2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

Spam Detection Protocol using Probabilistic Eshield Protocol

Masrath Parveen[1], Dr. Saurabh Pal[2], Dr. Venkateswara Rao CH[3]

[1]Research Scholar, Dept of CSE, V.B.S.Purvanchal University, Jaunpur
[2] Department of CSE, V.B.S.Purvanchal University, Jaunpur
[3]Department of CSE, Siddhartha Institute of Engineering and Technology, Hyderabad

Abstract – Due to its ongoing use of servers' and receivers' resources, email spam is becoming more and more popular. In this paper, we introduce ESHIELD, a unique email spam filtering protocol that employs big data analytics to protect against spam emails. In order to lighten the load on the server, the protocol is carried out at the receiver end for precise and quick filtering. ESHIELD's primary objective is to identify spammers and anyone else who sends spam emails. Using parameters like false positive rate, false negative rate, detection accuracy, and detection time, the performance of the ESHIELD implementation, which makes use of the Map Reduce feature of the Hadoop framework, has been assessed. By building probabilistic models on the suspected email, ESHIELD speeds up the spam detection process. Additionally, it applies similarity tests to the emails to reliably identify spam with a minimal amount of false positives and negatives.

Keywords: Feature, Cluster, Map reduce, Features, Spam Email, Similarity, Probabilistic

1. Introduction

One of the most common kind of computer-mediated communication is email [1]. Since sending an email is essentially free [2], email has taken the place of other forms of contact for millions of individuals. Spam, suspicious, fraudulent, and informal are some types of unwanted emails that can be identified [3]. Spam in email is the most well-known type, and it differs from spam in social networking sites, Usenet newsgroups, voice over Internet Protocol, mobile messaging, and instant messaging [4]. Large-scale email service providers are severely impacted, and they are unable to provide support for the actual users.

Email users use the Simple Mail Transfer Protocol (SMTP) protocol, which is described in RFC 821. Emails are typically transferred using the Multipurpose Internet Mail Extension (MIME), which allows anyone to style them along the way. Criminals may utilise email as a means of contact. Spam emails are illegal or unsolicited commercial emails that are frequently sent in mass and have a variety of purposes. Email spam detection is particularly important in a personalised as well as social environment because spam emails cost the online community billions of dollars annually.

Over 50% of all contemporary email is spam, according to studies, and 29% of users claim that because of spam, they use email less [5]. According to a recent survey, 70% of email traffic was spam [6]. Email spam consumes up computer resources such storage space, processing power, and network bandwidth, as well as causing traffic to be misused and other legal issues. While network administrators are always improving their technologies to keep email spam away from their subscribers, spammers are constantly coming up with new ways to get through filters. The researchers have so far developed a number of strategies against email spam [4], but these strategies fall short of offering a comprehensive answer. By making spammers who send spam emails subject to punishment, an anti-spam law [7] was put into effect.

By making spammers who send spam emails subject to punishment, an anti-spam law [8] was put into effect. Spam emails are continuously on the rise despite the CAN-SPAM Act [9], a federal regulation that was implemented in 2013. Current anti-spam measures are static, making it simple to get around them by slightly altering the content. Spammers simply research the most recent anti-spam techniques and figure out how to get around them to accomplish this. Spam must be effectively combated with an adaptable new strategy.

In this study, we describe a spam email detection system that more effectively filters out spam emails by performing user authentication, feature extraction, classification, and similarity detection utilising Big Data analytic techniques [10]. The recipient's message queue is first filtered by the user authorisation process to remove emails from ineligible users; the feature extraction process then utilises Map Reduce tools [11] to group the feature words of emails from valid users into three groups. A similarity detector that uses Map Reduce tools to determine the maximum feature similarity between the spam email being detected and any email available in the mail box ultimately results in the permanent removal of spam emails from the mail box. The classification process calculates the feature, cluster, and email probabilities for each email where the spam and legitimate emails are detected.

The dynamic spam email detection system that is being proposed makes it impossible for spammers to get around it since any alterations they make to the email content can be immediately detected. Big data analytics are being used to increase ESHIELD's effectiveness over time.

