Hello, Facebook! This is the stalkers' paradise!

Jinwoo Kim, Kuyju Kim, Junsung Cho, and Hyoungshick Kim
Sungkyunkwan University

Abstract

We introduce a new privacy issue on Facebook. We were motivated by the Facebook's search option, which exposes a user profile with his or her phone number. Based on this search option, we developed a framework to automatically collect Facebook users' personal data (e.g., phone number, location) by enumerating the (possibly) entire phone number range for the target area. To show the feasibility, we launched an attack for targeting the users who live in California, United States. Despite Facebook's best efforts to stop such attempts from crawling users' data with several security practices, 87,000 phone numbers were successfully tested and 20,371 actual users' personal data were obtained within a week by mimicking real users' search activities with three rogue accounts.

1 Introduction

Facebook is one of the most popularly used online social networking services, which has more than 1.94 billion monthly active users for March 2017 [1].

Naturally, Facebook has also become an attractive target of cyber crimes such as spam, phishing and misuse of personal data due to its popularity. For example, spammers often use tools and bots for harvesting people's contact information (e.g., phone numbers and email addresses) [4].

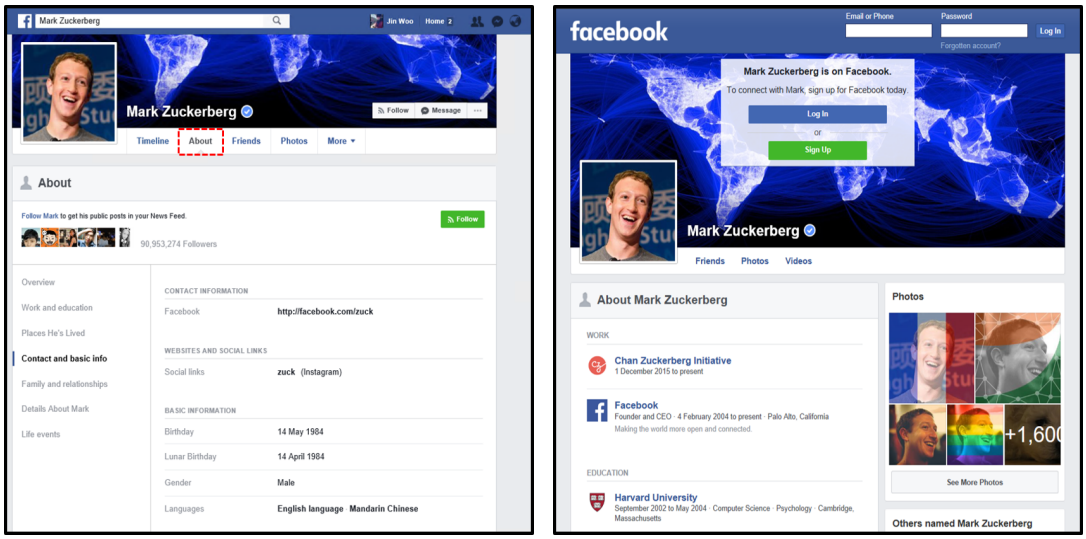
In this paper, we particularly focus on the security concerns raised by the friend search option with phone numbers, which is currently used in Facebook. Facebook provides various options to search for users. An option is to use a user's phone number. At first glance, this search option seems to be a proper compromise between privacy and utility, revealing a user's profile for his or her friends or acquaintances who only know the user's phone number. In this paper, however, we will show that this feature could potentially be misused for attackers who want to steal Facebook users' personal data such as their names, phone numbers, locations, education and even photos; those stolen data can be exploited for conducting additional cyber criminal activities such as spam, phishing and rogue accounts. To show the security risk of the search option, we developed a framework to automatically collect Facebook users' personal data (e.g., phone number, location) by enumerating the (possibly) entire phone number range for the target area. Our main contributions are summarized as follows:

- We introduce a novel *enumeration* attack using the search option by enumerating the (possibly) entire phone number range for the target area to harvest Facebook users' profile information including name, phone number, location, education, etc (read Section 3).
- We implement a automatic framework to perform our *enumeration* attack to show the feasibility and then analyze the collected user data. We also found how to bypass the defense mechanisms such as anomaly detection and CAPTCHA provided by Facebook (read Section 5).
- We discuss possible defense mechanisms to mitigate such *enumeration* attacks (read Section 6).

