Ghostbusting Facebook: Detecting and Characterizing Phantom Profiles in Online Social Gaming Applications

Atif Nazir, Saqib Raza, Chen-Nee Chuah, Burkhard Schipper
University of California–Davis
Davis CA

Abstract

A fundamental question when studying underlying friendship and interaction graphs of Online Social Networks (OSNs) is how closely these graphs mirror real-world associations. The presence of *phantom* or *fake* profiles dilutes the integrity of this correspondence. This paper looks at the presence of phantom profiles in the context of social gaming, i.e., profiles created with the purpose of gaining a strategic advantage in social games. Through a measurement-based study of a Facebook gaming application that is known to attract phantom profiles, we show statistical differences among a subset of features associated with genuine and phantom profiles. We then use supervised learning to classify phantom profiles. Our work represents a first-step towards solving the more general problem of detecting fake/phantom identities in OSNs.

1. Introduction

Over the last five years, Online Social Networks (OSNs) such as Facebook, MySpace, and Orkut have carved out a massive following of nearly 700 million users worldwide. The core reason for the success of OSNs is the frequency of fresh user-generated content and personal information (e.g., status updates, pictures and comments). As users have become accustomed to expect fresh content almost every single time they access the OSNs, social networking sites tend to enjoy a high level of engagement. This is especially true for Facebook, where 50% of its total user base returns every day[1].

User-generated content, although frequently updated, might not be as high quality or engaging as, say, creatively designed gaming content. Facebook's pioneering Platform paved the way for massive, viral distribution of online games. The Facebook Platform currently hosts 95,000+ applications for its over-400 million active users. The Facebook Platform breathed life into the virtual goods economy through social games especially in the US, which has now become more than a billion dollar a year market, and continues to grow at breakneck speeds. Open-API platforms such as Facebook's are becoming an increasingly large contributor to Internet traffic with social games exhibiting higher quality (in terms of user engagement and graphical content) with each passing month, especially as big players such as Electronic Arts enter the social gaming space[2].

Existing research on OSNs is mostly focussed on graph theoretic properties of social networks, user behavior and traffic patterns of OSNs and their applications. Given the growing contribution of social games to Internet traffic and user engagement in OSNs, it is important to consider the general trend in online user engagement, the approaches taken to achieve this engagement, and its effects on the structure of the underlying social graph. In this spirit, we focus on a previously unidentified issue with social games: the creation of *phantom* or *fake* user profiles on OSNs to target higher rewards in social games.

1.1 *What are Phantom User Profiles?*

User profiles on OSNs can be created to suit different purposes. Consider, for example, a user who creates separate profiles for work, friends and family, dating and for playing social games. She may also decide to abandon one profile (without disabling it) and create another profile for the same purpose. Furthermore, she might wish to indulge in activities that are untraceable to her real identity. OSNs such as Facebook prefer creation of one user profile per real user, and provide tools such as content filter lists to accommodate users that may potentially wish to separate groups of acquaintances.

In this paper, we consider the specific problem of users creating multiple *phantom gaming* profiles on OSNs in order to achieve a strategic advantage within social games. Our experience shows that the multiple phantom profiles

[1] http://www.facebook.com/press/info.php?statistics

[2] http://www.virtualgoodsnews.com/2010/01/quick-stat-us-virtual-goods-market-to-hit-16b-in-2010.html

are used by one or more gamers to achieve higher status within the social game. As an example, consider that social games tend to reward users for introducing friends to the game to achieve growth. Using these phantom identities, users fool the games' point-based systems into believing they are contributing to the growth of the application (in number of users), and are thus rewarded with higher privilege (i.e., points, higher rankings on 'Halls of Fame', unlocked items, new missions, etc.). Furthermore, to give the phantom identity the appearance of a real user (in order to avoid detection by the OSN administrators), the phantom identity is advertised to (but not restricted to) the creator's friends, resulting in a high number of OSN friendships, or trust relationships, being formed with the phantom user profile. This clearly distorts the underlying social graph, and contaminates user experience on the social game itself.

We identify three conditions relevant to OSNs themselves that would motivate the creation of phantom profiles. First, for every set of randomly drawn users, if there exists a user who is strictly better off if some additional users join the interaction, then the user will have incentive to add/invite new participants. This 'word-of-mouth' recruiting property is the essence of a successfully growing OSN and social games. The second condition is the low (or zero) cost for participation, e.g., for joining an OSN or subscribing to an OSN-based game. Lastly, the user identities are non-verifiable since most OSNs do not verify identities of users out of privacy concerns. Private information required during the sign-up phase is not shared publicly with other users.

1.2 *Motivation and Contributions*

We present a case study using almost **two years** of data from one of our popular social gaming applications on Facebook, Fighters' Club (FC) [11]. Similar to any multiplayer game, players in FC are rewarded/penalized as function of interactions with other players in the game. This characteristic of social games in general tends to motivate users to create phantom profiles in OSNs.

