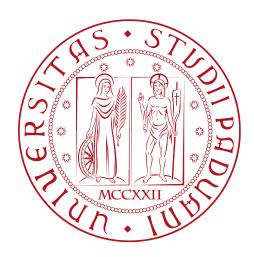# Università degli Studi di Padova

Dipartimento di Matematica "Tullio Levi-Civita"

Corso di Laurea in Informatica

Facebook Friend Requests' Analysis: development of tools for data collection and analysis

Tesi di laurea

Relatore
Prof.Mauro Conti

Laureando
Giada Zuccolo

Non come chi vince sempre
ma come chi non si arrende mai.

— Frida Khalo

Sommario

Il presente documento descrive il lavoro svolto durante il periodo di stage, della durata di circa trecento ore, dal laureando Pinco Pallino presso l'azienda Azienda S.p.A. Gli obbiettivi da raggiungere erano molteplici.

In primo luogo era richiesto lo sviluppo di ... In secondo luogo era richiesta l'implementazione di un ... Tale framework permette di registrare gli eventi di un controllore programmabile, quali segnali applicati Terzo ed ultimo obbiettivo era l'integrazione ...

Ringraziamenti

Innanzitutto, vorrei esprimere la mia gratitudine al Prof. Mauro Conti, relatore della mia tesi, per l'aiuto e il sostegno fornitomi durante la stesura del lavoro.

Per prima cosa voglio ringraziare la mia famiglia per il costante sostegno e per essermi stati vicini in ogni momento, sia gioioso che difficoltoso, durante questi gli anni di studio.

Voglio poi ringraziare i miei amici e compagni per questi anni passati insieme, da vicino e da lontano. In particolare ringrazio Sofia, per essere sempre stata una spalla e soprattutto un'amica su cui contare.

Ringrazio tutti i miei amici che sono qui a condividere questa gioia con me, soprattutto Clarissa, per l'appoggio che non mi ha fatto mai mancare e per la sua preziosa amicizia che per me è sempre stata un punto fermo.

Infine ringrazio Andrea, per avermi aiutata a vedere e a tirare fuori sempre il meglio di me anche quando io non ci riuscivo, per avermi supportato (e sopportato) fin dal primo istante e per tutto l'amore che mi dimostra.

Padova, Settembre 2021   Giada Zuccolo

Indice

Elenco delle figure

Elenco delle tabelle

Capitolo 1

Introduction

The exponential technological development and the internet have proliferated new criminally relevant conduct: new and more powerful IT tools have created new ways of committing a crime and have generated criminal phenomena.

For this reason, the area of primordial interest was the cybercriminal reality linked to the growing and hyperbolic development of the web and the technologies connected to it and how the types of crime (and criminals) and the most well-known and ordinary modus operandi have changed their physiognomy and have evolved with it.

Just think of online social networks and how they have an increasing impact on everyday life. Nowadays, social networks occupy a considerable part of daily life: they allow you to share links, photos, videos, thoughts, and to show your interest in some brands or companies. Many companies keep their customers updated through these channels, which can even be very useful for the growth of their customers, with shrewd and studied marketing skills. In a more circumstantial reality, social networks allow you to stay in touch with friends near and far and favor the birth and development of relationships even with unknown users.

It's precisely in this context that the project takes shape.

To start a relationship between two users, they must make contact, in some ways: generally or with a follow action (without some permissions) or with a friend request, where the user that receives the friend request has to accept (or not) the request to make contact with the other user.

In the case of the social network par excellence, namely Facebook, a profile A asks for friendship to a profile B, who can choose whether to accept, leave pending or reject.

In fact, the person behind a virtual profile could be unknown: a person may have created a profile that does not belong to any real person but simulating its existence, to get in touch with some specific users.

This mechanism has led to the birth of online crime aimed at targeting specific people who, initially unaware, find themselves becoming *victims*.

From here, a first project idea was born: put the focus from the attacker's point of view and then create a tool to analyze a victim and suggest to the criminal a series of information be used to create a profile that the victim would more likely accept.

1.1 Organizzazione del testo

Il secondo capitolo descrive ...

1

Il terzo capitolo approfondisce ...

Il quarto capitolo approfondisce ...

Il quinto capitolo approfondisce ...

