Big Data Privacy Issues in Public Social Media

Matthew Smith, Christian Szongott
Distributed Computing & Security Group
Leibniz Universität Hannover
Hannover, Germany
Email: {smith,szongott}@dcsec.uni-hannover.de

Benjamin Henne, Gabriele von Voigt
L3S Research Center
Hannover, Germany
Email: {henne, vonvoigt}@l3s.de

***Abstract*—Big Data is a new label given to a diverse field of data intensive informatics in which the datasets are so large that they become hard to work with effectively. The term has been mainly used in two contexts, firstly as a technological challenge when dealing with data-intensive domains such as high energy physics, astronomy or internet search, and secondly as a sociological problem when data about us is collected and mined by companies such as Facebook, Google, mobile phone companies, retail chains and governments. In this paper we look at this second issue from a new perspective, namely how can the user gain awareness of the personally relevant part Big Data that is publicly available in the social web. The amount of user-generated media uploaded to the web is expanding rapidly and it is beyond the capabilities of any human to sift through it all to see which media impacts our privacy. Based on an analysis of social media in Flickr, Locr, Facebook and Google+, we discuss privacy implications and potential of the emerging trend of geo-tagged social media. We then present a concept with which users can stay informed about which parts of the social Big Data deluge is relevant to them.**

I. Introduction

Big Data is becoming a hot topic in many areas where datasets are so large that they can no longer be handled effectively or even completely [15]. Or put differently, any task which is comparatively easy to execute when operating on a small but relevant set of data, but becomes unmanageable when dealing with the same problem with a large dataset can be classified as a Big Data problem. Typical problems encountered when dealing with Big Data include capture, storage, dissemination, search, analytics and visualisation. The traditional data-intensive sciences such as astronomy, high energy physics, meteorology, genomics, biological and environmental research in which peta- and exabytes of data are generated are common domain examples. Here even the capture and storage of the data is a challenge. But there are also new domains encroaching on the Big Data paradigm: data warehousing, Internet and social web search, finance and business informatics. Here datasets can be small compared to the previous domains, however the complexity of the data can still lead to the classification as a Big Data problem.

When looking at privacy issues in the Big Data domain we need to distinguish which of the many Big Data application domains we are discussing. The traditional Big Data applications such as astronomy and other e-sciences usually operate on non-personal information and as such usually do not have significant privacy issues. The privacy critical Big Data applications lie in the new domains of the social web, consumer and business analytics and governmental surveillance [6]. In these domains Big Data research is being used to create and analyse profiles of us, for example for market research, targeted advertisement, workflow improvement or national security. These are very contentious issues since it is entirely up to the controller of the Big Data sets if the information gleaned is used for nefarious purposes or not.

In particular in the context of the social web there is an increasing awareness of the value, potential and risk of the personal data which we voluntarily upload to the web. Where privacy is concerned there has been a lot of work in the small data area, i.e. how can users control who has access to what they post themselves. However, the Big Data issue in this area has focused almost entirely on what the controlling companies do with this information. These concerns are being addressed by calls for regulatory intervention, i.e. regulating what companies are allowed to do with the data we give them or what data they are allowed to gather about us. A topic which has not received as much attention is the effect other peoples' data has on us. This can be seen both in a social context, i.e. what happens if friends or acquaintances see this data and also what happens when companies with Big Data analytics harvest this information. Microsoft's Scott Charney offered a very good example during his Keynote speech at the RSA Conference 2012: *If a friend takes a picture of me during a volleyball game, shares this picture with other friends and one of them uploads the picture to the web, my insurance company can find and use that picture against me.*[1] There have been reports that insurance companies are looking for just such information which could raise premiums or even deny claims.[2] The same is true for banks and credit rating companies.[3]

In this paper we examine this side of the social media Big Data issue. We discuss how the growing proliferation and capabilities of mobile devices is creating a deluge of social media which can effect our privacy. Due to the vast amounts of data being uploaded every day it is next to impossible to be aware of everything which effects us. We also discuss a concept which can be used to regain control of some of the Big Data deluge created by other social web users.

