International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V69I9P103

Volume 69 Issue 9, 12-16, September 2021

Original Article

Classifying Unwanted Emails using Naïve Bayes Classifier

Victoria Oluwatoyin Oyekunle[1], Edward E. Ogheneovo[2]

[1]*Department of Computer Science, Federal Polytechnic of Oil and Gas, Bonny, Rivers State, Nigeria.*

[2]*Department of Computer Science, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria.*

Received Date: 18 August 2021
Revised Date: 20 September 2021
Accepted Date: 01 October 2021

Abstract *- In recent years, the increasing use of Electronic mail for fast and cheap personal, official, academic communication, and electronic commerce has led to the emergence and further widespread of problems caused by unsolicited and unwanted bulk e-mail messages. In this study, the objective is to enhance the classification of incoming e-mails-using the Naïve Bayes classifier-into unwanted and ham (legitimate) based on features in both the Subject text of the email and the Email body. The system segments the input email body into tokens and analyses its structure. The dataset is cleaned, and the total number of unique words are counted and extracted, and then compared with already learned unwanted words in the database. If email is classified as 'Unwanted with very high degree' or 'Unwanted with high degree', users are notified and advised to block unwanted emails. Some emails were classified as Ham. This means that users can view such messages as legitimate messages.*

Keywords *- Electronic mail, Machine Learning, Artificial Intelligence, Spam Filtering, Unwanted Emails, Ham, Phishing, junk Email.*

I. INTRODUCTION

Electronic mail (frequently called e-mail) is the method of exchanging electronic messages between computers over a network- usually the internet. It is undoubtedly one of the internet applications most commonly used. With the increased use of the internet, and the number of email users multiplying day by day, it has become one of the finest advertising ways to generate and send distinct messages. According to [1] [2], unwanted emails fit into the following three benchmarks:

- Anonymity (the sender's address and identity are hidden);
- Mass mailing (the email is usually sent to a w3large group of individuals); and
- Unsolicited (the recipients do not request the emails). Spammers use specialized computer programs to gather people's e-mail addresses from social media pages, websites, consumer lists, newsgroups, etc., and sell them to other spammers. The program looks at the code of every webpage; it looks for an email address and saves it to the spammers' database of harvested addresses [3].

Web-based mail systems (for example, Yahoo and Hotmail) have inbox quota limits of a couple of megabytes. These quotas may be exceeded on a daily basis by unwanted email, and legitimate messages will be rejected by the mail servers because the user's inbox is full. For businesses that depend on e-mail services for income, the loss of legitimate mail could prove very expensive and render the utilization of such e-mail services ineffective as a communication tool [10]

Some problems emerge from an email filtering model judging a legitimate email to be an unwanted email which is usually far worse than judging an unwanted email to be a legitimate one. Many current email filtering methods do well, but they must be frequently maintained and tuned as the characteristics of unwanted messages change sometimes.

Today's email packages typically enable the user to design rules to file emails automatically into folders and filter unwanted messages. Most users, however, do not create such rules as they believe it is difficult to use the software or essentially abstain from changing it [4]. To start with, systems that expect users to manufacture sets of rules to identify and differentiate unwanted messages suppose that their users have sufficient time and are wise and intelligent enough to build powerful rules. In addition, manually building a set of solid rules is a challenging job as users need to constantly refine the rules. This is because the characteristics of unwanted emails change over time, and spammers keep re-inventing new methods to bypass email filters. This is a tedious, time-consuming operation that may lead to loss of quality time.

The problems with rule-based systems point to the need for versatile and adaptive methods to block unwanted messages automatically. The filter should learn how to classify emails into a set of folders and should be able to adapt automatically over time to modifications in unwanted features. In addition, by creating a system that can learn straight from information in the email archive of a user, such an unwanted email filter can be tailored to the particular characteristics of unwanted and unwanted messages of a user. This can therefore prompt each user to build more accurate unwanted email filters [5].

[6]-[9] described the Artificial Intelligence approach as being more efficient than knowledge engineering because instead of specifying rules, a set of the training sample is used. These samples are a set of pre-classified e-mail messages. A specific algorithm is then used to learn from these email messages the classification rules. These rules are learned by enabling the model to learn the patterns in the training dataset and then using these learned rules to classify the test dataset.

