ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE

CORSO DI LAUREA IN INFORMATICA

PRIVACY SU FACEBOOK

TESI DI LAUREA IN
ARCHITETTURA DEGLI ELABORATORI

Relatore:
Prof. Vittorio Ghini

presentata da:
Simone Salucci

Sessione III
Anno Accademico 2011/2012

Indice

Capitolo 1

Introduzione

Da qualche anno, ormai, i social network hanno fatto irruzione su internet diventando col tempo sempre più utilizzati e sollevando di conseguenza dubbi sempre maggiori sulla sicurezza della privacy.
Tutti i social network si presentano come piattaforme in cui gli utenti possono interagire con amici, parenti o sconosciuti con cui si condividono gli stessi interessi, sviluppando così relazioni personali e, in alcuni casi, professionali.
Questo rapido ed esponenziale sviluppo ha attirato l'attenzione, non solo degli utenti, ma anche di malintenzionati.
Ormai la maggior parte degli utenti della rete reputa i social network uno strumento utile e indispensabile per stringere e mantenere legami nella vita di tutti i giorni, ma il prezzo da pagare potrebbe rivelarsi piuttosto caro.
Nel 2010 Andrea Vitali, nella sua tesi dal titolo “L'11 settembre telematico”, ha presentato un algoritmo di attacco che sfruttando le informazioni reperibili sui social network, avrebbe potuto portare alla violazione della privacy dell'intera popolazione del web 2.0 (oltre 500 milioni di utenti) e ad una violazione totale del 14% di questi, circa 70 milioni di persone.
Lo scopo di questa tesi è valutare se questo algoritmo sia valido ancora oggi o se le nuove politiche riguardo alla privacy di Facebook non ne permettano più l'utilizzo, valutando anche la possibilità di migliorare l'efficienza di questo attacco.
Dopo questa breve introduzione, nel secondo capitolo verrà illustrato lo

scenario di partenza per poi dettagliare, nel capitolo successivo, il funzionamento dell'algoritmo di attacco da esplicitare, nel capitolo quarto, gli obiettivi di questo elaborato.

Nel quinto capitolo partendo da un'analisi sulla sicurezza delle password relativamente alle abitudini degli utenti stessi, si passerà poi a una valutazione dei problemi che le nuove politiche di privacy e sicurezza di Facebook potrebbero creare all'utilizzo dell'algoritmo di attacco per poi trovare dei metodi per attualizzarlo ed una tecnica per migliorarne le prestazioni.

Capitolo 2

Scenario

2.1 Internet al giorno d'oggi

L'utilizzo di internet e la sua penetrazione all'interno della popolazione cresce di anno in anno: 428,9 su 652,1 milioni di persone in Europa sono online. Secondo i dati di una ricerca del 2012, eseguita da Mediascope Europe[6], la media europea di tempo trascorso online è di 14,8 ore la settimana, tendenza che ha registrato una crescita del 10% rispetto al 2010 e questo dato è destinato a crescere.

Oggi grazie alla diffusione sempre maggiore di tablet ed in particolare degli smartphone, internet è sempre più a portata di mano tanto che si conta che il numero di dispositivi connessi ad internet abbia superato il numero delle persone.

La penetrazione di internet all'interno della popolazione europea ha raggiunto il 65%, 19% in più rispetto al 2010.

Attività come guardare la televisione, ascoltare la radio, leggere le notizie o i giornali vengono svolte sempre più online: il 73% degli utenti di internet guarda la tv online, il 67% ascolta la radio online, il 91% legge le notizie sul web.

Solo in Europa il 37% degli utenti accedono ad internet utilizzando più di un device; nella maggior parte dei casi si tratta di pc e smartphone, ma da qualche anno anche le console di intrattenimento come Playstation e Xbox hanno

accesso ad internet così come i televisori di ultima generazione.

Appare quindi evidente che con il passare del tempo il numero di dispositivi collegati è destinato a crescere; se da un lato questa sempre maggiore connettività porta vantaggi e comodità, dall'altro vi è un aumento dei pericoli per la sicurezza di questi dispositivi e per la riservatezza dei dati sensibili.

2.2 I social network

Da qualche anno i social network sono sempre più popolari, se ne parla ovunque, dalle televisioni ai giornali, influenzando sempre di più la vita di tutti; Facebook, il più utilizzato tra i social network, da solo ha superato il miliardo di utenti, tanto che se fosse una nazione sarebbe la terza più grande del mondo, pari a due volte la popolazione degli Stati Uniti.

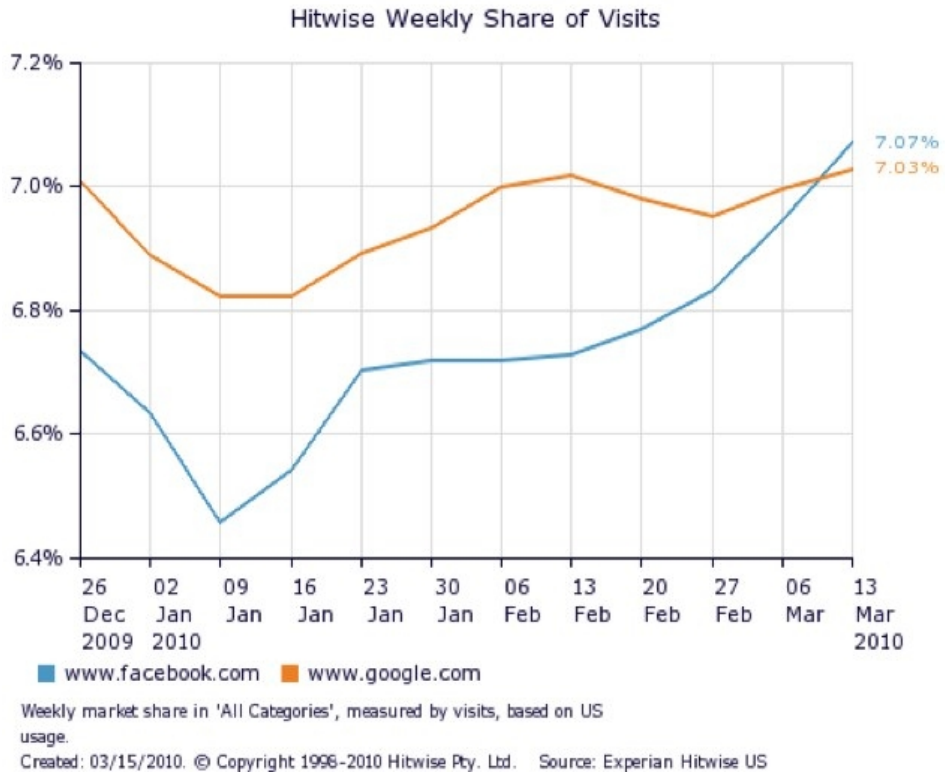

Figura 2.2.1: Grafico traffico Facebook vs Google

Nel marzo del 2010, come mette in evidenza il grafico, Facebook ha superato Google a livello di traffico diventando così il sito più visitato degli Stati Uniti; inoltre il tempo trascorso in media sui social network, in particolare su Facebook, ha superato il tempo trascorso su qualsiasi altro sito, Google compreso.

Gli americani passano in media 7,45 ore al mese sui social network, questi ormai influenzano sempre più le attività di tutti i giorni, infatti mentre guardiamo una serie televisiva veniamo invitati a commentarla sui social network, in particolare su twitter con un hashtag specifico.

Nell'ultimo anno anche il Papa ha attivato un account su Twitter e nelle ultime elezioni politiche in Italia, buona parte della campagna elettorale è stata portata avanti sui social network tanto che anche nei manifesti elettorali affissi per le strade era possibile vedere i simboli di Facebook e Twitter.

Nei videogiochi si ha la possibilità di condividere i propri punteggi sui social: addirittura nel controller della nuova playstation sarà presente il tasto "share" per condividere in tempo reale i propri punteggi.

I social network fanno parte della nostra vita e saranno sempre di più utilizzati per qualsiasi scopo.

List of continents on Facebook

#	Continent	Users	Penetration
1	Asia	267 816 640	6.61%
2	Europe	249 999 040	30.20%
3	North America	236 524 740	43.81%
4	South America	144 406 140	36.41%
5	Africa	50 438 620	4.97%
6	Oceania	14 627 180	39.47%

Figura 2.2.2: Statistiche utenti Facebook per continenti

Last 6 months | Last 3 months | Last month | Last 2 weeks | Last week

#	Country	Number of Users	Change	(± %)	Penetration
1.	United States	163 071 460	+1 509 120	+0.91%	52.56%
2.	Brazil	66 552 420	+9 847 500	+14.80%	33.09%
3.	India	61 499 220	+7 835 680	+12.74%	5.24%
4.	Indonesia	47 165 080	+2 998 420	+6.37%	19.41%
5.	Mexico	39 945 620	+2 368 620	+5.93%	35.52%
6.	Turkey	32 438 200	+1 384 020	+4.27%	41.69%
7.	United Kingdom	32 175 460	+459 100	+1.43%	51.61%
8.	Philippines	30 094 560	+989 660	+3.29%	30.12%
9.	France	25 307 820	+661 740	+2.61%	39.07%
10.	Germany	25 063 880	+763 540	+3.05%	30.64%
11.	Italy	23 028 220	+1 197 400	+5.19%	38.16%
12.	Argentina	20 403 520	+676 140	+3.30%	49.35%

Figura 2.2.3: Statistiche utenti Facebook per nazioni

I grafici riportano alcuni dati riguardo al numero di utenti di Facebook presi dal sito www.socialbakers.com[7].