The success of social gaming, as well as online social networking, is reliant on the integrity of the underlying social graph. In order to preserve this integrity, as well as the intended user experience on social games, it is imperative that phantom user profile detection and elimination is considered seriously. Facebook's current efforts at eliminating phantom user profiles rely on *manual examination* of individual user profiles, if they are at all reported by other OSN users. This paper, however, performs empirical characterization of phantom profiles and proposes methods to detect and potentially eliminate phantom user profiles used in social games. Our contributions are summarized as follows:

- We document our experiences with a real Facebook gaming application, Fighters' Club (FC), that is known to motivate phantom profile creations.
- We perform a measurement-based characterization of phantom and genuine user profiles interacting on FC and provide evidence of differences in some of the features recorded (Section 4).
- We explore how a learning technique using Support Vector Machines (SVMs) can be applied to classify phantom profiles.

The rest of the paper is structured as follows. We discuss related work in Section 2 and our methodology in Section 3. Section 4 presents the results of our measurement-based characterization of phantom and genuine user profiles. Section 5 discusses the results of applying SVM to classify phantom user profiles and Section 6 describes potential future work.

2. Related Work

Online social networking has garnered much academic interest recently [7]. Researchers have focused on various aspects of OSNs, including graph theoretic properties of social networks, usage patterns of individual OSNs [10, 3, 14], aggregated activity data from multiple OSNs [1, 5], as well as privacy and security issues [8]. We have previously studied network-level delays and user interaction on different types of social applications through home-grown social applications [11, 12]. A recent study [13] utilizes large ISP traces across two continents to study user interactions with various OSNs by examining actual user clickstreams.

Yu et. al's work on Sybil attacks in [15] is the most relevant to this paper. A Sybil attack is defined as a "malicious user obtaining multiple fake identities and pretending to be multiple, distinct nodes in the system." [15] used the insight that fake identities (in the case of Sybil attacks) are inherently less trusted than normal users, i.e., the number of trust relationships formed between fake identities might be high, but the number of trust relationships between fake identities and normal users must be low. This observation was used to develop an algorithm for identification of fake identities in Sybil attacks.

The problem introduced in this paper considers phantom identities formed by real users that wish to use the phantom profiles to achieve greater strategic advantage in a social game. Once created, the phantom identity is often advertised to one's real friends, who similarly wish to advance on the social game through the phantom profile. Moreover, since point-based social games require users target friends (because the application must spread in a social medium), phantom profiles can form a high number of trust relationships with genuine users and their friends. The assumption of lack of trust relationships in

Sybil attacks, and hence the solution developed in [15] is thus inapplicable in our scenario.

To the best of our knowledge, this paper is the first attempt to characterize and classify phantom profiles through online social games.

3. Measurement Methodology

We suspect that activities of fake/phantom users are inherently different from genuine users. Hence, a detailed empirical study of a large OSN user population that contains both genuine and phantom profiles can help reveal the distinct nature of the latter. An ideal approach is to analyze complete OSN-resident user data, such as user-entered data, uploaded photographs, activity patterns within the OSN, etc. While OSNs such as Facebook certainly do have this complete data, it is not publicly available. We adopted an alternate approach: by exploiting Facebook Developer Platform and launching home-grown social applications, we collected users' activity data within applications as well as limited (anonymized) OSN-specific user information. Our previous research utilized such data from eight highly popular Facebook applications that have collectively reached over 18 million users on Facebook to date [11, 12]. We analyzed social utility and social gaming applications. As discussed in Section 1.1, the incentives for creation of phantom user profiles exist mainly in social gaming applications. We launched two social games on Facebook, of which Fighters' Club (peak at 140,000+ daily users), became vastly more popular. We therefore employ Fighters' Club to gather relevant data for the case study in this paper.

3.1 Case Study: Fighters' Club

Fighters' Club (FC) was launched on Jun 19, 2007 on the Facebook Developer Platform. One of the first games to launch on Facebook, FC has been played by 6 million users on Facebook to date. The game's success was due to the addictive gameplay it fosters among friends and friends-of-friends. Specifically, FC enables users to pick virtual fights with their Facebook friends. These fights can last from 15 to 48 hours, during which time each player may request support from their Facebook friends, who then help the requesting individual's team defeat the opposing user's team through a series of virtual "hits" to decrease the strength of the target opponents. The team with the higher cumulative strength at the end of the fight is declared the winner. Players in a single fight on FC belong to one of the following three roles:

Offender: The user instigating the fight by choosing a friend to fight against and selecting a fight duration.

Defender: The Facebook friend fought by the offender.

Supporter: The offender and defender may advertise the fight to their Facebook friends, who then pick one side and support the chosen user's team.

Supporters may withdraw support or change sides until the last 2 hours of the fight. FC players accumulate virtual money, as well as 'street cred' points (which determine the 'strength' of the player). When the offender wins, they gain a percentage of the defender's strength (and vice versa). The winner (offender/defender) is awarded X money per opposing team member. Supporters gain/lose a fifth of the strength the Supportee gained/lost, but do not gain or lose money from the fight. FC also incorporates a leveling system (six levels) to reward user loyalty. Users need to win Y number of fights (as offender/defender) to advance to the next level in the game. Each higher level grants users higher upper-bound on strength, or 'street cred' points.

