigns that promote two different account markets. During our analysis, we identified 1,577 accounts (i.e., customers) that purchased followers on either of these Twitter Account Markets. Additionally, we also identified 1,041 spam tweets that were sent through the victim-accounts participating in these Twitter Account Markets. Twitter considers these markets a threat to the quality of their service, as they recently amended their terms of service with a clause that forbids their users to advertise such markets.

This paper makes the following contributions:

- We introduce the concept of Twitter Account Markets, and describe this phenomenon.
- We propose techniques to automatically identify customers of Twitter Account Markets.
- We discuss the threats associated with Twitter Account Markets, and discuss countermeasures that Twitter could take to mitigate the problem.

2. TWITTER ACCOUNT MARKETS

The services offered by a Twitter Account Market are typically accessible through a web page, similar to the one in Figure 1. The price for buying Twitter followers on these markets varies from about $20 to $100 for 1,000 followers. Furthermore, some services go as far as guaranteeing the "quality" of the advertised followers (e.g., each artificial follower has at least 500 followers itself). Intuitively, such "high-quality" followers are sold for more money than regular victims. As an alternative to buying Twitter followers, a customer can decide to purchase so-called *promoted tweets* (such tweets are not to be confused with promoted tweets that Twitter itself sells for advertising purposes). Promoted tweets are messages whose content is determined by the customer and are distributed by the victim accounts the service has control over. The price-range we identified for these promoted tweets is around $10 for 1,000 tweets.

All Twitter Account Markets that we have analyzed offer both "free" and a "premium" versions of their services. Premium customers pay a fee to get followed, or to spread content on Twitter. Free users, on the other hand, have to provide their account credentials to the market operators in exchange for followers. As these users of the "free" service give away their credentials to a third party, we also call them victims. At registration time, these victims are also commonly required to authorize a Twitter application under the control of the market operator. This application allows the market-operators to control the victim accounts through the Twitter API. Once registered, these victim accounts will

Figure 1: A Twitter Account Market

start following other victims to fulfill the promise of followers for other users of the "free" service. Simultaneously, victims will also follow a number of "premium" users who payed for the service.

As mentioned previously, the second stream of revenue for Twitter Account Markets results from the distribution of *promoted tweets* containing content determined by the paying market customers (typically, an advertisement containing a link to a website). Additionally, "free" users are periodically used to send tweets advertising the market itself, with the goal of luring more users into subscribing to the market.

Even though some account markets have specific terms of service that inform their users about the fact that their accounts could be used to follow third parties or send promotional tweets, such practices violate Twitter's terms of service [4]. As a result, Twitter shuts down any offending application that is used by Twitter Account Markets. However, since the market operators own the credentials to their victim accounts, they periodically create and authorize new applications, and keep performing their malicious activities.

Although the modus operandi of these markets is generally the same, they feature different pricing models and target different audiences. For instance, we identified some markets that target English-speaking users, and others that target a Portuguese-speaking audience (presumably Brazilian users, since all involved websites used by these markets are registered under `.br` top-level domains). This localization is important for the market customers: by having control over the predominant language for their target demographics, they can make sure that the promoted content will reach a suitable audience.

3. APPROACH

Our approach to analyzing Twitter Account Markets consists of three phases. First, we identify the victim accounts that gave away their credentials when they subscribed to the market for free. Second, we analyze the accounts these victims follow, which allows us to identify paying customers of the market. Third, we analyze the tweets that market

victims send after their account is compromised, to detect promoted tweets. In the following subsections, we describe these three phases in more detail.

3.1 Detecting Market Victims

As discussed in Section 2, market victims periodically send advertisement tweets. These tweets aim at attracting new users and luring them into subscribing to the service. Such tweets are easily recognisable, because they contain similar text, point to the same website, or contain the same hashtag (e.g., *#bigfollow*).

