Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.653 – 661

RESEARCH ARTICLE

A Customizable Content Based System to Filter Unwanted Messages from OSN User Wall

Ms. Mayuri Uttarwar [1], Prof. Pragati Patil [2], Prof. Yogesh Bhute [3]

[1] Department of Computer Science and Engineering, ABHA Gaikwad-Patil College of Engineering, RTMNU, India

[2] Department of Computer Science and Engineering, ABHA Gaikwad-Patil College of Engineering, RTMNU, India

[3] Department of Computer Science and Engineering, ABHA Gaikwad-Patil College of Engineering, RTMNU, India

[1] mayuriuttarwar@gmail.com; [2] pragatimit@gmail.com; [3] yog.bhute@gmail.com

Abstract – Today's online social networks have become a fairly common way to keep in touch with friends and family. The potential of this enhanced connectivity is very huge however the negative impacts are quite evident. Cyber bullying due to online social network (OSN) is also on rise therefore the need of present situation is not only to avoid unwanted wall post but also to avoid undesired creator. In this paper we have presented a system which prevents unwanted messages to get published on OSN user wall and at the same time messages from undesired creator can get blocked as well. Additionally we are implementing a strategy to blacklist user on a tentative or permanent basis with respect to bad post he wishes to post on OSN user wall. A machine learning technique and a customized filtering rule is used for this purpose.

Index Terms— Online Social Network, Customized filtering rule, Machine learning technique

I. INTRODUCTION

Online Social Networks are proliferated as a communication tool. It has become an integral part of our daily life offering new and varied ways of communicating with people. Both positive and negative impacts of online social network are evident. A wall is a portion in OSN user profile where others can post messages or can attach an image. This wall is a public writing space so others can view what has been written on wall. Users on a social networking site will have a profile page and will usually have some control over what they allow people to see. Some people share with everybody, while others take advantage of privacy settings and allow access only to those known to them. For example Facebook allows user to specify who is allowed to write a message on his wall, that is only specified list of friends, friends of friends or only friends. Superwall is facebook application that allows simple text messages, picture messages but no content preferences can be suggested by wall owner.

Axxis superwall also gives user the option to attach images, links, videos, MP3 songs, documents but no personalization and filtering of wall messages on the basis of content can be done. Recent studies have shown that the cyber bullying can occur due to undesired wall post of OSN user. In general, cyber bullying involves posting a harmful text and/or images using the Internet or other digital communication devices, such as cell phones. Cyber bullying in terms of flaming, harassment and stalking can occur due to undesired post on wall of OSN user. Therefore the major work of OSN today is to prevent undesired messages to be posted on user's wall and to avoid messages from undesired creators independent of their content.

II. RELATED WORK

A. Content based filtering

Marco Vanetti, Elisabetta Binaghi in [1] uses a hierarchical two step classifier. The first level of classification categorizes the wall messages into Neutral and Non neutral category whereas the second level classifies non neutral message into violence, vulgar; offensive, hate, sex categories depending upon its content. Every wall message is classified in terms of these categories and if it is non neutral it get blocked. The Filtered wall architecture conceptual architecture based on content based message filtering and a short text classifier is taken into consideration to avoid unwanted messages to get published on osn user wall. The system in this paper focuses mainly on filtering rule and message classification leaving blacklist user strategy as an enhancement.

B. Short Text Classification in Twitter

Bharath Sriram, David Fuhry in [2] classifies incoming tweets into categories such as News, Events, Opinions, Deals and Private Messages depending on the author information and features within the tweets. Such categorization of tweet requires the knowledge of source of information. A greedy strategy is used to select the feature set which follows the definitions of classes.The system implements an approach to classify tweets into general but important categories by using the author information and features within the tweets.User can subscribe or can view only certain types of tweets based on his interest.

C. Machine learning in text categorization

The problem of classification is a supervised learning approach as the learning step is supervised by training instances. Many different classifiers is used for classification problem in literature. Probabilistic methods of classification often generate numeric or quantities results and hence are sometime criticized as their effectiveness are not easily understandable or interpretable to human [3].A particular applicative context may exhibit very different characteristics and different classifiers may respond differently to these characteristics. In social network too different classifier such as radial basis function neural network [1], decision tree classifiers [4] are used depending on the result of classification that is desirable. Below table in [6] is the comparison of different classification method in pattern determination.

