Journal of Information Assurance and Security.
ISSN 1554-1010 Volume 16 (2021) pp. 117-124


# Spear-Phishing Emails Verification Method based on Verifiable Secret Sharing Scheme

**Gunikhan Sonowal[1], Aditi Sharma[2] and Latika Kharb[3]**

[1]Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation,
Vaddeswaram-522502, Guntur, Andhra Pradesh, India
*gunikhan.sonowal@gmail.com*

[2]Department of Computer Science and Engineering
Quantum University, Roorkee-247667, India
*aditi11121986@gmail.com*

[3]Department of Computer Science and Engineering
Jagan Institute of Management Studies (JIMS), Delhi-110085, India
*latika.kharb@jimsindia.org*

***Abstract***: **Phishing is a critical cybersecurity issue that differs from other attacks. This attack practices social engineering techniques to prompt users to disclose their credentials. Spear phishing is an advanced version of phishing attacks where the attacker investigates the online behavior of an individual or organization to gather the information for constructing an email that appears to be legitimate. As a result, spear-phishing holds a high success rate than traditional phishing emails since these emails can evade the standard security barriers and harvest the credentials. This paper presents an abstract method that collects features from different dimensions: phishing domain features, stylometric features, and others to detect spear-phishing emails. The auto-upgrade profile is additionally supplemented by the method to detect phishing emails within a second. Finally, the method employs a machine-learning algorithm to classify spear-phishing emails from legitimate emails. This paper owns the uniqueness of detecting traditional phishing emails as well as spear-phishing emails using multi-dimensional features. Finally, this paper applied the publicly verifiable secret sharing to verify the email whether the sender is genuine or not.**

***Keywords***: Spear Phishing, phishing, Social engineering, Machine learning algorithm, author's writing-style, cyber-security

## I. Introduction

In the digital world, email is selected as a primary communication channel for many organizations or individuals. It assists persons to share a file, information, link, or able to send multiple persons at a single time. It may have many benefits using emails but phishing email is another form of emails that are controlled by attackers for stealing sensitive information from victims. Phishing is the art of creating a fake website that appears to be a genuine website by collecting a logo, signature, or fonts [33]. Generally, the phishing website contains a login field for driving users to insert their credentials [24]. Once they create the fake website, then the link to the website is sent through phishing emails.

Phishing emails use social engineering techniques to prompt users to click the link to the website. Many users are unaware of phishing emails, accordingly, they click the link that is redirected to the phishing website for disclosing their credentials [4]. According to the Anti-phishing working group reports on February 2021, the number of unique phishing email's subject reports was 133,038 in December 2020 [39].

Recently, spear phishing is a new generation of phishing attacks where the attacker targets particularly a smaller target victim [8, 9]. The important distinction between spear-phishing and traditional phishing is merely the extent of using social engineering strategies with the goal that the victims are unable to think twice before revealing their credentials. Approximately, 68% of persons who trust the emails that receive from a friend or colleague by including names of friends, hobby interests, places visited and others [36].

In spear phishing, attackers investigate the personal information regarding the targeted users from different sources such as social media platforms and design the emails in such as way that the victims are unable to discriminate the phishing emails from legitimate emails. Usually, people fall prey to spear-phishing because spear phishing offers relevant information about the targeted users.

Overall, spear-phishing email appears more realistic than traditional phishing emails [16]. As per the report of Cloudmark, spear phishing is the largest cyberattacks on JPMorgan Chase & Co., eBay, Target, Anthem, Sony, and various departments within the U.S. Government. Approximately 71.4% of targeted attacks involved the

use of spear-phishing emails and targeted more than 400 organizations consistently, which depleting $3 billion in the most recent three years. Therefore, it is cleared that there is a need for an applicable approach to mitigating the current phishing pattern.

Even though a wide number of methods are as well as available to reduce the momentum of these threats, anti-phishing organizations are still awaiting a completely successful approach. Therefore, this paper gives a step toward developing a method that incorporates a large scale of features and machine learning algorithm to classify phishing or legitimate [19]. Besides, a verifiable secret sharing algorithm is added to the method for verifies the identity of the person that demands as genuine. Hereafter, the genuine person is included in the profile so that the next time allows this person to access the email without further verification.

