2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications

Social network content management through watermarking

Athanasios Zigomitros[1], Achilleas Papageorgiou, Constantinos Patsakis[2]

[1]Department of Informatics, University of Piraeus, Greece,
[2]Department of Computer Engineering & Mathematics,
Rovira i Virgili, Catalonia, Spain

***Abstract*—Due to the rise of social media, several new needs, problems and challenges have emerged in users' privacy and security policies. Two very serious problems that should be addressed are identity theft and unauthorized content sharing. In this work we propose a more secure scheme for privacy in social networks by the use of watermarking that manages to diminish these problems, at least inside current social media architectures, without the need for building them from scratch.**

***Index Terms*—Social networks, digital watermarks, identity theft, privacy**

I. Introduction

The recent technological advances in cloud computing and web services coupled with the increase of speed of Internet connection have created the necessary requirements to develop new services that allow the transmission and exchange of big multimedia files. In the new web 2.0, we have seen the development of many social media services. Blogs, wikis and social networks have evolved the everyday tools to communication and publication platforms for millions of users around the globe. Currently Facebook, Google+, LinkedIn, Twitter and other well known platforms count millions of subscribed users [18], drastically changing the way that people are connected and exchange information. Yet, many concerns have risen, regarding to the security and privacy that they can provide to their users [10], [11], [15], [16], [17].

A big part of social media consists of multimedia content that is published and exchanged between the registered users. Many problems stem from this flow of information as malicious users are able to exploit the features that these new services give them. New offenses ranging from identity theft to copyright infringement and from personal information exposure to medical history disclosure are being made everyday. It becomes obvious that reposting and republishing of images and multimedia content without any form of ownership, not necessarily about copyrights, can in many cases mislead many users, while on the same time harm the original owner socially and economically.

This work mainly focuses on the social networks (SNs) and the authentication of images that are being published on them. The lack of any authentication of the media content for sure enables malicious entities to publish content that bypasses the desired privacy policy of the users. Moreover, we discuss how could such policies be implemented using digital watermarks, allowing SNs to become more privacy aware, without the need to build them from scratch.

The term of ownership as it is going to be used throughout this work has more to do with privacy than with property. It is needless to say that people who share some of their digital content on a SN, do not want monetary exchange for it, otherwise they would sell it on such a platform. These people own a photo as they have taken it from their camera for example depicting and instance of their lives. This means that this thing is a part, or belongs to their private lives. Hence, owning the content on a SN, from the user side, has nothing to do with trade, but showing trust and applying privacy policies.

The structure of this work is the following. After this short introduction we discuss the problems of identity theft and leakage of personal information through SNs. Afterwards we illustrate the experimental results from tracing image distortions after uploading them on two of the biggest SNs, namely Facebook and Google+. The next section focuses in watermarking, just to give the necessary background for our proposal. The proposed method is illustrated in the following section and finally we end up with our main conclusions and ideas for future work.

II. The problem

Apart from the advantages of using social media, there are many disadvantages. In order to focus more on what this work is going to cover, we will refer to some of them, in order to discuss later probable solutions. One of the basic problems that have been augmented due to the rise of social media is identity theft [20]. In many cases people are creating fake profiles in order to maliciously manipulate other people or to harm the social image of others. The problem stems from the fact that one can upload a photo from another's profile and then using social engineering to send users to the fake profile.

Another problem that has risen is the leak of private information. In one way or another, all major social media platforms have implemented some sort of privacy control, yet in most cases they can be easily circumvented by the users. In order to state the problem more clear we give an example. User A has a personal photo that wants to share it only with users B, C and D, so he uploads the picture and sets the desired policy. Now user B can send the photo to everyone inside and outside of the SN, as in most cases the link is static, so even though user A wants only three people to see this picture, everyone else can have access to it after user's B publication. On the other hand, user C may download the photo and upload it to his profile, sharing it with everyone else, without any notification

978-0-7695-4745-9/12 $26.00 © 2012 IEEE
DOI 10.1109/TrustCom.2012.264

to user A. Finally, user D might make some alternations on the original photo and upload it on his profile. Clearly, user's A policy on his picture has been bypassed and he hasn't been warned about it at any point.