2.3 I rischi dei social network

All'interno dei social network sono memorizzate moltissime informazioni personali che rischiano di diventare accessibili a sconosciuti e a potenziali malintenzionati.

Questa problematica è stata affrontata in una Risoluzione dei 78 garanti mondiali per la privacy nel 2008, nella quale si legge:

"I dati personali divengono infatti disponibili pubblicamente e in modo globale, secondo schemi qualitativi e quantitativi che non hanno precedenti, anche attraverso enormi quantità di foto e video digitali".

"C'è il rischio di perdere il controllo dell'utilizzo dei propri dati una volta pubblicati in rete".

Infatti sui social network si possono reperire informazioni di vario tipo:

1. Dati anagrafici quali nome, cognome, sesso, data di nascita, città', residenza, etc.
2. Contatti: numero di telefono e di cellulare, email, sito web personale, instant messagging.
3. Percorso educativo: percorso accademico, titolo di studio, diploma / specializzazioni conseguite.
4. Altre info: orientamento politico, interessi generali, aggregazioni / gruppi.

Queste informazioni, che l'utente medio inserisce nel social network, pur apparendo di scarso interesse, possono però diventare molto pericolose nelle mani di malintenzionati che utilizzando tecniche di social engineering possono sfruttare questi dati per perseguire i propri scopi.

Di seguito riportiamo alcuni rischi presentati da Andrea Vitali nel suo lavoro di tesi[1] correlati all'utilizzo di servizi dei social network:

1. Niente oblio su Internet. Il concetto di oblio non esiste su Internet. I dati, una volta pubblicati, possono rimanerci letteralmente per sempre – anche se la persona interessata li ha cancellati dal sito "originario", possono esisterne copie presso soggetti terzi; appartengono a quest'ultima categoria i servizi di

archivistica e la funzione di "cache" disponibile presso un notissimo motore di ricerca. Inoltre, alcuni fornitori di servizi rifiutano di ottemperare (o non ottemperano affatto) alle richieste degli utenti di ottenere la cancellazione di dati e, soprattutto, di interi profili.

2. L'idea ingannevole di "comunità". Molti fornitori di servizi affermano di trasferire le strutture comunicative dal mondo "reale" al cyberspazio. Un'affermazione frequente è che non ci sarebbero problemi, per esempio, a pubblicare dati (personali)su queste piattaforme, perché è come se si condividessero informazioni con un gruppo di amici nel mondo reale.

Se però si vanno ad esaminare con più attenzione alcune caratteristiche di certi servizi, si vedrà che il parallelo non regge anche perché il concetto di "amici" nel cyberspazio può risultare assai diverso dall'idea più tradizionale di amicizia, e la comunità può essere assai estesa. Se non si informano gli utenti in modo trasparente sulle modalità di condivisione delle informazioni contenute nei loro profili, e sugli strumenti con i quali essi possono decidere tali modalità, può avvenire che l'idea di una "comunità" descritta nei termini sopra richiamati finisca per indurli a rivelare in modo sconsiderato informazioni personali che altrimenti non si lascerebbero sfuggire. Anche i nomi dati a talune di queste piattaforme come "MySpace", il mio spazio, creano un'idea illusoria di privacy e riservatezza sul web.

3. "Gratis" non sempre significa "a costo zero". In realtà, molti dei servizi di social network fanno "pagare" gli utenti attraverso il riutilizzo dei dati contenuti nei profili personali da parte dei fornitori di servizio, ad esempio per attività mirate di marketing.

4. La raccolta di dati di traffico da parte dei fornitori di servizi di social network, i quali hanno gli strumenti tecnici per registrare ogni singolo passo dell'utente sul loro sito e, in ultima analisi, comunicare a terzi dati personali di

traffico compresi gli indirizzi IP, che in taluni casi possono ricordare i dati relativi all'ubicazione. Ciò può avvenire, ad esempio, per finalità pubblicitarie, anche di tipo mirato. Si osservi che in molti Paesi i dati in oggetto devono essere comunicati, a richiesta, anche alle autorità giudiziarie o di polizia e/o ai servizi di intelligence, nonché con ogni probabilità, in base alle norme esistenti in materia di cooperazione internazionale, a soggetti stranieri.

5. Il bisogno crescente di finanziare i servizi e ricavare profitti può fungere da stimolo ulteriore per la raccolta, il trattamento e l'utilizzazione di dati relativi agli utenti, trattandosi dell'unico cespite patrimoniale dei fornitori di servizi di social network. I siti di social network non sono, contrariamente a quanto suggerito dal termine "social", un servizio pubblico. D'altra parte, il web 2.0 sta "diventando adulto" e le piccole aziende informatiche gestite, in certi casi, da gruppi di studenti meno interessati all'aspetto finanziario sono sostituite sempre più spesso da grandi soggetti di respiro internazionale. Tutto questo ha cambiato in qualche misura le regole del gioco, visto che molte delle imprese di cui sopra sono quotate in borsa e subiscono una pressione fortissima da parte dei rispettivi investitori nell'ottica di realizzare e massimizzare profitti.

Poiché per molti fornitori di questi servizi i dati contenuti nei profili degli utenti ed il numero di utenti esclusivi costituiscono gli unici veri beni patrimoniali di cui dispongono, possono sorgere rischi ulteriori per quanto riguarda la raccolta, il trattamento e l'utilizzo non proporzionati dei dati personali relativi agli utenti.

6. Rivelare più informazioni personali di quanto si creda. Ad esempio, le foto possono trasformarsi in identificatori biometrici universali all'interno di una rete ed anche attraverso più reti. Negli ultimi anni sono migliorate in misura notevole le prestazioni dei software di riconoscimento del volto, e risultati

ancora "migliori" arriveranno in futuro. Si osservi che, una volta associato un nome ad una foto, possono essere messe a rischio anche la privacy e la sicurezza di altri profili-utente, magari basati sull'uso di pseudonimi o addirittura di dati anonimi. Inoltre, l'ENISA, l'agenzia europea per la sicurezza delle reti e delle informazioni, ha richiamato l'attenzione su una tecnologia emergente (CBIR, content-based image retrieval) che offre ulteriori opportunità di localizzare gli utenti associando gli elementi identificativi di determinati ambienti o luoghi, ad esempio, un dipinto appeso in una stanza, o un edificio visibile nell'immagine, ai dati relativi all'ubicazione.

Infine, le funzioni dette di "grafo sociale", molto diffuse presso vari servizi di social network, di fatto rivelano informazioni sui rapporti intercorrenti fra i singoli utenti.

7. Utilizzo improprio dei profili utente da parte di soggetti terzi. Si tratta probabilmente del rischio potenziale più grave per quanto riguarda i dati personali contenuti nei profili utente dei servizi di social network. A seconda della configurazione disponibile rispetto alla privacy e dell'utilizzo o meno di tale configurazione da parte degli utenti, nonché del livello di sicurezza offerto dal servizio, le informazioni contenute nel profilo, comprese immagini, che possono ritrarre sia il singolo interessato, sia altri soggetti, diventano accessibili, nel peggiore dei casi, all'intera comunità degli utenti. Allo stesso tempo, sono assai scarse le salvaguardie oggi disponibili rispetto alla copia dei dati contenuti nei profili-utente ed al loro utilizzo per costruire profili personali e/o ripubblicare tali dati al di fuori dello specifico servizio di social network. Tuttavia, anche l'utilizzo "normale" dei dati contenuti nei profili-utente può impattare sull'autodeterminazione informativa degli utenti e, ad esempio, incidere gravemente sulle loro possibilità di carriera. Un esempio che ha suscitato interesse diffuso riguarda l'abitudine da parte dei dirigenti del

personale di singole società di consultare i profili-utente dei candidati all'assunzione e/o dei dipendenti.

Secondo quanto riferito da articoli di stampa, già oggi i due terzi dei dirigenti ammettono di utilizzare i dati ricavati da servizi di social network, ad esempio per verificare e/o completare i curricula dei candidati. Altri soggetti che possono trarre profitto da queste fonti di informazione sono le forze dell'ordine e i servizi segreti. Inoltre, alcuni fornitori di servizi di social network forniscono a terzi dati relativi agli utenti tramite interfacce di programmazione di applicativi, e i dati finiscono quindi per essere gestiti dai soggetti terzi in questione.

8. Il Gruppo di lavoro nutre particolari preoccupazioni rispetto al rischio ulteriore di furti d'identità causati dalla disponibilità diffusa di dati personali contenuti nei profili-utente e dall'abuso di tali profili da parte di soggetti terzi non autorizzati.

9. Utilizzo di un'infrastruttura la cui sicurezza lascia purtroppo molto a desiderare.

Si è molto parlato della (non) sicurezza di reti e sistemi informatici, compresi i servizi web. Casi recenti in merito riguardano fornitori di servizi molto conosciuti quali Facebook, Flickr, MySpace, Orkut, e StudiVZ. E' vero che i fornitori di servizi hanno preso misure atte a potenziare la sicurezza dei loro sistemi, ma molto resta ancora da fare. Allo stesso tempo, è probabile che in futuro emergano nuove falle nella sicurezza di questi sistemi, mentre è assai improbabile che si possa mai conseguire l'obiettivo di una sicurezza totale vista la complessità delle applicazioni software a qualunque livello dei servizi Internet.

