ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

Social Networks Privacy-Preserving On Collaborative Tagging and Spam Filter Using Naive Bayes Algorithm

L.Sundarrajan, S.Gunasekaran

PG Student, Department of Computer Science Engineering, V.S.B Engineering College, Karur, India

Associate Professor, Department of Computer Science and Engineering, V.S.B Engineering College, Karur, India

ABSTRACT: Collaborative tagging is one of the most popular services available in social networks, and it allows user to classify either online or offline resources based on their feedback, deliver in the form of tags. Although tags may not be secret information the wide use of collaborative tagging services increases the risk, thereby seriously compromising user privacy. In this paper, we make a contribution towards the development of a privactivey-preserving collaborative tagging service, by showing how a specific privacy-enhancing technology, namely tag suppression, can be used to protect end-user privacy.In most group key management protocols, group members are authenticated by the group leader "one by one." That is, n authentication messages are required to authenticate n group members. Then, these members share one common group key for the group communication. In our batch authentication protocols, users are simultaneously authenticated by the requester that is, one authentication message is required to authenticate n session peers.Spam is commonly defined as irrelevant comments or text, the goal of spam is to distinguish between irrelevant and relevant comments. Naive Bayes classifiers are among the most successful known algorithms for learning to classify text documents.Bayesian spam filtering has become a popular mechanism to distinguish illegitimate spam texts from legitimate texts

KEYWORDS*:* Policy-based collaborative tagging, tag annihilation,privacy-enhancing technology, social networking

I. INTRODUCTION

Data mining, *the extraction of hidden predictive information from large databases*, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- ✓ Massive data collection
- ✓ Powerful multiprocessor computers
- ✓ Data mining algorithms

The Scope of Data Mining

Data mining derives its name from the similarities between searching for valuable business information in a large database — for example, finding linked products in gigabytes of store scanner data — and mining a mountain for a vein

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities:

- **Automated prediction of trends and behaviors**. Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.

- **Automated discovery of previously unknown patterns.** Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors.

II. PREVIOUS WORK

The data disorder technology considered in tag annihilation, a technique that allows a user to Abstain from tagging certain resources in such a manner that the profile resulting from this disorder does not capture their interests so accurately. Our conceptually simple technique protects user privacy to a certain degree, but at the cost of the semantic loss incurred by annihilation tags. Other approaches based on data disorder include the submission of false tags.

In social network, users can allow friends to access their data, depending on their personal endorsement and privacy requirements. Although social network currently provide simple access control mechanisms allowing users to lead access to information contained in their profile, users unfortunately have no control over data. If a user posts a comment in a friend's space cannot specify which users can view the comment. In another case, when a user uploads a photo and tags friends who appear in the photo, the tagged friends cannot restrict who can see this photo, even though the tagged friends may have different privacy about the photo.

Drawbacks:

- Most of the current security protocols for P2P-based OSNs lack specific procedures.
- Each user has to be authenticated by OOB methods, which may slowdown the extension speed of social networks
- Do not consider the restrictions of underlying devices such as computing power and memory limitations.
- Illegal message will be posted in comments

III. PROPOSED SYSTEM

We pursue a systematic solution to facilitate collaborative management of shared data in social network. We begin by examining how the lack of tagging access control for data sharing in social network can undermine the protection of user information. Some typical data sharing honour to multiparty authorization in social network are also identified. Based on these sharing patterns, an tagging model is formulated to capture the core features of multiparty authorization requirements that have not been accommodated so far by existing access control systems and models for social network. Our models also contains a multiparty policy specification. Meanwhile, since conflicts are undeniable in multiparty authorization enforcement.

Spam is commonly defined as irrelevant comments or text, the goal of spam is to distinguish between irrelevant and relevant comments. Naive Bayes classifiers are among the most successful known algorithms for learning to classify text documents. Bayesian spam filtering has become a popular mechanism to distinguish illegitimate spam texts from legitimate texts

- The proposed framework reduces the communication cost required for authenticating users.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

- By incorporating different trust levels, the proposed protocols allow a user with a high trust level to help authenticate other users and achieve the sensibility of a social network.
- The proposed protocols support a one-to-many authentication, which is the basis of batch authentication, to simultaneously authenticate multiple users. To the best of our knowledge, this paper is the first study that offers one-to-many batch authentication in P2P-based social network.
- Spam classification method is used to filter the message as illegal or good message.

IV. SYSTEM MODULES

Collaborative tagging system is divided into five major modules:

1. User Interface Design.
2. User profile creation.
3. Post wall creation.
4. Spam filtering method.

1. WEB SEARCH ENGINE CREATION:

User interface design or user interface engineering is the design of computers, appliances, machines, mobile communication devices, software applications, and websites with the focus on the user's experience and interaction. The goal of user interface design is to make the user's interaction as simple and efficient as possible, in terms of accomplishing user goals—what is often called user-centered design. To run our remote control system we develop a GUI application in J2EE. User can easily execute the project with the help of GUI.

