https://doi.org/10.48009/2_iis_2013_217-224

Issues in Information Systems

Volume 14, Issue 2, pp.217-224, 2013

AN EXAMINATION OF MOBILE APP PRIVACY POLICIES AND THIRD-PARTY DATA SHARING

Matthew North, The College of Idaho, mnorth@collegeofidaho.edu

ABSTRACT

Mobile applications and social networking are an increasingly important, even central part of modern business strategy. Consumers have flocked to the smart phone and tablet market, driven there largely by the desire to be able to download and use a variety of mobile apps. These apps provide entertainment, utilities and services, and importantly to many, social networking. But at what cost do these apps come? Many are free to acquire, but what about costs in terms of personal privacy? This paper examines the top 50 free and top 50 paid apps downloaded from the Apple App Store in the context of their collection and sharing of user data. The study examines the terms of service documents associated with the targeted apps, and finds that the majority of apps in general, and specifically social networking apps, do collect user data and reserve the right to share that data with third parties. Based upon the results of the study, more research in the area of mobile app and social networking privacy is strongly recommended.

Keywords: Mobile Apps, Social Networking, Information Privacy and Text Mining.

INTRODUCTION

In the March, 2013 edition of *Communications of the ACM*, authors Nafaa Jebeur, Sherali Zeadally and Biju Sayed [10] make a bold, perhaps even audacious claim about mobile social networking apps: "They deliver the right social service to the right user anytime, anyplace, without divulging personal data." The authors go on to emphasize that mobile apps, particularly those that utilize or connect with social networking services, have become a mainstream strategic priority for organizations in developed and emerging economies around the globe, resulting in "intense competition among providers." [10] Ample evidence exists to support this claim as true [3, 11, 13], however, it then raises a question: Amid this intense competition, can businesses really employ mobile apps, especially in the area of social networking, in order to gain strategic advantage *without divulging personal data*, and if so (and perhaps more importantly), do they? This research will attempt to address these two questions.

Contextual Background Information

Certainly competition in the social networking space has intensified in the mobile-era, as an increasing number of companies look to these systems to connect with consumers, and to connect consumers with one another in an effort to build momentum, market share, product buzz, etc. [4, 6]. Numerous successful examples can be used to illustrate their central role in competitive business strategy. *Mashable* [12] recently highlighted five high profile companies that have used mobile apps and social networking technologies on handheld and tablet devices in innovative ways:

- Renault has provided auto show attendees with mobile technology allowing them to check in on Facebook and like certain models and displays, simultaneously allowing Renault to gather real-time reactions to their auto show presences while spreading their product message to a much broader audience that is *not even at the auto show.*

- Fashion brand *Diesel* has created QR codes that shoppers can scan in clothing and department stores that will post interests and purchases to shoppers' Facebook pages, allowing those shoppers to share their fashion choices with friends while effectively serving as a highly personal but low-cost marketing stream to other potential consumers, ostensibly with similar fashion preferences, for *Diesel*.

- Macy's department stores have begun using QR code apps as well. In their implementation, shoppers can scan QR codes that hook into YouTube accounts, enabling users to watch videos of the scanned product in

action, to post comments about the product, and to provide Macy's with input and ideas for accessories, complements or promotions surrounding the scanned product.

- Starbucks coffee has produced an app enabling 'regulars' to pay by mobile. Starbucks promoted their app as a way to "pay faster", something most morning coffee buyers would appreciate when trying to get to work on time, but while touting that benefit, Starbucks quietly enjoyed a different, and likely more valuable outcome from their app: pay by mobile enabled Starbucks to amass a tremendously large data set about their consumers' behaviors including purchase frequencies, product preferences, purchase timing, etc. Thousands of companies have since followed Starbucks into the mobile payment market.

- Coldwell Banker's realty division has embraced social mobile apps to market residential real estate. Using a mobile device's camera and cellular data network, home buyers can see virtual tours of homes on the market on YouTube, review home specifications on dedicated Facebook pages, gather information on schools, community resources and neighborhood profiles, and even follow Twitter feeds about the current status of the property for sale.