Given a collection of tweets, we identify Twitter Account Market victims as follows. First, we look at tweets that are similar. Note that simply grouping tweets that are identical does not work. The reason is that Twitter is known for having techniques in place to detect misuse. Consequently, to avoid being detected or blacklisted, miscreants slightly modify the content of the tweets or use different URLs (e.g., by leveraging URL shortening services). These minor modifications of their messages are commonly enough to avoid detection. Unfortunately, relying solely on hashtags to determine similarity is not sufficient either. The reason is that not all campaigns we observed made use of hashtags.

Thus, we developed a more robust measure to determine tweet similarity. More precisely, we use n-gram similarity to decide whether two tweets are similar, and, thus, belong to the same campaign. In particular, two tweets are considered similar if they share four consecutive words (words are consecutive characters separated by whitespaces). Note that this definition of word considers a URL as a word too. As we are looking into economic effects of Twitter Account Markets, we only consider markets advertised by at least 1,000 tweets. That is, we assume that the impact of smaller markets on the overall Twitter Account Markets is negligible.

After grouping tweets together based on the n-gram similarity measure, we manually look at the obtained groups, and we discard those that do not belong to campaigns promoting Twitter Account Markets. To assess this, we visit a randomly chosen URL among the ones advertised in each group. If the site pointed to by the URL is not a Twitter Account Market website, we discard the group. We also discard those groups that do not contain any URL in their tweets. We note that automatic identification of Twitter Account Market websites is probably feasible. However, such techniques are outside the scope of this paper.

3.2 Detecting Market Customers

After we have obtained a number of victims that participate in a Twitter Account Market, we want to identify a set of accounts that bought customers from the market.

As discussed in Section 2, the typical customer of a Twitter Account Market is an account that aims to promote goods, contents, or services, but that does not have an established network of contacts yet. It is possible to detect a market customer because, unlike other newly created accounts, a market customer has a large number of followers (i.e., the ones purchased through the market). Therefore, we consider a Twitter account A as being a customer who bought followers through a Twitter Account Market if:

- A is followed by at least t_v other accounts that we previously identified as victims of Twitter Account Markets. The rationale behind this threshold is that market customers are followed by many "free" subscribers (i.e., victims) of the market.
- A follows less than t_f other Twitter accounts. This threshold discards accounts that periodically follow a set of users, wait for a number of them to follow back, and unfollow them. This is a common behavior for accounts looking for followers (e.g., spammers) [28], but is not indicative on its own of being a Twitter Account Market customer.
- The ratio of friends to followers of A is lower than t_r. In Twitter terminology, friends are those accounts that A follows. This threshold is useful to discard those accounts that do not have many followers. Since Twitter Account Markets sell followers in batches of 3,000, such accounts are unlikely to be market customers.
- The influence of A on Twitter is below the threshold t_i. We define influence as a measure that indicates how engaged an account is on Twitter. This includes the number of followers the account has, but also the number of retweets, and the number of mentions. This threshold allows us to discard those accounts that belong to celebrities. Typically, such accounts have a very unbalanced friends to followers ratio. However, they also have high influence scores, because people interact with them a lot and share their content with their followers.
- The Twitter timeline of A does not contain any tweet indicative of A being a free subscriber (i.e., victim) to the market.

To assess the influence of an account on Twitter we use *Klout* [1]. Klout is a service that calculates how influential an account is, based on a number of features such as the probability that the content posted by the account is reshared, or the number of people that can be reached by the messages the account sends. We use Klout because it conveniently exposes its influence measure through a publicly-available API. However, any influence measure that gives the same type of information as Klout would work in this context.

By setting the thresholds t_v, t_f, t_r, and t_i accordingly, we can make sure to detect those accounts that have an anomalous number of followers, compared to other accounts with similar influence.

For our experiments, we set t_v to 50, t_f to 500, t_r to 0.3, and t_i to 20. We empirically determined these thresholds based on small scale-experiments. Furthermore, these experiments indicate that minor variations to these threshold values do not affect the overall results. If an account exceeds all four thresholds, we consider it as a Twitter Account Market customer.

3.3 Detecting Promoted Tweets

The second way Twitter Account Market victims are misused is for sending out tweets provided by market customers. These tweets usually promote content about the customers' business. We call such tweets *promoted tweets*. Unlike Twitter's legitimate promoted tweets, which are purchased by companies and appear in search results on the Twitter website, these tweets are sent by compromised profiles, and usually link to web sites with questionable content (i.e., spam sites).

