



Analyzing Tools and Algorithms for Privacy Protection and Data Security in Social Networks

A. Mohammadi^a, H. Hamidi^{*b}

^a Department of Information Technology Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran

^b Department of Industrial Engineering, Information Technology Group, K. N. Toosi University of Technology, Tehran, Iran

PAPER INFO

Paper history:

Received 07 January 2018

Received in revised form 27 January 2018

Accepted 08 February 2018

Keywords:

Privacy Protection

Social Networks

Information Leakage

Information Disclosure

ABSTRACT

The purpose of this research, is to study factors influencing privacy concerns about data security and protection on social network sites and its' influence on self-disclosure. 100 articles about privacy protection, data security, information disclosure and Information leakage on social networks were studied. Models and algorithms types and their repetition in articles have been distinguished and this study builds a research model to examine privacy concerns and the effect of it on self-disclosure. The need of having knowledge and skill about privacy protection seems to be necessary with social networks technology developments. Most of the researches have been studied about privacy protection scope related to users' privacy on social networks including women, men, children and adults in smartphone and E-health. Most of researches on this scope have been done in USA. Most studies were focused on privacy protection and security on social networks.

doi: 10.5829/ije.2018.31.08b.15

1. INTRODUCTION¹

One of the reasons of information self-disclosure is narcissism which they prefer to share their information in social networks to provide an attractive platform of their leaning [1]. Another reason is adolescents and specially children doesn't have enough information about how to protect privacy, that with no limitation they can use internet for completing their homework, checking emails, online chat, online gaming and access various social networks sites [2]. One of the networks that is highly considered in this scope, is health services network which along with technology is mentioned as enabler and guidance, named electronic health that can gather personal health information in all health's scopes [3].

Pervious findings are as follows, it's more probable that adults use privacy protection management strategies. Also, women use privacy protection management strategies with more probability than men and there is no meaningful effort between people's job and the way they protect their privacy [4]. Adolescents

disclosed their information easier and more than adults and they use less privacy protection settings [5]. Parents mediation doesn't have any relationship with privacy protection, beside training mediation have more effect on protecting privacy [6]. 50 articles have been studied about user's privacy protection in social networks that here are more related articles in Table 1.

Most of smart phone's applications wants user's location, this shows that we have too much distance with protecting data security [7].

The uniqueness of this research is that it uses empirical testing, combining all the privacy concern influencing factors and influence the behavior trends of self-disclosure of information on social network sites. The outcome of this study reveals an evaluative framework for measuring privacy concern and predicting the probability of disclosing private information.

This study is organized in five sections. Following on from this introduction, the second section reviews the theoretical background and research model with hypothesizes.

*Corresponding Author Email: h_hamidi@kntu.ac.ir (H. Hamidi)

TABLE 1. Concerns about privacy protection on social networks

No.	Ref.	Concentration	Findings
1	[8]	Concerns about privacy protection	Protoss test different variables to see if any one disobeys the rules or not
2	[9]	Privacy protection and children privacy	The need for reform the children safety rule and online privacy protection in Europe
3	[10]	Privacy protection	The effect of age and trust on understanding privacy risks
4	[11]	abuse of privacy protection	Direct relation of privacy protection with knowledge and information

The methods are established in section 3, followed by the data analysis, and results are discussed in section 4, and finally a conclusion and future research are presented.

2. LITERATURE SURVEY

2. 1. Security and Privacy of Smartphone and Mobile Commerce

In most of these studies, researchers have been tried to secure mobile commerce and simultaneously try to protect mobile user’s privacy security in commerce [12]. The most important part of bigdata security in mobile data centers is data size, traffic and delay which should be control from security aspect [13, 14]. Smartphones social networks are useful for connecting social communications with mobile phones [15].

The more parents are educated and have information technology skills the less children’s online privacy would be in danger [16]. In Table 2, 12 articles have been set about security and privacy of smart phone and mobile commerce.

2. 2. Privacy, Self-Protection and Security of E-Health

Smart phone’s health apps, collect their user’s health information.

TABLE 2. Security and confidentiality of smartphones and mobile commerce

No.	Ref.	Concentration	Findings
1	[17]	Mobile user’s privacy	Many mobile users say yes to advertisement cookies.
2	[18]	Mobile privacy	Introduce and analyze M-DGPS algorithm
3	[19]	Mobile privacy protection	Comparability of 4PR protocol with other privacy protection protocols in terms of efficiency

According to the studies women have more concern about their health information [20].

