Received: 04 July 2023 | Revised: 24 Sept 2023 | Accepted: 14 Dec 2023 | Published: 31 Dec 2023
Volume 1(2) Pages 93-108 DOI: https://doi.org/10.56279/jicts.v1i2.43

JICTS
Journal of ICT Systems

Security Risk Scale: A Case of Email Phishing Detection using Text Mining

Robert Karamagi[1] Said Ally

Mathematics and ICT Department, The Open University of Tanzania, Dar es Salaam, Tanzania

[1]Corresponding author
Email:
robertokaramagi@gmail.com

Funding information

This work was not funded.

Keywords

Security,
Phishing Email,
Risk Scale,
Social Engineering,
Text Mining

Abstract

The rise of cybersecurity defense has led to threat actors needing to deploy more resources to break into systems. However, the human factor remains the weakest link for system penetration through social engineering techniques, especially when phishing is used. While cybersecurity and risk management go hand in hand, a measure of the risk posed by the threats in our environment is a crucial control factor. In this study, an experimental test was conducted on 327 simulated phishing tests with probable responses of mail users. Our goal was to determine the emotion that triggers interaction when a false email is used to trick victims into unintentionally submitting data, providing unauthorized access to the mail server. Four major causes of successful phishing attacks where emotions are triggered were found to be the manipulation of curiosity, fear, authority, and empathy. For enhancing phishing detection, we proposed a framework that dynamically scales the security risks resulting from the social engineering attack through the content of the phishing email received in real time. Although the technical controls have proven to be far more effective in securing systems, the framework provides administrative techniques with risk scales that organizations with mail servers can use to train their staff and resolve the ever-growing security problem of social engineering attacks through phishing emails.

1. Introduction

Phishing is a social engineering technique where a malicious individual sends a fake message to a mail user. It is a deliberate action requesting the victim to perform some actions that will help the threat actor achieve an unethical mission, such as stealing login information or bank credit card details [1]. Phishing has financial and legal impacts on users [2] and is reported as one of the potential cybersecurity threats in Tanzania [3].

As one of the malicious acts, phishing has constricted e-commerce growth and led to losses of up to $85 million, according to the Tanzania Cybercrime Study Report of 2016 [4], with $30 million a year as the estimated cost of malicious insider threats [5]. Although phishing attacks can happen even in a short message service [5], they are considered a major threat for most first-time mail users as they are unaware of the dangers and prevailing threats [6, 7], especially in rural areas [9].

It is evident that web attacks are doubling every year [8], with 70.54% related to phishing [4]. This is massively influenced by the rise of the cashless economy [9], the usage of mobile money services [10], and the lack of sufficient and sophisticated protection techniques [11].

Various security risk scales exist for determining if a phishing email exists. The risk shows how possible it is to phish a user. The scales apply techniques that involve the analysis of the email header structure, URL information, script function, and psychological features to prepare a classification dataset. There are many risk scales in the market. Some of them include Virtual Risk Officer [12], Tessian Human Layer Rik Hub [13], multidimensional phishing susceptibility prediction model [14], dynamic scale to compute the risk [15], emailage risk scale [16], and phishing risk by observable characteristics called NIST[1] Phish Scale [17].

Despite all these risk scales, with evolving technology, cybercriminals are getting more sophisticated in their attacking mechanisms, and the cost of cyber defenses is skyrocketing. As of 2023, enterprise email phishing detection and prevention solutions had been charging at least $3 per user per month. The costs of blocking spam and phishing emails increase based on the number of incidents. Organizations aiming to optimize their cybersecurity expenditures may focus their budget on defending the higher-risk phishing emails revealed by the proposed email phishing security risk scale methodology. A high-risk phishing email should be countered by stern cybersecurity protection, whereas, for less-risk phishing emails, the security management may channel defense efforts where greater risk exists.

Many interesting efforts have been made to efficiently detect phishing attempts by examining the content of the email using artificial intelligence techniques. Improvements may be made to further classify the detected phishing emails based on the emotional sentiment within their context, then provide ratings of the risk posed by such a class of phishing emails.

This study proposes a scale to measure risk based on the content written by the hacker in the email. Although [17] proposed a similar type of scale, the technique has not critically dealt with emotional triggers in the email content. Since social engineering mainly aims at targeting human emotions, it is essential to focus on that aspect.

Despite the efforts made to accurately predict phishing emails using advanced machine learning (ML) models, a realization of the severity and risk

[1] NIST stands for the National Institute of Standards and Technology

posed by the detected phishing emails is yet to be uncovered. This is a challenge because, without a measurement of risk, security protection mechanisms are not implemented with consideration of threat levels and probable exposure. Thus, this paper proposes a security risk scale to enhance phishing detection in mail systems.

2. Related works

Various efforts have been made to sophisticate the detection of phishing emails using artificial intelligence (AI) techniques. A vast number of models have been proposed that leverage different ML algorithms to understand the content written in the email using text mining and natural language processing (NLP). Through AI applications, it has been possible to detect whether an email is a phishing one or not with motivating accuracy.

Several phishing detection models have been proposed in the literature. The first one is the Support Vector Machines (SVM) with an accuracy of 95%. The model is useful for classifying between phishing and non-phishing emails through analysis of the email-header structure, email-URL information, email-script function, and email psychological features used for the preparation of a classification dataset [18].

