İSTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ

KÖTÜ NİYETLİ ALAN ADLARININ VERİ MADENCİLİĞİ KULLANILARAK TESPİT EDİLMESİ

YÜKSEK LİSANS TEZİ

M. CİHAD TUNA

Bilişim Uygulamaları Anabilim Dalı

Bilgi Güvenliği Mühendisliği ve Kriptografi Programı

EYLÜL 2019

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ

KÖTÜ NİYETLİ ALAN ADLARININ VERİ MADENCİLİĞİ KULLANILARAK TESPİT EDİLMESİ

YÜKSEK LİSANS TEZİ

M. CİHAD TUNA
(707151017)

Bilişim Uygulamaları Anabilim Dalı

Bilgi Güvenliği Mühendisliği ve Kriptografi Programı

Tez Danışmanı: Prof. Dr. Muhammed Oğuzhan KÜLEKÇİ

EYLÜL 2019

İTÜ, Bilişim Enstitüsü'nün 707151017 numaralı Yüksek Lisans Öğrencisi M. CİHAD TUNA, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "KÖTÜ NİYETLİ ALAN ADLARININ VERİ MADENCİLİĞİ KULLANILARAK TESPİT EDİLMESİ" başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Prof. Dr. Muhammed Oğuzhan KÜLEKÇİ** .....
İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Doç. Dr. Behçet Uğur TÖREYİN** .....
İstanbul Teknik Üniversitesi

Dr. Mahmut Şamil SAĞIROĞLU .....
ERLAB Teknoloji AŞ.

Teslim Tarihi : **2 Mayıs 2019**
Savunma Tarihi : **10 Eylül 2019**

Tüm aileme,

ÖNSÖZ

Üzerimdekini haklarını ödeyemeyeceğim anne ve babama en derin saygı ve hürmetlerimi sunarım. Benim ve tüm kardeşlerimin her zaman en iyi eğitimi almamız için çabaladılar.

Lisans eğitimim sırasında hiçbir zaman desteğini esirgemeyen Sayın Prof. Dr. Erkan Türe'ye teşekkür ederim.

Yüksek lisans eğitimim boyunca her daim desteğini gösteren ve yüksek lisans eğitimimi tamamlamam konusundaki teşvikleri ile bu tezi yazmamı sağlayan hocam Sayın Prof. Dr. Muhammed Oğuzhan Külekçi'ye özel teşekkürlerimi sunarım.

Tez çalışması sırasındaki değerli katkıları dolayısıyla başta E. Delibaş, M. Macit, S. Aydın, A. Kartal olmak üzere tüm çalışma arkadaşlarıma teşekkür ederim.

Lisans ve yüksek lisans eğitimim sırasında en büyük destekçim olan hanımıma sürekli sabrı ve desteği için müteşekkirim.

Eylül 2019

M. Cihad TUNA
(Bilgisayar Mühendisi)

İÇİNDEKİLER

Sayfa

KISALTMALAR

2FA : Two Factor-Authentication
APWG : Anti-Phishing Working Group
ASCII : American Standard Code for Information Interchange
ccTLD : Country Code Top Level Domain
CERT : Computer Emergency Readiness Team
CSIRT : Computer Security Incident Response Team
DMARC : Domain-based Message Authentication, Reporting & Conformance
DNEV : Domain Name Extractor and Validator
DNS : Domain Name System
EPP : Extensible Provisioning Protocol
FDD : Fraudulent Domain Detection
FN : False Negative
FP : False Positive
FTP : File Transfer Protocol
GDPR : General Data Protection Regulation
gTLD : Generic Top Level Domain
HTML : Hypertext Markup Language
HTTP : Hypertext Transfer Protocol
HTTPS : Hypertext Transfer Protocol Secure
IDN : Internationalized Domain Name
IANA : Internet Assigned Numbers Authority
ICANN : Internet Corporation for Assigned Names and Numbers
IP : Internet Protocol
MD5 : Message Digest 5
ML : Machine Learning
MITM : Man in the Middle Attack
NLP : Natural Language Processing
NS : Name Server
OTP : One Time Password
REGEX : Regular Expressions
SOME : Siber Olaylara Müdahale Ekibi
SOC : Security Operations Center
SSL : Secure Sockets Layer
TLD : Top Level Domain
TN : True Negative
TP : True Positive
URL : Uniform Resource Locator
USOM : Ulusal Siber Olaylara Müdahale Merkezi
WWW : World Wide Web

ÇİZELGE LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

KÖTÜ NİYETLİ ALAN ADLARININ VERİ MADENCİLİĞİ KULLANILARAK TESPİT EDİLMESİ

ÖZET

İnternetin hayatımızın her alanında yer alması, yanında siber saldırıları da birlikte getirmiştir. Siber saldırılar; devlet kurumlarından özel sektöre, kritik altyapılardan kişisel verilere kadar birçok farklı alan ve boyutta meydana gelmektedir. Siber saldırganların günümüzde en çok kullandığı yöntemlerden biri olan oltalama saldırıları (phishing) her yıl yüz milyonlarca dolarlık zarara sebep olmaktadır. Oltalama saldırısı temelde, "-mış gibi davranmak" üzerine kurulmuştur. Müşterisi olunan bankadan gönderilmiş gibi gelen e-postalar, vatandaşı olunan ülkenin kamu kurumlarına aitmiş gibi görünen internet siteleri ve diğer örnekler.

Oltalama saldırılarını engellemek için birçok farklı çözüm geliştirilmektedir. Bunlardan bazıları bilgi ve bilinç seviyesini artırıcı çalışmalar, diğer bazıları ise oltalama saldırılarını engellemeye yönelik teknik çalışmalardır.

Bu tez çalışması kapsamında, oltalama amacıyla kullanılan internet sitelerinin tespiti ile ilgili yapılan çalışmalar incelenmiş, ticari ürünler denenmiş ve oltalama amacıyla kullanılabilecek kötü niyetli alan adlarının tespitine yönelik bir yöntem önerilmiştir.

Oltalama amacıyla kullanılabilecek kötü niyetli alan adlarının tespiti ile ilgili bu zamana kadar yapılan çalışmalar büyük çoğunlukla aynı/benzer veri setleri ile yapılmış çalışmalardır. Bu veri setleri gerçek hayat senaryoları ile çok örtüşmeyen ya da güncelliğini yitirmiş veriler olabilmekte ve elde edilen sonuçları tam olarak test etmeye imkan vermemektedir.

Tez çalışması kapsamında önerilen yöntemin test edilmesi için 18 ay boyunca gerçek verilerden oluşan bir veri seti hazırlanmış ve önerilen yöntemin doğruluğu kontrol edilmiştir. Tez çalışması sırasında, 18 ay boyunca kayıt edilen yaklaşık 140 milyon alan adı günlük olarak analiz edilmiştir.

Sonuç olarak önerilen yöntemin, kötü niyetli alan adlarını tespit etmede %89 başarı sağladığı görülmüş, bu başarı oranının nasıl artırılabileceği ve gelecek çalışmalarda neler yapılabileceği ortaya konmuştur.

DETECTING FRAUDULENT DOMAIN NAMES BY USING DATA MINING TECHNIQUES

SUMMARY

The presence of the Internet in every area of our lives has also brought cyberattacks to our lives. Cyber attacks are targeting public institutions, private sectors, critical infrastructure, personal data, and more. One of the most popular methods used by cyber attackers today is phishing, causing hundreds of millions of dollars of damage every year. Phishing is a cybercrime in which a target or targets are contacted by email, telephone or text message by someone posing as a legitimate institution to lure individuals into providing sensitive data such as personally identifiable information, banking and credit card details, and passwords. Phishing attacks are based on "fakeness". Fake emails that behave as coming from the banks, fake websites that act like the original website of public institutions, and more. The information is then used to access important accounts and can result in identity theft and financial loss.

Many different solutions are being developed to prevent phishing attacks. Some of these solutions aiming to increase the awareness and knowledge of phishing attacks, while others are technical solutions to prevent phishing attacks. Within the scope of this thesis, the studies on the detection of websites used for phishing purposes were examined, commercial products were tried, and a method for the detection of fraudulent domain names was proposed.

Mostly, the same/similar data sets were used on the researches done for the detection of fraudulent/malicious domain names that can be used for phishing purposes. These datasets may be data that do not overlap with real-life scenarios or are outdated and do not allow to test the results obtained thoroughly. The researches are also not focused only on the fraudulent domain names registered for phishing purposes, they include malware distribution, command & control (c2c) attacks, etc. However, the anatomy of phishing purposed domain names and others are different. So, in this thesis, only phishing purposed fraudelent domain names are considered to find and offer a better solution.

In this thesis mainly two questions are asked at the first and answers are tried to find during the thesis.

The first question is: Can we propose a better methodology than the current ones to detect fraudulent/malicious domain names registered for phishing? Can we support and test this methodology with real data other than the current limited test data? Is the methodology we would like to propose can be used to minimize the damage of phishing attacks?

The second question is: How can a system be developed to detect malicious domain names that can be used for phishing purposes faster than existing solutions? What is needed for the fastest detection, how much does the rapid detection reduce the impact of the attack?

In the scope of this thesis, more than forty banks, electronic payment systems, and payment gateways are tracked and the phishing domain names related to these organization are examined. It's known that phishing domain names are generally similar to the original/targeting website's domain name.

In order to detect fraudulent domain names, a method called FDD was proposed in this thesis. While detecting fraudulent domain names, all the domain names are filtered based on seven similarity labeling models that implemented for the thesis. These labeling models are as the following:

1. exact match
2. confusable exact match
3. contains match
4. fuzzy match
5. fuzzy contains match,
6. confusable fuzzy match,
7. confusable fuzzy contains match.

In order to test the proposed method on the thesis, a data set consisting of real data was prepared for 18 months, and the accuracy of the proposed method was checked. During the thesis study, approximately 140 million domain names registered for 18 months were analyzed daily.

The results found by the method proposed in this thesis was compared to a well-known blacklist called "USOM Zararlı Bağlantılar Listesi (USOM ZBL) (TR-CERT Malicious URL List) to understand how true this method.