Data Gathered: Third-party social applications such as FC allow us to gather and store an impressive amount of user data including their IPs, browser information, etc. in addition to user activity data (i.e., who picked a fight with whom, who hit whom, time of fight, time of hit, time of support, etc.) from the game itself. The case study presented here relies mainly on user activity data from the game itself, along with some anonymized Facebook-resident data. Table 1 summarizes the data set used in our study. We manually contacted a subset of users[3] to verify the existence of 545 phantom and 520 genuine user profiles (which we used as ground truth). We extracted the following attributes per profile:

1. FNJ: No. of Facebook Networks Joined
2. FW: No. of Posts on Facebook Wall visible to FC
3. FFC: No. of Friends using FC
4. FB: No. of Facebook Friends
5. ATFP: Average Time to Fight Participation
6. AFPs: Average No. of Fights Picked/sec
7. AFDur: Average Fight Duration
8. TFP: Total No. of Fights Picked
9. ASPF: Average No. of Supporters in Picked Fights
10. AOPF: Average No. of Opponents in Picked Fights
11. TFD: Total No. of Fights Defended
12. ASDF: Average No. of Supporters in Defended Fights
13. AODF: Average No. of Opponents in Defended Fights

4. Characterizing Phantom User Profiles

In this section, we perform empirical characterization of the various attributes described in Section 3.1 with the goal of identifying distinctive traits that can help differentiate phantom from genuine profiles. For each attribute, we computed and compared the conditional cu-

[3]Most phantom users were verified after reports by FC players, while all genuine users were randomly selected and verified.

Table 1: Data Set Analyzed

# of Fights	2,532,779
# of Support Requests	80,174,483
# of Unique Users	264,606
# of Unique Installed Users	30,990
# of Known Phantom Users	545
# of Known Genuine Users	520
# of Fights Instigated by Genuine Users	70,209
# of Fights Targeting Genuine Users	61,751
# of Fights Instigated by Phantom Users	105,704
# of Fights Targeting Phantom Users	341,389

mulative distributions (CDFs) associated with genuine and phantom profiles, respectively. Due to space constraints, we only show the empirical CDFs for a subset of the attributes in Figure 1.

We hinted at some of the characteristic features of phantom user profiles on OSNs in Section 1.1. Considering OSN-specific characteristics of phantom users, phantom profiles have the incentive to "blend in" by forming trust relationships with other genuine OSN users. As shown in Figures 1(a) and 1(b), the empirical distribution of the number of Facebook friends (FB) and number of friends playing FC (FFC) for the known phantom and genuine user profiles show virtually no difference. Similar observations were made for other OSN-specific features, such as number of Facebook networks joined (FNJ) and number of posts on a user's Facebook wall (FW). This observation that phantom profiles *form similar trust relationships* as genuine user profiles differentiates our problem from the Sybil attacks [15].

On the other hand, we found that phantom user profiles tend to exhibit different characteristics (compared to genuine profiles) on game-specific features of FC. For instance, the FC point-system and rewards are centered around winning or losing fights. Since the FC point-system encourages competition among users, enthusiastic FC players tend to use phantom profiles as a tool to follow one or multiple of the following strategies (which help explains some of the observed deviation of phantom user activities from the activities of genuine users).

Strategy 1: Fight weaker players to win fights and gain higher status in FC. Genuine users focus on achieving higher rank in FC to compete with friends. Phantom profiles, on the other hand, are not real entities and hence are not focussed on achieving higher rank solely for competition. They instead are used by genuine users as weaker opponents that are easier to defeat, and that do not carry any social cost as opponents. As a result, phantom users tend to pick, as well as defend in, slightly more fights than genuine users. This is shown in the CDF plots in Figures 1(c) and 1(d) for instigated and defended fights (TFP and TFD), respectively. Furthermore, phantom users tend to serve multiple genuine users as weaker opponents, and hence instigate slightly more fights in their lifetime of activity (AFPs) than do genuine users (see Figure 1(e)). Figure 1(f) shows they also experience a slightly larger number of opponents on average (AOPF) in fights.

Strategy 2: Support stronger players to accumulate more FC rewards. Genuine users wish to support stronger players to increase chances of victory to earn higher rewards. The large number of opponents (AOPF) to phantom users seen in Figure 1(f) is a result of this strategy as well. In contrast, phantom users experience smaller number of supporting users (ASDF) in fights as compared to genuine users (Figure 1(g)). This further reflects that the purpose of phantom profiles is to serve as weaker players in the game for the benefit of genuine players.

Strategy 3: Instigate or defend in fights of smaller duration. As described previously, FC fights last from between 15 to 48 hours. The large duration gap was introduced to accommodate for user response times in OSNs. Our analysis of the average duration of fights (AFDur) shows that phantom users tend to opt for smaller fight durations, as shown in Figure 1(h). Since there is a limit to the number of active fights one user may instigate or defend in simultaneously, phantom users choose smaller fight durations to increase their benefit to genuine users.

Strategy 4: Participate later in fights to improve chances of team victory. The structure of FC gameplay is geared towards fostering higher engagement in users. One major source of engagement is the ability to participate in fights towards the end of fights' durations. Specifically, the later a user participates in a fight, the higher the chances for their team's victory. Since phantom users have less incentive to win fights, phantom users tend to participate in fights regardless of the remaining duration of the fight. Figure 1(i) shows the resulting CDF plot for phantom and genuine user profiles, which hints at a slightly relaxed participation time for phantom user profiles.