Classifier	Comment
Nearest mean classifier	Fast testing, metric dependent
Binary decision tree	Fast testing, overtraining sensitive, needs pruning
RBFN	Overtraining sensitive ,may produce confidence value, may robust to outlier
Support vector classifier	Slow training, overtraining sensitive, metric dependent
Perceptron	Sensitive to training parameter, may produce confidence value
k-Nearest Neighbour	Slow testing, scale(metric) dependent, Asymptotically optimal

Table 1.Classification methods

D. Blacklist system

In OSN context we would like to blacklist certain misbehaved users in terms of bad wall messages post. But privacy of these of kind users must be preserved in the sense that this user is viewed as blacklist to only those

users who have blacklist this user because of frequent bad message post on their wall. In a system like nymble [7] the misbehaved users are blacklisted without compromising their anonymity.

III. PROBLEM STATEMENT

Today's OSN facilitates its user to prevent undesired wall messages to a small extent and no preferences for wall messages can be suggested for example one want to avoid vulgar post on his wall. No blacklist strategy has been implemented to block certain user on tentative or permanent basis. A machine learning based classifier in conjunction with filtering and blacklist strategy can overcome such problem and user can be able to avoid all undesirable post on his wall.

IV. EXISTING SYSTEM

Online network such as facebook facilitates its users to mention who is allowed to insert message on wall that is friends, friends of friends or defined group of friend but no content preferences such as vulgar or offensive can be suggested and thus cannot be prevented.

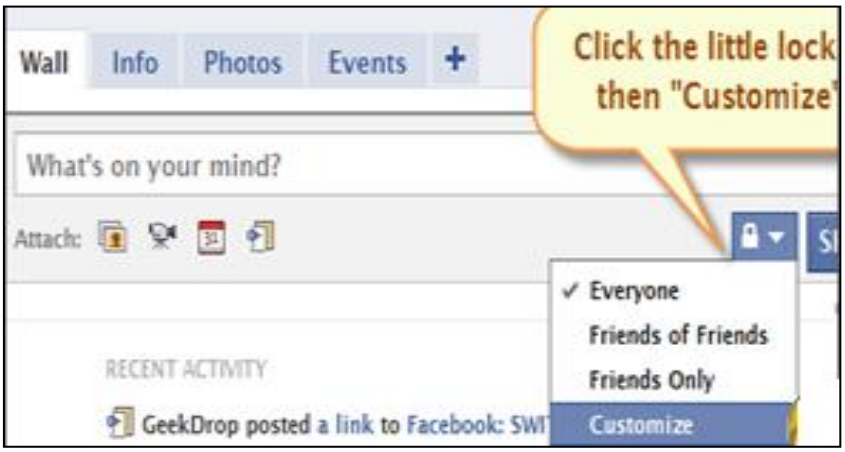

Fig 1.wall message setting in facebook

Many online social networking sites such as twitter, linkedin, facebook support blacklisting of user. For example for blocking a user on facebook one can specify such user by his name as in figure 2 below but no system alert on the basis of frequent bad wall messages can be seen to wall owner.

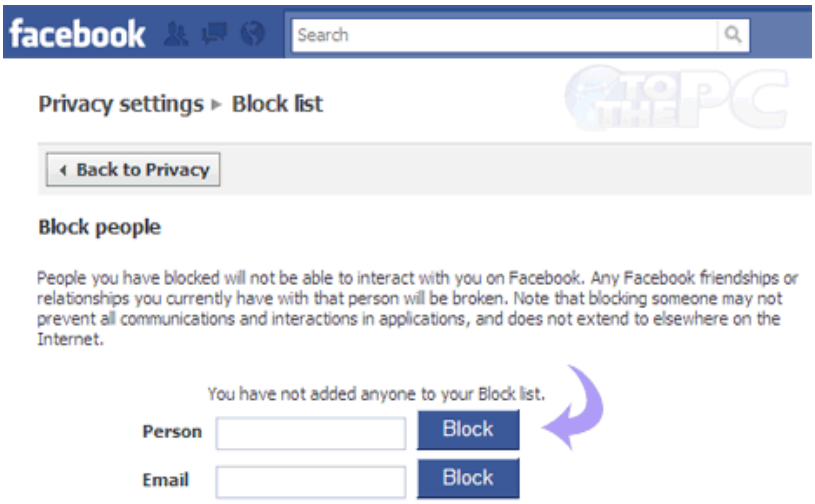

Fig2. Blocking a friend on facebook

V. PROPOSED SYSTEM

The motive of proposed system is to build an automated system to filter unwanted messages from OSN user wall. To develop a system that