10. I problemi tuttora irrisolti per quanto concerne la sicurezza dei servizi Internet costituiscono un rischio ulteriore connesso all'utilizzo dei servizi di

social network e, in certi casi, aumentano il livello complessivo di rischio, ovvero comportano "sfumature" di rischio specifiche di questo tipo di servizi. In un documento recente redatto dalla ENISA (European Network and Information Security Agency) vengono citati, fra l'altro, lo spam, lo scripting fra siti diversi, virus e "vermi", il phishing mirato (spear-phishing) e forme di phishing specifiche dei servizi di social network, l'infiltrazione della rete, l'utilizzo abusivo di profili-utente (profile-squatting) e attacchi reputazionali basati sul furto di identità, forme di persecuzione personale (stalking), il bullismo in rete, e lo spionaggio industriale (ossia, i cosiddetti "social engineering attacks" (strategie basate su interazioni interpersonali finalizzate a carpire informazioni riservate) compiuti attraverso i servizi di social network). Secondo l'ENISA, un rischio ulteriore per la sicurezza è rappresentato "dai fattori di aggregazione legati alle social network".

11. L'introduzione di standard di interoperabilità e interfacce di programmazione applicazioni (API: ad esempio, lo standard "open social" introdotto da Google nel mese di novembre 2007), allo scopo di consentire l'interoperabilità tecnica di servizi di social network tipologicamente diversi, comporta tutta una serie di rischi ulteriori.

Si rende infatti possibile una valutazione automatica di tutti i siti di social network che utilizzino lo standard prescelto. Attraverso le API è in pratica l'intera gamma di funzionalità del sistema ad essere passibile di valutazione automatica attraverso l'interfaccia web. Le applicazioni potenzialmente in grado di interferire con la privacy degli utenti (e forse anche con la privacy di soggetti che non sono utenti, ma i cui dati facciano parte di un profilo-utente) comprendono, ad esempio, l'analisi complessiva dei rapporti professionali e privati intrattenuti dal singolo utente, che può senz'altro travalicare i "confini" delle singole social network sulle quali l'utente interagisce volta per volta in

ruoli diversi; inoltre, l'interoperabilità può favorire in misura ulteriore il riutilizzo da parte di soggetti terzi delle informazioni e delle immagini contenute nei profili-utente.

Capitolo 3

L'attacco

Un paio di anni fa è stato progettato da Andrea Vitali [1] un nuovo tipo di attacco che mira a sfruttare i social network, in particolare le informazioni contenute dentro di questi, per attaccare anche altri sistemi.
In questi capitolo viene descritto questo tipo di attacco in tutte le sue fasi.

3.1 L'algoritmo e le fasi dell'attacco

Quattro sono le fasi da attuare al fine di compromettere l'intero sistema telematico:
1. VIOLAZIONE SOCIAL network: in questa fase andremo ad ottenere informazioni personali degli utenti da utilizzare come "grimaldello" per forzare l'account della "vittima". Inoltre, sempre in questa fase, andremo ad esplicare le modalità attraverso le quali l'algoritmo evolve e si espande per tutta la rete telematica. Le successive tre fasi sono subordinate alla principale;
2. VIOLAZIONE ACCOUNT DI POSTA: in particolare in questa fase, tenteremo di utilizzare le informazioni reperite nella fase precedente per tentare di violare i sistemi di mailling;
3. VIOLAZIONE DI ALTRI SISTEMI: nella terza fase ci espanderemo, invece, verso altri sistemi, come quelli di pagamento online o e-commerce;
4. VIOLAZIONE ATTRAVERSO IL RINVIO DELLE CREDENZIALI: in quest'ultima fase tenteremo di violare le misure di sicurezza di altri siti attraverso il rinvio delle credenziali di accesso per mail. Quest'ultima fase ovviamente è legata al successo della seconda.

La prima fase prevede la creazione di un profilo dal quale iniziare a sferrare l'attacco; nel giro di cinquanta giorni, tale profilo ha raggiunto le 5.000 amicizie, il limite massimo imposto da Facebook, a questo punto è stato possibile dare inizio all'attacco vero e proprio.
In questa fase grazie alle API di Facebook è stato possibile reperire e memorizzare in un database la lista degli amici con tutte le informazioni associate ad essi.
Di seguito riportiamo alcune delle informazioni che Facebook mette a disposizione:

• uid - The user ID corresponding to the user info returned. This is always returned, whether included in fields or not, and always as the first subelement.
• about_me - text element corresponding to Facebook ’About Mè profile section. May be blank.
• activities - User-entered "Activities" profile field. No guaranteed formatting.
• affiliations - list of network affiliations, as affiliation elements, each of which contain year, type, status, name, and nid child elements. If no affiliations are returned, this element will be blank. The user’s primary network (key: nid) will be listed first.
– type takes the following values: *** college: college network *** high school: high school network *** work: work network *** region: geographical network
– year may be blank, depending on the network type.
– name is the name of the network.
– nid is a unique identifier for the network. The user-to-network relation

may be stored.
– status describes the user's graduate status if the network is a college network. Otherwise, it is blank.
• birthday - User-entered "Birthday" profile field. No guaranteed formatting as it's based on the user's locale. Use birthday_date instead of you need to be sure of the format.
• birthday_date - The user's birthday, rendered as a machine-readable string. The format of this date never changes.
• books - User-entered "Favorite Books" profile field. No guaranteed formatting.
• contact_email - is a string containing the user's primary Facebook email address. If the user shared his or her primary email address with you, this address also appears in the email field (see below).
• current_location - User-entered "Current Location" profile fields. Contains four children: city, state, country, and zip.
– city is user-entered, and may be blank.
– state is a well-defined name of the state, and may be blank.
– country is well-defined, and may be blank.
– zip is an integer, and is 0 if unspecified by the user.
• education_history - list of school information, as education_info elements, each of which contain name, year, and concentration child elements. If no school information is returned, this element will be blank.
– year is a four-digit year, and may be blank.
– name is the name of the school, and is user-specified.
– concentrations is a list of concentration elements, and may be an empty list.
– degree is the name of the degree, and may be blank.

• email is a string containing the user's primary Facebook email address or the user's proxied email address, whichever address the user granted your application.
• family is an array containing the user's family members, listing the user ID and relationship for each member.
• has_added_app - [Deprecated] Bool (0 or 1) indicating whether the user has authorized the application. This value is now equivalent to is_app_user.
• hometown_location - User-entered "Hometown" profile fields. Contains three children: city, state, and country.
– city is user-entered, and may be blank
– state is a well-defined name of the state, and may be blank.
– country is well-defined, and may be blank.
• hs_info - User-entered high school information. Contains five children: hs1_name, hs2_name, grad_year, hs1_key, and hs2_key.
– hs1_name is well-defined, and may be left blank
– hs2_name is well-defined, and may be left blank, though may not have information if hs1_name is blank.
– grad_year is a four-digit year, or may be blank
– hs1_id is a unique ID representing that school, and is not zero if and only if hs1_name is not blank.
– hs2_id is a unique ID representing that school, and is not zero if and only if hs2_name is not blank.
• interests - User-entered "Interests" profile field. No guaranteed formatting.
• is_app_user - Bool (0 or 1) indicating whether the user has used the calling application.
• is_blocked - Bool that returns true if the user is blocked to the viewer/

logged in user. The user must be logged in or you must have a valid session key to get this value.

• locale - The current locale in which the user has chosen to browse Facebook. The basic format is LL_CC, where LL is a two-letter language code, and CC is a two-letter country code. Country codes are taken from the ISO 3166 alpha 2 code list. For instance, 'en_US' represents US English.

• meeting_for - list of desired relationship types corresponding to the "Looking For" profile element. If no relationship types are specified, the meeting_ for element is empty. Otherwise represented as a list of seeking child text elements, which may each contain one of the following strings: Friendship,
A Relationship, Dating, Random Play, Whatever I can get.

• meeting_sex - list of desired relationship genders corresponding to the "Interested In" profile element. If no relationship genders are specified, the meeting_sex element is empty. Otherwise represented as a list of sex child text elements, which may each contain one of the following strings: male, female .

• movies - User-entered "Favorite Movies" profile field. No guaranteed formatting.

• music - User-entered "Favorite Music" profile field. No guaranteed formatting.

• #name - User-entered "Name" profile field. May not be blank.

– first_name is generated from the user-entered "Name" profile field.

– middle_name is generated from the user-entered "Name" profile field.

– last_name - The user's last name, which is generated from the userentered "Name" profile field.

• notes_count - Total number of notes written by the user.

• pic - URL of user profile picture, with max width 100px and max height 300px. May be blank.
• pic_with_logo - URL of user profile picture with a Facebook logo overlaid, with max width 100px and max height 300px. May be blank.
• pic_big - URL of user profile picture, with max width 200px and max height 600px. May be blank.
• pic_small - URL of user profile picture, with max width 50px and max height 150px. May be blank.
• pic_small_with_logo - URL of user profile picture with a Facebook logo overlaid, with max width 50px and max height 150px. May be blank.
• pic_square - URL of a square section of the user profile picture, with width 50px and height 50px. May be blank.
• pic_square_with_logo - URL of a square section of the user profile picture with a Facebook logo overlaid, with width 50px and height 50px. May be blank.
• political - User-entered "Political View" profile field. It's a free-form text field.
• profile_blurb - A free-form text field under a user's profile picture.
• profile_update_time - Time (in seconds since epoch) that the user's profile was last updated. If the user's profile was not updated within the past 3 days, 0 is returned.
• profile_url - URL of the Facebook profile of the user. If the user has specified a username, the username is included in the URL, not profile.php?id=UID.
• proxied_email - A proxied wrapper alternative for contacting the user through email, instead of directly calling notifications.sendEmail. If the user shared his or her proxied email address with you, this address also

appears in the email field (see below).