USOM ZBL contains more than 20,000 malicious URLs come from three different sources:

1. USOM, itself.
2. CERTs.
3. Reports come from different sources such as individuals, etc.

It was seen that 89% of the phishing domain names listed on the USOM ZBL were detected by the proposed method:

- Source: USOM Detected by FDD: 85%
- Source: CERTs Detected by FDD: 92%
- Source: REPORTS Detected by FDD: 87%
- GENERAL Detected by FDD: 89%

Another comparison was the detection time difference between the method proposed in this thesis and the blacklist. The results are as the following:

- Source: USOM FDD Detection time: 92 hours (earlier)
- Source: CERTs FDD Detection time: 53 hours (earlier)
- Source: REPORTS FDD Detection time: 71 hours (earlier)
- GENERAL FDD Detection time: 64 hours (earlier)

It was seen that the proposed method achieved 89% success in detecting fraudulent domain names. Moreover, it identifies the fraudulent domain names 64 hours earlier than the well-known blacklist.

As a result, this thesis work shows how the success rate of fraudulent domain name detection can be increased and what could be done in future studies were also demonstrated in this thesis.

1. GİRİŞ

İnternetin hayatımızın her alanında yer alması, yanında siber saldırıları da birlikte getirmiştir. Siber saldırılar; devlet kurumlarından özel sektöre, kritik altyapılardan kişisel verilere kadar birçok farklı alan ve boyutta meydana gelmektedir. Siber saldırganların günümüzde en çok kullandığı yöntemlerden biri olan oltalama saldırıları (phishing) her yıl yüz milyonlarca dolarlık zarara sebep olmaktadır. Oltalama saldırısı temelde, "-mış gibi davranmak" üzerine kurulmuştur. Müşterisi olunan bankadan gönderilmiş gibi gelen e-postalar, vatandaşı olunan ülkenin kamu kurumlarına aitmiş gibi görünen internet siteleri ve daha birçok örnek.

İnternetin insan hayatını kolaylaştırdığı yadsınamaz bir gerçektir. Örneğin vatandaşlık işlemleri internet üzerinden yapılabilmekte, belediyelerden online hizmet alınabilmekte, abonelik ve fatura ödeme işlemleri yürütülebilmekte, e-ticaret siteleri aracılığıyla tüm dünyadan alışveriş yapılabilmekte, online para transferleri gerçekleşebilmektedir. Tüm bu online hizmet ve servisler insan hayatını kolaylaştırırken aynı zamanda siber saldırılara da kapı açmaktadır. Online hizmet ve servislerin bu kadar yaygın olması tüm bu hizmetleri hedef alan oltalama saldırılarının da aynı oranda artmasına sebep olmaktadır. Oltalama saldırıları finansal zarara sebep olur, güven ortamını zedeler ve itibar kaybına neden olur.

Oltalama saldırılarının bu kadar yaygın olmasının temel nedenlerinden biri insan faktörüdür. Dikkatsizlik, bilgi-bilinç eksikliği, acelecelik ve hırs gibi insana özgü duygular oltalama saldırılarının yaygın ve başarılı olmasının sebebidir. Bu yüzden oltalama saldırılarına karşı farkındalığın oluşturulması ve bilinç seviyesinin artırılması önemli olsa da bundan daha önemlisi oltalama saldırıları ile mücadele edecek sistemlerin geliştirilmesidir.

1.1 Tezin Amacı

Oltalama saldırıları, çoğunlukla hedef alınan gerçek internet sitelerinin alan adlarına benzer alan adları ile yapılmaktadır. Ayrıca, bir alan adı ile yapılan oltalama saldırısı çoğu zaman bir haftadan kısa sürmekte ve saldırganlar yeni alan adları ile saldırılarına devam etmektedirler. Bu durumda, oltalama amacıyla kullanılabilecek kötü niyetli alan adlarının en hızlı ve doğru şekilde tespit edilmesi, oltalama saldırıları ile mücadelede çok önem arzetmektedir.

Bu tez iki temel soruya cevap bulmak üzere hazırlanmıştır:

1. Oltalama amacıyla kullanılabilecek kötü niyetli alan adlarının tespitinde mevcut çalışmalardan daha iyi sonuçlar alabilecek ve gerçek verilerle desteklenmiş bir yöntem önerilebilir mi? Bu yöntem ile önleyici bir tedbir alınıp oltalama saldırılarının vereceği zarar en aza indirilebilir mi?
2. Oltalama amacıyla kullanılabilecek kötü niyetli alan adlarını mevcut çözümlerden daha hızlı tespit etmek için gerekli sistem nasıl hazırlanabilir? En hızlı tespit için ne gerekmektedir, hızlı tespit saldırının etkisini ne kadar azaltmaktadır?

1.2 Tezin Organizasyonu

Tez çalışması kapsamında öncelikle tezin amacı ve çalışmanın organizasyonı açıklanmış, tezin konusuna temel teşkil eden oltalama saldırılarını anlayabilmek için gerekli olan temel kavramlar ilk bölümde açıklanmıştır.

İkinci bölümde oltalama saldırıları ile ilgili ayrıntılı bilgi verilmiş, oltalama saldırılarının tarihçesi grafiklerle anlatılmış ve oltalama saldırılarına karşı geliştirilen ürün ve çözümlerin incelemesine yer verilmiştir.

Üçüncü bölüm olan Literatür Özeti'nde oltalama saldırıları ile ilgili bu zamana kadar yapılan ve tez konusu ile yakından ilgili olan çalışmaların kısa özetleri ve değerlendirmelerine yer verilmiştir.

Dördüncü bölümde tez çalışması kapsamında önerilen yöntem ayrıntılarıyla anlatılmış, veri temini sürecinden başlanarak önerilen yöntemde kullanılan benzerlik tespit yöntemlerine kadar detaylı şekilde açıklanmıştır.

Beşinci bölümde kötü niyetli alan adı tespit yönteminin gerçek verilerle karşılaştırması yapılmış, elde edilen sonuç, yöntemin öne çıkan başarılı ve başarısız tarafları açıklanmıştır.

Altıncı ve son bölümde tezin genel bir özeti sunulmuş, ileride yapılacak çalışmalar için önerilerde bulunulmuştur.

1.3 Temel Kavramlar

IP Adresi (Internet Protocol Address)

İnternetin genel kullanıma açıldığı 1980'li yıllarda, cihazlar arasındaki paket iletişimini belirleyecek bir protokol olarak geliştirilen IP (Internet Protocol) sayesinde, internete bağlı her bir cihazın 32 bitten ve 4 bloktan oluşan bir adresi vardır. Örnek bir IP adresi şudur: ***160.75.25.31***

1996 yılında IPv6 adıyla yeni bir versiyon duyurulmuş ve IP standartı olarak belirlenmiştir. IPv6'da, adresler 128 bit uzunluğundadır. Örnek bir IPv6 adresi şu şekildedir: ***fd84:69e1:1ce7::25***

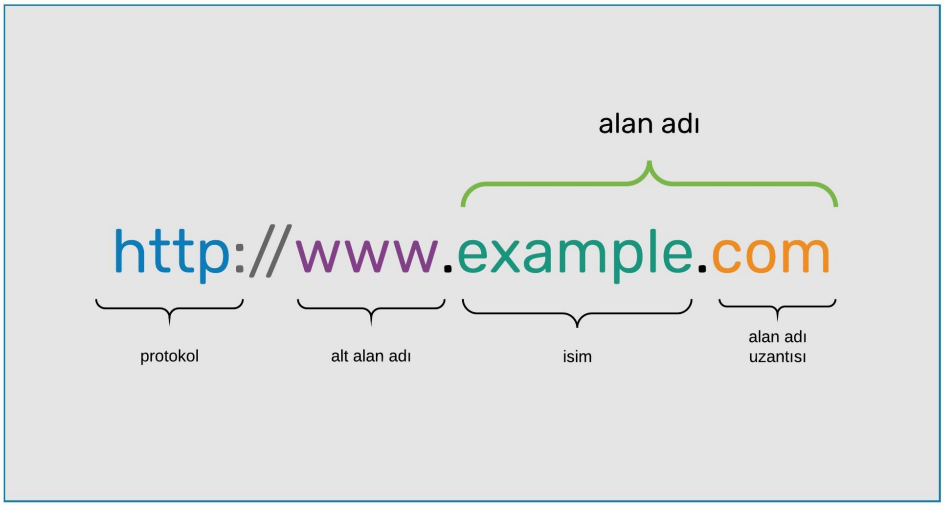
Tekdüzen Kaynak Bulucu (Uniform Resource Locator (URL))

İnternetteki herhangi bir dosyanın ya da kaynağın yerini göstermeye yarayan benzersiz adrese URL denir. Örneğin İstanbul Teknik Üniversitesi Bilişim Enstitüsü'nde görev alan akademisyenlerin yer aldığı Öğretim Üyeleri sayfasının URL'i ***http://www.be.itu.edu.tr/personel/akademisyenler*** dir.

Alan Adı (Domain Name)

URL'in ana kısmını oluşturan ve internette yayın yapan bir şirkete, kuruma, kişiye ya da herhangi bir organizasyona ait olan adresdir. Örneğin ***itu.edu.tr*** bir alan adıdır. Eğer alan adları olmasaydı internet sitelerine erişim sadece IP adresleri ile mümkün olurdu.

İlk alan adı 1985 yılında kayıt edilen ***symbolics.com*** dur [1]. Bir alan adının isim kısmı en fazla 63 karakterden oluşabilir ve Latin Alfabesi'ndeki 26 harf, 0 ile 9 arasındaki rakamlar ve tireden oluşabilir.



Şekil 1.1 : Alan adı anatomisi.

Alan Adı Kaydı (Domain Name Registation)

İlk alan adının kayıt edildiği 1985'den bu yana milyarlarca alan adı kaydedilmiştir. Günümüzde günlük ortalama 250,000 alan adı kayıt edilmekte ve 200,000'e yakın alan adı da süresi uzatılmadığı için silinmektedir. Haziran 2019 itibarıyla kayıtlı alan adı sayısı 351 milyondur [2].

Bir alan adı en az 1 en fazla 10 yıl sonrasına kadar kayıt edilebilir. Alan adı kayıt işlemleri alan adı kayıt firmaları (registrar) aracılığıyla yapılır.

Alan Adı Sistemi (Domain Name System (DNS))

Alan Adı Sistemi, IP adresleri ile alan adları arasındaki bağlantıyı sağlayan sisteme verilen isimdir. DNS sayesinde hangi alan adının hangi IP adresine bağlı olduğu anlaşılır. Böylece internet sitelerine erişmek için IP adresi yerine alan adları kullanılabilir.

İsim Sunucusu (Name Server)

İsim sunucuları alan adı ile alan adınının barındırıldığı sunucu arasında köprü görevi gören adreslerdir. Bir alan adı ziyaret edilmek istendiğinde o alan adının isim sunucularına bakılır, isim sunucuları hangi IP adresini işaret ediyorsa o IP adresine sahip olan sunucuya gidilir ve ilgili siteye erişim sağlanır.