Il sesto capitolo approfondisce ...

Nel settimo capitolo descrive ...

Riguardo la stesura del testo, relativamente al documento sono state adottate le seguenti convenzioni tipografiche:

* gli acronimi, le abbreviazioni e i termini ambigui o di uso non comune menzionati vengono definiti nel glossario, situato alla fine del presente documento;

* per la prima occorrenza dei termini riportati nel glossario viene utilizzata la seguente nomenclatura: *parola*[g];

* i termini in lingua straniera o facenti parti del gergo tecnico sono evidenziati con il carattere *corsivo*.

Capitolo 2

Description of the internship

In this chapter, there will explain the internship: the planning of work, the expected products, the objectives and methods of carrying out the activities

2.1 A brief introduction to the project

The general idea is studying the phenomenon of friend requests on Facebook, and how a profile with specific characteristics is accepted more easily by another profile with certain characteristics.
These characteristics, chosen after in-depth research in the literature, are defined by parameters (e.g. age, occupation, profile picture), which are divided into classes (e.g. age's class are defined by ranges, such as "age under 18", "age between 18 and 50 years old" or "over 50 years old").
Therefore, there will be developed specific tools capable to create a fake attacker profile with precise parameters and then send a friend request from this profile to more than one victim profile.
The objective is to create a decided number of attacker profiles, based on the various combination of the parameters that we chose to analyze; then, all of these profiles must ask a friend request to a specific number of victim profiles which are categorized into configuration based on the various combination of the parameters that we have decided to analyze. In this way, we are sure that every type of attacker will ask at least one friend request to every type of victim.
Consequently, an amount quantity of data will be collect and analyzed, in such a way as to be able to define a valid **model of acceptance**.
This is necessary to satisfy the concluding goal: define a functioning prototype for a final tool, which, after entering the URL of the victim, will take care of determining the characteristics that the attacker should include in his profile to be more likely accepted by the victim.

2.2 Expected products

The student had to keep track of the work done and the progress made against the expected progress day by day.

At the end of the internship period, the student had to produce a written report that keeps track of the work done and describes the results obtained, in particular:

1. what has been learned from the study and research in literature;

2. the study, development, and execution of the tool for automating the search for a profile with certain characteristics;

3. the study, development, and execution of the tool for the automation of the creation of an attacking profile with certain characteristics;

4. the study, development, and execution of the tool for the automation of friend requests from a profile with certain characteristics to another profile with certain characteristics;

5. consecutive data collection;

6. a preliminary analysis of the collected data, aided by some tools developed ad hoc;

7. a prototype of a model of acceptance.

2.3 Objectives

Now will be listed the objectives to achieve.

* implementation of the automatic search tool for people, given certain characteristics as input;

* implementation of creation attacker profile tool, giving as input the characteristics that it must have;

* implementation of the friend request tool from a profile to another profile, giving the necessary data as input;

* preliminary analysis on previously collected data.

2.3.1 Learning objectives expected

During this internship the student will have the opportunity to deepen and put his knowledge into practice, in particular:

* in the field of the automated management of browsers through the use of the Selenium framework;

* Python programming;

* tools and methodologies for data collection;

* statistical techniques and tools for the study and analysis of the collected data.

Capitolo 3

Background

Python

Selenium

Facebook

Facebook is a website providing social network service, launched in February 2004, operated and privately owned by Facebook Incorporation. Its goal is to give people the power to share, and make the world more open and connected. Facebook users may create a personal profile, add other users as friends, and exchange messages (including automatic feed notifications when they update their profile information. Additionally, users may share their status, news stories, notes, photos, videos, and allow their friends (or friends of friends) to comment on them. Furthermore, users may join common-interest groups, organize events, and create fans pages for a workplace/business, a school/college, or even a brand/product. However, it is unavoidable that this platform may also provide incentives for criminals to carry out illegal activities.

Capitolo 4

Literature search

The research in the literature and the study of founded projects was useful for the knowledge of the possible areas and above all, it was necessary to verify that no other project had as a focal point the attacker who wants to hit a specific victim and must make sure to be accepted as a possible friend.
Furthermore, these projects were useful for delineating the parameters from which to structure the attackers' and victims' profiles.