2 How to search for people on Facebook

To encourage a user to find and add other users as his/her friends (i.e., promotional purposes), Facebook provides several options to search for people on Facebook. To search for someone or some group/page, we can use a user's name, keyword, email address or phone number.

In general, user profile web pages display users' basic information such as work, education and location. Moreover, we found that the search results were significantly changed depending on whether we are logged in ("With login") or not ("Without login"). Figure 1 shows their differences. When we logged in (i.e., "With login"), the



(a) With login

(b) Without login

Figure 1: People search results ("With login" vs. "Without login").

search results include **About** which represents the user profile on Facebook. **About** page displays the detailed information about a user (e.g., the link to other services such as Instagram, birthday, gender, relationship with the user's partner, family members, life events, etc.). This information was not displayed when we did not logged in. Probably, Facebook believes that a user might be more trustworthy when the user already logins in. We will exploit this difference in displaying the search results to develop a framework for performing *enumeration* attacks on Facebook.

3 Enumeration attack using the people search with phone numbers

In this section, we present an automatic framework of *enumeration* attack to harvest user profile data using the search feature provided by Facebook. The proposed *enumeration* attack involves the following three steps: (i) enumerating target phone numbers in a random or sequential manner; (ii) checking whether the search results (including the user profile) are returned; (iii) extracting the interesting user data from the crawled user profile web page if the search results were correctly returned.

To conduct the *enumeration* attack using the people search with phone numbers, an attacker generates a range of phone numbers in a valid format and tries to search for people with that number. The phone numbers used for the *enumeration* attack can be generated for a specific target area. For example, the country code is 1 for United States; and the area codes are $\{209, 213, 279, 310, 323 \cdots\}$ for California. When the attacker wants to collect the information about the users who might live in California, the generated phone number format is like `+1209XXXXXXX`.

As described in Section 2, Facebook allows users to search for people with his or her phone number. Our framework performs the following procedure repeatedly (see also Figure 2):

1. An attacker signs in to Facebook using a rogue account because the search results are greatly changed depending on the login status as shown in Figure 1. For Facebook, a rogue account can simply be created by a temporary email service in an automatic manner.

2. A range of phone numbers in a valid format (e.g., `+1209XXXXXXX`) is generated and tries to search for people with that phone number on Facebook.

3. If the search results are successfully returned, the web page for user profile is crawled; otherwise, the below steps are skipped.

4. The crawled web page is parsed and the user data extracted from that web page is stored as the output of the attack.

As a result of our attack, an attacker can harvest victims' personal data (such as phone number, name, education level, the place user's lived, etc). Facebook already use the defense mechanism to protect their users' personal information. When a unusual or suspicious activity is detected, Facebook displays the "Security Check" error message, and asks the user to solve a CAPTCHA problem as shown in Figure 3.