Alt Alan Adı (Subdomain)

Bir alan adının altında oluşturulan isimlere alt alan adı denir. Örneğin ***www.example.com*** alan adında ***www*** kısmı alt alan adıdır. Alt alan adının yönetimi alan adının sahibindedir.

Alan Adı Uzantısı (Top Level Domain (TLD))

Alan adı uzantısı bir alan adında, isimden sonra gelen kısma verilen isimdir. Örneğin ***example.com*** alan adında, alan adı uzantısı ***com*** dur. Alan adı uzantısının bir diğer ismi de üst seviye alan adıdır.

Alan adı uzantıları; genel/jenerik alan adı uzantıları (generic top level domain (gTLD)) ve ülke kodlu alan adı uzantıları (country code top level domain (ccTLD)) olmak üzere ikiye ayrılır.

Genel Alan Adı Uzantısı (Generic Top Level Domain (gTLD))

Herhangi bir ülkeye ait olmayan ve genelde kayıt edilirken herhangi bir belgeye ihtiyaç duyulmayan uzantılardır. En popüler genel alan adı uzantıları ***com***, ***net*** ve ***org***'dur. Kayıtlı alan adlarının ***%60***'ını genel alan adı uzantıları oluşturmaktadır.

ICANN'in 2011 yılında başlatmış olduğu Yeni Genel Alan Adı Uzantıları Programı (New gTLD Program) ile birlikte 1000'den fazla yeni alan adı uzantısı kayda açılmıştır [3]. Bu alan adı uzantılarından bir kısmı genel kayda ve kullanıma açık iken (***.top***, ***.xyz*** gibi), bir kısmı sadece markaların kendi kullanımlarına mahsus, genel kayda açık olmayan uzantılardır (***.audi***, ***.aws***, ***.yandex*** gibi).

Çizelge 1.1 : En fazla alan adına sahip genel alan adı uzantıları [1].

Alan Adı Uzantısı	Kayıtlı Alan Adı Sayısı
.com	144 Milyon
.net	14 Milyon
.org	10 Milyon
.info	5 Milyon
.top	4 Milyon

Ülke Kodlu Alan Adı Uzantısı (Country Code Top Level Domain (ccTLD))

Ülkelere ya da bölgelere ait olan, iki harfli alan adı uzantılarıdır. Ülke kodlu alan adı uzantıları, ülkelerin ISO 3166-1 standardına göre [4] tanımlanmış iki harfli kısaltmasına göre oluşturulmuştur. Örneğin Türkiye'nin ülke kodlu alan adı uzantısı ***.tr***, Japonya'nınki ***.jp*** dir. Ülke kodlu alan adı uzantıları kısıtlı uzantılardır ve kayıt edilirken genelde bir takım belgeler istenmektedir.

Çizelge 1.2 : En fazla alan adına sahip ülke kodlu alan adı uzantıları [5].

Alan Adı Uzantısı	Kayıtlı Alan Adı Sayısı
.cn	23 Milyon
.tk	22.5 Milyon
.de	16.2 Milyon
.uk	13.3 Milyon
.tw	6.5 Milyon

Whois

Bir alan adına ait sahiplik verilerini içeren bilgiye whois denir. Whois bir alan adının kimlik kaydı olarak da nitelendirilebilir. Örnek bir whois bilgisi aşağıdaki gibidir:

```
** Domain Name: itu.edu.tr

** Registrant:
   İstanbul Teknik Üniversitesi
   İstanbul Tek. Üniv. Maslak- Sarıyer
   80626
   İstanbul,
     Türkiye
   hostmaster@itu.edu.tr
   + 90-212-2853930-
   + 90-212-2856936


** Administrative Contact:
NIC Handle    : itu4-metu
Organization Name : istanbul teknik üniversitesi
Address     : İstanbul Teknik Üniversitesi Maslak

        İstanbul,34390
        Türkiye
Phone     : + 90-212-2853930-
Fax     : + 90-212-2856936

...

** Domain Servers:
dns1.itu.edu.tr 160.75.25.1
dns2.itu.edu.tr 160.75.25.65

** Additional Info:
Created on..............: 2002-Dec-23.
```

Şekil 1.2 : Örnek bir whois bilgisi (itu.edu.tr).

Whois Sunucusu (Whois Server)

Whois sorgu isteklerinin gönderildiği sunuculardır. Her bir alan adı uzantısı için whois sunucuları farklıdır. Bazı alan adı uzantılarında whois bilgisi sadece alan adı tescil firmasında (registry) tutulmakta, bazılarında ise hem alan adı kayıt firmasında (registrar) hem de alan adı tescil firmasında tutulmaktadır.

Whois Gizlilik Koruması (Whois Privacy Protect)

Whois bilgisinde yer alan kişisel bilgilerin gizlenmesine yarayan özelliktir. Bu özelliğin aktif edildiği alan adlarının whois bilgisinde, alan adını kimin kaydettiği ve kaydeden kişi ya da kuruma ait iletişim bilgileri gizli tutulur.

Alan Adı Tescil Firması (Registry)

Bir alan adı uzantısını yöneten firmaya alan adı tescil firması (registry) denir. Tescil firmaları yönettikleri alan adı uzantısına ait tüm kayıtları saklamakla yükümlüdür. Bir tescil firması birden fazla alan adı uzantısını yönetebilir. Örneğin ***.com*** ve ***.net*** uzantılarının tescil firması ***Verisign*** firmasıdır [6].

Alan Adı Kayıt Firması (Registrar)

ICANN tarafından akredite edilen ve alan adı kayıt işlemini yapmaya yetkili kılınan firmalardır. Alan adı kayıt firmalarının herhangi bir alan adı uzantısı ile alan adı kaydı yapabilmeleri için ilgili alan adı tescil firması ile anlaşma yapmaları gerekmektedir.

Alan adı kayıt etmek isteyen kişi ve kurumlar, kayıt işlemini alan adı kayıt firmaları üzerinden gerçekleştirebilirler. ICANN tarafından akredite edilmiş iki binden fazla alan adı kayıt firması vardır [7].

Alan Adı Sahibi (Registrant)

Bir alan adını kayıt eden kişi ya da kuruma alan adı sahibi denir. Alan adı sahipleri, alan adı kayıt firmaları üzerinden kayıt işlemi yaparlar ve sahip oldukları alan adını yönetebilirler.

Alan Adı Kayıt Ücreti (Reg Fee)

Bir alan adı kaydı gerçekleştirmek için ödenen ücrete alan adı kayıt ücreti denir. Her bir alan adı uzantısının kayıt ücreti farklı olabileceği gibi, alan adı kayıt firmaları da belli kriterlere göre alan adı kayıt ücreti belirleme hakkına sahiptir. Alan adı kayıt işlemindeki ücret temelde üç kısma ayrılarak dağıtılır:

1. ICANN'e ödenen ücret.
2. Alan adı tescil firmasına ödenen ücret.
3. Alan adı kayıt firmasına ödenen ücret.

Alan Adı Transferi (Domain Name Transfer)

Kayıtlı bir alan adının bir kayıt firmasından diğer kayıt firmasına taşınması işlemine alan adı transferi denir. Alan adı transfer işlemi gerçekleştirilirken alan adı transfer kodu (domain name auth. code / epp code) adı verilen benzersiz bir kod gerekmektedir ve bu kod alan adı kayıt firmaları tarafından alan adı sahibine verilir.

Alan Adı Park (Domain Name Parking)

Kayıtlı bir alan adının aktif olarak kullanılmaması yani o alan adına ait herhangi bir servis/email gibi hizmetin olmaması ancak alan adının tek bir sayfada yayında olmasıdır. Bu tek sayfalık sitede genellikle reklam bulunur ya da gelecekteki bir kullanım için bilgi verilir.

Alan Adı Kök Dosyası (Domain Name Zone File)

Bir alan adı uzantısı ile kayıtlı tüm alan adlarının listelendiği dosyalara alan adı kök dosyası denir. Genel alan adı uzantısı tescil firmaları (gTLD registries) yönettikleri alan adı uzantılarına ait kök dosyalarını paylaşmaktadır ancak ülke kodlu alan adı uzantısı tescil firmaları (ccTLD registries) kök dosyalarını paylaşmamaktadır.

Alan adı kök dosyaları, internetin adres rehberi gibi düşünülebilir. Kayıtlı tüm alan adları ve o alan adına ait isim sunucuları alan adı kök dosyalarında yer alır.

Uluslararasılaştırılmış Alan Adı (Internationalized Domain Name)

Latin Alfabesi dışındaki karakterleri de içerebilen alan adlarıdır. Alan adı kök dosyalarına bu alan adlarının ASCII karşılığı ya da özel olarak punycode karşılığı kaydedilir. Tüm alan adı uzantıları IDN'i desteklemez.

Tüm uluslararasılaştırılmış alan adlarının punycode gösterimi ***xn--*** ile başlar. Örneğin ***örnek.com*** alan adının punycode gösterimi ***xn--rnek-4qa.com*** dur.

Uluslararasılaştırılmış Alan Adı Uzantısı (Internationalized Top Level Domain)

Latin Alfabesi dışındaki karakterleri de içerebilen alan adı uzantılarıdır. Örneğin Katar'ı temsil eden **.قطر** uzantısı ve **.РФ** uzantısı uluslararasılaştırılmış alan adı uzantılarıdır.

ASCII (American Standard Code for Information Interchange)

Bilgi Değişimi İçin Amerikan Standart Kodlama Sistemi anlamına gelen ASCII, Latin Alfabesi üzerine kurulu bir karakter kümesidir ve uluslarası bir standarttır [8]. ASCII olmayan karakterler içeren kök dosyaları alan adı kök dosyalarında ASCII formatına dönüştürülerek saklanır.

Evrensel Kod (Unicode)

Unicode Consortium adlı organizasyon tarafından geliştirilen ve Latin Alfabesi dışındaki harflerin ve karakterlerin de kullanılabilmesi için ortaya konulan bir endüstri standardıdır. Evrensel Kod ile Türkçe, Çince gibi onlarca farklı dildeki tüm karakterlere bir sayı değeri karşılığı verilmiş ve dünyadaki tüm yazım sistemlerinin tek bir standart altında temsil edilmesi sağlanmıştır. Evrensel Kod standardı uluslararasılaştırılmış alan adlarına temel oluşturan bir standarttır [9].