[1] Paraphrased from the Keynote at RSA 2012
[2] http://abclocal.go.com/kabc/story?section=news/consumer&id=8422388
[3] http://www.betabeat.com/2011/12/13/as-banks-start-nosing-around-facebook-and-twitter-the-wrong-friends-might-just-sink-your-credit/

978-1-4673-1703-0/12/$31.00 ©2013 IEEE

II. Environment & Problem Statement

The amount of social media being uploaded into the web is growing rapidly and there is still no end to this trend in sight. The ease-of-use of modern smartphones and the proliferation of high-speed mobile networks is facilitating a culture of spontaneous and carefree uploading of user-generated content. To give an idea of the scale of this phenomenon: Just in the last two years the number of photos uploaded to Facebook per month has risen from 2 billion to over 6 billion [9], [10]. From a personal perspective an overwhelming majority of these photos have no privacy relevance for oneself. Finding the few that are relevant is a daunting task.

While one's own media is uploaded consciously, the flood of media uploaded by others is so huge that it is almost impossible to stay aware of all media in which one might be depicted. This can be classified as a Big Data problem on the users side, however not on the provider side. Current social networks and photo-sharing sites mainly focus on the privacy of users' own media in terms of access control, but do little to deal with the privacy implications created by other users' media. There are ever more complex settings allowing a user to decide who is allowed to see what content but only content owned by the user [5], [7]. However, the issue of staying on top of what others are uploading (mostly in good faith), that might also be relevant to the user, is still very much outside the control of that user. Social networks which allow tagging of users usually inform affected users when they are tagged. However, if no such tagging is done by the uploader or a third party, there are currently no mechanisms to inform users of relevant media.

A second important emerging trend is the capability of many modern devices to embed geo-data and other metadata into the created content. While the privacy issues of location-based services such as Foursquare or Qype have been discussed at great length, the privacy issues of location information embedded into uploaded media have not yet received much attention. There is one very significant difference between these two categories. In the first, users reveal their current location to access online services, such as Google Maps, Yelp or Qype or the user actually publishes his location on a social network site like Foursquare, Google Latitude or Facebook Places. In this category, the user mainly affects his own privacy. There is a large body of work examining privacy preserving techniques to protect a user's own privacy, ranging from solutions which are installed locally on the user's mobile device [2], to solutions which use online services relying on group-based anonymisation algorithms, as for instance mix zones [3] or k-anonymity [13].

The second category is created by media which contains location information. This can have all the same privacy implications for the creator of the media, however, a critical and hitherto often overlooked issue is the fact, that the location and other metadata contained in pictures and videos can also affect other people than the uploader himself. This is a critical oversight and an issue which will gain importance as the mobile smart device boom continues.

III. Threat Analysis

We categorise privacy issues into two classes. Firstly, home-grown problems: Someone uploads a piece of compromising media of himself with insufficient protection or forethought which causes damage to his own privacy. A prime example of this category is someone uploading compromising pictures of himself into a public album instead of a private one or onto his Timeline instead of a message. The damage done in these cases is very obvious since the link between the content and the user is direct and the audience (often the peer circle) has direct interest in the content. One special facet of this problem is that what is considered damaging content by the user can and often does change over time. While this is a serious problem, especially amongst the Facebook generation, this issue is a small data problem and thus is not the focus of this work.

Secondly we have the Big Data problems created by others: An emerging threat to users' online privacy comes from other users' media. What makes this threat particularly bad is the fact that the person harmed is not involved in the uploading process and thus cannot take any pre-emptive precautions and the amount of data being uploaded is so vast it cannot be manually sighted. Also there are currently no countermeasures, except post-priory legal ones, to prevent others from uploading potentially damaging content about someone. There are two requirements for this form of privacy threat to have an effect: Firstly, to cause harm to a person a piece of media needs to be able to be associated/linked to the person in some way. This link can either be non-technical, such as being recognisable in a photo, or technical such as a profile being (hyper-)linked to a photo. There is also the grey area of textual references to a person near to the photo or embedded in the metadata of the photo. This metadata does not directly create a technical link to a profile, but it opens the possibility for search engines to index the information and make it searchable, thus creating a technical link. Secondly, a piece of media in question must contain harmful content for the person linked to it. This can again be non-technical such as being depicted in a compromising way. However, more interestingly it can also be technical. In these cases metadata or associated data causes harm. For instance time and location data can indicate that a person has been at an embarrassing location, took part in a political event, or was not where he said he was.

