CEUR-WS.org/Vol-3646/Short_5.pdf

The Method for Determining the Degree of Suspiciousness of a Phishing Url

Serhii Toliupa, Serhii Buchyk, Anastasiiya Shabanova and Oleksandr Buchyk

Taras Shevchenko National University of Kyiv, 60 Volodymyrska St., Kyiv, 01033, Ukraine

Abstract

In the context of rapidly outpacing the influx of phishing attacks in the digital environment, three key methods of phishing detection are considered: the Levenshtein distance in the similarity formula and the cosine similarity and Jaccard similarity algorithms. The paper examines both theoretically described general types of phishing and technological aspects of applying these mathematical methods, taking into account the importance of a balanced approach. It was found that each method has its advantages and limitations, and the key aspect is to find the optimal combination of human expertise and technological tools to effectively counter the ever-growing threat of phishing.

Keywords [1]

Phishing, Levenshtein distance, cosine similarity, Jaccard similarity, threat detection, digital security.

1. Introduction

In the era of high online activity in the modern digital world, addressing network security is becoming an extremely important and challenging task. The rapid development of technology is leading to an increase in threats, with phishing attacks coming to the fore. Phishing, defined as a type of deceptive scheme aimed at extracting confidential information such as passwords, banking data or personal identifiers, is becoming a key aspect of cybersecurity and remains a serious threat to network users.

According to the Incident Response Threat Summary report [1], such attacks have increased by 30% and are characterized by the main type of interference to obtain sensitive information. The healthcare sector is particularly vulnerable, accounting for 22% of all incidents, and compromised credentials are used in nearly 40% of cases. This underscores the urgent need to strengthen security measures and develop effective strategies to counter phishing attacks. Responding to these threats requires an integrated approach that encompasses technical, social, and educational measures to detect and prevent malicious activity.It should be noted that at the outbreak of the war in Ukraine in 2022, there was a significant increase in phishing attacks. In the first quarter of that year, fraudulent websites appeared that redirected funds under the guise of humanitarian aid for Ukraine. These resources referred to well-known individuals and charitable organizations, offering to make donations in cryptocurrency, which in itself raises suspicion (Figure 1).

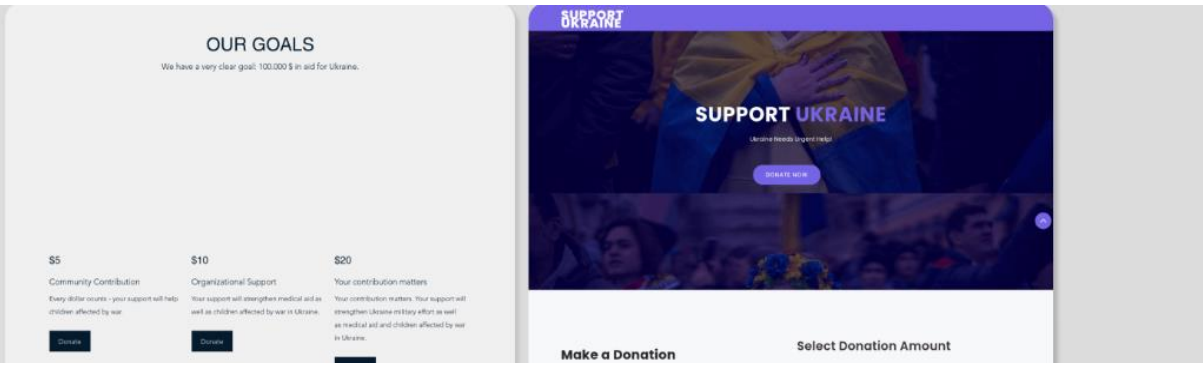
Throughout the year, investigations were conducted to analyze links in emails [2] aimed specifically at English-speaking users, offering to transfer money to help the affected Ukrainians using a direct mechanism of transactions on bitcoin wallets, as it is more difficult to identify the recipient in cryptocurrency than in bank transactions. A characteristic difference is that previously blackmail prevailed in spam demanding payment in cryptocurrency, while now the attackers have started collecting bitcoins for charity (see Figure 2). Since the success of phishing is determined by the scale of its audience, an attack on supporters of the occupation side was developed to reach more victims of the attack: in early July, a mailing of 300,000 emails was blocked, where fraudsters asked to help a Russian millionaire invest money to avoid sanctions (see Figure 3). Another mailing said that the European Commission had decided to distribute a fund created by Russian oligarchs, and the recipient of the letter could become the happy owner of a share of these funds.

Information Technology and Implementation (IT&I-2023), November 20-21, 2023, Kyiv, Ukraine

EMAIL tolupa@i.ua (A. 1); buchyk@knu.ua (A. 2); nastiash.2003@gmail.com (A. 3); alex8sbu@gmail.com (A. 4)
ORCID: 0000-0002-1919-9174 (A. 1); 0000-0003-0892-3494 (A. 2); 0009-0008-4962-569X (A. 3); 0000-0001-7102-2176 (A. 4)

© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073



OUR GOALS

We have a very clear goal: 100.000 $ in aid for Ukraine.

$5

Community Contribution

Every dollar counts - your support will help children affected by war

Donate

$10

Organizational Support

Your support will strengthen medical aid as well as children affected by war in Ukraine.

Donate

$20

Your contribution matters

Your contribution matters. Your support will strengthen Ukraine military effort as well as medical aid and children affected by war in Ukraine.