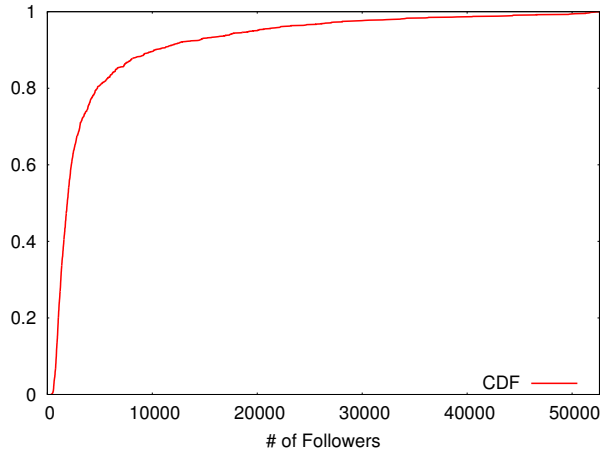

Figure 2: Cumulative Distribution Function of the followers of the customers of the two Twitter Account Markets

To detect promoted tweets, we proceed as follows: for each account in a group of Twitter Account Market victims, we first retrieve the 200 most recent tweets sent by that account. Subsequently, we look for similar tweets in the set of all these downloaded tweets. As before, we use use n-grams to identify groups of similar tweets. The rationale for searching for these groups is that we expect the Account Market operators to use their victim accounts to post promoted tweets.

For each group of similar tweets obtained in this way, we consider it as consisting of promoted tweets if:

- None of the tweets in the group passes the n-gram similarity with the tweets that are periodically sent to advertise that same Twitter Account Market. In other words, we are not interested in tweets that promote the market itself. Additionally, we discard groups that contain tweets that advertise alternate Account Markets. The reason for this is that many users subscribe to the free offerings from multiple Account Markets.

- Similar tweets have been sent by at least t_p other accounts in the victim group. The rationale behind this is that the Twitter Account Market administrators will likely have several of their victims sending out the same promoted tweet.

4. EVALUATION

We analyzed a tweet dataset composed of 1.4 billion tweets. These tweets have been collected from Twitter's Gardenhose stream in the three month period between May 13, 2011 and August 12, 2011. This corresponds to a random sample of 10% of the overall activity generated on Twitter during that time.

We identified two large scale campaigns advertising Twitter Account Markets in this dataset. Additionally, we identified two smaller campaigns, which we discarded, as discussed in Section 3.1.

As we discussed in Section 2, the first market we analyzed was advertised by tweets in English, while the second one was advertised by tweets in Portuguese. In total, 39,004 accounts were sending tweets promoting these two account markets. In particular, we detected 22,003 victims for the English-speaking market and 17,001 victims for the Portuguese-speaking one. We consider these victim accounts as our baseline for finding customers who bought followers on a market, as well as promoted tweets.

4.1 Analysis of Market Customers

By applying the techniques described in Section 3.2, we identified 1,577 accounts that bought followers on a Twitter Account Market. By manually analyzing these customer accounts, we found many instances of accounts that were promoting their own business. Examples include accounts linking to online pharmacies, but also Search Engine Optimization (SEO) and marketing companies.

Figure 2 shows the cumulative distribution function (CDF) of the number of followers for the market customers we identified. Both of the markets we analyzed sell followers in multiples of 3,000. However, 70% of the identified customers had less than 3,000 followers. Thus, even if they purchased followers, many of these customers are followed by less than 3,000 accounts. We identified two explanations for this discrepancy. First, Twitter Account Markets typically state that it will take some time until the full number of followers is reached. Presumably, to avoid detection, artificial followers are added gradually and this can take up to 30 days. Thus, many of the market customers might not have reached their full advertised amount of followers by the time we detected them. Another reason could be that some victims get annoyed with the market customers content they see on their timeline and therefore unfollow these accounts. In any case, Figure 2 shows that it rarely happens that market customers buy more than one batch of followers.

