User Awareness and Defenses Against Sockpuppet Friend Invitations in Facebook

Sajedul Talukder
Southern Illinois University, USA
sajedul.talukder@siu.edu

Nestor Hernandez
FIU, Miami, USA
nestorghh@gmail.com

Mozhgan Azimpourkivi
Bloomberg LP, USA
mojganaz@gmail.com

Bogdan Carbunar
FIU, Miami, USA
carbunar@gmail.com

ABSTRACT

Friend relations in Facebook have been used to access private and sensitive user data, post false or abusive information on the timelines and news feeds of victims, and scam and influence the perceptions of victims. In a qualitative study with 35 participants who use Facebook we found that they often confirm invitations received from sockpuppet accounts. Some participants claimed prior exposure to such invitations and exhibited increased awareness. However, in our study several such participants have also confirmed at least one sockpuppet invitation. To take steps toward addressing this problem, in this paper we introduce a new interface to inform, educate and nudge users to inspect their Facebook friend invitations before making a decision, and avoid confirming invitations received from suspicious accounts. We propose a classifier to detect undesirable friend invitations and use it to prioritize a must-see pending friend list. In studies with 145 participants, we found that when compared to a control, Facebook-like interface, our solution reduced confirmed friend invitations from sockpuppet accounts by 42.6 percentage points, reduced blind confirmations of such accounts by 22.6 percentage points, and increased the inspected profiles of such accounts by 54.67 percentage points. We show that when trained on only a few user decisions from these studies, our classifier predicted the user decisions for pending friends with an F1-measure of 96.52%.

CCS CONCEPTS

• **Security and privacy** → **Privacy protections**; *Social aspects of security and privacy*; *Spoofing attacks*;

KEYWORDS

Social network friend abuse, friend spam, sockpuppet, classifier

ACM Reference Format:
Sajedul Talukder, Nestor Hernandez, Mozhgan Azimpourkivi, and Bogdan Carbunar. 2022. User Awareness and Defenses Against Sockpuppet Friend Invitations in Facebook. In *The 37th ACM/SIGAPP Symposium on Applied Computing (SAC '22), April 25–29, 2022, Virtual Event, .* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3477314.3507012

SAC '22, April 25–29, 2022, Virtual Event,

ACM ISBN 978-1-4503-8713-2/22/04…$15.00
https://doi.org/10.1145/3477314.3507012

1 INTRODUCTION

Social networks like Facebook are used to collect and infer private and sensitive information from users [15], inject content to change user perception [22], scam and spam users [1], mislead stock market predictions [8], and distribute fake news, misinformation, propaganda and malware [16, 18]. Friend relationships are one gateway through which such attacks can take place. This is because in sites like Facebook, where users tend to disclose honest self-representations [12], their data are often shared by default with their friends.

Previous work [4, 9] has shown that less than half of cyber abuse victims subsequently adopt self-protective behaviors (SPBs) [30]. The rational choice perspective (RCT) [6] interpretation of this fact is that victims are more likely to adopt SPBs when they perceive the benefits of preventing repeat victimization to be higher than the cost of prevention [17]. In this paper we further posit that in social networks like Facebook, users repeatedly accept friend invitations from the accounts of potentially abusive strangers [1] for three additional reasons. First, Facebook provides users with an insufficient background to make an informed decision about the accounts that seek to befriend them, and also pressure users into confirming them. Second, some users lack the knowledge required to (1) understand the potential for abuse of often appealing stranger accounts and (2) identify and adopt appropriate SPBs. Third, Facebook's interface clutters user's pending requests in a single screen which increases their perceived cognitive load, which can be approximated by the number of interactions that the users think will be required by the page [10].

Case in point, in a qualitative study (n = 35) we asked participants who use Facebook to either confirm, delete or skip decisions for a mix of their real pending friends in Facebook and five *synthetic friends*, i.e., account profiles that we fabricated. We found that 24 participants have each confirmed at least one out of the five synthetic friends shown. We further report vulnerability even among educated participants who claimed and exhibited increased awareness to invitations received from sockpuppet accounts.

In this paper we introduce FriendLock [1], a new UI design to process Facebook pending friends. FriendLock seeks to complement existing solutions that detect and eliminate abusive accounts [13, 33],

[1] [Urban Dictionary]: Feeling pressured to add as friend someone who you typically would not, due to association with one of your friends.

by involving users in the decision process for accounts that escape abuse detection. FriendLock provides soft-paternalistic mechanisms or "nudges" [5, 31], for users to investigate their pending requests, and reduce their impulse to blindly confirm them as friends, without affecting their autonomy.

FriendLock further leverages the above synthetic pending friend concept to educate users about the dangers of blindly accepting friend invitations. We develop educational material to enable users to identify and avoid synthetic friends, and raise awareness about the importance of carefully considering each pending friend.

In addition, to reduce the cognitive load when processing pending friends, FriendLock leverages a filtering device to help users cope with "overmanning" [32]. Specifically, we introduce a novel combination of a pending friend classifier and a *must-see pending list*: The classifier identifies friend invitations that are non-suspicious and are unlikely to be ignored by the user. The user is shown with priority these must-see invitations; the order may change in time as more information becomes available through the social network.