In the face of these and countless other examples, we can concede that mobile apps and hooks into associated social networks are now considered part of modern business and will undoubtedly become even more deeply entrenched as an integral part of 21st century marketing and strategic operations [6]. Jabeur and colleagues agree: "Even more services are expected soon, along with numerous challenges and questions about privacy and data security." With this established, we return to the research question previously stated: Are companies actually employing social networking strategies on the mobile platform "without divulging personal data"? [10]

RELEVANT LITERATURE

The modern mobile app environment as we know it today arguably began with the launch of the Apple iPhone in the year 2007 and has grown steadily and rapidly ever since [6]. The introduction of the Android operating system the next year solidified and accelerated the adoption of mobile apps, introducing them to a wider audience and enabling a more open development environment for business and individuals to experiment with apps for a variety of purposes [8]. Already established enterprises found that they could expand their user bases by mobile-enabling their offerings, often with very little incremental cost. Given that the current mobile environment is only about seven years old, it ranks among the most revolutionary of technological innovations in human history [2, 9].

But at what cost has this revolutionary innovation come? Internet and privacy advocates were quick to react to mobile adoption, at least in terms of voicing concern, if not in subjecting to the technology to immediate and intense scrutiny [2]. Certainly, efforts have been made in those areas, but without a centralized and empowered body governing the right and wrong on mobile data privacy, the movement has had to adopt a more grass-roots advocacy approach, rather than one of enforcement [5, 9].

The question of data privacy however cannot solely be addressed at the level of those who are already concerned with privacy in general. These individuals are already conscientious and informed about risks and mitigation strategies. The best-laid protections against privacy abuses can be circumvented unwittingly by an uninformed, perhaps blissfully ignorant consumer base [2, 6, 9]. Users of any data collecting technology who either don't know, or don't care, that usage and/or personally identifiable information is being collected can undermine even the best privacy protections. Unfortunately, due to the highly socialized nature of mobile app adoption, most apps are downloaded without users even reading terms of service agreements or otherwise informing themselves of the potential consequences of adopting an app. Apps are often adopted by users as a result of a friend or other relation recommending the app [9, 11, 13]. Many app adopters admit to having adopted apps without having read the provider's terms of service agreement. With such laissez-faire adoption practices as the norm, businesses and organizations, particularly if they happen to be unethical in any measure, could easily abuse data privacy to the detriment of consumers [9], and might even justify doing so in light of their consumers' seeming indifference.

Unfortunately in modern society, consumers have become accustomed to, and even demanding of being protected from their own ignorance or poor choice. Ample product liability lawsuits illustrate the expectation by many

consumers that manufacturers and service providers must bear the majority of the burden in protecting people from loss or harm resulting from adoption of goods and services. In 2003, in an effort provide such protections in the Internet age, the State of California passed California Commercial Code § 22575-22579: The California Online Privacy Protection Act [1]. While the act was originally passed before the current mobile app age, it has been successfully and recently applied to app privacy cases [7]. In 2012, the state successfully sued Delta Airlines under this statute because the company collected personally identifiable information via its mobile app, but did not provide app users with a terms of service agreement within the app. Delta settled the case with the state and quickly updated their app to include their terms of service, however legislation and enforcement of this type can only go so far. Many app makers are outside the jurisdiction of the state of California, and indeed even the United States, and many countries neither have nor want such legislation [6, 8]. Further, so long as consumers continue to download, install and use apps without paying close attention to terms of service, no amount of legislation would prevent data from being used in ways that could harm privacy [9, 13]. Considering the few examples given in the introduction to this paper, it is clear that usage and personally identifiable information are being shared from mobile apps to social networking sites such as Facebook, YouTube and Twitter [11]. All three of these companies do operate in the state of California, and as such, are subject to the aforementioned statute. An inspection of all three of these companies' terms of service agreements reveals that they do collect and store user data, and they do share it with third-party partners. Thus we undertake in this work to determine whether or not organizations using social networking in the mobile platform really are doing so without sharing private data as Jebeur, et al. suggest [10].

RESEARCH METHODOLOGY

As stated, our research questions are two-fold:

Research Question 1: Do organizations that provide mobile apps, and particularly those related to social networking, do so without requiring users to divulge personal information?

Research Question 2: Do organizations that provide mobile apps, and particularly those related to social networking, share personal information with third parties?