The major contributions of this paper as follows:

- *To detect spear-phishing emails using a machine learning algorithm with a large scale of features such as domain features, stylometric features, and others.*
- *To verify the email whether the persons with genuine or not using a verifiable secret sharing algorithm that shares the discrete logarithm.*
- *To detect zero-hour spear-phishing emails using an auto-upgrade profile that automatically upgrades the email-id of the genuine persons*

## II. Related works

An immense number of phishing email detection methods are available, but few of them only concentrate on spear-phishing emails [2, 10]. This section reviewed some of the methods, which assist to develop the abstract of the proposed method as shown below:

One interesting model called EmailProfiler is proposed by Duman et al. for detecting spear phishing using the metadata of the senders [6]. This model incorporates two operations; one was assessing incoming emails based on recipient-trained profiles, and another was generating profiles at the sender and creating the profile obtainable for querying at a trusty server. To design the profiles, the approach extracted 222 features, including body, header, sending time of the emails. The result shows that the approach evaluated with accuracy rates between 67% and 100%.

Stringhini et al. proposed another approach called IdentityMailer includes building a profile for the email-sending practice of a user [38]. Three types of features were used including writing habits, composition habits, and interaction habits. The extracted features from the emails are compared with a behavioral profile to identify whether the sender is genuine or fake. In case, the sender is found as fake then terminates the verification, or it is genuine then performs an identity-verification using answering a security question or a more advanced method.

An affinity graph-based semi-supervised learning approach with email profiling features such as origin features, text features, attachment features, and recipient features is proposed by Han et al. for detecting spear-phishing emails. An experiment is carried out with the spear-phishing emails and unknown emails using a k-nearest-neighbor (KNN) graph in order to measure the distance between email profiling. Using the random forest algorithm shows that the model achieved 0.9 F1 scores with a 0.01 false-positive rate.

Technical and non-technical approaches are as well as used in order to defend from spear-phishing attacks. Pilli et al. [42] presented a detailed survey on network secirity and related challanges and suspicion level modules of email's contents and message-id are used is confidential domain. Besides this model maintained immunological memory cells (IMCs) to detect easily the subsequent attacks from the earliest known or detected phishing sources and the User's information on this phase was additionally updated.

Stembert et al. proposed another rich prototyping approach to detect spear-phishing attacks using the combination of warnings, blocking, educational messages, and reporting [37]. This approach is designed and implemented three mockups like reporting button, blocking and warning of suspicious emails, and providing educative tips. This model contains two types of sensors; one was user-generated alerts, and another was an intrusion-detection system (IDS) generated alerts [18].

Aycock et al. mentioned in their article that Spear phishers attack in a single organization requires two ways to achieve their goals: external and internal [1]. The external approach is targeted to the organization from outside and internal from inside the organization. The external spear-phishing emails are recognized by the distributed checksum clearinghouse (DCC) and the internal spear-phishing emails are detected by generating individual user profiles of the organization.

One novel model that blends both stylometric features was extracted from emails, and social features of the online social network to identify targeted spear-phishing emails [13, 25]. To prove this statement, Dewan et al. experimented with 27 features including 18 stylometric, and nine social features, using 10-fold cross-validation. The experimental result shows that the model evaluated better accuracy 98.28% without using social features like LinkedIn.

To lessen spear-phishing attacks through the document authorship method, a novel model Anti-Spear phishing Content-based Authorship Identification (ASCAI) was proposed by khonji et al. [15]. To identify the Authorship of the senders, a profile of regular users is created without relying on the sender's user IDs and calculated the write-print of the newly arrived message using Jaccard's similarity index. The experiment of extraction of Writeprint module via the use of an effective source-code authorship method, namely SCAP and evaluated the accuracy 87%.

As far as it is known, no previous research has investigated both traditional phishing and spear-phishing emails together and one way to overcome these problems is to assemble all features to detect all categories of phishing attacks. Accordingly, this paper gathers all the features such as email-id features, social engineering features, readability features, writing-style features, and others and develops an abstract method that identifies phishing emails and spear-phishing one by one.