2. Literature review

Email, often known as electronic mail, is frequently used abusively. Spam emails are a strain on mailing systems because they waste millions of subscribers' priceless time and resources. 71.9% of email traffic is spam, according to the Symantec Intelligence Review (Symantec Intelligence Report, 2013) [5]. Two main challenges are involved in spam email detection: the first significant work is to identify the likely sender of an incoming email, and the second important goal is to identify spam emails from a collection of received emails. To finish these tasks, numerous studies have been done. This section analyses the current spam email detection techniques, and based on that study, we suggest a desirable and helpful solution to the problems with spam email detection listed above.

In order to identify the likely author in cyberspace, reference [1] has created a feature selection model based on genetic algorithms that leverages writeprint features. The writeprint has many characteristics, comparable to a fingerprint, including keyword sequence, grammatical and spelling errors, composition and layout, sentence length, and language richness. These writeprint traits are utilised to determine the writing style of the author, which aids in determining who the likely author is. Reference [12] has proposed a data mining technique to identify the likely author of each email sent by analysing the sender's individual write-print. For the identification of the likely author of an email, a cluster-based classification model (CEAI) [13] uses stylometric data with an enlarged feature set.

Reference [2] has created a framework for authorship identification of email messages that extracts four different sorts of data and then constructs classification models based on these aspects. New methods have been developed [14] to locate botnets, which are networks of spammers involved in the distribution of spam emails. Additionally, the work does email grouping based on fingerprints.

These currently used methods for authorship identification combined larger feature sets with stylometric traits to get promising outcomes. However, because to a higher rate of false positives and false negatives, they were unable to attain high detection accuracy. Furthermore, in order to achieve high detection accuracy, the stylometric features in the current approaches are insufficient. Because there are an infinite number of email senders, a scalable model is necessary to obtain very good accuracy.

For the purpose of categorising email contents into several folders, a classification model based on NGram was described in [2]. A spam-based classification approach for email spam detection was suggested in reference [15]. Reference [6] has offered a comparative analysis of several learning-based techniques, including filters based on the bag-of-words feature, hybrid filters, language-based filters, and filters based on non-content features.

Several machine learning techniques, including the Bayesian classifier, KNN, SVM, and ANN, have been used for email spam filtering, and comparisons of their performance have been made. A supervised machine learning-based approach for successfully identifying spammers has been given in reference [16].

Reference [17] conducted a comparison of several classification techniques, including Naive Bayes, SVM, J48, and neural networks for the detection of spam emails, and they came to the conclusion that J48 is the most effective classification approach. These machine learning techniques need a set of emails that have already been pre-classified, commonly referred to as training samples, in order to learn the classification rules from those emails. Due to their extensive maintenance requirements, these approaches are unable to reach very high precision.

For the purpose of detecting spam emails, Reference [18] suggested using word hashing rather than just content-based words and demonstrated how effective their method is. Reference [19] suggested SOAP, which takes advantage of social relationships between email correspondents to allow them to automatically and adaptably identify spam emails.

An ontology-based spam filtering approach based on the J48 classifier was proposed in [20]. Spambayes, a keyword-based method that detects spam email using Bayes' theorem, was utilised by Microsoft Outlook. [12] suggested a router-level spam filtering method employing thin operating system signatures. suggested spam detection method [13] based on features at the network level. [22] suggested employing neural networks to create a rule-based spam detection system.

For the purpose of identifying spam content on the web, [21] presented a structural, content similarity metric. [22] provided a way for detecting web spam. Web spam was identified using email spam detection methods and a Webb Spam Corpus data collection, according to [23]. Spam emails were categorised using the Enron corpus dataset [24] and the Enron and SRI Corpora datasets [25].

3. Proposed model to perform defense

The two main objectives of this paper is ESHIELD are: It first determines whether the email being received is coming from an authorised user. By examining the email's content, it determines if an email is spam or not. The first task is completed while the user authorization process is in progress, and the second task is completed while the feature extraction, classification, and similarity detection processes are all in progress.