To examine these research questions we selected the top 50 free and top 50 paid apps for Apple's iOS as identified by appadvice.com. Appadvice.com tracks top apps by download rate through the Apple App Store. Each app was then reviewed and entered into a database, recording the following eight attributes:

- App Name
- Publisher
- Date Released
- Price
- Category
- Terms of Service
- User Data
- Third Party Sharing

An app's Category attribute is entered as defined by the publisher at the time that the app was made available through the App Store. This attribute is of particular importance for this study because it enables us to examine apps specifically associated with social networking services. The App Name, Publisher, Date Released and Price attributes are also all defined as listed in the App Store. The Terms of Service for each of the apps were collected manually and recorded verbatim, in their entirety in the study's database. This component is critical for this research, as we have used a text mining approach to determine the extent to which app publishers acknowledge that they collect and/or share personal data. Somewhat surprisingly, for several apps, no terms of service, end user agreement or other form of privacy policy could be found. This will be discussed further in the results section of this paper. The User Data and Third Party Sharing attributes are Boolean values in the research database. Each terms of service document was reviewed, and if the terms of service specifically contained a section addressing user data collection or third party data sharing, these attributes were flagged as 'yes' where relevant.

Procedure

The data comprised 100 observations; one for each app on the top 50 free and top 50 paid lists. The data were reviewed for accuracy and completeness. In six instances, apps were found to contain missing values in at least one attribute, and these were corrected. The data were then imported in the RapidMiner software environment, where both descriptive statistics and text mining models could be utilized.

Descriptive statistics revealed much about our two research questions. Drawing upon the Category attribute, we have used histograms, along with Pearson correlations to examine relationships between user data, third-party sharing, and other app attributes.

The study also undertakes to text mine the terms of service documents that were compiled in the study database. A text mining model was constructed to identify frequent words or word patterns which might reveal typical patterns on the part of app publishers to gather and share user data. The model also seeks for words or phrases which might indicate specific behaviors by app publishers that could impact user privacy. The model is constructed first by tokenizing the terms of service documents. It then eliminates stopwords such as 'a', 'the', 'it', etc., reduces words to their stems (e.g. 'Private', 'Privacy', and 'Privately' are all part of the 'Priva_' word stem), and then seeks to identify n-grams (i.e. phrases or strings of terms used in conjunction with one another.). The output of this text mining model is presented and explained. Where necessary, post hoc evaluation of the output was conducted to more specifically understand the meaning of the terms generated by the model.

RESULTS

Of the 100 apps examined (top 50 free and top 50 paid for iOS), the vast majority were games. Figure 1 depicts the distribution of app categories in this study.

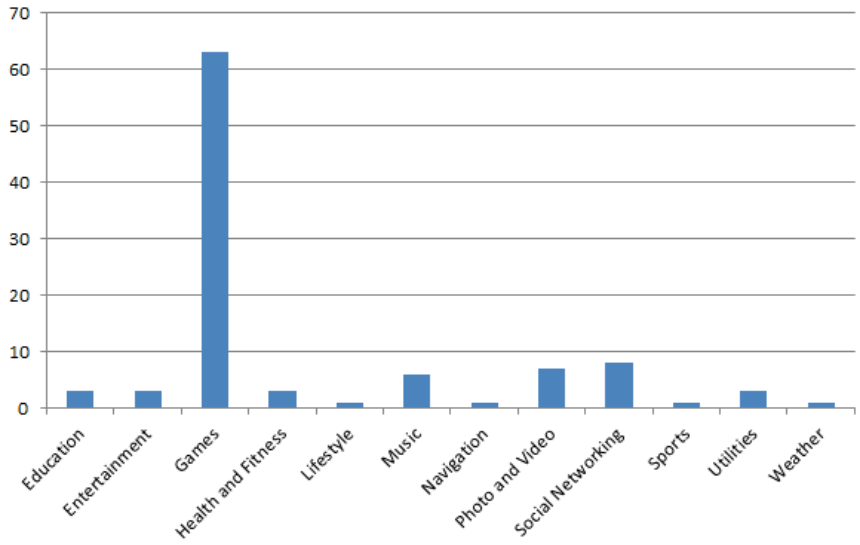

Figure 1. App distribution by Category.

While games clearly enjoy a sizeable majority (63 apps), it is notable in the context of this study that the second most prevalent category is social networking (8 apps). We examine here the terms of service for all apps combined, and also break out those associated with social networking specifically. Interestingly, only 37% of game apps in this study are distributed free of charge, while 88% of social networking related apps are free. The cost of developing an app must be borne in some way, and social networking apps certainly have the capacity to generate large amounts of valuable data. It is possible that these apps will have a higher proportion of user data and third-party sharing concessions in their terms of service than will games? Figure 2 represents the portion of paid vs. free apps for each category.