- Enable OSN users to have a control on the messages posted on their walls.
- Avoid messages from undesired creators (blacklist user), irrespective of their contents.
- Allow users to specify filtering rules for wall messages

Whenever a friend frequently tries to post bad wall message, a wall owner should get a system alert whether he want to blacklist such a friend on a tentative or permanent basis showing his bad scrap count.

VI. SYSTEM FLOW

The flow of system can be summarized by a flowchart shown in fig 3.Whenever a registered user login to a system and navigate to friend's wall to post a message it is intercepted by system. The system first checks whether the user is in blacklist list of friend. If yes, the user gets an alert that he can't post else the user is checked against the filtering rule (FR) created by wall owner. If the user is matched against the parameter in FR of his friend then he can't post a message else the user can post a message as shown in fig below.

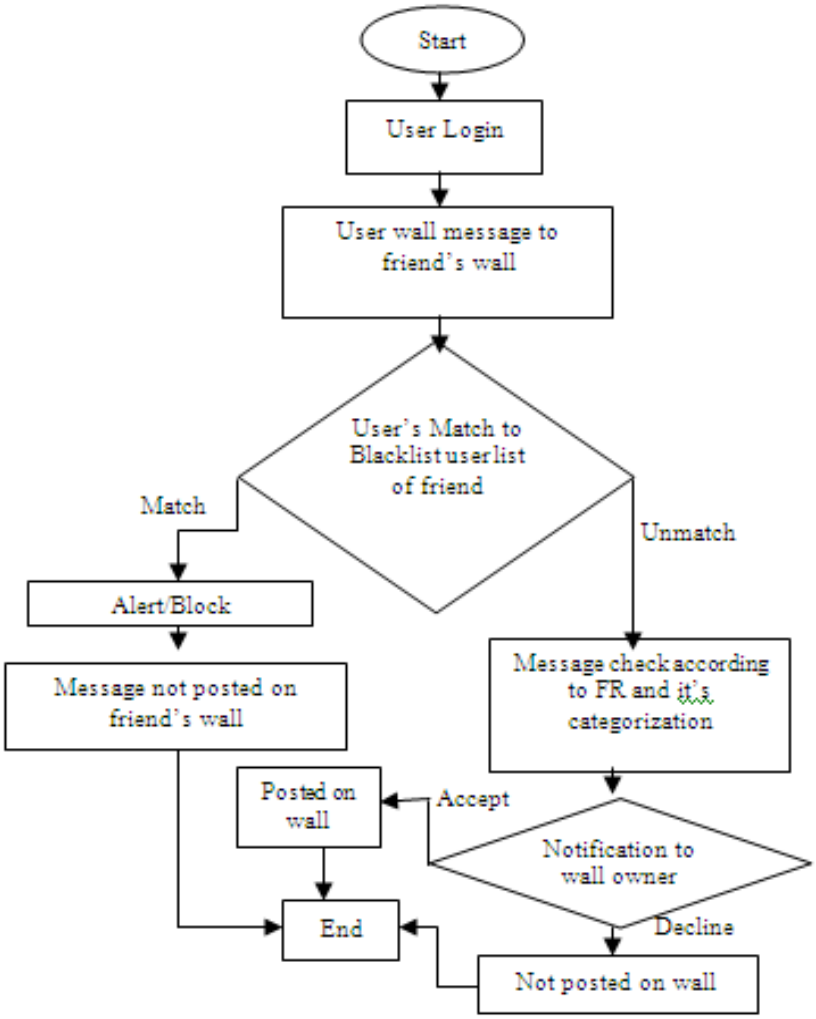

Fig 3. System Flow

The categorization of message posted by user is done and sends as to notification to friend in which the friend can accept or deny a user's message. If it is accepted the message get published on friend's wall else is rejected. From this message path wall messages get filtered twice in our system. These two types are

- System Filtering
- OSN user Filtering

The system categorizes the wall messages whereas the messages get published on wall only on the will of wall owner irrespective of its content.