III. Experiments

A. The process

In order to test whether current social media have any form of watermarking on the digital content they receive, we conducted some experiments on two of the most widely used SNs Facebook and Google+. In the scope of this experiment we used two groups of images (Test set 1 and Test set 2) and two user accounts. Each image group had the necessary characteristics and properties that could point any hint of image distortion when sharing images on a SN. Therefore, these images were divided by their color, their source (computer generated vs camera) and their resolution.

The first group of images, Test set 1, consists of 40 gray scale computer generated test images, that are being used in image processing test [9]. Twenty of them are 1200x1200 pixels and the other 20 are 600x600 pixels. The second group of images, Test set 2, consists of 40 color images that could be characterized as typical user images. The group consists of 20 images above 1200x1200 pixels, which range from 2048x1536 pixels to 3648x2736 pixels. From these high resolution pictures, 7 were taken from the camera of an Apple iPhone 3GS, 6 had been taken from a Casio EX-Z1050 camera, 4 from a LG KU 990i mobile and 3 with a Cannon IXUS 130 camera. The other 20 images were again from TESTIMAGES [9], 10 images of 1200x1200 pixels and 10 of 600x600 pixels.

The process we used in our experiment to trace the changes in our samples can be seen in figure 1 and is divided in the following steps.

1) Two accounts were created for users A and B, one in each SN.
2) The images were uploaded on 27/10/2011 on each user account. In the case of Facebook High Resolution uploading was applied.
3) Each image was then downloaded several hours afterwards.
4) Firstly the downloaded images were compared each against the other's user, in order to trace probable user traces.
5) Then the images were tested for differences compared to the original ones (filesize, resolution).
6) The next step was to use Matlab in order to trace differences in basic image characteristics: Mean of Mean Square Error, Mean of Peak Signal to Noise Ratio, Mean of Normalized Cross-Correlation, Mean of Structural Content, Mean of Average Difference, Mean of Maximum Difference and Mean of Normalized Absolute Error.

B. Image distortion

The comparison between the downloaded users' images showed that there was no difference in their size and resolution. The next test was regarding the differences of the

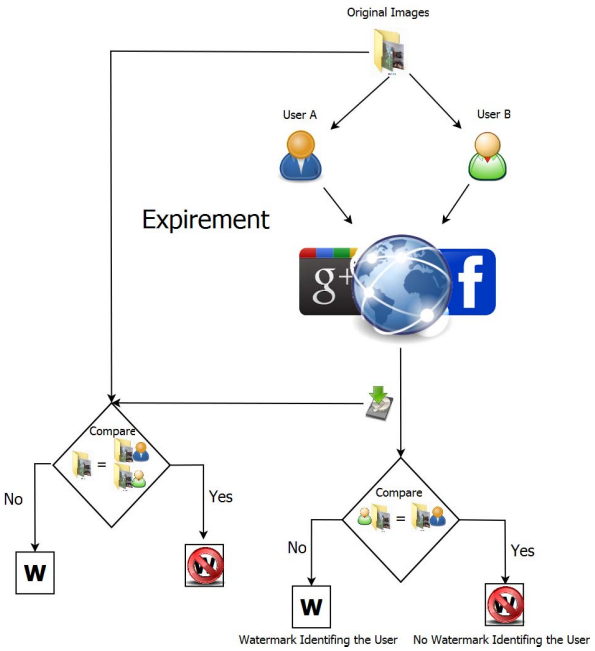


Figure 1: The experimental process.

	Test Set 1		Test Set 2	
Value	Facebook vs. Original	Google+ vs Original	Facebook vs. Original	Google+ vs Original
Mean Square Error	17,7139	0	14,8287	0
PSNR	42,3189	Inf.	41,4107	Inf.
Normalized Cross-Correlation	1,0014	1	0,9992	1
Structural Content	0,9974	1	1,0005	1
Aver. Difference	-0,5496	0	-0,0476	0
Max. Difference	34,025	0	55,5926	0
Normalized Absolute Error	0,0137	0	0,0261	0

Table I: Mean values of basic image characteristics,
The table refers to the images that had no change in their resolution.

downloaded images compared to the original ones. In figure 2 we present the histogram regarding the differences in file sizes for Test Set 1. It is obvious that the test set images had no difference with the original ones in their filesize when they were uploaded on Google+. On the other hand, in most of them we notice a reduction on their filesize, when they were uploaded on Facebook.