A social networking service is a platform to build social networks or social relations among people who, for example, share interests, activities, backgrounds or real-life connections. A social network service consists of a representation of each user (often a profile), his social links, and a variety of additional services.

Privacy Implications

Privacy implications associated with online social networking depend on the level of identifiability of the information provided, its possible recipients, and its possible uses. Even social networking websites that do not openly expose their users' identities may provide enough information to identify the profile's owner. This may happen, for example, through face reidentification. A15% overlap in 2 of the major social networking sites they studied. Since users often re-use the same or similar photos across different sites, an identified face can be used to identify a pseudonym profile with the same or similar face on another site. Similar re-identifications are possible through demographic data, but also through category-based representations of interests that reveal unique or rare overlaps of hobbies or tastes. We note that information revelation can work in two ways: by allowing another party to identify a pseudonymous profile through previous knowledge of a subject's characteristics or traits; or by allowing another party to infer previously unknown characteristics or traits about a subject identified on a certain site.

Data Visibility and Privacy Preferences

For any user of the Facebook, other users fall into four different categories: friends, friends of friends, non-friend users at the same institution and non-friend users at a different institution.14 By default, everyone on the Facebook appears in searches of everyone else, independent of the searchers institutional affiliation. In search results the users' full names (partial searches for e.g. first names are possible) appear along with the profile image, the academic institution that the user is attending, and the users' status there. The Facebook reinforces this default settings by labelling it "recommended" on the privacy preference page. Also by default the full profile (including contact information) is visible to everyone else at the same institution.

Profile Search ability

We first measured the percentage of users that changed the search default setting away from being searchable to everyone on the Facebook to only being searchable to CMU users. We generated a list of profile IDs currently in use at CMU and compared it with a list of profile IDs visible from a different academic institution. We found that only 1.2% of users (18 female, 45 male) made use of this privacy setting.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

Profile Visibility

We then evaluated the number of CMU users that changed profile visibility by restricting access to CMU users. We used the list of profile IDs currently in use at CMU and evaluated which percentage of profiles was fully accessible to an unconnected user (not friend or friend of friend of any profile). Only 3 profiles (0.06%) in total did not fall into this category.

Online social networks are both vaster and looser than their offline counterparts. It is possible for somebody's profile to be connected to hundreds of peers directly, and thousands of others through the network's ties. Many individuals in a person's online extended network would hardly be defined as actual friends by that person; in fact many may be complete strangers. And yet, personal and often sensitive information is freely and publicly provided.

Our study quantifies patterns of information revelation and infers usage of privacy settings from actual field data, rather than from surveys or laboratory experiments. Still, the relative importance of the different drivers influencing Facebook users' information revelation behavior has to be quantified. Our evidence is compatible with a number of different hypotheses. In fact, many simultaneous factors are likely to play a role.

2. USER PROFILE CREATION:

A user profile (user profile, or simply profile when used in-context) is a collection of personal data associated to a specific user. A profile refers therefore to the explicit digital representation of a person's identity. A user profile can also be considered as the computer representation of a user model. A user profile is a visual display of personal data associated with a specific user, or a customized desktop environment. A profile refers therefore to the explicit digital representation of a person's identity. A user profile can also be considered as the computer representation of a user model. A profile can be used to store the description of the characteristics of person. This information can be exploited by systems taking into account the persons' characteristics and preferences. The user personal data store in ONLINE social networks (OSNs) database that details contain informs like first name, last name, username, password, email Id, gender etc.

New User Problem

The user has to rate a sufficient number of items before a content-based recommender system can really understand the user's preferences and present the user with reliable recommendations. Therefore, a new user, having very few ratings, would not be able to get accurate recommendations.

Collaborative Methods

Several recommendation systems use a hybrid approach by combining collaborative and content-based methods, which helps to avoid certain limitations of content-based and collaborative systems. Different ways to combine collaborative and content-based methods into a hybrid recommender system can be classified as follows:

1. Implementing collaborative and content-based methods separately and combining their predictions,
2. Incorporating some content-based characteristics into a collaborative approach,
3. Incorporating some collaborative characteristics into a content-based approach, and
4. Constructing a general unifying model that incorporates both content-based and collaborative characteristics.

Adding Content-Based Characteristics to Collaborative Models Several hybrid recommender systems, including Fab and the "collaboration via content" approach, described in are based on traditional collaborative techniques but also maintain the content-based profiles for each user. These content-based profiles, and not the commonly rated items, are then used to calculate the similarity between two users. Recommender systems made significant progress over the last decade when numerous content-based, collaborative, and hybrid methods were proposed and several "industrial strength" systems have been developed. However, despite all of these advances, the current generation of recommender systems surveyed in this paper still requires further improvements to make recommendation methods more effective in a broader range of applications.