II. RELATED WORK

[9] implemented a three-way spam filtering decision-making strategy based on Bayesian decision-making theory, which offers customers more sensitive feedback on the precautionary handling of their incoming messages, thus reducing the likelihood of misclassification. The primary benefit of their strategy is that it enables rejection possibilities, i.e., refusal to make a choice. By gathering extra data, the undecided instances must be re-examined. [10] also provided a Naive Bayes Classifier spam filter method for spam email. It worked by assessing the likelihood of distinct phrases appearing in lawful and spam emails and then classifying them based on those likelihoods.

[11] also debated technological alternatives to block spam mail. This technological solution consisted of an adaptive mixture of origin-based filters (OBF) and content-based filters (CBF). The CBF Filter included two parts of the classifier based on machine learning (MLC) and semantic resemblance to the classifier based on top (SSC). On the conventional dataset, this technological solution was tested on the normal dataset like Enron, Ling Spam, and Personal Email (PEM) posts. It was found that the general output of the OBF and CBF mixture exceeded the output of the person. [12] used machine learning algorithms to classify emails. They used Naïve Bayes and J48 Decision Tree, which have been evaluated for their effectiveness in spam or ham classification of messages. In conjunction with pre-processing methods and text categorization ideas, the experiment concentrated on classification.

[13] [14] also evaluated several spam information mining techniques to determine the best classifier for email sorting. When we integrated the feature selection strategy into the classification method, they evaluated that classifiers function well. Using the word count algorithm, they used the Bayesian Naïve classifier and extracted the phrases. They discovered the naive Bayesian classifier is more precise than the vector machine after computing. When the Bayesian Naïve Classifier was used, the error rate was very small. [15] attempted to improve the various spam filtering methods available by suggesting one' UBSF' method that could be very helpful. Their document studied the numerous spam-related issues and numerous methods and techniques that attempt to address them. [16] investigated the use of random woods for automatic email filtering in folders and spam email filtering. They also demonstrated that random forests are a great option for these assignments because it operates quickly on big and high-dimensional databases, it is simple to adjust and extremely precise, outperforming common algorithms such as decision trees, vector supporting computers and naive Bayes.

[17], on the basis of the Bayesian decision, the theory presented a performance evaluation of several term selection techniques for reducing the dimensionality of the spam filtering domain. They compared the output obtained by seven distinct Naive Bayes spam filters applied to classify emails from six well-known, true, public, and big email information sets, after a dimension reduction step employed by eight common term selection methods that varied the number of terms chosen. Their findings also checked the performance of Boolean characteristics better than those of the term frequency. [18] also evaluated some of the most common machine learning methods (Bayesian classification, k-NN, ANNs, SVMs, Artificial Immune System, and Rough Sets) and their applicability to the spam classification problem. Algorithm descriptions were provided, and their performance comparison was displayed on the Spam Assassin spam corpus.

III. MATERIALS AND METHODS

The proposed system is an Unwanted Email Filtering System. We used the Naïve Bayes classifier to create an Email classification system that classified unwanted emails into classes ‘Unwanted with very high degree’, ‘Unwanted with high degree’, and Ham.

A. The architecture of the Proposed System

This explains the suggested scheme, explaining how to integrate the modules and elements to bring about the suggested system's working implementation. The architecture of the proposed system is illustrated in figure 1.