In the case of Test Set 2 we can see many image differences. The main reason seems to be the fact that both SNs have a bound on the resolution of the images that can be hosted. The bound, at the time of the experiment was 2048x1536, or the reverse 1536x2048, depending on the orientation of the picture. Above this bound, the images are resized to fit the optimal resolution within it. Again, in figure 3, we observe that Google+ does not make any change in the image size if the image is within the boundary. In the Facebook case, we see a big reduction in the filesize, even if the image was of the appropriate size. The distortion on several image characteristics is summarized in Table I.

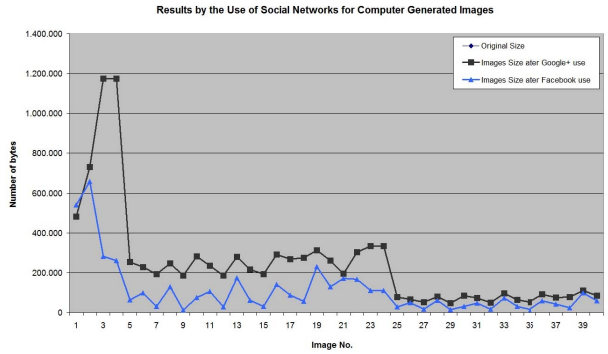


Figure 2: Test set 1, image sizes.

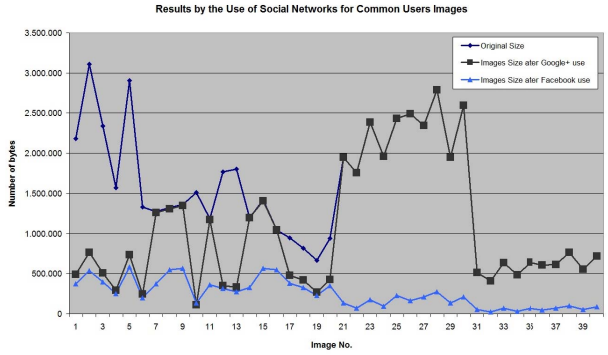


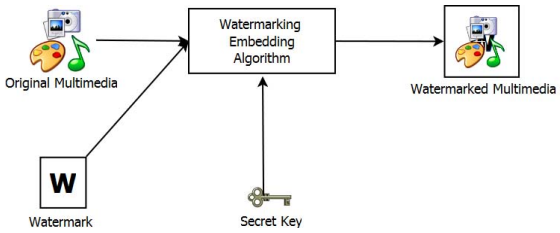
Figure 3: Test set 2, image sizes.

C. Comments on the results

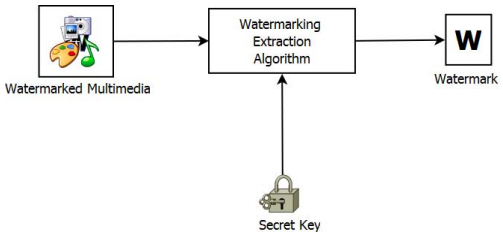
The results that were taken from both these SNs can lead us to several helpful conclusions. In Google+, when the image resolution does not pass a certain threshold, the uploaded image is exactly the same with the original one. So no watermarking is being applied. Moreover if the image resolution exceeds the aforementioned threshold, the image is resized, yet, the image is exactly the same for both users. Hence we may assume that no watermarking is being applied by Google+ on the uploaded images at any case.