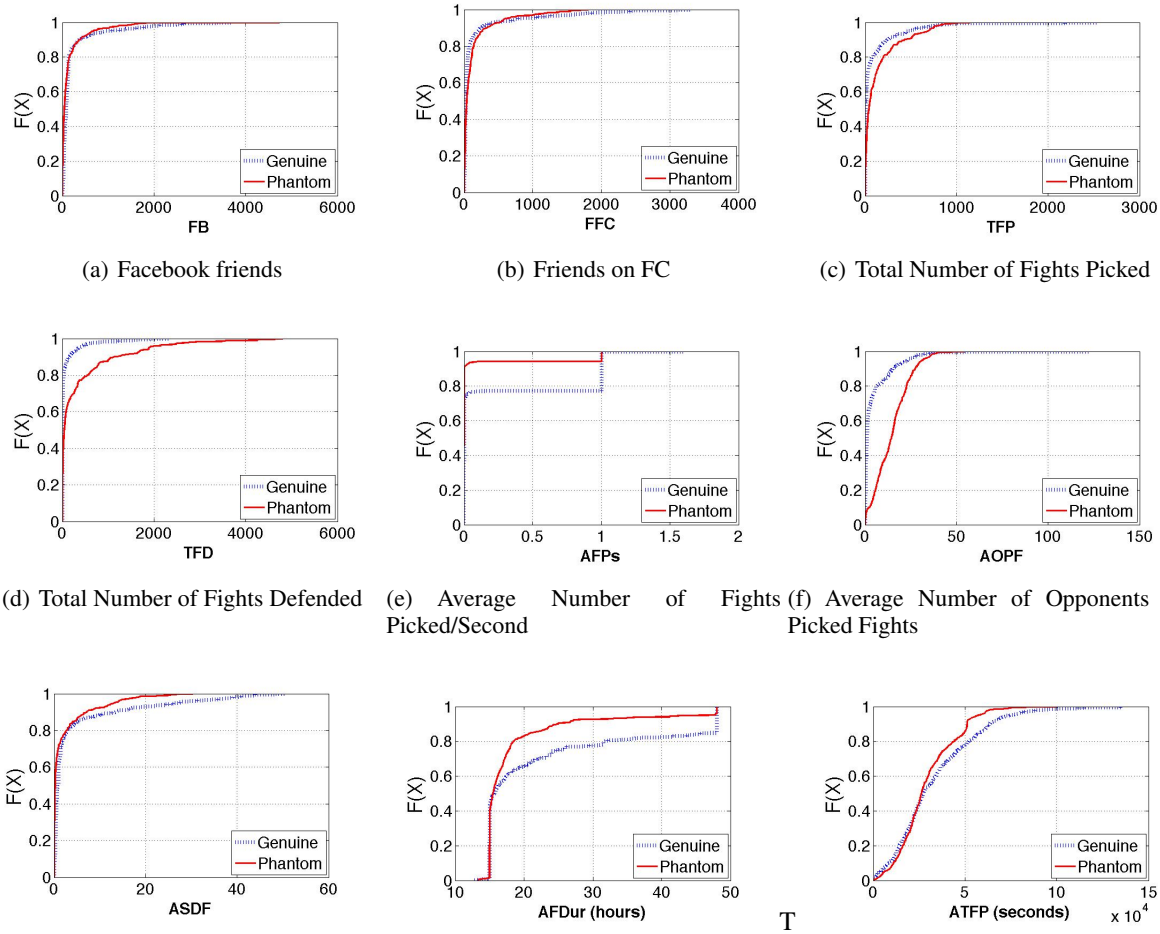
To explore if any of the attributes mentioned above can provide a basis for accurate discrimination between phantom and genuine user profiles, we use 'Kullback-Leibler (K-L) Divergence' metric. This metric measures the difference between the conditional distributions of an attribute associated with two classes, phantom vs. genuine profiles, as denoted by P and G, respectively.

The K-L divergence of G from P is defined as:

$$D_{PG}(P\|G) = \sum_i P(i) log \frac{P(i)}{G(i)} \quad (1)$$

D_{PG}, however, is asymmetric. Instead, we present our results in terms of symmetrized divergence metric, sD_{PG}, which is simply defined as

$$sD_{PG} = D_{PG} + D_{GP} \quad (2)$$



(a) Facebook friends (b) Friends on FC (c) Total Number of Fights Picked

(d) Total Number of Fights Defended (e) Average Number of Fights Picked/Second (f) Average Number of Opponents in Picked Fights

(g) Average Number of Supporteds in Defended Fights (h) Average Fight Duration (i) Average Time Between Fights

T

Figure 1: Empirical Distributions for Selected Attributes for Genuine vs. Phantom Profiles

Table 2 reports the sD_{PG} for each of the 13 attributes that we analyzed. Although a few of our attributes' K-L Divergence values stand out, e.g., in the case of average time to fight participation (ATFP), the apparent large divergence in distributions is due to the large variance (noise) and does not indicate meaningful differentiation between phantom and genuine user profile distributions. We also analyzed the Entropy E for the various attributes of phantom (P) and genuine (G) user profiles, as well as the standard deviation among values for the phantom and genuine users' data sets. Unfortunately, for attributes that exhibit high sD_{PG}, its entropy and variance are also very high. Hence, no individual attribute qualifies as a distinct metric for discrimination between phantom and genuine user profiles.