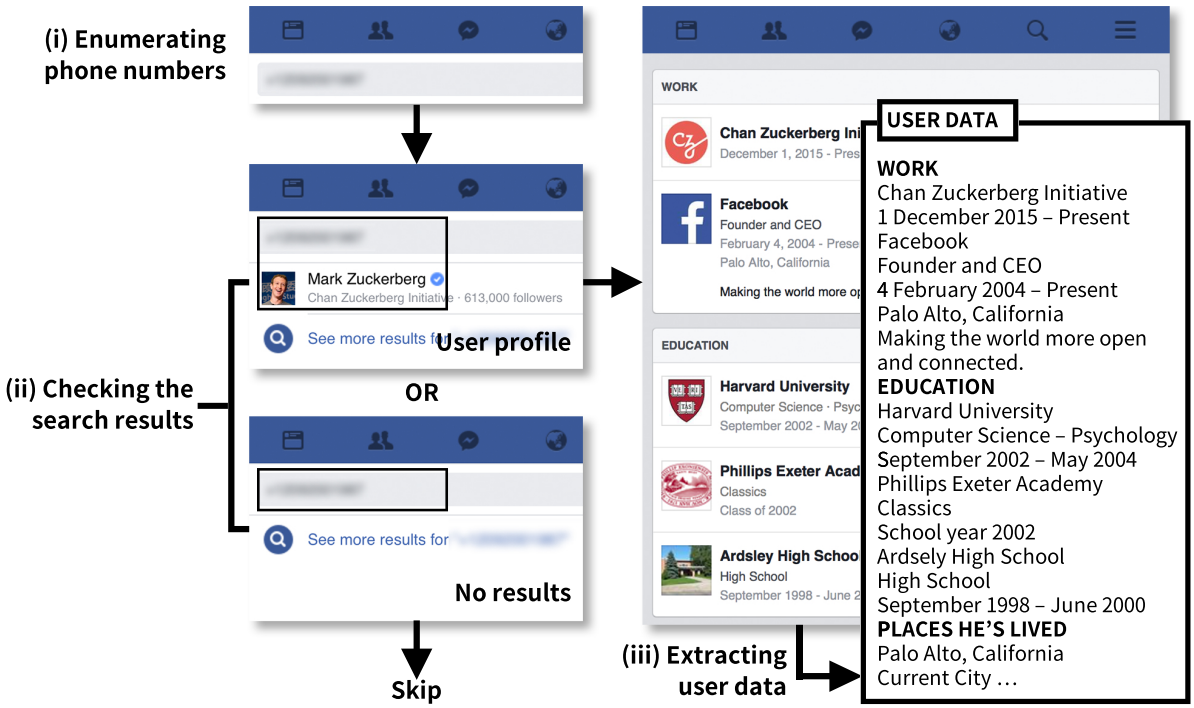


Figure 2: Overview of the enumeration attack using the people search with phone numbers.

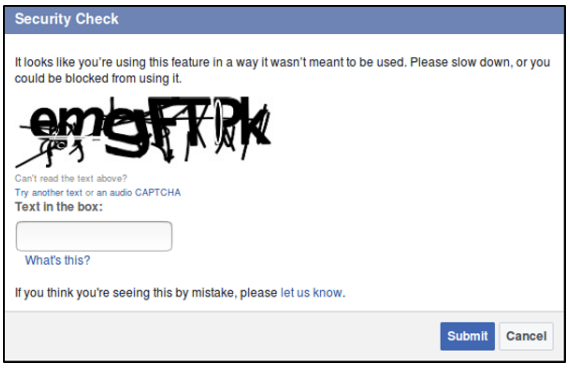


Figure 3: Example of CAPTCHA used in Facebook.

This policy seems effective against such *enumeration* attacks or web crawling. However, we found that this anomaly detection can be bypassed by using a few rogue accounts and performing attack attempts with an intentional delay (see Section 4).

4 How to evade the anomaly detection by Facebook

As mentioned in Section 3, we demonstrated our *enumeration* attack via people search option in Facebook.

We present our main idea to evade Facebook's defense mechanism in this section. To prevent personal data leakage from the service, Facebook has mechanisms for detection and protection against anomaly searching behavior. When someone searches many user profiles via people search by phone number quickly, Facebook generates CAPTCHA in page (see Figure 3) to check either these quick search behaviors are conducted by bot or normal user.

If an attacker tried to search people by phone numbers many times and CAPTCHA occurrence caused by these searches repeated for a long time, Facebook may occasionally block the searching ability of the user's account for a while (about one hour). In this case, we use the term block because even if we use a valid phone number to search for a user, Facebook search does not provide any profile information but only the message "There is no result" on some occasions.

Despite the existence of a such protection mechanism, because Facebook performed anomaly detection by account-based behavior, these defense mechanisms can be exploited easily by multiple accounts. In other words, if an attacker's attack was blocked (security checking occurred) by protection mechanism in Facebook, then the attacker changed his/her account to another and conducted attack continually, it is possible to conduct *enumeration* attack without blocking. Facebook can not aware that two attacks before blocking (security checking) and after blocking are conducted by the same attacker.