## III. The proposed method overview

This paper presents a novel method, and the architecture of the method is shown in Fig. 1. The method receives the emails from the user's mail inbox, then examines whether the emails belong to phishing emails or legitimate emails. Initially, the proposed method validates the receive emails with their profile, whether the users earlier communicate with the persons or not, who sends the emails as explained in section IV. If the method found the previous communication, then terminates the further investigation and allows the users to access the emails. Otherwise, the method extracts features from the email-id domain and stylometric features. The features are trained to the machine learning algorithm to classify the phishing and legitimate**??**. If the method classifies the email as phishing then inform the user regarding the phishing email.
If the method considers the email as legitimate, then the method applies the verification algorithm to verify the email whether the sender person is genuine or not. If the person is genuine then the method upgrades the profile of the method.

## IV. Generating Profile

A profile alludes to the whitelist database where only genuine email-ids are included. As mentioned above a spear-phishing attack is conducted on a small number of users, therefore, the profile of authors would be an applicable mechanism. On the other way, the weakness of this profile is that the legitimate email-id that not in the profile regards as a phishing email. Hence, the auto-upgrade profile is Incorporated in this method. The auto-upgrade profile frequently upgrades the genuine email-id, once the method perceives that email-id as genuine. On the more positive side of this approach is that the next time the identical email-id would get permission to move into the user's mailbox without further examination by other filters.

## V. Feature Extraction

Although the various collection of features are used by several researchers to detect phishing emails, a spear-phishing email is different from the traditional phishing emails because the spear-phishing email mimics the specific persons rather than the random persons. In this scenario, a person's behavior plays an important role in this detection approach. This paper analyzes the various features from the traditional phishing features as well as the user stylometric features which are explained in the remaining parts of this section.

### *A. Domain analyze*

The domain is used by attackers to control the users and appears as legitimate domains. Attacker employs various ways to manipulate users to promote the conviction since most of the attacker exploits the username of the specific persons but the domain of the email-id belong to the phishing domain. This attack is also called domain spoofing [34]. To detect domain spoofing, this paper collects domain features in the feature's corpus to detect spear phishing. These features are primarily inspired by the existing literature as explained below:

- Dots: This feature is very common in the phishing email domain. Attackers manipulate the legitimate domain by adding extra dots that create the legitimate domain to several subdomains. An example of this attack is the PayPal domain like name@pay.pal.com [7, 21, 28].

- Domain age: Age of the domain is used to identify the validation of the domain because the phishing domain lives only for a day. Within these days, the attacker quickly finishes their task and disconnect because of avoiding an act of catching by anti-phishers organizations. This feature is as well used in [7, 28, 43]

- Typosquatting domain: Typosquatting domain is a type of cybersquatting where attackers manipulate the domain, so it resembles the legitimate domain. Various models are used by attackers to fulfill their tasks such as adding more characters, Homoglyph, Insertion, Omission, Repetition, Transposition. This feature is also used in [3, 15].

- Dotted-decimal IP address: It is one of the common methods where an attacker uses an IP address instead of the domain name. For example, name@61.129.33.10. This feature is also used in [5, 28, 43].

- Domain length: The domain length of the phishing site is different from the legitimate domain because the attacker inserts some more characters with the legitimate domain to hide the phishing parts. This feature is also used in [23, 40].

- Special character: Most of the phishing domains are used with special characters like "-", and in most of the legitimate domains avoid this character in their domains. This feature is also used in [20, 21, 43].

- Encode in the domain: Encoding alphabets into their corresponding ASCII codes. For example, name@%36%2E%23%36%2E%31%39%35%2E%36%31:%36%39%30%33/%6C%69%6E%64%65%78%2E%68%74%6D. This feature is also used in [41].

- Abnormal domain: Most of the legitimate website domains are registered in the whois lookup database. If the domain name is absent in the whois lookup, then the method regards as an abnormal domain. This feature is also used in [9, 17, 18].

- Search engine query: The Search engine provides the rank of the domain based on their traffic. It can be seen that the legitimate domain always appears in the top rank of the search engine result, unlike the phishing domain. This feature is also used [22, 43].

- Domain Alexa rank: Alexa is used to ranking the domain over a while based on traffic information, access levels, connections to other websites, and the refreshed data. This feature is also used in [14, 22].

