Pamukkale Univ Muh Bilim Derg, 29(5), 468-480, 2023

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

PhisherHunter: Module design for automatic detection of phishing websites and preventing user abuse

OltalamaAvcısı: Oltalama internet sitelerinin otomatik tespiti ve kullanıcı istismarının önüne geçilmesi için modül tasarımı

Samet GANAL[1*], *Ecir Uğur KUCUKSILLE*[2], *Mehmet Ali YALÇINKAYA*[3]

[1]HSBC, Cyber Security Management, Istanbul, Turkey.
sametganal@hotmail.com

[2]Department of Computer Engineering, Faculty of Engineering, Suleyman Demirel University, Isparta, Turkey.
ecirkucuksille@sdu.edu.tr

[3]Department of Computer Engineering, Faculty of Engineering and Architecture, Ahi Evran University, Kırşehir, Turkey.
mehmetyalcinkaya@ahievran.edu.t

Received/Geliş Tarihi: 05.06.2022 Revision/Düzeltme Tarihi: 17.09.2022 doi: 10.5505/pajes.2022.22470
Accepted/Kabul Tarihi: 22.10.2022 Research Article/Araştırma Makalesi

Abstract

One of the most common cyber-attacks that users encounter on the internet are phishing websites. In the attacks that are performed on phishing websites, real websites are duplicated and published on different domain names, and users are directed to these fake websites through various social engineering techniques. Through to the website to which users are directed, they transmit some personal and confidential data such as credit card, username-password details to attackers. In this study, the establishment of the infrastructure and content of phishing internet sites has been explained, a tool named PhisherHunter created, and four different methods have been developed so as to detect such websites. Through the examination of newly registered websites, which is the main detection method, a successful detection rate of 95.4% has been achieved. Three different methods have been used in the active defense part of the study. Firstly, the hosting company has been automatically determined to stop the publication of the phishing website and a notification has been sent with a success rate of 98%. As the second active defense method, the active honeypot technique has been developed. The active honeypot method aims to enter a marked information on the phishing website and to track this information on the real website. And as the last active defense method, the method of poisoning phishing websites by using fake data has been developed. It has been observed that poisoning methods by using the techniques of active honeypot and fake data have achieved a success of 92%.

Keywords: Phishing, Phishing website, Active defense.

Öz

Günümüz dünyasında bilgisayar ve mobil cihazların kullanımının yaygınlaşması internet kullanımının giderek artmasına neden olmaktadır. Kullanıcıların internet ortamında en çok karşılaştığı siber saldırılardan biri oltalama internet siteleridir. Oltalama internet siteleri üzerinden gerçekleştirilen saldırılarda, gerçek internet siteleri kopyalanıp farklı alan adları üzerinden yayın yapılmakta ve kullanıcılar bu sahte internet sitelerine çeşitli sosyal mühendislik teknikleriyle yönlendirilmektedir. Kullanıcılar yönlendirildikleri internet sitesine göre; kredi kartı, kullanıcı adı-şifre bilgileri gibi kişisel ve gizli verilerini saldırgana iletmiş olmaktadır.Bu çalışmada; oltalama internet sitelerinin altyapısının ve içeriğinin oluşturulması anlatılmış, bu tür internet sitelerini tespit etmede kullanılacak 4 farklı metot geliştirilmiştir. Ana tespit yöntemi olan yeni kayıtlı internet sitelerinin incelenmesi ile beraber %95.4'lük başarılı tespit oranına ulaşılmıştır. Çalışmanın aktif savunma kısmında üç farklı yöntem kullanılmıştır. İlk olarak oltalama internet sitesi yayınının durdurulması için yer sağlayıcı firma otomatik olarak tespit edilmiş ve %98 başarı oranıyla bildirim gönderilmiştir. İkinci aktif savunma yöntemi olarak aktif bal küpü (honeypot) tekniği geliştirilmiştir. Aktif bal küpü yöntemi oltalama internet sitesine işaretli bir bilgi girilmesi ve bu bilginin gerçek internet sitesinde takibini amaçlamaktadır. Bu yöntem ile saldırgana ait pek çok veri elde edilebilmektedir. Son aktif savunma yöntemi olarak, oltalama internet sitelerini sahte veriler ile zehirleme metodu geliştirilmiştir. Bu yöntem ile oltalama internet sitelerinin girdi alanları otomatik tespit edilmekte ve çok fazla sahte veri gönderilerek gerçek kullanıcı bilgilerinin saldırganlar tarafından ayırt edilmesi önlenmeye çalışılmıştır. Aktif bal küpü ve sahte veri ile zehirleme yöntemlerinin %92 başarı elde ettiği görülmüştür.

Anahtar kelimeler: Oltalama, Oltalama internet sitesi, Aktif savunma.

1 Introduction

The use of the Internet is becoming widespread day by day, and its fields of use and the number of its users are continuously increasing. The active participation of the users in the internet environment requires them to transfer their very critical information to this area, depending on the intended use. The biggest factor in the increasing trend of cybercrime is that attackers get high amount of financial gains. The loss of people, just in the USA, in 2019 on account of cyber-crimes is 3.5 billion dollars [10]. There are many attack techniques in today's cyber world. Phishing attacks carried out on websites is one of them. In this type of attack, attackers aim to obtain users' confidential information such as identity, login and credit card details [47]. In attacks performed through phishing websites, attackers duplicate real websites and publish them on different domains and wait for users to take the bait. Attackers expect users, through various channels, to be directed and to log in to

*Corresponding author/Yazışılan Yazar

phishing websites that they have prepared before. And when users enter this duplicated website, they log in without detecting an abnormality due to their familiar content, and transmit their information to attackers. By this way, attackers obtain the login information of users and achieve their goals [48].

Phishing attacks carried out of the internet sites are widely seen in Türkiye and cause users to incur huge financial and emotional damages. USOM which means in english CERT (Computer Emergency Response Team), authorized in terms of such kind of cyber-attacks, particularly in Türkiye and some private institutions detect phishing web sites, through various methods and users' denunciations, to ensure the internet security. According to the figures of CERT, in the period of one-year from July 2019 to July 2020, 17626 web sites that are used with the aim of phishing were detected in Türkiye [19]. Following the detection, the state authorities inform the hosting company where this phishing website is hosted and request that access of such site is blocked, and have access to such web site within the country blocked in the ISP (Internet Service Provider). However, since the application of all these protection methods is a relatively long process, attackers obtain a lot of users' information and transfer the material assets to themselves. Through blocking access of the used phishing website, the attackers activate a new phishing website and continue their attacks. Within this cycle carried out, a requirement for a tool that quickly and automatically detects malicious websites activated by attackers and develops an active defense mechanism. In this study, a module having the following features is developed based on the aforementioned requirement;

- Four different methods have been developed for the detection of phishing websites. The first method aims to detect phishing websites before they are active, and it makes inferences from the domain names of newly registered websites for this. The second and third methods inspect the twitter posts and google ads used by attackers to direct users to phishing websites. The fourth, and final, detection method is to receive information of web sites that are used for the purpose of phishing shared by the competent authority, CERT, in Türkiye,
- Potential phishing websites obtained as a result of detection methods are kept in a list. Access to domain names contained in this list is regularly checked, and if the requested result is successful (200), the content of the relevant website is checked. Based on the words in the web site content, the status of phishing publication is checked. Provided that the website is publishing phishing, it is removed from the checklist and transferred to the active defense part, but if it does not publish phishing, it is kept to control again,
- The active defense part against phishing websites consists of three parts. First of all, a whois query of a phishing website is carried out, the abuse address is separated from the returned result, and an e-mail is sent to this address stating that the relevant website serves cybercrime and that access to the relevant web site should be blocked. As the second method, active honeypot technique has been used and the phishing website was entered with a signed username-password combination. The poisoning method was used as the third and last active defense method. A method that browses the phishing website, finds input fields and writes and posts the necessary data in these fields has been developed. This process is carried out over SOCKS proxy IPs and thus any user image is provided. Thanks to all these methods, it is desired to make it much more difficult for attackers to access real users' data.

