About privacy and phishing on social networks and the case of Facebook

Paolo Di Sia
School of Engineering & School of Medicine, Stradella S.
Nicola 3, 36100 Vicenza, Italy

e-mail address: paolo.disia@gmail.com
ORCID: 0000-0002-6405-0483

DOI: https://doi.org/10.15503/emet.v5i5.429

Abstract

Aim. In recent years, social networks have multiplied on the Internet, becoming more and more used, and consequently raising doubts about the security of privacy. This exponential development has attracted the attention of bad-intentioned too. The aim of the research is to understand how "attack algorithms" can violate the privacy of millions of people, despite privacy policies which do not allow their use.

Methods. Considering an analysis of password security on Facebook, I evaluate the problems connected with the use of an attack algorithm in relation to privacy and security.

Results. Over the years, Facebook privacy policies have been changed, but with new services it is still possible to trace personal information. Using special phishing techniques, it is possible to get the access credentials of a good percentage of users. This allows attackers to perform online transactions, view bank accounts and their transactions, call details, credit card numbers and many other personal data.

Conclusions. Waiting for the power of the future quantum Internet, it is unfortunately possible today to launch an attack exploiting the analysed techniques and even improve them, making them more effective and reaching even higher success rates, thus placing a very high number of users in serious danger.

Key words: Social network, Privacy, Algorithm of attack, Password, Phishing, Facebook.

Introduction

For some years social networks have invaded the Internet, becoming more and more used and consequently raising many doubts on the security of privacy. They are platforms in which users can interact with friends, relatives or strangers, developing personal and professional relations. This exponential development attracted the attention not only of users, but also of mal-intentioneds. Most of the users of network consider social networks as a useful and indispensable tool for maintaining relations, but the price to pay could be rather high.

Many attack algorithms have been created during these years; exploiting the information available on social networks, they can lead to violations of privacy of entire populations of the web, i.e. millions of people. It is true that new policies regarding privacy are created over time, but it is also true that the efficiency of attacks improves (Vitali, 2010).

The use of the Internet and its penetration within the population grows year by year: 428.9 of 652.1 million people in Europe are online. According to data from a 2012 survey carried out by Mediascope Europe, the European average time spent online is 14.8 hours per week, a trend that grew by 10% compared to 2010 and it continuously increases.

Considering the rising spread of tablets and smartphones, the Internet is always closer to everyone's hands; the number of devices connected to the Internet has exceeded the number of people. The Internet penetration within the European population has reached 65%, 19% more than in 2010. Activities such as watching TV, listening to the radio, reading newspapers are being carried out more and more online: 73% of internet users watch online TV, 67% listen to radio online, 91% read news on the web (Mediascope Europe, 2012).

Only in Europe, 37% of users access the Internet using more than one device; in most cases they are PCs and smartphones, but for some years even the entertainment consoles such as Playstation and Xbox have Internet access as well as the last generation TVs. While on one hand this increased connectivity brings advantages and convenience, on the other one there is an increase in security danger for these devices and for the confidentiality of sensitive data.

In the following paragraphs the starting scenario will be illustrated, considering then in more detail the operation of attack algorithms. I will then consider an analysis on password security regarding user habits, and an assessment of problems that the new privacy and security policies of Facebook could create to the use of attack algorithms.

Social networks and associated risks

Social networks are increasingly popular and influence the lives of everyone; Facebook, the most used among them, has exceeded two billions users (Figure 1).

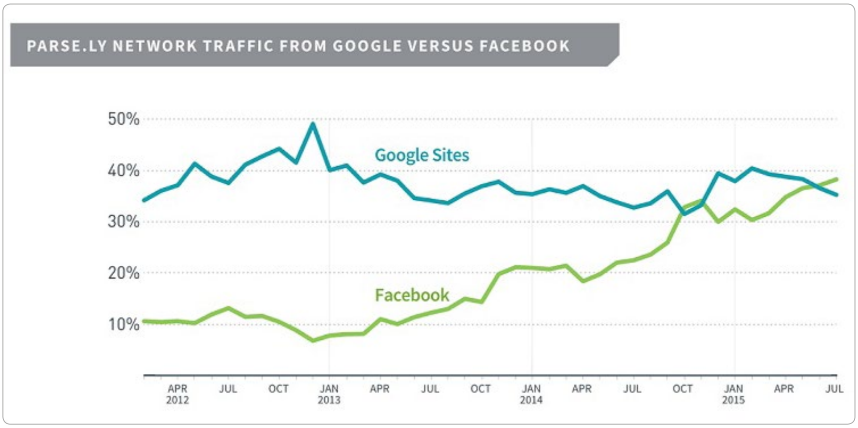


Fig. 1. Facebook vs Google. Source: fortune.com, 2018.