In quantitative user studies with 145 participants we found that when compared to the control Facebook-like interface, FriendLock reduced the percentage of confirmed synthetic friends to 8% from 50.6%, increased the number of inspected synthetic friend profiles to 82% from 27.33%, and decreased the percentage of blind confirms for synthetic friends from 55.9% to 33.3%. Further, when trained with only five past decisions per user and using externally-accessible features alone, FriendLock's pending friend classifier achieved an F1 score of 84.08%. The F1 score increased to 96.52% when trained on 25 past decisions per user. In summary, we introduce the following contributions:

- **Study user perceptions and vulnerabilities**. We analyze user perceptions and reasons to confirm and reject friend invitations, that reveal both vulnerabilities and abuse awareness. We confirm through quantitative and qualitative studies that participants often confirm blindly friend invitations received from synthetic accounts.
- **Interface designs for pending friends**. We introduce a new UI design to process received friend invitations, that encourages users to carefully evaluate pending friends, and raises awareness to suspicious friend invitations.
- **Predict user decisions**. We develop a classifier to identify invitations from suspicious or likely to be ignored accounts, and introduce a must-see pending invitation list to help users regulate their relationship boundaries.
- **Experimental results**. We introduce metrics to evaluate responses to pending friends. We provide ground truth evidence that FriendLock reduces the user vulnerability to invitations from synthetic accounts.

2 RELATED WORK

Previous work on detecting abusive accounts includes using graph structure and features of friend accounts [25, 34] and studying spam propagation patterns [19]. Yu [35] further surveys efforts to detect sockpuppet or fake accounts in social networks, many of which also use social graph connections [14]. Facebook uses a variety of techniques to detect sockpuppet and abusive accounts. For instance,

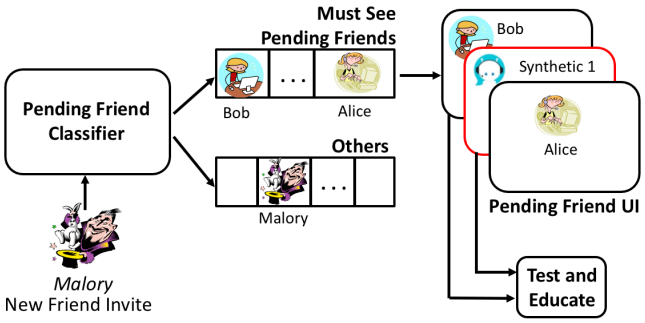


Figure 1: System architecture. Upon receiving a friend invitation, the pending friend classifier decides if it must be seen by the user. The pending friend UI nudges the user to inspect each pending profile. The UI leverages synthetic profiles to test and educate the user.

Xu et al. [33] reveal that Facebook employs deep learning-based techniques to detect abusive accounts, while Kozlov et al. [13] provide insights into techniques used at Facebook to detect fake accounts and protocols used to evaluate such techniques. With FriendLock we leverage the observation that one way to address the unavoidable false negatives of classifiers that detect sockpuppet and abusive accounts is to raise the awareness of potential victims through education and nudges, thus use them as a second line of defense.

Talukder et al. [26, 27] introduce AbuSniff, a system that identifies Facebook friends perceived as strangers or abusive, and protects the user by unfriending, unfollowing, or restricting the access to information for such friends. Unlike AbuSniff, FriendLock seeks to educate users against suspicious pending friends at the time of the request, and avoid the harm that may be inflicted if such requests are confirmed. Onaolapo et al. [20] fabricate Facebook accounts and expose them to attackers to understand the effects of demographic attributes on attacker behavior in stolen social accounts. In this paper, we introduce synthetic friend invitations to collect ground truth on fake friend acceptance and educate users via their incorporation in new UI designs.

To validate suspected sockpuppet accounts, previous work used challenges that include e-mail verification, phone confirmation, puzzles, e.g., CAPTCHAs, IP restrictions, and liveness detection systems [29]. FriendLock can be considered a challenge-based sockpuppet account validation solution, where the challenge is implicit and is not shown to the account owner but rather inferred from the behavior of other people which they seek to befriend.

3 SYSTEM AND ADVERSARY MODEL

Facebook users form *friend* relationships, initiated when users send invitations to other Facebook users, see Figure 1. Friend invitations are stored in a *pending friend list*. Pending friends cannot access any of the non-public information of the user. Users can inspect their pending friend list in order to make a decision, and view the profile of any pending friend, by tapping on its entry in the pending friend list. A user can confirm, delete or skip a pending friend. Upon confirmation, the pending friend is moved to the user's friend list, where it will be able to access all the account information that the user shares with friends.

Adversary Model. We consider adversaries who control multiple social network accounts, a.k.a sockpuppets, created using fraudulent information (e.g., photos collected online), and perhaps maintained with the help of bots [2]. Adversaries use sockpuppet accounts to send friend invitations to other, victim accounts. After such an invitation is confirmed, the adversary can engage in further abusive behaviors that include scams [1], spam [23], and the distribution of fake news, misinformation, propaganda and malware [16, 28].

4 FRIENDLOCK

FriendLock consists of three main components (1) the pending friend processing UI, (2) the pending friend classifier, and (3) the education component, see Figure 1. In the following we first introduce the concept of synthetic friend profiles, which is pivotal for FriendLock, then detail each component.

Synthetic Friends. To evaluate the willingness of users to accept spam friend invitations, the apps randomly mix among the real pending friends, 5 synthetic pending friends, i.e., fake profiles (3 female, 2 male if the user is male; 3 male, 2 female if the user is female). We personalized synthetic friends to be from the same country and live in the same city as the user, and have a randomly chosen number of mutual friends. We created a personalized list of names [3] and profile photos for the synthetic friends, for each country of the participants, see Figure 2(b) for a screenshot of a synthetic friend. We chose the number of followers of a synthetic friend, to be a random value between zero and half of the user's number of friends. We used identity-neutral images (e.g., nature photos) to display as part of the synthetic friend's profile.