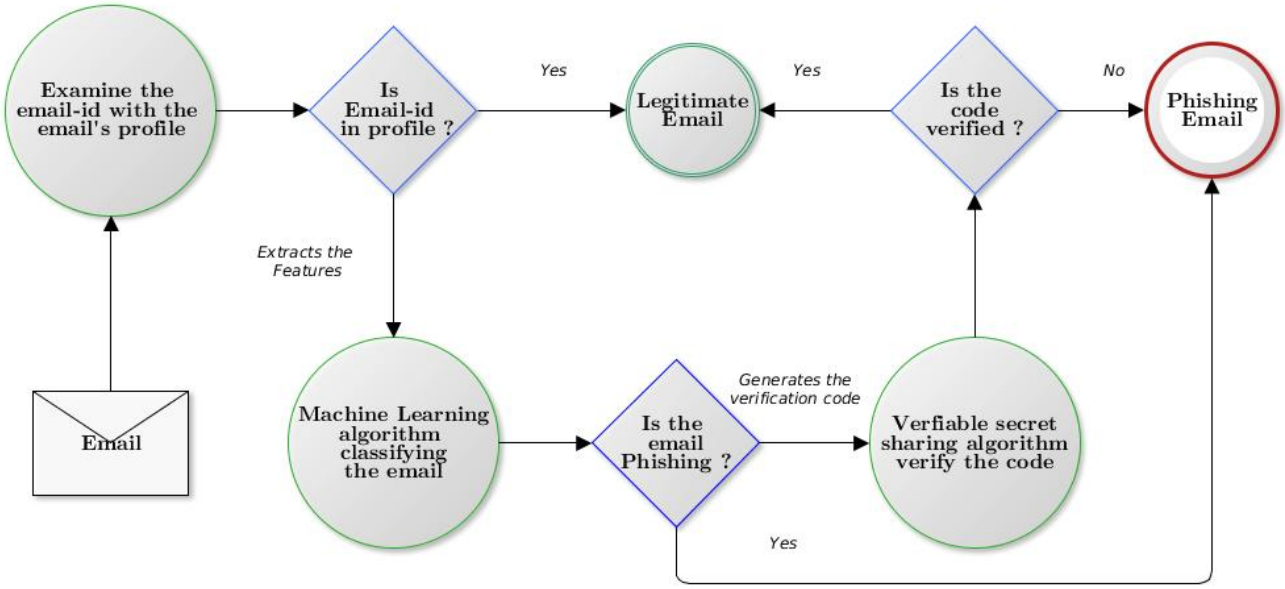


**Figure. 1**: The proposed method Architecture

The method collected 10 features by analyzing the domain of the email-id. The data-type used in these features is Boolean data-type; that is, if the feature presents in the domain then the method assigns 0, otherwise 1.

### *B. Stylometric features*

Stylometric features attempt to recognize patterns of writing-style in text [11]. This feature is also important to identify the authorship of genuine persons. In a spear-phishing email, the patterns of the actual person's style are impersonated to promote victims to believe in it. However, According to Forensic Stylistics [27], there are two premises of writing-style; 1) Two writers' writing-style always dissimilar, 2) The author does not write similarly constantly. On the behalf of these premises, the method incorporates novel and existing features to distinguish spear phishing. Several researchers provide many features to identify the authorship of the particular people [32].

### *C. Readability Features*

Readability scores help individuals to compute how hard to peruse a piece of text. Usually, companies or organizations maintain their standard of writing text in emails and before sending any specific emails to their customers. All organizations or companies have their writing-style of the emails. Therefore, the proposed method accompanies readability features to distinguish spear-phishing emails from legitimate emails and selects eight algorithms (8 features) that are,*automatic readability index, Coleman Liau Index, Flesch reading ease score, Flesch–Kincaid Grade Level, RIX, LIX, SMOG Index and Gunning Fog Index*.

### *D. Syntactic features*

Syntactic features refer to the grammar rules, and three features are selected to extract the syntactic features from the emails.

- Frequency of punctuation: Punctuation is a symbols that indicates where pauses, stops, questions, omissions, introductions, and other forms of expression occur in writing. Punctuations are to help understanding and correct reading, accordingly, every writer has their way of using punctuation. The proposed method employs eight punctuation that are, ",",".","?","!"," : ",";","'",""" to discriminate authors.
- Frequency of stop words: Stop words are frequently used in almost all sentences such as a, an, the, and, etc. Author use of stop words varies from each other. NLTK provides a collection of stop words based on their research. Total 127 words are gathered as features for the proposed method.
- Part of Speech: Every single word belongs to one of the eight parts of speech in the English language including noun, pronoun, verb, adverb, adjective, conjunction, preposition, and interjection. This part of speech is used to distinguish phishing emails, or spam from legitimate emails.