Punycode

Evrensel kod temelli olan uluslararasılaştırılmış alan adlarının, alan adı sistemi (DNS) için belirlenmiş kısıtlı sayıda ASCII karakteri ile gösterimine punycode denir.

Punycode gösterimi Internationalized Domain Names in Applications (IDNA2008) [10] adlı standarta uygun şekilde geliştirilmiştir. Örneğin ***türkçe*** kelimesinin punycode gösterimi ***xn--trke-2oa7j*** dir.

Typo Alan Adı (Typo Domain Name)

Bir alan adının yanlış yazılmış versiyonuna typo alan adı denir. Örneğin ***google.com*** için ***goolge.com*** bir typo alan adıdır.

Typo alan adları genelde şu yöntemlerle oluşturularlar:

1. **Harflerin yer değiştirmesi:** paypal.com – ***papyal.com***
2. **Harf/rakam/tire eklenmesi:** paypal.com – ***paypal1.com***
3. **Harf çıkartılması:** paypal.com – ***paypa.com***
4. **Duble harf kullanılması:** paypal.com – ***paypall.com***
5. **QWERTY metodu:** paypal.com – ***paypak.com***
6. **Homoglyph metodu:** paypal.com – ***paypa1.com***
7. **IDN:** paypal.com – ***paypâl.com***
8. **Farklı uzantı kullanılması:** paypal.com – ***paypal.co***

Typo alan adlarının kötüye kullanılmasına ise ***typosquatting*** denmektedir.

Avrupa Birliği Genel Veri Koruma Yönetmeliği (General Data Protection Regulation (GDPR)

Avrupa Birliği vatandaşlarının kişisel verilerini ve gizliliğini korumaya yönelik bir takım düzenlemelerin yer aldığı ve 25 Mayıs 2018'de yürürlüğe giren yönetmeliktir [11].

İnternet Tahsisli Sayılar ve İsimler Kurumu (Internet Corporation for Assigned Names and Numbers (ICANN))

1998 yılında kurulmuş uluslararası düzeyde organize olmuş bir kurum olan ICANN, İnternet Protokolü (IP) adresi alanı tahsisi, genel alan adı uzantıları (gTLD) ve ülke kodlu alan adı uzantıları (ccTLD) yönetimi ve kök sunucu sistemi yönetimi gibi sorumlulukları olan kâr amacı gütmeyen bir kurumdur [12].

İnternet Tahsis Edilen Numaralar İdaresi (Internet Assigned Numbers Authority (IANA))

1988 yılında kurulmuş ve Alan Adı Sistemi (DNS) ve IP adresi yönetimi gibi görevleri olan uluslararası bir kurumdur [13].

Siber Olaylara Müdahale Ekibi (SOME)

Kamu kurum ve kuruluşları ile özel sektör kuruluşlarının kendi bünyelerinde oluşturmuş oldukları ve görevi kurumlarına doğrudan ya da dolaylı yapılan siber saldırıları tespit ve bertaraf etmek olan, zarar önleyici ve engelleyici fonksiyonları olan ekiplerdir. Bilgisayar Acil Durum Hazırlık Ekibi, Bilgisayar Güvenliği Olay Müdahale Ekibi gibi isimlerle de anılır.

Ulusal Siber Olaylara Müdahale Merkezi (USOM)

Türkiye'de, Bilgi Teknolojileri ve İletişim Kurumu (BTK) [14] bünyesinde kurulan, Türkiye'nin siber güvenliğine karşı siber ortamda ortaya çıkan tehditlerin belirlenmesi, muhtemel saldırı ve olayların etkilerinin azaltılması veya ortadan kaldırılmasına yönelik önlemlerin geliştirilmesi ve belirlenen aktörlerle paylaşılması gibi görevleri olan bir merkezdir.

USOM, ulusal ve uluslararası seviyede siber ortamda ortaya çıkan tehditler ile ilgili kendisine ulaştırılan ihbarları da değerlendirmekte, söz konusu tehditlerin tespit ve bertaraf edilmesi için Kamu Kurumları ve özel kişiler ile koordinasyonunu sağlamaktadır [15].

2. OLTALAMA SALDIRILARI (PHISHING)

2.1 Giriş

Tez çalışması kapsamında önerilen yöntemin daha iyi anlaşılması ve bu alanda yapılan çalışmaların daha iyi değerlendirilebilmesi için oltalama saldırılarının iyi anlaşılması gerekmektedir. Çünkü bu tezin amacı, oltalama amacıyla kullanılabilecek kötü niyetli alan adlarını önceden tespit etmeye yönelik bir yöntemin ortaya konmasıdır.

2.2 Oltalama Saldırılarının Tarihçesi

İlk defa 1987 yılında kullanılmaya başlanan phishing terimi; password ve fishing yani parola ve balıkçılık kelimelerinden ihtira yoluyla oluşturulmuş yeni bir terimdir [16]. Türkçe'ye oltalama ya da yemleme olarak çevrilmiştir. Oltalama saldırılarının amacı, kişisel bilgiler, parolalar, kredi kartı bilgileri gibi hassas bilgilerin ele geçirilmesidir.

Genel olarak sahte e-posta ve sahte internet siteleri aracılığıyla yapılan oltalama saldırılarında gerçek hizmet taklit edilir ve bu taklit işlemi büyük çoğunlukla asıl internet sitesinin alan adına benzer isim kullanılarak yapılır. Oltalama amacıyla kullanılmak üzere kayıt edilen kötü niyetli alan adlarına örnekler aşağıdaki tabloda verilmiştir:

Çizelge 2.1 : Oltalama amacıyla kullanılmak üzere kayıt edilen bazı alan adları.

Gerçek Alan Adı	Oltalama Amacıyla Kayıt Edilen Alan Adı
turkiye.gov.tr	turkiyegovtr.com
turktelekom.com.tr	turktelekon1.net
garanti.com.tr	garantı.com
teb.com.tr	ceptebsube.com
gib.gov.tr	gelir-idaresi-odeme.com
denizbank.com	denîzbank.com
vodafone.com.tr	vodafone-fatura.com

Oltalama saldırıları ile ilgili çalışmalar yapan Anti-Phishing Working Group [17]'un yayınladığı ***Phishing Activity Trends Report 1st Quarter 2019*** raporuna göre 2019'un ilk çeyreğinde en çok oltalama saldırısına uğrayan sektörler, ödeme sistemleri, finansal kuruluşlar ve SaaS/Webmail sağlayıcılarıdır [18].

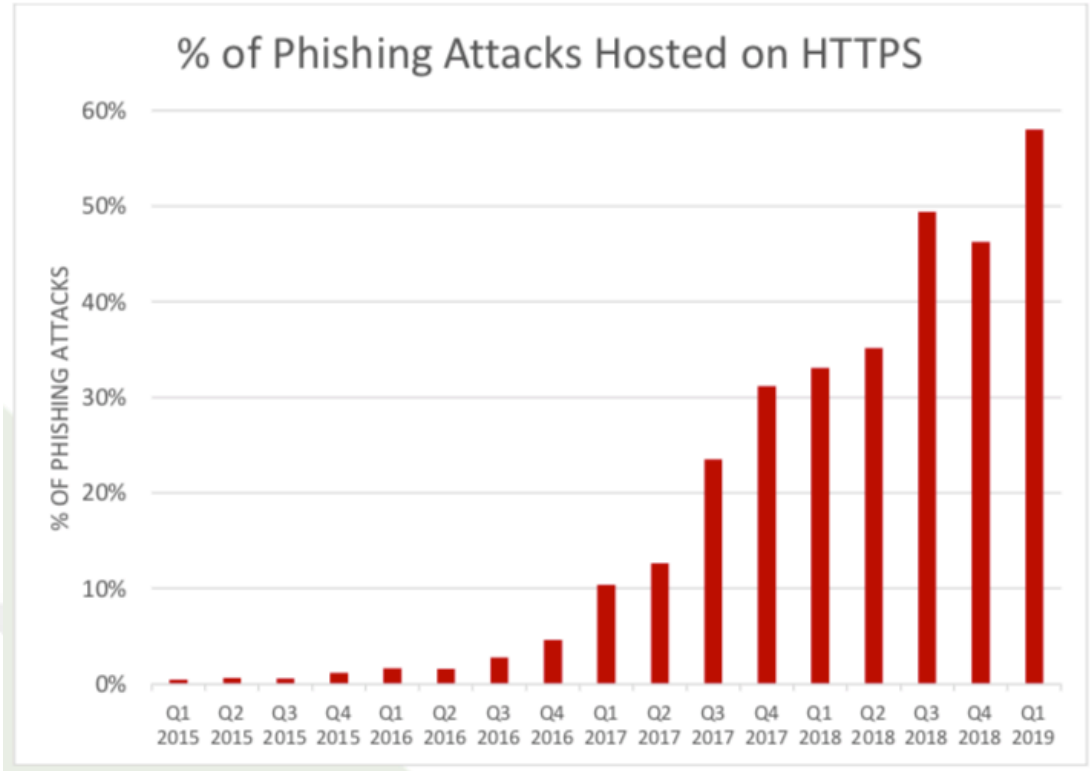
Çizelge 2.2 : APWG 2019 1. Çeyrek Raporu'ndaki istatistikler.

	Ocak	**Şubat**	**Mart**
Tespit edilen oltalama sitesi	48,663	50,983	81,122
Tespit edilen oltalama e-posta kampanyası	34,630	35,364	42,399
Oltalama saldırılarında hedef alınan marka sayısı	327	288	330

Rapora göre APGW tarafından tespit edilen oltalama sitelerinin sayısı aylık ortalama 50,000 olarak belirtilmiştir.

Şekil 2.1 : Ekim 2018 – Mart 2019 arasında tespit edilen oltalama site sayısı.

Uzun yıllar boyunca öne çıkartılan ve bir güvenlik-güven kriteri olarak sunulan SSL sertifikalarının oltalama saldırılarında çokça kullanılmaya başlandığı ve APWG tarafından tespit edilen oltalama sitelerinin %60'ının https üzerinde çalıştığı görülmektedir [18]. Yani bir sitenin SSL sertifikasına sahip olması o sitenin güvenli olduğuna anlamına gelmemekte hatta oltalama saldırılarında kullanılan alan adlarında meşru sitelere oranla HTTPS kullanımının daha yaygın olduğu görülmektedir.



Şekil 2.2 : Oltalama sitelerinin HTTPS kullanma oranları.

Phishlabs [19] tarafından yayınlanan ***2019 Phishing Trends and Intelligence*** raporuna göre ise Türkiye'de de özellikle 2018 yılında oltalama saldırıları bir önceki yıla göre %905 artmış [20], bankalar, kamu kurumları, telekom operatörleri gibi online işlem yapılan servisleri kullanan vatandaşlar hedef alınmıştır.