• quotes - User-entered "Favorite Quotes" profile field. No guaranteed formatting.

• relationship_status - User-entered "Relationship Status" profile field. Is either blank or one of the following strings: Single, In a Relationship, In an Open Relationship, Engaged, Married, It's Complicated, Widowed.

• religion - User-entered "Religious Views" profile field. No guaranteed formatting.

• sex - User-entered "Sex" profile file. This is a translated string, so the gender returned depends on the user's locale. This field may be blank.

• significant_other_id - the id of the person the user is in a relationship with. Only shown if both people in the relationship are users of the application making the request.

• status - Contains a "message" child with user-entered status information, as well as a "time" child with the time (in seconds since epoch) at which the status message was set.

• timezone - offset from GMT (e.g. California is -8).

• tv - User-entered "Favorite TV Shows" profile field. No guaranteed formatting.

• username - The user's Facebook username, if one was specified. Otherwise, this field is blank.

• wall_count - Total number of posts to the user's wall. Note that this does not include items with attachments, i.e. wall photos, wall videos, posted links, etc. Only items that show up on http://www.Facebook.com/wall.php are included in this count.

• website - User-entered personal website profile field. No guaranteed formatting.

• work_history - list of work history information, as work_info elements, each of which contain location, company_name, position, description, start_date and end_date child elements. If no work history information is returned, this element is blank.
– location is user-entered, and has a similar format to current_location and hometown_location above.
– company_name is user-entered, and does not necessarily correspond to a Facebook work network.
– description is user-entered, and may be blank.
– position is user-entered, and may be blank.
– start_date is of the form YYYY-MM, YYYY, or MM. It may be blank.
– end_date is of the form YYYY-MM, YYYY, or MM. It may be blank.

Una volta raccolte tutte le informazioni sugli utenti si passa all'individuazione delle credenziali di accesso, verranno infatti effettuati fino ad otto tentativi per ogni "amico", utilizzando l'email scoperta in precedenza e provando le password statisticamente più probabili secondo il seguente schema:

Informazioni Rappresentanti	Password	
	123456	
	password	
	querty	
	abc123	$\approx 2\ \%$ [1]
#name	andrea	
#gg#mm#aaaa	20061984	
#gg#mm#aa	200684	
#name#aa	andrea84	
#cell	3407721002	$> 6\ \%$ [2]

Tabella 3.1.1: Password utilizzate con più frequenza

Secondo i dati statistici questo tipo di attacco dovrebbe portare ad individuare

circa l'8% delle password, al termine della prova eseguita su un campione di 100 individui, questa tecnica ha portato alla scoperta del 14% delle password.
Per ogni account violato questa operazione viene ripetuta ricorsivamente.
Nello schema seguente viene mostrata un'analisi della possibile espansione dell'algoritmo.
Partendo da un utente con 1000 amici, considerando la percentuale di password trovate del 6% e ipotizzando che ogni utente colpito abbia a sua volta una media di 130 amici, lo schema seguente analizza la possibile espansione dell'attacco.

	%	Pwd Found	Amici	Totale persone
1.000	6 %	60	130[3]	7.800
7.800	6 %	450	130	≈ 60.000
60.000	6 %	3.600	130	≈ 500.000
500.000	6 %	30.000	130	≈ 4.000.000
4.000.000	6 %	250.000	130	≈ 32.000.000
32.000.000	6 %	2.000.000	130	≈ 260.000.000
260.000.000	6 %	15.600.000	130	> 400.000.000
400.000.000	6 %	24.000.000		

Tabella 3.1.2: Analisi statistica espansione

L'immagine sottostante mostra graficamente l'espansione di questo attacco.

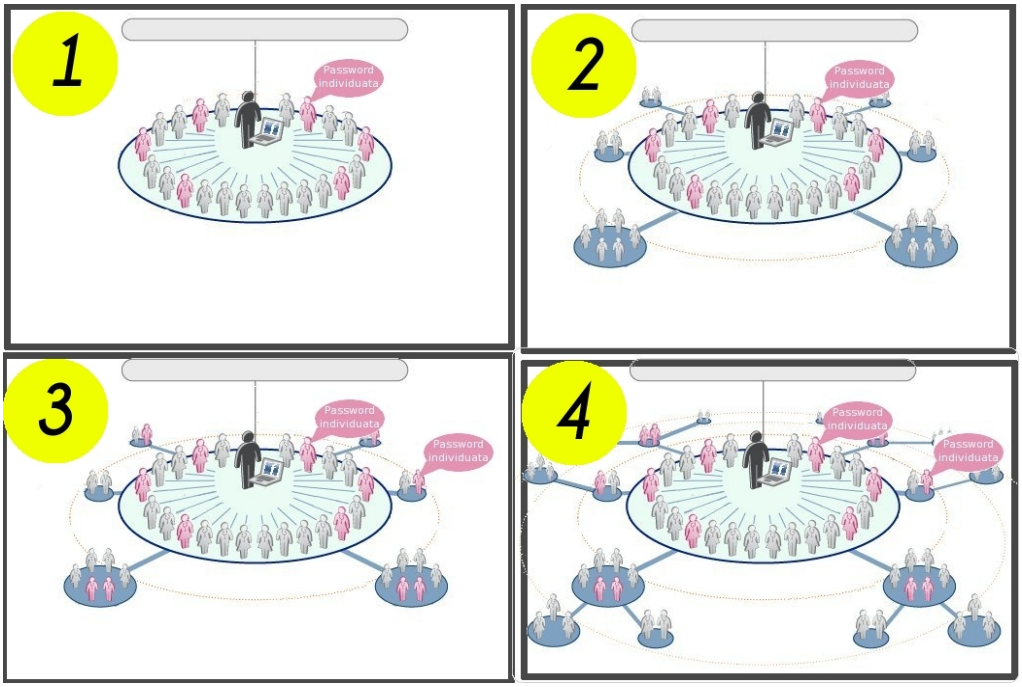

Figura 3.1.3: Grafico di espansione

Terminata questa fase si passa alla violazione dell'account di posta come illustrato da Andrea Vitali[1]:

Un metodo molto diffuso, ma anche molto rischioso, è quello di utilizzare la stessa password per servizi online differenti.

1. Per ogni account violato: provare la stessa password per gli account mail (oltre il 61 % di successo)

2. Una volta avuto l'accesso effettuare operazioni di ricerca all'interno della casella mail (post ricevuta / inviata / cestinata) alla ricerca di credenziali di accesso, numeri di carte, bank account, mail compromettenti, etc

Una diffusa cattiva abitudine degli italiani è quella di archiviare nelle proprie mail password di altri sistemi.

Nella terza fase si passerà alla violazione di altri sistemi:

Per ogni account violato: provare la stessa pwd per account paypal, ebay, altri siti. Secondo uno studio della Symantec, la maggior parte dei clienti bancari online riutilizza le proprie credenziali per accedere a siti web molto meno sicuri. Un nuovo report di Trusteer rivela che sono molte le persone a riutilizzare le stesse password e username. Il report rivela dati a dir poco inquietanti:

• Password Sharing: il 73% degli utenti riutilizzano la password di accesso al proprio servizio di banking online in siti non finanziari.

• User ID Sharing: Quando una banca consente la scelta del proprio user id, il 65% degli utenti riutilizza lo stesso username in siti non finanziari; nel caso in cui la banca assegni automaticamente un nome utente, verrà riutilizzato dall'utente (il 42% degli utenti) almeno in un sito non finanziario.

• User ID and Password Sharing: 47% degli utenti riutilizza i propri username e password di accesso alla banca su siti non finanziari

Una volta acquisito login e password vengono testati su siti di servizi finanziari per commettere frodi. Utilizzare le credenziali rubate resta il modo più semplice per aggirare le misure di sicurezza attuate dalle banche per proteggere le loro applicazioni.

Nella quarta ed ultima fase si provvederà al rinvio delle password: richiedere l'invio di password per email attraverso la funzionalità di "Password forget". Sistemi come Facebook, Gmail, Paypal, Vodafone, Trenitalia, etc rinviano la password o permettono di reimpostarla tramite link inviato per mail.

Capitolo 4

Obiettivi

Lo scopo di questa tesi è valutare se l'algoritmo presentato nel 2010 da Andrea Vitali nella tesi “L'11 settembre telematico” sia valido ancora oggi o se dopo due anni le abitudini degli utenti in ambito di sicurezza delle password, le nuove tecnologie e le nuove politiche in ambito di sicurezza e privacy di Facebook non permettano più di utilizzare questo algoritmo.

In particolare verrà analizzata la prima fase dell'attacco, ossia la fase in cui vengono raccolte le informazioni degli “amici” su Facebook e si tenta di impossessarsi dei loro account.

Questa è la fase più importante dal momento che le tre sono subordine al successo della prima.

Dopo avere valutato la validità di questo tipo di attacco si valuterà anche un eventuale metodo per migliorare le prestazioni dell'algoritmo analizzato.

Capitolo 5

Valutazioni

5.1 La complessità delle password

Secondo gli studi eseguiti prima di testare l'algoritmo in esame, risultava che l'8% delle password era facilmente individuabile provando prima semplici password banali poi utilizzando password mirate in base ai dati raccolti sui social network; in particolare il dato utilizzato in questa fase era la data di nascita che, concatenata in diversi modi con l'username, dava origine alle password da testare.
Dopo la prova su un campione di 100 utenti reali, la percentuale risultava pessimistica in quanto l'algoritmo aveva portato alla scoperta del 14% delle password.
Analizzando articoli sulla sicurezza pubblicati negli ultimi anni risulta che la percentuale di password banali facilmente rintracciabili con l'algoritmo descritto è sempre dell'8% circa.
Nel 2012 alcuni hacker hanno reso pubbliche le password di molti utenti iscritti a vari siti tra cui Yahoo voice, Linkedin, eHarmony e Last.fm; grazie a questi dati “SplashData”, un'azienda che produce software per la gestione di password, ha pubblicato la lista della peggiori password utilizzate dagli utenti nel 2012[8].
Nell'elenco sotto riportato delle prime dieci password pubblicate, si può notare come quattro delle prime cinque siano proprio quelle provate dall'algoritmo.