Studies have clearly shown that in what category of audience can see their personal information. But they have got little information about the domain of shared information with users [21]. Table 3 summarized literature regarding confidentiality and security in e-health scope on social networks and information leakage concerns.

As can be seen in Figure 1, 36% of articles are focusing on “concerns about privacy”. The subject that have lowest percentage coverage in the articles is “security and confidentiality in mobile commerce” which is in last place with 2%. According to analyzes which have been done, it seems that providing a flowchart based on how we have chosen articles could be useful.

2. 4 Research Model The research model presented in Figure 2 has three sources of privacy protection: antecedents, privacy concern with social networking sites and applications, and self-disclosure behavior in social networking sites and applications.

2. 4. 1. Privacy Awareness Social networks sites have become part of people’s daily life, their expansion will present new issues in privacy and self-disclosure. Privacy worth have positive effect on self-disclosure in social networks sites [22].

TABLE 3. Confidentiality and security in e-health scope, self-protection on OSN, anonymity on social networks, information leakage concerns and user’s profile privacy

No.	Ref.	Concentration	Findings
1	[23]	Protection of privacy and personal information	Development of privacy protection systems based on mobile phones
2	[24]	Privacy and anonymity on mobile social networks	Sometimes people aren't aware of their information disclosure and doesn't protect it.
3	[12]	Information distribution, information leakage	Privacy information leakage even when users have set the privacy setting truly.

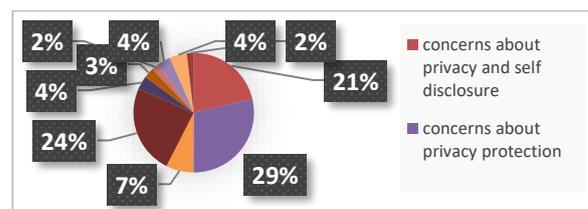


Figure 1. Percentage of covered topics

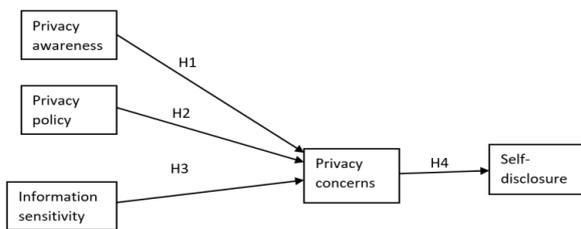


Figure 2. Research model

Research suggests that privacy concerns of individuals may come into play when individuals become aware of how their data is collected and used, and of what their individual rights are concerning that data usage and acknowledgement about privacy have positive effect on its value [6].

Privacy awareness is a construct that measures how much a person is informed about privacy practices in social networking system (SNS). The literature suggests that privacy awareness is related to self-disclosure, privacy value and privacy concerns. This study proposes the following hypothesis:

H1: Privacy awareness has a negative impact on privacy concerns.

2. 4. 2. Privacy Policy Privacy policy is a construct that incorporates users' opinions on how their privacy is protected with regards to SNS privacy policy and if their information is confidential. Nowadays with internet development and electronic messages, privacy is a big concern and the users who are more acquainted with the privacy policy of a website tend to disclose less information [25]. With internet pervasiveness in recent century people in the society have different understanding of privacy therefore they have less limitation [26]. As reported in literature, previous research suggests that privacy policy has had an influence on privacy concerns. This study proposes the following hypothesis:

H2: Privacy policy has a negative impact on privacy concerns.

2. 4. 3. Information Sensitivity Information type sensitivity deals with individual differences with their privacy concerns when the type of information is different. According to the new environment of clouds, no strong rules have been ratified for protecting privacy and data security [27]. Methods of protecting security of data in cloud are not efficiently successful and delaying time is another problem [27]. Based on arguments above, we postulate the following hypothesis:

H3: Users with more information sensitivity are more concerned about their privacy.

2. 4. 3. Privacy Concerns Study on self-disclosure for SNSs indicated the negative relationship between

privacy concerns and self-disclosure. The more secure privacy setting, the more reliable connections would be [28]. Based on previous research, the impact of privacy concerns on self-disclosure is supported and the research shows that the more users are concerned about privacy, the less information they are going to disclose on SNSs. Therefore, this study proposes the following hypotheses:

H4: Privacy concerns have a negative impact on self-disclosure.