Worldwide, there are over 2.13 billion monthly active Facebook users; there is a 14% increase year on year. The average time spent on social media, in particular on Facebook, has exceeded the time spent on any other site (*zephoria.com*, 2018). Social networks influence everyday activities; while watching a TV series, for example, people are invited to comment on social networks, for example on twitter with a specific hashtag. Considering a world map of social networks, Facebook wins by far the challenge (Figure 2).

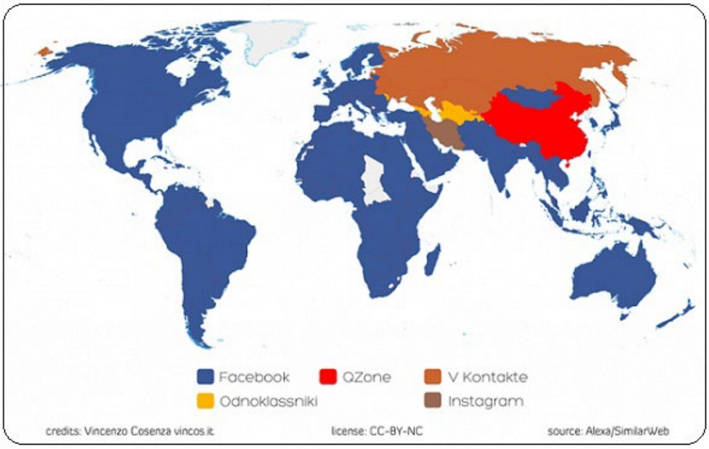
In social networks a lot of personal information is stored that risks becoming accessible to strangers and potential criminals. This problem has been repeatedly considered in resolutions of world privacy sponsors, in which it has highlighted that personal data has become publicly available in a global way, according to qualitative and quantitative unprecedented schemes, also through huge quantities of photos and digital videos.

On social networks it is possible to get information of various types:

a) personal data such as name, surname, gender, date of birth, city, residence, etc.;
b) contacts such as phone number, mobile phone number, e-mail address, personal website, instant messaging;
c) educational path: academic path, qualification, diploma, achieved specializations;
d) further info: political orientation, general interests, aggregations, groups.

This information, inserted by the average user into the social network, can become very dangerous in the hands of malicious people who, using social engineering techniques, can exploit this data to pursue their own goals.

Among the main risks (Salucci, 2012; Hove, Ray, Roberts, Urbanska, &



credits: Vincenzo Cosenza vincos.it license: CC-BY-NC source: Alexa/SimilarWeb

Fig. 2. World map of social networks in January 2018. Source: vincos.it, 2018.

Byrne, 2012) related to the use of social network services, we underline:

- *the data:* once published, data can remain forever, even if the person has deleted it from the "original" site; in fact there may be copies to third parties. In addition, some service providers refuse to comply with user requests for deletion of data and entire profiles;
- *the misleading idea of "community":* many service providers claim to transfer the communication structures from the "real world" to "cyberspace" and this is referred to as sharing information with a group of friends in the real world. However, examined more closely the features of some services, the parallel does not hold up because the concept of "friends" in cyberspace can be very different from the traditional idea of friendship. Without transparency on how to share the information contained in the profiles, it may happen that the idea of "community" described in the aforementioned terms ends up inducing them to reveal in inconsiderate ways personal information. Even the names given to these platforms (for example "MySpace") create an illusory idea of privacy and discretion on the web;
- *"free":* it does not always mean "at no cost". Many of the social network services charge users through the reuse of data contained in personal profiles by service providers, for example for targeted marketing activities;
- *collection of traffic data:* social network service providers have the technical means for recording on their site every single step of the user and can communicate to third parties personal traffic data including IP addresses, for example for advertising purposes. For many providers of these services, data contained in the profiles of users and the number of exclusive users are the only real assets they own;