### *E. Lexical feature*

The lexical feature is the combination of two categories including word-based features and character-based features as explained below:

- The word-based features refers to nine features including the number of words, short words, characters in words, average word length, average sentence length in terms of character, average sentence length in terms of word, frequency of once-occurring words, frequency of twice-occurring words, frequency of words in different length.
- The character-based feature refers to eight features including the number of characters, alphabetic characters, uppercase characters, digit characters,

white-space characters, tab spaces, frequency of letters, frequency of special characters.

### F. Structural features

Structural features consist of personal identifiers of the authors including the number of lines, number of sentences, number of paragraphs, number of sentences per paragraph, number of characters per paragraph, number of words per paragraph, greeting content, farewell content, email, telephone, URL, the indentation of the paragraph. Total 12 features are used in structural features.

### G. Social-Site feature

The social site gives a platform to online users for publishing themselves publicly. This feature includes education, communication location, permanent location, career experience, languages, projects, any certificate, jobs, summary field. Total 10 features collect from the social site.

#### 1) Content-specific features

Content-specific feature refers to keywords that are used to confuse users to disclose their sensitive information [26].These keywords include *"account, access, bank, credit, click, identity, inconvenience, information, limited, log, minutes, password, recently, risk, social, security, service, and suspended"*. Total 18 keywords are selected for features. Overall, 218 features are selected for the proposed method, and a machine learning algorithm is applied to classified whether phishing or legitimate. If it is classified as phishing then terminate the proposed and warn the users regarding phishing emails or it if is classified as legitimate then one more filter is used to verify. This filter is discussed in details in section VII.

## VI. Machine learning algorithm

Once the features are collected, then machine learning algorithm is employed to classify the phishing emails from legitimate emails. Kumar et al. [12] proposed a deep learning based model that can be utilized make the system secured. The proposed method employs supervised learning where the method assigns classes to every user as a decision attribute.
Although several machine learning classifiers are available, a random forest classifier is widely used in phishing detection because it gives superior accuracy rate [17, 35]. It is an ensemble classifier where multiple decision trees are collected to classify a new class. A voting system is utilized in the classifier where multiple trees give votes and select that class that got more votes. Suppose, a dataset contains N number of training sets which gives N decision trees and these trees are made randomly. On more variable M is selected as input dataset for testing. Using the voting system, the classifier selects "m" where $m < M$ randomly from M. The best split of these "m" is used to split the node.
If the id of the user is considered as phishing terminates the method otherwise the verifies the id using the publicly verifiable secret sharing scheme as mention in the following section.

## VII. Verifying the Email-id

This step is additionally added to the proposal to confirm the sender is a genuine person. The proposed method employs the publicly verifiable secret sharing scheme to verify the secret when the user receives the email from the unknown person [19]. The receiver requests the secret code from the sender, if the sender's secret code can be verified through the verifiable secret sharing scheme then, the email-id is upgraded to the profile; otherwise, the method warns the user regarding the phishing email.
In this scheme, the method applies the general monotone access structure and threshold scheme to verify the secret message. Assume, $s$ be the secret message in a threshold scheme with threshold $k$ and $x_i \in \mathbf{Z}_p$ where $x_i \neq 0$ is assigned to the sender $P_i$. The user employs the known value $g^s = S$ where $g \in \mathbf{Z_p}$ be the generator or primitive root of $\mathbf{Z_p}$ if for every $a \in \mathbf{Z_p}$ it has $g^r = a$ for some integer $r$. In addition, the user choose random elements $f_j \in \mathbf{Z_p}$, $j = 1 \ldots k-1$ and publishes the values $g^s = S$ and $g^f = F$.
The user sends the message($s_i$) to the sender to verify the message

$$s_i = s + \sum_{j=1}^{k-1} f_i {x_i}^j (mod\ p) \tag{1}$$

For the example, assume $s = 2$ and generator is $g = 2$ of the $F_{19}$, threshold $k = 2$ and random element $f_1 = 3$, the user publishes the values $S = g^s$; that is $S = 2^2 = 4$ and $F = g^f = 2^3 = 8$. Suppose the id of the receiver $i = 1$ to whome the user send the message; so $x_1 = 1$ and computes the equation (1); $s_1 = 2 + 3 * 1 = 5$.