Halgaš et al [20] proposed an improved recurrent convolutional neural network (RCNN) model with multilevel vectors, word embedding, and phishing classifiers in comparison to long-short-term memory (LSTM) layers. The model has an accuracy of 99.8%.

Olayemi [21] proposed a Naive Bayes, K-Nearest Neighbor algorithms, and Decision Tree (J48) classification classify the word embedding and remove noise and non-words. The model has an accuracy of 99%. Castillo et al. [22] proposed a back propagation model with an accuracy of 95.7%. The model works through a classical feed-forward network with multiple hidden layers to detect phishing emails.

Lee et al. [23] proposed a Bidirectional Encoder Representations from the Transformers (BERT) model, which uses email content and context features to detect phishing. The model has an accuracy of 87%. Bountakas et al. [24] proposed a Natural language processing (NLP) model for lexical and semantic analysis of email content using random forest, decision tree, and logistic regression. This model offers an accuracy of 98.95%. Franchina et al. [25] proposed a text mining and text analytics model with an accuracy of 99.2%. The model uses text categorization, information extraction, clustering, and text summarization. The authors analysed email metadata and content, including the body and subject, to detect phishing [25].

Salahdine et al. [26] proposed an Artificial Neural Networks (ANN) model with two hidden layers extract content features present in the email body and header. The model gives an accuracy of 94.5%. Ahmed et al. [27] proposed a multi-layer perceptron (MLP) neural network and Random Forest classifiers with feature selection to extract headers and hyperlinks in the email. The model gives an accuracy of 99.5%. Noah et al. [28] built upon a stochastic gradient descent classifier (SGD) with an accuracy of 96% to predict a phishing email by analyzing the content such as the subject line, email address, and body.

With regards to measuring the phishing risk, various authors have made appreciable contributions, but their methods are limited when it comes to providing risk measurements based on the psychological manipulation or social engineering technique observed within the content of the phishing email. Most of the risk scales append a

risk score based on the user that is tricked by the phishing email. Such risk scales include [12-15].

In [15], authors developed a 10-point single-dimensional phishing risk scale to attach a respective risk score to an individual based on the actions they perform when they are subjected to a phishing attack. They proposed the factors responsible for either an increase or decrease in the phishing score, in which the risk of all individuals is initially taken as neutral with a score of five. On this scale, there is no change in the risk score of a user who simply reads a phishing email. However, the risk score shall increase when a user clicks a link in the phishing email and increase more if the user gives away classified information in the process. On the other hand, the risk score decreases when a phishing email is left untouched and decreases more if the phishing email is reported to the responsible authority [23].

In [14], authors devised a technique to predict the risk of a user being susceptible to phishing using a multidimensional phishing susceptibility prediction model (MPSPM). The model is based on multiple supervised machine learning experiments using both legitimate and phishing emails, with decision factors being demographic, personality, knowledge experience, security behavior, and cognitive factors to determine or classify if the email is susceptible or easily tricked (high-risk) or non-susceptible (low risk) [24].

In [12], authors designed the Virtual Risk Officer (VRO) to aid organizations in determining the vulnerability of their staff to phishing attacks. Using this approach, the risk score of a user is dynamically allocated using a deep learning neural network algorithm based on AI factors such as phishing-prone percentage, security awareness training status, breach data, job function, user risk booster, and group risk booster [25].

In [13], authors developed the Risk Hub to provide a granular insight into the email users' levels of risk and risk enhancers, where the risk score increases when bad security actions are performed by the user [26]. The platform focuses on the behavior of users to convey an extensive spectrum of risk assessments over incoming and outgoing email threats, phishing attacks, and data breaches. A contextually rich risk profile of a user is provided by a unique data modeling technique called the Behavior Intelligence Model. Table 1 compares risk scales, showing their types and measured objects.

Table 1. Comparison of risk scales.

Type of Risk Scale	Name of Risk Scale	Scale Focus	Measured object
Affinity IT Security Services (2019)	'Phishing Risk' scale	User Interaction	User
Yang et al. (2022)	Multidimensional Phishing Susceptibility Prediction Model (MPSPM)	User Personality	User
KnowBe4® (2017)	Virtual Risk Officer (VRO)	User Personality and Interaction	User
Tessian® (2021)	Tessian Human Layer Rik Hub	User Personality and Interaction	User
LexisNexis® Risk Solutions (2017)	Emailage	Email Identity	Email
Steves et al. (2020)	NIST Phish Scale	Email Content	Email

Natural language processing and text mining models have evolved to perceive the context of email messages and determine whether they are malicious or not. The introduction of ML and AI techniques into the cybersecurity regime has significantly improved the mechanisms to detect phishing emails. However, modern techniques on how to predict and classify phishing emails are still required. The existing risk scales have limitations. For instance, a risk [14] deals with the personality of the user, while that [15] works with how the user interacts with the email. On the other hand, the risk scales [12] and [13] combine the effects of user personality and user interaction with the phishing email to rate the risk. Other risk scales use identity features of the email [16] and analysis of the actual email content [17]. Generally, all the risk scales focus on the user's behavior or email address.

This study provides risk input for the enterprise risk management program and defines the associated phishing key risk indicators (KRIs). The level of phishing risk may be objectively quantified, alerting organizations in advance of potential phishing risks that could cause damage.

Thus, this study attempts to address the social engineering techniques used by hackers in phishing emails to emotionally manipulate their victims and to find out if they have any effect on the probability of a user interacting with that phishing email. Furthermore, the study determines the impact of demographic factors on the probability of a user interacting with a phishing email for a given social engineering technique and its frequency. Lastly, the study assesses the accuracy level of the security risk scale of the proposed phishing detection framework.