Since the uploading of this type of damaging media cannot be effectively prevented, awareness is the key issue in combating this emerging privacy problem.

A. Awareness of Damaging Media in Big Datasets

Most popular social networks and media sharing sites allow users to tag objects and people in their uploaded media. Additionally, some services also extract embedded metadata and use this information for indexing and linking. Media is annotated with names, comments, or is directly linked to users'

978-1-4673-1703-0/12/$31.00 ©2013 IEEE

profiles. In particular the direct linking of profiles to pictures was initially met with an outcry of privacy concerns since it greatly facilitated finding information about people. For this reason, social networks quickly introduced the option to prevent people from linking them in media. However, there is also a positive side to this form of direct linking since the linked person is usually made aware about the linked media and can then take action to have unwanted content removed or restrict the visibility of the link. While the privacy mechanism of current services are still limited, hidden and often confusing, once the link is made the affected people can take action.

A more critical case in our view is the non-linked tagging of photos. In this case a free text tag contains identifying information and/or malicious comments. However there is no automated mechanism to inform a user that he was named in or near a piece of media. The named person might not even be a member of the service where the media was uploaded. The threat of this kind of linking is significantly different to the one depicted above. While the immediate damage can be smaller since no automated notification is sent to friends of the user, the threat can remain hidden far longer. The person can remain unaware of this media whereas others can stumble upon it or be informed by peers.

The final case of damaging media does not contain any technical link. Without any link to the person in question this kind of media can only cause harm to the person if someone having some contact to that person stumbles across it and makes the connection. While the likelihood of causing noticeable harm is smaller, it is still possible. The viral spreading of media has caused serious embarrassment and harm in real world cases. The critical issue here is that there is currently no way for a person to pro-actively search for this kind of media in the Big Data deluge to mitigate this threat.

IV. Analysis of Service Privacy

The following section overviews our privacy analysis of different media hosting sites. It includes media access control as well as metadata handling.

Flickr provides the most fine-grained privacy/access control settings of all analysed services. Privacy settings can be defined on metadata as well as the image itself. One particularly interesting feature of Flickr is the geo-fence. The geo-fence feature enables users to define privacy regions on a map by putting a pin on it and setting a radius. Access to GPS data of the user's photos inside these regions is only allowed for a restricted set of users (friends, family, contacts). Flickr allows its users to tag and add people to images. If a user revokes a person tag of himself in an image, no one can add the person to that image again.

Facebook extracts the title and description of an image from metadata during the upload process of some clients. All photos are resized after upload and metadata is stripped off. Facebook uses face recognition for friend tagging suggestions based on already tagged friends. Access to images is restricted by Facebook's ever changing, complex and sometimes abstruse privacy settings [5], [7].

Picasa Web & Google+ store the complete EXIF metadata of all images. It is accessible by everyone who can access the image. The access to images is regulated on a per-album base. It can be set to public, restricted to people who know the secret URL to the album, or to the owner only. A noteworthy feature is that geo-location data can be protected separately. Google+ and Picasa Web allow the tagging of people in images.

Locr is a geo-tagging focused photo-sharing site. As such, location information is included in most images. By default all metadata is retained in all images. Access control is set on a per image basis. Anybody who can see an image can also see the metadata. There are also extensive location-based search options. Geo-data is extracted from uploaded files or set by people on the Locr website. Locr uses reverse geocoding to add textual location information to images in its database.