In the second chapter of the study, the results of the literature reviews are included, and in the third chapter of the study, attacks carried out of phishing websites, infrastructure preparation and directing users are discussed. And in the fourth chapter, the module detection methods and active defense methods that have been developed are explained under separate headings. Furthermore, under this heading, how all code blocks are made into a single piece and timed are explained as well. Finally, results and future plans are included in the fifth chapter.

2 Related work

Because the mentioned study is complicated and includes methods that serve many different fields, no exactly similar study has been found in the literature. There is a study that uses fuzzy logic, in detecting phishing websites, with the use of six different metrics that it compiles from the web site by controlling its content [1]. There are many different modules in a successful study that detects phishing web sites by controlling its content [20]. Phishing web sites were able to be detected by using the CCS information contained in web sites, [14]. In another original study, an application, called as Adaptive Neuro-Fuzzy Inference System (ANFIS) has been developed to detect phishing websites. According to this, it has been aimed to detect them by using texts, pictures and frames in internet web sites [2]. There is a study where the hyperlinks in the HTML source code are divided into 12 different categories and machine learning is applied, thus enabling detection [23]. Another study using hyperlinks also aimed to reduce the number of false detections with two different methods by verifying from search engines [24]. In another study using hyperlinks, "page linking" data was collected and an accurate determination was made using graph-theory approach [30]. In addition, there is a study that includes the development, distribution, taxonomy and comparison of phishing websites with other phishing websites from day zero [21]. There are studies in which various machine learning algorithms are used together, including random forest [RF], neural network [NN], bagging, support vector machine, Naïve Bayes and k-nearest neighbor, and thus achieving a high success rate [31]. There is a study on detecting phishing websites by using artificial intelligence algorithms of Meta-Learners and Extra-Trees Algorithm [32].

The phishing website detection has been tried by taking the logos and favicons in the website and searching them on Google [6],[7]. Another detection method carried out using a search engine is based on dividing domain names into pieces and searching them in the search engine and making inferences based on the returned results [18]. Except this, an application that aims to detect phishing website according to the results returned following the search of the data in the website in search engines has also been developed [15]. Detection methods including not only search engine but also intuitive approach and logistic regression have been developed [8]. In addition, in the study conducted by combining search engine

and heuristic analysis, a very high rate of accurate detection was achieved regardless of language [22]. With the convolutional neural network approach from URL text, almost 100% successful detection has been achieved, including fast and zero-day attacks [27]. CrawlPhish application automatically detects and categorizes client-side obfuscation used by known phishing websites [29].

Phishing internet sites have been tried to be determined by using various algorithms over ready data sets [5],[11],[17]. There are also studies that combine machine learning with the helical feature selection method and produce a new methodology in the detection of phishing websites [4]. There is a study aiming to detect phishing websites by using artificial intelligence algorithms including machine learning, deep learning and hybrid learning together with scenario-based techniques [25]. In a study utilizing 1278 user experiences, the "phishing funnel model" was developed and phishing websites were detected based on user movements [26].There is one study that performs the detection of phishing websites that are shared during direction process of users over Twitter [3]. There are also some methods that try to detect phishing websites by imitating human behaviors. One of them is trying to log in the login panel of the website, and suggests that, providing that the login is successful, the phishing website is detected [15]. Even though there are many studies aiming to detect phishing websites, there has been no study regarding the active defense methods. Thus, this study features the first one in the literature in terms of active defense.

3 Cyber attacks through phishing websites

The most important factor, for attackers, in attacks carried out over phishing websites is the high similarity of the phishing website, which they designed, to the real website in terms of such issues as domain name and content. Attackers try to make all details, as much as they can, similar to those in the original website, including the domain name that they will use to make this happen. Because the domain name is a unique value, the same name cannot be taken, and thus, closely similar ones are preferred. Users pay attention to three points to detect whether a website is real or duplicated;

- The first and most important one of these is the web site content. Users are familiar with the website content that they regularly use and they expect to see this similarity when they log in,
- The second important point is the situation that the website publishes SSL. Many users think that there is no problem when they see the lock icon or green color in the address bar,
- The third and last point of attention is the domain name of the website.

While attackers are aiming to create a phishing website, they first purchase the use of domain names that are similar to the domain names in their target to be duplicated [33]-[35],[43]. As an example of this situation, domain names that are similar to the "disneyplus.com" domain name have been searched in DNSTwister. According to the result obtained, it has been seen that there are 377 records similar to the relevant domain name. Another method that is used by attackers to choose a domain name is internationalized (IDN) domain names [44]-[46]. The purpose of IDN is to enable domain names to be created with letters in all international alphabets. Attacker was able to, by using this method, create a domain name "adwords-googíe.com", which is very similar to the real "adwords-google.com" website. The attacker replaced the letter "l" in the original website domain name with the letter "í" in the Faroese alphabet, thereby creating this domain name that looks very similar to the eye. As shown in Figure 1, when the related domain name is analyzed, the expression of "xn--adwords-googe-7ib.com" appears.

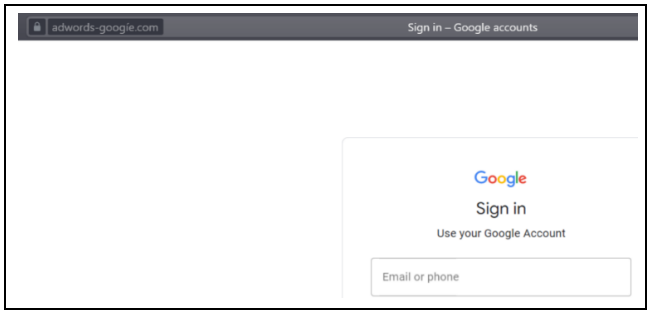

Figure 1. The phishing website having an internationalized domain name to hack Google account information (https://xn--adwords-googe-7ib.com).

The second step in creating a phishing website is to benefit from the tools that are used to duplicate the Httrack-style website to duplicate the target website [36]. This kind of tools, with a single click, make the target website ready to be duplicated completely and published on a different domain [37]. After attackers have duplicated the website of the target institution, they change some of the codes in the background so that the information entered by users is directed to them, but not to the real institution. Because these changes made have no effect on the theme, when users access this fake website, they will see the same website as that of the real institution.

The third and last step in creating a phishing website is to install a certificate and make the fake website link reliable. The main purpose of cyber attackers in using the encryption method with certificate is that the type of attack cannot be detected due to the fact that encrypted traffic cannot be inspected. However, the purpose of using the certificate in phishing websites is to give user a safe environment image rather than protecting the traffic. When users see the green color or lock icon in the address bar, they are more convinced that the website is authentic. It is essential that the certificate be signed by an authorized company in order for the certificate used on websites to be reliable. Before 2016, such cases were rare because authorized companies were rather reluctant to sign certificates that could potentially be used in situations such as phishing websites. However, especially with the recently emerged free certificate signing platform "Let's Encrypt", which aims to secure all the internet traffic, a great change has been experienced [38],[39]. Attackers turned this situation designed for security purposes in their favor and produced reliable certificates for phishing websites. Attackers have installed these reliable certificates, which they created free of charge and without supervision, on phishing websites and provided a more convincing environment for users [40],[41]. Even though the phishing website is ready to be used following the installation of the certificate, it does not make sense to exist unless users access this website. Moreover, users are unlikely to find this phishing website by their own efforts. Because this site is not indexed in any search engine or is indexed at very low results. Therefore, Attackers direct users to these phishing websites through methods such as campaigns, sweepstakes, financial gain and advertising. Nowadays, Twitter posts and Google ads

are often used to direct users to phishing websites [42]. Attackers follow the phishing websites that they create live as of the moment users are directed to those websites. Their main purpose here is to assess user information without wasting time. Because users can realize the situation, and change their information and fail the attack for attackers. Another case is that instant monitoring is mandatory to access the systems having a secondary verification. It is recommended by cyber security experts to use two different entry methods while entering the fields containing valuable information [49]. In the internet sites containing financial assets such as websites of banks, it is mandatory that two different entry methods exist in accordance with the regulations.

As shown in Figure 2;

1. User logs in to phishing website with credentials,
2. Attacker gains user's credential via phishing website,
3. Attacker logs in the real website with user's credential,
4. Real website send secondary verification code to user,
5. User enters secondary verification code to phishing website,
6. Attacker gains user's secondary verificiation code via phishing website,
7. Attacker enters secondary verification code the real website and successfully enters the users account.

