A Literature Survey on Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites

Samiksha Jaiswal[1], Dr. Babita Pathik[2]
[1]M-Tech Research Scholar, [2]Research Guide, Department of Computer Science & Engineering
Lakshmi Narain College of Technology Bhopal

Abstract - Online social networks such as Friendster, MySpace, or the Face book has experienced exponential growth in membership in recent years. These networks offer attractive means for interaction and communication, but also raise privacy and security concerns. In this study we survey a representative sample of the members of the Face book (a social network for colleges and high schools) at a US academic institution, and compare the survey data to information retrieved from the network itself. We look for underlying demographic or behavioral differences between the communities of the network's members and non-members; we analyze the impact of privacy concerns on members' behavior; we compare members' stated attitudes with actual behavior; and document the changes in behavior subsequent to privacy-related information exposure. We find that an individual's privacy concerns are only a weak predictor of his membership to the network. Also privacy concerned individuals join the network and reveal great amounts of personal information. Some manage their privacy concerns by trusting their ability to control the information they provide and the external access to it. However, also find evidence of members' misconceptions about the online community's actual size and composition, and about the visibility of members' profiles.

Keywords - Content Sharing, Image Sharing, Privacy of Content Sharing, User Uploaded Image, Sharing Protection.

I. INTRODUCTION

Doing so would support more sophisticated software systems that share knowledge, information and data on the Web just as people do by publishing text and multimedia. Under the stewardship of the W3C, a set of languages, protocols and technologies have been developed to partially realize this vision, to enable exploration and experimentation and to support the evolution of the concepts and technology.

The current set of W3C standards are based on RDF (Lassila & Swick 1998), a language that provides a basic capability of specifying graphs with a simple interpretation as a "semantic network" and serializing them in XML and other popular Web systems (e.g., JSON). Since it is a graph-based representation, RDF data are often reduced to a set of 'triples' where each represents an edge in the graph or alternatively, a binary predication. The use of OWL to define policies has several important advantages that become critical in distributed environments involving coordination across multiple organizations. First, most policy languages define constraints over classes of targets, objects, actions and other constraints (e.g., time or location). A substantial part of the development of a policy is often devoted to the precise specification of these classes, e.g., the definition of what counts as a 'student' or a 'entertainment activity'. This is especially important if the policy is shared between multiple organizations that must adhere to or enforce the policy even though they have their own native schemas or data models for the domain in question. Second, OWL is based on description logic, a well understood subset of logic for which powerful and efficient reasoning systems are available. By constraining our use of OWL to the right subset, we can exploit existing OWL reasoners. A third advantage is that OWL's grounding in logic facilitates the translation of policies expressed in OWL to other formalisms, either for analysis or for execution. Finally, OWL is designed of and for the Web, making sharing policies and the ontologies they use both natural and easy.

There has been a lot of work done to develop access control frameworks (Moses 2005), (?), (Jajodia et al. 1997). Rein (Rei and N3) (Kagal & Berners-lee 2005) is a distributed framework for describing and reasoning over policies in the Semantic Web. It supports N3 rules (Berners-Lee & Connolly 2008), (Berners-Lee et al. 2005) for representing interconnections between policies and resources and uses the CWM forward-chaining reasoning engine (Berners-Lee), to provide distributed reasoning capability over policy networks. AIR (Kagal, Hanson, & Weitzner) is a policy language that provides automated justification support by tracking dependencies during the reasoning process. It uses Truth Maintenance System to track dependencies. Policies and data are represented in Turtle (Beckett 2007), whereas the reasoning engine is a production rule system (Waterman & Hayes-Roth 1978) with additional features for improved reasoning efficiency such as goal direction. Rei and AIR consider rules defined over attributes of classes in the domain including users, resources, and the context.