The ESHIELD user authorisation procedure is shown in Fig. 1. When a sender delivers an encrypted message to any number of receivers, the user authorisation procedure begins. The HMAC value of certain private information, including the sender's email address EA_O (32 bits), the email message E (variable length), the sender's private key KS_O (32 bits), the sending time of the email T_{send} (16 bits), a nonce value N (32 bits) generated by the sender, and the recipient's email address EA_R, make up the encrypted message. The recipient's public key is used to encrypt the message.

$$E_{KUR}\,(HMA(EA_O,\, E,\, KS_O,\, T_{send}\,,\, N,\, EA_R)) \qquad (1)$$

T_{send} is used to determine how recent a message is, while N is utilised to thwart replay attempts. When T_{send} and KS_O are XORed with a 32-bit random integer that increases by 1 each time a message is sent, the result is the nonce value. Thus, the receiver takes the encrypted message out of the message queue and uses its private key to decrypt it.

$$D_{KSR}\,(E_{KUR}\,(HMA(EA_O,\, E,\, KS_O,\, T_{send}\,,\, N,\, E_{AR}))) \qquad (2)$$

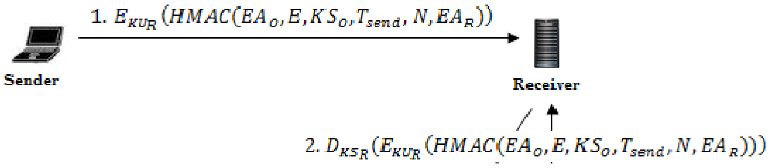
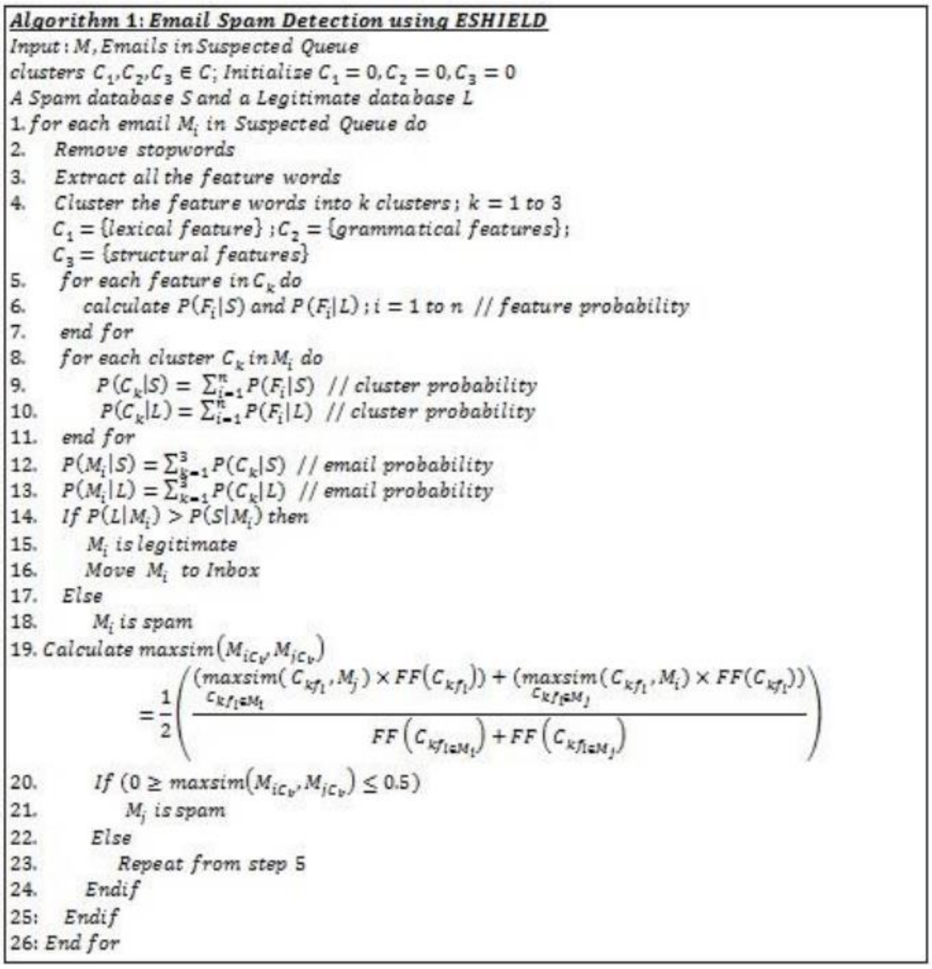

1. $E_{KU_R}(HMAC(EA_O, E, KS_O, T_{send}, N, EA_R))$
Sender
Receiver
2. $D_{KS_R}(E_{KU_R}(HMAC(EA_O, E, KS_O, T_{send}, N, EA_R)))$

Fig.1.Authorization process of an user [1]

Now that emails from authorised users are in the suspected queue and being further examined to see if they are real or spam, they are referred to as suspected emails. The message must be classified as spam or not before being stored in the recipient's mailbox. As a result, only the valid emails (i.e., suspicious emails) are subjected to the feature extraction procedure for further processing during the user authorization process. The algorithm for ESHIELD's use in email spam detection is shown in Figure 2. The feature extraction method, which starts by removing the stop words from the suspicious emails in the suspected queue and subsequently extracts all feature words, must be applied.