3. Method

The study is both exploratory and experimental, proposing a phishing security risk framework that applies text mining and phishing simulation attacks to generate a security risk scale. The framework provides an assessment of the risk posed by the content forged in the phishing email (Figure 1).

The entire process is based on the following steps:

- Step 1: A phisher or external threat actor targets the organization with email phishing attacks that are received by the organization's mail server.
- Step 2: The inbound emails from the mail server are directed to the AI layer. Text mining is performed using natural language processing models after an initial pre-processing of the content in the emails to create a vectorized form of the words. Pre-processing involves sentence segmentation, tokenization, stemming, lemmatization, removal of noise, i.e., stop words (such as 'a', 'the', 'and'), special characters and punctuation marks, dependency parsing, and part of speech tagging. The corpus contains datasets to be used in the model's training algorithms.
- Step 3: Datasets are fed into the text mining block. The corpus is connected to available cloud services to receive new datasets.
- Step 4: ML algorithms are used to classify the vectorized words and detect if the content in the email is phishing or not.
- Step 5: Detected phishing emails are taken in for sentiment and emotion analysis.
- Step 6: Once the emotion is detected from the contextual data, it increments its respective emotion counter in the frequency count matrix.

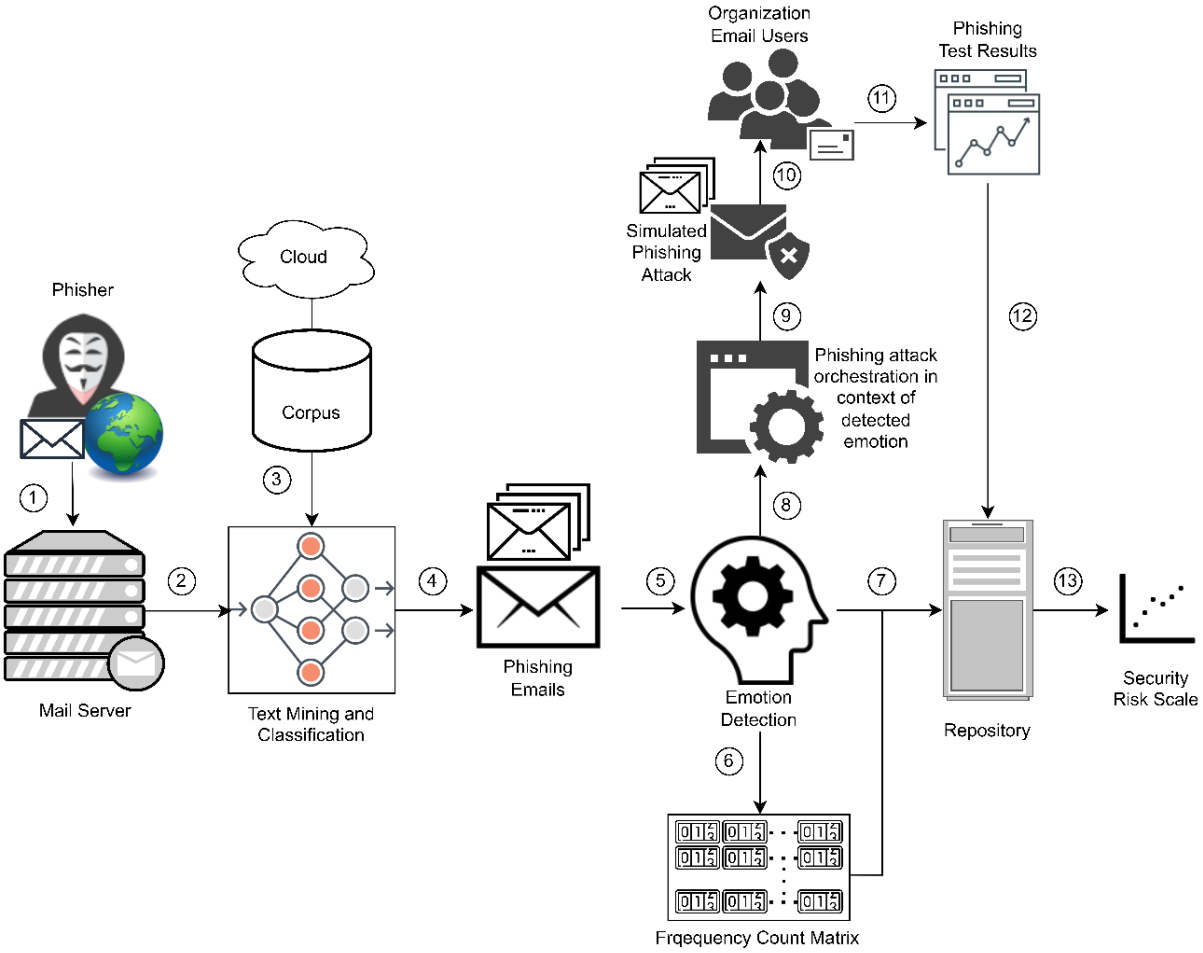

Figure 1. Architecture of email phishing security risk scale.