Emniyet Genel Müdürlüğü Siber Suçlar Daire Başkanlığı'na bağlı ekipler; 2018 yılında, yüzden fazla siber saldırganı oltalama saldırıları yapmak ve hassas verileri ele geçirmek suçlamasıyla göz altına almıştır [21].

Oltalama saldırıları, saldırının yapısı itibarıyla en kolay gerçekleştirilebilen siber saldırıların başında gelmektedir. Bir kurum, kendi iç ağını, internet sitelerini, sunucularını, uygulamalarını ve akla gelebilecek diğer tüm dijital varlıklarını korumak için çaba sarfetse de kendini hedef alan bir oltalama saldırısı ile itibar ve güven kaybına uğrayabilir, finansal zarar görebilir. Çünkü oltalama saldırıları tüm bunlardan bağımsız olarak yapılan dış saldırılardır.

2.3 Oltalama Saldırıları ile İlgili Açık Veritabanları ve Uygulamalar

PhishTank

OpenDNS [22] tarafından yürütülen bir proje olan PhishTank [23], dünya genelinde binlerce gönüllü kullanıcısı olan ve kullanıcıların oltalama saldırıları için kullanılan URL'leri ekledikleri, onayladıkları ve yayınladıkları bir platformdur. Oltalama saldırıları ile ilgili yapılan çalışmaların çoğunda veri kaynağı olarak Phishtank kullanılmaktadır.

OpenPhish

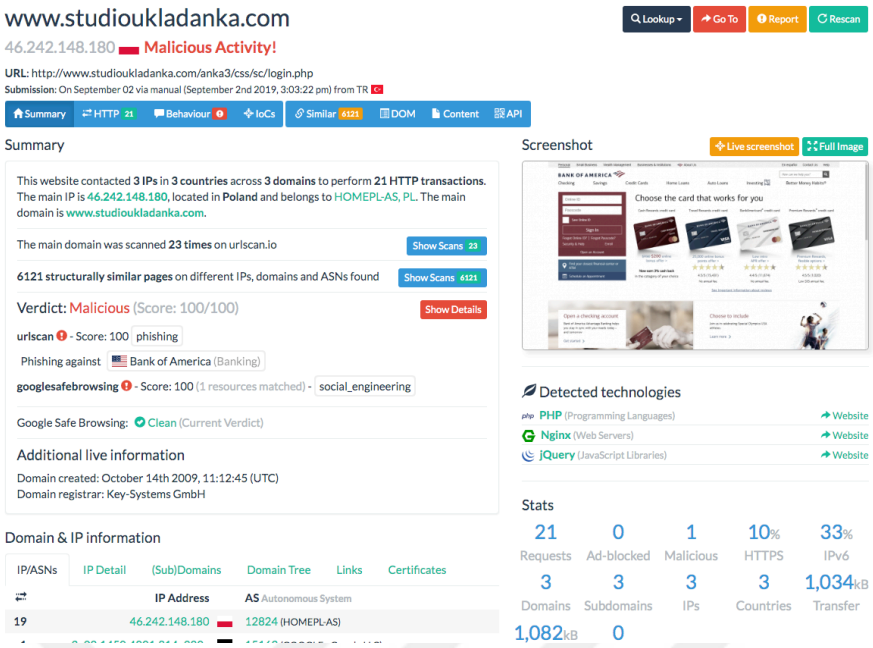
Birçok farklı ağ kaynağından aldığı URL'leri oltalama saldırısı tespit motoru aracılığıyla analiz eden ve tespit ettiği zararlı URL'leri ücretsiz olarak yayınlayan bir platformdur [24]. OpenPhish'de sadece yeni ve aktif URL'ler yayınlanmaktadır.

PhishBank

Comodo Tehdit Araştırma Laboratuvarları (COMODO Threat Research Labs) [25] tarafından yürütülen bir proje olan PhishBank [26]; kullanıcı adı, parolalar, kredi kartı bilgileri gibi hassas bilgileri ele geçirmeye yönelik açılan internet sitelerini tespit edip yayınlandığı bir platformdur.

URLScan.io

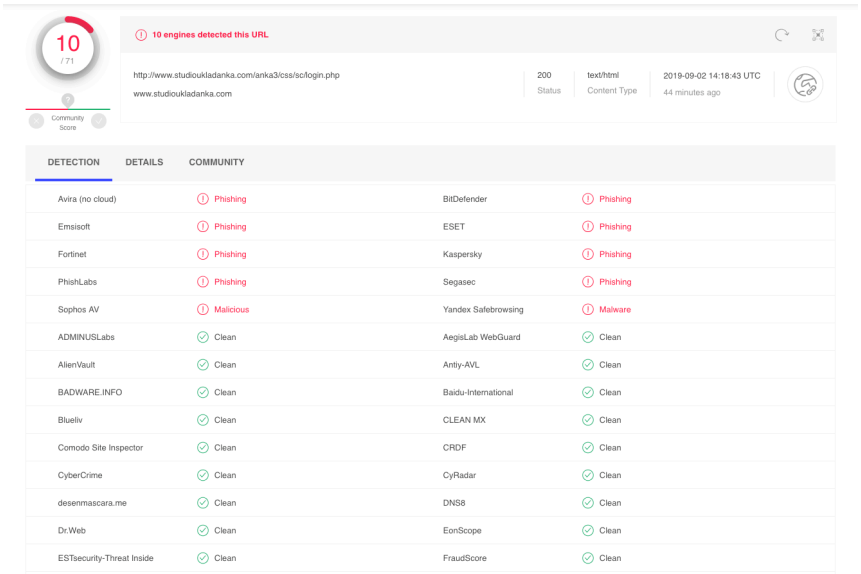
URL'leri tarama ve analiz etme hizmeti sunan URLScan.io [27], taranan URL'in IP adresi, JavaScript ve CSS kodları, ekran görüntüsü gibi özellikleri ile Google Safe Browsing gibi veri kaynaklarının verilerini birlikte sunan bir platformdur.



Şekil 2.3 : URLScan.io adresinde yapılan örnek bir sorgunun sonucu.

VirusTotal

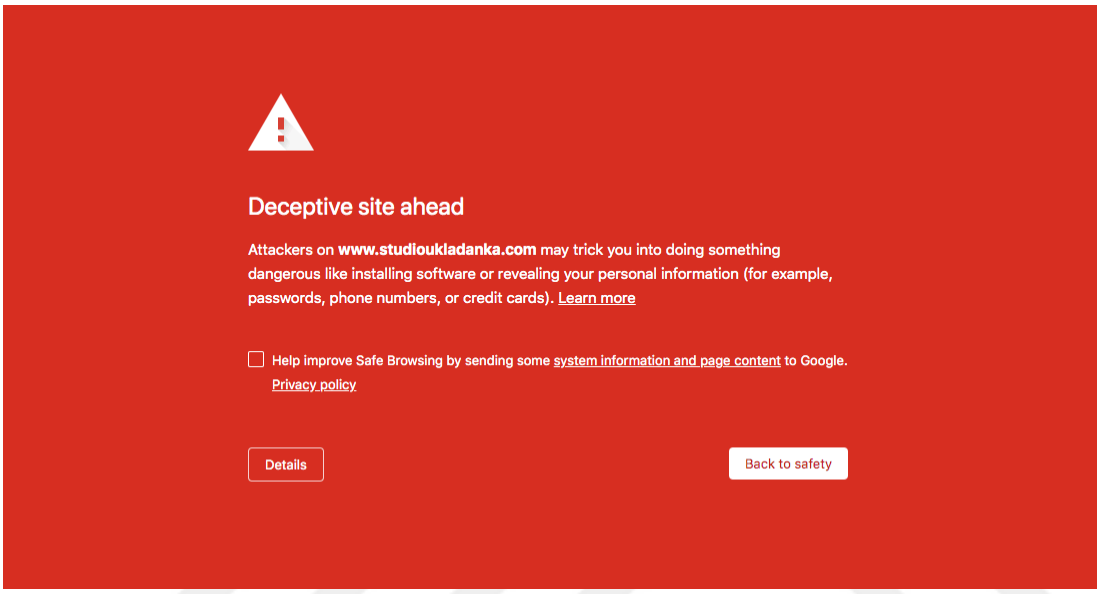
2004 yılında kurulan ve 2012 yılından bu yana Google [28] tarafından yönetilen VirusTotal [29]; dosya ve URL taranmasına imkan sağlayan ve 50'den fazla antivirüs yazılımında tarama gerçekleştiren bir ücretsiz bir araçtır. VirusTotal'e eklenen URL'ler antivirüs programlarında taranır ve bu programların taranan URL'i nasıl sınıflandırdığı gösterilir.



Şekil 2.4 : VirusTotal.com adresinde yapılan örnek bir sorgunun sonucu.

Google Safe Browsing

2007 yılından bu yana Google tarafından yürütülen bir proje olan Google Safe Browsing [30], internet kullanıcılarını başta oltalama saldırıları ve zararlı yazılımlar olmak üzere tehditlerden korumayı amaçlayan bir platformdur. Birçok web tarayıcısı, kullanıcılarının ziyaret ettiği URL'lerin zararlı olup olmadığına dair kontrolü Google Safe Browsing veritabanı ile yapmaktadır.



Şekil 2.5 : Google Safe Browsing tarafından zararlı bulunan bir websitesi.

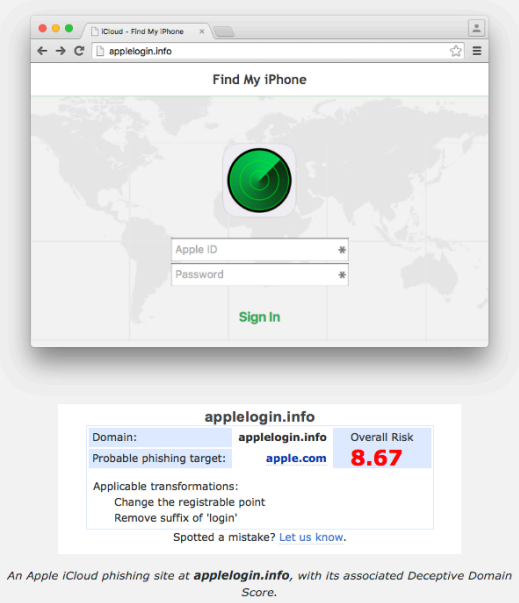
2.4 Oltalama Saldırılarını Engellemeye Yönelik Ürünler, Çözümler ve Firmalar

Oltalama saldırılarını engellemeye yönelik ürünler genellikle e-posta yoluyla yapılan saldırıları engellemeye yönelik ürünlerdir. Bu bölümde oltalama saldırıları ile mücadeleye yönelik ürünlerden tez çalışması kapsamına girenler incelenecektir.