In the next section, we explore how more elaborate learning techniques may be applied on *multiple* attributes to classify phantom profiles.

5. Classification of Phantom User Profiles

We now turn our attention to using the attributes (features) detailed in Section 4 to distinguish between phantom and genuine user profiles. We employ a supervised learning method that generalizes information about user profiles known to be phantom or genuine, and use it to determine the authenticity of other profiles. We focus on Support Vector Machines (SVMs) for this classification.

We use the following standard statistical metrics [4] to characterize the performance of our classifier:

- *Accuracy* (ACC): The fraction of correctly classified profiles.
- *Positive Predictive Value* (PPV): The ratio of correctly classified phantom profiles to the total number of profiles classified as phantom.
- *Negative Predictive Value* (NPV): The ratio of correctly classified genuine profiles to the total number

	Attribute	sD_{PG}	$E(P)$	$E(G)$	$E(P+G)$	Std. Dev (P)	Std. Dev (G)	Std. Dev ($P+G$)
1	FNJ	0.4595	0.0492	0.5009	0.3119	0.0741	0.3164	0.2329
2	FW	0.7497	0.4620	1.0462	0.8146	4.3201	7.0376	6.0537
3	FFC	8.0768	6.9674	6.6666	7.1439	260.3930	413.1073	343.4374
4	FB	10.1985	7.1288	7.5132	7.7262	334.5041	459.8524	401.4987
5	ATFP	17.9409	8.3813	8.6048	9.1234	1.58e04	2.82e04	1.96e04
6	AFPs	1.6235	1.2663	1.70667	1.5741	0.2317	0.4215	0.3485
7	AFDur	9.72	5.5135	4.4706	5.3616	7.8763	12.2083	10.4386
8	TFP	8.6535	7.1032	5.5456	6.7091	207.0386	222.4157	216.0614
9	ASPF	3.6993	1.2342	3.3766	2.5450	101.9091	181.5344	146.3723
10	AOPF	3.6795	5.0210	3.0993	4.4248	9.5909	9.4245	10.7106
11	TFD	9.4761	7.3296	4.8066	6.5118	732.3210	236.1243	568.8518
12	ASDF	1.3978	2.0042	2.6408	2.4079	4.5865	8.9286	7.1097
13	AODF	1.8512	4.7619	3.8897	4.5156	8.5263	7.6326	8.7494

Table 2: KL Divergence, Entropy and Standard Deviations for our data set for FC. Note that none of the attributes on its own stands out for accurate discrimination between phantom and genuine user profiles.

of profiles classified as genuine profiles.

- *True Positive Rate* (TPR): The ratio of phantom profiles classified as phantom to the total number of phantom profiles.
- *False Positive Rate* (FPR): The ratio of genuine profiles that are classified as phantom to the total number of genuine profiles.

5.1 SVM Classifier

Support Vector Machines (SVMs) represent a general class of supervised learning methods that have been successfully applied in fields ranging from text recognition to bio-informatics [2]. The general idea behind an SVM is to project the training inputs to a high (possibly infinite) dimension space. The SVM can then construct hyper-planes that divide the high dimension space into regions that correspond to the classification categories. Our objective is to find hyper-planes that optimally or maximally separate phantom and genuine user profiles.

We use the Rapid Miner learning environment [9] to conduct our SVM based classification experiment. We use the *libsvm* learner [6] supported by Rapid Miner, and employ the Radial Basis Function (RBF) as our kernel function so that our classifier can create a non-linear classification[4]. Furthermore, depending upon the SVM parameter, the RBF kernel can model the linear and sigmoid kernels [6].

Training Data for SVM Classification: Our training data consists of input-label pairs. The label is a binary variable indicating whether a profile is a phantom profile or not. Our inputs are represented by a vector of real values. We use the 13-element vector of features detailed in Section 4. Our data set has 545 feature vectors for phantom user profiles, and 520 feature vectors for genuine user profiles. Each feature is distributed between some range. Since features in greater numerical ranges can dominate those in smaller numerical ranges [6], we linearly scale each attribute between -1 and 1.

We define different categories of input sets, where each category consists of a specific number of phantom and genuine user profiles. We generate 10 input sets by randomly selecting 50, 100, ... 500 phantom profiles. Each input set also contains a fixed number (500) of genuine user profiles. Therefore, the ratio between phantom profiles to genuine profiles ranges between 0.1 to 1 for our input set categories.

Feature Selection: Some of the 13 features we use may not contribute to a good discrimination between phantom and genuine user profiles. Such features can decrease the accuracy of our classification, and should be removed. We, therefore, conduct a preliminary feature-selection step using the *backward-elimination* algorithm. This algorithm starts with all 13 features and iteratively removes the feature that results in the largest increase in performance. It continues doing so until removing a feature does not result in any performance increase.