Before applying this vulnerability to the attack, we

attempted to generate CAPTCHA through a number of consecutive number searches in Facebook people search option. This test was conducted with 30 accounts in Facebook and the results of the test are as follows. For generating CAPTCHA the average number of search counts was 362.6, standard deviation was 78.82 and the minimum and maximum count were 300 and 694.

In order to bypass Facebook's protection method by blocking account, an attacker should make the *enumeration* attack look like a normal behavior as humans did it. Based on the previous experimental results, we set the threshold for changing the login session like *k=300*, this count was the minimum count of searches for generating CAPTCHA. For the infinite *enumeration* attack without any blocking, with the threshold in the test, we conducted an *enumeration* attack. The attack process is described in the Figure 4. In this attack, each account search people via consecutive phone numbers until the number of search reached the white circle (Threshold, 300). Once searched count reached the point of the threshold or CAPTCHA occurred in the searching process, the attacker should change account to another it. In this way, an attacker can bypass attack detection against Facebook.

5 Experiments

We implemented an *enumeration* attack framework prototype in Section 3 to show the feasibility against Facebook and evaluate its attack performance in the real-world environment.

5.1 Implementation

For the *enumeration* attack using the phone number search, we used a virtual machine (VMware Workstation 12.0.0) installed on an Ubuntu 16.04 LTS desktop computer (with two 2.7 GHz CPU and 2.4 GB RAM) and equipped with a non-congested 100 Mbit/s WiFi connection to a LAN that was connected to the Internet. We also used a software-testing framework called `Selenium` to

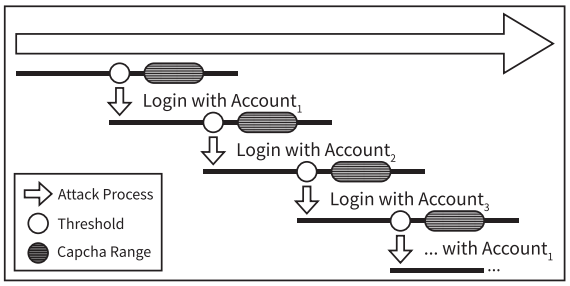


Figure 4: Using multiple rogue accounts to mimic real users' search activities.

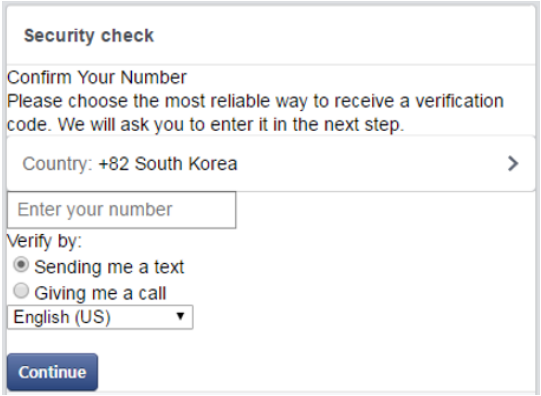


Figure 5: Security check via a user's phone number.

automate our *enumeration* attack attempts.

For conducting the *enumeration* attack, at least three authenticated (by e-mail or phone number) accounts on Facebook are necessary. But, Facebook service has a protection mechanism in the account creation phase as well as searching friend service. When we try to create more than one account in the same environment, Facebook service required users to authenticate by phone number additionally, like Figure 5.

For creating multiple Facebook accounts against from these protections, we used rooted mobile phone and tools for bypassing detection (e.g., DonkeyGuard, Android ID changer, IMEI Changer and Turbo VPN). It is possible to bypass detection of multiple account creation by changing some internal values in a smartphone. Once any identity checking occurs during the account creating process, we change the several values such as Android id and IMEI. We then continue to create more accounts using this method. The reason behind making changes to these values is to make Facebook aware that each of newly created accounts have been created in different environments. With these bypassing methods and temporary mail service, we can easily create multiple accounts for the attack with only mail authentication.