4.1 Pending Friend Processing UI

We conjecture that the Facebook UI for processing pending friends biases users toward accepting pending friends thus increasing their vulnerability to sockpuppet invitations. We further conjecture that two design choices contribute to this bias, see Figure 2(c). First, the emphasized, blue-colored "Confirm" button vs. the gray "Delete" button of Facebook. Second, the crammed listing of all pending friends on a single screen that makes it difficult for users to make informed decisions for each pending friend.

To evaluate this hypothesis and provide a first-line, user-centric defense against sockpuppet account invitations, we designed the FriendLock user interface to process pending Facebook friends. The UI seeks to reduce the clutter and cognitive load by displaying each pending friend on a single screen, with a large, centrally-placed profile photo. Further, we transform the "Confirm" button into an inhibitive attractor [7], by displaying it in the same gray color of the "Delete" button. To address the case where the user does not feel comfortable making a decision, we also include a "Skip" button shown in the same gray color. Figure 2(a) shows an example for a synthetic profile.

Similar to Facebook, when the user taps on the profile photo of a pending friend, a screen is shown that includes the profile of the pending friend, see Figure 2(b). Users can navigate their pending friend list using the next and previous buttons on the sides of the profile photo, see Figure 2(a). FriendLock logs the time of each user action.

4.2 Education Component

We conjecture that Facebook users can learn to detect and carefully evaluate friend requests received from stranger accounts with the potential for cyber abuse. To evaluate and leverage this hypothesis, we have extended FriendLock with an education component: when the user accesses the pending friend list, FriendLock displays m (m = 3 in our studies) synthetic profiles in succession, using the UI of Figure 2(a).

For each such synthetic friend, if the user blindly confirms the invitation, i.e., without exploring the user's profile, FriendLock displays the message "*We generated this synthetic profile to help you make better decisions in the future. Such profiles could belong to a scammer who wants access to your profile data, timeline and newsfeed*". If the user chooses to delete, FriendLock displays the message "*Great choice! This was a synthetic profile, that we generated to help you make safer decisions in the future*".

Following the first m synthetic profiles FriendLock continues to display the participant's actual pending friends, randomly mixed with other synthetic friends (different from the first m). However, at this stage FriendLock no longer shows the warning or congratulation messages following the user processing of synthetic friends.

4.3 Classifying Pending Friend Invites and the Must-See List

More than 22% of the participants in our studies had more than 20 pending friends. Participants in our online studies ignored, i.e., skipped 24.11% (550 out of 2281) of friend invitations. 86.66% (104 out of 120) participants made at least one skip decision. To help reduce the cognitive load in evaluating pending friend requests, we introduce a pending friend classifier that seeks to identify undesirable invitations, i.e., friend invitations that are likely to be ignored by the user or friend invitations that should be ignored by the user.

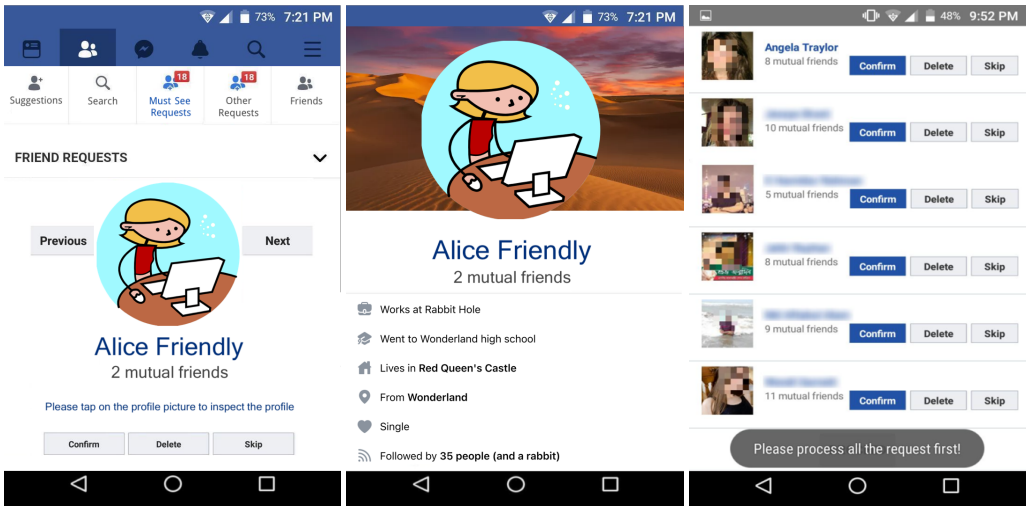
We trained a supervised learning model to predict the user decisions on pending friend requests using features that we extracted from the data that we collected in our user studies from Facebook users and their pending friends. The data was collected through various ways including our questionnaire, APIs and profile crawler. The features we extracted from the user include gender, age, region, occupation, education, sharing preferences, frequency of invitation reception, account use, number of friends, number of pending friends, device price, device age, and percentage of past confirm, delete and skip decisions.

The features we extracted from the user and the pending friend include pending friend gender and age, the number of mutual friends, whether they live in the same city or have the same hometown, and profile inspection count.

We can use this classifier to split the list of pending friends into a *must-see* and an *undesirable* list. Both lists are sorted in decreasing order of the likelihood that the invite will not be ignored by the user. FriendLock displays profiles from these lists when the user taps the Must-See Requests and Other Requests buttons respectively, shown at the top of the screen, see Figure 2(a). In section Evaluation of the Friend Invite Classifier we experimentally evaluate this classifier.