- *revealing more personal information than you think:* photos can turn into universal biometric identifiers within a network and even across multiple networks. In recent years, the performance of face recognition software has improved dramatically, and even better results will come in the future. Content-based image retrieval (CBIR) technology provides additional opportunities to locate users by associating the identifying elements of some environments or locations (such as a painting hanging in a room, a building visible in the image) to location data (Yue, Li, Liu, & Fu, 2011). The functions called "social graphs" actually reveal information on the relations between the individual users;
- *improper use of user profiles by third parties:* this is probably the most serious potential risk with regard to personal data contained in the user profiles of social network services. Depending on the available configuration and the level of security offered by the service, the information contained in the profile (including images that can portray both the concerned and others) may become accessible, in the worst case, to the entire user community. At the same time, the safeguards available today are weak compared to the copy of data contained in the user profiles and to their use for building personal profiles and/or re-publish such data outside the specific social network service;
- *security of Internet services:* recent cases of problems concern well-known service providers such as Facebook, Flickr, MySpace, Orkut, StudiVZ. The goal of total security is very difficult, given the complexity of software applications at any level of Internet services. ENISA (European Network and Information Security Agency) (enisa.europa.eu, 2018) cites spam, scripting between different sites, viruses and worms, spear-phishing and specific phishing forms of social network services, network infiltration, abusive use of user profiles (profile-squatting) with attacks based on identity theft, forms of personal persecution (stalking), network bullying, industrial espionage, i.e. the so-called "social engineering attacks" performed through social network services;
- *introduction of inter-operability standards and application-programming interfaces*, in order to allow the technical inter-operability of typologically different social network services, involves a whole series of additional risks. In fact, an automatic evaluation of all social network sites that use the chosen standard becomes possible. Applications potentially able to interfere with the users privacy include, for example, the overall analysis of professional and private relationships maintained by the individual user; furthermore, inter-operability may promote the re-use by third parties of the information and images contained in the user profiles.

Algorithms and attack phases

There are always new types of attacks designed to exploit social networks, particularly the information contained within them, for attacking other systems as well. The steps to be implemented in order to compromise the entire telematic system are:

a) *violation of the social network:* in this phase we get personal information from users to be used as a "picklock" for forcing the "victim" account. Furthermore, the modalities through which the algorithm evolves are made and this is expanded throughout the entire telematic network. The next three phases are subordinated to this one;

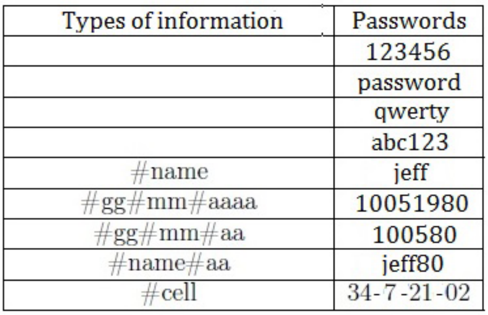
b) *violation of the mail account:* in this phase we use the information found in the previous one to try to violate the mailing systems;

c) *violation of other systems:* in this phase we expand to other systems, such as online payment and e-commerce;

d) *violation through the referral of credentials:* in the last phase we try to violate the security measures of other sites by resending access credentials by email.

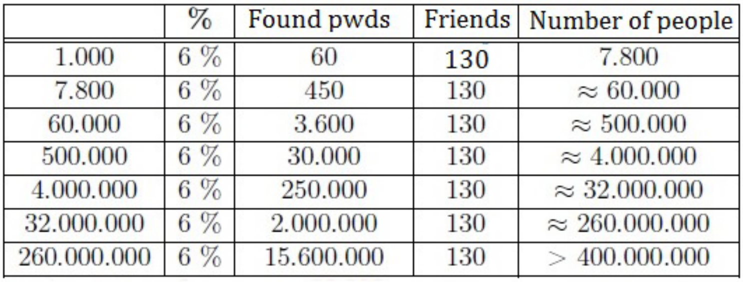
a) In the case of Facebook, a profile from which to start the attack is created; the idea is to reach the maximum limit of friendships before starting the attack. Thanks to the Facebook API it is possible to find and store in a database the list of friends with all information associated with them.

Some of the information that Facebook makes available are: user ID, about me, activities, affiliations, birthday dates, books, contact email, current location (city, state, country), education history, family members, interests, list of desired relationship types corresponding to the "Looking For" profile element, list of desired relationship genders corresponding to the "Interested



Types of information	Passwords
	123456
	password
	qwerty
	abc123
#name	jeff
#gg#mm#aaaa	10051980
#gg#mm#aa	100580
#name#aa	jeff80
#cell	34-7-21-02

Fig. 3. Most frequently used passwords.