4.1   List of papers

Many papers are dealing with this social network, each of which opened to the perspective of many other projects. The most important and useful papers found in the great research phase are now shown and discussed:

1. INVESTIGATIVE TECHNIQUES IN THE DIGITAL AGE - CYBER-CRIME AND CRIMINAL PROFILING
   In this paper, it is made clear how cyberspace offers the possibility of carrying out illicit acts with the perception of going unpunished, showing the new various types of crimes such as cyberstalking, cyberbullying, online sexual offenses, etc. Cyberspace has allowed the criminal to evolve the techniques and approaches to the victim, so much so that it is not always immediately obvious that the intentions of one account towards another can be malicious.
   This new world, moreover, places investigators to consider the crime scene differently: a crime born of a victim's approach on the web places the computer itself as a victim or as a witness, but above all as a crime scene and therefore as a space to be analyzed. that could help (or penalize?) the investigators.
   LINK: https://oapub.org/soc/index.php/EJSSS/article/view/821/1403

2. OSNs AS SUPPORTING EVIDENCE: A DIGITAL FORENSIC INVESTIGATION MODEL AND ITS APPLICATION DESIGN
   This paper states that digital crime investigations are performed without adequate guidelines, as there isn't a consistent standard and model, only a set of procedures and tools, and above all, there is no model built specifically for social networks. This paper, therefore, proposes a standard survey model to be used for social networks, incorporating existing traditional frameworks and strategies..
   LINK: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6125728

3. INTEGRO: LEVERAGING VICTIM PREDICTION FOR ROBUST FA-
KE ACCOUNT DETECTION IN OSNs
The *Integro* project is a scalable defense system that helps OSNs detect fake
accounts using a meaningful user classification scheme. This paper, in addition to
providing an explanation of how *Integro* works, provides interesting explanations
on the behaviors that differentiate the attacker and the victim.
LINK: https://www.sciencedirect.com/science/article/pii/S0167404816300633

4. SOCIALSPY: BROWSING (SUPPOSEDLY) HIDDEN INFORMATION
IN OSNs
This paper highlights how current privacy settings in social networks are not as
effective as users might think, focusing on Facebook. It shows how easy it is to
retrieve information that a user thinks they have set as private.
LINK: https://arxiv.org/pdf/1406.3216.pdf

5. EVALUATION OF THE LIKELIHOOD OF FRIEND REQUEST AC-
CEPTANCE IN OSNs
This paper explains how OSN users often run into breaches or security issues
due to rash acceptance of the friend request, which can lead to the disclosure of
personal information and vulnerability to an attack. The document proposes a
method to evaluate the probability of becoming a friend having defined a model
data: a future friend and incoming friend requests are evaluated with reference
to this model which takes into account the attributes (such as common interests)
and behavioral properties (such as seat frequency).
LINK: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8731968

6. CAN FRIENDS BE TRUSTED? EXPLORING PRIVACY IN OSNs
This article presents a case study describing the privacy and trust inside a small
population of online social network users. Taking Facebook as a reference, the
frequency with which people are willing to disclose personal data to an unknown
online user was determined. While most of the users sampled did not share
sensitive information when requested by the stranger, it turned out that several
users were willing to divulge personal details to a stranger if there is a mutual
friend.
LINK: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5231850

7. TO BEFRIEND OR NOT: A MODEL OF FRIEND REQUEST ACCEP-
TANCE
This paper was the most useful, as it provided a valid model for the choice of
parameters: it explains how a friend request acceptance model was developed
that explains how various factors influence user acceptance behavior. This paper
highlighted how there are 4 decision-making factors to which a user appeals
when he has to accept a friend request, in particular, **"friendship factors"**,
i.e. all the visible aspects of the profile (name, gender, profile picture, city of
origin, residence, school, mutual friends, interests, etc.); **"privacy factors"**,
dictated by a personal concern of a user and/or awareness of her, perhaps also
due to personal experiences or friends; **"environmental factors"**, when friend
requests are accepted without actually considering those who arrive due to lack
of concentration and/or time to check; in the end **"capacity of interface"**,
that means that some types of users spend a lot of time to understand what
kind of person has asked them for friendship, with careful evaluation of all the

information and photos of that profile, until sending a private message. Some users keep the friend request pending and reserve the right to check the profile occasionally, to see changes within it.