$$S_i = S \prod_{j=1}^{k-1} F_j^{({x_i}^j)} (mod\ p) \tag{2}$$

$$S_i = g^{s_i} \tag{3}$$

The receiver verifies the id through comparing the equations (2) and (3) and if the both the equations are identical then the email are verified. The receiver receives the publicly values ($S = 4$, $F = 8$ and $s_1 = 5$), then the receiver computes the equation (2); $S_1 = 4 * 8 * 1 = 32 \bmod 19 = 13$ and verify using the equation (3); $S_1 = 2^5 \pmod{19} = 32 \bmod 19 = 13$. Now, the equations (2) and (3) are identical; therefore, the message is verified.

## VIII. Discussion

This paper proposed a novel method that detects the spear-phishing emails as well as verifies the email-id. It is believed that the proposed method tends to detect spear-phishing emails with high accuracy. However, lack of a dataset of spear-phishing emails, it is unable to evaluate in a real-time scenario. Below discuss some of the issues of the method.
Most of the existing literature indicates the use of a whitelist for phishing detection is an inefficient approach [30]. However, the proposed method employs a whitelist-based profile since the spear-phishing attack usually occurs inside a small circle, attackers impersonate victim's friends, colleague, and others so that the victims easily believe in it

and uncover their credentials. As a result, the dimension of the profile would be small.
Another benefit of using a whitelist-based profile is that it reduces the detection time. The user receives a message from the unknown email-id then the method initially verifies the unknown message and if the message is found as legitimate then upgrades the profile. Again, the user receives the identical email-id then the method allows the user to access the message by investigating only on the profile.
The dimension of features is also an important challenge because a large dimension may increase the processing time in the training dataset and sometimes the irreverent features decrease the detection accuracy. O the other hand, a small number of features leads to difficulties to distinguish between two persons. This method collected 218 features and it is considered as large. In the future, a feature selection algorithm such as a wrapper, ranker, and others is applied to reduce the dimension of the features and find the best features set for the method.
The proposed method applies to an organization because it uses the verifiable secret sharing scheme to verify the person. If the person is genuine then the share code is verifiable; otherwise, the code is unverifiable. The users within the company have a unique id and use some publicly known elements to verify the secret. According to us, the verifiable secret sharing scheme is the strongest concept of this paper to detect spear-phishing emails and it is believed, this method has the potential to detect spear phishing in real-time.

## IX. Conclusion

This paper proposed an abstract method that detects both traditional phishing and spear-phishing emails. This method contains auto-upgrade profile, domain validation, writing-style to detect all categories of phishing attacks. The machine learning algorithm is as well as used in the method to classify spear phishing. This method adds one more filter to verify the genuine emails using the verifiable secret sharing scheme. The literature shows that this method has the uniqueness of detecting traditional phishing emails as well as spear-phishing emails using multi-dimensional features and the proposed method employed the verifiable secret sharing scheme so that the attacker unable to cheat the users.
In the future, this method would be implemented in the real-time scenario, and more features will be added to the method to reduce the momentum of phishing trends.

## References

[1] Aycock, J. (2007, September). A design for an anti-spear-phishing system. In Virus Bulletin Conference September 2007.

[2] Kim, B., Lee, D. Y., & Kim, B. (2020). Deterrent effects of punishment and training on insider security threats: a field experiment on phishing attacks. Behaviour & Information Technology, 39(11), 1156-1175.

[3] Buber, E., Demir, Ö., & Sahingoz, O. K. (2017, September). Feature selections for the machine learning based detection of phishing websites. In 2017 international artificial intelligence and data processing symposium (IDAP) (pp. 1-5). IEEE.

[4] Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. Computers & Security, 68, 160-196.

[5] Dhamija, R., Tygar, J. D., & Hearst, M. (2006, April). Why phishing works. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 581-590).

[6] Duman, S., Kalkan-Cakmakci, K., Egele, M., Robertson, W., & Kirda, E. (2016, June). Emailprofiler: Spearphishing filtering with header and stylometric features of emails. In 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC) (Vol. 1, pp. 408-416). IEEE.

[7] Fette, I., Sadeh, N., & Tomasic, A. (2007, May). Learning to detect phishing emails. In Proceedings of the 16th international conference on World Wide Web (pp. 649-656).