3. POST WALL CREATION:

The Website wallpost is the most social network is enabling with photo sharing activities. Protected albums allow users to set their albums with access protection. This is one of the beneficial features from wallpost that who fear with photo

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

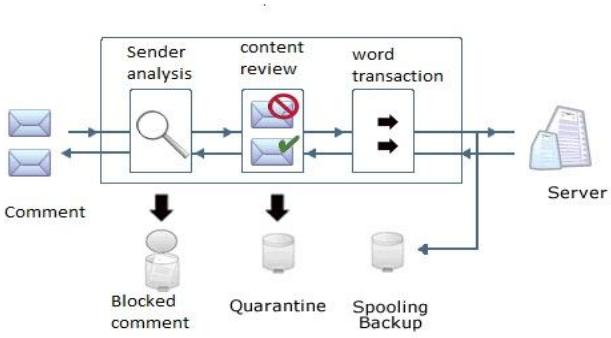
(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

scams on photo sharing websites. Photo tagging the option makes the photo search easier after a long period of time. Here ruse can give the names or keywords for photos that related to the photo in better to recognize easily. Although social network currently provide simple access control mechanisms allowing users to govern access to information contained in their own spaces, users, unfortunately, have no control over data residing outside their spaces.

In this module user can add their or interested photos in their wall. This wall posting contains the photo, photo description, tag information are given by the user that details are stored in the social network database.

4. SPAM FILTERING METHOD:
Spam is to distinguish between irrelevant and relevant comments. Naive Bayes classifiers are among the most successful known algorithms for learning to classify text documents.Bayesian spam filtering has become a popular mechanism to distinguish illegitimate spam texts from legitimate texts



The training stage of the spam detector includes following steps:
1. Preparation of Training Set.
Training Set is divided into positive set (spam comments) and negative set (ordinary comments).
2. Generating word lists.Preparation of Testing/classifying Set.
All comments in the testing set are pre classified and mixed together**.**
Generating word list.
A tokenizer tokenizes main bodies into word lists. A stop word list is used to delete stops words from word lists.
Generating word maps.
A word map is a list of words that appear both in a given. One word map contains words that appear both in the given commentand the spam vocabulary table. Another word map contains words that appear both in the given comments and the comments vocabulary table. These two word maps can be different from each other since some words appear in spam comments may not appear in ordinary comments according to the training set

V. CONCLUSIONS

Collaborative tagging is currently an extremely popularonline service. Although nowadays it is basically used tosupport resource search and browsing, its potential is still tobeexploited. One of these potential applications is theprovision of web access functionalities such as contentfiltering and discovery. For this to become a reality, however,it would be necessary to extend the architecture of currentcollaborative tagging services so as to include a policy layerthat supports the enforcement of user preferences.On the other hand, as collaborative tagging has beengaining popularity, it has become more evident the need forprivacy protection; not only because tags are sensitiveinformation per se, but also because of the risk of crossreferencing. In a nutshell, collaborative tagging would alsobenefit from a layer helping users protect their privacy.

Motivated by all this, our first contribution is anarchitecture that incorporates two layers on support ofenhanced and private collaborative tagging. More specifically,the proposed architecture consists of a bookmarkingservice and two additional services built on it. The formerservice enables users to specify policies both to blockundesired web content and to denote resources of interest.The latter implements tag suppression, a privacy-preservingtechnology based on

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

data perturbation. The combinationof these two services allows us then to broaden thefunctionality of collaborative tagging systemsand, at thesame time, provide users with a mechanism to preservetheir privacy while tagging.

REFERENCES

1. P. Mika, "Ontologies Are Us: A Unified Model of Social Networks and Semantics," Proc. Int'l Semantic Web Conf. (ISWC '05), Y. Gil, E. Motta, V. Benjamins, and M. Musen, eds., pp. 522-536, 2005.
2. X. Wu, L. Zhang, and Y. Yu, "Exploring Social Annotations for the Semantic Web," Proc. 15th Int'l World Wide Web Conf. (WWW), pp. 417-426, 2006.
3. B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and S. Gerd, "Evaluating Similarity Measures for Emergent Semantics of Social Tagging," Proc. 18th Int'l Conf. World Wide Web (WWW), pp. 641-650, 2009.
4. C. Marlow, M. Naaman, D. Boyd, and M. Davis, "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read," Proc. 17th Conf. Hypertext and Hypermedia (HYPERTEXT),
5. pp. 31-40, 2006.
6. Z. Yun and F. Boqin, "Tag-Based User Modeling Using Formal Concept Analysis," Proc. IEEE Eighth Int'l Conf. Computer Information Technology (CIT), pp. 485-490, 2008.
7. B. Carminati, E. Ferrari, and A. Perego, "Combining Social Networks and Semantic Web Technologies for Personalizing Web Access," Proc. Fourth Int'l Conf. Collaborative Computing:

BIOGRAPHY

Sundarrajan.L is a PG Student in Computer Science Engineering, V.S.B College Of Engineering, karur.

Gunasekaran.S Assistant Professor, Department of Computer Science and Engineering, V.S.B Engineering College, karur