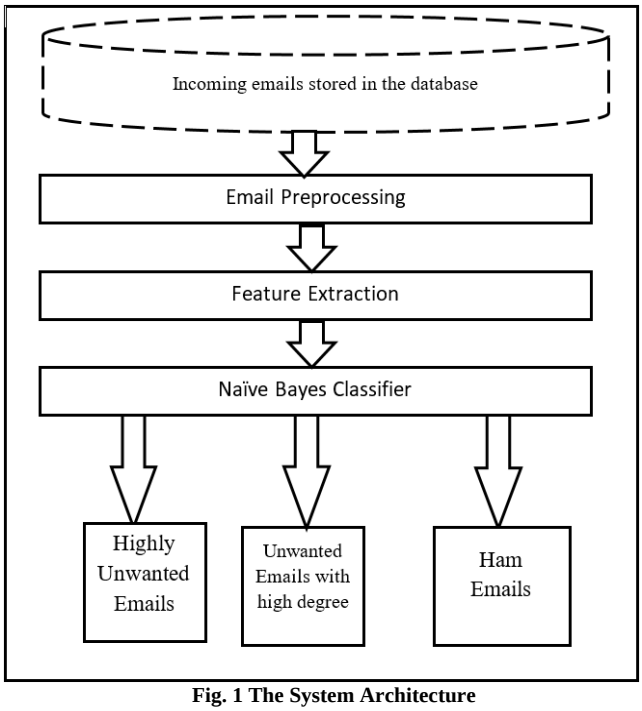

Fig. 1 The System Architecture

B. Algorithm of the Unwanted Email Classification System

Algorithm 1: Algorithm for improved email violent word filtering

Input:

Output:

1. begin
2. for i=w do
3. count how many of the unwanted words emails contain w
4. end for
5. Compute probability of unwanted email using
7. Create a set (w1,.....,wn)of the distinct words in the email
8. Compute $P(w1, \ldots wn|S)P(S) = \frac{P(w1, \ldots wn|S)P(S)}{P(w1, \ldots wn|S)P(S)+P(w1, \ldots wn|\dot{S})P(\dot{S})}$
9. if (P(S| w1,.....,wn) 0.5)
10. output " Unwanted Email"
11. else
12. output "ham"
13. end if
14. end

Fig. 2 Algorithm of the Unwanted Email Classification System

C. Detailed Algorithm steps

Step 1: The content of the email is received through our software and saved in a database table called the "EMAIL" table with fields for "Subject", "Body", "From," and "ID" as Primary keys. This table will hold all emails with their category.

Another Table was created called the "word frequency" table that has the fields "ID", "Word", and "Count" and "Category" fields. This table will hold all the words seen so far, along with their count and category.

Step 2: Feature Extraction

The body of the email is pre-processed by removal of duplicates, case folding, commonly used words and special characters through a process called "Tokenization" and an algorithm called Word Count Algorithm. During tokenization, the algorithm takes a paragraph of text as input and returns an array of keywords by removing characters that are not letters, converts them to lower case.

Step 3: Training the dataset with Naïve Bayesian Classifier Algorithm.

We train our classifier using the training dataset included in the LingSpam dataset, which is the source dataset used for this project.

The actual words extracted from the first training dataset are stored in the "word frequency" table, and it is given a count of 1. The "Count" field represents the frequency of the word in the email. The "Category" field represents the category of the email (If the training email tokenized is unwanted, the category of all the extracted words is stated as Unwanted). When a new email comes in, the algorithm checks if each extracted keyword is already present in the "word frequency" table. If already present, it updates Count = Count + 1.

The classifier will implement the following pseudocode;

```
If (P (Ham | bodyContent) > P (Unwanted | bodyContent))
{
Return 'Ham';
} else
{
Return 'Unwanted';
}
```

Step 4: Testing the Dataset

The next step is to test new email data with the trained Naïve Bayesian Classifier for the calculation of the probability of Unwanted and Ham mails and make a prediction of which value is higher. If unwanted words are greater than Ham words in a mail, then the mail is an Unwanted email; otherwise, it is a Ham email.

IV. IMPLEMENTATION AND RESULTS

The Naïve Bayes classifier embedded in Matlab's Classification learner app was used in the training and classification of sample datasets. The emails were classified into unwanted or ham by selecting message features and checking the word count of each word in the message to determine their number of occurrences. The classification system had a prediction speed of 36sec, training of 26.651 sec, and accuracy of 67%.

A. Training results

Fig. 3. and Fig. 4. represent the training result of the Unwanted Email filter using the Naïve Bayes Algorithm. X and Y-axis consist of the message body. Unwanted and Ham are specified in dotted lines with three colors. Blue color represents labels, orange color represents not spam, and yellow color represents unwanted messages. Ham messages are "I'm going back to try", and "I'm busy", spam messages are "To claim your prize", "your email address is selected to claim the sum of $500,000.00 in the 2014 European lottery.