	%	Found pwds	Friends	Number of people
1.000	6 %	60	130	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

Fig. 4. Statistical analysis of the expansion.

In" profile element, "Favorite Movies" profile field, "Favorite Music" profile field, notes written by the user, pics, "Political View" profile field, URL of the Facebook profile of the user, with username included in the URL, "Favorite Quotes" profile field, "Relationship Status" profile field, "Religious Views" profile field, ID of the person the user is in relationship with, time-zone, total number of posts to the user's wall, "Personal website" profile field, list of work history information.

Once all the information about users is collected, the access credentials are identified, making attempts for each "friend", using the previously discovered e-mail and proving the most probable passwords (Figure 3).

Statistically this type of attack should lead to identify about 8% of passwords or more; for each violated account this operation is recursively repeated, expanding the algorithm (Vitali, 2010). Starting as example by a user with 1000 friends, considering the percentage of found passwords of 6% and assuming that each hit user has an average of 130 friends, the possible expansion of the attack follows numbers shown in Figure 4.

b) Once this phase is over, the mail account is violated; a very widespread risky method is to use the same password for different online services:

b1) for each violated account, try the same password for e-mail accounts (over 61% successful);

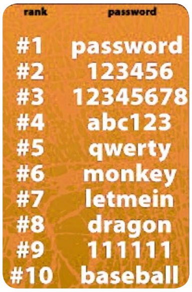
b2)once logged in, perform search operations in the mailbox (receipt/sent/ trashed mails) looking for login credentials, card numbers, compromising mails, etc. A widespread bad habit is to store in their mail the passwords of other systems.

c) In the third step, we go to the violation of other systems; for each account violated the idea is to try the same password for paypal account, ebay, and other sites. Most online banking customers re-use their credentials to access much less secure websites.

d) In the fourth phase, passwords are re-sent with the request of sending passwords by email through the "Password forget" action. Systems like Facebook, Gmail, Paypal, Vodafone, re-send the password or allow to reset it via a link sent by email.

About the complexity of passwords

The percentage of easily identifiable passwords is about 8%, first trying with simple trivial passwords and then using targeted passwords based on the data collected on social networks, in particular starting by the date of birth which, linked in different ways with the username, gives rise to passwords to be tested. Tests carried out on a sample of 100 real users showed that the percentage was pessimistic because the algorithm led to the discovery of 14% of passwords (Bonneau, 2018).



rank	password
#1	password
#2	123456
#3	12345678
#4	abc123
#5	qwerty
#6	monkey
#7	letmein
#8	dragon
#9	111111
#10	baseball

Fig. 5. Worst passwords of 2012
Source: splashdata.com, 2018.

"SplashData", a company that produces software for password management, published the list of the worst passwords used by users in 2012. We can note that many passwords (Figure 5) are exactly those previously indicated (*splashdata.com*, 2018).

From studies on the perception of risk by those who do online shopping (Hove, Ray, Roberts, Urbanska, & Byrne, 2012), 85% of people say they use very strong passwords with combinations of letters, numbers and symbols; in fact 24% use words that are part of a dictionary and another 24% of passwords contain personal information. We understand that the user does not have a clear conception of the true meaning of secure password. Continuing the analysis we get other interesting information on user behaviour: only 22% have never re-used a password on another site and 51% say they never change their password or change it very rarely.

Analyzing data from another study (Devillers, 2010) about the stolen passwords of 32 million users by a hacker in 2009 from the "RockYou" database, we note again how users use unsafe passwords; most of them are made up of lowercase characters, numbers or simple concatenations of these two, while the passwords considered as "stronger" are only 9%.

Although users are informed that the databases containing their passwords have been compromised and therefore their passwords are made public, the percentage of them who decide to change the password remains very small. Then, to improve the password search algorithm, one could insert in the new database also all usernames and passwords that in the course of these years have been published by hackers.

Since the percentage of users who re-use the same password on different sites is about 80%, we could search if in the created database is present the user of which we are looking for the password and in this case to insert in the list of passwords to be tested those that she/he has already used on other websites.

Password choice policies do not help security; in fact the only given limit to new subscribers is normally that the chosen password is at least 6 characters and that these are not all the same; on the contrary, obliging the user at the time of registration to enter a password with alphanumeric characters (both uppercase and lowercase), numbers and special characters, many attacks would be completely ineffective.