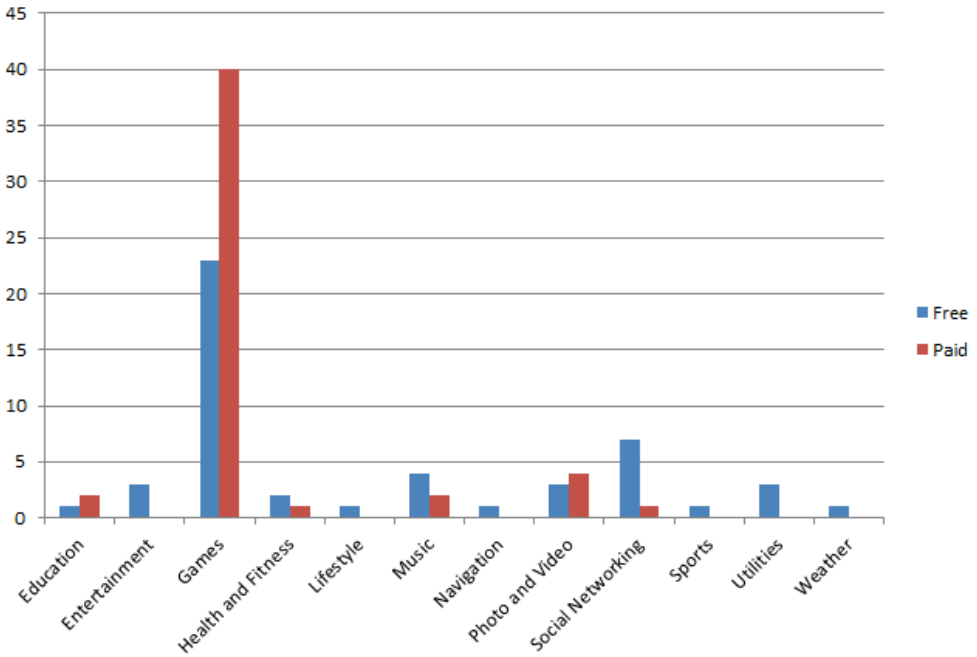

Figure 2. Proportion of free and paid apps by category.

In examining the app data statistically, we find evidence that the ability to collect user data and potentially share it with third-parties may be an important motivating factor for app providers. Using a Pearson correlation ($\alpha = .05$), we find a strong, statistically significant correlation between the User Data and Third Party attributes for social networking apps with a coefficient (p) of .78. By comparison, the Pearson coefficient between User Data and Third Party for game apps is still significant, but lower and slightly weaker at .68. Because there was only one paid social networking app in our data set, we cannot test for correlations between these two attributes for paid vs. unpaid apps. Our sample does enable us to examine the User Data and Third Party attributes for paid vs. unpaid games, and yields what may be a very telling result: the p value between these two attributes for paid game apps is only .54, while it rises all the way to .91 for free games. While this correlation does not prove that app providers are sharing user data with third parties in order to fund the cost of their app development and distribution, it is compelling evidence that such practices may be occurring, and it is a reasonable conclusion at which to arrive. This then represents an opportunity for future research. With a larger data set containing both more paid and free social networking apps, we could examine whether or not the correlation identified on game apps holds true for social networking apps as well. Further investigation into causation would also be warranted.

With this evidence in mind, and unable to pursue additional correlational investigation with the present data set, we turn to text mining to further examine the relationship of user data collection and third party sharing. For 27 of the 100 apps examined in this study, no terms of service, user agreement, or privacy policy could be located. While this is in violation of the aforementioned California statute [1, 7, 8], and though Apple's App Store operates within that jurisdiction, it appears that Apple has not taken upon itself the task of enforcing that law within its App Store publishing policies. Six of these 27 are published by individuals unassociated with a corporation or employer. Nineteen of the 27 are from developers outside United States jurisdiction. Interestingly, 15 of the 27 are paid apps, yet despite charging a fee, they do not provide terms of service. With these observations in mind, our text mining model was applied to the 83 terms of service documents that were located.

Collectively, the terms of service agreements for the 83 apps examined comprised some 516 pages of information containing more than 1.5 million characters in nearly a quarter of a million words. 259 of these pages were for

terms of service for free apps, containing 123,088 words; while 257 pages were required for paid app terms of service, representing 127,558 words. Figure 3 represents the text mining model constructed to mine these terms of service documents in RapidMiner. It contains all operators for tokenization of the documents' words, along with stop word removal, word stemming and n-gram (phrase) generation.