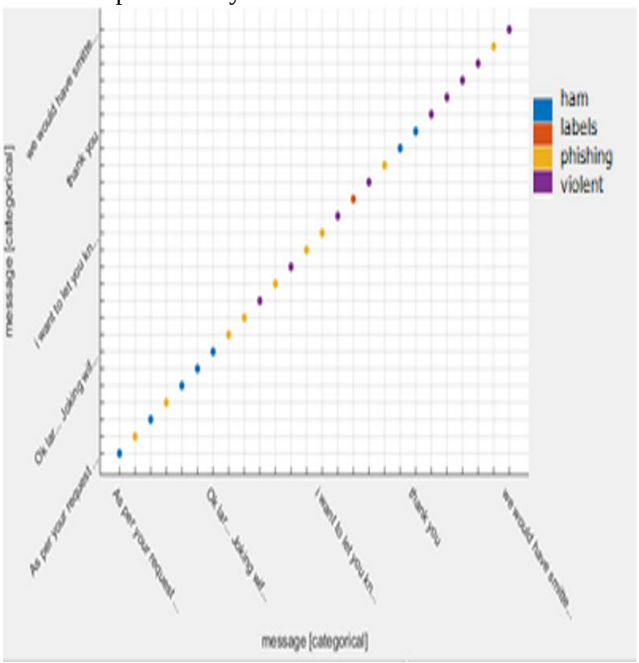

Fig. 3 Training results of the unwanted email filter with Naïve Bayes Algorithm

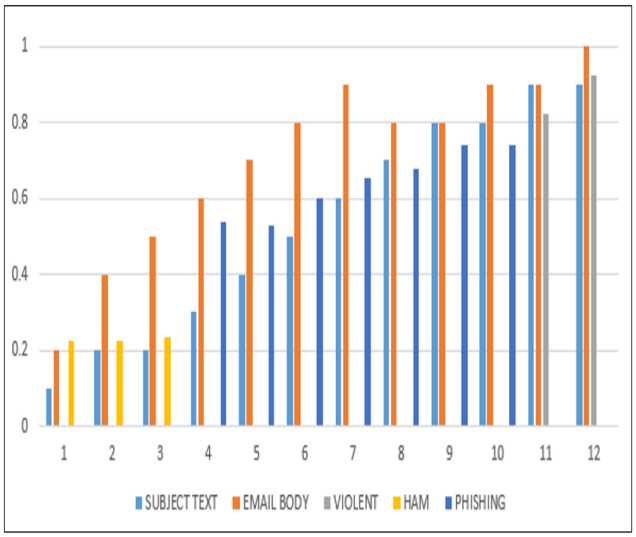

Fig. 4 Another view of classified emails

B. Testing results

To assess the efficiency of the model, an interface was designed to display the results of the classification. The Interface consists of the "Email address" text field, "Subject" text field, "Main body" text area, and "Submit" button. The results are shown in a message box showing if messages are unwanted words or ham words together with the degree (very high, high, low, very low) as shown in Fig. 5., Fig. 6., and Fig. 7. If the Email address text field, Subject text field, or Main body text area is empty, the user will be prompted to type in an Email address, Subject text, or main body; else system will prompt the user with the message “fill the out this field”. When the user clicks the submit button, the system searches the database contents and displays results of the classification and the degree; otherwise, the system will display “This is a correct email”.