- Step 7: The information regarding the emotion used by the threat actor and its count are stored in the repository. At this point, the frequency of a user receiving a phishing email targeting a specific emotion is obtained from the live environment or real-world.
- Step 8: The detected emotion is fed to the phishing simulation layer.
- Step 9: A phishing email campaign is orchestrated in the context of the emotion detected. A simulated phishing attack is devised and staged for launch.
- Step 10: The organization's users receive test phishing emails to determine the probability of a user interacting with a phishing email and triggering the emotion detected.
- Step 11: The interaction of the organization's users with the orchestrated phishing email that triggers the detected emotion is recorded in the results.
- Step 12: The results portraying the impact of the phishing email, or its probability of exploitation are stored in the repository.
- Step 13: The phishing security risk scale is derived from the repository data of multiple emotions.

For phishing detection, the ML models are trained across a dataset of emails marked as “phishing” and "not-phishing." This classification must occur first to identify the status of the received email. At this point, natural language processing transforms the email in the mailbox into a machine-readable format that the model can process. The non-phishing emails are considered legitimate and are allowed to reach the mailboxes of the users. Only emails that are discovered to be phishing are used to create the security risk scale by moving them into the block of detecting the emotion. The difference between the detection of phishing emails and the detection of emotion is the dataset used to train the ML model, labeled as *Authority, Commitment, Contrast, Curiosity, Empathy, Fear, Liking, Reciprocity, Scarcity,* and *Social Proof.* The emotions were discovered through an exploratory research methodology.

The mean frequency of a user receiving a phishing email targeting a specific emotion is plotted against the mean probability of a user interacting with a phishing email triggering the emotion detected on the proposed security risk matrix.

In this study, the security risk was determined as a product of the probability of occurrence of the risk event and the risk impact in a qualitative sense. This means that the more the subject receives the phishing email, the greater the chances of them being exploited, since it fully depends on the likelihood of a user clicking the phishing link. This can further be interpreted as indicating that a qualitative impact that contributes to the security risk is only possible when a phishing link is clicked.

An experimental test was conducted in a bank where three separate phishing attacks were launched for social engineering. The social engineering techniques applied include authority, commitment, and reciprocity. The click rates were measured for a timing window of three days for a single phishing campaign. The interval for each campaign was two weeks to give independence to the results.

- Campaign 1 (Authority Test): A p*hishing email pretends to be sent by the bank CEO, requesting the staff navigate to the bank’s brand page via a link in the email.*
- Campaign 2 (Commitment Test): A p*hishing email pretends to be from the human resources training unit and motivates the employees to take a learning course.*
- Campaign 3 (Reciprocity Test): A p*hishing email convinces bank staff to navigate to a social movement page illustrating the decent work the bank has done for the staff community, so they in turn deserve a reciprocating hand of support for the initiative.*

The test involved mail users with at least bachelor's-level education working in the banking sector. The test was carried out using the KnowBe4 phishing simulator. The probability of a mail user interacting with a phishing email was determined using a Friedman statistical test on 5-point scale nominal descriptors: *very unlikely (1), unlikely (2), not sure (3), likely (4),* and *very likely (5)*. Data was populated from a questionnaire in which users were asked two questions each for 10 phishing techniques. For each phishing technique, the user was asked about the frequency with which they have received phishing emails and the likelihood of them interacting with such emails for the given technique. We collected 327 user responses to determine the emotion that triggers interaction with a false email.

The Friedman test was also used to determine if the social engineering technique used by the hacker in the phishing email has any effect on the frequency with which the user will receive that phishing email. The technique was used to find out

the frequency of receiving a phishing email using various forms of social engineering techniques:

- *Authority: humans will typically conform when an eminent authority confronts them.*
- *Commitment: the desire to work hard with effort—can allow hackers to convince a victim to follow their instructions.*
- *Contrast: email has two choices that contradict each other. If the target disagrees with one option, they may select the other.*
- *Fear: when people are frightened, they tend to do things they do not necessarily want, so the threat actor scares them in the email.*
- *A likeable hacker may act like they are someone the victim cares about to get them to perform their demands.*
- *Reciprocity: one pretends to have done a good deed, knowing people will be inclined to return the favor.*
- *Scarcity: when there is very little time or few opportunities offered, a victim may quickly agree to the phishing request.*
- *Social proof: usually, people feel better doing something if everyone else is doing it.*
- *Curiosity: Someone is more likely to follow the hacker's request if they are very interested in finding out more about it.*
- *Empathy makes a victim more vulnerable to accepting the demands in the phishing email.*

Similarly, a Kruskal-Wallis H Test was used to determine if demographic factors affect the probability of a user interacting with a phishing email for a given social engineering technique used by the hacker. To measure accuracy, measurements were evaluated using the following formulae:

$$\text{Error Rate} = \frac{\text{PST Value-SRS Value}}{\text{SRS Value}} \text{x}100\% \quad (1)$$

$$\text{Accuracy} = 100\% - \text{Error Rate} \quad (2)$$

where PST denotes Phishing Simulation Test and SRS denotes Security Risk Scale.

4. Results

Based on the analysis of the phishing emails that emotionally manipulate victims, all ten studied techniques, including *authority*, *commitment*, *contrast*, *curiosity*, *empathy*, *fear*, *lying*, *reciprocity*, *scarcity*, and *social proof*, were found to be the key social engineering techniques used by hackers.