SUPPORT UKRAINE

SUPPORT UKRAINE

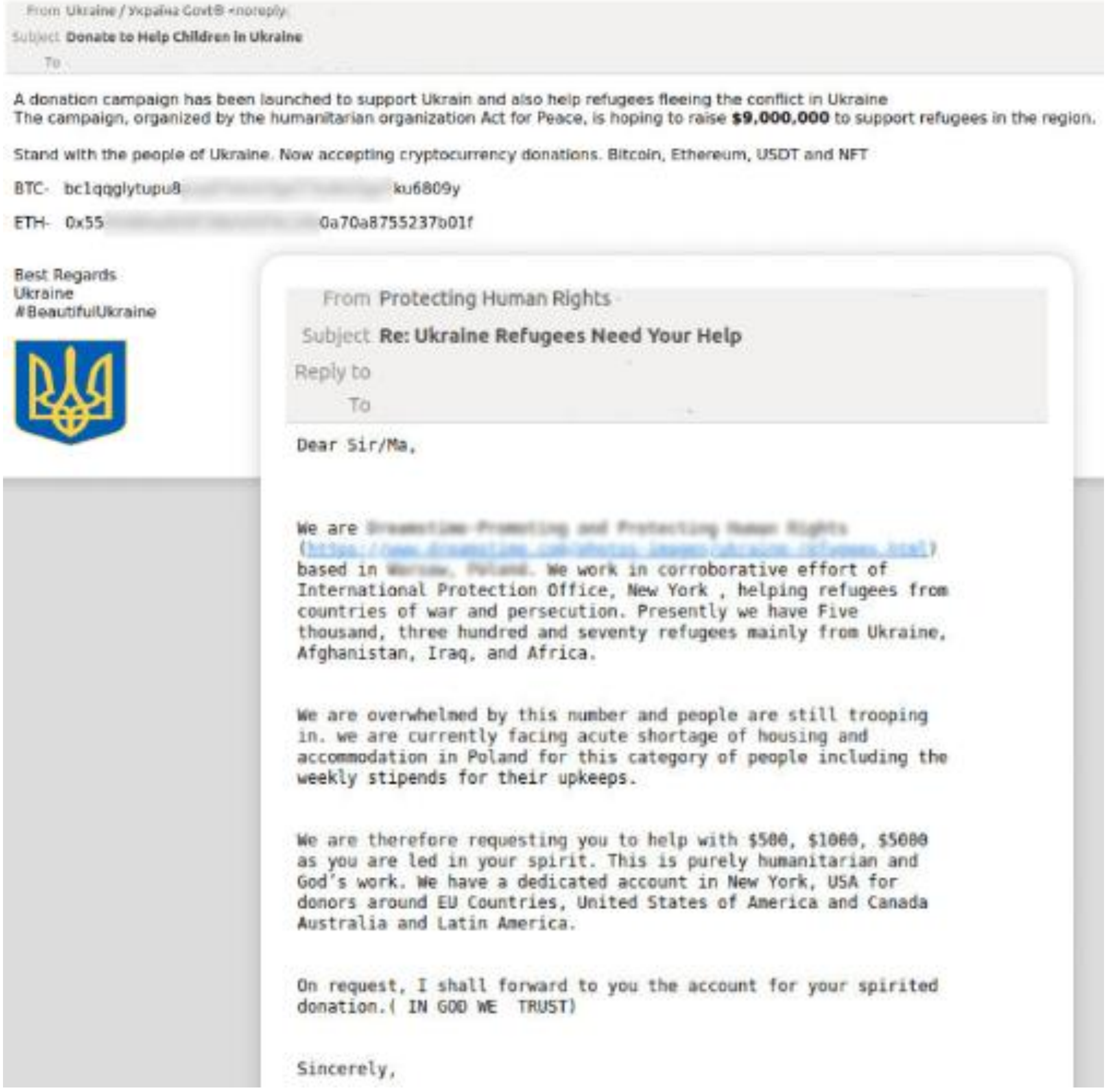
Ukraine Needs Urgent Help!

DONATE NOW

Make a Donation

Select Donation Amount

Figure 1: An example of a phishing form



From Ukraine / Україна Govt® <noreply

Subject **Donate to Help Children in Ukraine**

To

A donation campaign has been launched to support Ukrain and also help refugees fleeing the conflict in Ukraine
The campaign, organized by the humanitarian organization Act for Peace, is hoping to raise **$9,000,000** to support refugees in the region.

Stand with the people of Ukraine. Now accepting cryptocurrency donations. Bitcoin, Ethereum, USDT and NFT