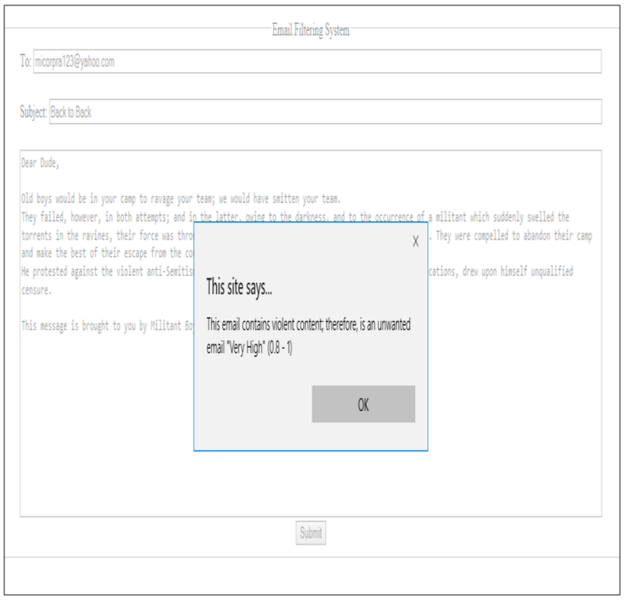

Fig. 5 An Email classified as an unwanted email with a Very High Degree

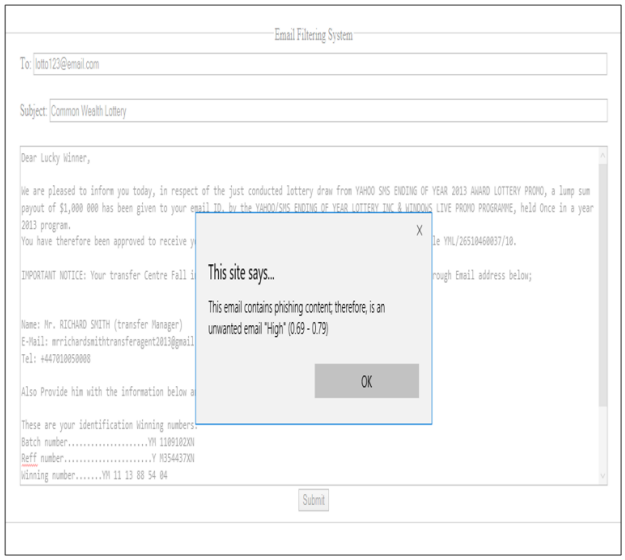

Fig. 6 An Email classified as unwanted with a High Degree

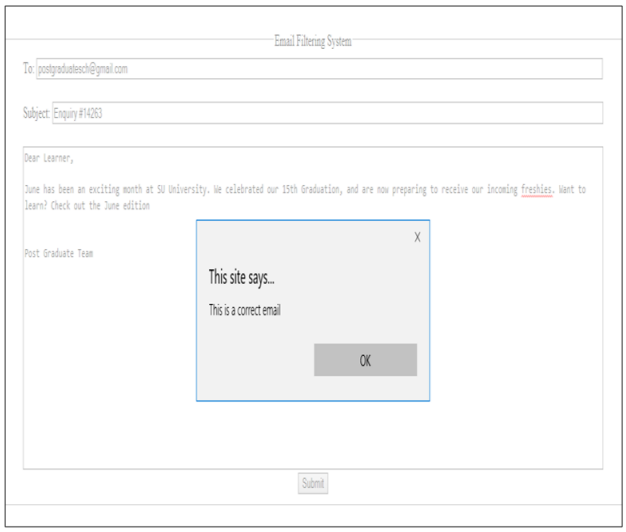

Fig. 7 An Email classified as Ham

V. CONCLUSION

In this study, the objective was to enhance the classification of incoming e-mails using-Naïve Bayes classifier-into unwanted and ham (legitimate) based on features in both the Subject text of the email and Email body. The system was trained in Classification learner; corpus was gotten online. In preprocessing, the input email body was segmented into tokens, and its structure was analyzed. Message blocks that are likely to contain typical unwanted words were marked. Feature Extraction stage implemented word count algorithm which provides a flexible result. After preprocessing the dataset to remove the stopwords and non-words, the system counted the total number of unique words out of the total word and found the frequency of that word in a particular document, and then compared with already learned unwanted words in the database. If the probability of the summation word count was greater than or equal to 0.5, $P(\sum WC) \geqslant 0.5$, then unwanted email exists, and users are notified. This enabled users to understand and block unwanted emails. Some emails were classified as Ham. This means that users can

view such messages as legitimate messages. The model performed well with 89% accuracy. The proposed approach will work only for e-mails having Subject text and E-mail body as plain text. But today, spammers also include multimedia content and HTML links in e-mails sent to users. Our future work aims at detecting and filtering emails with such content.

REFERENCES

[1] W. L. Sushma, D. Shailaja, D. Ganesh and B. Bipin Shinde. Overview of Anti-Spam Filtering Techniques. International Research Journal of Engineering and Technology (IRJET), 04(01) (2017). p-ISSN: 2395-0072.

[2] S. Geerthik and T. P. Anish. Filtering Spam: Current Trends and Techniques. International Journal of Mechatronics, Electrical and Computer Technology,3(8) (2013) 208–223.

[3] P. Pantel and D. Lin. SpamCop: A Spam Classification & Organization Program. In Proceedings of Workshop for Text Categorization, AAAI-98 (1998) 95–98.