3. TOOLS AND ALGORITHMS COLLECTIONS

3. 1. Measurement Instruments The items for all the constructs were collected from relevant literature, namely: privacy awareness reported in literature [29, 30]; Privacy policy also stated in literature [29, 31]; information sensitivity is given [29]; Privacy concerns – reported [29, 31-33]; and self-disclosure stated [31, 34, 35]. Based on literature, a questionnaire was developed and distributed online through Islamic azad university south Tehran branch information technology students, using google forms. A five-point quantitative scale was used to measure all the items, that 1 was 'strongly disagree', and 5 was 'strongly agree'.

4. ANALYZING HOW TO COLLECT INFORMATION

A valid and measured questionnaire was obtained. The online questionnaire was shared between IT (Information Technology) students of Islamic Azad University of South Tehran Branch. The number of feedbacks were 360 replies which 300 of them were valid. Based on information on Table 4 and Figure 3 following information was obtained. Respondents' age was between 25 to older than 45 years. Most respondents were between 35 to 40 years old and the rate of bachelor degree was higher than other degrees among respondents. From the collected information most respondents had incomes between 10 to 20 million Rials; rated as low social income level. 77.7% of respondent were males and 22.3% of respondents were female and also 48.7% of them were single and 51.3% of them were married. As it can be seen men had more cooperation in responding to online questionnaire than women. As it is obtained most respondents had bachelor's and master degree.

4. 1. Data Analyze Method In this study PLS (Partial Least Squares) method was used. Both measurement model and structural model were used to test the research model. PLS is a variance-based method and the software used for applying the technic was PLS Smart 2.0.

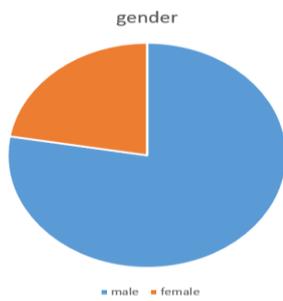


Figure 3. Gender distribution

TABLE 4. Demographic statistics

	Frequency	Percent
Degree		
Diploma	48	16.0
Associate	20	6.7
Bachelor	135	45.0
M. Sc.	89	29.7
Ph.D.	8	2.7
Total	300	100
Age		
25 to 30	9	3.0
30 to 35	65	21.7
35 to 40	113	37.7
40 to 45	81	27.0
Older than 45	32	10.7
Total	300	100
Marital status		
Single	146	48.7
Married	154	51.3
Total	300	100

4. 2. Measurement Model As it can be seen from Table 5 all loadings, composite reliabilities, Cronbach alpha and average variance extracted for each construct were measured. Internal consistency was measured by verifying whether all constructs are above 0.7, in Cronbach’s alpha (CA), based on each indicator inter-correlation (assuming that all are equally reliable); and on composite reliability (CR), based on the quantification of internal consistency and reliability of each construct (assuming that the indicators have different loadings). For indicator reliability, it is important that factor loadings are statically significant, and greater than 0.7. In order for latent variables to be able to explain more than half of the indicators, it is important that average variance extracted (AVE) should be above 0.5; thus guaranteeing convergent validity [36].

TABLE 5. Factor Loading, Composite Reliabilities, Cronbach Alpha and Average Variance Extracted (n=300)

Constructs	Loadings	CR	CA	AVE
PA				
PA1	0.756875			
PA2	0.892673	0.916340	0.886275	0.687775
PA3	0.869422			
PA4	0.747676			
PA5	0.868468			
PP				
PP1	0.816869			
PP2	0.866477	0.922252	0.894680	0.703598
PP3	0.814303			
PP5	0.850871			
PP6	0.844311			
IS				
IS1	0.730092			
IS2	0.818805	0.882981	0.834858	0.601817
IS3	0.755532			
IS4	0.783814			
IS5	0.787664			
PC				
PC1	0.854602			
PC2	0.792875	0.901633	0.863454	0.647381
PC3	0.803729			
PC4	0.758405			
PC6	0.810402			
SD				
SD1	0.861964			
SD2	0.816853	0.882309	0.822107	0.652630
SD4	0.797708			
SD5	0.750967			

Note: Privacy Awareness (PA), Privacy Policy (PP), Information Sensitivity (IS), Privacy Concerns (PC), Self-Disclosure (SD).

The other criterion was that the square root of AVE needs to be greater than the correlations between the constructs, which can be seen in Table 6. As it can be seen from the results the measurement model is verified and meaningful statistically. All constructs are useable to test the structural model.