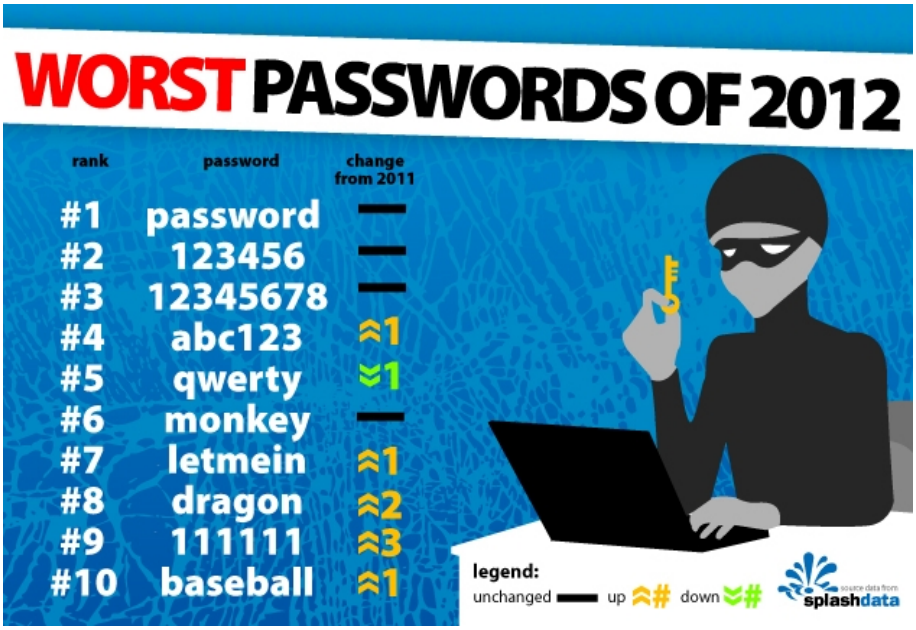

Figura 5.1.1: Worst password of 2012

Proseguendo in questa analisi, nell'articolo[3] "The psychology of Security for the Home Computer User", presentato nel 2012 al "IEEE Symposium on Security and Privacy", viene riportato uno studio sulla percezione dei rischi da parte di chi fa shopping online, dal quale si evince che l' 85% delle persone afferma di usare password molto robuste utilizzando combinazioni di lettere, numeri e simboli; proseguendo nelle ricerche però questi dati vanno in contraddizione perchè il 24% utilizza parole che fanno parte di un dizionario e un altro 24% delle password contengono informazioni personali.

Da questa differenza si comprende che l'utente non ha una concezione chiara del vero significato di password sicura.

Proseguendo ulteriormente l'analisi si arriva ad ottenere utili informazioni sul comportamento dell'utente: solo il 22% non ha mai riutilizzato una password in un altro sito e il 51% afferma di non cambiare mai la sua password o cambiarla molto raramente.

Analizzando i dati di un altro studio[9] " Analyzing password strength" di

Martin Devillers, in cui vengono studiate le password di 32 milioni di utenti rubate da un hacker nel 2009 dal database di RockYou, vediamo nuovamente come gli utenti utilizzino password poco sicure.
Le password di questo sito erano state salvate in chiaro nel database, senza alcuna funzione di hash e grazie a questo è stato possibile eseguirne un'accurata analisi.
Guardando il grafico sottostante è possibile notare che la maggior parte delle password sono costituite da caratteri minuscoli, numeri o semplici concatenazioni di questi due, mentre le password considerate più “robuste” sono solo al 9%.

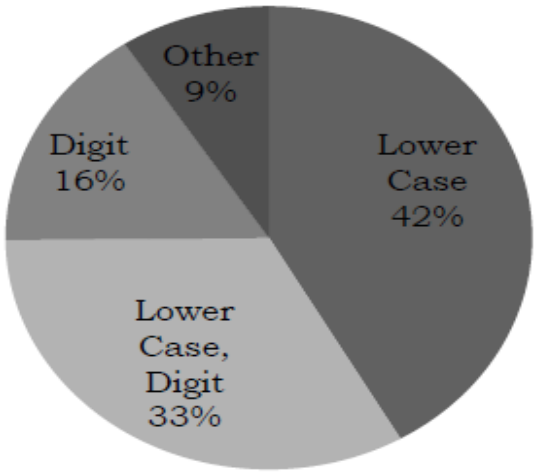

Figura 5.1.2: Diagramma composizione password del sito "RockYou"

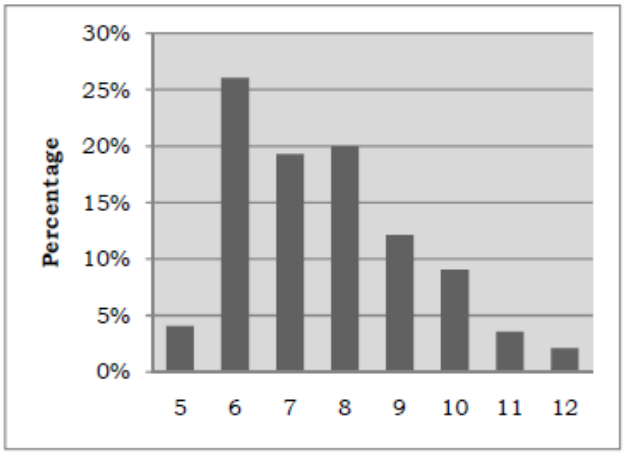

Figura 5.1.3: Grafico lunghezza password del sito "RockYou"

Se poi si analizza il grafico che mostra la lunghezza delle password, si può notare che più della metà delle password ha una lunghezza compresa tra i 6 e gli 8 caratteri.

Confrontando queste password con le liste delle password più comuni si ottengono risultati interessanti: in media ogni 100 password una è costituita dalla password “123456”, e le top 10, 100, 1000, 10000 password coprono rispettivamente il 2%, 5%, 11% e 22% delle password utilizzate.

Un altro dato interessante è che nonostante gli utenti siano stati informati che i database contenenti le loro password sono stati compromessi e di conseguenza le loro password rese pubbliche, la percentuale di utenti che decide di cambiare la password rimane bassissima, inferiore all'1% secondo la ricerca di Matt Weir e Sudhir Aggarwal.

Per migliorare l'algoritmo di ricerca delle password quindi si potrebbe sfruttare questo dato, inserendo nel nuovo database anche tutti gli username e le password che nel corso di questi anni sono state pubblicate da hacker; dato che la percentuale di utenti che riutilizzano la stessa password su siti diversi è circa dell' 80%, basterebbe cercare se nel database creato è presente l'utente del

quale stiamo cercando la password e in tal caso inserire nella lista delle password da testare le password che ha già utilizzato su altri siti web.
I risultati ottenuti potrebbero essere considerati buoni.

Le politiche di scelta della password di Facebook non sono di aiuto alla sicurezza, infatti gli unici limiti che vengono dati ai nuovi iscritti sono che la password scelta sia di almeno 6 caratteri e che questi non siano tutti uguali; obbligando invece l'utente, al momento dell'iscrizione, ad inserire una password con caratteri alfanumerici sia maiuscoli che minuscoli, numeri e caratteri speciali, un attacco come quello analizzato sarebbe completamente inefficace.

5.2 Le nuove API e il problema dell'email

Le nuove impostazione di Facebook hanno reso l'email degli "amici" un dato nascosto a meno che l'utente non specifichi che vuole renderlo pubblico.
Questa nuova politica crea dei seri problemi all'algoritmo perché le API non permettono più di ottenere l'indirizzo e-mail degli "amici".
Inoltre le funzioni friends.get e users.getInfo oggi sono deprecate e sono state rimpiazzate dalle nuove "Graph API"[10] che sono più sicure dal punto di vista della sicurezza e della privacy; infatti molti dati che prima erano disponibili senza alcun problema adesso richiedono un "access token" specifico per accedervi.
Ogni utente ha i seguenti campi:

Name	**Description**	**Permissions**
`id`	The user's Facebook ID	No `access_token` required

name	The user's full name	No access_token required
first_name	The user's first name	No access_token required
middle_name	The user's middle name	No access_token required
last_name	The user's last name	No access_token required
gender	The user's gender: female or male	No access_token required
locale	The user's locale	No access_token required
languages	The user's languages	user_likes
link	The URL of the profile for the user on Facebook	No access_token required
username	The user's Facebook username	No access_token required
age_range	The user's age range; only returned if specifically requested via the fields URL parameter	Requires access_token
third_party_id	An anonymous, but unique identifier for the user; only returned if specifically requested via the fields URL parameter	Requires access_token
installed	Specifies whether the user has installed the application associated with the app access token that is used to make the request; only returned if specifically requested via the fields URL parameter	Requires app access_token
timezone	The user's timezone offset from UTC	Available only for the current user
updated_time	The last time the user's profile was updated; changes to the languages,	Requires access_token

	link, timezone, verified, interested_in, favorite_athletes, favorite_teams, and video_upload_limits are not not reflected in this value	
verified	The user's account verification status, either true or false (see below)	Requires access_token
bio	The user's biography	user_about_me or friends_about_me
birthday	The user's birthday	user_birthday or friends_birthday
cover	The user's cover photo (must be explicitly requested using fields=cover parameter)	Requires access_token
currency	The user's currency settings (must be explicitly requested using a fields=currency URL parameter)	Requires access_token
devices	A list of the user's devices beyond desktop	User access_token required; only available for friends of the current user
education	A list of the user's education history	user_education_history or friends_education_history
email	The proxied or contact email address granted by the user	email
hometown	The user's hometown	user_hometown or friends_hometown
interested_in	The genders the user is interested in	user_relationship_details or friends_relationship_details
location	The user's current city	user_location or friends_location

political	The user's political view	user_religion_politics or friends_religion_politics
payment_pricepoints	The payment price-points available for that user	User access_token
favorite_athletes	The user's favorite athletes; this field is deprecated and will be removed in the near future	user_likes or friends_likes
favorite_teams	The user's favorite teams; this field is deprecated and will be removed in the near future	user_likes or friends_likes
picture	The URL of the user's profile pic (only returned if you explicitly specify a 'fields=picture' param)	access_token required for pages with whitelisting/targeting restrictions, otherwise no access_token required
quotes	The user's favorite quotes	user_about_me or friends_about_me
relationship_status	The user's relationship status: Single, In a relationship, Engaged, Married, It's complicated, In an open relationship, Widowed, Separated, Divorced, In a civil union, In a domestic partnership	user_relationships or friends_relationships
religion	The user's religion	user_religion_politics or friends_religion_politics
security_settings	Information about security settings enabled on the user's account (must be explicitly requested using a fields=security_settings URL parameter)	Available only for the current user

significant_other	The user's significant other	user_relationships or friends_relationships
video_upload_limits	The size of the video file and the length of the video that a user can upload; only returned if specifically requested via the fields URL parameter	Requires access_token
website	The URL of the user's personal website	user_website or friends_website
work	A list of the user's work history	user_work_history or friends_work_history

E le seguenti connessioni:

Name	Description	Permissions
accounts	The Facebook apps and pages owned by the current user.	manage_pages yields access_tokens that can be used to query the Graph API on behalf of the app/page
achievements	The achievements for the user.	user_games_activity or friends_games_activity.
activities	The activities listed on the user's profile.	user_activities or friends_activities.
albums	The photo albums this user has created.	user_photos or friends_photos.
apprequests	The user's outstanding requests from an app.	Requires app access_token.
books	The books listed on the user's profile.	user_likes or

		friends_likes.
checkins	The places that the user has checked-into.	user_checkins or friends_checkins.
events	The events this user is attending.	user_events or friends_events.
family	The user's family relationships	user_relationships.
feed	The user's wall.	read_stream
friendlists	The user's friend lists.	read_friendlists.
friendrequests	The user's incoming friend requests.	user_requests.
friends	The user's friends.	Any valid access_token of the current session user.
games	Games the user has added to the Arts and Entertainment section of their profile.	user_likes
groups	The Groups that the user belongs to.	user_groups or friends_groups.
home	The user's news feed.	read_stream.
inbox	The Threads in this user's inbox.	read_mailbox.
interests	The interests listed on the user's profile.	user_interests or friends_interests.
likes	All the pages this user has liked.	user_likes or friends_likes.
links	The user's posted links.	read_stream.
locations	Posts, statuses, and photos in which the user has been tagged at a location, or where the user has authored content (i.e. this excludes objects with no location information, and objects in which the user is not tagged). See documentation of the location_post table for more detailed information on permissions.	user_photos, friend_photos, user_status, friends_status, user_checkins, or friends_checkins.

movies	The movies listed on the user's profile.	`user_likes` or `friends_likes`.
music	The music listed on the user's profile.	`user_likes` or `friends_likes`.
mutualfriends	The mutual friends between two users.	Any valid `access_token` of the current session user.
notes	The user's notes.	`user_notes`.
notifications	App notifications for the user.	Any valid `access_token` of the current session user.
outbox	The messages in this user's outbox.	`read_mailbox`.
payments	The Facebook Credits orders the user placed with an application. See the Credits Api for more information.	app `access_token`
permissions	The permissions that user has granted the application.	None.
photos	Photos the user (or friend) is tagged in.	`user_photo_video_tags` or `friends_photo_video_tags`.
Photos/uploaded	All of the updates photos of a user. Cursor based pagination.	`user_photos`
picture	The user's profile picture.	No `access_token` required.
pokes	The user's pokes.	`read_mailbox`.
posts	The user's own posts.	Any valid `access_token` or `read_stream` to see non-public posts.
questions	The user's questions.	`user_questions`
scores	The current scores for the user in games.	`user_games_activity` or `friends_games_act`

		ivity.
sharedposts	Returns shares of the object. Cursor based pagination.	read_stream
statuses	The user's status updates.	read_stream.
subscribedto	People you're subscribed to.	Any valid access_token
subscribers	The user's subscribers.	Any valid access_token
tagged	Posts the user is tagged in.	read_stream
television	The television listed on the user's profile.	user_likes or friends_likes.
updates	The updates in this user's inbox.	read_mailbox.
videos	The videos this user has been tagged in.	user_videos or friends_videos.

Nonostante queste nuove limitazioni la maggior parte delle informazioni sono ancora reperibili senza alcun problema: è sufficiente infatti richiedere gli apposti token dei dati che interessano (ad esempio per ottenere le date di nascita degli amici basterà richiedere il token “friends_birthday” e otterremo questi dati senza alcun problema).

Con poche righe di codice utilizzando le nuove “Graph API” è possibile ottenere le seguenti informazioni sugli amici:


```
{
 "id": "100004883953359",
 "name": "Tony Mazzini",
 "first_name": "Tony",
 "last_name": "Mazzini",
 "link": "http://www.facebook.com/tony.mazzini.52",
 "username": "tony.mazzini.52",
 "birthday": "12/18/1983",
```

```
"hometown": {
  "id": "110340245656034",
  "name": "Bologna, Italy"
},
"location": {
  "id": "110340245656034",
  "name": "Bologna, Italy"
},
"bio": "Utente di prova",
"quotes": "Citazione di prova",
"work": [
  {
    "employer": {
      "id": "120851971326024",
      "name": "pizzeria \"da mimmo\""
    },
    "start_date": "0000-00",
    "end_date": "0000-00"
  }
],
"education": [
  {
    "school": {
      "id": "194641867213525",
      "name": "Liceo scientifico \"Enrico Fermi\" Bologna"
    },
    "year": {
      "id": "324576644221180",
      "name": "2000"
    },
    "type": "High School"
  },
  {
    "school": {
      "id": "102183273156950",
      "name": "University of Bologna"
    },
    "year": {
      "id": "120960561375312",
      "name": "2013"
    },
    "type": "College"
  }
],
```

```
"gender": "male",
"relationship_status": "Single",
"religion": "Buddista",
"political": "Anarchy",
"locale": "it_IT",
"languages": [
  {
    "id": "162930927096211",
    "name": "Old Spanish"
  }
],
}
```


L'unico vero problema è dato dal campo “email”; consultando queste nuove API vediamo infatti che è scritto esplicitamente “A user's email is a protected property and access to that information must be specifically requested by the app and granted by the user”.

Questo è un problema che impedirebbe l'uso dell'algoritmo dato che, non potendo conoscere le email degli utenti che l'algoritmo va ad attaccare, non sarebbe possibile tentare il login su Facebook.

Esiste tuttavia un metodo che permette di aggirare questo ostacolo grazie all'utilizzo di un servizio offerto da Yahoo: collegandosi all'indirizzo http://address.yahoo.com/ basta creare un account su yahoo e avremo la possibilità di importare i contatti di facebook.

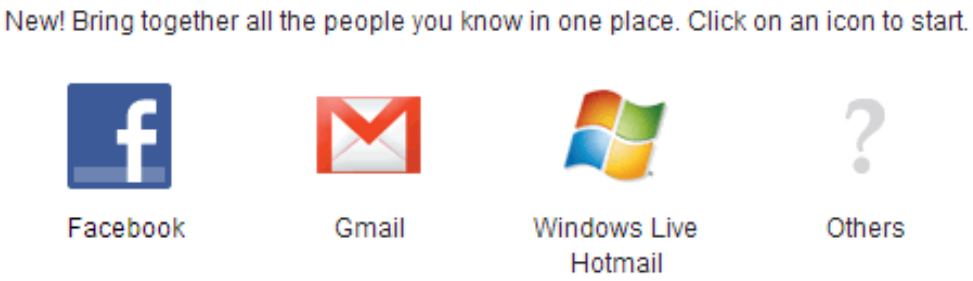

Figura 5.2.1: Schermata per importare i contatti di Facebook

Selezionando l'icona di Facebook ci verrà chiesto di loggarci con le nostre credenziali e, una volta dato il permesso a Yahoo di accedere alle informazioni dei nostri contatti, avremo disponibile la lista di tutti i nostri "amici" con le rispettive e-mail.