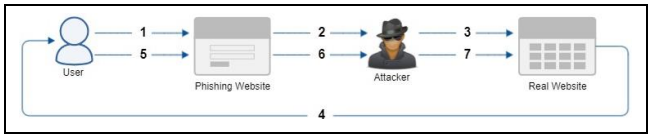

Figure 2. Flowchart of circumvention of secondary verification method with information from phishing websites.

4 Developed tool: phisherhunter

In this chapter, developed PhisherHunter module content and methods are discussed along with all necessary background technical details.

4.1 Detection of phishing websites

In this chapter, a tool has been developed to detect phishing websites as soon as possible and to prevent any undetected website. In the developed tool, four different methods have been used for the detection of phishing websites, one of which is used before the phishing website starts to publish, two of which are used while users are being directed to the phishing website, and the last of which is used by obtaining information from the authorized institution of the state.

4.1.1 Checking newly registered domain names

Domain name details on the Internet are kept by registrar companies. For example, ".com" and ".net" extensions belong to Verisign company and this company keeps information about its extension. ".Ist" and ".istanbul" extensions belong to İstanbul Dijital Medya Tic. A.Ş. and this company keeps the domains registered with this domain name. There are more than 1500 domain name extensions around the world and the number of extensions is increasing day by day. Registrar companies keep information of their domains up-to-date and share the information of all registered domain names of their extensions every day. Considering this logic, providing that the information belonging to the previous day is known, the differences can be deduced and the websites that are newly registered in the last 24 hours can be detected. In this study, the information of the domain names registered in the last 24 hours has been obtained through API of the Zonefiles company, since it would require a lot of effort to process the information received from more than 1500 registrar companies and find new registered websites. With this way, a domain name that can be used for phishing purposes can be detected before it starts publishing. In the light of data shared by CERT, 5240 domain names used by attackers in phishing website attacks in the first 6 months of 2020 have been collected in a list. Among these domain names, the names of the most frequently targeted institution websites were determined and a "potential phishing website targets list" covering 23 corporate websites in total has been prepared. Moreover, the most used word fragments in the 5240 domain names have been extracted by using the same data set with the help of the "Keyword Extraction" website [12]. As a result of the operation made, 36 word-fragments have been detected and included in the "phishing vocabulary" list.

$$Similarity = \left[100 \times \left(\frac{2 \times ncm}{los\ a + los\ b}\right)\right] + (10 \times kmc) \qquad (1)$$

The similarity calculation has been made by using the formula shown above, (ncm=Number of characters matched, los=lenght of string, kmc=keyword matched count), and each newly registered domain name and the potential phishing website targets list values have been compared and a similarity value has been obtained. Those with a similarity score above 70 have been assessed as potential phishing websites and included in the "Potential phishing website" list.

4.1.2 Twitter post controls

Social media websites are frequently preferred by attackers because there is a large amount of data flow and the accuracy of this data is not questioned. Twitter is the most preferred among such websites due to the fact that posts can reach everyone. It is necessary that tweets be read in order to detect the direction of users to phishing websites carried out on Twitter. First of all, a permission has been taken from Twitter to create an application for this process, and a test application has been created following the permission obtained. With the help of Tweepy, it has become possible to search for the desired words on Twitter [50]. The words in the "Phishing vocabulary" list have been searched in tweets and the website addresses have been parsed in the results obtained and added to the "Potential phishing websites" list.

4.1.3 Ad control as a result of google searches

Attackers enable their own phishing websites to be listed ahead of the real websites, which they have duplicated, among the results that are shown to users by paying money to search engines for advertising. In such a case, most of users click and are directed to the firstly listed result without paying attention to the domain name. Thus, attackers can direct the desired users directly to the phishing website. Because the most common search engine in which this situation occurs is Google, it has been studied in this chapter of the study. For this process, the words are basically searched on Google at regular intervals, the ad internet addresses listed among the returned results are parsed and their cases whether they are for phishing or not is checked. A web driver has been used to get the websites shown as advertisements in Google results. The reason for this is that the search case made by users cannot be simulated exactly in

the event of searching on Google over direct code. Through the Chrome web driver that is used, the website opens in the same method as the one for users and it show exactly the same results. The keywords in the "Phishing words" list have been searched in a loop for ad control in Google, and the returned results have been parsed with the help of the BeautifulSoup Library. Upon examining the results returned from Google, it was seen that the ad results were kept in the "ads-ad" class, and a regex was used to extract URL addresses from the data in this field. Because institutions also give ads to Google that are valid on their names, a protection step has been applied in this point. According to this, the links to which the advertisements given by institutions to Google are directed, have been first learned and added to a whitelist. The ad addresses detected in the subsequent studies of the function have been compared with this whitelist, thus ensuring that real advertisements are not considered as phishing websites.

4.1.4 CERT black list controls

The black list shared by CERT unit (Computer Emergency Response Team), established in Türkiye, and including phishing websites, has been checked in the scope of this heading. As being particular to Türkiye, in the event of malicious attacks cases via the Internet, CERT is the institution which is directly addressed. CERT has the authority to make the internet service provider to block the traffic to the addresses, where malicious attacks are detected. Therefore, it is aimed that all attacks to be made against our citizens are reported to this address and that the domestic access to the related websites is blocked as soon as possible. Considering the perspective of phishing websites, CERT is the first unit to be notified when a phishing website is detected. CERT publishes malicious websites that it has received and verified the accuracy, under its own domain name, in a black list. Yet, there is no time indicator in this malicious list, and newly added threats cannot be taken separately from the old ones. CERT adds newly added URLs over old ones, and thus, a parsing method is required to be applied.

In order to distinguish the newly added malicious URL addresses, an updated CERT list is kept and whenever the function runs, the incoming data is compared with this list and the differences between them are determined. When the work process is finished, the CERT list is updated. When the parsing process of the newly added addresses are considered as coding, the current CERT malicious address information is written to a list in the function. Whenever the function runs, the current CERT malicious address information is taken and compared with the previous list, and the differences among them are extracted. And when the process is finished, the current malware information is saved in the function so as to be used in the next comparison. With the work of this function, all phishing website detection methods have been completed and all detected domain names have been collected under the "Potential phishing websites" list.

4.1.5 Valid website whitelisting controls

Although unlikely, the system may still detect valid websites as phishing, start active defense against them, and can cause a major problems. To prevent this ASN (Autonomous System Number) based control step has been added to module to whitelist valid websites before active defense starts. Within the framework of the developed method, the ASN IP ranges of the companies within the scope are added to the system manually. In order to detect valid websites, the IP addresses of the hostnames in the list of potential phishing websites are resolved. The obtained IP addresses are checked for being among the previously added ASN IP ranges, and if the IP is in one of these ranges, it is marked as valid and added to the whitelist in order not to cause any problems. In this way, the websites that companies host in their own IP blocks are secured.

4.2 Control of active publication status of phishing websites

Through the phishing website detection methods that are used in the study, some phishing websites are detected before they start active phishing publication, while others are detected in case of active publication. A method has been developed to control the active publication status of potential phishing websites. As shown in Figure 3, the flowchart of the method in use is shown below, and as a first step, a post request is made to the potential phishing website.