In the website-version of Facebook, a users information (e.g., birthday, gender, relationship, etc) is displayed in separate pages. In mobile version. However, all information is displayed in a single page. Due to this simple structure, crawling and parsing data from Facebook can be conducted much faster on mobile version compared to website-version.

5.2 Attack results

As a result of our attack, we were able to find 20,371 valid accounts out of 87,000 consecutive phone numbers used for searching during 7 days, and from those 20,371 accounts, we could successfully obtain a large amount of

Table 1: Summary of the collected users' personal data.

Types	Phone Numbers	Basic Info	Friends	Current City	Home Town	Education	Family Members	Relationship	Place User's lived
Count	20,371	19,249	14,702	10,338	9,701	8,732	6,770	6,336	11,723

personal data (The ratio of valid data from the searching to total searching number ,valid/total was 23.41%). In the process of deriving these results, the processing time was slightly different when the search results are valid and not. In the case of valid results, the time for searching and extracting data taken 4.78 seconds, when the searching results had not valid, the time for searching taken 6.49 seconds on average.

All data we have obtained from the performed attack have a plenty of information related to each target user. The table as shown in the Table 1, there are many critical personal data such as family members, friends, relationship, life events, workplace, level and place of education, residence location, etc. It is very important to notice that relationship, friends and family members in particular, have the profiles of other users connected with the user. These profiles can be our next possible targets of the attack [5].

In this work, we used three different accounts to bypass anomaly behavior detection against Facebook. In each of these three accounts, it took 35 minutes for searching 300 phone numbers on average. As described in Figure 4, after continuously searching for 600 phone numbers in other two accounts, an attacker needs to re-log in to the previous account and continue the searching process again from there. The average time taken to return to original account is 70 minutes.

6 Defences

In this section, we discuss some possible mitigation techniques against such attacks. We found that some protection mechanisms are already implemented in Facebook service. However, as we mentioned in Section 5, the protection mechanisms are not enough to protect our attacks, and personal data leakage is quite possible yet in Facebook with easy bypassing methods.

6.1 Changing privacy settings

Unfortunately, even if Facebook offers users sufficient options to protect them from personal data leakage, users often do not fully comprehend the security options, and are therefore still vulnerable to data leakage without even realizing it. Figure 6 shows the privacy settings in Facebook related with user personal data such as "Who can see my stuff?" and "Who can look me up?". The options

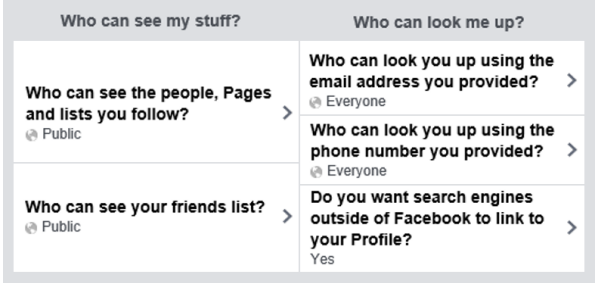


Figure 6: Privacy settings for personal data protection.

in this privacy setting (Everyone, Yes) are the default options those are set when users create the account. With these default options set, *enumeration* attack could be performed effectively for obtaining abundant personal data in the user's profile. In order for users to protect them against personal data leakage caused by malicious behaviors such as an *enumeration* attack, they need to realize that checking the privacy settings at Facebook security options is the top priority. The best way to protect user data against the attack is setting the options for viewing user information to Friends only or Only me, not Everyone or Public by default. These options are placed in the Facebook's menu, *Account Settings - Privacy*.

6.2 Disallowing multiple accounts

In our attack, we generate arbitrary rogue accounts to avoid protection mechanism such as CAPTCHA. Therefore, one of possible countermeasure is making an attacker hard to create multiple fake accounts.