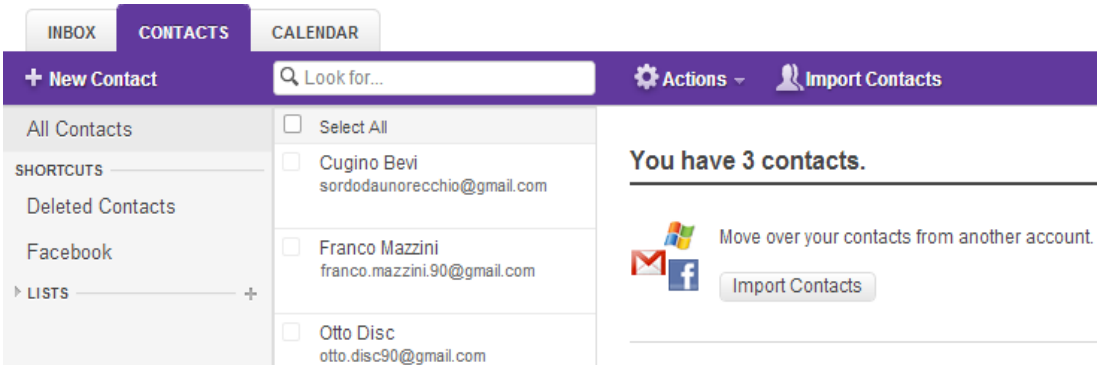

Figura 5.2.2: Schermata contatti importati con rispettivi indirizzi e-mail

5.3 Il "social phishing"

5.3.1 "Phishing" e "spear phishing"

Il "Phishing" è un tentativo di acquisire informazioni riservate come username, password e numeri di carta di credito, fingendosi un'entità fidata in una comunicazione elettronica.
Il processo standard delle metodologie di attacco di phishing può riassumersi nelle seguenti fasi[11]:

1. L'utente malintenzionato (*phisher*) spedisce al malcapitato e ignaro utente un messaggio email che simula, nella grafica e nel contenuto, quello di una istituzione nota al destinatario (per esempio la sua banca, il suo provider web,

un sito di aste online a cui è iscritto).

2. L'e-mail contiene quasi sempre avvisi di *particolari situazioni o problemi* verificatesi con il proprio conto corrente/account (ad esempio un addebito enorme, la scadenza dell'account, ecc.) oppure un'offerta di denaro.

3. L'e-mail invita il destinatario a seguire un link, presente nel messaggio, per evitare l'addebito e/o per regolarizzare la sua posizione con l'ente o la società di cui il messaggio simula la grafica e l'impostazione (Fake login).

4. Il link fornito, tuttavia, *non* porta in realtà al sito web ufficiale, ma a una *copia fittizia* apparentemente simile al sito ufficiale, situata su un server controllato dal phisher, allo scopo di richiedere e ottenere dal destinatario dati personali particolari, normalmente con la scusa di una conferma o la necessità di effettuare una autenticazione al sistema; queste informazioni vengono memorizzate dal server gestito dal phisher e quindi finiscono nelle mani del malintenzionato.

5. Il phisher utilizza questi dati per acquistare beni, trasferire somme di denaro o anche solo come "ponte" per ulteriori attacchi.

Figura 5.3.1.1: Esempio di email phishing

Figura 5.3.1.2: Esempio sito phishing

Gli attacchi di questo tipo non hanno in genere alte probabilità di successo (dall'1 al 3%) a causa del fatto che spesso queste email di phishing sono scritte con traduttori automatici, per cui sono presenti errori ma soprattutto perché gli utenti ai quali vengono offerte somme di denaro sul conto corrente o offerte troppo vantaggiose tendono a non fidarsi; i browser, inoltre, hanno dei sistemi di sicurezza che sono in grado di riconoscere e bloccare molte pagine di phishing e contribuiscono così a tenere bassa questa percentuale.

Esiste anche una tecnica particolare di phishing, chiamata “spear phishing”, che è diretto a una particolare persona o a un gruppo ristretto; l'attaccante cerca di guadagnare la fiducia della vittima servendosi di informazioni su di lei, studiando le sue abitudini, i suoi interessi, cosa compra online, qual'è la sua banca, il nome dei genitori, della moglie, dei figli.

Tutte queste informazioni si possono recuperare con delle ricerche su internet,

e con l'avvento dei social network, reperire queste informazioni è diventato ancora più facile.

5.3.2 Social phishing

"Social Phishing"[4], ricerca pubblicata nel dicembre del 2005, rivela dati molto interessanti relativi al phishing in collaborazione con i dati raccolti sui social network.

Scopo di questa ricerca è capire quanto facilmente può un phisher servirsi dei dati dei social network per incrementare il rendimento del proprio attacco.

In questo studio sono stati semplicemente acquisiti dati da vari social network ed è stato facilmente e velocemente creato un database con decine di migliaia di relazioni usando crawling e parsing tools come le "Perl LWP library", accessibili a chiunque.

L'attacco è stato lanciato su un gruppo di studenti dell'Indiana University di età tra i 18 e 24 anni.

Nella figura seguente viene presentato graficamente l'attacco in cui l'indirizzo e-mail del mittente viene alterato e il destinatario viene invitato a collegarsi a una pagina dell'università in cui gli vengono chieste le proprie credenziali, nel primo caso il mittente è uno sconosciuto con un indirizzo email dell'università, mentre nel secondo caso l'indirizzo è di un amico, il tutto grazie alle informazioni ricavate dai social network e salvate in precedenza nel database.

Nel primo caso il 16% degli utenti, 15 su 94, hanno fornito le proprie credenziali, ma il dato veramente interessante è relativo al secondo attacco, quello che si è servito delle informazioni dei social network; in questo caso la percentuale di successo è stata del 72%, con 349 vittime su 487 utenti.

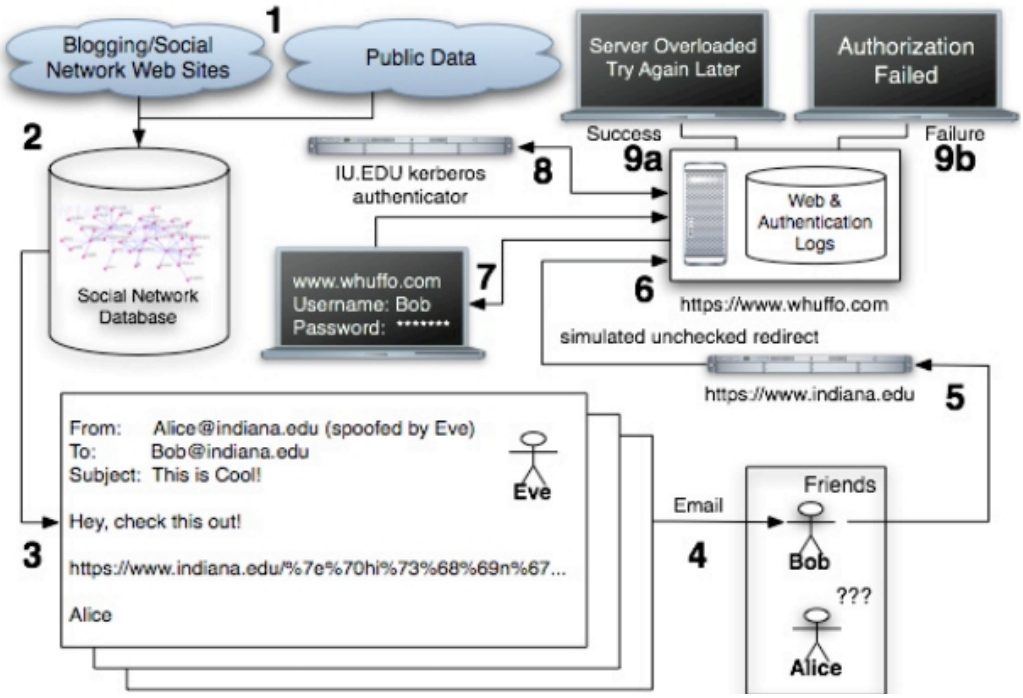

Figura 5.3.2.1: Schema attacco "Social Phishing"

5.3.3 Il nuovo algoritmo di attacco

Data la percentuale di successo del tipo di attacco appena descritto si potrebbe pensare di ottimizzare l'algoritmo sfruttando questa tecnica.

Una volta terminata la prima fase dell'attacco, in cui vengono testate le 8 password, si procede con l'algoritmo normale con gli utenti di cui si sono scoperte le password, mentre per il resto degli utenti, si potrebbero utilizzare i dati raccolti in precedenza nel database per creare messaggi personalizzati in cui gli utenti vengono invitati a cliccare su un determinato link che connette l'utente ignaro a una finta pagina di login di Facebook nella quale se la vittima inserisce i propri dati questi vengono inviati all'attaccante e alla vittima compare un messaggio in cui si comunica che le credenziali inserite sono errate venendo poi reindirizzato alla vera pagina di Facebook per cercare di

non creare sospetti.

Secondo i dati della ricerca si potrebbe attaccare con successo una percentuale molto elevata di utenti, nella tabella è possibile osservare come nel giro di poche iterazioni dell'algoritmo il numero di utenti colpiti risulterebbe molto alto.