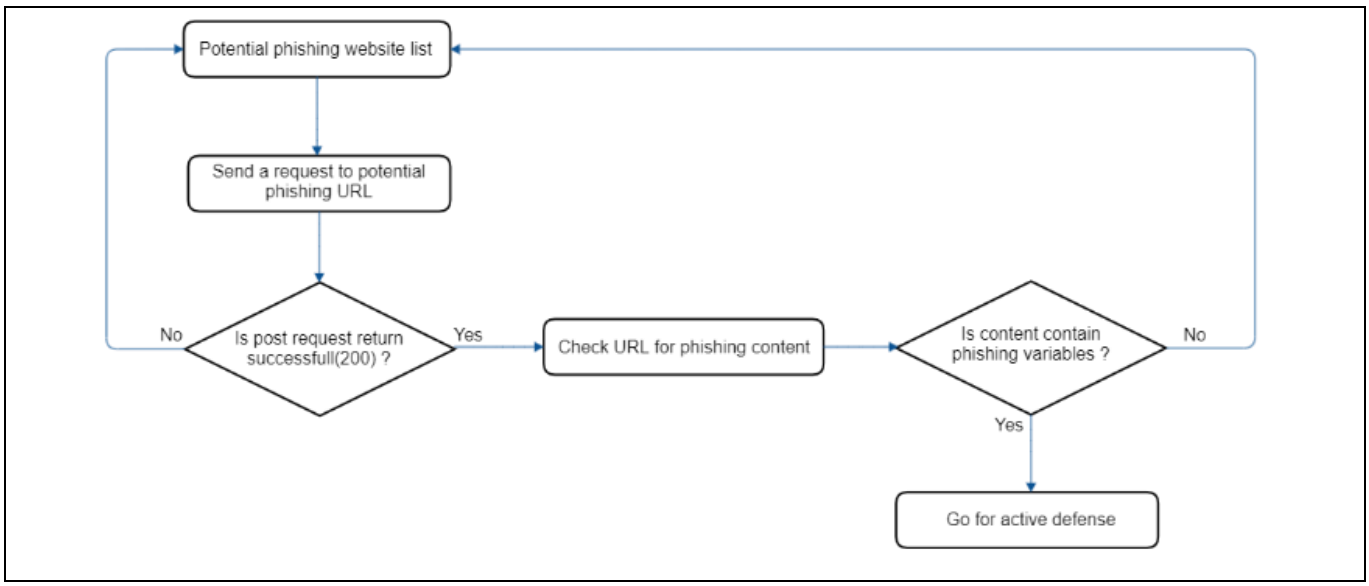

Figure 3. Flowchart showing the active phishing publication status of potential phishing websites.

Provided that the returned value as a result of the post request is successful (200), the flow progresses, but if not, it is kept for re-checking later. The content of potential phishing websites that have been verified to be accessible is taken into the application and their texts are checked. Provided that the texts are suitable for the phishing website model, it is decided that this website makes active phishing publication. But if the content has not yet been loaded or removed, it is converted to the waiting list for later re-checking.

Phishing websites can have two different accesses such as "https" and "http". Therefore, in all these steps above, two kinds of connection requests are checked. Thanks to all these steps, potential phishing websites are separated as active and passive.

4.3 Performance of active defense against phishing websites

Upon the detection of potential phishing websites and then sorting out the active ones, the active defense step, which is the innovative aspect of the work, starts. According to this, users' abuse is tried to be prevented in the period until the access to active phishing websites is blocked. Active defense consists of three titles, each of which is independent.

4.3.1 Sending a notification to the hosting provider

The first step in active defense is to ensure that access to the phishing website is blocked. The most accurate method of blocking access to the website is to send a notification to the hosting provider. For this purpose, the information of the phishing website is learned through a whois query and a notification is sent to the hosting company requesting to block access to the related website. The process of sending notification to the hosting provider consists of two parts. The first part is the detection of the host provided in which the malicious domain name exists and the second part is to send the notification to the related place. These two steps work sequentially for each phishing website. For the phishing website transmitted in this cycle, first a whois query is made and the address to be notified in case of abuse is learned and then a notification is sent to the address.

Whois query results may differ. But almost all of them have a malicious report email address. Furthermore, the only e-mail address returned as a result of whois query is the e-mail address to be used for this malicious report. Therefore, an email parser is used to parse the relevant mail address. With the parser used, the e-mail address to which the notification will be made can be accessed directly and separated from other data. The obtained e-mail address information has been assigned as the e-mail address to be made sending. In this example, because sending is made via gmail, necessary information such as e-mail address and password has been entered. Moreover, the subject and content of the e-mail to be sent have been prepared. In order to facilitate the identification of the relevant company, details are given in the message and the domain name information is included in the mail. Because the sending will be made via the application, the security settings of the relevant gmail address have been adjusted and automatic sending is enabled.

As shown in Figure 4, the notification made is user-friendly and forwarded in an understandable format. A special character has been added to prevent accidental clicks on the domain name address included in it. Upon this step, the notification sending process is completed and the process of blocking the website for access is started.

4.3.2 Active honeypot method

The notification made is user-friendly and forwarded in an understandable format. A special character has been added to prevent accidental clicks on the domain name address included in it. Upon this step, the notification sending process is completed and the process of blocking the website for access is started. The active honeypot method meets the industry and literature for the first time with this study. On the other hand, the honeypot method is a subject that has previously found a large place in both the industry and the literature. The main purpose in the honeypot method is to deceive attackers and wait for them to reveal their own information. Therefore, honeypots remain stationary and are only stimulated by attackers' interactions. With the developed active honeypot method, the interaction of attackers with the honeypot is not expected, this process becomes triggered by security experts. This method has been processed through the sample of financial institutions websites and can be adapted to almost any kind of phishing internet scenario.

As shown in Figure 5;

1. Cyber security engineer logs in to the phishing website with previously marked credential,
2. Attacker gains marked credential via phishing website,
3. Attacker login the real website with marked credential,
4. Security device detects marked credentials,
5. Attacker's session redirects to the honeypot website,
6. Cyber security engineer gains various valuable informations about attacker such as IP and account number.

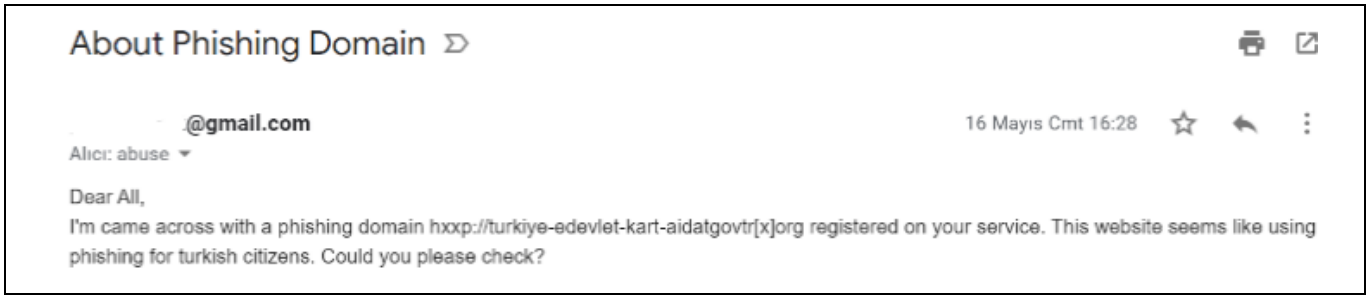

Figure 4. An example of a notification to be sent, for the block of access to the phishing website, to the abuse email address coming as a result of Whois query "hxxp://turkiye-edevlet-kart-aidatgovtr[x]org" phishing.

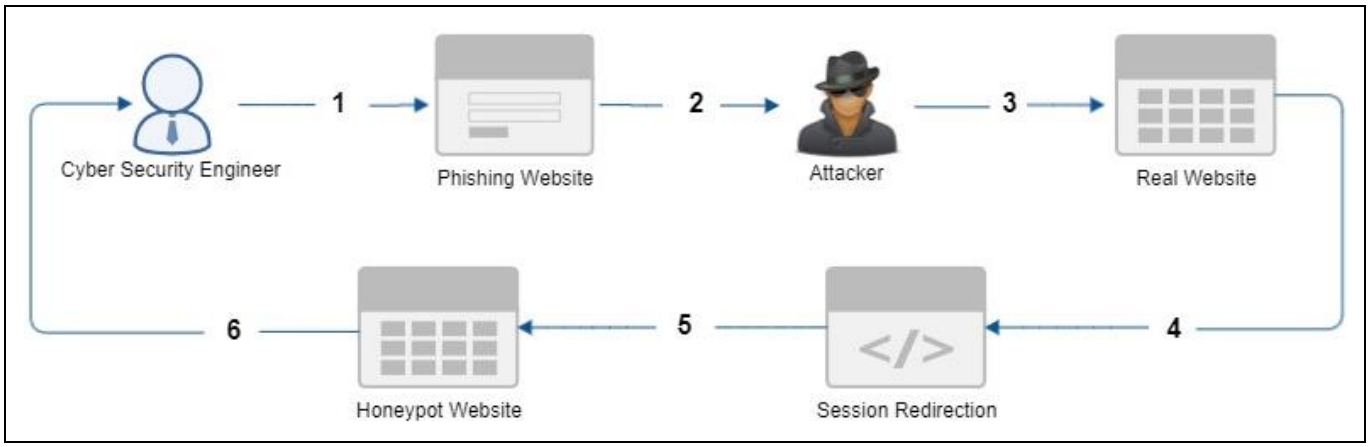

Figure 5. Flowchart of active honeypot method. In order to start this process, security experts get access to the phishing website with a previously marked and known username and password combination.