NetCraft

Oltalama saldırıları ile mücadeleye yönelik çalışan en eski firmalardan biri olan NetCraft [31] bu alanda birçok çözüm geliştirmiştir.

Bu çözümlerden biri olan "Deceptive Domain Score" da, bir alan adının aldatıcı olup olmadığı ile ilgili bilgi verilmekte ve alan adı skorlanmaktadır. Bu skorlama; uluslararasılaştırılmış alan adı (IDN), benzer karakterler, çok kullanılan kötü niyetli kelimeler (update, login, secure) gibi farklı kriterlere göre yapılmaktadır.



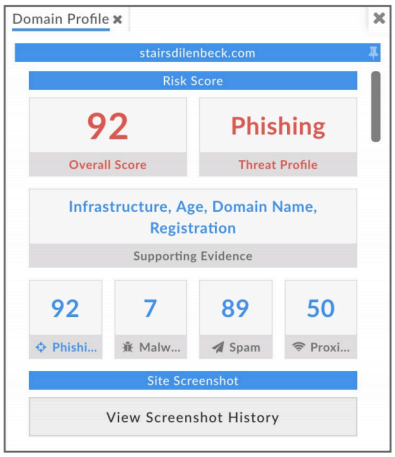
Şekil 2.6 : NetCraft tarafından tespit edilen kötü niyetli bir alan adı.

PhishLabs Domain Monitoring

Sadece oltalama saldırılarına yönelik ürün ve servisler sunan PhishLabs [32]; Domain Monitoring ürünü ile 2,000'den fazla alan adı uzantısını takip ettiğini, park edilmiş alan adlarını da izlediğini ve müşterilerine yönelik saldırılarda hızlı aksiyon aldığını belirtmektedir.

DomainTools PhishEye ve DomainScore

Alan adı endüstrisindeki en eski firmalardan biri olan DomainTools [33]'un PhishEye ve DomainScore ürünleri, kötü amaçlı kullanılabilecek alan adlarını önceden tespit edip önlemeye yönelik çözümler sunmaktadır.



Şekil 2.7 : Domaintools tarafından tespit edilen kötü niyetli bir alan adı.

CSIS PhishDB

Kendisini tam otomatik bir anti-phishin ürünü olarak tanıtan CSIS PhishDB [34]; hızlı ve doğru tespit, sürekli izleme, ayrıntılı analiz ve hızlı engelleme özelliklerini öne çıkartmaktadır.

MarkMonitor AntiFraud - Phishing

En eski marka koruma firmalarından biri olan ve dünyanın en büyük firmaları tarafından tercih edilen MarkMonitor'ün Anti Fraud – Phishing [35] ürününün rakiplerinden en az %50 daha fazla oltalama saldırısı tespit ettiği iddia edilmekte ve erken uyarı sistemi ile kurumları hedef alan saldırıların önceden engellendiği vurgulanmaktadır.

RSA Fraud Action - Phishing Protection

Siber güvenlik alanında birçok ürün ve servis sunan RSA'nın FraudAction [36] ürününün, her saat 120'den fazla oltalama saldırısı tespit ettiği, mobil uygulamalar ve sosyal medyayı sürekli takip ettiği ve partnerleri aracılığıyla her gün milyonlarca URL'i taradığı belirtilmektedir.

Proofpoint Web Domain Fraud Monitoring

Yeni kayıt edilen alan adlarını takip ederek, müşterilerini siber saldırganlardan, oltalama kampanyalarından ve alan adlarının her türlü kötüye kullanımından koruduğunu vurgulayan Proofpoint [37] ayrıca alt alan adı, URL ve diğer birçok farklı kritere göre keşif taraması yaptığını belirtmektedir.

Looking Glass Technical Threat Indicator

Oltalama amacıyla kullanılan URL'leri gerçek zamanlıya yakın şekilde tespit ettiğini ve yeni kayıt edilen alan adlarını takip ettiğini öne çıkartan Looking Glass [38] oltalama saldırılarına karşı bütüncül bir tehdit istihbaratı ürünü olduğunu belirtmektedir.

RedMarlin

Sıfırıncı gün oltalama saldırılarına karşı yapay zeka ile güçlendirilmiş tespit sistemi sunduğunu söyleyen RedMarlin [39], derin öğrenme, görsel karşılaştırma ve doğal dil işleme ile oltalama saldırılarına karşı güçlü bir çözüm olduğunu belirtmektedir.

Segasec

Alan adı kayıtlarından SSL sertifikalarına, sosyal medyadan e-posta sağlayıcılarına kadar milyarlarca veri kaynağını sürekli analiz ettiğini söyleyen Segasec [40], izleme ve engelleme sistemlerinin markaları en etkin biçimde koruduğunu belirtmektedir.

3. LİTERATÜR ÖZETİ

Oltalama saldırılarının tespiti ile ilgili yapılan çalışmalar incelendiğinde özellikle son yıllarda; oltalama saldırılarının makine öğrenmesi ve yapay zeka kullanılarak tespit edilmesi ile ilgili çokça çalışma yapıldığı görülmektedir. Bu tez çalışmasından farklı olarak alan adı, onlarca tespit kriteri arasından sadece biri olarak görülmektedir. Sadece kötü niyetli alan adlarının tespiti ile ilgili yapılan çalışmalarda ise oltalama saldırıları ile birlikte botnet, zararlı yazılım dağıtımı gibi farklı siber saldırı yöntemleri için kullanılabilecek alan adlarının tespitine yönelik çalışmalar yapıldığı görülmektedir. Oltalama amacıyla kullanılabilecek kötü niyetli alan adları ile diğer kötü niyetli alan adları yapısal olarak birbirinden farklı olduğu için tek bir çalışma ve yöntem ile tüm kötü niyetli alan adlarının tespit edilmesinin önerilmesi yazar tarafından önemli bir sorun olarak görülmektedir.

Bu bölümde kötü niyetli alan adı tespitinde alan adlarını ön plana çıkartan çalışmalar incelenmiş ve her biri ile ilgili kısa değerlendirmeler yapılmıştır.

D. Chiba ve diğerleri tarafından kötü niyetli alan adlarına yönelik tehdit istihbaratı çözümü geliştiren çalışmada [41], kötü niyetli (malicious) alan adları, ele geçirilmiş ve adanmış olarak ikiye ayrılmış; ele geçirilmiş alan adlarının normalde kötü amaçlı olmayan ama siber saldırganlar tarafından ele geçirilip kötü amaçlı kullanılan alan adları olduğu, adanmış alan adlarının ise kayıt amaçlarının en baştan beri kötü olduğuna işaret edilmiştir. 1.6 milyon alan adının analiz edildiği ve DOMAINCHROMA adı verilen bir yöntem sunulan çalışmada, adanmış kötü niyetli alan adları dokuz alt kategoride incelenmiş ancak oltalama amacıyla kullanılabilecek alan adları için özel bir çalışma yapılmamış, daha çok zararlı yazılım ve komuta kontrol gibi kötüye kullanımlar değerlendirilmiştir.

Kötü niyetli alan adlarının tespitinde yüksek başarı oranının yakalanması için çok fazla özellik/kritere bakılması gerektiğini ve bunun çok maliyetli olduğunu belirten çalışmada [42], W. Wang ve K. E. Shirley, sundukları yöntemde sadece alan adlarının yapısal özelliklerine bakıldığını ve kelime segmentasyonu (word segmentation) yapılarak iyi sonuçlar alınabileceğini ifade etmişlerdir.

H. Zhao ve diğerleri tarafından kötü niyetli alan adlarının tespiti için N-Gram tabanlı bir algoritma önerilen çalışmada [43], Alexa [44] verilerine göre en popüler 100,000 internet sitesi seçilerek 3,4,5,6 ve 7'li n-gram ayrımı yapılmış ve bir tespit yöntemi önerilmiştir. Ancak çalışmada da belirtildiği üzere bu yöntem uluslarasılaştırılmış alan adları (IDN) ve karıştırılabilir karakterlere sahip alan adlarını bulmak için uygun değildir.

DNS analizi ile kötü niyetli alan adlarını tespit etmeye yönelik EXPOSURE adında bir sistem ortaya konulan çalışmada [45], araştırmacılar L. Bilge ve diğerlerinin; botnet, zararlı yazılım, komuta kontrol merkezi (C2C) ve oltalama saldırıları gibi atakları on beş farklı kritere göre değerlendirip tespit ettikleri belirtilmiştir. Çalışmanın öne çıkan noktalarından biri verinin yüz milyar gerçek DNS sorgusunun analiz edilmesi ile oluşturulmasıdır. Ancak bu çalışmada alan adının kötü niyetli olup olmadığına bakılırken göz önünde bulundurulan on beş farklı kriterden sadece ikisi alan adına hastır ve diğer on üç kriter DNS bilgilerine bağlıdır. Ayrıca alan adına has kriterler olarak alan adının içerdiği rakamların oranı ve anlamlı kelimelerin uzunluğunun toplam alan adı uzunluğuna oranı göz önünde bulundurulmuştur. Bunun sebebi olarak iyi amaçlı internet sitelerinin anlamlı alan adlarını kullanması ama kötü niyetli siber saldırganların alan adlarının anlamlı olması gibi amaçları olmadığı belirtilmiştir. Ancak oltalama saldırıları için kullanılan alan adlarında anlamlı alan adları kullanıldığı bilinmektedir. Bu yüzden rakam kriteri de anlamlı kelime oranı kriteri de oltalama saldırılarında kullanılan kötü amaçlı alan adlarını bulmada anlamlı sonuçlar doğuracak kriterler değildir. Bu çalışmaya bu alandaki diğer çalışmalarda çokça referans verilmiştir.

P. Agten ve diğerleri tarafından yapılan ve yanlış yazılmış (typo) alan adlarını inceleyen çalışmada [46], Alexa verilerine göre dünya genelindeki en popular 500 internet sitesi incelenmiş ve bu internet sitelerinin alan adlarına benzeyen alan adı kayıtları yedi ay boyunca takip edilmiştir. En popüler internet sitelerine benzer alan adı kayıtlarının incelenmesi önemli bir çalışmadır ancak bu çalışmada oltalama saldırıları ile ilgili özel bir çalışma yapılmamış daha çok trafik kaçırma ve paraya çevirme yöntemleri ile ilgilenilmiştir.