For feature extraction, ESHIELD uses the emails present in the suspicious queue as input. Each email's feature words are divided into three clusters by the ESHIELD using an enhanced feature set. The first cluster is under the heading of lexical feature and includes the quantity of keywords, characters per keyword, common words, special words, and negative words per sentence, line, link, and hyperlink. The number of verbs, adverbs, nouns, adjectives, tense, conjunctions, interjections, prepositions, articles, and gerunds make up the second cluster, which is categorised as a grammatical feature. The Stanford CoreNLP system [25] supports these capabilities.

The author identification information, including the author's email address, message type, email size, email type, and the time the email was sent, is included in the third cluster, which is categorised as a structural characteristic. Because it decreases the total detection time and enhances the functionality of ESHIELD, the email is divided into three groups based on their attributes.



```
Algorithm 1: Email Spam Detection using ESHIELD
Input: $M$, Emails in Suspected Queue
clusters $C_1, C_2, C_3 \in C$; Initialize $C_1 = 0, C_2 = 0, C_3 = 0$
A Spam database $S$ and a Legitimate database $L$
1. for each email $M_i$ in Suspected Queue do
2.     Remove stopwords
3.     Extract all the feature words
4.     Cluster the feature words into k clusters; $k = 1$ to 3
       $C_1 = \{lexical\ feature\}$; $C_2 = \{grammatical\ features\}$;
       $C_3 = \{structural\ features\}$
5.     for each feature in $C_k$ do
6.         calculate $P(F_i|S)$ and $P(F_i|L)$; $i = 1$ to $n$ // feature probability
7.     end for
8.     for each cluster $C_k$ in $M_i$ do
9.         $P(C_k|S) = \sum_{i=1}^{n} P(F_i|S)$ // cluster probability
10.        $P(C_k|L) = \sum_{i=1}^{n} P(F_i|L)$ // cluster probability
11.    end for
12.    $P(M_i|S) = \sum_{k=1}^{3} P(C_k|S)$ // email probability
13.    $P(M_i|L) = \sum_{k=1}^{3} P(C_k|L)$ // email probability
14.    If $P(L|M_i) > P(S|M_i)$ then
15.        $M_i$ is legitimate
16.        Move $M_i$ to Inbox
17.    Else
18.        $M_i$ is spam
19. Calculate $maxsim(M_{iC_v}, M_{jC_v})$
       $= \frac{1}{2}\left(\frac{(\underset{C_{kf_1 \in M_i}}{maxsim}(C_{kf_1}, M_j) \times FF(C_{kf_1})) + (\underset{C_{kf_1 \in M_j}}{maxsim}(C_{kf_1}, M_i) \times FF(C_{kf_1}))}{FF(C_{kf_1 \in M_i}) + FF(C_{kf_1 \in M_j})}\right)$
20.        If $(0 \geq maxsim(M_{iC_v}, M_{jC_v}) \leq 0.5)$
21.            $M_j$ is spam
22.        Else
23.            Repeat from step 5
24.        Endif
25:    Endif
26: End for
```

Fig.2.Proposed Spam detection protocol model

As a result, the feature extraction procedure is carried out by the receiver, and the feature set is used as input in the classification process to determine if the email is spam or not. ESHIELD determines if a feature will be obtained as valid or spam within each cluster. ESHIELD computes the feature probabilities utilising the Bayesian Classifier approach. In other words, using the data gathered from the prior probabilities, we compute the posterior probabilities. The choice of the Bayesian concept is made because it decreases the quantity of false positives and false negatives, increasing the detection accuracy.

To determine if a cluster is real or spam, the feature probabilities of all the characteristics in each cluster are now added together. In order to determine if an email is real or spam, the cluster probabilities of each cluster are pooled at the end. When the likelihood of receiving an email as legitimate exceeds the likelihood of receiving it as spam, the email is classified as legitimate; however, when the likelihood of receiving an email as legitimate exceeds the likelihood of receiving it as spam, the email is classified as spam.