[2]Facebook estimated that 13% (i.e., 270 million) of their accounts are either bots or clones [11].

[3]https://www.behindthename.com/random/



(a) (b) (c)

Figure 2: (a) FriendLock UI decision page for a synthetic pending friend shown in the must-see list. (b) FriendLock profile of the pending friend shown when the user chooses to inspect the synthetic account further. (c) Control app: the friend request interface emulates the one of the Facebook app.

This approach serves as a filtering device that helps users cope with "overmanning" [32]. The must-see and undesirable lists are generated online each time the user chooses to process pending friends. Thus, the membership and order in this list can change, as additional information becomes available to the classifier.

5 METHODS

We now detail our studies, including in-person interviews to investigate user perception and reaction to pending friends and quantitative studies with online participants to evaluate FriendLock.

5.1 Qualitative Study

In order to study the reasons to accept synthetic friends and observe whether an intervention can influence the desire to blindly accept such requests, we conducted semi-structured interviews with 35 consenting participants. We have recruited 20 of these participants in-person, at various locations in the US (e.g., supermarkets, grocery stores, restaurants, social events). We recruited the other 15 participants online, from JobBoy.com, and performed the interviews over Skype.

The 21 male and 14 female participants were between 18-65 years old (M = 32, SD = 10.74), with diverse backgrounds ranging from accountant to waitress, and elementary education to graduate degree. The participants had between 36 and 3,131 Facebook friends with an average of 505 friends (M = 308, SD = 694.11).

During the interview we used a FriendLock version that displays each of the pending and synthetic friends, in random order, one by one, on the screen, but does not include confirm, delete and skip buttons. Instead, we asked each participant to either confirm or delete each pending friend, and recorded their answers.

The participants took an average 38 minutes (M = 35, SD = 8.43) to complete the interview, and we paid them 15 USD.

Annotation. We recorded the audio of all interview sessions, with the consent of the participants, and transcribed relevant segments. We also took field notes during each interview. We reviewed field notes immediately following each interview, and noted tentative insights in field memos. Analysis of the data was an ongoing process until the responses reached saturation. We labeled the surveys and audio records with each participant ID, to link participant responses and audio transcripts. Two coders independently and thematically analyzed audio recordings for the discussion periods, and performed "macro-level lumper coding" [24]. The coders individually reviewed 40% of all responses, and any disagreements about code assignments were resolved through discussion. The 43 coded responses produced an excellent inter-rater reliability, with an overall Cohen's kappa (k) coefficient of 0.89 with an agreement of 96.12%. We then single-coded the remaining responses.

5.2 Quantitative Investigation

To evaluate FriendLock we have conducted user studies with 145 online participants that we recruited from JobBoy.com. The jobs we posted asked each participant to install one of our apps from the Google Play store, use it to login to their Facebook accounts and follow the instructions on the screen. We have only recruited participants who had at least 30 Facebook friends and at least one pending friend, had access to an Android device, and were at least 18 years old. We applied a cryptographic hash function to the Facebook

account IDs and used the anonymized IDs to ensure that each human participated only once in our studies.

To evaluate the impact of various FriendLock components we divided the 145 participants randomly into three groups. Participants in the first group (n = 35) were asked to install and use a version of FriendLock without the education component. Participants in the second group (n = 68) were asked to install FriendLock with the education component. Participants in the third group (n = 42) were asked to install a control app, described below.

Control App. To evaluate the impact of FriendLock we have designed a control app that emulates the Facebook interface for processing pending friends, see Figure 2(c). We have shown this app only to a control group. The app shows a mix of all the pending friends of the user and five synthetic friends. The real and synthetic friends are listed in random order, and shown on a single screen. When the user taps the "Proceed" button at the bottom of the screen, the app displays the next screen only if the user has made a decision for each entry in the list. Otherwise, the app changes the current screen to include an additional, "Skip" button for each pending friend. The app also pops up a message asking the user to make a decision for each pending friend. The reasons for this design are that we want the initial app to look similar to Facebook, however, we also want participants to make an explicit decision for each friend, without forcing them to decide between "Confirm" and "Delete". As mentioned above, the user can inspect the profile of each pending friend by tapping on the profile photo or name of the friend.

Both FriendLock and the control app have the following functionality.

Attention check and tutorial. Before login, FriendLock displays an instructional manipulation check [21], to verify that the participant understands English, reads all the text, understands and follows simple instructions. Following login, while loading the Facebook data of the user's pending friends, the app displays a tutorial screen. To ensure the quality of the data collected, we discarded the data of 25 participants (out of the total of 145 participants) who failed the attention check or declared an incorrect location, i.e., inconsistent with both the Facebook listed location and the device IP address.

Post-study questionnaire. After processing pending friends, FriendLock displays screens with multiple-choice questions about the user's age range, gender, occupation, the highest level of education, the age of their Facebook account, the frequency of Facebook use, and the frequency of received friend invitations.

Demographics. Of the 120 participants that remained after the above selection process, 69.16% were male and 30.83% female, and were between 18-70 years old (M = 25, SD = 7.56). The top 5 countries of the participants are Bangladesh 25.0%, India 22.50%, USA 11.66%, Nepal 5.83% and Pakistan 5.0%. The participants had diverse occupations, and education levels, with high-school and bachelor's degrees being most frequent (30.83% and 45.0% respectively).