Tüm bunların yanı sıra kötü niyetli alan adı tespiti ile ilgili sadece .eu uzantılı alan adları üzerinde yapılan bir çalışma [47]**,** yine Alexa verileri ile oltalama URL veritabanlarını karşılaştıran bir çalışma olan DomainProfiler çalışması [48], algoritmik olarak üretilen kötü niyetli alan adlarını bulmaya yönelik bir yöntem sunan çalışma [49] ve spam amacıyla kullanılabilecek kötü niyetli olabilecek alan adlarını kayıt edildikleri an keşfetmeye yönelik bir yöntem sunan PREDATOR adlı çalışma [50] bu alandaki diğer çalışmalara örnektir.

Sonuç olarak literatürde kötü niyetli alan adlarını tespit etmeye yönelik birçok çalışma yapılmış olsa da, sadece oltalama amacıyla kullanılabilecek kötü niyetli alan adlarını tespit etmeye yönelik özel bir çalışmaya rastlanmamıştır. Alan adlarını tespit etmeye yönelik yapılan çalışmalarda, alan adları ile birlikte birçok farklı kriter de değerlendirilmiş ve kötü niyet kapsamı olarak hem oltalama saldırıları, hem zararlı yazılımlar hem de komuta kontrol merkezi gibi yapısı birbirinden farklı saldırı türleri ele alınmıştır. Ayrıca yapılan çalışmalarda çoğunlukla aynı/benzer veritabanı ve veri setlerinin kullanıldığı görülmüştür. Çalışmalarda sıklıkça dile getirilen ve yapılan çalışmanın kısıtlı olmasına sebep olarak gösterilen; gerçek verinin ***çok büyük ve dinamik olması ve bu veriyi işleyebilmek için fazla kaynak gerekmesi*** gibi hususlar bu tez çalışması sırasında bir sorun olarak görülmemiş, tam tersine tezde önerilen yöntemin diğerlerinden farkını ortaya koyan ve yöntemin sağlaması için gerekli bir kıstas olarak ele alınmıştır.

4. KÖTÜ NİYETLİ ALAN ADI TESPİTİ İÇİN ÖNERİLEN YÖNTEM (FDD)

Oltalama amacıyla kullanılan alan adları temelde ikiye ayrılabilir:

1. Kötü niyetli alan adları (fraudulent).
2. Ele geçirilmiş alan adları (compromised).

Bu tez çalışması kapsamında sadece kötü niyetli alan adlarının tespitine yönelik geliştirilen ve FDD adı verilen bir yöntem önerilmiştir. Kötü niyetli alan adları, bir firmanın ya da organizasyonun müşterilerini/kullanıcılarını aldatma amaçlı olan ve hedef alınan organizasyonun alan adına benzeyen alan adlarıdır. Yani, ele geçirilmiş alan adından farklı olarak kötü niyetli doğan alan adıdır. Örneğin orijinal alan adı ***bankofexample.com*** olan Bank of Example için kötü niyetli alan adı ***bankoffexample.com*** olabilir ve bu alan adı ile oltalama saldırısı yapılabilir. Aynı zamanda tamamen ilgisiz bir alan adı olan anotherdomainname.com alan adı ele geçirilerek de oltalama saldırısı yapılabilir.
Kötü niyetli alan adı tespit süreci temelde üç aşamadan oluşmaktadır:

1. Veri temini
2. Veri üzerinde önerilen tespit algoritmalarının çalıştırılması
3. Tespit veriminin ölçülmesi

4.1 Veri Toplama Süreci ve Ön Hazırlık

Tez çalışması kapsamındaki en önemli süreçlerden biri veri toplama sürecidir. Kısıtlı üçüncü kaynak verilerden ziyade tüm tez çalışması süresince kullanılan verileri asıl kaynaklarından temin etmek için özel çaba sarf edilmiştir. Buna bağlı olarak alan adı kayıt verilerini temin etmek için alan adı kök dosyaları üzerinden bir veri toplama süreci planlanmıştır.

Alan adı kayıt verilerine ulaşmak için alan adı kök dosyalarına erişim gerekmektedir. Bir alan adı uzantısı ile kayıtlı tüm alan adlarının listelendiği dosyalara kök dosyası denmektedir. Alan adı tescil kuruluşları gerekli sözleşmeler imzalandıktan sonra kök dosyalarını ücretsiz olarak sunmaktadır ancak ülke alan adı uzantısını yöneten kurumlar (.us harici) bu dosyaları hiçbir şekilde paylaşmamaktadır. Tez çalışması kapsamında başta .com ve .net olmak üzere binden fazla alan adı uzantısının kök dosyalarına erişim sağlanmıştır.

Farklı alan adı uzantıları için verinin; paylaşılma metodu, güncellenme sıklığı, veriye erişim prosedürleri ve veri formatı farklıdır. Örneğin bir alan adı uzantısı kök dosyası her gün saat 03:30'da güncellenirken bir diğeri değişken saatlerde olmak üzere günde iki kez güncellenebilmektedir. Bir alan adı uzantısına ait kök dosyalarına herhangi bir IP adresinden erişim sağlanabilirken bir diğeri sabit IP isteyebilmektedir. Ayrıca veri formatları da farklı olabilmektedir.

Alan adı kök dosyaları ile ilgili yapılan temelde iki aşamadan oluşmaktadır:

1. Kaynaktaki verinin kontrolü, çekilmesi ve depolanması
2. Verinin işlenmesi ve FDD için hazır hale getirilmesi

Alan adı tescil firmaları alan adı kök dosyalarına erişimi FTP ve HTTP metodları ile vermektedir. Bundan dolayı her bir metod için ayrı tanımlamalar yapılmıştır.

Çizelge 4.1 : Alan adı kök dosyalarına erişim metodu.

Alan Adı Uzantısı	Metod
.com	FTP
.name	FTP
.org	FTP
.info	FTP
.biz	FTP
.us	FTP
Yeni alan adı uzantıları (newgTLD)	HTTP

Kötü niyetli alan adı tespiti ya da marka koruma ile ilgili ürün ve servislerin neredeyse tamamında alan adı kök dosyaları günde bir kez ve aynı/sabit saatte kontrol edilmektedir.

Tez çalışması kapsamında önerilen yöntemde öne çıkan iki geliştirme bulunmaktadır: kaynaktaki verinin kontrolü, çekilmesi ve depolanması işlemindeki **dinamiklik** ve verinin işlenmesi ve FDD için hazır hale getirilmesi işlemindeki **hız**.

4.1.1 Dinamiklik (kaynaktaki verinin kontrolü, çekilmesi ve depolanması)

Dinamik veri kontrolü yeni kaydedilen alan adlarına en hızlı şekilde ulaşmak için büyük önem arz etmektedir. Her bir alan adı kök dosyasının güncellenme saati hem diğer alan adı kök dosyalarından farklıdır, hem de dosyanın güncellenme saati aynı kök dosyası için de farklı için farklı saatlerde olabilmektedir.

Tez çalışması kapsamında geliştirilen yöntemde her bir alan adı kök uzantı dosyası on dakikada bir kontrol edilmektedir. FTP metodu kullanılanlarda dosyanın güncellenme saati ve MD5 dosyası, HTTP metodu kullanılanlarda ise güncellenme saati kontrol edilmektedir. Eğer herhangi bir güncellenme tespit edilirse bir sonraki aşama olan kaynaktan verinin çekilmesi ve depolanmasına geçilmektedir.

1. FTP sunucusuna bağlan.
2. Dosyaları kontrol et.
3. Eğer yeni dosya varsa indir ve döngüye bir saat ara ver.
4. Eğer yeni dosya yoksa on dakika sonra tekrar kontrol et.
5. Eğer indirme işlemi sırasında bir hatayla karşılaşırsan bir dakika sonra tekrar dene.

Her bir alan adı kök dosyası şu şekilde isimlendirilerek depolanmaktadır:

[alan-adi-uzanti-adi]-[yil][ay][gun]-[saat].zone.[depolanma-formati]

Örneğin 30 Haziran 2018 tarihinde, saat 05:00'de çekilen .com uzantısına ait kök dosyası ***com-20180630-05.zone.gz*** ismi ile depolanmıştır.

4.1.2 Hız (verinin işlenmesi ve FDD için hazır hale getirilmesi)

Alan adı kök dosyasının en hızlı şekilde kontrol edilip depolanması kadar önemli olan bir diğer işlem bu verinin en hızlı biçimde işlenmesidir. Alan adı kök dosyalarında, o alan adı uzantısındaki mevcut tüm alan adları ve isim sunucuları yer almaktadır. Örneğin 30 Haziran 2018 gününe ait .com uzantısı kök dosyası ***341,152,261*** satırdan oluşmaktadır.

Verinin işlenmesi süreci üç adımdan oluşmaktadır:

1. Veri temizliği
2. Tekilleştirme
3. Karşılaştırma ve dosya oluşturma

Veri temizliği

Alan adı kök dosyasında yapılan ilk işlem veri temizliği işlemidir. Örneğin, ***.com*** alan adı uzantısına ait kök dosyasının ilk kırk satırı Şekil 4.1'de gösterildiği gibidir.