[8] Sarginson, N. (2020). Securing your remote workforce against new phishing attacks. Computer Fraud & Security, 2020(9), 9-12.

[9] Gupta, S., & Kumaraguru, P. (2014, September). Emerging phishing trends and effectiveness of the anti-phishing landing page. In 2014 APWG Symposium on Electronic Crime Research (eCrime) (pp. 36-47). IEEE.

[10] Han, Y., & Shen, Y. (2016, April). Accurate spear phishing campaign attribution and early detection. In Proceedings of the 31st Annual ACM Symposium on Applied Computing (pp. 2079-2086).

[11] Harpalani, M., Hart, M., Singh, S., Johnson, R., & Choi, Y. (2011, June). Language of vandalism: Improving wikipedia vandalism detection via stylometric analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (pp. 83-88).

[12] Kumar, N., & Sukavanam, N. (2020). An improved CNN framework for detecting and tracking human body in unconstraint environment. Knowledge-Based Systems, 193, 105198.

[13] Dewan, P., Kashyap, A., & Kumaraguru, P. (2014, September). Analyzing social and stylometric features to identify spear phishing emails. In 2014 apwg symposium on electronic crime research (ecrime) (pp. 1-13). IEEE.

[14] Kausar, F., Al-Otaibi, B., Al-Qadi, A., & Al-Dossari, N. (2014). Hybrid client side phishing websites detection approach. International Journal of Advanced Computer Science and Applications (IJACSA), 5(7), 132-140.

[15] Khonji, M., Iraqi, Y., & Jones, A. (2011, December). Mitigation of spear phishing attacks: A content-based authorship identification framework. In 2011 International Conference for Internet Technology and Secured Transactions (pp. 416-421). IEEE.

[16] Laszka, A., Lou, J., & Vorobeychik, Y. (2016, February). Multi-defender strategic filtering against spear-phishing attacks. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 30, No. 1).

[17] Li, Y., Yang, Z., Chen, X., Yuan, H., & Liu, W. (2019). A stacking model using URL and HTML features for phishing webpage detection. Future Generation Computer Systems, 94, 27-39.

[18] Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2009, June). Beyond blacklists: learning to detect malicious web sites from suspicious URLs. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1245-1254).

[19] Harinahalli Lokesh, G., & BoreGowda, G. (2020). Phishing website detection based on effective machine learning approach. Journal of Cyber Security Technology, 1-14.

[20] Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007, July). Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In Proceedings of the 3rd symposium on Usable privacy and security (pp. 88-99).

[21] Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Intelligent rule-based phishing websites classification. IET Information Security, 8(3), 153-160.

[22] Nguyen, D. T., Shen, Y., & Thai, M. T. (2013). Detecting critical nodes in interdependent power networks for vulnerability assessment. IEEE Transactions on Smart Grid, 4(1), 151-159.

[23] Aggarwal, A., Rajadesingan, A., & Kumaraguru, P. (2012, October). PhishAri: Automatic realtime phishing detection on twitter. In 2012 eCrime Researchers Summit (pp. 1-12). IEEE.

[24] Mohammad, R. M., Thabtah, F., & McCluskey, L. (2015). Tutorial and critical analysis of phishing websites methods. Computer Science Review, 17, 1-24.

[25] Ali, W., & Malebary, S. (2020). Particle swarm optimization-based feature weighting for improving intelligent phishing website detection. IEEE Access, 8, 116766-116780.

[26] Patil, S. M., & BR, P. (2019). Security Analysis of Proxy Cryptography Based Group Key Management Schemes for Dynamic and Wireless Networks Under Active Outsider Attack Model. Journal of Information Assurance & Security, 14(2).

[27] Pavelec, D., Justino, E., & Oliveira, L. S. (2007). Author identification using stylometric features. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, 11(36), 59-65.

[28] Basnet, R., Mukkamala, S., & Sung, A. H. (2008). Detection of phishing attacks: A machine learning approach. In Soft computing applications in industry (pp. 373-383). Springer, Berlin, Heidelberg.

[29] Kumar, N., & Sukavanam, N. (2017). Deep Network Architecture for Large Scale Visua Detection and Recognition Issues. Journal of Information Assurance & Security, 12(6).