Payment screen. The final screen of the app displays a code, which the user needs in order to prove completion of the app experience and redeem their payment. We have paid each participant, including the ones whose data we discarded, $3, for an average job completion time of 11.18 minutes (M = 11.58, SD = 3.18).

Data. We have collected from each participant the total number of friends, gender, country of origin (selected from a drop-down menu), answers to the post-study questionnaire, and, for each processed pending friend, its type (true pending or synthetic), the user decision and timing.

Evaluation Metrics. We used several metrics to evaluate user behaviors concerning their pending friends: (1) *Decision split*, i.e., the number of pending friends confirmed, deleted and skipped, (2) *Inspected profiles*, the number of pending friend profiles inspected, and (3) *Blind confirmations*, the number of pending friends confirmed without prior profile inspection.

5.3 Ethical Considerations

We have developed our protocols to interact with participants and collect data in an ethical, IRB-approved manner. For analysis purposes, we only stored anonymized data that we extracted from the participants or their pending friends. The actual raw data was only displayed to the participants during the study. Further, we did not exfiltrate and store Facebook listed locations of the participants or their IP addresses, but only the user-selected country of domicile, if consistent with the IP address. Since we do not preserve this information, its handling does not fall within the PII definition of NIST SP 800-122 [3]. Under GDPR [2], the use of the information without context, e.g., name or personal identification number, is not considered to be "personal information". We communicated our data collection process during recruitment and in the consent form, and we recruited only consenting participants. We paid all the recruited participants, regardless of them passing the attention check screen.

6 FINDINGS

We first present our findings from the qualitative study, then compare user behaviors when using FriendLock with and without the education component, against the control app.

6.1 Qualitative Study: Perceptions of Friend Invitations

All participants in the qualitative study were shown five synthetic profiles mixed randomly with their real pending friends. During the first round, 37% (i.e., 65 of the total 175) of the synthetic pending friends were confirmed by the 35 participants. In 36 of these cases, the participants confirmed such synthetic friends even though they later said that they did not trust them, had a negative feeling about them, or thought they would be abusive in the future. 24 of the 35 participants have confirmed at least one synthetic pending friend, with five participants confirming all such friends.

In 14 instances, participants claimed to have confirmed a synthetic account because they found them attractive. In seven cases participants claimed to have a common background with the synthetic friend, even though none was possible. For instance, P3 (42 years old male grocery store owner) said about a synthetic friend:

"She looks like I have seen her in my store before. I don't remember exactly when."

Five participants accepted all shown synthetic friends. Three of them revealed low technical expertise to be the reason for their

vulnerability to such invitations. For instance, P16 (27 years old unemployed male with a high school degree) said:

"I normally accept all requests at once without checking. Every few months I go over all of my friends and delete the ones that seem to be fake.".

P33 (53 years old male businessman) said:

"I am still a little scared about the technology. My nephew helped me to create this account. Most of the time I press this (confirm) button when I see a request."

Two participants however claimed a perceived tight control over their activities. For instance, P34 (32 years old male engineer) said:

"My Facebook privacy settings are very strict. Only my close Facebook friends can see my personal information. That's why I normally don't mind adding people."

P27 (65 years old male professor) motivated his decision to accept 3 synthetic friends, by a perceived safety of his posts:

"I'm obsessive about what I post on Facebook, only pictures that I'd be comfortable with the world seeing, and innocuous posts. That's why I don't mind accepting strange friend requests."

62.8% (110 of 175) of the synthetic friend invitations were rejected by the 35 participants. 30 of the 35 participants have rejected at least one synthetic pending friend, with 11 participants rejecting all such friends. Most of the synthetic requests were rejected because the participants identified the profile to be fake. For instance, P2 (27 years old male grad student) said:

"This account looks like a fake account to me. The picture of the profile is a popular actress in Hong Kong.".

Some participants exhibited an increased awareness due to past experiences. For instance, P13 (26 years old female grad student) said:

"I recently got a friend request from a profile picture that was artificially generated. Now I became so concerned about accepting any request from an unknown face."

Some participants rejected several of their actual pending friend invitations, for reasons that include past abuse, perceived lack of connection and vanity. For instance, P15 (34 years old female office clerk) said:

"I was once harassed by a group of men in my former workplace. Since then I am scared of accepting any man as Facebook friend."

P32 (30 years old female high school teacher) said:

"I think this man is desperate. We met just yesterday at a bus ticket station and chatted for 2 minutes. Now he added me on Facebook the next day.", while P2 said *"I only accept a friend if I know him or her personally and if I have a strong connection."*

This reveals participant exposure to undesirable invitations in their everyday Facebook interactions, and suggests benefits for FriendLock's invitation classifier and must-see pending list. In the following we evaluate the protection provided by FriendLock against synthetic profiles, when compared against the control app.

6.2 Quantitative Studies: Impact of Our Interfaces

In the following we use a significance level of $\alpha = 0.05$ in all the statistical tests. We used the Bonferroni correction method to account for Type I errors when performing multiple comparisons across groups.

Facebook usage details. 50% of the participant Facebook accounts were between 3 and 5 years old, while 19.16% were over 10 years old. Participants had between 30 and 2,720 Facebook friends, with an average of 574 (M = 323, SD = 588.84). The participants declared diverse frequency of Facebook account use, ranging from frequent (every hour 24.16%, several times a day 31.66%, once a day 7.50%), to less frequent (several times a week 10.83%, once a week 8.33% and occasionally 17.50%). None of the participants declared to rarely access their accounts. Most participants declared that they received friend invitations either once a week (36.66%), once a month (35.83%), or occasionally (25.83%). Participants had an average of 12 pending friends (M = 13, SD = 6.71). 27 participants had each at least 20 pending friends.