The cheaper price point for 3,000 followers we observed on these two markets is $65. Thus, the two markets in our analysis earned at least 1,577 customers times $65 (i.e., $102,505). This estimate is conservative because, as mentioned above, we only received 10% of all tweets sent to the Twitter platform during the observation period. Thus, the number of victims not present in our dataset might be substantial.

4.2 Analysis of Promoted Tweets

We also analyzed the collected data for evidence of promoted tweets. To this end, we set the parameter t_p to the value of two. That is, we consider a tweet promoted, if two distinct victims post a similar tweet. Applying the technique detailed in Section 3.3, we identified 29 campaigns consisting of 1,041 individual messages total. The largest single campaign consisted of 60 tweets. Manually inspecting these campaigns showed that most promoted tweets were used to advertise "free money" opportunities, "click bank" websites, or coupon programs. Of course, we would expect to identify larger campaigns of promoted tweets. However, as stated above, a significant number of victims might not be present in our dataset because of the limited coverage we had while collecting the data.

5. DISCUSSION

Twitter Account Markets are considered a problem by Twitter. This phenomenon is so wide spread that the social network added a specific clause to their terms of service forbidding to send tweets that advertise Twitter Account

Markets [4]. The concerns by Twitter are understandable: on the one hand, Twitter Account Markets inflate the number of followers of their users (customers), exposing market victims to unwanted tweets and generating a false sense of trustworthiness around the accounts that bought followers. On the other hand, they spread spam through Twitter, by making their victims send promoted tweets.

As stated in Section 2, Twitter promptly suspends any OAuth application involved in Twitter Account Market operations. However, Twitter does not take any countermeasure against the activities of market customers. Even though mitigating the effects Twitter Account Markets have on their victims helps reducing the problem, we assume that taking countermeasures against market customers would be beneficial too. For example, Twitter could apply techniques similar to the ones presented in this paper to detect Twitter Account Market customers and promoted tweets. As a countermeasure, they could suspend the market customer's account, and delete promoted tweets, thus severely limiting the stream of revenue for Twitter Account Market operators.

A limitation of our approach is that it is focused on detecting newly-created customer accounts. This avoids false positives that would arise if non-misbehaving accounts are detected. However, such false positives could also be avoided by leveraging additional information, (e.g., when did victims start following an account). Unfortunately, this information is only available to Twitter and therefore, we were unable to use it in our approach.

6. RELATED WORK

Twitter, and online social networks in general, have attracted the interest of researchers over the last few years [10, 19, 26]. In particular, a wealth of research has been conducted to discuss the threats associated with online social networks [13,14]. Results show that spam and phishing campaigns performed through social networks have a higher success rate compared to traditional email campaigns [7, 15].

Researchers have developed a number of systems to mitigate threats on social networks. Several systems aim to detect fake profiles that spread malicious content on social networks [6, 8, 12, 20, 23, 24, 28]. Other systems look at suspicious URLs in social network messages [21, 25], or try to detect distributed threats such as a worm outbreak [9, 27]. However, one of the most important threats social networks are facing nowadays is legitimate accounts that get compromised and misused [13]. Reliably detecting such accounts is an open research problem.

A number of papers that analyze the underground economy behind cyber-crime have been published [11, 16, 17, 22]. However, we are the first to identify and analyze the threats that arise from Twitter Account Markets, which treat Twitter accounts as commodities, and leverage these accounts for their abusive operations.

7. CONCLUSIONS

In this paper, we introduced the concept of Twitter Account Markets, and analyzed how these markets work. We developed techniques to automatically detect the accounts involved in the activity of such markets, as well as the spam tweets that are generated by them. Although this is just a first step, we hope that the insights provided by this paper will help to fight this phenomenon, and make Twitter a safer place.

8. REFERENCES

[1] klout. http://klout.com.

[2] Oauth community site. http://oauth.net.

[3] Twitter blog: #numbers. http://blog.twitter.com/2011/03/numbers.html.