The absolute detection of spam mail requires M_j to go through the classification process if the highest similarity between M_i and M_j exceeds 0.5. As a result, similar spam emails are blocked at the recipient as a result of the similarity test. When comparable spam emails are blocked, the spam email's author is also blocked.

4. Experimental Results

The suggested ESHIELD protocol has been implemented and evaluated using the Java programming model. 1.4 million emails were included in the dataset that was used for the experiments.

ESHIELD is implemented using the Map Reduce feature of the Hadoop framework, and the results are presented using it. We extract all the necessary features using the Map class, which implements the map method, and then cluster the data as necessary. The reduction function is implemented by the reduction class, which also calculates the likelihood of an email. In terms of False Positive Rate (FPR), False Negative Rate (FNR), Detection Accuracy, and Detection Time, we assess ESHIELD's performance. Furthermore, we demonstrate that our suggested method surpasses the existing methods by contrasting its performance with that of MLP [3] and FEDM [25].

Out of all legitimate emails received at the recipient, the FPR is the proportion of legitimate emails that are incorrectly labelled as spam.

TABLE I displays the number of false positives created for 10,000, 20,000, 30,000, 40,000, and 50,000 emails, which are split equally between valid and spam emails.

TABLE I: False Positive Rates created

No.of. Emails	Number of False Positives			False Positive Rate		
	ESHIELD	MLP	FEDM	ESHIELD	MLP	FEDM
10000	800	2025	2547	0.16	0.47	0.54
20000	2349	4367	5774	0.21	0.49	0.59
30000	4564	7232	9089	0.26	0.52	0.64
40000	6235	10056	13021	0.31	0.56	0.68
50000	8897	13598	17556	0.37	0.58	0.77

It has been discovered that as the number of emails received increases, the number of false positives increases and decreases with the number of emails received. The FPR rises as the number of false positives rises, and vice versa. Figure 3(a) compares the FPR for ESHIELD with that of MLP and FEDM. According to the data, when the number of incoming emails rises from 10,000 to 50,000, ESHIELD generates an FPR of 0.16 and 0.37, whereas MLP generates an FPR of 0.47 and 0.58 and FEDM generates an FPR of 0.54 and 0.77. It has been found that ESHIELD performs better than the other two methods currently in use and produces fewer false positives.

TABLE II displays the number of false negatives created for 10,000, 20,000, 30,000, 40,000, and 50,000 emails, which are split equally between valid and spam emails.

TABLE II: False Negative Rates created

No.of. Emails	Number of False Negatives			False Negative Rate		
	ESHIELD	MLP	FEDM	ESHIELD	MLP	FEDM
10000	175	254	424	0.032	0.052	0.081
20000	520	732	1531	0.053	0.073	0.153
30000	1056	1545	3067	0.065	0.108	0.206
40000	1668	3151	5178	0.087	0.156	0.243
50000	2872	5099	8867	0.119	0.207	0.377

Figure 3(b) compares the FNR for ESHIELD with that of MLP and FEDM. According to the results, when the number of incoming emails rises from 10,000 to 50,000, ESHIELD generates a FNR of 0.032 and 0.119, whereas MLP generates a FNR of 0.052 and 0.207 and FEDM generates a FNR of 0.081 and 0.377. It has been found that ESHIELD performs better than the other two methods currently in use and produces fewer false negatives. This is as a result of ESHIELD's similarity detection process's accuracy in identifying spam emails and its decreased incidence of false negatives.