Decision split. Figure 3(a) compares the percentage of synthetic and real pending friends that were confirmed, deleted or skipped by the participants who used the control, FriendLock without education and FriendLock with education apps. We observe that 50.6% (76/150) of synthetic pending invitations were confirmed through the control app, compared with 24.7% (37/150) through FriendLock w/o education, and 8% (24/300) through FriendLock with education.

A two-proportion one sided z-test ($H_0 : p_{cb} = p_{fb}$ vs. $H_a : p_{cb} > p_{fb}$ where p_{cb} and p_{fb} are the proportion of synthetic friends confirmed under the control and FriendLock w/o education treatment respectively) produced a $p\_value < 0.05$ ($z = 4.64$), suggesting that the proportion of synthetic pending friends confirmed is statistically significantly lower for the FriendLock w/o education group. The portion of accepted synthetic friend requests in the FriendLock with education group was further significantly lower than that in the FriendLock w/o education ($z = 4.86$, $p\_value < 0.05$), suggesting significant impact of the education component. However, in the case of actual pending friends, in FriendLock and FriendEdu studies, we did not observe a significant decrease in the number of confirmed requests or increase in the number of skipped requests compared to the control study. This suggests that, our FriendLock versions would not impair the user's ability to correctly decide about their actual pending friends.

In addition, in both FriendLock groups, participants confirmed fewer actual pending friend invitations, compared to the control: 30.66% in FriendLock with education and 27.16% in FriendLock w/o education vs. 37.46% in the control group. A two-proportion one sided z-test revealed the portion of confirmed actual pending friends in both FriendLock groups to be significantly lower than that of control study ($z = 2.78$, $p\_value < 0.05$ and $z = 2.35$, $p\_value < 0.05$ respectively). However, this test did not show any significant impact for FriendLock's education component against FriendLock ($z = 1.073$, $p\_value = 0.85$).

Inspected profiles. Figure 3(b) compares the percentage of profiles inspected, grouped by the type of friend request (synthetic or actual pending), in the control and FriendLock groups. Overall, we note that while in the control Facebook-like UI, 15.33% (94/613) of the

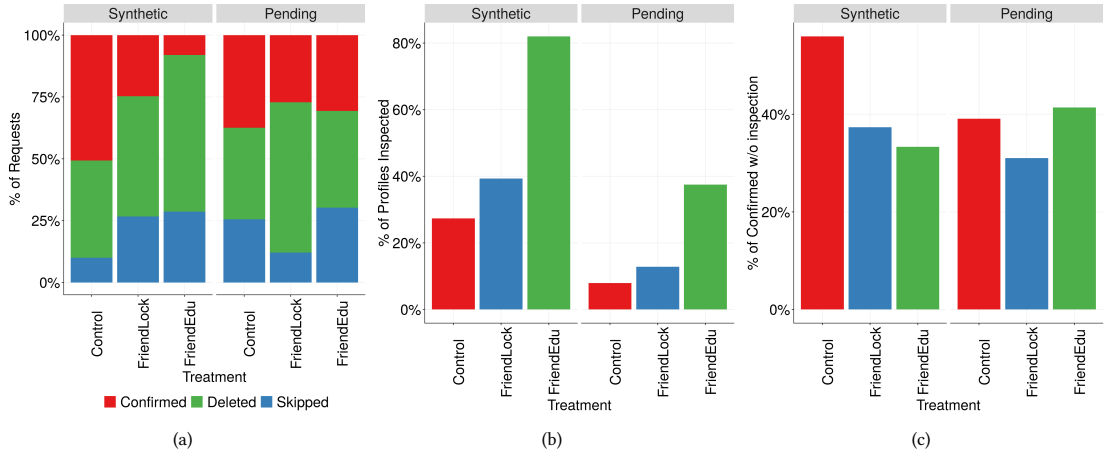


Figure 3: Comparison of control, FriendLock w/o education (labeled FriendLock) and FriendLock with education (labeled FriendEdu). (a) Per-treatment distribution of confirmed, deleted, and skipped requests. FriendEdu reduced the percentage of confirmed synthetic friends from 50% (control) to 8.8%. (b) Per-treatment distribution of profiles inspected by type of friend request. Both FriendLock flavors significantly increased the percentage of friend invites whose profiles were inspected. (c) Per-treatment distribution of profiles confirmed without inspection. FriendEdu reduced the percentage of synthetic friends confirmed blindly to 33.3% from control's 55.9%.

pending friend invitations were inspected, both FriendLock without education and with education increased the number of inspected profiles, to 25% (122/475) and 53% (643/1193) respectively. A two-sample test showed that the proportion of inspected profiles in the FriendLock without education study is statistically significant higher than that of control ($z = -4.24, p\_value < 0.05$), while the proportion of inspected profiles in further statistically increased by the education module ($z = -10.43, p\_value < 0.05$). These results are also statistically significant per friend type (i.e., actual pending and synthetic).

Blind confirmations. Figure 3(c) shows the percentage of pending friends who were confirmed blindly, i.e., without prior profile inspection, by type of friend request. On the whole, we observe that in the control study 44.12% (229/519) of invites were confirmed blindly, while both FriendLock flavors reduced the proportion of blind confirmations, to 36.15% (128/354) and 41.56% (229/551) respectively. A hypothesis test showed that the difference between the control and FriendLock without education is significant ($z = 2.35, p\_value = 0.0113$), but it did not show a significant difference with the education component ($z = 0.84, p\_value = 0.216$). As can be seen from Figure 3(c), this is due to the actual pending friends, where the number of blind confirmations is higher: participants are likely to have seen the actual pending requests in the past, thus were more familiar with pending friends. However, we observe a statistically significant reduction in the number of blind confirmations of synthetic invitations, from the control (55.96%) to FriendLock without education (37.36%; $z = 2.62, p\_value < 0.05$), and also impact of education (33.33%, $z = 2.72, p\_value < 0.05$).