LINK: https://www.usenix.org/conference/soups2014/proceedings/presentation/

4.2 Summary and starting points

Continuing with the research, we have increasingly focused on the type of projects that could have turned out to be similar to the idea of putting oneself on the side of the attacker who wants to hit a specific victim and must ensure that he is accepted as a possible friend.

All the papers found were useful to outline the parameters from which to start to structure the attacking and victim profiles.

AGGIUNGEREEEE

Capitolo 5

Progettazione e codifica

Breve introduzione al capitolo

5.1 Definition of the project's line guide and the chosen parameters

AGGIUNGERE CONTENT

The attacker knows that it is not easy to be accepted by his victim and, for this reason, he wants to understand what choices he must take to create the best profile possible to be accepted without too much delay by the victim. Therefore, the parameters must be defined base on the victim, to be accepted with greater possibility.

Thanks to the study of some specific papers mentioned above and thinking about how much to deepen the search for the best profile, the chosen parameters are now defined and discussed.

5.1.1 1st parameter: gender

Gender is an almost decisive factor. Many papers of those cited report that a friend request from a female user profile is accepted easier than a friend request from a man user profile, regardless of the gender of the profile that receives it.

Classes

The parameter `gender` ∈ {`female, male`}, where:

* `female/F`, if the victim's profile user declares to be a *woman*;

* `male/M`, if the victim's profile user declares to be a *man*.

5.1.2 2nd parameter: image profile

The profile picture is the first thing, along with the name, a user sees when they have a friend request. As reported in some above-mentioned papers, the profile image already serves to give an idea of the person who is asking for friendship, because it shows some details that do not need to be checked (such as gender or an indicative age range). A

profile with a hidden or fake image (eg. a landscape) is more mysterious in the eyes of
a user who receives the friend request, beacuse the real person behind this profile is
not clearly identifiable.

Therefore, the main idea is to verify the presence of at least one person in the profile
picture and classify the user profiles based on the outcome of this verification.

Facebook uses a technology that allows recognizing the profile image's content (faces,
objects, ...), and automatically creating a description that will be enclosed in the `alt`
tag. The presence of one or more people (and occasionally other details about the
panorama and/or some objects present) is always specified within the `alt` tag. On
the other hand, if the image is fictional (cartoons, drawings, etc.) or is an image
without a person (eg. landscapes), the `alt` tag contains "`No description of the
photo available`".

This technology has been exploited to classify each victim's profile picture.

Classes

The parameter `real_img` $\in$ {`true`,`false`}, where:

* `true`, if in `alt` tag reported the presence of one or more people;

* `false`, if in `alt` tag not reported the presence of one or more people;

5.1.3   3rd parameter: age

Age is a very relevant factor. From the literature, it appears to be one of the first
factors that a user who receives a friendship goes to check the profile from which the
friend request started, if the age is not deducible from the image profile. The more
similar age, the more likely it is that the friend request will be accepted.

A person who wants to register on Facebook, must declare a real birth date (and must
be at least 13 years old) and, once the profile has been created, can choose whether
to make the date public or not. Some profiles publish only the month and day, other
profiles only the year, others the complete date, others hide everything.

To classify profile user's age, the date of birth that it reports will be taken, and the age
will be calculated based on the current year, so as to be able to assign to this profile
its membership class.

Classes

Age is classified into 3 ranges, so the parameter `age_range` $\in$ {2, 3, 4}, where:

* 2, if the age victim's profile user is between 18 and 50 years;

* 3, if the age victim's profile user is greater than 50 years;

* 4, if the age victim's profile user is hidden;

5.1.4   Consideration

Other parameters that could be interesting to study are "working occupation" and
"current town". In particular, often the friend request is dictated by work/study
interests, so checking the job occupation that the profile declares to employ. Same for
the current town: maybe a profile is more secure to accept friend requests only for
those profiles that indicate a current town like its o near its, or is this indifferent?

Capitolo 6

Verifica e validazione

Capitolo 7

Conclusioni

7.1   Consuntivo finale

7.2   Raggiungimento degli obiettivi

7.3   Conoscenze acquisite

7.4   Valutazione personale

Appendice A

Appendice A

Bibliografia