BTC- bc1qqglytupu8[redacted]ku6809y

ETH- 0x55[redacted]0a70a8755237b01f

Best Regards
Ukraine
#BeautifulUkraine

From Protecting Human Rights

Subject **Re: Ukraine Refugees Need Your Help**

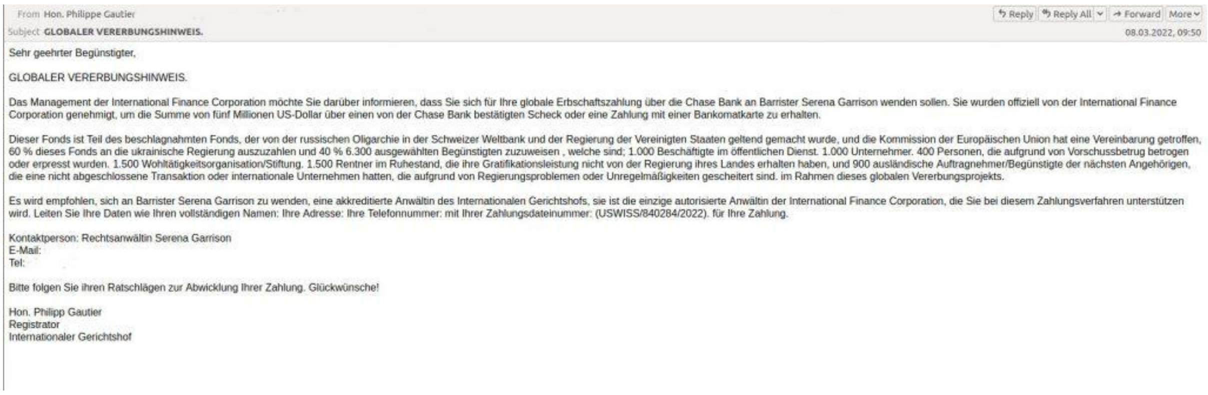
Reply to

To

```
Dear Sir/Ma,


We are [redacted]
([redacted])
based in [redacted]. We work in corroborative effort of
International Protection Office, New York , helping refugees from
countries of war and persecution. Presently we have Five
thousand, three hundred and seventy refugees mainly from Ukraine,
Afghanistan, Iraq, and Africa.


We are overwhelmed by this number and people are still trooping
in. we are currently facing acute shortage of housing and
accommodation in Poland for this category of people including the
weekly stipends for their upkeeps.


We are therefore requesting you to help with $500, $1000, $5000
as you are led in your spirit. This is purely humanitarian and
God's work. We have a dedicated account in New York, USA for
donors around EU Countries, United States of America and Canada
Australia and Latin America.


On request, I shall forward to you the account for your spirited
donation.( IN GOD WE  TRUST)


Sincerely,
```

Figure 2: An example of a phishing email

To summarize, in 2022, the share of spam from Russia continued to grow, from 24.77% to 29.82%. Germany (5.19%) swapped places with mainland China (14.00%), whose share increased by 5.27 percentage points. The third place went to the United States (10.71%). The Netherlands remained in fifth place (3.70%). Japan (3.25%) and Brazil (3.18%) took the sixth and seventh places, with their shares increasing by 0.89 and 3.77 percentage points, respectively. This is followed by the United Kingdom (2.44%), France (2.27%), and India (1.82%) (Figure 4). In 2022, email anti-virus programs detected about 166,187,118 malicious attachments in emails, an increase of 18 million compared to the previous year - the highest number of detections in February, March, and June 2022, which is proof of the decisive impact on the transformation in cyberspace, confirming the close connection between external events and characteristic changes in the digital environment (Fig. 5).



From Hon. Philippe Gautier
Subject GLOBALER VERERBUNGSHINWEIS.
08.03.2022, 09:50

Sehr geehrter Begünstigter,

GLOBALER VERERBUNGSHINWEIS.

Das Management der International Finance Corporation möchte Sie darüber informieren, dass Sie sich für Ihre globale Erbschaftszahlung über die Chase Bank an Barrister Serena Garrison wenden sollen. Sie wurden offiziell von der International Finance Corporation genehmigt, um die Summe von fünf Millionen US-Dollar über einen von der Chase Bank bestätigten Scheck oder eine Zahlung mit einer Bankomatkarte zu erhalten.

Dieser Fonds ist Teil des beschlagnahmten Fonds, der von der russischen Oligarchie in der Schweizer Weltbank und der Regierung der Vereinigten Staaten geltend gemacht wurde, und die Kommission der Europäischen Union hat eine Vereinbarung getroffen, 60 % dieses Fonds an die ukrainische Regierung auszuzahlen und 40 % 6.300 ausgewählten Begünstigten zuzuweisen , welche sind; 1.000 Beschäftigte im öffentlichen Dienst. 1.000 Unternehmer. 400 Personen, die aufgrund von Vorschussbetrug betrogen oder erpresst wurden. 1.500 Wohltätigkeitsorganisation/Stiftung. 1.500 Rentner im Ruhestand, die ihre Gratifikationsleistung nicht von der Regierung ihres Landes erhalten haben, und 900 ausländische Auftragnehmer/Begünstigte der nächsten Angehörigen, die eine nicht abgeschlossene Transaktion oder internationale Unternehmen hatten, die aufgrund von Regierungsproblemen oder Unregelmäßigkeiten gescheitert sind. im Rahmen dieses globalen Vererbungsprojekts.

Es wird empfohlen, sich an Barrister Serena Garrison zu wenden, eine akkreditierte Anwältin des Internationalen Gerichtshofs, sie ist die einzige autorisierte Anwältin der International Finance Corporation, die Sie bei diesem Zahlungsverfahren unterstützen wird. Leiten Sie Ihre Daten wie Ihren vollständigen Namen: Ihre Adresse: Ihre Telefonnummer: mit Ihrer Zahlungsdateinummer: (USWISS/840284/2022). für Ihre Zahlung.

Kontaktperson: Rechtsanwältin Serena Garrison
E-Mail:
Tel:

Bitte folgen Sie ihren Ratschlägen zur Abwicklung Ihrer Zahlung. Glückwünsche!