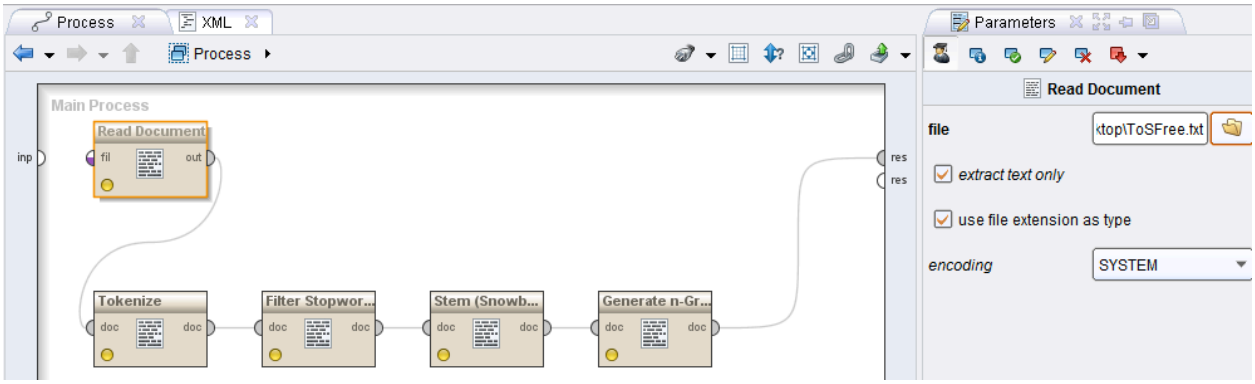

Figure 3. Terms of Service Text Mining model in RapidMiner.

This model was run against the terms of service documents for free and paid apps separately. It extracted the high frequency words and terms from each document, allowing us to review and compare them, seeking possible indicators of policies affecting the gathering and sharing of user data. Table 1 depicts the primary words and phrases for free and paid apps in this study. These are listed in order of frequency in the respective documents.

Table 1. Frequent terms mined from free and paid app terms of service documents.

Free App Frequent Terms		Paid App Frequent Terms	
Term	**Average Occurrences**	**Term**	**Average Occurrences**
Satisfaction	77	Inform	28
Account	35	User	22
Inform	19	License	19
Law	15	Person	16
Access	15	Right	16
Copyright	13	Law	16
Right	12	Third Party	13
Infringe	10	Privacy	12
Person	9	Access	11
Third Party	8	Copyright	10

While exercising caution not to read too much into these mined terms, we can see some interesting patterns emerge. Between free and paid terms of service agreements, there are some shared common words. Words such as 'Inform', 'Law' and 'Person' are found on both lists, though with differing frequencies. The different frequencies are consistent across terms, as the paid apps have a larger number of word tokens than do the free apps. While the terms shown in Table 1 are the most frequent in each of these documents, there were 5,722 unique terms identified in the free apps' terms of service, and 5,469 unique terms on the paid side. Thus, the paid app terms of service have a wider variety of topics included, and broader range of words included in them.

Perhaps most interesting in the context of this paper is the word 'Privacy'. While it appears an average of 12 times per document in the terms of service agreements for paid apps and is the 8th most popular term in Table 1, this term does not appear in Table 1 for free apps at all. In fact, through post hoc investigation, we found that 'Privacy' appears an average of only two times per document in the terms of service agreements for free apps. This would indicate that when a user is paying to use an app, privacy is not only more thoroughly discussed, but also enjoys a

more prominent place in the agreement with the end user. It would appear that when paying a provider for an app, the consumer's privacy is more carefully regarded.

Both free and paid apps have 'Third Party' frequently in their terms of service, however again, post hoc analysis of the documents reveals that this term takes on different meanings in the context of the two kinds of apps. Within free app terms of service, the term 'Third Party' and its variants almost always refers to the publisher reserving the right to share user data *with* third parties; while in the paid terms of service agreements, the term most usually refers to the fact that the publisher *may* work with third parties, and will share user data only by consent. It should also be noted that the text mining of these documents revealed through data what has long been assumed in practice: terms of service documents exist primarily to protect the publishers of apps, not the users of them.

Now consider the most frequent terms associated with only those apps included in this study in the Social Networking category. There were eight such apps in the study's database. These are depicted in Table 2.

Table 2. Frequent terms mined from social networking terms of service documents.