Instagram and *PicPlz* are services/mobile apps that allow posting images in a Twitter like way. Resized images stripped of metadata but with optional location data are stored by the services. Additionally they allow the posting of photos to different services like Flickr, Facebook, Dropbox, Foursquare and more. Depending on the service used metadata is stored or discarded. For instance, when uploading a photo to Flickr metadata is stripped from the actual file, but title, description as well as geo location are extracted from the image and can be set by the user. In contrast, the *Hipstamatic* mobile app preserves in-file metadata when uploading images to Flickr.

V. Survey of Metadata in Social Media

To underpin the growing prevalence of privacy-relevant metadata and location data in particular and to judge potential dangers and benefits based on real-world data we analysed a set of 20,000 publicly available Flickr images and their metadata. Flickr was chosen as the premiere photo-sharing website, because it can be legally crawled, offers the full extent of privacy mechanisms and does not remove metadata in general. We crawled one photo each from 20k random Flickr users. Of these, 68.8% were Pro users where the original file could be accessed as well. For the others only the metadata available via the Flickr API was accessed. This includes data automatically extracted from EXIF data during upload and data manually added via website or semi-automatically set by client applications. 23% of the 20k users denied access to their extracted EXIF data in the Flickr database. We also took a set of 3,000 images made with a camera phone from 3k random mobile Flickr users. 46.8% of the mobile users were Pro users and only 2% denied access to EXIF data in the Flickr database.

GPS location data was present in 19% of the 20k dataset and in 34% of the 3k mobile phone dataset. While Flickr hosts many semi-professional DSLR photos, mobile phones are becoming the dominant photo generation tool with the iPhone 4 currently being the most common camera on Flickr [12]. Textual location information like street or city names are currently not used much on Flickr. However, as reverse geocoding becomes more common in client applications this will change (cf. Locr in Figure 2). To evaluate the potential privacy impact, we manual checked which photos contained

978-1-4673-1703-0/12/$31.00 ©2013 IEEE

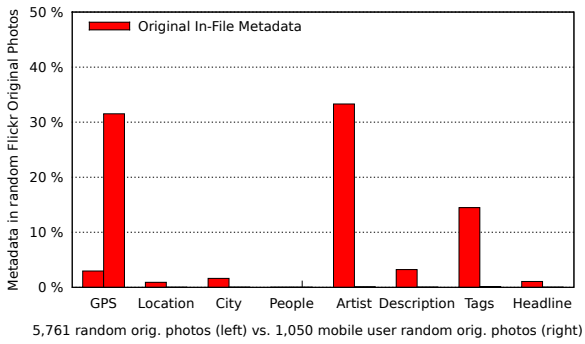

Fig. 1. Public privacy-related metadata in 5.7k random and 1k mobile user original Flickr photos

people and geo-reference, but no user profile tags – i.e. images which could contain people who are unaware of the photo. In the set of 20k images we found 16% and in in the set of 3k mobile photos we found 28% fulfilling this criteria.

We further analysed the subset of images which were available from Pro users, since these can contain the unaltered metadata from the camera. From the 20k dataset, 5761 images contained in-file metadata. From the 3k dataset, 1050 images contained in-file metadata. Figure 1 shows the percentage values for the different types of metadata contained in the files. For the rest of the images metadata was either manually removed by the uploader or the image never has had any in the first place. Of the 20k dataset only 3% of the in-file metadata contained GPS data compared to 32% from the mobile 3k dataset. This shows a clear dominance of mobile devices when it comes to publishing GPS metadata. This itself is unsurprising since most compact and DSLR cameras currently do not have GPS receivers and only few photographers add external GPS devices to these cameras – but this will likely change with future cameras. However, combined with the fact that mobile phones are becoming the dominant type of camera where it comes to the number of published pictures, it is to be expected that the amount of GPS data available for scrutiny, use and abuse will rise further.