VII. SYSTEM DESIGN

The system in this paper aims to have filtered wall that is the wall contained only those posts that are desirable to OSN user. The system consist of mainly following components as in [8]

- Filtered wall interface
- Social network application
- Social network manager

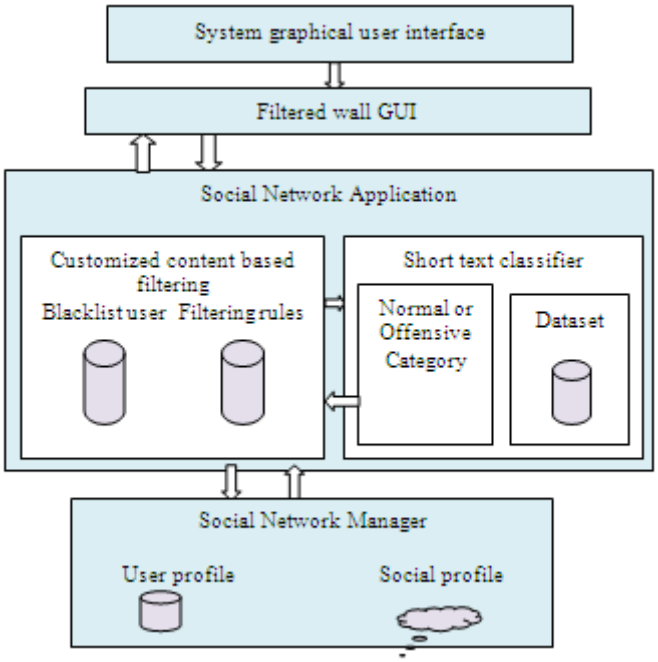

Fig 4. System Design

The system graphical user interface composed of interface to insert user credentials to login into system as well as new user registration. The Filtered wall interface consist of components to post a message on user wall which on submission go through social network application layer and social network manager layer before being published on user wall. The social network manager layer extract data from user social profile and provide it to the social network application layer to impose filtering rules. The social network application layer consists of two main components customized content based filtering and short text classifier. The modules in customized content based filtering checks whether the friend is in blacklist friend list and impose the filtering rules provided by wall owner on friend's message. The short text classifier categories the message into normal or controversial category based on its content. The trained short text classifier uses the pre-classified data to categorize a message. Depending upon the result provided by the underlying layers and user's will the message will be published or blocked.

VIII. Machine Learning Classifier

Wall messages categorization into controversial category and a normal one is a task of supervised machine learning based classifier. A supervised learning algorithm analyzes the trained data and produces an inferred function on the basis of which new examples can be mapped.

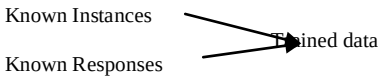

Fig 5 Trained data creation

The classifier must be trained with large and accurate data set to go well. As the underlying domain is dynamic, there are more chances of training data becoming stale. Therefore we have added dynamic dictionary module to the system which can be used to update data as needed by OSN user.

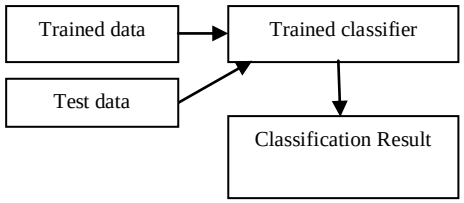

Fig 6 Training and testing of classifier

Decision tree classifier is used in this context for classification task. The main objectives of decision tree classifiers as in [9] are listed below

- to classify correctly as much of the training sample as possible
- generalize beyond the training sample so that unseen samples could be classified with as high of an accuracy as possible
- be easy to update as more training sample becomes available
- Have as simple a structure as possible.

There are numerous advantages of decision tree that are remarkable in literature.