Hon. Philipp Gautier
Registrator
Internationaler Gerichtshof

Figure 3: An example of a phishing email

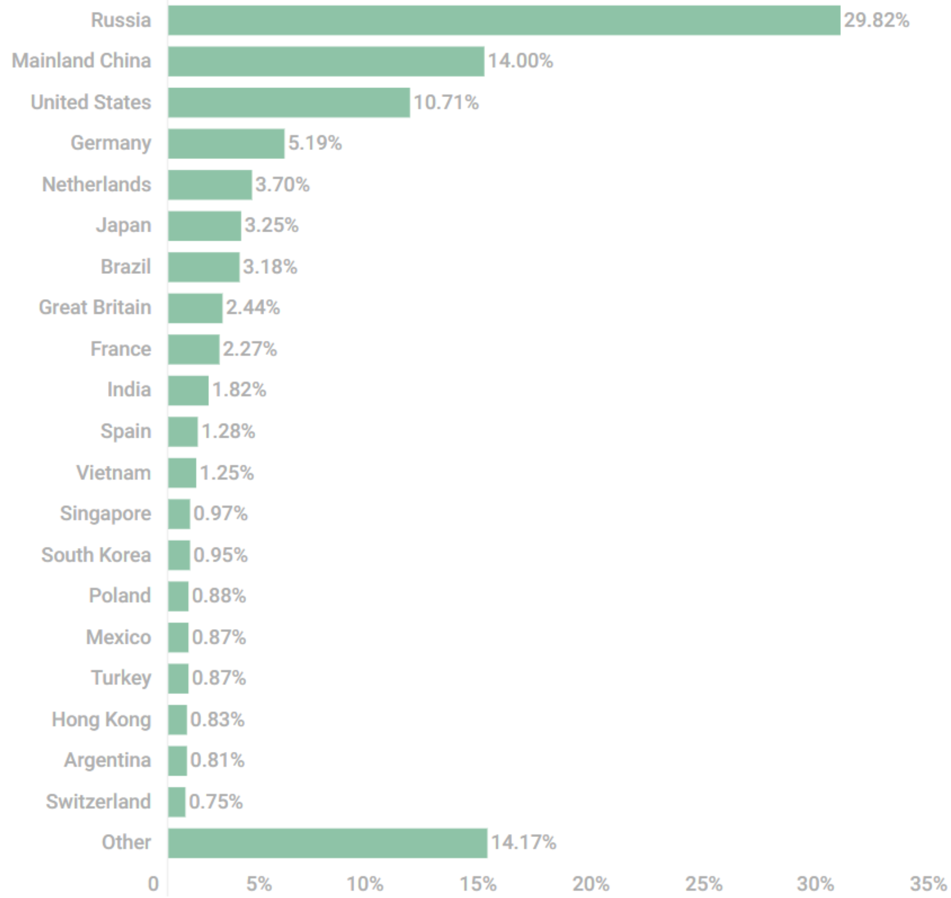


Figure 4: TOP 20 countries and territories - sources of spam

2. Types of phishing

In today's information world, there is a variety of types of phishing, each targeting different aspects of the consequences and using different methods to influence the potential victim [3]. Some of them are based on mass mailings or standard attacks, while others are refining guaranteed designed and personalized approaches to make sure they work on their target audience. Phishing as a modern cyber threat is growing in importance, requiring a deeper understanding of its various aspects. This threat can take the form of emails, text messages, social media, and even phone calls (vishing), so a deeper understanding of different types of phishing is important to develop effective strategies to counter these threats and ensure cybersecurity in the online environment [4].

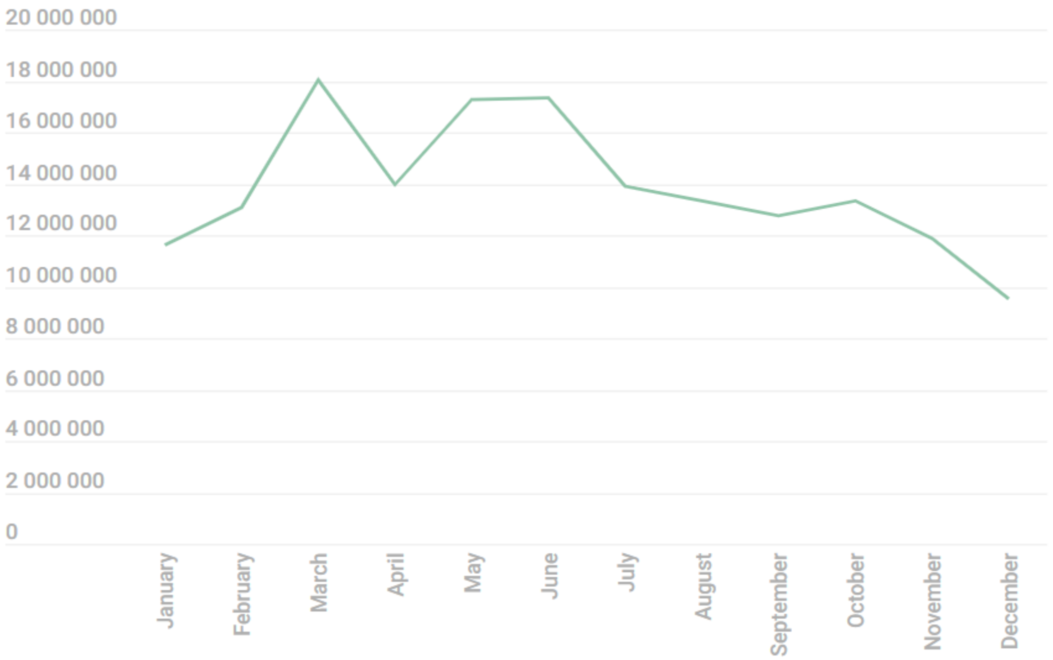


Figure 5. Number of malicious email attachments detected by the antivirus

Targeted phishing is a prominent cyberattack method that targets a specific individual, organization, or company. In this type of attack, attackers use important information about a potential victim, such as name, job title, professional contact details, and other internal aspects, to create personalized and credible messages. The characteristics of this attack method are the use of personalization, in-depth analysis and selection of the necessary information about the potential victim, allowing attackers to create fictitious but plausible situations and messages specifically adapted to the nature and activities of the target person. This allows for the acquisition of sensitive information such as passwords, financial data, commercial and other confidential data. Attackers may try to look like employees of the organization or other trusted individuals, using social engineering and psychological techniques to influence the potential victim. Specialized phishing is particularly dangerous due to its high level of personalization and difficulty in detection for the general user. To avoid falling under a specialized phishing attack, it is important to be extra cautious and attentive, checking the messages you receive, especially if they are unusual or request confidential information.

Vishing, short for voice phishing, involves the use of the telephone to deceive and manipulate people into giving up confidential information such as passwords, credit card numbers, and other personal data. The characteristic features of vishing are the use of voice messages or direct phone calls, during which attackers pretend to be representatives of banks, companies or government agencies. They try to arouse certain emotions in the victim, put pressure on him or her, and encourage him or her to provide personal information. Vishing can take many forms, including imitating customer service of banks, companies, and even government agencies, which makes the attack particularly dangerous because of the ability to convince the victim that the call is legitimate and provide important information. To prevent falling victim to a phishing attack, it is important to be cautious and realize that legitimate organizations will never request confidential information over the phone from unknown or questionable sources.