[4] Twitter terms of service. http://support.twitter.com/articles/18311-the-twitter-rules.

[5] Twitter finally released a "stalkers" app? no, it's a phishing scam. http://nakedsecurity.sophos.com/2011/08/12/twitter-finally-released-a-stalkers-app-no-its-a-phishing-scam/, 2011.

[6] F. Benvenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting Spammers on Twitter. In *Conference on Email and Anti-Spam (CEAS)*, 2010.

[7] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks. In *Wold Wide Web Conference (WWW)*, 2009.

[8] Z. Cai and C. Jermaine. The latent community model for detecting sybils in social networks. In *Symposium on Network and Distributed System Security (NDSS)*, 2012.

[9] Y. Cao, V. Yegneswaran, P. Possas, and Y. Chen. Pathcutter: Severing the self-propagation path of xss javascript worms in social web networks. In *Symposium on Network and Distributed System Security (NDSS)*, 2012.

[10] M. Cha, H. Haddadi, F. Benvenuto, and K. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *International AAAI Conference on Weblogs and Social Media*, 2010.

[11] J. Franklin, V. Paxson, A. Perrig, and S. Savage. An inquiry into the nature and causes of the wealth of internet miscreants. In *ACM Conference on Computer and Communications Security (CCS)*, 2007.

[12] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary. Towards online spam filtering in social networks. In *Symposium on Network and Distributed System Security (NDSS)*, 2012.

[13] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao. Detecting and Characterizing Social Spam Campaigns. In *Internet Measurement Conference (IMC)*, 2010.

[14] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: the underground on 140 characters or less. In *ACM Conference on Computer and Communications Security (CCS)*, 2010.

[15] T. Jagatic, N. Johnson, M. Jakobsson, and T. Jagatif. Social phishing. *Comm. ACM*, 50(10):94–100, 2007.

[16] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage. Spamalytics: An empirical analysis of spam marketing conversion. In *ACM Conference on Computer and Communications Security (CCS)*, 2008.

[17] C. Kanich, N. Weaver, D. McCoy, T. Halvorson, C. Kreibich, K. Levchenko, V. Paxson, G. Voelker, and S. Savage. Show Me the Money: Characterizing Spam-advertised Revenue. *USENIX Security Symposium*, 2011.

[18] R. King. How companies use Twitter to bolster their brands. In *BusinessWeek Online*, 2008.

[19] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Wold Wide Web Conference (WWW)*, 2010.

[20] K. Lee, J. Caverlee, and S. Webb. Uncovering social spammers: social honeypots + machine learning. In

International ACM SIGIR Conference on Research and Development in Information Retrieval, 2010.

[21] S. Lee and J. Kim. Warningbird: Detecting suspicious urls in twitter stream. In *Symposium on Network and Distributed System Security (NDSS)*, 2012.

[22] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Félegyházi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, and Others. Click Trajectories: End-to-End Analysis of the Spam Value Chain. In *IEEE Symposium on Security and Privacy*, 2011.

[23] J. Song, S. Lee, and J. Kim. Spam filtering in twitter using sender-receiver relationship. In *Symposium on Recent Advances in Intrusion Detection (RAID)*, 2011.

[24] G. Stringhini, C. Kruegel, and G. Vigna. Detecting Spammers on Social Networks. In *Annual Computer Security Applications Conference (ACSAC)*, 2010.

[25] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song. Design and Evaluation of a Real-Time URL Spam Filtering Service. In *IEEE Symposium on Security and Privacy*, 2011.

[26] C. Wilson, B. Boe, A. Sala, K. Puttaswamy, and B. Zhao. User Interactions in Social Networks and Their Implications. In *ACM European conference on Computer systems (EuroSys)*, 2010.

[27] W. Xu, F. Zhang, and S. Zhu. Toward worm detection in online social networks. In *Annual Computer Security Applications Conference (ACSAC)*, 2010.

[28] C. Yang, R. Harkreader, and G. Gu. Die Free or Live Hard? Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers. In *Symposium on Recent Advances in Intrusion Detection (RAID)*, 2011.