- Global complex decision regions can be approximated by the union of simpler local decision regions at various levels of the tree [9].
- A sample is tested against only certain subsets of classes, thus eliminating unnecessary computations.
- It gives the flexibility of choosing different subsets of features at different internal nodes of the tree

The disadvantages of decision tree such as overlap when numbers of classes are large, errors accumulation in large tree from level to level can be overcome in our system as the number of classes as well as tree is not too much large. There are also difficulties involving design of optimal decision tree. The accuracy of decision tree is also dependable on various parameters such as appropriate choice of tree structure, choice of feature subsets at each internal node of tree and the choice of decision rules to be used at each internal node of the tree. Considering the success parameter of the tree we have designed the binary decision tree in fig 7 and 8.

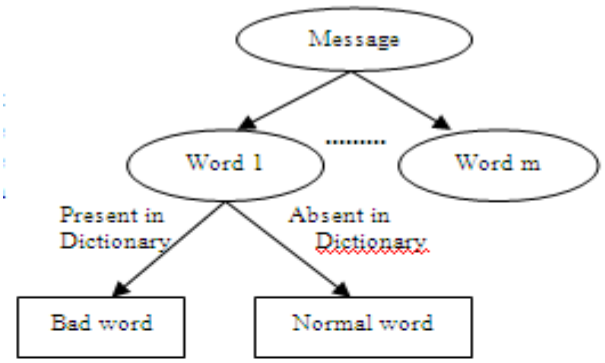

Fig 7 Decision tree 1 for detection of bad word in message

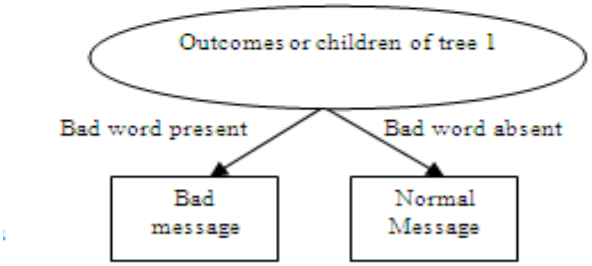

Fig 8 Decision tree classifier for classifying message into normal or bad category

The controversial words are present in dictionary. The wall message may be composed of one or more words. Each word in message is checked for its direct or indirect presence in dictionary of controversial words. If it is present in dictionary the message is classified as bad message else it is classified as normal one. In general the average depth of tree reflects the weight given to classifier efficiency [9].

IX. CUSTOMIZED FILTERING OF WALL MESSAGES

Customized Filtering of wall messages is achieved by means of Filtering rules. Filtering rules are the rules specified by the wall owner with an objective of preventing messages from undesirable creators. Once the user has been created his account he has to create his filtering rules to prevent messages to get published on his wall directly. This rule may be dependent on various factors such as age, sex, country, religion, occupation etc. In our system we have focused on the age, gender, trustworthiness of friend parameters and the action to be taken for matching parameter friend. The trustworthiness of friend can be detected by the relationship with the friend. For example direct friend implies trusted friendship or relationship [10] whereas Indirect indicate less trusted relationship. Therefore filtering rule can be defined as a tuple(creator, (age<numeric value), gender, friend type, action) where

- creator is OSN user or wall owner
- age is the numeric age of friends of wall owner below which he wants to take an action
- friend type refer indirect or direct friend
- Action is the activity block or notifies that should be taken for matched friends.

Notify action implies that the friend's message matching with the above parameter should be transferred to the notification window of wall owner where the owner can accept or decline the friend's message whereas block action indicate that the message from friends matching with the parameters should be blocked and thus can be prevented from publishing on wall of OSN user. On the basis of this parameters different rules can be created.
For example Suppose that Alice is an OSN user and she wants to always notify messages coming from teenagers boys who are indirect friends of her, hence her rule will be as (Alice,age<16,Male,Indirect,notify).

X. BLACKLIST USER

Blacklist users are blocked users whose messages are prevented from publishing irrespective of their content [1]. Blacklist users are blocked on tentative basis or permanently. OSN user can blacklist or block his friend and can be blacklist by a system too with the help of blacklist system strategy. OSN user can blacklist a friend by observing his friend's profile, relationship and number of bad messages coming from the particular friend as well as the system defined strategy counts the number of bad or controversial messages coming from a friend in certain time span and if it exceed certain threshold then a notification goes to wall owner whether he wish to blacklist such friend on tentative or permanent basis. Tentative blocking will blacklist a friend for 15 days whereas permanent blocking implies end of relationship with the friend. Whenever a blacklist friend tries to post a message on his friend's wall he will get an alert blacklist person can't send the message. Taking privacy of blacklist person into consideration such a friend is blacklist and can be seen as blacklist only to a friend who has blacklisted him and other OSN users will not see such a friend as a blacklist one. Thus the privacy of blacklist person is preserved. Additionally the wall owner is able to view a graph of messages coming from his friends (graph 1). By observing bad message statistics in comparison with good one from specific friend OSN user can decide whether to blacklist a friend or not.