The attacker thinks that this information is real and enters the real institution website with this username and password combination. As soon as the attacker tries to login, its session is detected and transferred to a honeypot server. The attacker thinks that he has accessed the real system at this point, but he has been trapped in a fake system. In this way, the IP details of the attacker and the bank account information that the user will use to transfer the material assets, in the event that the target website is financed, are obtained. Attackers can easily prepare new phishing websites and can easily carry out users' direction to those websites. The most valuable object, for the attacker, in phishing website attacks is the bank account number to which users will transfer their money. Because the bank account number cannot be imitated like other elements, and it features a real data that has an organic link with the attacker. When the attacker sends the material asset in the user's account to its own bank account number, the attacker discloses its own information and the movements regarding this account are started to be monitored by the financial institutions. Through the active honeypot method, the most valuable information that the attacker has is obtained. In the active honeypot method, honeypot servers and username and password combinations can be easily created. The important thing is to capture the session of the attacker and direct it to the honeypot system. Two different simulations have been created for this process; one in the website codes and the other in the cloud WAF device.

- Additions have been made for predetermined marked users for the aim of routing in the website codes. According to this, if one of these marked users appears in the username field, the session is directed to the honeypot server block, but not to the default server block,
- For the routing in the WAF device, the "http_header" field containing the username information in the packages that reach the institution WAF device has been checked. In the event that one of the marked usernames is found in this field, the routing is made to the honeypot server block, but not to the default server block.

The transfer of the marked data to the detected phishing website is carried out automatically in the developed module. In this process, fake data and the codes developed for the poisoning method were used, the data requested by the phishing website were determined, the relevant fields were filled and sent. Furthermore, the attacker was allowed to use this marked data for 5 minutes, and afterwards the poisoning process was initiated. The reason for the waiting for 5 minutes is to give the attacker time for secondary verification and money transfer transactions while entering the fake site.

4.3.3 Poisoning method with fake data

Every information entered on the phishing website is transmitted to the attacker, and the attacker uses this information to do things for his own benefit. In general terms, every information transferred to the attacker carries a potential financial income value. In this respect, the poisoning method with fake data has been developed to reduce the reality and validity of the information transmitted to the attacker. The poisoning method is based on the principle that too much fake data is transferred to the attacker and real information cannot be selected among these fake data.

In this method, the phishing website is accessed as if a real user does, and the form fields prepared by the attacker are filled with fake information and sent. As shown in Figure 6, provided that too much fake data is entered on the phishing website and forwarded to the attacker, the attacker will be unable to distinguish the real user information from the fake one. Therefore, the attacker will not be able to acquire users' financial assets, or will not reach his target and will be eliminated by security experts. The content of the website that is currently publishing active phishing is received for the poisoning method with fake data. The input fields in the website codes are determined and the "input_name" information is learned. Through the "Pagefiller" function written, the correct data is entered into these input fields on the phishing website, and the structure of the posted website is arranged with the "Httpprocedures" function so as to enable sending process. Because the websites duplicated by attackers are clear and their content does not change, attackers copy a website and use the same content over and over again in different domains. 976 of the 2018 phishing websites verified by CERT in June 2020 are phishing websites that want to obtain credit card information. Attackers used only 6 themes for 976 phishing websites during the said month. The fact that attackers use a limited number of themes facilitates the application of poisoning method which has been developed in the scope of this study.

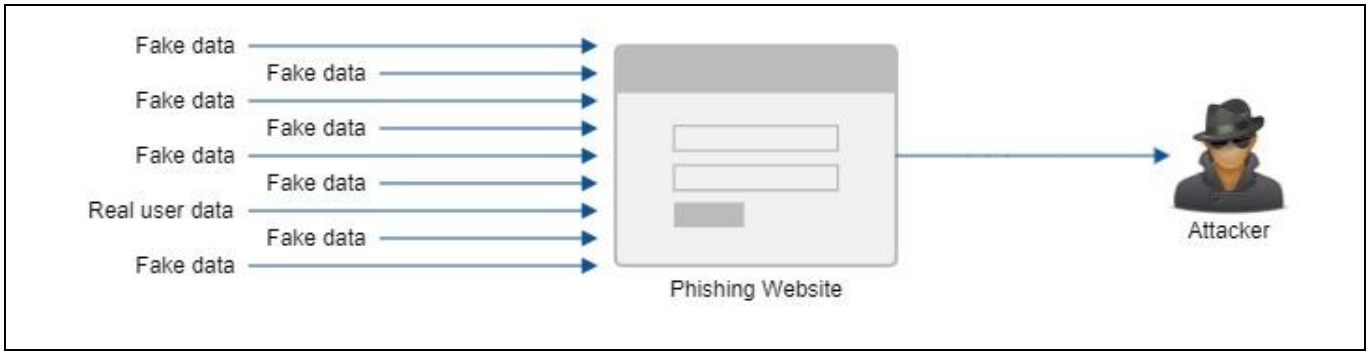

Figure 6. Flowchart of the poisoning process with fake data.

In order to learn what values are taken in the input fields of phishing websites, 25 phishing websites operating in Türkiye has been copied using the WinHTTrack application. Among these copied phishing websites, there are 6 themes used to obtain the credit card information of users during the month of June 2020 and 7 themes used to obtain login username and password information of private institutions. The phishing internet sites duplicated have been restarted on a local computer by using the XAMMP application and their contents have been examined. According to this, it has been observed that 25 websites that were restarted have requested 23 different name field entries. Based on the purpose of poisoning, some methods have been prepared so as to enter different and real-like data into each input field. Attackers check the accuracy of some information such as credit card and Turkish Republic Identity Number and their compliance with algorithms in phishing websites, and, if they are incorrect data, they do not accept them. Therefore, all fields are planned to generate fake data by using correct algorithms. Based on this situation, the following functions, preparing the fake information that attackers want to capture in phishing websites, have been produced;

- Fake credit card data,
- Fake credit card validity date,
- Fake credit card security code,
- Fake Turkish Republic identity number,
- Fake e-mail address,
- Fake password,
- Fake username,
- Fake name and surname combination,
- Fake phone number,
- Fake credit card limit.

The methods formed are placed according to the "input_name" fields in the website. According to this and as shown in Figure 7, provided that there is an entry called "ccnumber" in the website, a fake credit card number will be generated and entered in this field. When the key in the phishing websites is clicked, the information entered into the website is saved in the database. So as to test the poisoning process, information of key actions in phishing websites are arranged to be saved in the database on the local computer, but not in the attacker database. For this operation, a Php file was created and shown as a key action. All required fields are specified in the local database in the php file. For real-life simulation, phishing websites started to be published on the local computer with the help of the Xampp software. Again, the database was re-started through the Xampp software and the necessary links were made. Following these procedures, poisoning trials were initiated. According to this, random fake data was generated according to the field names and accurate submission was provided.

As shown in Figure 8, the data sent within the poisoning process are inseparable from their originals. However, in the event that the poisoning process is performed over a single IP or several IPs, the attacker can exclude the data from these IPs and distinguish the real user data. SOCKS IP addresses were used to overcome this situation. SOCKS is an internet protocol that enables the transmission of internet packets with the help of a proxy server. In this study, it is used as a proxy with the aim of transmitting the data desired to be sent to the other party over different IPs. The data set desired to be sent is prepared for this and delivered to the proxy IP. The proxy IP transmits this package prepared to the target website over its own information. Therefore, the other party are not able to know who the original sender is, and the poisoning process will be successful due to constant different data flow from different IPs, and the real user data will be prevented from distinguishing. A global website has been created and a registration form has been placed in it so as to test the possibility of performance of poisoning over SOCKS IPs. A database link has been made to save the data posted on the website and all incoming data, including IP information, is logged in this field. SOCKS IP and port information that would be used in the process was taken from the "https://free-proxy-list.net" website and typed in a list on the application. The package content desired to be sent to the website was prepared and was transmitted to a random one of the SOCKS IP and ports in the list instead of sending it directly. Check was typed for SOCKS accesses that failed to perform request or could not respond to it and were skipped in case of their failure [51].