Geo-social networking systems

A mobile social network identifies and tracks the geo-spatial locations of a user and other people in her social network and typically can display them on a map interface. So not only can a user share information, media and updates with her friends, but one can also find out exactly where everyone is. Facebook has recently launched a location-based feature "Places Check-in". It lets you check in on the place you are currently at and when you check in, it allows you to tag friends who are with you, just as you can tag a friend in a status update or photo. You can post an update along with your check-in to tell people more about what you are doing. The tricky part here is if you have set your privacy control to "Everyone", other Facebook users will know that you and your friend are in specific location at the current time. Bright kite, one of the popular mobile geo-social networking applications, provides two different modes. In public mode, information shared with everyone with full accuracy, and in private mode, it allows users to share information with people at three different trust levels (Trusted friends, friends and everyone else) and with three different levels of visibility (hidden, city, and exact). Four Square, another popular service, shows a user's current geo-location to her friends, even when the application is not open. Google Latitude also has similar privacy settings wherein a user can hide her location, show exact location, or share the city where she is in with all the invited friends on Google Latitude. Privacy is an important issue with these services, they all have some option and information protection options, but they don't provide strong control while allowing a user to share her geo-location or current activity information.

II. LITERATURE SURVEY

Anna Cinzia Squicciarini, Member, IEEE, Dan Lin, Smitha Sundareswaran, and Joshua Wede [1], have proposed an Adaptive Privacy Policy Prediction (A3P) system that helps users automate the privacy policy settings for their uploaded images. The A3P system provides a comprehensive framework to infer privacy preferences based on the information available for a given user. They also effectively tackled the issue of cold-start, leveraging social context information. Their experimental study proves that A3P is a practical tool that offers significant improvements over current approaches to privacy

Shane Ahern, Dean Eckles*, Nathan Good, Simon King, Mor Naaman, Rahul Nair [2] have discussed a taxonomy of privacy considerations that was surfaced by their study. These considerations can be classified according to four main themes: security, social disclosure, identity, and convenience. For each of these themes, users may consider implications for themselves or for others. They expand on this taxonomy and demonstrate how different users' privacy considerations fall within it. In addition, they show initial evidence that many users have content- and context-derived patterns in making privacy decisions. For example, patterns of "location-based privacy" emerged, showing that, as one user phrased it, "some locations are more private than others".

Alessandro Acquisti1 and Ralph Gross [3], their study is based on a survey instrument, but is complemented by analysis of data mined from the network before and after the survey was administered. They show that there are significant demographic differences between FB member and non-members; that although FB members express, in general, significant concern about their privacy, they are not particularly concerned for their privacy on FB; that a minority yet significant share of the FB population at the Campus we surveyed is unaware of the actual exposure and visibility of the information they publish on FB; and we document that priming about FB's information practices can alter some of its members' behavior.

Jia Deng, Alexander C. Berg, Kai Li1, and Li Fei-Fei [4] have presented the first large scale recognition experiments on 10,000+ categories and 9+ million images. They show that challenges arise from the size and density of the semantic space. The ordering of NN and Linear classification approaches swap from previous datasets to very large scale experiments.They produce a measure of category distance based on the WordNet hierarchy and show that it is well correlated with the difficulty of various datasets.They presented a hierarchy aware cost function for classification and show that it produces more informative classification results. These experiments point to future research directions for large scale image classification, as well as critical dataset and benchmarking issues for evaluating different algorithms.

Simon Jones and Eamonn O'Neillr [5] research addresses the considerable burden of exercising fine-grained control for sharing content with social network contacts. They have presented a novel approach to understanding relationships between properties of SN contacts, content, context and sharing decisions.Their work underlines the complexity of variables affecting sharing decisions but presents a manageable approach to quantifying their effects and takes a step towards better recommendation systems. They have shown that preconfigured privacy based groups which are intended to simplify and reduce the burden of controlling disclosure with individual contacts do not always provide a suitable mechanism for sharing in *all* contexts, as the groups often have to be adjusted. Furthermore, they have identified some factors that affect users' tendency to adjust these groups. They expect that social network users will benefit from the ability to easily to share content with different people based on both their differing relationships and the specific context relating to that content.