4. 3. Structural Model A bootstrapping with 300 resamples was used to estimate the path significance levels, based on t-statistic values to test research model, and the results are presented in Table 7.

TABLE 6. Square root of AVE, and correlations between constructs

	PA	PP	IS	PC	SD
PA	0.8293				
PP	0.3378	0.8388			
IS	0.1143	0.3055	0.7757		
PC	0.3519	0.5855	0.4713	0.8045	
SD	0.1330	0.4226	0.3785	0.5026	0.8078

Note: Privacy Awareness (PA), Privacy Policy (PP), Information Sensitivity (IS), Privacy Concerns (PC), Self-Disclosure (SD).

TABLE 7. Hypothesis Conclusion

Hypothesis	Beta	T-Value	Conclusion
H1	0.169902	3.690223	99% Significant
H2	0.430264	6.731588	99% significant
H3	0.320447	5.749243	99% significant
H4	0.502624	9.298297	99% significant

The research model explains 46% of variation in privacy concern. The hypotheses of privacy awareness ($\beta=0.169$; $P<0.01$), privacy policy ($\beta=0.430$; $P<0.01$), and information sensitivity ($\beta=0.320$; $P<0.01$) are all statistically significant for explaining privacy concerns. Therefore, hypotheses H1, H2, and H3 are supported. 25% of variation in self-disclosure is explained in the research model. The hypothesis of privacy concerns ($\beta=0.502$; $P<0.01$) is statistically significant to explain self-disclosure. Therefore, the hypothesis H4 is supported for explaining self-disclosure.

4. 4. Limitations During the study the limitations found were mainly concerned with the constructs that we wanted to measure, because, however the amount of theoretical background related with privacy concerns and self-disclosure is extensive, experimental data combining these two subjects were not easy to come by. Furthermore, even though the sample size was 360, only IT students were included. The population and sample were drawn from only a single university of higher learning in a specific country. Thus, the findings could not be generalized to the entire university.

5. CONCLUSION AND FUTURE REASERCH

In conclusion, this study explains the privacy factors behind self-disclosure on social networking sites and applications. The model presented in our research was constructed based on previously developed models. This model attempted to present a clearer view on the issue of self-disclosure on social networking sites and

applications among IT students of Iran's Islamic Azad university of south Tehran branch and privacy variables that have a significant impact on self-disclosure.

Findings shows that nowadays privacy concerns on social networks are epidemic. Educating people to have knowledge about probable risk of information leakage and personal information disclosure could be recommended as an effective solution. Parents can play an efficient rule in their children perception of protecting privacy. Different societies have their own definition of privacy and its boundaries, to this regard, each person has her/his definition of privacy depend on her/his personality. Smartphones security is challenging and so many software and techniques have been analyzed.

The developed model could serve as a basis for other models to be developed and the model could also be extended to some other platform, not only social networking sites and applications.

6. REFERENCES

- Cheng, P.H. and Chen, L.W., "Peer learning efficacy analysis on undergraduate software design course", *Computer Applications in Engineering Education*, Vol. 26, No. 1, (2018), 5-16.
- Lwin, M.O., Miyazaki, A.D., Stanaland, A.J. and Lee, E., "Online usage motive and information disclosure for preteen children", *Young Consumers*, Vol. 13, No. 4, (2012), 345-356.
- Hamidi, H. and Jahanshahifard, M., "The role of the internet of things in the improvement and expansion of business", *Journal of Organizational and End User Computing (JOEUC)*, Vol. 30, No. 3, (2018), 24-44.
- Li, H., Yimin, Z., Mengshi, C., Xun, L., Xiulan, L. and Ying LIANG, H.T., "Development and validation of a disease severity scoring model for pediatric sepsis", *Iranian Journal of Public Health*, Vol. 45, No. 7, (2016), 875-884.
- Christofides, E., Muise, A. and Desmarais, S., "Hey mom, what's on your facebook? Comparing facebook disclosure and privacy in adolescents and adults", *Social Psychological and Personality Science*, Vol. 3, No. 1, (2012), 48-54.
- Shin, W. and Kang, H., "Adolescents' privacy concerns and information disclosure online: The role of parents and the internet", *Computers in Human Behavior*, Vol. 54, No., (2016), 114-123.
- Ajami, R., Al Qirim, N. and Ramadan, N., "Privacy issues in mobile social networks", *Procedia Computer Science*, Vol. 10, (2012), 672-679.
- A., P. and H., H., "An approach to managing and organizing text documents using intelligent text analysis", *Journal of Information Processing and Management*, Vol. 32, No. 4, (2017), 1171-1202.
- Stol, W.P., Kaspersen, H., Kerstens, J., Leukfeldt, E. and Lodder, A., "Governmental filtering of websites: The dutch case", *Computer Law & Security Review*, Vol. 25, No. 3, (2009), 251-262.
- Gerlach, J., Widjaja, T. and Buxmann, P., "Handle with care: How online social network providers' privacy policies impact users' information sharing behavior", *The Journal of Strategic Information Systems*, Vol. 24, No. 1, (2015), 33-43.