We measure the performance as the accuracy (ACC) averaged over our 10×10 input sets using a 10-fold cross validation for each set. Our cross-validation sets use stratified sampling so that the ratio of phantom to genuine profiles is preserved in each set. We set the SVM parameters to certain default values. The backward-elimination heuristic eliminates three features: 1) TFP, 2) TFD, and 3) ATFP. Without these features, the average accuracy (ACC) for our selected SVM increases from 69.43% to 81.05%. Furthermore, repeated experiments with different SVM parameter settings also result in the elimination

[4]We expect non-linear classification methods to provide satisfactory accuracy in discriminating between phantom and genuine user profiles.

of the same features. We, therefore, reduce our feature set to exclude these features.

Parameter Optimization: Our SVM is characterized by two parameters, the penalty parameter C and the RBF kernel parameter $\gamma > 0$ [6]. Our default settings for the feature selection experiment had $C = 0$ and $\gamma = \frac{1}{13}$ (reciprocal of the size of the global feature set). This resulted in an average accuracy (ACC) of 81.05%. Since tuning the SVM parameters can potentially improve performance, we use a two-phased grid search approach to ascertain the optimal values of C and γ.

In the first phase, we select C from $\{2^{-5}, 2^{-3}, ..., 2^{15}\}$, and γ from $\{2^{-15}, 2^{-13}, ..., 2^{3}\}$. Figure 2(a) shows the accuracy (ACC) of different combinations of C and γ averaged over 10-fold cross validation tests for all our input sets. The average accuracy across the points on our search grid is 76.31%. We use this step to further narrow down the parameter search space to the sets $\{0, 5, 10, ..., 50\}$ and $\{1\times10^{-6}\}\cup\{2\times10^{-5}, 2\times10^{-5}, ..., 2\times10^{-4}\}$ for C and γ, respectively. This improves the average accuracy across points on our search grid to 85.321%. Figure 2(b) shows that the accuracy does not vary much across the points of our refined grid. Since the highest accuracy is 86.14% for $C = 150$ and $\gamma = 2 \times 10^{-5}$, we select these values for our SVM for all the following results.

5.2 Classification Results

We now look at the classification performance of our learner in greater detail. In Figure 3(a), we show the accuracy (ACC), positive predictive value (PPV), and the negative predictive value (NPV) for the 10-fold cross validation experiments for each category. The results are averaged over the 10 random input sets of each category. Figure 3(a) shows that our SVM has a high ACC for all input categories. On the other hand, the PPV and NPV values vary. We see that the PPV is very low when there are only 50 phantom profiles compared to 500 genuine profiles in the input set. Our SVM classifier, therefore, is biased towards classifying more profiles as genuine resulting in more false negatives. However, we see that the PPV steadily improves as we increase the representation of phantom profiles in the input set. Understandably, the NPV decreases slightly as this representation of phantom profiles increases.

We report similar results for a slightly different experiment setting. That is, instead of looking at the cross-validation results, we use the entire input set to train our SVM. We then test the learned model on the remaining labelled data that was not included in the corresponding input set. The results are shown in Figure 3(b).

The major trend seen in Figure 3(a) is also seen in Figure 3(b), i.e., PPV monotonically increases with the ratio of phantom profiles to genuine profiles in the input set. Furthermore, we see that the PPV value for each category is close to the corresponding PPV value for our cross-validation experiments. However, we do not see any clear trend for the NPV. This is because the testing set only contains $520 - 500 = 20$ genuine profiles. The results are, therefore, not statistically significant. Since phantom profiles dominate the test set, we see that the overall accuracy (ACC) closely mirrors the PPV.

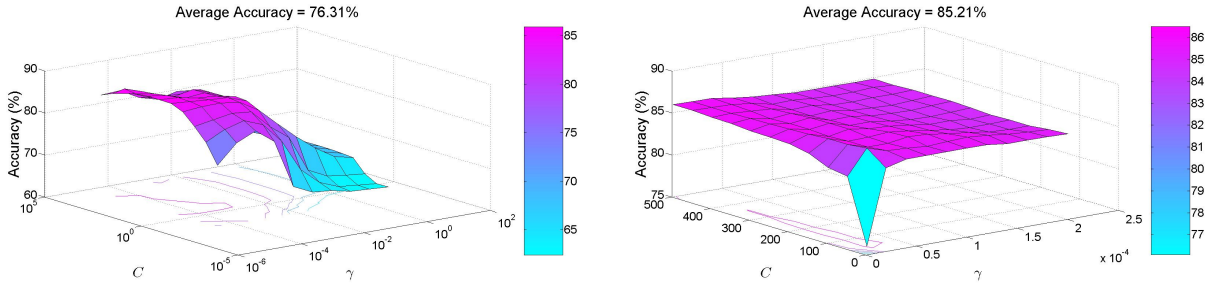
Figure 4(a) and Figure 4(b) plot the receiver operating characteristic (ROC) [4] for our cross-validation and our segregated testing set experiments. We would like the ROC to be as close to the upper-left corner i.e, false positive rate (FPR) close to 0%, and true positive rate (TPR) close to 100%. The ROC for the cross-validation experiments are close to this ideal scenario for all our input categories. Figure 4(a) suggests that a good tradeoff between the FPR and TPR is when we have 250 phantom profiles and 500 genuine profiles in our input set (FPR=13.4%, TPR=86.4%). Our classification using separate testing sets show similar TPR values. However, as indicated previously, the FPR has greater variance because of the very low number of genuine profiles.