SR. NO.	TITLE	AUTHORS	YEAR	METHODOLOGY
1	Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites	Anna Cinzia Squicciarini, Dan Lin, Smitha Sundareswaran, and Joshua Wede	2014	Proposed an Adaptive Privacy Policy Prediction (A3P) system that helps users automate the privacy policy settings for their uploaded images.
2	Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing	Shane Ahern, Dean Eckles, Nathan Good, Simon King, Mor Naaman, Rahul Nair		A taxonomy of privacy considerations that was surfaced by their study.
3	Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook	Alessandro Acquisti1 and Ralph Gross	2006	Study is based on a survey instrument, but is complemented by analysis of data mined from the network before and after the survey was administered.
4	What Does Classifying More Than 10,000 Image Categories Tell Us?	Jia Deng, Alexander C. Berg, Kai Li1, and Li Fei-Fei	2010	Presented a hierarchy aware cost function for classification and show that it produces more informative classification results.
5	Simon Jones and Eamonn O'Neillr	Contextual Dynamics of Group-Based Sharing Decisions	2011	Presented a novel approach to understanding relationships between properties of SN contacts, content, context and sharing decisions.
6	The PViz Comprehension Tool for Social Network Privacy	Alessandra Mazzia, Kristen LeFevre, Eytan Adar	2012	Introduced the PViz policy comprehen-sion tool for social network privacy.
7	Motivations for Annotation in Mobile and Online Media	Morgan Ames,Mor Naaman	2007	Content annotation has been an important research area in the field of human computer interaction.
8	Photos: Concerns, Perceptions, and Protections	Andrew Besmer, Heather Lipford	2009	Photo sharing through online social networking sites is allowing huge numbers of people to upload and socially communicate around photos.
9	Retagging Social Images Based on Visual and Semantic Consistency	Dong Liu, Xian-Sheng Hua, Meng Wang, Hong-Jiang Zhang,	2010	Introduced an image retagging scheme that aims at improving the quality of the tags associated with social images in terms of content relevance.
10	Fast algorithm for mining algorithm	Rakesh Aggarwal, Ramkrishnan shrikant	2008	Presented two algorithm apriori, apriori Tid for all significant association rule between items in large database.

Alessandra Mazzia, Kristen LeFevre, Eytan Adar [6]has introduced the PViz policy comprehen-sion tool for social network privacy. The tool is designed to be more directly aligned with users' mental models of privacy, which often involve natural and user-speci_ sub-groups of friends within their local networks, while allowing users to investigate and assess group membership. They conducted an extensive user study comparing PViz to the current state of the art. The study indicated that PViz results in significantly better accuracy than existing tools for group tasks and provides support for single tasks that is comparable to the existing Audience View interface. They made further modifications to the interface based on par-ticipant feedback, and demonstrated PViz on several non-synthetic networks during a follow-up study.In designing PViz, They focused primarily on the privacy comprehension problem (resolving one's mental model of pri-vacy and publicity with the existing con_guration). Thereare future opportunities to provide improvements in this re-gard (e.g., improved community detection and labeling al-gorithms). They believe that PViz also provides anatural platform for privacy control. In the future, They planto extend the PViz tool to include support for policy mod-ification. They believe that the extension will be straightfor-ward; one possible approach involves attaching

drop-downboxes to the various communities in the visual display, asproposed in prior work.

Morgan Ames,Mor Naaman [7] worked on topic of content annotation has been an important research area in the field of human computer interaction. They conducted qualitative studies in a real-world system where users annotate their data – in this case, photos. The interviews, although focused on a narrow set of users, showed that more than one set of motivations comes into play when users tag their photos. They hypothesize that having these multiple motivations is a determining factor in users' decision to annotate. In particular, social incentivesfor tagging appear to be surprisingly important in motivating users to tag their photographs.Under these conditions, we showed that it is possible to motivate users to annotate content. Point-of-capture annotation (*e.g.*, on the mobile device) can encourage the addition of tags. Tag suggestions and other methods of assisting mobile annotation proved to have broader implications that just assistance in text-entry. In some cases, the suggestions can inspire users to tag their photos andgive them guidance for how best to annotate. Based on ourobservations, we believe that people are more inclined to tag their content when they are given the right incentivesand affordances for annotation.