Utenti	%	Password trovate	Amici	Totale persone
940	72	676	130	87800
7350	72	5292	130	687960
56400	72	40608	130	5279040
470000	72	338400	130	43992000
3750000	72	2700000	130	351000000
30000000	72	21600000	130	2808000000

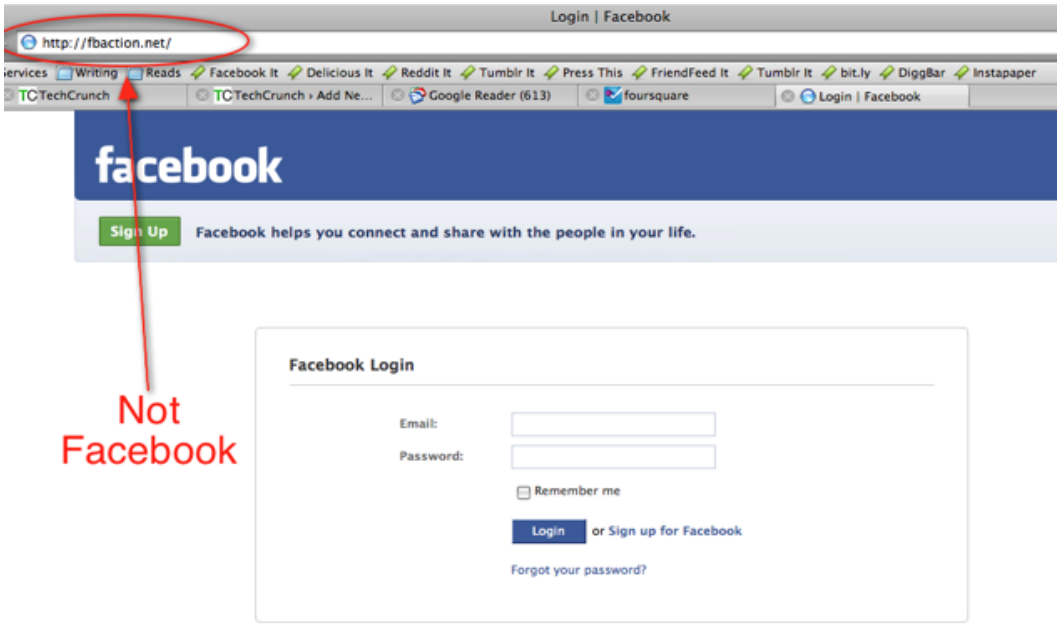

Figura 5.3.3.1: esempio sito phishing di Facebook

A queste "nuove" vittime si potrebbe poi sferrare di nuovo un attacco del

primo tipo e ai “superstiti” riprovare con la tecnica di “social phishing” descritta in precedenza e proseguendo infine alternando i due attacchi in questione. Se i dati ipotizzati trovassero riscontro anche in un test reale si arriverebbe ad attaccare un numero altissimo di utenti nel giro di poche iterazioni dell'algoritmo.

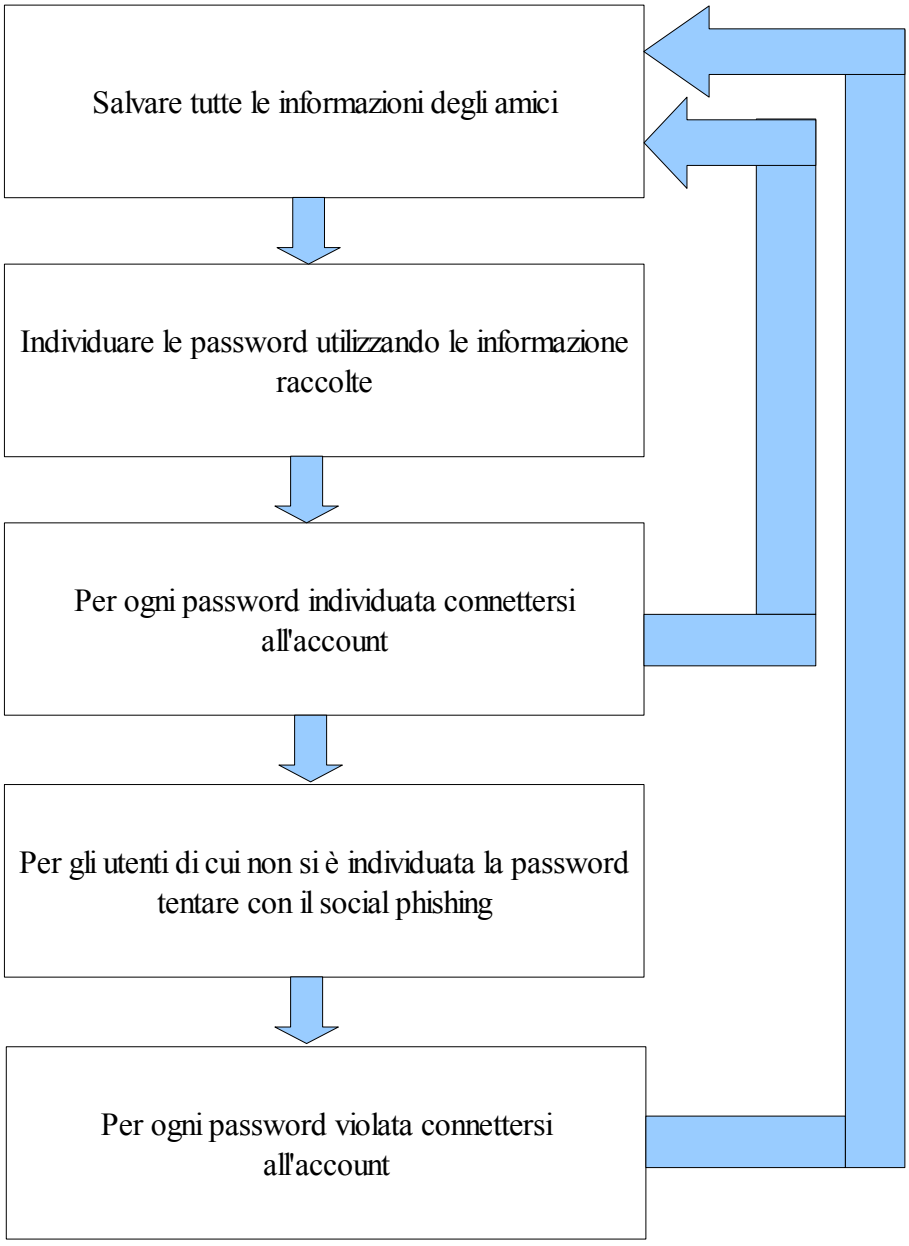

Figura 5.3.3.2: schema nuovo algoritmo di attacco

Capitolo 6

Conclusioni

Lo scopo di questa tesi è valutare l'efficienza del tipo di attacco presentato nella tesi “L'11 settembre telematico” di Andrea Vitali.
Nel 2010 era stata provata una violazione totale del 14% degli utenti attaccati; alla luce dei nuovi dati sulla sicurezza delle password utilizzate dagli utenti, questo dato risulta confermato anche ad inizio 2013.
E' stato riscontrato un problema che potrebbe rendere inutilizzabile questa tecnica di attacco in quanto dal 2010 ad oggi le politiche sulla privacy di Facebook sono state cambiate e l'email degli amici non è più un dato accessibile, tuttavia grazie a un servizio di Yahoo è ancora possibile risalire agli indirizzi di posta elettronica dei contatti.
Come ultimo punto è stata presentata una tecnica che potrebbe portare a migliorare notevolmente la percentuale di successo dell'algoritmo: utilizzando una tecnica particolare di phishing, che si avvale delle informazioni raccolte in internet e sui social network riguardo agli utenti e le loro relazioni, si potrebbe così arrivare ad ottenere le credenziali di accesso del 72% degli utenti.
Questo dato è particolarmente interessante perché tale percentuale di successo significherebbe poter avere il controllo totale di quasi 3 utenti su quattro, una percentuale altissima.
Si tratta quindi di poter eseguire transazioni on-line, visualizzare i conti bancari e i loro movimenti, i dettagli delle chiamate, i numeri di carte di credito e un'infinità di dati personali.

Si può quindi concludere che ancora oggi sia possibile sferrare un attacco sfruttando la tecnica analizzata e che tale tecnica potrebbe essere migliorata e e resa più efficace arrivando a percentuali di successo ancora più elevate, mettendo in serio pericolo un numero altissimo di utenti.

6.1 Sviluppi futuri

In questo lavoro non è stato possibile trovare un modo valido di implementare script per bypassare il problema degli indirizzi email non più visualizzabili con le API di Facebook, bisognerebbe testare se effettivamente sia possibile crearlo per poter sferrare così questo tipo di attacco.
Sarebbe inoltre interessante provare ad eseguire test su un campione reale e abbastanza significativo di persone per verificare se, attraverso la tecnica del "Social phishing", si possa realmente arrivare ad avere una percentuale di successo così alta.

Bibliografia

[1] Andrea Vitali. L'11 settembre telematico. Rischi di sicurezza causati dai social network. Master's thesis, Alma Mater Studiorum- Università di Bologna, 2009/2010.

[2] Andrea Vitali, Vittorio Ghini, Fabio Panzieri. E-apocalypse now: the Massive Multi-Way Attack.

[3] Adele E.Hove, Indrajit Ray, Mark Roberts, Malgorzata Urbanska, Zinta Byrne. The Psychology of Security for the Home Computer User. 2012 IEEE Symposium on Security and Privacy.

[4] Tom N. Jagatic, Nathaniel A. johnson, Markus Jakobsson, Filippo Menczer. SOCIAL PHISHING. Comunication of the Acm, October 2007, vol. 50, no. 10.

[5] Joseph Bonneau, Computer Laboratory, University of Cambridge. The science of guessing: analyzing an anonymized corpus of 70 million password.

[6] Mediascope Europe. Pan-European Launch Presentation Summary. May 2012

[7]SocialBakers. URL:http://www.socialbakers.com/facebook-statistics/

[8]Splashdata. Worst Password of 2012 – and How to Fix Them. URL:http://splashdata.com/press/PR121023.htm

[9] Martin M.A. Devillers. Analyzing Password Strength. July 2010.

[10]FacebookDevelopers
URL:https://developers.facebook.com/docs/reference/api/
URL:https://developers.facebook.com/docs/reference/api/user/

[11]Phishing. Wikipedia. URL:http://it.wikipedia.org/wiki/Phishing