We use SVMs trained on input sets in this category to classify close to $30,000$ unlabeled users whose activity and information are captured in our traces. Figure 5.2 shows the percentage of these unlabeled profiles that are classified as phantom. The SVM computes a probability that a certain profile is phantom. By default, if this probability is greater than 0.5, the profile is classified to be phantom. However, we see that this default threshold results in $33\%+$ profiles being classified as phantom for most of our learned SVM models. We suspect this corresponds to a very high FPR even though in some random samples we found the number of phantom profiles to be close to 20%. We can set a higher confidence threshold to lower the number of profiles classified as phantom. Figure 5.2 shows the percentage of our unlabeled profiles that are classified as phantom for different thresholds.

6. Discussion and Conclusion

Our experience in hosting a major Facebook social gaming application (FC) has granted us anecdotal evidence of phantom users created to gain strategic gaming advantage. The presence of such phantom users have significant measurement, security, and trust implications.

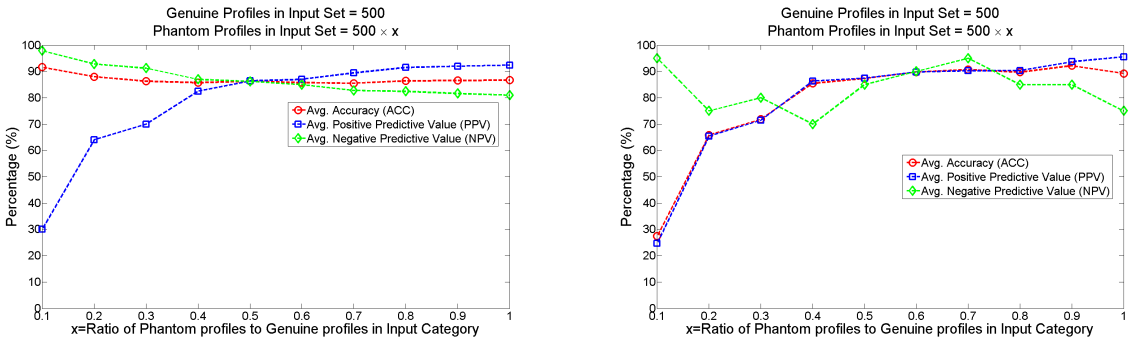
We identified phantom and genuine profiles and singled out features representing users' general Facebook activities, information about their social network friends, as well as more specific information about users' application usage. However, a statistical characterization of these features does not suggest any obvious discriminants between phantom and genuine users. We, therefore, require more elaborate machine learning techniques to detect phantom users.



(a) First Phase Parameter Grid Search

(b) Second Phase Parameter Grid Search

Figure 2: Parameter Tuning Experiment



(a) Cross-Validation Results

(b) Segregated Test-Set Results

Figure 3: Classification Performance for different Input Categories

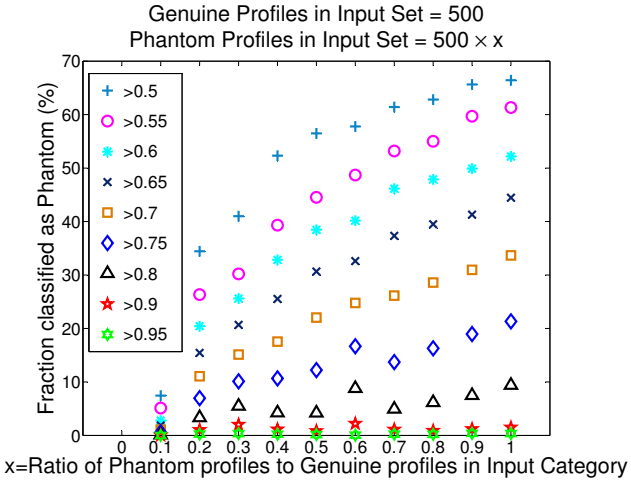
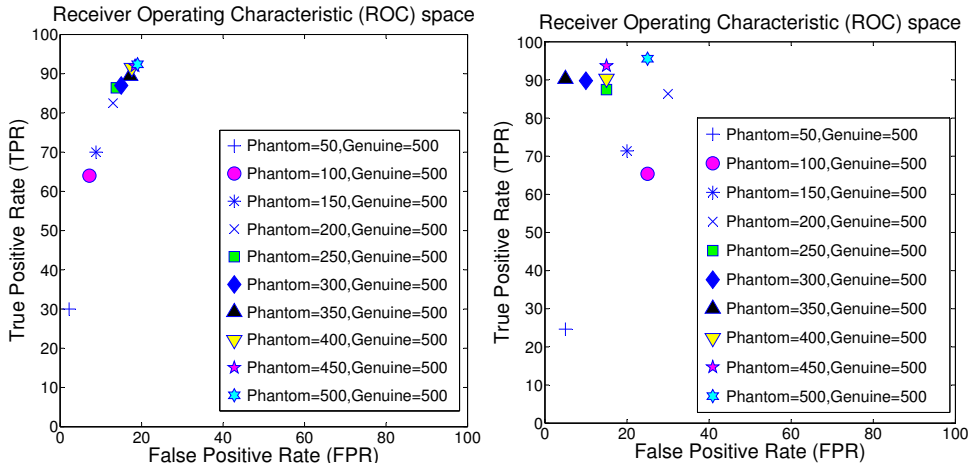