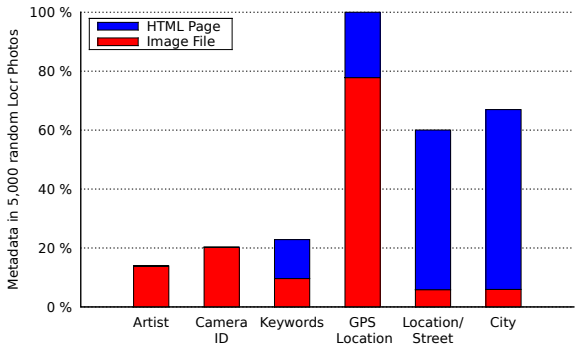

Fig. 2. Public privacy-related metadata of 5k random photos from Locr service

We also collected a 5k dataset of random photos from Locr and analysed the metadata in the images plus the images' HTML pages built from the Locr database. Figure 2 shows the results from this dataset. Particularly interesting is the high rate of non-GPS based location information. This is a trend to watch since most location stripping mechanisms only strip GPS information and leave other text-based tags intact. Furthermore, the amount of camera ID meta-data is notable since these IDs can be used to link different pictures and infer meta-data even if data has been stripped from some of the photos.

To summarise, one third of the pictures taken by dominant camera devices contains GPS information. About one third of these images depict people on it. Thus, about 10% of all the photos could harm other peoples' privacy without them knowing about it.

VI. Location Based Big Data Handling

In the previous sections we discussed some of the privacy issues created by the other peoples' media in particular the privacy issues created by location information. In this section we discuss how this information can also be used to improve a user's privacy by raising awareness about potentially compromising media and thus enabling the user to stay on top of the Big Data wave. We propose leveraging the location tracking capability of modern smartphones to create smart privacy zones in which the user is informed about media events. As was shown above, the number of images which contain location information is already substantial and it is likely to grow further. If a user's phone keeps a GPS record of where the person was at which time, these two pieces of information can be combined with the location data stored in the media to significantly reduce the amount of data which could be relevant to the individual person.

A. Design

The privacy awareness concept consists of a watchdog client a server side watchdog service. Using a GPS-enabled mobile device a user can activate the watchdog client to locally track his position during times he considers relevant to his privacy. Then his device can request the privacy watchdog to show him media that could potentially affect him whenever the user is interested in the state of his online privacy. For this the watchdog client sends the location traces to the watchdog server or servers which then respond with a list of media which was taken in the vicinity (both spatially and temporally) of the user.

The watchdog service can be operated in three different ways. The first two would be value-added services which can be offered by the media sharing sites or social networks (SN) themselves. In both these cases the existing services would need to be extended by an API to allow users to search for media by time and location.

The first type of service would do this via the regular user account. Thus, it would able to see all public pictures, but also all pictures restricted in scope but visible to the user. The benefit of this service type is that through the integration

978-1-4673-1703-0/12/$31.00 ©2013 IEEE

with the account of the user pictures which aren't publicly available can be searched. These private pictures are typically from social network friends and thus the likelihood of pictures involving the user is higher and the scope of people able to view the pictures more relevant. This type of service also has the benefit-of-sorts that the location information is valuable to the SN, so it has an incentive to offer this kind of value-added-service. The privacy downside of searching with the user's account is that the SN receives the information of when and where a user was. While there are certainly users who would not mind this information being sent to their SN if it means they get to see and remove embarrassing photos in a timely manner, there are also certainly users who do not wish their SN to know when and where they were, particularly amongst the clientele who wish to protect their online privacy.

The second type of watchdog service would also be operated by the SN. However, it does not require a user account to do the search and can be queried anonymously. This type of service would thus have a smaller scope, since it can only access publicly available media. A further drawback of this type of service is that there is less of an incentive for the SN provider to implement such a service. While there are sites such as Locr that allow such queries, most SN sites do not. Without outside pressure there is less intrinsic value for them to include such a service compared to the first type.

The third type of service would be a stand-alone service which can be operated by a third party. The stand-alone service operates like an indexing search machine, which crawls publicly available media and its metadata and allows this database to be queried. Possible incentive models for this approach include pay-per-use, subscription, ad-based or community services. The visibility scope would be the same as for the second type of service.