In the case of Facebook, we notice that we have several differences when comparing to the original one. This of course could hint the existence of a watermarking scheme. Yet, we observe that if we upload the same image on two different user profiles, the hosted image is the same in both cases. Hence either in both cases we have the same watermark, or we do not have any watermark at all. It is obvious that if such watermark existed, then it would contain something like user ID or timestamps, photoID etc. Since in the conducted experiment we had two user accounts, uploading their photos at different time, nor the user ID nor the timestamps can be the same. Therefore, we can safely assume that Facebook does not watermark the uploaded images, but compresses them in order to gain some storage space on its servers.

By looking at figure 3 again, we can see that for the images that exceed the resolution threshold, both Google+ and



(a) Embedding process.



(b) The extraction process

Figure 4: Watermarking.

Facebook apply a similar algorithm for image resizing and reconstruction, as the sizes are almost identical, with the one from Facebook being a bit more efficient.

Modern cameras have very good image resolutions that exceed the threshold that is set by both of the SNs that were tested. Taking into consideration that most images that are submitted on SNs are taken from cameras and the fact that they will be resized by SNs to be hosted, the assumption that users do not mind for minor image distortions, as long as it will not destroy their content is valid. Hence, they would be willing to trade some of their image resolution or of their content quality generally, if this could give them some extra features from the service.

IV. Watermarking

Steganography is the art of hiding content in media, so that it cannot be detected by not trusted entities. On the other hand, digital watermarking is the process of embedding information into media in order to prove the origin of the content. Since images and audio can be very well described as signals, we embed the information in the signal, in order to achieve the least possible distortion of the valuable information, e.g. LSB, DCT coefficients etc. Nowadays watermarking has been proposed as a solution mostly for proving ownership but also for copy control, fingerprinting and tamper detection, therefore we see it applied in many Digital Rights Management (DRM) implementations [2].

In figure 4, a typical structure of a watermarking system is presented. A secret key is used to scatter the watermark in the multimedia content, so that the content is not significantly distorted, it cannot be removed and only the secret key can prove its existence and origin.

In this work in order to achieve the objectives of our proposal the watermarks should have the following properties.

a) Invisible: The embedded watermarks should be invisible. In contrast to visible watermarking in the invisible

watermarking the original multimedia must change in a way that would be imperceptible by the human visual system or the auditory system in case of sound.

b) Blind: Watermarking algorithms which do not require access to the original multimedia for detecting and extracting the watermark are called blind, if they need access then they are called not-blind. In a non-blind algorithm for a SN the space to store the original and the watermarking multimedia can double the needed storage and in case that we decide to save storing space by embedding the watermark on-the-fly, that can have an extreme computational cost. In our approach we suggest the use of a blind algorithm and the original content to be only in the user's "hands".

c) Robustness : Depending on the application the watermark can be fragile, semi-fragile or robust.

- Fragile watermarks are used when the concern are the complete integrity of the image. Even the slightest modification results to an alert of the watermarking system.
- Semi-Fragile watermarks are used when the concern are only the malicious attacks on the host image and not the common image processing as lossy compression and/or random noise. Any process that has an effect on the content of the image, as cropping or insertion of a new object in the host image, should be noticed by the watermarking system.
- Robust watermarks are mostly used for proving of ownership and that is why they cannot be removed easily and without great degradation of the host image. They must be able to defend against in a wide range of possible attacks.

For more on watermarking and possible attacks the reader may refer to [1], [12], [13], [14], [7], [8].

V. The Proposed Solution

SNs in which users trust their multimedia files, can be considered either as open or closed systems, that have full access to alter the uploaded files. The majority of the users seem not to mind about this kind of distortions, as long as the content is available and without visible distortions to proper correspondents, issued by them. The SNs' approach in a conflict of multimedia ownership and misuse is so far to let users report the offenders. This approach obviously has many disadvantages, as it lets anyone reporting everyone, whether they are the original owners of the content or not. Moreover, a user can report such misuse only when he becomes aware of it by others or by sheer luck. The SNs currently do not have a sort of policy of notifying users and taking precautional measures about such problems.

Addressing to this problem, we propose a dual watermarking scheme as the first line of defense for both SNs and users [5], [6]. Of course user reporting still remains a valuable function in SNs, but must be used for problems that really require human interference, for example if someone takes a picture of someone else without his consent, or if the uploaded photo is offensive and misuses the service. As we will see a lot of problems can be solved automatically and with users' notification and awareness.