4.4 Automatic operation of functions by combining them

An arrangement has been made for the methods created in separate parts, independent from one another, to come together in a common structure and to adjust the timing periods. Code pieces that consist of independent parts are defined in a way that they can be a method by themselves, and "schedular" class is used in order to perform timing. All the codes developed are collected in a single file. According to this file;

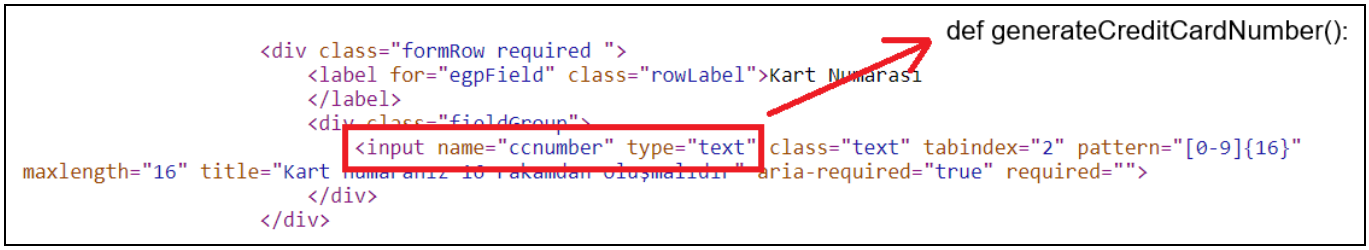

Figure 7. Display of phishing website source code.

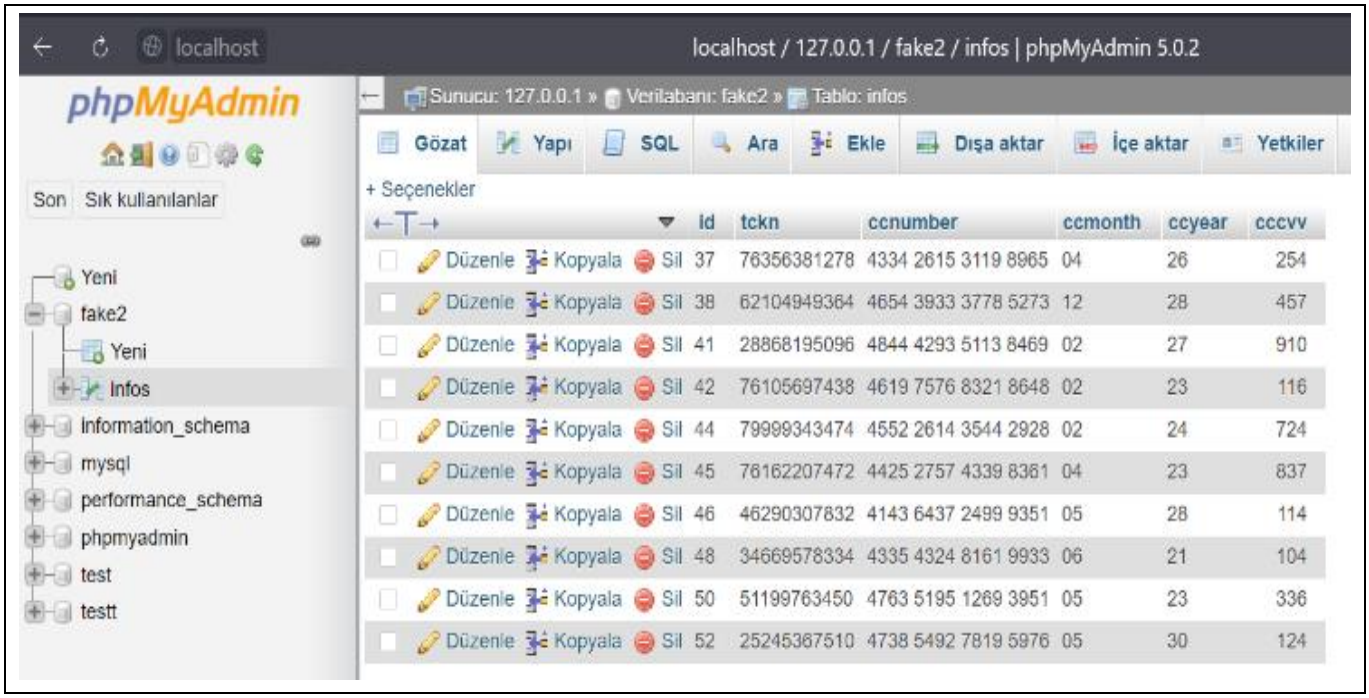

Figure 8. Phishing website database display as a result of poisoning with fake data.

- Daily new registered domain names information is taken every day at 22.10 through the Zonefiles API and saved to the database. The data of newly registered websites use network/storage space between 4-6 mb values,
- The function of detection of potential phishing websites, with the help of similarity, among domain names recorded in the last 24 hours works 10 minutes after the data is received over API, in other words, at 22.20. No network is used during this process, and although this process is the most resource consuming among all module average server easily meets the load of the task,
- Checking black list shared by CERT and the retrieval of newly added logs are made at the 50th minute of every hour of the operation. The data of CERT blacklist use 3 mb of network traffic and 6mb of storage disk,
- The process of searching for potential phishing words on Google search engine and checking whether there are any advertisements for phishing websites among the listed results are made at 40th minute of every hour. This process creates a network traffic of less than 1 mb,
- The process of searching for potential phishing words on Twitter and separating URL addresses from the listed result tweets are done every 30th minutes of every hour. The network traffic of this process can vary according to the content, and it has been observed that it used a maximum of 20 mb.
- The process of whitelist valid websites around potential phishing website list via ASN-based IP control task are done every 55th minutes of every hour. This process creates a network traffic of less than 1 mb,
- The process of notifying abuse addresses via e-mail, in order to block access to active phishing internet sites, is made at the 5th minute of every hour. Since this process is only SMTP traffic, it uses a network less than 1 mb,
- The process of logging into phishing websites with marked username-password combinations, in order to use the active honeypot method, is made at the 5th minute of every hour. Since this process only includes login, less than 1mb of data is used,
- The process of poisoning by sending fake data continues as long as the phishing website is online. The newly added phishing website is only given some

time for the active honeypot method to work, and the poisoning process starts at the 10th minute of every hour. In the poisoning process, which uses the most network by far, one transmission uses a network between 100 kb-1 mb. In case of actively poisoning 5 phishing websites twice a minute, the hourly average network usage is around 300 mb.

5 Results and actions to be taken ın the future

In this study, a tool has been developed for automatic detection of phishing websites and active defense against them. The codes developed and the tool created feature to be the first in the literature in terms of various aspects. Comparison of the study made with other studies in the literature will be processed separately for each method. Studies concerning phishing websites in the literature focused on a single detection method and made use of ready-made data sets. In this study, real phishing websites have been detected by using four different methods that back up each other. There is no open source application that has been developed to detect phishing websites. Therefore, this study features to be the first in this regard in the literature. The detection of potential phishing websites among the websites registered in the last day, developed within the scope of this study, features to be the first in the literature. Many domain names detected by using this method belong to phishing websites that are still in preparation process and have not started publication yet. Through this development, it will be the first time in the literature that the phishing website can be detected before it is published. The similarity used to measure the status of newly registered domain names as being potential phishing websites, via this method, and the 12-day test results made as a result of the word-based controls, it includes, are shown below. According to Table 1, the average successful detection rate is 95.4%.