All three types of service are mainly focused on detecting relevant media events and breaking down the Big Data problem to humanly manageable sizes. The concept is mainly focused on bringing possibly relevant media to the attention of the user without overburdening him. The system does not explicitly protect from malicious uploads with which the uploader is intentionally trying to harm another while attempting the activity at the same time. Even though the watchdog service proposed here could make the subterfuge harder for the malicious uploader. But even without full protection from malicious activity we believe that such a watchdog would improve the current state of the art by enabling users to gain better awareness of the relevant part of the social media Big Dataset.

1) Privacy Analysis: These different types of watchdog service can help to reduce the number of relevant pieces of media a user needs to keep an eye on if they don't want uncontrolled media of themselves to be online. However, the devil is in the detail since this form of service can also have serious privacy implications itself if designed in the wrong way. Care must be taken to facilitate the different usage and privacy requirements of different users. The critical issue is the fact that to request the relevant media a user must send location information to the watchdog service.

When using the first type of service, there is little which can be done to protect the location privacy of the user since the correlation between the location query and the user account is direct. One option to protect the privacy to a degree can be an obfuscation approach. For every true query a number of fake queries could also be sent, making it less easy (but far from impossible) for the SN provider to ascertain the true location. However, this approach does not scale well for two reasons. Firstly, it creates a higher load on the SN. However, it is more critical that the likelihood deducing the true location rises, if many queries are sent unless great care is taken in creating the fake paths and masking the source IP addresses. As such this type of service should only be used if the user is willing to "swap" their location data for the best possible update on uploaded media.

Protecting the user's privacy in the second case is simpler. Since the queries do not require an account the only way a user could be tracked directly is his IP address. Using an anonymising service such as TOR sequential queries cannot be linked together and creating a tracking profile becomes significantly harder. The anonymous trace data can of course still be used by the SN, but the missing link to the user makes it less critical for the user himself. The third type of service is probably the most interesting privacy wise, since the economic model behind the service will significantly impact the privacy techniques which can be applied to this model. Most payment models would require user credentials to log in and thus would allow the watchdog service provider to track the user. In this case it would have to be the reputation of the service provider which the user would have to trust, similar to the case of commercial anonymising proxies. In an ad-based approach no user-credentials are needed, thus it would be possible to use the service anonymously via TOR. If so, the watchdog client would need to be open source to ensure that no secret tracking information is stored there. In community-based approaches a privacy model like that in TOR can be used, to ensure none of the participating nodes gets enough information to track a single user.

Each of these proposed service types has different privacy benefits and disadvantages and the trade-off between the two is an interesting area for research.

VII. Related Work

Two services worth mentioning which collect the type of information needed for a privacy watchdog are SocialCamera and Locaccino. SocialCamera [17] is a mobile app that detects faces in the picture and tries to recognise them with the help of Facebook profile pictures of persons that are in your friends list. Recognised people can be automatically tagged and pictures instantly uploaded to Facebook. Locaccino [8] is a Foursquare type application which allows users to upload location-based information into Facebook. These two apps show the willingness of users to share this kind of information in the social web.

978-1-4673-1703-0/12/$31.00 ©2013 IEEE

Ahern et al. analyse in their work [1] privacy decisions of mobile users in the photo sharing process. They identify relationships between the location of the photo capture and the corresponding privacy settings. They recommend the use context information to help users to set privacy preferences and to increase the users' awareness of information aggregation. Work by Fang and LeFevre [11] focuses on helping the user to find appropriate privacy settings in social networks. They present a system where the user initially only needs to set up a few rules. Through the use of active machine learning algorithms the system helps the user to protect private information based on the individual behaviour and taste. In [14] Mannan et al. address the problem, that private user data is not only shared within social networks, but also through personal web pages. In their work they focus on a privacy-enabled web content sharing and utilise existing instant messaging friendship relations to create and enforce access policies.