Figure 5: Predicted phantom profiles

Our use of SVMs resulted in very accurate classification for the phantom and genuine profiles we have identified. One of the major challenges in employing supervised learning techniques is correctly procuring the ground truth used to train the classifier. Our identification of phantom and genuine users is through a tedious manual process. A specific drawback of our method is that we do not know the relative proportion of phantom users and genuine users in the global population. Furthermore, there is potential for our training data to be biased if the users that are manually identifiable by us have selective properties that are different from the unidentifiable users. There is some evidence of such bias in that our SVM classifier determines an unexpectedly high number of our unlabeled users to be phantom (Figure 5.2). Mitigating this bias is an area of future work. We plan to diverge from our binary classification (i.e., into phantom or genuine users) by defining multiple classes corresponding to our confidence levels in manual identification of users.

We also intend to explore other supervised learning methods for the phantom user detection problem. Moreover, we considered a restricted set of features in this paper. Mining more detailed demographic and behavioral data about users has the potential to provide significant information that can be used for classification. It is also interesting to explore if (and how) different OSN applications can collaborate to detect phantom profiles. An interesting and associated problem is to identify the genuine users who create or benefit from such phantom users. Furthermore, an accurate detection technique for phantom users in gaming applications may also help us identify/explain patterns in their off-game social activi-



(a) Cross-Validation Results (b) Segregated Test-Set Results

Figure 4: Receiver Operating Characteristic (ROC)

ties. Lastly, our data can help support sociological studies, e.g., understanding incentives or personal traits that influence cheating vs. complying behaviors.

7. References

[1] BENEVENUTO, F., RODRIGUES, T., CHA, M., AND ALMEIDA, V. Characterizing user behavior in online social networks. In *Proc. ACM Internet Measurement Conference (IMC)* (2009).

[2] BURGES, C. J. C. A tutorial on support vector machines for pattern recognition. *Journal of Data Mining and Knowledge Discovery 2*, 2 (1998).

[3] CHUN, H., KWAK, H., EOM, Y., AHN, Y.-Y., MOON, S., AND JEONG, H. Comparison of online social relations in volume vs. interaction: A case study of cyworld. In *Proc. ACM Internet Measurement Conference (IMC)* (2008).

[4] FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters 27*, 8 (2006), 861–874.

[5] GARG, S., GUPTA, T., CARLSSON, N., AND MAHANTI, A. Evolution of an online social aggregation network: An empirical study. In *Proc. ACM Internet Measurement Conference (IMC)* (2009).

[6] HSU, C.-W., CHANG, C.-C., AND LIN, C.-J. A practical guide to support vector classification. Tech. rep., National Taiwan University, Taipei.

[7] KRISHNAMURTHY, B. A measure of online social networks. In *Proc. International Conference on Communication Systems and Networks (COMSNETS)* (2009).

[8] KRISHNAMURTHY, B., AND WILLS, C. On the leakage of personally identifiable information via online social networks. In *ACM SIGCOMM Workshop on Online Social Networks* (2009).

[9] MIERSWA, I., WURST, M., KLINKENBERG, R., SCHOLZ, M., AND EULER, T. Yale: Rapid prototyping for complex data mining tasks. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (2006).

[10] MISLOVE, A., MARCON, M., GUMMADI, K. P., DRUSCHEL, P., AND BHATTACHARJEE, B. Measurement and Analysis of Online Social Networks. In *Proc. ACM Internet Measurement Conference (IMC)* (2007).

[11] NAZIR, A., RAZA, S., AND CHUAH, C.-N. Unveiling facebook: A measurement study of social network based applications. In *Proc. ACM Internet Measurement Conference (IMC)* (2008).

[12] NAZIR, A., RAZA, S., GUPTA, D., CHUAH, C.-N., AND KRISHNAMURTHY, B. Network level footprints of facebook applications. In *Proc. ACM Internet Measurement Conference (IMC)* (2009).

[13] SCHNEIDER, F., FELDMANN, A., KRISHNAMURTH, B., AND WILLINGER, W. Understanding online social network usage from a network perspective. In *Proc. ACM Internet Measurement Conference (IMC)* (2009).

[14] TORKJAZI, M., REJAIE, R., AND WILLINGER, W. Hot today, gone tomorrow: On the migration of myspace users. In *ACM SIGCOMM Workshop on Online Social Networks* (2009).

[15] YU, H., KAMINSKY, M., GIBBONS, P. B., AND FLAXMAN, A. Sybilguard: Defending against sybil attacks via social networks. In *Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (2006).