No study was found in the literature to detect phishing websites that use the advertisements of search engines. A method has been developed in this study to detect phishing websites that advertise on Google. Through this method, the advertisements given to Google were checked at regular intervals and their status as being a phishing website was assessed. Through this development, a method that is used by today's cybercriminals to attract users to their own phishing websites has been introduced to the literature and the first study example on it has been presented. 2 phishing websites that advertise on Google were detected, by using this method, in the testing process of the study. When the control methods, in the literature, of phishing websites in the social media websites are examined; Aggarwal examines tweets via API and browser plug-in, through the PhishAri application that it developed in 2012, and performs detection of phishing websites. In the study in question, a successful detection rate of 92.52%. Was reached. Liew, in the study developed in 2019, reached a successful detection rate of 97.5% by processing 11 features in tweets with the Random Forest Algorithm. Jeong, in the study made in 2016, had a success rate between 88% and 99% using machine learning algorithms on tweets. Because these studies in the literature wish to detect all phishing websites on Twitter, their scope is very wide and they have developed API and browser plug-in for this purpose. Because the scope in the method developed in this study covers only phishing websites operating on Türkiye, searches are made with key words and all the listed results are regarded as potential phishing web sites. Therefore, the success rate of this method is related to the correct configuration of the searched words. In the testing process of the study, more than one million tweets were examined using this method and 6 phishing websites were detected. As being the last part of the detection process, the internet site of CERT institution authorized by the government against phishing internet sites in Türkiye is checked. Through these controls, internet sites that cannot be detected by other methods are learnt, and most of phishing websites operating in Türkiye can accessed.

The words used during the controls of the websites, search engines and social media websites registered in the last day have been taken from the last 6 months data obtained from CERT. Therefore, the methods which are used by attackers are monitored up-to-date and it is enabled that the correct detection rate is stable. Active defense techniques, which is the innovative aspect of this study, have been introduced to the literature for the first time. Three different active defense techniques have been developed in the study. The first of these is automatic mail sending so as to block access to the detected phishing websites. The hosting providers of phishing websites that are detected to be publishing phishing were found for this transaction, their notification addresses were determined and automatic notification was sent via e-mail. During the tests performed, this process was tried 50 times and a success rate of 98% was achieved. The process of blocking phishing internet sites detected with this automation was shortened and the possibility of human error was eliminated. Another active defense method developed is the active honeypot technique. Honeypot techniques are frequently used in today's cyber world and aim to deceive attackers and betray themselves.

Table 1. Correct detection rate by date.

Days	Newly Registered Domains	Detection	True Positive	False Positive	Correct Detection Rate
Day 1	227910	24	21	3	87.5
Day 2	151412	15	15	0	100
Day 3	239270	36	34	2	94.4
Day 4	239118	38	36	2	94.7
Day 5	122617	40	38	2	95
Day 6	190784	30	29	1	96.6
Day 7	336973	50	47	3	94
Day 8	155358	54	54	0	100
Day 9	141268	60	59	1	98.3
Day 10	211544	53	53	0	100
Day 11	238570	51	48	3	94.1
Day 12	291346	21	19	2	90.4

However, in this technique, the honeypot is fixed and awaits actions of attackers. And in the method developed within the scope of this study, marked data is sent to the phishing website monitored by attackers and active trigger operation is made. When attackers enter the real institution website with the marked data, their sessions are directed to the honeypot system. Therefore, valuable information of attackers can be accessed.

The last active defense method is the poisoning method based on the principle of sending a lot of fake data to the phishing website and thus enabling the real victim not to be distinguished. This method and the active honeypot method use the same data transmission functions and, therefore, have the same success rates. During the development of data transmission functions, 25 phishing websites were arranged to publish on the local computer and what information they tried to obtain was examined. It was observed that these websites requested 12 different data types such as credit card number, T.R. ID number, telephone number, and functions that generate fake data were created for each data type by adhering to their real algorithms. Furthermore, the name fields that phishing websites accept as data were examined and appropriate data were defined in 23 name input fields. In this way, attackers see data that looks as if it is sent, by a real user, for each submission. Automatic data sending process is constantly repeated with different data and a situation is created in which attackers cannot distinguish the real victim data. Furthermore, because it is known that provided that attackers list the source IP addresses to which the transmitted data is sent, they can distinguish the fake data, sending via SOCKS proxy addresses is included in the code. Thanks to this method, the web page content prepared is forwarded to attackers over different SOCKS IPs, making it impossible for attackers to distinguish the real information. Through this function, success rate of 92% was achieved in poisoning phishing websites and sending username-password data marked for active honeypot. By using the phishing website detection methods developed within the scope of this study, the detection status of real phishing websites was tested. Based on the results obtained, some improvements have been made and the detection part has taken its final form. And again, among the active defense methods developed within the scope of this study, automatic notification sending process was tested on real phishing websites detected. Test laboratories were established for poisoning and active honeypot methods, which are other active defense methods, and the whole process was carried out here. Timing processes are adjusted for the automatic operation of the entire system and their interoperability is tested. There is no such a module in the literature that includes both detection and active defense.

When all the improvements are summarized, the phishing website detection methods developed within the scope of this study are considered to be completely appropriate for real life scenarios. Through the methods developed, it has been possible to detect the phishing website even before it becomes active. And other methods were used to detect phishing websites that could not be detected by this method at the time of publication. Moreover, details of all phishing websites operating in Türkiye has been obtained with the help of CERT. The module developed takes, with all these processes, all phishing internet sites on its system. Through the active defense module developed, the process of blocking the related phishing websites and abusing users in this process was prevented. Furthermore, additional precautions were taken by obtaining information such as the IP and account number of attackers. And in the future, necessary efforts will be made to change the "phishing words" used in the detection of phishing websites together with the domain names of new phishing websites and to keep themselves up-to-date

6 Author contribution statements

In the scope of this study, the Author 1 in the coding the module, design of the architecture, test the application and the writing of the manuscript; Author 2 and Author 3 in the literature review, examining the results, spelling and checking the article in terms of content were contributed.

7 Ethics committee approval and conflict of interest statement

There is no need to obtain permission from the ethics committee for the article prepared.

8 References

[1] Aburrous M, Hossain MA, Thabatah F, Dahal K. "Intelligent phishing website detection system using fuzzy techniques". *In 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications*, Damascus, Syria, 7-11 April 2008.

[2] Adebowale MA, Lwin KT, Sanchez E, Hossain MA. "Intelligent web-phishing detection and protection scheme using integrated features of Images, frames and text". *Expert Systems with Applications*, 115, 300-313, 2019.

[3] Aggarwal A, Rajadesingan A, Kumaraguru P. "PhishAri: Automatic realtime phishing detection on twitter". *In 2012 eCrime Researchers Summit*, Las Croabas, PR, USA, 23-24 October 2012.

[4] Ali W. "Phishing website detection based on supervised machine learning with wrapper features selection". *International Journal of Advanced Computer Science and Applications*, 8(9), 72-78, 2017.

[5] Ali W, Ahmed AA. "Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting". *IET Information Security*, 13(6), 659-669, 2019.

[6] Chiew KL, Chang EH, Tiong WK. "Utilisation of website logo for phishing detection". *Computers & Security*, 54, 16-26, 2015.

[7] Chiew KL, Choo JSF, Sze SN, Yong KS. "Leverage website favicon to detect phishing websites". *Security and Communication Networks*, 2018, 1-11, 2018.

[8] Ding Y, Luktarhan N, Li K, Slamu W. "A keyword-based combination approach for detecting phishing webpages". *computers & security*, 84, 256-275, 2019.

[9] Federal Bureau of Investigation. "2019 Internet Crime Report". https://pdf.ic3.gov/2019_IC3Report.pdf (03.03.2020).

[10] Jeong SY, Koh YS, Dobbie G. "Phishing detection on Twitter streams". *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Auckland, New Zealand, 19 April, 2016.