Andrew Besmer, Heather Lipford[8] has worked on Photo sharing through online social networking sites is allowing huge numbers of people to upload and socially communicate around photos. However, users have lost control over their identity and disclosures as other users can upload and tag undesired photos.

Additionally, users are struggling to manage their identity through the contents of photos across multiple audiences and the many people in their social networks. Users desire and need more tools to allow them to regain control over their privacy, and manage their privacy decisions over time. They have explored some of the concerns that users have in participating in these communities. While this study focused on Facebook in particular, other social network sites are adding similar features. For example, MySpace recently added the ability to tag other users in photos. The concerns, issues and mechanism we discussed will likely be applicable to this and other general social network sites with photo sharing. As these sites continue to grow in popularity and users add more and more photos, meeting users' privacy needs is important to allow safe and comfortable participation on these online communities. They will continue to investigate privacy concerns and new mechanisms to improve privacy management inonline social networking communities.

Dong Liu, Xian-Sheng Hua, Meng Wang, Hong-Jiang Zhang,[9] they have introduced an image retagging scheme that aims at improving the quality of the tags associated with social images in terms of content relevance. Experiments on real-world social image dataset have demonstrated its effectiveness.An effective iterative bound optimization algorithm is applied to learn the optimal tag assignment. In addition, as many tags are intrinsically not closely-related to the visual content of the images, we employ a knowledge-based method to differentiate visual content related from unrelated tags and then constrain the tagging vocabulary of our automatic algorithm within the content related tags.

Rakesh Aggarwal, Ramkrishnan shrikant[10] have presented two algorithm apriori, apriori Tid for all significant association rule between items in large database. They compared these algorithm with previous known algorithm, tha AIS andSETM algorithm.They presented experimental results with both synthetic data and real life data showing that these algoritm will outperform AIS and SETM. They showed that best feature of two algorithm can be combined into hybrid algorithm called apriorihybrid.

III. PROBLEM IDENTIFICATION

Our mobile devices are becoming the dominant way we communicate with people, access information, and consume services. As they become more intelligent, they can and will model our interests, activities and behaviour in order to understand our current context and using it, better serve our needs. When appropriate, aspects of this learned context may be shared with other devices in order to collaborate and provide enhanced service. This development introduces a strong need to allow users greater control of what information is shared with whom and with what level of detail. context encompasses more than just the user's location, because other things of interest are also mobile and changing (Schilit, Adams, & Want 1994). Other important aspects include the ambiance, resources and people nearby, and the activities in which they are engaged. The rise of online social networking systems along with recent improvements in mobile technology, smartphones, and sensor networks presents a unique opportunity for context-aware systems.

A very important but often overlooked issue in most social networking systems is that of privacy. The existing research addressing privacy issues (Acquisti & Gross 2006), (Dwyer, Hiltz, & Passerini 2007), (Gross & Acquisti 2005), (Jones & Soltren 2005), brings out various concerns and emphasizes the need of strong privacy control mechanisms. Fur-thermore, the recent emergence of context-aware geosocial networking services demand more robust access control mechanisms. A context-aware infrastructure should provide the end user with a (logically) central place of privacy control and trust

management, contrary to point solutions within different, possibly not trusted, applications (van Sinderen 2006). Thus, users should be able to define their privacy policies and the context-aware system should be able to protect their information from illegal access as per privacy policies regardless of the application.