The Friedman test revealed a significant effect of social engineering techniques on the probability of a subject interacting with a phishing email (9, n=100) = 52.306, p < 0.001, W =0.058, and a significant influence on the frequency of a subject receiving a phishing email (9, n=100) = 89.573, p < 0.001, W =0.100. The Friedman test returned an asymptotic significance of less than 0.05, which means the probability of a subject interacting with a phishing email is affected by social engineering techniques. The chance of a subject interacting with a phishing email provoking authority on the age range was statistically significant with (4, n=100) = 14.172, p=0.007; on the professional status, (3, n=100) = 12.979, p =0.005.

Results from the Kruskal-Wallis H test reveal that the education level of a subject influences the probability of clicking on a phishing link themed with the commitment phishing variable. The probability of a subject interacting with a phishing email-provoking commitment on the age range and education level was found to be (4, n=100) = 10.378, p=0.035, and (2, n=100) = 6.166, p=0.046, respectively, where the variables N, p, and W respectively represent total number of samples, probability value/asymptotic significance, and Kendall's coefficient of concordance.

Regarding the professional status of a subject, results show that the subject influences the probability of clicking on a phishing link themed

with the authority phishing variable. Specifically, someone who is unemployed or retired would react differently to clicking a phishing link with an authority theme than someone who is employed. A Kruskal-Wallis H test was also used to check if the gender of a subject has any effect on the frequency of receiving a phishing email. Results show that the gender of the subject influences the frequency of receiving a phishing email that triggers the social proof phishing variable. Table 2 summarizes the calculation of the phishing security risk by taking the product of the frequency and probability of the subject receiving a phishing email.

Figure 2 shows the derived phishing security risk scale that plots the mean values found from the phishing questionnaire survey for each social engineering technique. Table 3 summarizes the phishing test results for the three phishing campaigns for the emotional triggers of authority, reciprocity, and commitment. The number of recipients and the number of phishing emails delivered, opened, clicked, and reported for each phishing social engineering technique are tabulated. The risk rating distribution of the sampled phishing social engineering techniques (authority, reciprocity, and commitment) in our simulated phishing attack is compared with that of the designed security risk scale.

Figure 3 shows the distribution of risk ratings from the designed security risk scale and the distribution of risk ratings from the simulated phishing test for the authority, reciprocity, and commitment social engineering techniques. The plot shows that the distributions are similar in nature. This implies that a relationship exists between the risk ratings of the phishing social engineering techniques derived from the designed security risk scale and the risk ratings from the simulated phishing test.

Table 4 shows the derived risk ratings of the authority, reciprocity, and commitment social engineering techniques by taking the probability of a subject interacting with a phishing email as per the designed security risk scale. It refers to the values from Table 2 that tabulate the phishing security risk evaluation.

Table 5 tabulates the derived risk ratings for Authority, Reciprocity, and Commitment Social Engineering Techniques by taking the percentage of users that interacted with the phishing email corresponding to the associated phishing technique in the simulated phishing test. It refers to the values from Table 3 that tabulate the phishing test results for the sampled phishing social engineering techniques.

The phishing test using the authority technique was the first to be conducted in the series of phishing tests and had the least error (5.263%). The error was found to increase significantly in the second test, i.e., the reciprocity technique (50.820%). The test with the largest error was the final test, i.e., the commitment technique (80.882%). The increase in errors through subsequent phishing tests can be justified by users gaining awareness and suspicion of the possibility of phishing attempts following the significant success of the first phishing test.

$$\text{Error Rate} = \frac{|\text{PST Value} - \text{SRS Value}|}{\text{SRS Value}} \times 100\%$$

$$\text{Error Rate (Authority)} = \frac{|80 - 76|}{76} \times 100\% = 5.263\%$$

$$\text{Error Rate (Reciprocity)} = \frac{|30 - 61|}{61} \times 100\% = 50.820\%$$

$$\text{Error Rate (Commitment)} = \frac{|13 - 68|}{68} \times 100\% = 80.882\%$$

Table 6 shows a comparison of the risk ratings derived from the security risk scale and those from the phishing test. The error and accuracy are tabulated as well. The calculations for obtaining the error and accuracy are shown below.