[11] Karabatak M, Mustafa T. "Performance comparison of classifiers on reduced phishing website dataset". *6th International Symposium on Digital Forensic and Security,* Antalya, Turkiye, 22-25 March 2018.
[12] Keyword Extractor. "Keyword Extraction". http://keywordextraction.net/keyword-extractor (01.07.2020).
[13] Liew SW, Sani NFM, Abdullah MT, Yaakob R, Sharum MY. "An effective security alert mechanism for real-time phishing tweet detection on Twitter". *Computers & Security*, 83, 201-207, 2019.
[14] Mao J, Tian W, Li P, Wei T, Liang Z.. "Phishing website detection based on effective CSS features of Web pages". *In International Conference on Wireless Algorithms, Systems, and Applications,* Guilin, China, 19-21 June 2017.
[15] Rao RS, Pais AR. "Jail-Phish: An improved search engine based phishing detection system". *Computers & Security*, 83, 246-267, 2019.
[16] Srinivasa RR, Pais AR. "Detecting phishing websites using automation of human behavior". *In Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security*, Abu Dhabi, United Arab Emirates, 2 April 2017.
[17] Subasi A, Molah E, Almkallawi F, Chaudhery TJ. "Intelligent phishing website detection using random forest classifier". *In 2017 International Conference on Electrical and Computing Technologies and Applications*, Ras Al Khaimah, United Arab Emirates, 21-23 November 2017.
[18] Tan CL, Chiew KL, Wong K. "PhishWHO: Phishing webpage detection via identity keywords extraction and target domain name finder". *Decision Support Systems*, 88, 18-27, 2016.
[19] Ulusal Siber Olaylara Müdahale Merkezi. "Zararlı Bağlantılar". https://www.usom.gov.tr/url-list.xml (02.07.2020).
[20] Zhuang W, Jiang Q, Xiong T. "An intelligent anti-phishing strategy model for phishing website detection". *In 2012 32nd International Conference on Distributed Computing Systems Workshops*, Macau, China, 18-21 June 2012.
[21] Jain AK, Gupta BB. "A survey of phishing attack techniques, defence mechanisms and open research challenges". *Enterprise Information Systems*, 16(4), 527-565, 2021.
[22] Gupta BB, Jain AK. "Phishing attack detection using a search engine and heuristics-based technique". *Journal of Information Technology Research*, 13(2), 94-109, 2020.
[23] Jain AK, Gupta BB. "A machine learning based approach for phishing detection using hyperlinks information". *Journal of Ambient Intelligence and Humanized Computing*, 10(5), 2015-2028, 2019.
[24] Jain AK, Gupta BB. "Two-level authentication approach to protect from phishing attacks in real time". *Journal of Ambient Intelligence and Humanized Computing*, 9(6), 1783-1796, 2018.
[25] Basit A, Zafar M, Liu X, Javed AR, Jalil Z, Kifayat K. "A comprehensive survey of AI-enabled phishing attacks detection techniques". *Telecommunication Systems*, 76(1), 139-154, 2021.
[26] Abbasi A, Dobolyi D, Vance A, Zahedi FM. "The phishing funnel model: A design artifact to predict user susceptibility to phishing websites". *Information Systems Research*, 32(2), 410-436, 2021.
[27] Wei W, Ke Q, Nowak J, Korytkowski M, Scherer R, Woźniak M. "Accurate and fast URL phishing detector: a convolutional neural network approach". *Computer Networks*, 178, 1-9, 2020.
[28] Oest A, Safaei Y, Zhang P, Wardman B, Tyers K, Shoshitaishvili Y, Doupé A. "PhishTime: Continuous longitudinal measurement of the effectiveness of anti-phishing blacklists". *In 29th Security Symposium,* Boston, MA, USA, 12–14 August 2020.
[29] Zhang P, Oest A, Cho H, Sun Z, Johnson RC, Wardman B, Ahn GJ. "CrawlPhish: Large-scale Analysis of Client-side Cloaking Techniques in Phishing". *In Proceedings of the IEEE Symposium on Security and Privacy*, San Francisco, CA, USA, 24-27 May 2021.
[30] Tan CL, Chiew KL, Yong KS, Abdullah J, Sebastian Y. "A graph-theoretic approach for the detection of phishing webpages". *Computers & Security*, 95, 1-47, 2020.
[31] Zamir A, Khan HU, Iqbal T, Yousaf N, Aslam F, Anjum A, Hamdani M. "Phishing web site detection using diverse machine learning algorithms". *The Electronic Library*, 38(1), 65-80, 2020.
[32] Alsariera YA, Adeyemo VE, Balogun AO, Alazzawi AK. "Ai meta-learners and extra-trees algorithm for the detection of phishing websites". *IEEE Access*, 8, 1-12, 2020.
[33] Zeng Y, Zang T, Zhang Y, Chen X, Wang Y. "A comprehensive measurement study of domain-squatting abuse". *In ICC 2019-2019 IEEE International Conference on Communications*, Shanghai, China, 20-24 May 2019.
[34] Loyola P, Gajananan K, Kitahara H, Watanabe Y, Satoh F. "Automating Domain Squatting Detection Using Representation Learning". *In 2020 IEEE International Conference on Big Data*, Atlanta, GA, USA, 10-13 December 2020.
[35] Spaulding J, Upadhyaya S, Mohaisen A. "The landscape of domain name typosquatting: Techniques and countermeasures". *In 2016 11th International Conference on Availability, Reliability and Security*, Salzburg, Austria, 31 August-02 September 2016.
[36] Marill JL, Boyko A, Ashenfelder M, Graham L. "Tools and techniques for harvesting the World Wide Web". *In Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries*, Tucson, Arizona, USA, 7–11 June 2004.
[37] Tan, C. L., Chiew, K. L., Yong, K. S., Abdullah, J., & Sebastian, Y. (2020). "A graph-theoretic approach for the detection of phishing webpages". *Computers & Security*, 95, 1-47, 2020.
[38] Aas J, Barnes R, Case B, Durumeric Z, Eckersley P, Flores-López A, Warren B. "Let's Encrypt: an automated certificate authority to encrypt the entire web". *In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, London, United Kingdom, 11-15 November 2019.
[39] Robinson M. "How to Get Free HTTPS Certificates from Let's Encrypt". *Journal of Intellectual Freedom & Privacy*, 2(1), 11-12, 2017.
[40] Kim D, Cho H, Kwon Y, Doupé A, Son S, Ahn GJ, Dumitras T. "Security Analysis on Practices of Certificate Authorities in the HTTPS Phishing Ecosystem". *In Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*, Hong Kong, 7-11 June 2021.
[41] Holub A, O'Connor J. "COINHOARDER: Tracking a ukrainian bitcoin phishing ring DNS style". *In 2018 APWG Symposium on Electronic Crime Research,* San Diego, CA, USA, 15-17 May 2018.
[42] Liew SW, Sani NFM, Abdullah MT, Yaakob R, Sharum MY. "An effective security alert mechanism for real-time phishing tweet detection on Twitter". *Computers & Security*, 83, 201-207, 2019.

[43] Szurdi J, Kocso B, Cseh G, Spring J, Felegyhazi M, Kanich C. "The long "taile" of typosquatting domain names". *In 23rd {USENIX} Security Symposium*, San Diego, CA, 20-22 August 2014.

[44] Krammer V. "Phishing defense against IDN address spoofing attacks". *In Proceedings of the 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services*, Markham, Ontario, Canada, 30 October-1 November 2006.

[45] Fu AY, Deng X, Liu W. "A potential IRI based phishing strategy". *In International Conference on Web Information Systems Engineering*, New York, NY, USA, 20-22 November 2005.

[46] Hu H, Jan ST, Wang Y, Wang G. "Assessing Browser-level Defense against IDN-based Phishing". *In 30th {USENIX} Security Symposium,* Anaheim, CA, USA, 11-13 August 2021.

[47] Aburrous M, Hossain MA, Dahal K, Thabtah F. "Experimental case studies for investigating e-banking phishing techniques and attack strategies". *Cognitive Computation,* 2(3), 242-253, 2010.

[48] Qabajeh I, Thabtah F, Chiclana F. "A recent review of conventional vs. automated cybersecurity anti-phishing techniques". *Computer Science Review,* 29, 44-55, 2018.

[49] Amin A, Haq I, Nazir M. " Two factor authentication". *International Journal of Computer Science and Mobile Computing*, 6(7), 5-8, 2017.

[50] Roesslein J. "Tweepy Documentation". http://tweepy.readthedocs. io/en/v3, 5 (20.06.2020).

[51] Free Proxy List. "Free Proxy List". https://free-proxy-list.net (04.07.2020).