Spear phishing is a term that originates from the combination of the words "SMS" and "phishing". The term refers to a specific type of cyberattack in which attackers use text messages (SMS) to deceive and manipulate individuals. The characteristic features of spear phishing are the use of text messages in which attackers try to look like trusted sources, such as banks, companies, or government agencies. They send short and persuasive messages that encourage victims to provide personal information or visit malicious websites. Spoofing can take many forms, including imitating official messages from banks about the danger or providing possible lottery winnings. This makes the attack particularly dangerous due to the ability of the attackers to convince the victim of the legitimacy of the message and to cause emotional impact. To prevent falling victim to a smishing attack, it is important to be cautious and attentive to the text messages you receive, especially if they come from unknown or suspicious sources. Victims should realize that legitimate organizations will never request confidential information via SMS.

Pharming is a form of cyberattack aimed at redirecting network traffic in order to obtain sensitive information from users. This method of attack is particularly dangerous because it directs traffic to fake websites where users inadvertently disclose their confidential information. One of the key characteristics of pharming is the spoofing of DNS queries that users receive when they try to access a particular website. This allows them to redirect traffic to their fake resources, where they carry out fraudulent transactions. Farming takes many forms, including DNS farming, IP address routing manipulation, malware, etc. This makes the attack extremely difficult to detect and avoid for the average user. To protect yourself from pharming attacks, it is crucial to have up-to-date anti-virus software, use reliable firewalls, and be especially careful when accessing websites and entering sensitive information.

Spear phishing is a phishing attack aimed at influencers and executives of large organizations. The main goal of this type of attack is to obtain confidential information that is important to the organization. Attackers who carry out warehousing use specific techniques and social engineering to hack into and gain access to critical data. They can use sophisticated methods, including phishing, malware distribution, and specialized attacks on security systems. The aim of a vectoring strategy is to cause the maximum possible damage to an organization, including financial loss, data breaches, and reputational damage. To prevent warehousing attacks, it is important to implement effective measures to strengthen cybersecurity, train staff in preventive measures, and use appropriate technical means to protect important data and information systems.

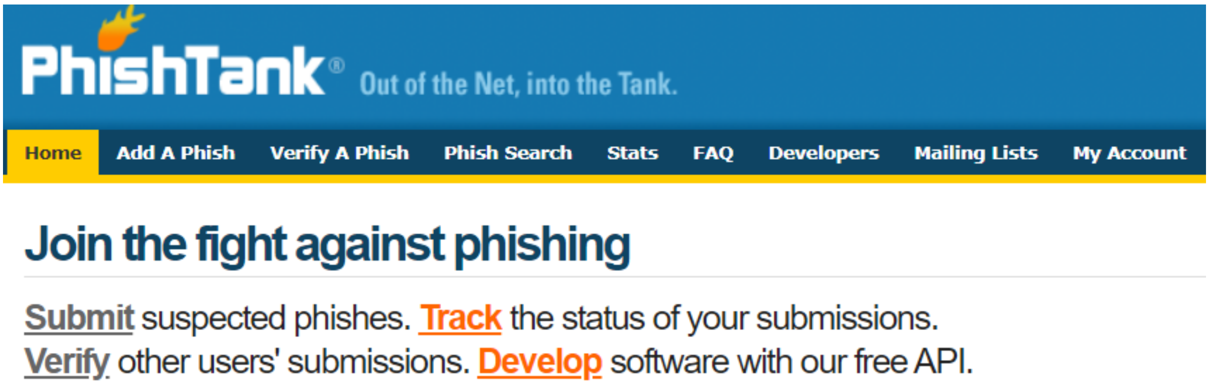
Cloning is a cyberattack where attackers create a fake copy of a valid message or website to deceive recipients. This type of attack is aimed at obtaining confidential information, which can include passwords, banking information, and personal information. Fake messages usually look as authentic as possible, including logos and graphics from original sources to appear as similar as possible to the original. The goal is to get potential victims to disclose their confidential information. This can often lead to financial losses and privacy violations. The trick to cloning is to use valid communications or websites and change only certain elements to achieve malicious purposes. It is important to check the official websites of organizations and analyze the details of communications before disclosing personal data [5]. Unfortunately, fake messages try to imitate original sources as authentically as possible, including their logos and graphic design, in order to create an impression of authenticity. The main goal is to try to obtain important confidential information from potential victims. In the event of a successful attack, this can lead to serious financial losses and privacy violations. Thus, there should be a heightened awareness and caution in relation to the messages received, especially given the risk they pose to personal data privacy. In the modern period, there is an initiative focused on countering such forms of fraud. One example of such measures is the PhishTank platform (see Figure 6), which is a community where users can work together to detect and block phishing attacks by leaving links to phishing sites. This helps to improve the level of protection against cybercriminals and ensures overall security on the global Internet, contributing to the protection of the confidentiality of users' personal data [6].

3. General algorithm for determining the degree of URL suspiciousness based on the current Phishtank page

Fuzzy logic is a mathematical approach for dealing with fuzzy and uncertain concepts, so it effectively manages the ambiguity of this type of threat in phishing detection. It can be used to expand the understanding of the similarity between phishing and legitimate elements, especially when considering ambiguities in text attributes such as URLs or headers [7]. The use of vaguely described values allows us to determine the degree of suspicion of a link - in our case, the main basis for analysis and comparison will be the current page with a list of new phishing URLs on the PhishTank website. The cosine similarity measure, used in the context of phishing attack detection, helps to determine the degree of similarity between text fragments that are typical of phishing schemes by considering the relationship in vector forms. Also, fuzzy logic can be used to analyze the mutual intersection of character sets or phishing patterns, where the Jaccard coefficient acts as a fuzzy measure of similarity between sets, helping to avoid hard dividing lines between phishing and legitimate elements.

3.1. Let's consider the first method, based on the Levenshtein distance [8], which is able to determine the minimum number of edits required to transform one string into another. These edits include insertions (adding characters), deletions (removing characters), and replacements (replacing one character with another). In the Damerau-Levenstein algorithm used to calculate this distance, the recurrent formula (1) is as follows (cost is 0 if the characters are equal and 1 if the characters are different).