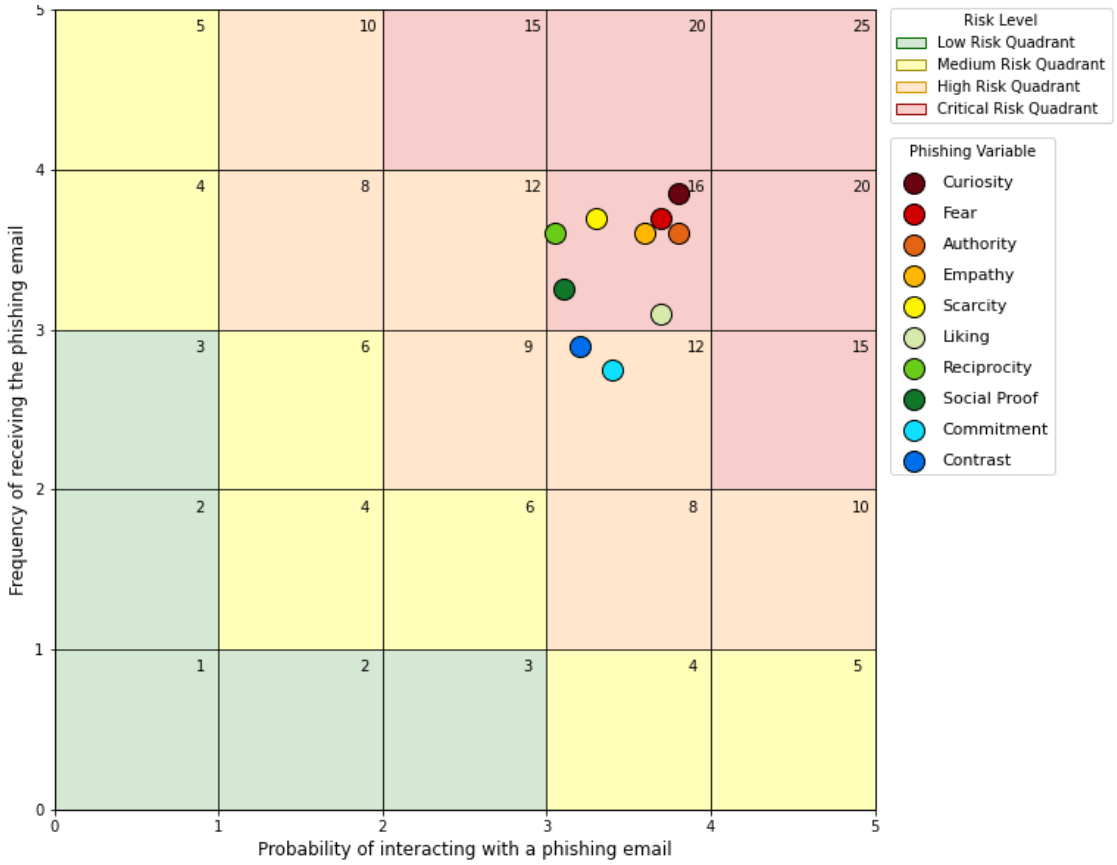

Figure 2. Phishing security risk scale.

Table 2. Security risk calculations.

Social Engineering Technique	Probability of a subject interacting with a phishing email	Frequency of a subject receiving a phishing email	Phishing Security Risk Score
Curiosity	3.8	3.85	14.63
Fear	3.7	3.7	13.69
Authority	3.8	3.6	13.68
Empathy	3.6	3.6	12.96
Scarcity	3.3	3.7	12.21
Liking	3.7	3.1	11.47
Reciprocity	3.05	3.6	10.98
Social Proof	3.1	3.25	10.075
Commitment	3.4	2.75	9.35
Contrast	3.2	2.9	9.28

Table 3. Phishing tests results for the sampled phishing social engineering techniques.

Phishing Technique	Recipients	Delivered	Opened	Clicked	Reported
Authority	4236	4227	3213	2561	16
Reciprocity	4229	4102	1908	567	45
Commitment	4233	4094	1527	195	87

Table 4. Risk ratings for social engineering techniques.

Social Engineering Technique	Probability of a subject interacting with a phishing email as per the designed security risk scale (%)
Authority	$\frac{3.8}{5} = 0.76$
Reciprocity	$\frac{3.05}{5} = 0.61$
Commitment	$\frac{3.4}{5} = 0.68$

Table 5. Risk ratings for authority, reciprocity and commitment social engineering techniques.

Phishing Technique used in the simulated phishing test	Number of users that opened the phishing email received	Number of users that clicked on a link in the phishing email received	Percentage of users that interacted with the phishing email in the simulated phishing test (%)
Authority	3213	2561	80
Reciprocity	1908	567	30
Commitment	1527	195	13

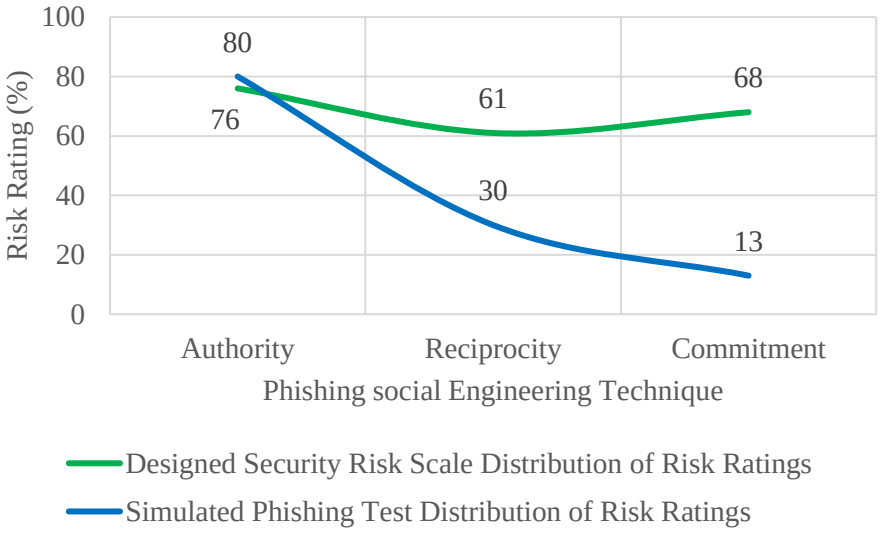

Figure 3. Comparison of the distributions of risk ratings with respect to the sampled social engineering techniques

Table 6. Performance measurement of the security risk scale.

Phishing Technique	Probability of a subject interacting with a phishing email as per the designed security risk scale (%)	Percentage of users that interacted with the phishing email in the simulated phishing test (%)	Error (%)	Accuracy (%)
Authority	76	80	5.263	94.737
Reciprocity	61	30	51	49
Commitment	68	13	80.882	19.118

5. Discussion

If we were to ask ourselves, what follows once we can tell with satisfactory accuracy that the email is a phishing one? Conventional measures involve blocking the email from the users' mailboxes and the domain from which the phishing email came. This study suggests that phishing emails do not exhibit the same content composition. Hackers concoct the phishing emails in different ways to help them achieve their objectives. The main goal of a hacker is to psychologically manipulate the mind of the victim into performing a poor security decision, regardless of the form of social engineering. The content of the phishing email can be in the form of the voice of the hacker, theme, or tone.