The limitation of Facebook's protection mechanism is that the protection mechanism monitor the malicious behavior in only one account environment. Therefore, once an attacker has several accounts on Facebook, it is possible to perform *enumeration* attack continuously. In actually, explained in the Section 5.1, an attacker can create searchable new accounts easily. Authentication of newly created Facebook accounts can be implemented in a temporary mail service easily like Figure 7. The vulnerability like this can be exploited for *enumeration* attack. To solve this vulnerability, the service needs to have solutions for detecting the use of numerous accounts by one user. The more restrictions on creating multiple accounts, the more difficult for an attacker to perform an *enumeration* attack.

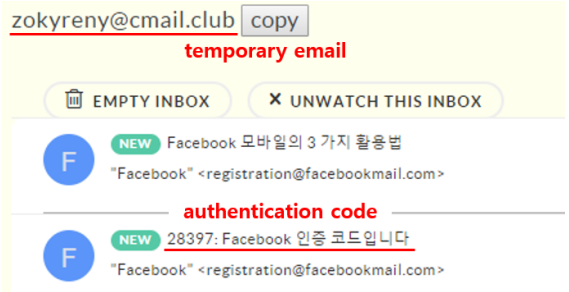


Figure 7: Use of a temporary email account for a rogue Facebook account.

6.3 Using dynamically changed webpages

In the *enumeration attack*, an attackers can infinitely crawl users personal data by leveraging the static sources of a Facebook web pages, such as ID and classes on Facebook. Designing Facebook web pages with dynamic structure can protect service against any malicious crawling behavior. We could find some elements in the Facebook's page sources such as *'mSideSearch', 'timelineBody'*. By using these elements in the Facebook web page, we implemented the *enumeration* attack (i.e., search people via phone number input, parsing users' personal data in profile) automatically.

For preventing automatic crawling against an attacker, Facebook need to change static elements in the page to be dynamic. It is possible to make these predictable elements in a page difficult to use by an attacker.

By using server-side scripts in the page, if the service is created with the user specific data based on dynamic resources which an attacker cannot predict in real-time, Facebook could be more secure service against *enumeration* attack.

6.4 Optimizing the detection of suspicious activities

As described in the Figure 1(a) and the right side of the Figure 2, the main target of *enumeration* attack is the user profile page in general. Hence, in here, conducting an *enumeration* attack also means that the searchers who are not in a friend relationship with users viewing the users' profiles indiscriminately. To prevent these *enumeration* attack, Facebook needs to create a policy in seeing profiles.

For a normal user, it is not uncommon to see a large number of profiles of anyone who are not the friend of the user in this social networking by using Facebook service. By limiting the number of profile page requests to user profiles in a specific time unit, it is possible that once unnecessary access to a profile are found, service providers can block this malicious behavior at the web application firewall level.

7 Ethical issues

The objective of our work was not obtaining personal information data by using the people search option on Facebook. Our academical motivation was evaluating the possible threat which can occur when abusing this social network service and suggesting the countermeasure against any malicious behavior.

In the services which having a huge number of users in it, exploiting the service for any data leakage may cause the critical privacy issues. After this work had completed, hence, all obtained user data were discarded.

At the end of the study, we proposed the ideas for preventing any attack using this service. The countermeasures for protecting service against our attack methods are suggested in the Section 6.

8 Related work

The *enumeration* attack proposed in our work is still available because many users in Facebook overlook the necessity of changing privacy settings provided in Facebook and leave their setting in default status like Public.

According to the work implemented by Malduzzi, Marco, at el [2], a small number of users who can see their profile by changing their privacy settings. Because a default value for the privacy setting is public and many users do not change default privacy settings. Actually on Facebook, there are more than 99% open profiles.

In previous works, there had been research on *enumeration* attacks using the telephone numbers in messenger by using address book synchronization functions. Schrittwieser et al [9]. analyzed the enumeration attack in WhatsApp. They uploaded 10 million phone numbers on the WhatsApp server and then took 21,095 valid numbers via WhatApp application within 2 hours.