Translated with www.DeepL.com/Translator (free version)



Join the fight against phishing

Submit suspected phishes. Track the status of your submissions.
Verify other users' submissions. Develop software with our free API.

Figure 6. PhishTank website appearance

$$matrix[i][j] = \min \begin{pmatrix} matrix[i-1][j]+1: \text{deletion} \\ matrix[i][j-1]+1: \text{insertion} \\ matrix[i-1][j-1]+\cos t: \text{replacement} \end{pmatrix}$$

Formula 1: Damerau-Levenstein algorithm

The training strategy involves the creation of a matrix where the value in each cell represents the distance between rows and strings. The maximum length specified in the code is used to normalize the Levenshtein distance to obtain a value between 0 and 1, and the coefficient indicates that the contribution to the overall similarity score is only 40%, allowing other methods to also influence the comparison, as this method is chosen as the base method due to the more accurate results in Listing 1:

```
def calculate_similarity_score(template, url):
levenshtein_dist = levenshtein_distance(template, url)
max_length = maximum_length(template, url)

# Applying the specified influence factors
similarity_score = 0.4 * (1 - levenshtein_dist / max_length )
```

Listing 1: Similarity evaluation

It is particularly effective for detecting similarities between URLs, where small differences can point to potentially suspicious schemes, but it is important to consider the limitations of this method - semantic understanding and URL structure, which are key aspects in similarity analysis. The method is also sensitive to character order and can be time-consuming to process large amounts of data: if one URL is significantly longer or shorter than another, it can affect the result, regardless of the actual similarity (it does not distinguish between different parts of the URL (domain, path, parameters), which can lead to it treating different URLs with similar structures as less similar). Compared to other methods, such as Cosine Similarity and Jaccard Similarity, which take into account the similarity of sets and vectors, Similarity Score may be less accurate in certain cases where it is important to take into account the semantic content and structure of the URL.

3.2. Cosine Similarity [9] is a method of measuring the similarity between two vectors in a vector space. In the context of textual comparison, using ASCII character codes, text strings, in this case url, are converted into numerical vectors. The similarity itself is calculated as the cosine of the angle between these vectors, which allows you to ignore the absolute size of the vectors and focus on their direction. A factor of 0.3 in the code is used to ensure that cosine similarity contributes significantly, but not dominantly, to the overall score, as the choice of this particular factor is the result of empirical research and more generally experimentation during the practical application testing of Listing 2.

This method is effective in general for comparing text documents or words because it takes into account semantic similarity and context. Compared to the method based on Levenshtein distance, which measures the "order of precedence" between two strings, cosine similarity allows for semantic relationship and context, which is important for textual information. However, it can be vulnerable to short links, and does not always reflect the specifics of the particular operations that need to be performed to convert one text string to another, which is what the previous method does well.

```
# Cosine Similarity
min_length = min(len(template), len(url))
vector_a = [ord(char) for char in template[:min_length]]
vector_b = [ord(char) for char in url[:min_length]]
cosine_similarity = 0.3 * (dot(vector_a, vector_b) / (norm(vector_a) *
norm(vector_b)))
```

Listing 2: Cosine similarity

3.3. The Jaccard Similarity method [10] is a measure of the distance between sets, described as the number of common elements divided by the total number of unique elements. In the code under consideration, this method is used to determine the similarity between a pattern word and a URL. In particular, the weight of Jaccard similarity in the calculations is 0.3, which indicates its importance compared to other methods on a par with the previous one, since this method has an absolutely similar but reverse methodological characteristic (see Listing 2):

```
# Jaccard Similarity
set_a = set(template)
set_b = set(url)
jaccard_similarity = 0.3 * (len(set_a.intersection(set_b)) /
len(set_a.union(set_b)))
```

Listing 3: Jacquard's similarities

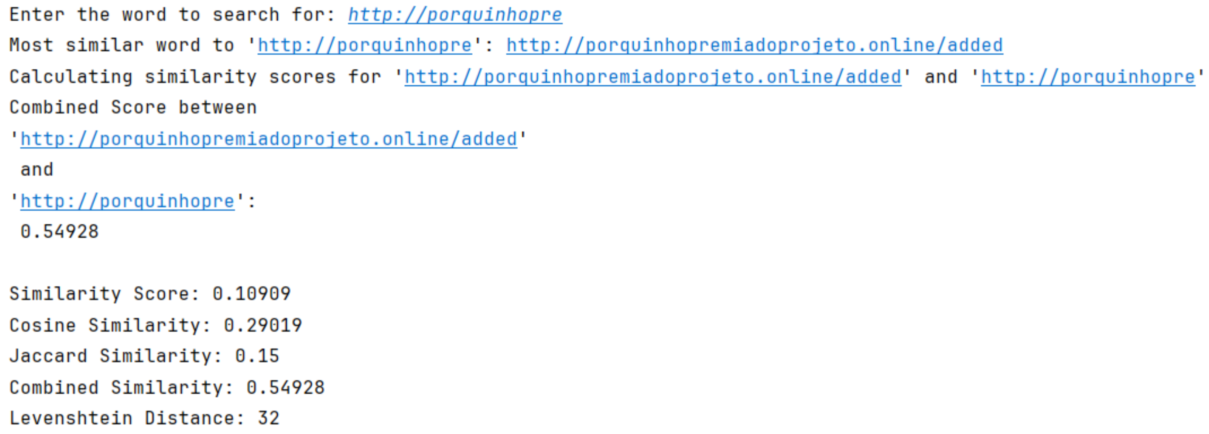
The key important feature of Jaccard similarity is that with just a few operations on sets, an effective similarity measure can be obtained, and the method is effective for comparing large sets, making it applicable to large amounts of data. When comparing texts or sets where the order is not important, this is an advantage over the Levenshtein distance, which takes into account the character order. In general, the above practically-based distribution of coefficients is preferable to the mathematical expectation of their values, as it allows assigning weight to different comparison methods according to their effectiveness and relevance to a particular context. A statistical average uses the same weighting for all methods, which can lead to an unfair consideration of the contribution of any one chosen methodology. This measure is classified as a weighted weighting strategy that allows for the importance of each approach to be noted in comparison to the others, which can be represented by denoting the effect of each method as E1, E2, E3 and their weights as W1, W2, W3, the overall effectiveness can be expressed as Formula 2:

$$E_{\text{general}} = W_1 \cdot E_1 + W_2 \cdot E_2 + W_3 \cdot E_3$$

Formula 2: Overall efficiency

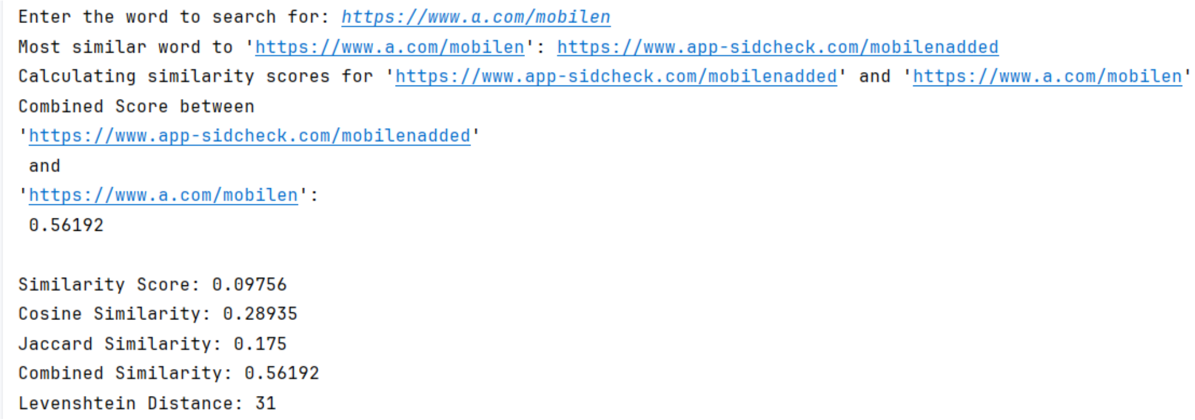
3.4. This context of mathematical analytics is determined to be of great value in tracking and investigating the most recently added URLs on the PhishTank website. Such an approach can help to recognize and avoid potentially dangerous sites in time, focusing on the underlying mechanism of operation: the above algorithm of actions began with obtaining the HTML content of the PhishTank page using the get_page_content function, after which beautifulsoup was used to parse the HTML and create the phishtank_soup object, which allows convenient interaction with HTML elements on the site itself. The find_all function was used to find HTML elements with the class that contain similar URLs relative to the similarity coefficient introduced to the almost equal level, which were stored and sequentially processed in the phishtank_urls list. Then, the user-entered URL (base_url) is algorithmically compared with each URL in the phishtank_urls list, and the level of similarity between the two URLs is calculated using mathematical methods. If the similarity is greater than the current maximum, the most_similar_url is changed. The last step is to calculate the above three link analysis methodologies together with the addition of the "/added" security element to prevent accidental URL clicks. Examples of the results of using this mechanism are shown in Figures 7-8, the general detection algorithm is shown in Figure 9.

This algorithm is designed to perform a comparative analysis of the content of a web page based on a given URL and a user word, taking into account similarity factors. The process includes user input of the URL and the word to be compared, fetching the page content via an HTTP request, parsing the HTML using BeautifulSoup to obtain the textual content. The entered word is then compared with all the words from the page, and the most similar word is identified as a "pattern" to calculate various similarity measures such as Levenshtein distance, cosine similarity, and Jaccard index. Weighting factors are applied to determine the influence of each factor on the final similarity score. As a result, the algorithm generates comprehensive similarity scores, taking into account various aspects of the comparison, and displays the results directly to the user.



```
Enter the word to search for: http://porquinhopre
Most similar word to 'http://porquinhopre': http://porquinhopremiadoprojeto.online/added
Calculating similarity scores for 'http://porquinhopremiadoprojeto.online/added' and 'http://porquinhopre'
Combined Score between
'http://porquinhopremiadoprojeto.online/added'
 and
'http://porquinhopre':
 0.54928

Similarity Score: 0.10909
Cosine Similarity: 0.29019
Jaccard Similarity: 0.15
Combined Similarity: 0.54928
Levenshtein Distance: 32
```

Figure 7. An example of the first result of using this mechanism



```
Enter the word to search for: https://www.a.com/mobilen
Most similar word to 'https://www.a.com/mobilen': https://www.app-sidcheck.com/mobilenadded
Calculating similarity scores for 'https://www.app-sidcheck.com/mobilenadded' and 'https://www.a.com/mobilen'
Combined Score between
'https://www.app-sidcheck.com/mobilenadded'
 and
'https://www.a.com/mobilen':
 0.56192

Similarity Score: 0.09756
Cosine Similarity: 0.28935
Jaccard Similarity: 0.175
Combined Similarity: 0.56192
Levenshtein Distance: 31
```

Figure 8. An example of the second result of using this mechanism

4. Conclusions

The key functionality of this mechanism lies in the successful implementation of the mathematical approach and its practical application in the field of detecting and determining the degree of suspiciousness of phishing links. It has been theoretically and practically proven that any textual information values can be considered as an array, a vector representation, or a set of elements. These methods will be effective in any derivative or initial, but more developed, code representations, since they are universal due to the ability of versatile algorithmic adjustment and the metamathematical basis of laws and axioms. At the same time, one should rely on direct advantages and situational necessities of application due to irrelevance of use at critical points, as a last resort - weigh them with the help of efficiency proportionality coefficients or categorize the information presented for research, and then correlate them with the best application of one of the methods in the given case. The structure of general actions can be further developed through additional content and further checks by other technical or physical methods of processing the information flow in the form of textual information, or by including it in the sequence of checks of in-depth analysis at the domain level and transition to the URL. Comparison of the code execution examples at the selected addresses shows that the mechanism is equally effective (about 0.5), despite the sequence or removal of a part of the link (subdomain removal) at different synthetic URL depths. Also, each similarity finding principle increased by a characteristic index, thus describing its stability under different actions, but the partial change method did not affect the quality of the results.