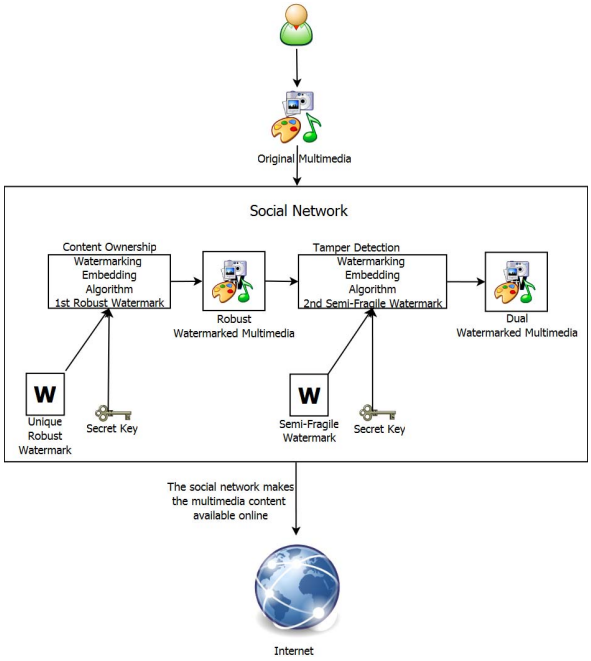


Figure 5: Dual watermarking Scheme

The proposed scheme for SNs, as mentioned previously uses dual watermarking for the multimedia content, one robust and one semi-fragile. In order to clarify the scheme we use the following scenario. User A provides the original multimedia content to the SN. Then the SN embeds a robust watermark, which identifies the multimedia content uniquely and relates it to the user A. Afterwards a second semi-fragile watermark is embedded in the content. The dual watermarked content is then stored on the SN and shared among the users, according to privacy settings set by the user.

The role of the first watermark, the robust, is to identify the owner of the content, so that it can be traced even if the content is tampered. The semi-fragile watermark does not break the first watermark, but on the same time enables the detection of possible alternations from other entities. Figure 5 describes the dual watermarking embedding process.

Let's suppose that user A is the original owner of some multimedia content and he uploads it to a SN. The SN embeds the dual watermark to his content and then makes it available to the users according to users privacy settings, e.g. public, friends of friends etc. User B is a user that has the right to access the multimedia file and since a user can access it he can store it to his computer. Meanwhile, the link that user B has for the content is not static. The use of dynamic links for contents in SNs must be used at any time, as the static links are the most obvious way that users can bypass any sort of privacy policy. The case now is "you see it once you own it forever" no matter if one changes his privacy policy. Moreover, a static link is easy to be copied and shared not only inside a SN, but in the whole Internet as well.

Now let's assume that user B tries to re-upload the content, to the same SN, with or without making any changes to it. The

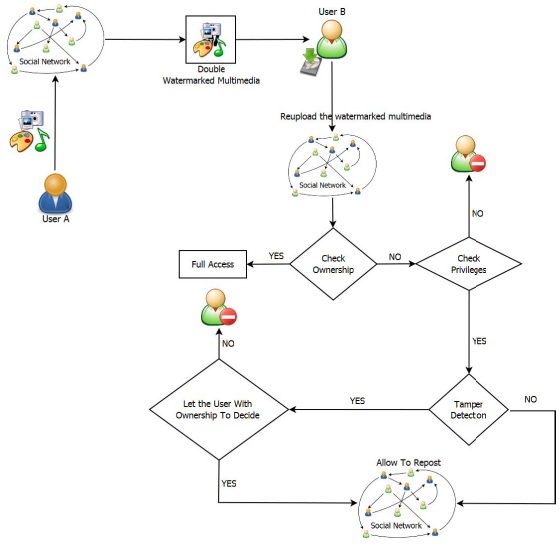


Figure 6: The proposed scheme.

submitted file is checked from the SN for the existence of the first watermark, the robust, which shows owner of the content. If there isn't one, the file is watermarked and the ownership goes to user B. If there is one then the SN checks the privacy settings of the user A, who has the ownership and according to it, the content is allowed or not to proceed to next check. If user B has the privileges to repost the content then the last step is to check the integrity of the content with the semi-fragile watermark. If the content has been tampered, then an alarm is triggered for user A, showing the altered version of his content requiring his consent for resubmission. In order to avoid possible problems, a logical timeframe for this answer is being applied, so that if user A hasn't answered for e.g. a month then this means that he doesn't care for this post, therefore it is automatically published. In case user B has the right to submit the file, user A just receives a notification for the event.