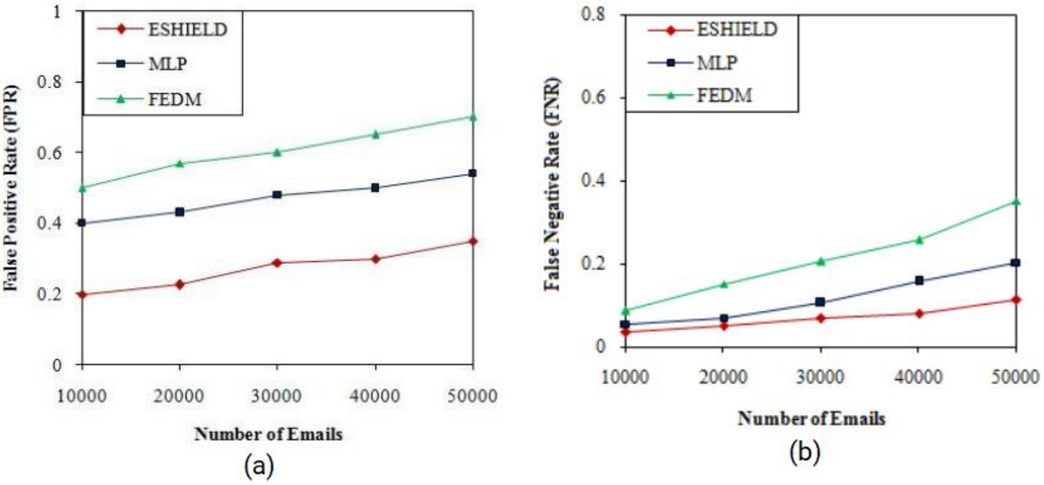


Fig.3. (a) Comparison of False Positive Rate (b) Comparison of False Negative Rate

The findings indicate that ESHIELD achieves an average detection accuracy of 0.98, while MLP achieves an average detection accuracy of 0.885, and FEDM achieves an average detection accuracy of 0.882. ESHIELD works better than the other two methods currently in use and produces fewer false negatives. This is due to ESHIELD's similarity detection process's accuracy in identifying spam emails and its decreased occurrence of false positives and false negatives.

5. Conclusion

As a result, the unique email spam filtering protocol ESHIELD has been introduced. Big data analytics are used by the ESHIELD, which is deployed at the receiver's side, to protect against spam emails. ESHIELD finds both the spammers and the uninvited senders of spam emails. The performance of ESHIELD has been assessed using metrics like false positive rate, false negative rate, detection accuracy, and detection time. ESHIELD has been implemented using the Map Reduce paradigm. The performance findings unambiguously show that ESHIELD improves detection effectiveness by producing fewer false positives and false negatives. Additionally, it quickly and reliably detects spam emails.

References

[1] AlMadahkah AM, “Big Data in computer Cyber Security Systems”, International Journal of Computer Science and Network Security, Vol.16, No.4, 2016,pp: 56-65

[2] Mohammed Ali Shaik, Praveen Pappula, T Sampath Kumar, "Predicting Hypothyroid Disease using Ensemble Models through Machine Learning Approach", European Journal of Molecular & Clinical Medicine, 2022, Volume 9, Issue 7, Pages 6738-6745. https://ejmcm.com/article_21010.html

[3] Alsmadi I, Alhami I, “Clustering and classification of email contents”, Computer and Information Sciences, Journal of King Saud University, Vol.27, 2015, pp: 46-57.

[4] M. A. Shaik, S. k. Koppula, M. Rafiuddin and B. S. Preethi, (2022), "COVID-19 Detector Using Deep Learning”, International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 443-449, doi: 10.1109/ICAAIC53929.2022.9792694.

[5] A. Subasi, S. Alzahrani, A. Aljuhani and M. Aljedani, "Comparison of Decision Tree Algorithms for Spam E-mail Filtering," 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2018, pp. 1-5, doi: 10.1109/CAIS.2018.8442016.

[6] Mohammed Ali Shaik and Dhanraj Verma, (2022), "Prediction of Heart Disease using Swarm Intelligence based Machine Learning Algorithms", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020025-1–020025-9; https://doi.org/10.1063/5.0081719, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020025-1 to 020025-9.

[7] M. A. Shaik and Dhanraj Verma, (2022), "Predicting Present Day Mobile Phone Sales using Time Series based Hybrid Prediction Model", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020073-1–020073-9; https://doi.org/10.1063/5.0081722, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020073-1 to 020073-9

[8] in Srinidhi, Jia Yan & Giri Kumar Tayi 2015, ‘Allocation of resources to cyber-security: The effect of misalignment of interest between managers and investors’, Decision Support Systems, July 2015, http://dx.doi.org/10.1016/j.dss.2015.04.011, vol. 75, pp. 49-62.

[9] Mohammed Ali Shaik, MD.Riyaz Ahmed, M. Sai Ram and G. Ranadheer Reddy, (2022), "Imposing Security in the Video Surveillance", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020012-1–020012-8; https://doi.org/10.1063/5.0081720, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020012-1 to 020012-8.