11. Johnson, R.D., Li, Y. and Dulebohn, J.H., "Unsuccessful performance and future computer self-efficacy estimations: Attributions and generalization to other software applications", *Journal of Organizational and End User Computing (JOEUC)*, Vol. 28, No. 1, (2016), 1-14.
12. Piao, C., Li, X. and Pan, X., "Research on the user privacy protection method in mobile commerce", in 2014 IEEE 11th International Conference on e-Business Engineering (ICEBE), IEEE., (2014), 177-184.
13. Thayanathan, V. and Albeshri, A., "Big data security issues based on quantum cryptography and privacy with authentication for mobile data center", *Procedia Computer Science*, Vol. 50, (2015), 149-156.
14. Abad, F., Saeed, A. and Hamidi, H., "An architecture for security and protection of big data", *International Journal of Engineering*, Vol. 30, No. 10, (2017), 1479-1486.
15. Najafloo, Y., Jedari, B., Xia, F., Yang, L.T. and Obaidat, M.S., "Safety challenges and solutions in mobile social networks", *Systems Journal*, Vol. 9, No. 3, (2013), 1-21.
16. Daraei, A. and Hamidi, H., "An efficient predictive model for myocardial infarction using cost-sensitive j48 model", *Iranian Journal of Public Health*, Vol. 46, No. 5, (2017), 682.
17. Liu, Y., Tan, C.-H. and Sutanto, J., "Selective attention to commercial information displays in globally available mobile application", *Journal of Global Information Management (JGIM)*, Vol. 24, No. 2, (2016), 18-38.
18. Sun, Y., Wang, N., Shen, X.-L. and Zhang, J.X., "Location information disclosure in location-based social network services: Privacy calculus, benefit structure, and gender differences", *Computers in Human Behavior*, Vol. 52, (2015), 278-292.
19. Wu, J., Ding, F., Xu, M., Mo, Z. and Jin, A., "Investigating the determinants of decision-making on adoption of public cloud computing in e-government", *Journal of Global Information Management (JGIM)*, Vol. 24, No. 3, (2016), 71-89.
20. Rowan, M. and Dehlinger, J., "A privacy policy comparison of health and fitness related mobile applications", *Procedia Computer Science*, Vol. 37, (2014), 348-355.
21. Hamidi, H., Vafaei, A. and Monadjemi, S.A., "A framework for abft techniques in the design of fault-tolerant computing systems", *EURASIP Journal on Advances in Signal Processing*, Vol. 2011, No. 1, (2011), DOI: 10.1186/1687-6180-2011-90.
22. Zlatolas, L.N., Welzer, T., Heričko, M. and Hölbl, M., "Privacy antecedents for sns self-disclosure: The case of facebook", *Computers in Human Behavior*, Vol. 45, (2015), 158-167.
23. Bimonte, S., Sautot, L., Journaux, L. and Faivre, B., "Multidimensional model design using data mining: A rapid prototyping methodology", *International Journal of Data Warehousing and Mining (IJDWM)*, Vol. 13, No. 1, (2017), 1-35.
24. Hamidi, H. and Mousavi, R., "Analysis and evaluation of a framework for sampling database in recommenders", *Journal of Global Information Management (JGIM)*, Vol. 26, No. 1, (2018), 41-57.
25. Esposito, C. and Ficco, M., "Recent developments on security and reliability in large-scale data processing with mapreduce", *International Journal of Data Warehousing and Mining (IJDWM)*, Vol. 12, No. 1, (2016), 49-68.
26. Hamidi, H. and Chavoshi, A., "Analysis of the essential factors for the adoption of mobile learning in higher education: A case study of students of the university of technology", *Telematics and Informatics*, Vol. 35, No. 4, (2018), 1053-1070.
27. Hamidi, H., Vafaei, A. and Monadjemi, A., "Algorithm based fault tolerant and check pointing for high performance computing systems", *Journal of Applied Science*, Vol. 9, No. 22, (2009), 3947-3956.
28. Trepte, S. and Reinecke, L., The social web as a shelter for privacy and authentic living, in Privacy online. 2011, Springer.61-73.
29. Hamidi, H. and Moradi, S., "Analysis of consideration of security parameters by vendors on trust and customer satisfaction in e-commerce", *Journal of Global Information Management (JGIM)*, Vol. 25, No. 4, (2017), 32-45.
30. Malhotra, N.K., Kim, S.S. and Agarwal, J., "Internet users' information privacy concerns (iupc): The construct, the scale, and a causal model", *Information systems research*, Vol. 15, No. 4, (2004), 336-355.
31. Karimzadeh-Farshbafan, M. and Ashtiani, F., "Semi-myopic algorithm for resource allocation in wireless body area networks", *IET Wireless Sensor Systems*, Vol. 8, No. 1, (2017), 26-35.
32. Al Shaheen, H. and Takruri-Rizk, H., "Improving the energy efficiency for the wbsn bottleneck zone based on random linear network coding", *IET Wireless Sensor Systems*, Vol. 8, No. 1, (2017), 17–25, DOI: 10.1049/iet-wss.2017.0056.
33. Hamidi, H. and Hashemzadeh, E., "An approach to improve generation of association rules in order to be used in recommenders", *International Journal of Data Warehousing and Mining (IJDWM)*, Vol. 13, No. 4, (2017), 1-18.
34. Chen, R., "Living a private life in public social networks: An exploration of member self-disclosure", *Decision Support Systems*, Vol. 55, No. 3, (2013), 661-668.
35. Echeverría, L., Cobos, R., Machuca, L. and Claros, I., "Using collaborative learning scenarios to teach programming to non-cs majors", *Computer Applications in Engineering Education*, Vol. 25, No. 5, (2017), 719-731.
36. Hamidi, H. and Daraei, A., "Analysis of pre-processing and post-processing methods and using data mining to diagnose heart diseases", *International Journal of Engineering-Transactions A: Basics*, Vol. 29, No. 7, (2016), 921-930.