For instance, consider a healthcare context-aware system where sensor-enabled mo-bile phones can be used to collect in situ sensor data and context data such as patient's and caretakers' personal information, current location and current activity. In this case the user can specify privacy policies like "allow Dr. Nash detailed information at all time" and "allow access to caretaker's location only in case of emergency". Consider another scenario of university campus; a student user may be willing to let her teachers see where she is between 9:00am and 6:00pm on weekdays but not over the weekend. Further, she may not be willing to let her teachers know about her sleeping activity during the daytime. Additionally, a user may want to control the granularity or accuracy of the answer, depending on current context of her and the requester. For instance, a user might want to share the room-level location to some people and city-level location to others. She might want to share the exact room number to anyone who is in same building as she. On the similar lines, a user may not want to disclose her location if she is at some sensitive place like a nightclub. To incorporate all such privacy policies, the system needs to generalize the contextual data and provide an option to specify policies over different granularity levels of the context data. Overall, we are motivated by the need of privacy control models to control the information flow in collaborative context-aware geo-social networking applications based on the context of both owner and requester. None of the existing models allow users to specify the privacy preferences based on this information in a subtle way. Therefore, in this analysis presented a policy based framework to constrain the information flow based on the con-textual information along with profile information. It can be extended and incorporated in existing social networks including location based mobile social networks. Validate of this architecture in an on-campus context-aware prototype system that aggregates information from a variety of sensors on the phone, online sources, and sources internal to the campus intranet, and infers the dynamic user context.

PROPOSED METHODOLOGY

The proposed system is shown in the Fig. 1. The major components of this system are client devices, server side modules and the Internet services that provide social media. The client devices are location aware smart phones. Today's smart phones are programmable and come with a large set of cheap powerful embedded sensors, such as a camera, GPS, accelerometer, digital compass, gyroscope, microphone, and many more. These sensors are enabling the emergence of personal, group and community scale sensing applications. These client devices as well as the server side modules contain a user profiles repository, a privacy control module and content preferences. The server side also contains a content aggregator, a learn and share module and a privacy control module. The content aggregator combines social media like event updates, photos, and videos from Internet services like YouTube, Flickr, Facebook or university information portals. The learn and share module infers the user's dynamic context using sensor data collected by a variety of sensors on the phone, the information from the content aggregator and online sources such as user's calendar. The inferred context is shared with corresponding client device so that the device along with server can handle further context sharing queries from other clients. The requester queries are passed through the privacy control module to constrain the information flow and hence to protect the user privacy. The privacy control module provides the access control mechanisms and aids in controlling the information flow within system. On the client device, it enables privacy sensitive and resource sensitive reasoning over sensed data along with privacy enforcement between peer devices sharing contextual information. The interaction between various components of our system can be described as follows: The system user has a client device to collect the sensor data periodically.

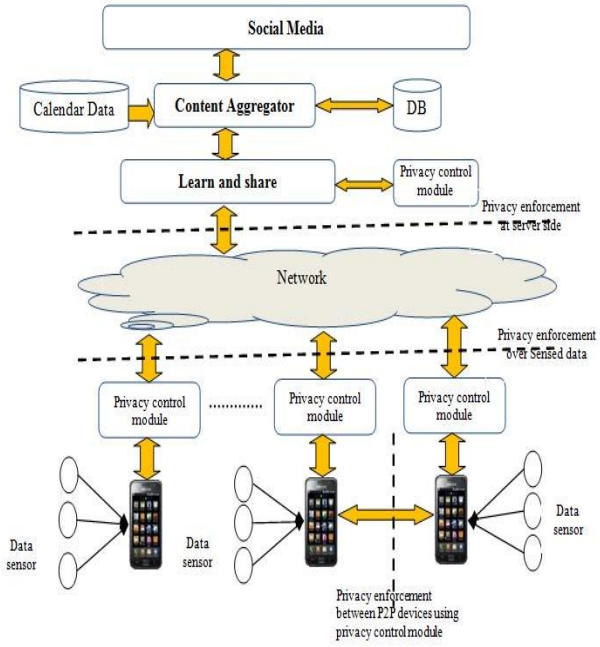

Fig. 3 The architectural view of the system

This data is passed to the learn and share module residing on the server through the privacy control module of client device. The privacy control module decides the specific sensor data that can be shared with the server based on user-specified privacy policies. The learn and share module

infers the user context using sensor data and information from the content aggregator and other online sources. This context consists of current location, activity and additional surrounding information like nearby people.