These techniques are emotional manipulation techniques that are variable depending on the choice of the hacker. The detection of phishing

emails was possible through reading the content using AI techniques, which shed light on the need to distinguish the manipulative style in the content.

Imagine a hacker trying to trick you by telling you that you have won 1 million dollars, or the same hacker telling you that you will be taken to trial and face a lawsuit against you. Which scenario is more likely to get you to submit to the demands of the hacker? Technical controls have proven to be far more effective in securing systems in comparison to administrative controls. By being able to measure the risk of the various phishing emails, organizations can optimize their cybersecurity expenditure. They may allocate funds to proactively detect, block and innovatively respond against the high-risk phishing emails, and go further into threat hunting across the domains and IP addresses observed. Lower risk attacks may be dealt with conventionally, however higher risks attacks should be dealt with innovatively.

The proposed solution is a user-based risk scale that measures the likelihood that a user is not vigilant enough to avoid a phishing attack. Since phishing mainly depends on exploiting the human emotional factor, it is important to find out the risk of manipulating a specific emotion that the user has. This conforms to a changing world where the level of temptations based on the emotions of an individual through email contents is alarming, with the greatest risk of successful exploitation.

The solution provides administrative controls to reduce the risk that evolves from the user receiving the phishing email, similar to the risk scales of LexisNexis® Risk Solutions. Again, regarding risk measurement based on the characteristics of the phishing email, the study conforms to the technique by [17].

As part of the content analysis, a plethora of historical and conduct details are used from the email address, making the risk associated with it visible. The frequency of use and composition of emails used by hackers categorizes them into related patterns. This is possible because email addresses exhibit a similar name structure, such as "*xyz@domainname*", which may be looked up through a list of flagged or risky emails to match against any high-risk known names. It is simpler for a user to discover that the email is a phishing one if there are many cues available in the email. This concurs with the study by [17], which revealed that a phishing email with more premise alignment, such as matching the target's work surroundings, is harder to realize.

Despite a good number of interesting efforts made to efficiently detect phishing by examining the content of the email using AI techniques, the solution provided in this study improves the detection of phishing emails as it uses emotional sentiment within its context. Furthermore, it provides ratings of the risk posed based on the proposed risk scale. Since social engineering mainly aims at targeting human emotions, it is essential to focus on that aspect.

In addition, the risk scale of the phishing email is practically assessed in real time, making it a dynamic risk scale. This helps to observe how the relationships between various phishing techniques used by hackers vary over time. On the static risk scale, we found that the curiosity technique had the highest risk score, followed by the fear technique. Over months, these values could change, and it would be necessary to know of these changes. So, the study proposes a setup that makes this possible.

6. Conclusion and Future Work

In this study, an email phishing security risk framework capable of plotting a security risk scale dynamically is proposed. The scale is based on the detection of the emotion manipulated by the hacker through phishing emails.

The scale is based on critical, high, medium, and low severity levels. Through the text mining approach, the emotions that trigger users to interact with phishing emails have been determined from the analysis of user responses and the experimental test. Major causes of successful phishing attacks are the manipulation of curiosity, fear, authority, and empathy emotions.

A relationship exists between the social engineering technique used by the hacker in the phishing email and both the frequency of receiving the email and the probability of a user interacting with that phishing email. The best accuracy of the risk scale in measuring the risk of exploiting a user was found to be 94.737%. Improvements in the experimental process can be made in future as users are social beings. Once tested upon social gatherings and conversations, they leak out the plot of the test, making them appear possibly more secure than they would rather be. As the hacker will always be hit by surprise, future work can involve scenarios where measurements are only taken by surprise and not repeated for the same population, unless excessive time has passed for them to forget.

For further research, an automated email phishing detection framework that is contextually aware of the risk posed by the content forged in the phishing email may be constructed to complement this study. For data-sensitive applications where the frequency of inbound emails is high, the application of AI can be used to determine phishing or external threat actors, which normally target the mail servers.

ACKNOWLEDGEMENT

We thank the referred bank for giving access to test mail servers for real time experimental tests.

CONTRIBUTIONS OF CO-AUTHORS

Robert Karamagi	[ORCID: 0000-0002-1036-1745]	Conceived the idea and wrote the paper
Said Ally	[ORCID: 0000-0002-4419-6004]	Reviewed the paper and provided technical support

REFERENCES

[1] M. Rosenthal, *Must-Know Phishing Statistics:* Updated 2021, Tessian, 2021

[2] A. M. Salim, *Assessment of Mobile Money Transaction Frauds and Consequences Confronting Zanzibar Telecom Service Providers*, Asian J. Econ. Bus. Account., **22** (20): p. 16–31, 2022.

[3] Kavishe, A. M., *Exploring the Experience of Cyberstalking among Female Students in Tanzanian Universities: A Case Study of the University of Dar es Salaam*, University of KwaZulu-Natal, Durban, South Africa, https://ukzn-dspace.ukzn.ac.za/handle/10413/20010, 2021.

[4] Msaki, L., *Assessment of the Challenges on E-Commerce Engagement: The Case of Selected Traditional Retailers,* Mzumbe University, 2019.