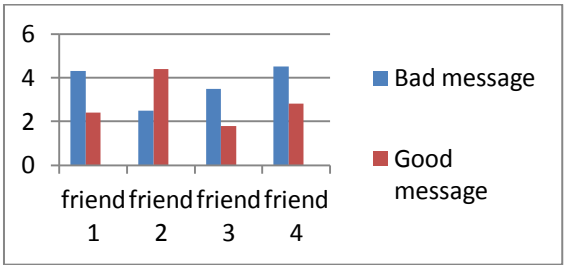

Fig 9. Wall message graph

XI. NUMERICAL RESULT

In a classification, the precision for a class is the number of true positives (TP) divided by the total number of elements labeled as belonging to the positive class that is the sum of true positives and false positives (FP) which are items incorrectly labeled as belonging to the class whereas Recall is defined as the number of true positives divided by the total number of elements that actually belong to the positive class that is the sum of true positives and false negatives(FN), which are items which were not labeled as belonging to the positive class but should have been.

$$\text{Precision} = \frac{TP}{TP + FP} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

F-measure is a measure that combines precision and recall is the harmonic mean of precision and recall.

$$\text{F-measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad \ldots\ldots\ldots\ldots\ldots (3)$$

On the basis of dataset of wall messages we found the results as in table 2.

Metric	Good Message	Bad Message
Precision	90.3%	79.7%
Recall	91.5%	80.1%
F-measure	90%	80%

Table 2. Accuracy of classifier

XII. CONCLUSION

In this paper we have presented a system to prevent undesirable messages to get published on OSN user wall. With the help of filtering rules and blacklist user strategy we prevent messages from undesired creator. Marco Vanetti and Elisabetta in [1] classify non neutral messages into four categories that is violence, vulgar; offensive, hate, sex. With a view to prevent all these four category word we considered only single category that is controversial or bad which covers all these four category words and calculates results on the basis of two category that is normal and bad. The success of the system is dependent on the data that is used for training purpose in machine learning based classifier and therefore is made incremental in our work.

REFERENCES

[1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo,"A System to Filter Unwanted Messages from OSN User Walls," IEEE Transaction on Knowledge and Data Engineering,vol. 25, 2013.

[2] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," in Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, 2010, pp. 841–842.

[3] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[4] Hongyu Gao, Yan Chen, Kathy Lee, Diana Palsetia, Alok Choudhary, "Towards Online Spam Filtering in Social Networks" in Proceedings of the 19th Annual Network & Distributed System Security Symposium,February 2012.

[5] S. Pollock, "A rule-based message filtering system," ACM Transactions on Office Information Systems, vol. 6, no. 3, pp. 232–254,1988.

[6] A. K. Jain, R. P.W. Duin, and J. Mao, "Statistical pattern recognition:A review," IEEE Transactions on Pattern Analysis and MachineIntelligence, vol. 22, pp. 4–37, 2000.

[7] Patrick P. Tsang, Apu Kapadia, Cory Cornelius, and Sean W. Smith, "Nymble: Blocking Misbehaving Users in Anonymizing Networks," Ieee transactions on depedable and secure computing,2009.

[8] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari,"Content-based filtering in on-line social networks," in Proceedings of ECML/PKDD Workshop on Privacy and Security issues in Data Mining and Machine Learning (PSDML 2010), 2010.

[9] S. Rasoul Safavian and David Landgrebe, "A Survey of Decision Tree Classifier Methodology" IEEE Transactions on Systems, Man, and Cybernetics, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 21, No. 3, pp 660-674, May 1991.

[10] J. A. Golbeck, "Computing and applying trust in web-based social networks," Ph.D. dissertation, PhD thesis, Graduate School of the University of Maryland, College Park, 2005.