A different approach would be embedding of watermarks on the fly, every time an authorized user is granted access to the content, figure 7. This allows the use of non-blind algorithms, which are more robust, while enhancing the watermark system with fingerprinting capabilities. On the fly watermarks could include the user ID of the user that gains access on the content. Therefore if the content leakage has been made by some user to another SN or the Internet generally, it can be traced and settled more easily. The obvious trade-off of this approach is of course the computational cost on the server side.

VI. Conclusions

The experimental results from two of the biggest SNs indicate the lack of any watermarking mechanism. Moreover, the fact that SNs have gained so much attention and have such an effect on our daily lives, make at least essential the development of new security and privacy policies, as the current ones prove to be inefficient. Towards this end our work aims to introduce a very well known technique,

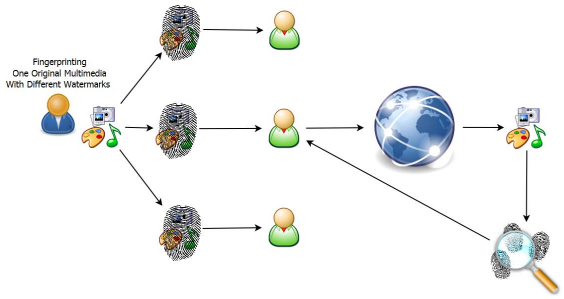


Figure 7: User fingerprinting.

digital watermarking to SNs making them more secure and trustworthy. Obvious techniques like dynamic links to content have not yet been implemented, exposing users sensitive data, while allowing malicious entities to do their work easier. Of course there are other researchers pointing towards the use of watermarking in SNs [19], yet to our knowledge no experiments on the existence of watermarks in SNs content has been published, nor has a formal protocol or policy has been proposed as in this work, neither has it been implemented by SNs.

The implementation of the proposed solution, does not require the complete redesign of current SNs, therefore it can be easily adopted. One could argue that this leads to DRM practices in SNs, yet the scope is not corporate development or tracking user's traffic . The watermarking is only made to protect users' content, moreover there is no reason why users shouldn't be allowed to opt in or out of this service for some of their shared content.

A problem of the proposed solution seems to be who has "owned" first the content, meaning that if an image for example belongs to user A, yet user B uploaded first, it would seem that it belongs to user B, so user A should report it in order to settle the dispute. For sure such a scenario is realistic, yet we can see that the balance of what can be automated from the proposed solution and what is left on the human factor is drastically changed, leaving far less problems to be solved manually. Additionally, current solutions do not offer at any time such alarms.

A wider view on the subject could be on the future create a link between all major SNs, so that users, even if they are registered to one network, become aware of misuse of their content in other networks as well, providing a more holistic approach to privacy for everyone.

Appendix

Figures

The icons for the figures were taken

Figures 1, 5, 6, 7. David Vignoni, http://en.wikipedia.org/wiki/File:User.svg

Figures 1, 6, 7. The Tango Icon Team, https://commons.wikimedia.org/wiki/File:User_icon_2.svg

Figures 1, 6. Mikael Haggstrom, http://en.wikipedia.org/wiki/File:Download.svg

Figure 1. Mikadiou, http://en.wikipedia.org/wiki/File:Interdit_forbidden.svg

Figures 1, 5, 7. The Tango Desktop Project, http://commons.wikimedia.org/wiki/

Figures 4, 5, 6, 7. IvanLanin, http://commons.wikimedia.org/wiki/File:Nuvola_multimedia.png

Figure 6. David Vignoni, Stannered,Korrigan, Available: http://en.wikipedia.org/wiki/File:Blocked_user.svg

Figure 7. Wilfredor, http://en.wikipedia.org/wiki/File:Fingerprint_picture.svg

Figure 7. http://openclipart.org/detail/94687/fingerprint_search_enhanced

REFERENCES

[1] I. J. Cox and M. L. Miller, "The first 50 years of electronic watermarking", EURASIP J. Appl. Signal Process, 2002, 2, pp.126-132, Feb. 2002.