As a similar study with [9], Kim et al [7]. implemented an *enumeration* attack via contacts sync in the Kakaotalk, which has a huge share in South Korea. And finally, they obtained numerous personal information from it successfully. We have expanded this research into the social network service. The attack target in our work is Facebook, which has a large number of users around the world. Because of Facebook's popularity, the amount of personal data included in this social network service is also enormous.

And there was also an email *enumeration* attack on Facebook. Balduzzi, Marco, et al [2]. collected mass mailings using automatic queries from social network service providers. By using 10.4 million emails for searching

input, they conducted *enumeration* attack, then automatically collected 1.2 million user profiles and personal information. Proposed attack in this work already fixed by Facebook, but it also can be conducted using our bypassing methods described in the subsection 4.

Previous researches on the leakage in personal information about social network services have been going on steadily. Mahmood at al [8]. analyzed that personal information leaks on Facebook and Twitter with 1 billion active users. In particular, on Facebook, they analyzed how email addresses are mapped to real names through account recovery service, and how to rebuild a friend list. Bing-Zhe He et al [6]. showed that the attacker can create a false identity account for that user by collecting the victim's personal information on Facebook.

Recently, Inti De Ceukelaire [3] proposed the small attack for finding someone's phone number by using contact sync in Facebook service. In his work, he tried the experiments for discovering someone's phone number by only using Facebook service. And finally, he has found the person whose phone number is unknown successfully. These attacks for obtaining someone's personal data is still quite possible now. Many people live in the social network world tend to overlook the abundant user data in Facebook and do not try hard to hide their personal data on Facebook.

9 Conclusion

This paper analyzes a security issue about the people search option provided by Facebook, which is the most popular social networking service worldwide. The people search option with phone number could potentially be misused to leak a user's sensitive personal data. Based on this feature, we developed a framework to automatically collect Facebook users' personal data by enumerating all the valid phone numbers for the target area. To show its feasibility, we implemented an attack for targeting the users who live in California, United States. Our implementation can bypass the Facebook's defense mechanisms; 87,000 phone numbers were tested and 20,371 Facebook users' personal data were collected within a week.

To mitigate such enumeration attacks, we also suggest several practical defense mechanisms. As part of future work, we plan to implement those mechanisms and evaluate their performance against the attacks.

References

[1] Number of monthly active facebook users worldwide as of 1st quarter 2017 (the statistics portal, statista). `https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/`.

[2] BALDUZZI, M., PLATZER, C., HOLZ, T., KIRDA, E., BALZAROTTI, D., AND KRUEGEL, C. Abusing social networks for automated user profiling. In *Proceedings of the 13th International Symposium on Recent Advances in Intrusion Detection* (2010).

[3] CEUKELAIRE, I. D. How i got your phone number through facebook. `https://hackernoon.com/how-i-got-your-phone-number-through-facebook-223b769cccf1/`.

[4] GAO, H., HU, J., WILSON, C., LI, Z., CHEN, Y., AND ZHAO, B. Y. Detecting and characterizing social spam campaigns. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement* (2010).

[5] HALEVI, T., LEWIS, J., AND MEMON, N. D. Phishing, personality traits and facebook. *Computing Research Repository abs/1301.7643* (2013).

[6] HE, B.-Z., CHEN, C.-M., SU, Y.-P., AND SUN, H.-M. A defence scheme against identity theft attack based on multiple social networks. *Expert Systems with Applications 41*, 5 (2014), 2345–2352.

[7] KIM, E., PARK, K., KIM, H., AND SONG, J. Design and analysis of enumeration attacks on finding friends with phone numbers: A case study with kakaotalk. *Computers & Security 52* (2015), 267–275.

[8] MAHMOOD, S., AND DESMEDT, Y. Your facebook deactivated friend or a cloaked spy. In *Proceedings of the 10 th IEEE International Conference on Pervasive Computing and Communications Workshops* (2012).

[9] SCHRITTWIESER, S., KIESEBERG, P., LEITHNER, M., MULAZZANI, M., AND HUBER, M. Guess whos texting you? evaluating the security of smartphone messaging applications. In *Proceedings of the 19th Annual Symposium on Network and Distributed System Security* (2012).