Identity-Based Encryption An Alternative to Public Key Encryption In Cloud Computing.

Mr. Sanap Bhima S  
PG Student  
Department of Computer Engineering.  
S.N.D.COE and RC, Yeola

Prof. Chandgude A. S.  
Asst. Professor  
Department of Computer Engineering.  
S.N.D.COE and RC, Yeola

ABSTRACT

Identity-Based Encryption (IBE) which simplifies the public key and certificate management at Public Key Infrastructure (PKI) is an important alternative to public key encryption. However, one of the main efficiency drawbacks of IBE is the overhead computation at Private Key Generator (PKG) during user revocation. Efficient revocation has been well studied in traditional PKI setting, but the cumbersome management of certificates is precisely the burden that IBE strives to alleviate. In this paper, aiming at tackling the critical issue of identity revocation, we introduce outsourcing computation into IBE for the first time and propose a revocable IBE scheme in the server-aided setting. Our scheme offloads most of the key generation related operations during key-issuing and key-update processes to a Key Update Cloud Service Provider, leaving only a constant number of simple operations for PKG and users to perform locally. This goal is achieved by utilizing a novel collusion-resistant technique: we employ a hybrid private key for each user, in which an AND gate is involved to connect and bound the identity component and the time component. Furthermore, we propose another construction which is provably secure under the recently formulated Refereed Delegation of Computation model. Finally, we provide extensive experimental results to demonstrate the efficiency of our proposed construction.

Keywords  
Identity-based encryption, Revocation, Outsourcing, Cloud computing.

1. INTRODUCTION

The popularity of on-line media is increases day by day. As millions of people are connected to social media at a time with their friends, family, colleague etc. People were discuss different issues, news or leading topics with all their circle on on social sites using comments. People use to comments as opinions or their views. There were different comments present on social media sites. Some of them are spam comments. So need to detect that spam comments and spammer. There are several methods that are used for detecting spam comments. Some of them are listed as term frequency-inverse document frequency (tfidf), Naïve Bayes classifier and SVM. The disadvantage of these techniques are they do not consider semantic information of the spam. Due to this issue we get incomplete result of spam detection. So there is need to detection take place on the basis of spam word as well as semantic. Natural Language processing (NLP) is a technique refers to the Artificial intelligence. The NLP Algorithm contains different phases which are given below: 1. Lexical analysis 2. Syntactic analysis 3. Semantic analysis 4. Disclosure Integration and 5. Pragmatic analysis. Text mining is nothing but the text data mining. It is the process of extraction or deriving the high quality information from large text data. The goal of system is to detect spam comments. There are two phases of proposed system first is pre-processing and second is feature extraction. The system will give overall better results than previous techniques. Further, We discuss the details of proposed system.

2. REVIEW OF LITERATURE

M. Cataldi and C. Schifanella [2] they are proposes measurement of spam. For this purpose they propose two measurements such as page rank and term frequency. Term frequency measure weightage of each word this is known as busy keywords. This keyword is used to detect spam and then eliminate it. Hurst, Maykov and Sayyadi [4] It proposes clustering of the word approaches for extracting the words. In this paper the author used twitter comments as a input data for formalizing and frequency count is done after mining the words. This frequency is counted for specific time interval and clustered data. Backstorm and Kleinberg [5] It gives the information of scalable algorithm in clustering which helps to checking the phases or words. It is the algorithm which is used for analysing matching the contents. For this purpose it analyse the large data of social media sites. Andrew Y. Ng and David M.Blei[6] It gives the information and use of model named as Latent Dirichlet Allocation (LDA) model. It is model which is works on the basis of word frequency of given document. This is used to find the accuracy of data according to reasonable data set.

3. SYSTEM OVERVIEW

![Fig No 1. System Overview](image-url)
The system takes the input as comments from social media sites. then these comments are preprocessed in pre-processing module or phase. The output of preprocessing is simple plane text. This plane text is give as an input to feature extraction module. In this phase actual spam is detected by comparing text with self-extensible spam word dictionary. If spam word detected then it is replace with star(***). Finally we get the output as spam has been detected or not.

4. IMPLEMENTATION DETAILS

4.1 Pre-processing

Pre-processing is the initial phase or first phase of proposed system. In this phase we filter comments and convert comments into simple plane text. In this phase we use the concept of natural language processing (NLP) algorithm. We mainly use the lexical analysis phase in NLP.

- **Lexical analysis:** Lexical analysis is used as analyzer or as a parser. It converts sentence into words and also words into character.
- **Elimination of Punctuations:** In this sub module we are removing punctuations. for example comma, full stop etc.
- **Elimination of Symbols:** In this sub module we are eliminating symbols like $ , #, , % etc.
- **Elimination of stop words:** In this sub module we remove words that break the sentence. For example for, and, is, the, of, in etc.