Social phishing

"Phishing" is an attempt to acquire confidential information such as username, passwords and credit card numbers, pretending to be a trustworthy entity in an electronic communication. The standard process of phishing attack methods can be summarized in the following steps (Jagatic, Johnson, Jakobsson, & Menczer, 2007):

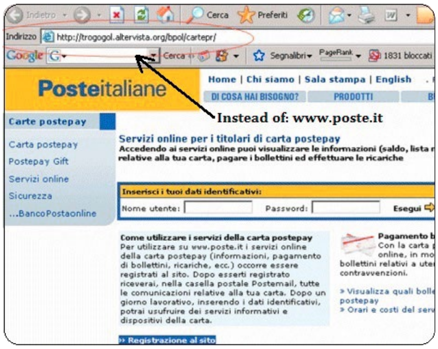
a) the bad-intentioned user (phisher) sends to the user an e-mail message that simulates, in graphics and content, that of an institution known to the recipient (for example, her/his bank, her/his web provider);

b) the e-mail contains normally notices of "particular situations" or "problems" occurred with the current account of the person (for example a huge charge, the expiry of the account) or a money offer;

c) the e-mail invites to follow a link, present in the message, for avoiding the charge and/or for regularising the position with the institution or company whose message simulates the graphics and the setting;

d) the provided link does not lead to the official website, but to a fictitious copy apparently similar to the official website, located on a server controlled by the phisher, in order to request and obtain particular personal data, normally with the excuse of a confirmation or the need to perform the system authentication. This information is stored by the phisher server;

e) the phisher uses this data to buy goods, transfer money or even as a "bridge" for further attacks (Figure 6).



Instead of: www.poste.it

Servizi online per i titolari di carta postepay
Accedendo ai servizi online puoi visualizzare le informazioni (saldo, lista r
relative alla tua carta, pagare i bollettini ed effettuare le ricariche

Inserisci i tuoi dati identificativi:
Nome utente: Password:

Come utilizzare i servizi della carta postepay
Per utilizzare su www.poste.it i servizi online della carta postepay (informazioni, pagamento di bollettini, ricariche, ecc.) occorre essere registrati al sito. Dopo esserti registrato riceverai, nella casella postale Postemail, tutte le comunicazioni relative alla tua carta. Dopo un giorno lavorativo, inserendo i dati identificativi, potrai usufruire dei servizi informativi e dispositivi della carta.

Pagamento b
Con la carta p
online, in mo
bollettini relativi a ute
contravvenzioni.

» Visualizza quali bolle
postepay
» Orari e costi del serv

Fig. 6. Example of e-mail phishing.

Such attacks have generally not high probability of success, because often these e-mails are written with automatic translators, so there are some errors, but also because users who are offered money or too advantageous offers tend not to trust. In addition, browsers have security systems that are able to recognize and block many phishing pages and thus contribute to keeping this percentage low.

There is also a special phishing technique, called "spear phishing", directed to a particular person or group (*trendmicro.com*, 2018); the attacker tries to gain the victim's confidence by using information about her/him, studying habits, interests, what she/he buys online, what her/his bank is, the name of her/his parents, wife, children. All this information can be retrieved with research on the Internet, and with the advent of social networks to find this information has become even easier.

"Social Phishing" deals with phishing linked to data collected on social networks, i.e. how easily a phisher can use social network data to increase the performance of his attack. In a study, data were acquired from various social networks and a database with tens of thousands of relationships was quickly and easily created, using crawling and parsing tools such as the "Perl LWP library", accessible to anyone (Gupta, Arachchilage, & Psannis, 2018; *search.cpan.org*, 2018).

The sender's e-mail address is altered and the recipient is invited to connect to a university page where she/he is asked for her/his credentials; in the first case the sender is a stranger with a university e-mail address, in the second case the address is a friend, thanks to the information obtained from social networks and previously saved in the database. In the first case, 16% of users have provided their own credentials, but the really interesting data is related to the second attack, the one that used the information of social networks; in this case the percentage of success was 72%.

Given the success rate of the type of attack described above, one could think to optimize the algorithm by exploiting this technique. Once the first phase of the attack (in which the 8 passwords are tested) is over, the process goes with the normal algorithm with the users whose passwords have been discovered, while for the rest of users the previously collected data in the database could be used for creating personalised messages in which users are invited to click on a specific link, that connects the unsuspecting user to a fake Facebook login page. In that page, if victims enter their data, they are sent to the attacker and to the victims appears a message stating that the entered credentials are incorrect and then redirected to the real Facebook page, for not creating suspicions.