6.3 Evaluation of the Friend Invite Classifier

We trained the pending friend classifier model (see Classifying Pending Friend Invites and the Must-See List) using data that we collected from our control study. We chose the control study data as we wanted to use data from Facebook-like interface without having extra component or modified UI. For the prediction task, our dataset consists of tuples (X_i , Y_i), where X_i is the feature vector and Y_i is the user decision on the i-th friend invitation, i.e., skip, delete or confirm.

We have used Gradient Boosting Machine (GBM), Random Forrest (RF), Naive Bayes (NB), Multinomial Logistic Regression (MLR) and K-Nearest Neighbor (KNN) using the 613 tuples from the 30 control participants that had the unadulterated, Facebook-like view of the friend invitation interface. We used k-fold cross-validation for time series, where we train on the first k decisions of each user, and test on the rest of the invitations. k is between 1 and the maximum number of pending friends per participant.

Algorithm	k	Precision	Recall	F1
GBM	5	81.47%	78.75%	80.08%
	24	**96.75%**	**96.29%**	**96.52%**
RF	**5**	**85.49%**	**82.72%**	**84.08%**
	25	92.22%	88.88%	90.52%
NB	5	81.02%	79.67%	80.34%
	24	94.23%	92.59%	93.40%
MLR	5	70.91%	70.66%	70.79%
	24	86.17%	85.18%	85.67%
KNN	5	65.96%	59.78%	62.72%
	22	77.77%	77.77%	77.77%

Table 1: Precision, Recall, and F1 measure for different machine learning classifiers as a function of the history length k. RF achieves an F1 of 84.08% considering only 5 previous friend requests while GBM achieves an F1 of 96.52% when trained on the past 24 decisions.

Table 1 shows the Precision, Recall, and F1 measure for different machine learning classifiers. When trained on only the first five decisions of each participant, RF outperformed GBM, NB, MLR, and KNN. Specifically, RF achieved an F1 of 84.08%, with a precision of 85.49% and a recall of 82.72%. However, when trained on the first 25 decisions of a participant, GBM outperformed all the other classifiers with a precision of 96.75%, recall of 96.29% and F1 measure of 96.52%. RF achieved an F1 measure of 90.52%.

7 CONCLUSIONS

In this paper we have analyzed user perceptions and reasons to confirm and reject friend invitations in Facebook, and found both participant vulnerabilities and abuse awareness. We confirmed through quantitative and qualitative studies, that participants often blindly and quickly confirm friend invitations received in Facebook from synthetic accounts. We have developed classifiers and new designs of interfaces to process pending invitations. We believe that FriendLock's substantial reduction in the total number of confirmed synthetic friends and in the number of participants who confirmed at least one such friend, is essential in protecting against fake friend invitations and subsequent attacks. We further emphasize the importance of FriendLock's increase in the number of inspected pending friend profiles. This suggests that users can be trained to avoid making automatic decisions with important security implications. We consider this study as a preliminary research. To further confirm the validity of our findings, we plan to conduct future large-scale user studies with a larger number of participants.

8 ACKNOWLEDGMENTS

This research was supported by NSF grants CNS-2013671 and CNS-2114911, and CRDF grant G-202105-67826. This publication is based on work supported by a grant from the U.S. Civilian Research & Development Foundation (CRDF Global). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of CRDF Global.

REFERENCES

[1] The colossal scam of Finland's fake lovers. BBC Reel, https://www.bbc.com/reel/video/p083tr4z/the-colossal-scam-of-finland-s-fake-lovers, February 2020.

[2] General data protection regulation (gdpr). https://gdpr-info.eu/, 2021. Accessed: 2021-02-12.

[3] Guide to protecting the confidentiality of personally identifiable information (pii). https://tinyurl.com/ylyjst5y, 2021. Accessed: 2021-02-12.

[4] Margit Averdijk. Reciprocal effects of victimization and routine activities. *Journal of Quantitative Criminology*, 27(2):125–149, 2011.

[5] Rebecca Balebako, Pedro G Leon, Hazim Almuhimedi, Patrick Gage Kelley, Jonathan Mugan, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. Nudging users towards privacy on mobile devices. In *Proceedings of the CHI Workshop on Persuasion, Nudge, Influence and Coercion*, 2011.

[6] Cesare Beccaria. On crimes and punishments. *Criminology Theory: Selected Classic Readings*, page 367, 1764.

[7] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W. Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. Your Attention Please: Designing Security-decision UIs to Make Genuine Risks Harder to Ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, pages 6:1–6:12. ACM, 2013.

[8] Stefano Cresci, Fabrizio Lillo, Daniele Regoli, Serena Tardelli, and Maurizio Tesconi. Fake: Evidence of spam and bot activity in stock microblogs on twitter. In *International AAAI Conference on Web and Social Media*, 2018.

[9] Maeve Duggan. Online Harassment. Pew Research Center, https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/, 2017.

[10] Simon Harper, Eleni Michailidou, and Robert Stevens. Toward a Definition of Visual Complexity as an Implicit Measure of Cognitive Load. *ACM Transactions on Applied Perception (TAP)*, 6(2):1–18, 2009.