[30] Sonowal, G., & Kuppusamy, K. S. (2020). PhiDMA–A phishing detection model with multi-filter approach. Journal of King Saud University-Computer and Information Sciences, 32(1), 99-112.

[31] Khonji, M., Iraqi, Y., & Jones, A. (2011, December). Mitigation of spear phishing attacks: A content-based authorship identification framework. In 2011 International Conference for Internet Technology and Secured Transactions (pp. 416-421). IEEE.

[32] Zheng, R., Li, J., Chen, H., & Huang, Z. (2006). A framework for authorship identification of online messages: Writing-style features and classification techniques. Journal of the American society for information science and technology, 57(3), 378-393.

[33] Sonowal, G., & Kuppusamy, K. S. (2018). Mmsphid: a phoneme based phishing verification model for persons with visual impairments. Information & Computer Security.

[34] Kumar, N., & Sharma, A. (2019). A Spoofing Security Approach for Facial Biometric Data Authentication in Unconstraint Environment. In Progress in Advanced Computing and Intelligent Engineering (pp. 437-448). Springer, Singapore.

[35] Sonowal, G., & Kuppusamy, K. S. (2018). Smidca: an anti-smishing model with machine learning approach. The Computer Journal, 61(8), 1143-1157.

[36] Steer, J. (2017). Defending against spear-phishing. Computer Fraud & Security, 2017(8), 18-20.

[37] Stembert, N., Padmos, A., Bargh, M. S., Choenni, S., & Jansen, F. (2015, September). A study of preventing email (spear) phishing by enabling human intelligence. In 2015 European intelligence and security informatics conference (pp. 113-120). IEEE.

[38] Stringhini, G., & Thonnard, O. (2015, July). That ain't you: Blocking spearphishing through behavioral modelling. In International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (pp. 78-97). Springer, Cham.

[39] Rao, R. S., & Pais, A. R. (2019). Detection of phishing websites using an efficient feature-based machine learning framework. Neural Computing and Applications, 31(8), 3851-3873.

[40] Yadav, S., Reddy, A. K. K., Reddy, A. N., & Ranjan, S. (2010, November). Detecting algorithmically generated malicious domain names. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement (pp. 48-61).

[41] Zhang, D., Yan, Z., Jiang, H., & Kim, T. (2014). A domain-feature enhanced classification model for the detection of Chinese phishing e-Business websites. Information & Management, 51(7), 845-853.

[42] Pilli, E. S., Joshi, R. C., & Niyogi, R. (2010). Network forensic frameworks: Survey and research challenges. digital investigation, 7(1-2), 14-27.

[43] Zhang, Y., Hong, J. I., & Cranor, L. F. (2007, May). Cantina: a content-based approach to detecting phishing web sites. In Proceedings of the 16th international conference on World Wide Web (pp. 639-648).

## Author Biographies

**Gunikhan Sonowal** Dr. Sonowal was born on October 20, 1988, at Tinsukia City, Assam, India. He received his Bachelor of Science degree(B.Sc.) from the Sibsagar College (Affiliated to Dibrugarh University) in the year 2010. After that he completed Master of Computer Application at University of Hyderabad, India. On completion of his PG, he joined the Ph.D. program in computer science and engineering at Pondicherry University.

**Aditi Sharma** is associate Professor in Deptt. of Computer Science & Engineering at Quantum University Roorkee, India. She received her PhD in Cybersecurity from Department of Computer Sc. Engineering M.B.M. College Jodhpur, India in 2019. She has technical and key leads of several renewed IEEE and Springer National and International Conferences. She also was the convener of IEEE conference *ICFIRTP*2020. She has supervised graduate and master students in the field of computer science. Her research area includes Cyber security analytics, IoT, Machine learning and VLSI technologies on wireless sensor network.

**Latika Kharb** is Profesor been working as a Professor in the Jagan Institute of Management Studies (JIMS), Delhi, India since 2013. Dr. Kharb served as a reviewer of Elsevier, Springer and IGI Global journals and conferences and worked as an advisory board member in national and international conferences. Her career accomplishments over the past sixteen years include more than 164 peer reviewed papers/articles with approx. 190 citations in national/international journals and conferences. She has contributed many Chapters in Scopus Indexed Books with Elsevier, Emerald, AAP-CRC, Taylor & Francis, EAI-Springer and many other with International and National Editors.