To these "new victims" one could then launch an attack of the first type and to "survivors" to try again with the "social phishing" technique described above, then alternating the two attacks. It would thus arrive to attack a very high number of users within a few iterations of the algorithm.

Conclusions

Many works have been written about the evaluation of the efficiency of telematic attacks. From 2010 to date, Facebook privacy policies have been changed, but thanks to new services it is still possible to trace personal information. With special phishing techniques, that use information collected on the Internet and on social networks about users and their relationships, it is possible to get the access credentials of a good percentage of users. This allows attackers to perform online transactions, view bank accounts and their transactions, call details, credit card numbers and many other personal data.

Waiting for the power of the future quantum Internet and applications of quantum information, it is therefore unfortunately possible today again to launch an attack exploiting the analyzed techniques and even improve them, making them more effective and reaching even higher success rates, thus placing a very high number of users in serious danger (Di Sia, 2011; Di Sia, 2017).

REFERENCES

Bonneau, J. (2018). *The science of guessing: analyzing an anonymized corpus of 70 million password.* Retrieved May 01, 2018, from: http://www.jbonneau.com/doc/B12-IEEESP-analyzing_70M_anonymized_passwords.pdf.

Devillers, M. M. A. (2010). *Analyzing Password Strength.* Retrieved May 01, 2018, from: https://www.cs.ru.nl/bachelorscripties/2010/Martin_Devillers___0437999___Analyzing_password_strength.pdf.

Di Sia, P. (2011). Extreme Physics and Informational/Computational Limits, *Journal of Physics: Conference Series*, 306, 012067 (8 pp.).

Di Sia, P. (2017). Looking at the Quantum Internet. *E-methodology, 4*, 31-35, doi: 10.15503/emet2017.31.35.

enisa.europa.eu. Retrieved May 01, 2018, from: https://www.enisa.europa.eu.

fortune.com. Retrieved May 01, 2018, from: http://fortune.com/2015/08/18/facebook-google/ Gupta, B. B., Arachchilage, N. A. G. & Psannis, K. E. (2018). Defending against phishing attacks: taxonomy of methods, current issues and future directions, Telecommunication Systems, 67, 247. https://doi.org/10.1007/s11235-017-0334-z.

Hove, A. E., Ray, I., Roberts, M., Urbanska, M., Byrne, Z. (2012). The Psychology of Security for the Home Computer User, *IEEE Symposium on Security and Privacy*. Retrieved May 01, 2018, from: http://www.cs.colostate.edu/psysec/papers/SSP.pdf.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., Menczer, F. (2007). Social Phishing, *Communications of the ACM, 50(10)*, 94-100, doi: 10.1145/1290958.1290968.

Mediascope Europe. (2012). *Pan-European Launch Presentation Summary*. Retrieved May 01, 2018, from: https://iabspain.es/wp-content/uploads/Mediascope_europe_2012_pan-european_launch_presentation_summary.pdf.

Salucci, S. (2012). Privacy su Facebook (Privacy on Facebook). *Thesis*. Bologna: Alma Mater Studiorum, University of Bologna (Italy).

search.cpan.org. Retrieved May 01, 2018, from: http://search.cpan.org/dist/libwww-perl/lib/LWP.pm.

splashdata.com. Retrieved May 01, 2018, from: http://splashdata.com/press/releases.htm.

trendmicro.com. Retrieved May 01, 2018, from: https://www.trendmicro.com/vinfo/us/security/definition/spear-phishing.

vincos.it. Retrieved May 01, 2018, from: http://vincos.it/world-map-of-social-networks/.

Vitali, A. (2010). L'11 settembre telematico. Rischi di sicurezza causati dai social network (Telematic September 11th. Security risks caused by social networks), *Thesis*, Alma Mater Studiorum, University of Bologna (Italy).

Yue, J., Li, Z., Liu, L., Fu, Z. (2011). Content-based image retrieval using color and texture fused features, *Mathematical and Computer Modelling, 54(3-4)*, 1121-1127, https://doi.org/10.1016/j.mcm.2010.11.044.

zephoria.com. Retrieved May 01, 2018, from: https://zephoria.com/top-15-valuable-facebook-statistics/.