[4] I. Koprinska, J. Poon, J. Clark and J. Chan., Learning to Classify E-mail. Information Sciences 177 (2007) 2167–2187.

[5] M. Sahami, S. Dumais, D. Heckerman and Horvitz, E. A Bayesian Approach to Filtering Junk E-mail, In Proceedings AAAI Workshop on Learning for Text Categorization (1998).

[6] W. A. Awad and S. M. ELseuofi Machine Learning Methods for Spam Email Classification. Proceedings of the International Journal of Computer Science & Information Technology (IJCSIT), 3(1) (2011) 273-284.

[7] I. Ismaila, S. Ali, N. ThanhNguyen, S. O. Omatu, and M. P. KamilKuca. A Combined Negative Selection Algorithm–Particle Swarm Optimization for an Email Spam Detection System. Engineering Applications of Artificial Intelligence 39 (2015) 33-44.

[8] V. Christina, S. Karpagavalli and G. Suganya. Email Spam Filtering Using Supervised Machine Learning Techniques. (IJCSE) International Journal on Computer Science and Engineering. 2(9) (2010) 3126-3129.

[9] N. Andrew and D. Jeff. Building High-level Features Using Large-Scale Unsupervised Learning. Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK, 1-13 (2013).

[10] C. Wu. Behavior-Based Spam Detection Using a Hybrid Method of Rule-Based Techniques and Neural Networks. Expert Systems with Applications 36 (2009) 4321–4330.

[11] M. N. Marsono, M. W El-Kharashi, Fayez Gebali., Targeting Spam control on middleboxes: Spam detection based on layer-3 E-mail content classification. Elsevier Computer Networks 53 (2009) 835– 848.

[12] Z. Bing, Y. Yiyu and J. Luo. A Three-Way Decision Approach to Email Spam Filtering. Canadian AI 2010, LNAI 6085 (2010) 28–39.

[13] S. Roy, A. Patra, S. Sau, K. Mandal, S. Kunar An Efficient Spam Filtering Techniques for Email. American Journal of Engineering Research (AJER) e-ISSN: 2320-0847 p-ISSN: 2320-0936, 2(10) (2013) 63-73.

[14] M. N. Marsono, El-Kharashi, M. W. and F. Gebali Targeting Spam Control on Middleboxes: Spam Detection Based on Layer-3 E-mail Content Classification. Elsevier Computer Networks 53 (2009) 835–848.

[15] S.S. Shinde, and Patil, P. R. Improving Spam Mail Filtering Using with Discretization Filter. International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS) (2014) 82–87.0

[16] N. Mirza, Mirza, T. and Auti, B. R. Evaluating Efficiency of Classifier for Email Spam Detector Using Hybrid Feature Selection Approaches. IEEE, International Conference on Intelligent Computing and Control Systems ICICCS, 978-1-5386-2745 (2017).

[17] A. Almeida, J. Almeida and A. Yamakami Spam Filtering: how the Dimensionality Reduction Affects the Accuracy of Naive Bayes Classifiers. Journal of Internet Services and Applications, Springer London, 1 (2011) 183–200.

[18] P. Revar, S. Arpita, P. Jitali and K. Pimal. A Review on Different Types of Spam Filtering Techniques. International Journal of Advanced Research in Computer Science, 8 (5) (2017) 2720-2723.

[19] S. S. Shinde, and P.R. Patil. Improving Spam Mail Filtering Using with Discretization Filter. International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS) (2014) 82–87.

[20] V. O. Oyekunle, P. O. Asagba, F. Egbono. Detection of Violent E-mails Using Fuzzy Logic. International Journal of Computer Trends and Technology, 69(3) (2021) 79-84.

[21] V. O. Oyekunle, M. Nwanyanwu, M. A. Ide., Efficient Method of Mining Sequential pattern in the retail database. Journal of Scientific and Engineering Research, 8(5) (2021) 65-74.

[22] Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, High-resolution fiber distributed measurements with coherent OFDR, in Proc. ECOC'00, 11.3.4 (2000) 109.

[23] Surendiran,R., and Alagarsamy,K., 2013. "Privacy Conserved Access Control Enforcement in MCC Network with Multilayer Encryption". International Journal of Engineering Trends and Technology (IJETT), 4(5), pp.2217-2224.