[5] Oreku, G. S., *A Rule-based Approach for Resolving Cybercrime in Financial Institutions: The Tanzania case*, Huria Journal, **27**(1): p. 93–114, 2020.

[6] Mambina, I. S., Ndibwile, J. D., & Michael, K. F., *Classifying Swahili Smishing Attacks for Mobile Money Users: A Machine-Learning Approach*, IEEE Access, **10**: p. 83061–83074, 2022.

[7] Ndibwile, J. D., Luhanga, E. T., Fall, D., Miyamoto, D., & Kadobayashi, Y., *A comparative study of smartphone-user security perception and preference towards redesigned security notifications,* AfriCHI '18: Second African Conference for Human Computer Interaction: Thriving Communities, p. 1–6, 2018.

[8] Bishel, E., *Africa, China, and the Development of Digital Infrastructure Governance: A Case Study of Ghana and Tanzania*, Central European University School of Public Policy, 2022.

[9] Panga, R. C. T., Marwa, J., & Ndibwile, J. D., *A Game or Notes? The Use of a Customized Mobile Game to Improve Teenagers' Phishing Knowledge, Case of Tanzania*. Journal of Cybersecurity and Privacy, **2**(3): p. 466–489, 2022.

[10] A. N. Ntigwigwa, *Factors that contribute to Cybercrime in Mobile Money Services in Tanzania: A Case of Kibaha Town*, Mzumbe University, 2019.

[11] E. Mwabukojo, *Technology Transfer Strategy: A Neglected Approach in Tanzania*, Munich Pers. RePEc Arch., 2020.

[12] KnowBe4®, *Virtual Risk Officer (VRO) and Risk Score Guide*, https://support.knowbe4.com/hc/en-us/articles/360001358728-Virtual-Risk-Officer-VRO-and-Risk-Score-Guide, 2018.

[13] Tessian®, *Tessian Risk Hub: Email Security and Risk Management, Made Easy*, https://www.tessian.com/resources/product-datasheet-human-layer-risk-hub/, 2021.

[14] Yang, R., Zheng, K., Wu, B., Li, D., Wang, Z., & Wang, X., *Predicting User Susceptibility to Phishing Based on Multidimensional Features*, Computational Intelligence and Neuroscience, 2022.

[15] Affinity IT Security Services, *Measuring Phishing Risk*, https://affinity-it-security.com/measuring-phishing-risk/, 2019.

[16] LexisNexis® Risk Solutions, *the critical role of email risk scoring in fraud prevention*. https://risk.lexisnexis.com/global/en/insights-resources/white-paper/email-risk-scoring-for-fraud-prevention, 2017.

[17] Steves, M., Greene, K., & Theofanos, M., C*ategorizing human phishing difficulty: a Phish Scale*, Journal of Cybersecurity, **6**(1): p. 1–16, 2020.

[18] Z. Yang, C. Qiao, W. Kan, and J. Qiu, *Phishing Email Detection Based on Hybrid Features, in IOP Conference Series: Earth and Environmental Science*, **252**(4): p. 1-10, 2019.

[19] Y. Fang, C. Zhang, C. Huang, L. Liu, and Y. Yang, *Phishing Email Detection Using Improved RCNN Model with Multilevel Vectors and Attention Mechanism*, IEEE Access, **7**: p. 56329–56340, 2019.

[20] L. Halgaš, I. Agrafiotis, and J. R. C. Nurse, *Catching the Phish: Detecting Phishing Attacks Using Recurrent Neural Networks (RNNs)*, Information Security Applications: 20th International Conference, p. 219–233, 2019.

[21] O. Olayemi, *Text Analysis and Machine Learning Approach to Phished Email Detection*, Int. J. Comput. Appl., **182**(36): p. 11–16, 2019.

[22] E. Castillo, S. Dhaduvai, P. Liu, K.-S. Thakur, A. Dalton, and T. Strzalkowski, *Email Threat Detection Using Distinct Neural Network Approaches*, Workshop on Social Threats in Online Conversations: Understanding and Management, p. 48–55, 2020.

[23] Y. Lee, J. Saxe, and R. Harang, *CATBERT: Context-Aware Tiny BERT for Detecting Social Engineering Emails*, KDD '21 Workshop on AI-enabled Cybersecurity Analytics, 2021.

[24] P. Bountakas, K. Koutroumpouchos, and C. Xenakis, *A Comparison of Natural Language Processing and Machine Learning Methods for Phishing Email Detection*, 16th International Conference on Availability, Reliability and Security, 2021.

[25] L. Franchina, S. Ferracci, and F. Palmaro, *Detecting phishing e-mails using text mining and features analysis*, IEEE International Conference on Computational Intelligence and Computing Research, p. 106–119, 2021.

[26] F. Salahdine, Z. El Mrabet, and N. Kaabouch, *Phishing Attacks Detection: A Machine Learning-Based Approach*, IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), p. 0250–0255, 2021.

[27] D. S. Ahmed, H. A. A. A. Allah, and I. Abbas, *Effective Phishing Emails Detection Method*, Turkish J. Comput. Math. Educ., **12**(14): p. 4898–4904, 2021.

[28] N. Noah, A. Tayachew, S. Ryan, and S. Das, *PhisherCop: Developing an NLP-Based Automated Tool for Phishing Detection*, HFES 66th International Annual Meeting, p. 2093- 2097, 2022.