Analyzing Tools and Algorithms for Privacy Protection and Data Security in Social Networks

A. Mohammadi^a, H. Hamidi^b

^a Department of Information Technology Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran

^b Department of Industrial Engineering, Information Technology Group, K. N. Toosi University of Technology, Tehran, Iran

PAPER INFO

چکیده

Paper history:

Received 07 January 2018

Received in revised form 27 January 2018

Accepted 08 February 2018

Keywords:

Privacy Protection

Social Networks

Information Leakage

Information Disclosure

هدف این تحقیق مطالعه عوامل مؤثر بر نگرانی‌های حریم خصوصی در مورد امنیت داده‌ها و حفاظت از آن‌ها در سایت‌های شبکه‌های اجتماعی و تأثیر آن بر خودافشایی است. صد مقاله درباره‌ی حفاظت حریم خصوصی، امنیت داده‌ها، افشای اطلاعات و درز آن در شبکه‌های اجتماعی مورد مطالعه قرار گرفت. انواع مدل‌ها و الگوریتم‌ها و تکرار آن‌ها در مقالات مستخرج بررسی شد. در نتیجه، یک مدل تحقیقاتی برای ارزیابی نگرانی‌های حریم خصوصی و تأثیر آن بر خودافشایی ساخته شد. نیاز به داشتن دانش و مهارت در مورد حفظ حریم خصوصی با توجه به توسعه روز افزون شبکه‌های اجتماعی بسیار ضروری به نظر می‌رسد. بیشتر مطالعات در زمینه حریم خصوصی کاربران شبکه‌های اجتماعی خصوصاً در تلفن‌های هوشمند و سلامت الکترونیک میان زنان، مردان و کودکان است. بیشتر این مطالعات در آمریکا انجام شده است. این مطالعه بر دانشجویان فناوری اطلاعات دانشگاه آزاد اسلامی واحد تهران جنوب به عنوان نمونه تمرکز دارد. مطالعات بر حفظ حریم خصوصی و امنیت در شبکه‌های اجتماعی تمرکز دارند.

doi: 10.5829/ije.2018.31.08b.15