[11] Alex Heath. Facebook quietly updated two key numbers about its user base. [Business Insider], tinyurl.com/y76s8rvs, 2017.

[12] Kokil Jaidka, Sharath Chandra Guntuku, Anneke Buffone, H Andrew Schwartz, and Lyle Ungar. Facebook vs. Twitter: Differences in Self-disclosure and Trait Prediction. In *Proceedings of the International AAAI Conference on Web and Social Media*, 2018.

[13] Fedor Kozlov, Isabella Yuen, Jakub Kowalczyk, Daniel Bernhardt, David Freeman, Paul Pearce, and Ivan Ivanov. Evaluating changes to fake account verification systems. In *23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020)*, pages 135–148, 2020.

[14] Srijan Kumar, Justin Cheng, Jure Leskovec, and V.S. Subrahmanian. An army of me: Sockpuppets in online discussion communities. The Web Conference 2017, 2017.

[15] Issie Lapowsky. Cambridge Analytica Execs Caught Discussing Extorsion and Fake News. [Wired] https://tinyurl.com/yaagbe9h, 2018.

[16] David Lee. Facebook, Twitter and Google berated by senators on Russia. [BBC Technology] tinyurl.com/ybmd55js, 2017.

[17] Bill McCarthy and Ali R Chaudhary. Rational Choice Theory and Crime. *Encyclopaedia of crime and criminal justice*, pages 2–12, 2014.

[18] Alfred Ng. How Russian trolls lie their way to the top of your news feed. [CNET], https://tinyurl.com/yag4t8zq, 2017.

[19] Shirin Nilizadeh, François Labrèche, Alireza Sedighian, Ali Zand, José Fernandez, Christopher Kruegel, Gianluca Stringhini, and Giovanni Vigna. POISED: Spotting Twitter Spam Off the Beaten Paths. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1159–1174, 2017.

[20] Jeremiah Onaolapo, Nektarios Leontiadis, Despoina Magka, and Gianluca Stringhini. Socialheisting: Understanding stolen facebook accounts. In *30th USENIX Security Symposium (USENIX Security 21)*, 2021.

[21] Daniel M Oppenheimer, Tom Meyvis, and Nicolas Davidenko. Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power. *Journal of Experimental Social Psychology*, 45(4):867–872, 2009.

[22] Barbara Ortutay and Anick Jesdanun. How Facebook likes could profile voters for manipulation. [ABC News] https://tinyurl.com/yaaf3lws, 2018.

[23] Elissa M. Redmiles, Neha Chachra, and Brian Waismeyer. Examining the demand for spam: Who clicks? In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 212:1–212:10, 2018.

[24] Johnny Saldaña. *The coding manual for qualitative researchers*. Sage Publications, 2021.

[25] Gianluca Stringhini, Pierre Mourlanne, Gregoire Jacob, Manuel Egele, Christopher Kruegel, and Giovanni Vigna. EVILCOHORT: Detecting Communities of Malicious Accounts on Online Services. In *24th USENIX Security Symposium (USENIX Security 15)*, pages 563–578, 2015.

[26] Sajedul Talukder and Bogdan Carbunar. AbuSniff: Automatic Detection and Defenses Against Abusive Facebook Friends. In *Twelfth International AAAI Conference on Web and Social Media*, 2018.

[27] Sajedul Talukder and Bogdan Carbunar. A study of friend abuse perception in facebook. *Transactions on Social Computing*, 1(1), 2020.

[28] K. Thomas, D. Akhawe, M. Bailey, D. Boneh, E. Bursztein, S. Consolvo, N. Dell, Z. Durumeric, P. Kelley, D. Kumar, D. McCoy, S. Meiklejohn, T. Ristenpart, and G. Stringhini. Sok: Hate, harassment, and the changing landscape of online abuse. In *IEEE Symposium on Security and Privacy (SP)*, pages 473–493, 2021.

[29] Erkam Uzun, Simon Pak Ho Chung, Irfan Essa, and Wenke Lee. rtcaptcha: A real-time captcha based liveness detection system. In *Proceedings of the Network and Distributed Systems Security Symposium*, 2018.

[30] Zarina I Vakhitova, Rob I Mawby, Clair L Alston-Knox, and Callum A Stephens. To SPB or not to SPB? A mixed methods analysis of self-protective behaviours to prevent repeat victimisation from cyber abuse. *Crime Science*, 9(1):1–18, 2020.

[31] Yang Wang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. Privacy nudges for social media: An exploratory facebook study. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 763–770. ACM, 2013.

[32] Pamela Wisniewski, Heather Lipford, and David Wilson. Fighting for My Space: Coping Mechanisms for SNS Boundary Regulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 609–618. ACM, 2012.

[33] Teng Xu, Gerard Goossen, Huseyin Kerem Cevahir, Sara Khodeir, Yingyezhe Jin, Frank Li, Shawn Shan, Sagar Patel, David Freeman, and Paul Pearce. Deep entity classification: Abusive account detection for online social networks. In *30th USENIX Security Symposium (USENIX Security 21)*, 2021.

[34] Chao Yang, Robert Chandler Harkreader, and Guofei Gu. Die Free or Live Hard? Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers. In *International Workshop on Recent Advances in Intrusion Detection*, pages 318–337. Springer, 2011.

[35] Haifeng Yu. Sybil Defenses via Social Networks: A Tutorial and Survey. *ACM SIGACT News*, 42(3):80–101, 2011.