[2] I. J. Cox, M. L. Miller and J. A. Bloom, "Watermarking applications and their properties", Proc. The International Conference on Information Technology: Coding and Computing (ITCC' 00), Washington, DC, USA, IEEE Computer Society, 2000, pp.6-10.

[3] F. Hartung and M. Kutter, "Multimedia watermarking techniques" Proc. of the IEEE , vol.87, no.7, July 1999, pp.1079-1107.

[4] M. D. Swanson, M. Kobayashi and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies", Proc. of the IEEE , vol.86, no.6, Jun. 1998, pp.1064-1087.

[5] F. Mintzer and G. W. Braudaway, "If one watermark is good, are more better?", Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999, vol.4, no., 15-19 Mar. 1999, pp.2067-2069.

[6] N. P. Sheppard, R. Safavi-Naini, and P. Ogunbona, "On multiple watermarking", Proc. ACM Workshop on Multimedia and Security 2001, New York, NY, USA, 2001, pp. 3-6.

[7] F. A. P. Petitcolas , "Watermarking schemes evaluation", Signal Processing Magazine, IEEE , vol.17, no.5, pp.58-64, Sep. 2000.

[8] S. Voloshynovskiy, S. Pereira, T. Pun, J. J. Eggers and J. K. Su, "Attacks on digital watermarks: classification, estimation based attacks, and benchmarks", Communications Magazine, IEEE , vol.39, no.8, pp.118-126, Aug 2001.

[9] TESTIMAGES. [Online]. Available: http://sourceforge.net/projects/testimages/files/

[10] ENISA, (2007, Oct.) "Security Issues and Recommendations for Online Social Networks", ENISA Position Paper No.1., 2007. [Online]. Available: http://www.enisa.europa.eu/

[11] R. Gross and A. Acquisti., "Information Revelation and Privacy in Online Social Networks", Proc. Workshop on Privacy in the Electronic Society (WPES), Alexandria, Virginia, USA, 2005, pp. 71-80.

[12] I. Cox, M. Miller, J.Bloom, J. Fridrich, T. Kalker, "Digital Watermarking and Steganography", Morgan Kaufmann, 2007.

[13] P. Wayner, "Disappearing Cryptography, Information Hiding: Steganography & Watermarking", Morgan Kaufmann, 2008.

[14] S. Katzenbeisser, and F. A. Petitcolas, "Information Hiding Techniques for Steganography and Digital Watermarking", Artech House, 2000.

[15] N. Li, N. Zang and S. K. Das, "Preserving Relation Privacy in Online Social Network Data", IEEE Internet Computing Magazine, vol. 15, pp.35-42, May-Jun 2011.

[16] K. Strater and H. Richter, "Examining privacy and disclosure in a social networking community", Proc. 3rd Symposium on Usable Privacy and Security, SOUPS '07, ACM, NY, USA, 2007, pp. 157–158.

[17] L. Backstrom, C. Dworkand and J. Kleinberg, "Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography", Proc. The 16th International Conference on World Wide Web, WWW '07, 2007, pp. 181-190.

[18] Wikipedia list of Social Networking websites. [Online]. Available: http://en.wikipedia.org/wiki/List_of_social_networking_websites

[19] P. Chomphoosang, P. Zhang, A. Durresi and L. Barolli, "Survey of Trust Based Communications in Social Networks", Proc. The 14th International Conference on Network-Based Information Systems (NBiS-2011), Tirana, Albania, 2011, pp. 663-666.

[20] Identity Theft Resource Center, http://idtheftcenter.org/