IV. CONCULESION

System Level Policies The context-aware systems are used by individuals to organization and from social-networking application to military domains. In case of military domains or organizations, the user may not be the sole owner of client device and there is a strong need of robust security mechanisms. It can be in the form of multi-level secure systems where the system-level policies must override the user-level policies. This highlights the need of system-level policies along with the user-specified policies. The system-level policies should be defined by the system-administrator to ensure that the sensitive resources are always protected from illegitimate access. Consider a system-level policy as "Do not share the user's context if she is inside a military building BuildingXYZ."

REFERENCES

[1] Anna Cinzia Squicciarini, Member, IEEE, Dan Lin, Smitha Sundareswaran, and Joshua Wede "Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites" 1041-4347 _ 2014 IEEE.

[2] Shane Ahern, Dean Eckles*, Nathan Good, Simon King, Mor Naaman, Rahul Nair. Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing.

[3] Alessandro Acquisti1 and Ralph Gross. "Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook" PET [2006].

[4] Jia Deng, Alexander C. Berg, Kai Li1, and Li Fei-Fei, What Does Classifying More Than 10,000 Image
Categories Tell Us? [2010]

[5] Simon Jones and Eamonn O'Neillr, Contextual Dynamics of Group-Based Sharing Decisions [2011]

[6]Alessandra Mazzia, Kristen LeFevre, Eytan Adar, The PViz Comprehension Tool for Social Network Privacy [2012]

[7]Morgan Ames,Mor Naaman, Why We Tag: Motivations for Annotation in Mobile and Online Media [2007]

[8] Andrew Besmer, Heather Lipford Tagged Photos: Concerns, Perceptions, and Protections [2009]

[9] Dong Liu, Xian-Sheng Hua, Meng Wang, Hong-Jiang Zhang, Retagging Social Images Based on Visual and Semantic Consistency∗[2010]

[10] Rakesh Aggarwal, Ramkrishnan shrikant, fast algorithm for mining algorithm [2008]

[11]Beckett, D. Turtle - Terse RDF Triple Language. Technical report. 2007.

12] Berners-Lee, T., and Connolly, D. Notation3 (N3): A readable RDF syntax. Technical report. 2008.

[13] Berners-Lee, T. Cwm - a general purpose data processor for the semantic web.

[14] Berners-Lee, T.; Connolly, D.; Prud'hommeaux, E.; and Scharf, Y. 2005. Experience with n3 rules. In Rule Languages for Interoperability.

[15] Carroll, J. J.; Dickinson, I.; Dollin, C.; Reynolds, D.; Seaborne, A.; and Wilkinson, K. 2004. Jena: implementing the semantic web recommendations. 74–83. New York, NY, USA: ACM.

[16] Cheverst, K.; Davies, N.; Mitchell, K.; Friday, A.; and Efstratiou, C. 2000. Developing a context-aware electronic tourist guide: some issues and experiences. In CHI, 17–24.

[17] Doyle, J. 1978. Truth maintenance systems for problem solving. Technical report, Cambridge, MA, USA.

[18] Dwyer, C.; Hiltz, S. R.; and Passerini, K. 2007. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In Proceedings of the Thirteenth Americas Conference on Information Systems (AMCIS).

[19] Gross, R., and Acquisti, A. 2005. Information revelation and privacy in online so-cial networks. In Proceedings of the 2005 ACM workshop on Privacy in the electronic society, WPES '05, 71–80. New York, NY, USA: ACM.

[20] Hayes, G. R.; Patel, S. N.; Truong, K. N.; Iachello, G.; Kientz, J. A.; Farmer, R.; and Abowd, G. D. 2004. The personal audio loop: Designing a ubiquitous audio-based memory aid. In In Proceedings of Mobile HCI, 168–179. Springer Verlag.