4.2 Feature Extraction

Feature extraction is second phase of proposed system. In this system we take input as simple plane text from pre-processing. Then we detect the spam word by comparing it with spam word dictionary. In this phase we use iteration algorithm for creating the spam word dictionary. Remove of slang word: In this sub module we control the spamming in comment. for this task we create a dictionary of slang words and comparing word of comments with this dictionary if word is available then this word detected as spam word and then it is replaces by star(***).

4.3 Algorithm for Spam Word Dictionary

a) Procedure : Construct the Spam words Dictionary.

b) Input: spam words of AD dictionary and Basic vulgar or added in the result of the previous iteration.

c) Compare semantic of the most similar words from dictionary with our spam word selected;

d) And spam word is added into the candidate spam word dictionary;

e) Delete words of candidate spam dictionary if they exist in basic vulgar dictionary;

f) Calculate the average weight of same words in candidate spam word dictionary.

g) For each spam word in candidate spam word dictionary do Acquire 4 most similar words for spam word by comparing the semantic similarity between them;

h) If there are more than 4 words among exist in Basic vulgar and AD dictionary then Add given spam word into final dictionary otherwise Drop it;

i) Empty the candidate dictionary

j) Output: the newly added spam words in this iteration.

Remove Repeated comments: In extracting the data from document remove ambiguity in result.

5. REQUIREMENT SPECIFICATION

**Software Requirements:**

Software requirement for the system are given below:

a. Operating System: Windows 8
b. Front End: JSP/JAVA
c. Back End: MySQL 5.0
d. Netbeans 8.2

d) Netbeans 8.2

**Hardware Requirements:**

Hardware requirement for the system are given below:

e. Hard disk: 20 GB (min)
f. RAM: 2GB (min)
g. Processor: Intel Core i3
h. Input device: Keyboard and Mouse
6. RESULT ANALYSIS

6.1 Data Set And Expected Output

The expected input of proposed system is comments. The constraints is comment must be in textual format only. The expected output is the given comment is spam or not. We take twitter comments as a input data sets. From these comments spam word are detected and spam word is replaced by *. The overall result of proposed system is spam detected means comment is spam or not.

6.2 Experimental Result

We use classification algorithm to measure a result of our proposed system. The confusion matrix of system is define as,

Table No 1 Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>Spam</th>
<th>Not Spam</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Spam</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Not Spam (Normal comments)</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

There are three threshold values such as Precision Rate, Recall Rate and F1.

Precision Rate

\( P = \frac{a}{a+b} \)

Recall Rate

\( R = \frac{a}{a+c} \)

F1 is the balanced value of P and R used to evaluate the overall result of classification.

\( F = \frac{2PR}{P+R} \)

Consider as we compile 10,000 comments taken from twitter as input file.

Table No 2 Confusion Matrix With Sample Data Set

<table>
<thead>
<tr>
<th></th>
<th>Spam</th>
<th>Not Spam</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Spam</td>
<td>4350</td>
<td>1048</td>
</tr>
<tr>
<td>Not Spam (Normal comments)</td>
<td>1354</td>
<td>3248</td>
</tr>
</tbody>
</table>

By calculating value of a Precision Rate, Recall Rate and F1 we get,

Table No 3 Confusion Matrix With Precision Rate, Recall Rate And F1

<table>
<thead>
<tr>
<th></th>
<th>Precision Rate</th>
<th>Recall Rate</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Spam</td>
<td>4350</td>
<td>1048</td>
<td>78.36</td>
</tr>
<tr>
<td>Not Spam (Normal comments)</td>
<td>1354</td>
<td>3248</td>
<td>88.04</td>
</tr>
</tbody>
</table>

Table No 4 Test Cases

<table>
<thead>
<tr>
<th>Test ID</th>
<th>Expected Result</th>
<th>Actual Result</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>GUI should be display</td>
<td>GUI displayed</td>
<td>Pass</td>
</tr>
<tr>
<td>02</td>
<td>Input file must be opened</td>
<td>Input file is open successfully</td>
<td>Pass</td>
</tr>
<tr>
<td>03</td>
<td>Preprocessing must be done</td>
<td>Preprocessing is done successfully</td>
<td>Pass</td>
</tr>
</tbody>
</table>
CONCLUSION

The main aspects of proposed work are detection of spamming occurred in social media sites. This work is divided into two approaches pre-processing and feature extraction. Preprocessing works on filtering data and convert it into simple plain text and feature extraction extract the spam word by comparing it with self-extensible spam word dictionary. The Next work will be research of spam detection based on order and format of spam words also combination of spam words. And also to implement on-line security and current event detection.

REFERENCES


ACKNOWLEDGMENT
I sincerely express my thanks to my guide Prof. Dhakane V N. for his invaluable guidance, supervision and encouragement throughout the work. I express my gratitude to Prof. V. Dhakane (P.G. Co-coordinator), Prof. A. S. Chandgude (HOD Department of Computer Engineering), and Dr. H. N. Kudal (Principal of S.N.D C.O.E and RC, Yeola) for their kind help and co-operation.