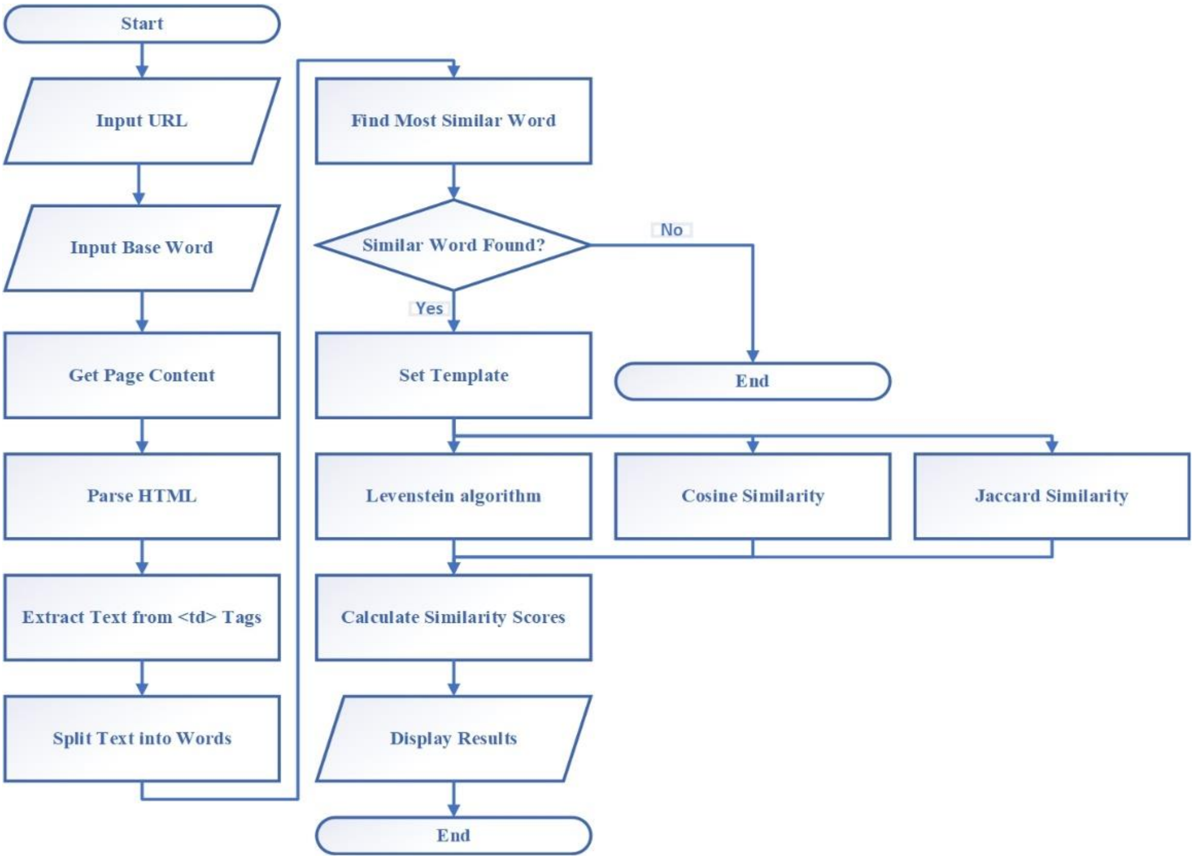


Figure 9: General detection algorithm

5. References

[1] "Spam and phishing in 2022". URL: https://securelist.com/spam-phishing-scam-report-2022/(date of access: 21.10.2023).

[2] "Email spoofing: how attackers impersonate legitimate senders". URL: https://securelist.com/email-spoofing-types/102703/(date of access: 21.10.2023).

[3] Types of phishing - What it is an how to prevent it. Everything About Online Reputation Management. URL: https://blog.reputationx.com/guest/whats-phishing (date of access: 21.10.2023).

[4] Phishing - what is it, its essence, definition, types and examples of phishing. URL: https://termin.in.ua/fishynh/(date of access: 21.10.2023).

[5] Buchyk S., Shutenko D., Toliupa S., (2022) Phishing Attacks Detection. IX International Scientific Conference "Information Technology and Implementation" (IT&I-2022), Workshop Proceedings, Kyiv, Ukraine, November 30 - December 02, 2022., Kyiv, Ukraine, pp. 193–201. URL: https://ceur-ws.org/Vol-3384/Short_7.pdf.

[6] PhishTank site. URL: https://www.phishtank.com/index.php.

[7] Guanrong Chen, Trung Tat Pham. ""Introduction to Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems" by Guanrong Chen, Trung Tat Pham" (2019). URL: "https://books.google.pl/books?id=cj0kNFa1ZZgC&printsec=frontcover&hl=pl#v=onepage&q&f=false".

[8] A.Y. Minyailo, V.A. Turchina . "Using Levenshtein distance for data similarity analysis" (2015). URL: https://pm-mm.dp.ua/index.php/pmmm/article/view/111/111.

[9] "Cosine similarity: How does it measure the similarity, Maths behind and usage in Python". URL: https://towardsdatascience.com/cosine-similarity-how-does-it-measure-the-similarity-maths-behind-and-usage-in-python-50ad30aad7db (date of access: 21.10.2023).

[10] "How to Calculate Jaccard Similarity in Python". URL: https://www.geeksforgeeks.org/how-to-calculate-jaccard-similarity-in-python/(date of access: 21.10.2023).