The three works shown above all focus on protecting a user's privacy based on dangers created by the user himself while sharing media. They do not discuss how users can be protected from other peoples' media. This is prevalent for most of the research work done in this area.

Besmer et al [4] present work which allows users that are tagged in photos to send a request to the owner to hide the linked photo from certain people. This approach also follows the idea that forewarned is forearmed and that creating awareness of critical content is the first step towards the solution of the problem. However the work relies on direct technical tags and as such does not cover the same scope as the privacy watchdog suggested in this paper.

Work that also takes into account other users' media is presented by Squicciarini et al. [16]. They postulate that most of the shared data does not only belong to a single user. Therefore they propose a system to share the ownership of media items and by that strive to establish a collaborative privacy management for shared content. Their prototype is implemented as a Facebook app and is based on game theory, rewarding users that promote co-ownerships of media items. While this work does take into account other users' media, unlike our approach it does not cope with previously unknown and unrelated users.

VIII. Conclusion

In this paper we presented an analysis of the threat to an individual's privacy that is created by other peoples' social media. For this we presented a brief overview of privacy capabilities of common social media services regarding their capability of protecting users from other peoples' activities. We also conducted an analysis of privacy related metadata, particularly location data contained in social media and analysed over 28k real world images from popular social media sites. Based on this survey we analysed the Big Data privacy implications and potential of the emerging trend of geo-tagged social media. We then presented three concepts how this location information can actually help users to stay in control of the flood of potentially harmful or interesting social media uploaded by others.

References

[1] S. Ahern, D. Eckles, N. Good, S. King, M. Naaman, and R. Nair. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 357–366, 2007.

[2] C. a. Ardagna, M. Cremonini, S. De Capitani di Vimercati, and P. Samarati. An Obfuscation-Based Approach for Protecting Location Privacy. *IEEE Transactions on Dependable and Secure Computing*, 8(1):13–27, June 2009.

[3] A. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, Jan. 2003.

[4] A. Besmer and H. Richter Lipford. Moving beyond untagging. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 1563, Apr. 2010.

[5] D. Boyd. Facebook's privacy trainwreck: Exposure, invasion, and social convergence. *Convergence: The International Journal of Research into New Media Technologies*, 14(1):13–20, 2008.

[6] d. Boyd and K. Crawford. Six Provocations for Big Data. *SSRN eLibrary*, 2011.

[7] D. Boyd and E. Hargittai. Facebook privacy settings: Who cares? *First Monday*, 15(8), 2010.

[8] Carnegie Mellons Mobile Commerce Laboratory. Locaccino - a user-controllable location-sharing tool. http://locaccino.org/, 2011.

[9] E. Eldon. New Facebook Statistics Show Big Increase in Content Sharing, Local Business Pages. http://goo.gl/ebGQH, February 2010.

[10] Facebook. Statistics. 2011. http://www.facebook.com/press/info.php?statistics.

[11] L. Fang and K. LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web - WWW '10*, page 351. ACM Press, Apr. 2010.

[12] Flickr. Camera Finder. http://www.flickr.com/cameras, October 2011.

[13] B. Gedik and L. Liu. Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms. *IEEE Transactions on Mobile Computing*, 7(1):1–18, Jan. 2008.

[14] M. Mannan and P. C. van Oorschot. Privacy-enhanced sharing of personal content on the web. In *Proceeding of the 17th international conference on World Wide Web - WWW '08*, page 487. ACM Press, April 2008.

[15] Metadata Working Group. Gartner special report - pattern-based strategy: Getting value from big data. www.gartner.com/patternbasedstrategy, 2012.

[16] A. C. Squicciarini, M. Shehab, and F. Paci. Collective privacy management in social networks. In *Proceedings of the 18th international conference on World wide web - WWW '09*, page 521. ACM Press, Apr. 2009.

[17] Viewdle. SocialCamera. http://www.viewdle.com/products/mobile/index.html.

978-1-4673-1703-0/12/$31.00 ©2013 IEEE