[10] M. A. Shaik, Geetha Manoharan, B Prashanth, NuneAkhil, Anumandla Akash and Thudi Raja Shekhar Reddy, (2022), "Prediction of Crop Yield using Machine Learning", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020072-1–020072-8; https://doi.org/10.1063/5.0081726, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020072-1 to 020072-8.

[11] Mohammed Ali Shaik, Dhanraj Verma, (2021), Agent-MB-DivClues: Multi Agent Mean based Divisive Clustering, Ilkogretim Online - Elementary Education, Vol 20(5), pp. 5597-5603, doi:10.17051/ilkonline.2021.05.629

[12] Byrnea, DJ, David Morganb, Kymie Tana, Bryan Johnsona & Chris Dorrosa 2014, ‘Cyber Defense of Space-Based Assets: Verifying and Validating Defensive Designs and Implementations’, Conference on Systems Engineering Research (CSER 2014), Science Direct, vol. 28, pp. 522-530.

[13] Mohammed Ali Shaik and Dhanraj Verma, (2020), Enhanced ANN training model to smooth and time series forecast, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022038, doi.org/10.1088/1757-899X/981/2/022038

[14] M. A. Shaik, Dhanraj Verma, P Praveen, K Ranganath and Bonthala Prabhanjan Yadav, (2020), RNN based prediction of spatiotemporal data mining, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022027, doi.org/10.1088/1757-899X/981/2/022027

[15] Ashish Malviya, Glenn A Fink, Landon Sego & Barbara Endicott-Popovsky 2011, ‘Situational Awareness as a Measure of Performance in Cyber Security Collaborative Work’, Eighth International Conference on Information Technology: New Generations, pp. 937-942.

[16] Mohammed Ali Shaik and Dhanraj Verma, (2020), Deep learning time series to forecast COVID-19 active cases in INDIA: A comparative study, 2020 IOP Conf. Ser.:Mater.Sci.Eng. 981 022041, doi.org/10.1088/1757-899X/981/2/022041

[17] Mohammed Ali Shaik, “Time Series Forecasting using Vector quantization”, International Journal of Advanced Science and Technology (IJAST), ISSN:2005-4238,Volume-29,Issue-4 (2020), Pp.169-175.

[18] Arunabha Mukhopadhyay, Samir Chatterjee, Debashis Saha, Ambuj Mahanti & Samir K Sadhukhan 2013, ‘Cyber-risk decision models: To insure IT or not?’, Decision Support Systems, http://dx.doi.org/ 10.1016/j.dss.2013.04.004, Volume 56, December 2013, pp. 11-26.

[19] Mohammed Ali Shaik, “A Survey on Text Classification methods through Machine Learning Methods”, International Journalof Control and Automation (IJCA), ISSN:2005-4297,Volume-12,Issue-6 (2019), Pp.390-396.

[20] Andreas GK Janecek, Wilfried N Gansterer & Ashwin Kumar, K 2008, ‘Multi-Level Reputation-Based Greylisting’, in proc. of Third International Conference on Availability, Reliability and Security ARES 08, 4-7 March 2008, Barcelona, Spain.

[21] Mohammed Ali Shaik, P. Praveen, T. Sampath Kumar, "Integration and application of Fog, IoT and Edge Computing", Fog Computing: Concepts, Frameworks, and Applications (FCCFA) Aug-2022, CRC Press, ISBN: 9781003188230.

[22] Praveen, P, Mohammed Ali Shaik, T. Sampath Kumar, Choudhury T, "Smart Farming: Securing Farmers Using Block Chain Technology and IOT", Aug-2021, EAI/Springer Innovations in Communication and Computing, ISBN: 978-3-030-65690-4

[23] M. A. Shaik, "Protecting Agents from Malicious Hosts using Trusted Platform Modules (TPM)," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp. 559-564, doi: 10.1109/ICICCT.2018.8473278.

[24] Amani Mobarak & AlMadahkah 2016, 'Big Data In computer Cyber Security Systems', IJCSNS International Journal of Computer Science and Network Security, vol. 16, no. 4, pp. 56-65.

[25] Aakash Atul Alurkar; Sourabh Bharat Ranade; Shreeya Vijay Joshi; Siddhesh Sanjay Ranade, Piyush A Sonewar, Parikshit N Mahalle & Arvind V Deshpande 2017, 'A proposed data science approach for email spam classification using machine learning techniques', Internet of Things Business Models, Users, and Networks, pp. 1-5.