Social Networking App Frequent Terms	
Term	**Average Occurrences**
User	24
Inform	18
Third Party	18
Access	14
Right	12
Agreement	12
Account	12
Limit	11
Response	11
Website	10

Keeping in mind that all but one of the social networking apps is free, we would expect the frequent terms to be similar to the free app list in Table 1. Indeed several of the terms are the same, and yet some strong terms such as 'Agreement' and 'Satisfaction' do not coincide. In the context of our research questions, the term 'Third Party' does coincide, and is much more prevalent with the social networking apps than it is with the free apps generally. Based on the data, there is ample evidence that social networking apps do share user data with third parties. The only exception found in post hoc analysis of the terms of service agreements was the Facebook Messenger app, which does not claim to share user data with third parties. This app directly plugs into Facebook however, which does share user data. So at least indirectly, this app also shares user data with third parties.

CONCLUSIONS

The research questions posed in this study were based on Jebeur, et al.'s claim that mobile users could use social networking apps *without divulging personal information* [10]. Based on the information found in this study, this statement cannot be defended. While this study is admittedly exploratory in nature, the data examined here are relatively consistent in sending the message that when mobile apps collect user data, the publishers of those apps, at the very least, are reserving the right to share that data with third parties. Of the apps for which we were able to locate a terms of service agreement, fully 80% of them collect user data, and 71% of them are engaged in third party data sharing. Strong correlational evidence exists to support the relationship between user data gathering and third party data sharing, especially with social networking apps.

As integration of mobile apps and social networking into business strategy continues and even accelerates, the trends identified in this paper will likely continue. Advanced mobile devices such as Google glass, which move beyond increasingly powerful smart phones and tablets now prevalent in most developed economic markets will likely enable even more detailed and extensive data gathering [13]. Many such devices are even GPS enabled, allowing user location tracking to complement behavioral data capture. In an age of Big Data, data is valuable, and it can be

amassed quickly and effectively through the creation and dissemination of mobile apps. Businesses employing this tactic are clearly already protecting themselves legally as they collect and share this data. Much additional research is warranted in this area, in order to both increase the body of evidence regarding mobile app and social networking data gathering and sharing, and to shape information-driven decisions about the apps we adopt and use in our daily lives.

REFERENCES

1. California Online Privacy Protection Act. (2003). *State of California*. [Electronic Version]. Retrieved on 27 March 2013 from: http://oag.ca.gov/privacy/COPPA
2. Camenisch, J. (2012). Information Privacy?! *Computer Networks, 56*(18), 3834-3848.
3. Christin, D., Sánchez López, P., Reinhardt, A., Hollick, M. & Kauer, M. (2013). Share with Strangers: Privacy bubbles as user-centered privacy control for mobile content sharing applications. *Information Security Technical Report, 17*(3), 105-116.
4. Gaggioli, A. (2012). CyberSightings. CyberPsychology, Behavior & Social Networking, 15(9), 512-513.
5. Hasan, O., Brunie, L., & Bertino, E. (2012). Preserving Privacy of Feedback Providers in Decentralized Reputation Systems. *Computers & Security, 31*(7), 816-826.
6. Holbrook, E. (2011). Mobile Apps and Hidden Risks. Risk Management, 58(6), 6-8.
7. Kirk, J. (2012). California Sues Delta Airlines over App Privacy Policy. *PCWorld*. [Electronic Version]. Retrieved on 26 March 2013 from: http://www.pcworld.com/article/2018966/california-sues-delta-airlines-over-app-privacy-policy.html
8. Krasnow, M. J. (2013). Mobile Application and Website Privacy Policies - It's Not Just About California. *Financial Executive, 29*(2), 65-66.
9. Mylonas, A., Kastania, A. & Gritzalis, D. (2013). Delegate the Smartphone User? Security awareness in smartphone platforms. *Computers & Security, 34(1)*, 47-66.
10. Nafaâ, J., Sherali, Z. & Biju, S. (2013). Mobile Social Networking Applications. *Communications of the ACM, 56*(3), 71-79.
11. Shehab, M., Squicciarini, A., Ahn, G. & Kokkinou, I. (2012). Access Control for Online Social Networks' Third Party Applications. *Computers & Security, 31*(8), 897-911.
12. Wasserman, T. (2011). 5 Innovative Mobile Marketing Campaigns to Learn From. *Mashable*. [Electronic Version]. Retrieved on 27 March 2013 from: http://mashable.com/2011/06/21/innovative-mobile-marketing-campaigns
13. Wicker, S. B. (2012). The Loss of Location Privacy in the Cellular Age. *Communications of the ACM, 55*(8), 60-68.