Görülebileceği gibi kök dosyalarında alan adları ve isim sunucuları haricinde farklı veriler de bulunmaktadır. Bazı alan adı uzantısı kök dosyalarında alan adları uzantı olmaksızın listelenmektedir

```
; The use of the Data contained in Verisign Inc.'s aggregated
; .com, and .net top-level domain zone files (including the checksum
; files) is subject to the restrictions described in the access Agreement
; with Verisign Inc.

$ORIGIN COM.
$TTL 900
@ IN     SOA     a.gtld-servers.net. nstld.verisign-grs.com. (
                  1566879722 ;serial
                  1800 ;refresh every 30 min
                  900 ;retry every 15 min
                  604800 ;expire after a week
                  86400 ;minimum of a day
                  )
$TTL 172800
 NS A.GTLD-SERVERS.NET.
 NS G.GTLD-SERVERS.NET.
 NS H.GTLD-SERVERS.NET.
 NS C.GTLD-SERVERS.NET.
 NS I.GTLD-SERVERS.NET.
 NS B.GTLD-SERVERS.NET.
 NS D.GTLD-SERVERS.NET.
 NS L.GTLD-SERVERS.NET.
 NS F.GTLD-SERVERS.NET.
 NS J.GTLD-SERVERS.NET.
 NS K.GTLD-SERVERS.NET.
 NS E.GTLD-SERVERS.NET.
 NS M.GTLD-SERVERS.NET.
COM. 86400 DNSKEY 257 3 8 AQPDzldNmMvZFX4NcNJ0uEnKDg7tmv/
F3MyQR0lpBmVcNcsIszxNFxsBfKNW9JYCYqpik8366LE7VbIcNRzfp2h9OO8HRl+H+E08zauK8k7evWEmu/6od+2boggPoiEfGNyvNPaSI7FOIroDsnw/
taggzHRX1Z7SOiOiPWPNIwSUyWOZ79VmcQ1GLkC6NlYvG3HwYmynQv6oFwGv/KELSw7ZSdrbTQ0HXvZbqMUI7BaMskmvgm1G7oKZ1YiF7O9ioVNc0+7ASbqmZN7Z98EGU/
Qh2K/BgUe8Hs0XVcdPKrtyYnoQHd2ynKPcMMlTEih2/2HDHjRPJ2aywIpKNnv4oPo/
COM. 86400 DNSKEY 256 3 8 AQPVjPYYiWkWvFhogUC0Q8L0oenMR6EDF5n7veG6sXTs1FcRzOnI9k0dWSe5g7VFgzZ6Zo/W/
gXIsyrli5oL0jbzz2zHX7z7RbnrnTlt5p2fbvezMTJrsqiNRCsx1/YiLyGLk52o+u+265Hi4SM1Zu4bPsjYjCJChxuwZWOvQZ3sZQ==
COM. 86400 NSEC3PARAM 1 0 0 -
COM. 900 RRSIG SOA 8 1 900 20190903042202 20190827031202 17708 COM.
uofJVLJ+aH4bpHJgJzTbj3V+pqenmfdkGtKZRFtAFoBl58XNLZQQWHx3sv+rKLQRKri1jmGOvv5j86CQi+vHrYSzvy/cRWb5vQmcCizQw/
xFPxCXtoBrvgClyRtj+uH4GzqmepNnGfpzcnC9GuOwM9/EAZ88MTlYT9qP/n622M4=
COM. 86400 RRSIG NSEC3PARAM 8 1 86400 20190831044353 20190824033353 17708 COM. AeWkSdZAKngqBb9m9hCdhM3349nFvyu0sMATNBNUtQ7hXOthtGhss6
lzbrFaCacQm3wyx7fR+Chj14w4YvFfDWNaNefs3vu1fKD8ixbP2P9xmXaigNQhvTkT8APNsaPCBVN0v+L+yonenZ2M6g1wAAgP0B1SjDziw3f4jZm0FL4=
COM. RRSIG NS 8 1 172800 20190831044354 20190824033354 17708 COM. W3X4Eu4kfheo9dFM4EvegxelgjpYSYBHo/Fm9KV0VWsSpsE0qrlqUIKC4rdO1DjwoKh
2VoNIJFWNNnHIBPEG2MrYG2gvlsk+TCJDu3WO7sVybaO4HbS8NwRSVfyVZNSdHkj84JHPBvT9+UGAAx8E+MUeAq7QlQCjE073U3skVmU=
COM. 86400 RRSIG DNSKEY 8 1 86400 20190904182533 20190820182033 30909 COM. lrMHl10F3/dxkJreDIqCq/2txBVzAXsokhyPb/
A7fW81AJIEFeXGnQ5c66WKiL15UAMKygAvULJtBIGnYN6LZJ7i7Up+ed4APImUJHnl7XOOEViniUSz5q1NDD4+MiTXe2EFJeC9ynG/bomBUg9G5GuU803de+ZzsKJIR6294wr
BXgTHstKI1/PBypoCiQifSbQuX3YPDYSwKmvqDJfp+TQeORvDwxbp2JcwpIpFTesCzHcydP9NODesON9xxYL1p3+10fwa9X861IX5qgWD99vYn63Ul2HyoJRhDmMU/
oVYK4Ab9tBt8s6yYvO/KYZ4w8KMfU6nin6mw+XxnubYlA==
KITCHENEROKTOBERFEST NS NS1.UNIREGISTRYMARKET.LINK.
KITCHENEROKTOBERFEST NS NS2.UNIREGISTRYMARKET.LINK.
KITCHENFLOORTILE NS NS1.UNIREGISTRYMARKET.LINK.
KITCHENFLOORTILE NS NS2.UNIREGISTRYMARKET.LINK.
KITCHENTABLESET NS NS1.UNIREGISTRYMARKET.LINK.
```

Şekil 4.1 : .com uzantısına ait örnek kök dosyasından bir kesit.

Kök dosyaları üzerinde yapılan veri temizliği işleminde uzantı bazlı düzenli ifadeler (REGEX) kodu hazırlanmıştır. Bu işlem sonucunda sadece alan adları ile isim sunucularının yer aldığı bir sonuç elde edilmektedir.

...

KITCHENEROKTOBERFEST NS NS1.UNIREGISTRYMARKET.LINK.

KITCHENEROKTOBERFEST NS NS2.UNIREGISTRYMARKET.LINK.

KITCHENFLOORTILE NS NS1.UNIREGISTRYMARKET.LINK.

KITCHENFLOORTILE NS NS2.UNIREGISTRYMARKET.LINK.

...

Tekilleştirme

Alan adı kök dosyasında yapılan ikinci işlem tekilleştirme işlemidir. Tekilleştirme işlemi ile her bir alan adının sadece bir kez yer aldığı sıralı bir liste oluşturulur. Burada, normalize işlemi gerçekleştirilmekte, sol tarafta grup anahtarı olarak alan adı (uzantısız), sağ tarafta da o alan adının isim sunucuları bulunmaktadır. İsim sunucuları kendi içlerinde sıralı halde tutulmaktadır.

...

kitcheneroktoberfest [ns1.uniregistrymarket.link. | ns2.uniregistrymarket.link.]

kitchenfloortile [ns1.uniregistrymarket.link. | ns2.uniregistrymarket.link.]

...

Karşılaştırma ve dosya oluşturma

Verinin işlenmesi aşamasındaki en kritik aşama karşılaştırma aşamasıdır. Bu aşamada, mevcut tekilleştirme aşamasında oluşturulan liste ile bir önceki kök dosyasının işlenmesi sırasında oluşturulan liste karşılaştırılmaktadır. Karşılaştırma işlemi, özellikle büyük ölçekli dosyalar için çok fazla kaynak tüketen bir işlemdir. Örneğin yüz kırk milyondan fazla alan adına sahip .com uzantısı için iki dosyanın karşılaştırılmasında en hızlı ve verimli yolu bulmak oldukça önemlidir.

Karşılaştırma işleminin en hızlı ve en verimli bir şekilde gerçekleşmesi için birçok farklı yöntem, teknoloji ve kaynak kullanımı test edilmiş ve sonuç olarak büyük veri üzerinde paralel işleme yapmaya imkan sağlayan açık kaynak kodlu bir çözüm olan Apache Spark üzerinde karar kılınmıştır. Böylece .com alan adı uzantısı kök dosyası gibi büyük dosyalarla işlem yapmak mümkün olmakta, "***full outer join***" özelliği ile iki dosya arasında karşılaştırma yapılabilmektedir. Spark işlemi için ***64 GB RAM, 16CPU***'luk makineler kullanılmıştır.

Böylece, tekilleştirme aşamasında elde edilen iki sıralı dosya satır satır karşılaştırılarak yeni kaydedilen alan adları, silinen alan adları ve isim sunucusu güncellenen alan adları çıkartılmaktadır.

Burada iki dosya satır satır karşılaştırılarak üç yeni dosya oluşturulur:

1. Yeni kayıt edilen alan adları
2. Silinen alan adları
3. Güncellenen alan adları

Yeni oluşturulan dosyalar ***bz2*** formatında saklanmaktadır. Bunun sebebi bz2 sıkıştırma formatının büyük veri işlemleri için uyumlu olması, parçalara bölünüp okunabilmesi ve varsayılan olarak indeks bilgisi saklıyor olmasıdır.

Oluşturulan her bir dosya şu şekilde isimlendirilerek depolanmaktadır:

[alan-adı-uzantı-adı]-[yıl][ay][gün]-[saat].[tür].[depolanma-formati]

Örneğin 30 Haziran 2018 tarihinde, saat 05:00'de çekilen .com uzantısına ait kök dosyasından oluşturulan yeni dosyalar şu şekilde isimlendirilmiştir:

com-20180630-05.new.bz2

com-20180630-05.deleted.bz2

com-20180630-05.updated.bz2

4.1.3 Kötü niyetli alan adı tespiti için takip edilecek listenin hazırlanması

Kötü niyetli alan adı tespiti için; Türkiye'deki tüm bankalar, ödeme altyapısı sağlayıcıları, kamu kurumları, e-ticaret siteleri, online servisleri içeren bir liste hazırlanmıştır. Bu liste oltalama saldırısı yapılabilecek potansiyel kurum ve kuruluşları içermektedir.

Liste hazırlığı şu şekilde yapılmıştır:

1. Oltalama saldırısına uğrayabilecek organizasyonların belirlenmesi.
2. Listedeki her bir organizasyona ait ana alan adının listeye eklenmesi.
3. Her bir organizasyona ait diğer alan adlarının RDN (Related Domain Names) algoritması yardımıyla bulunması.
4. Her bir organizasyona ait ana alan adı ve diğer alan adlarının analiz edilmesi ve takip edilecek kelimelerin çıkartılması.
5. Kelime listesinin FDD algoritmalarına verilmesi.

Bankacılık Düzenleme ve Denetleme Kurumu internet sitesinde yer alan bankalar, elektronik para kuruluşları ve ödeme kuruluşları listeye dahil edilmiştir [51].

Listedeki organizasyonların bir kısmı birden fazla internet sitesine sahip olduğu için öncelikle ana operasyonun yürütüldüğü alan adı belirlenmiş ve listeye eklenmiştir.

Listedeki her bir organizasyonun sahip olduğu diğer alan adlarının tespit edilmesi için İlgili Alan Adları (RDN) adı verilen bir algoritma geliştirilmiştir. Bu algoritma şu şekilde çalışmaktadır:

1. RDN'ye bir email adresi girilir (organizasyonun ana alan adına ait whois bilgisinde yer alan email adresi).
2. RDN, kayıtlı 350 milyon alan adı arasında whois taraması yapar. Tersine whois (reverse whois) adı verilen bu yöntemde girilen email adresi ile kayıtlı olan diğer alan adları bulunur ve listelenir. GDPR dolayısıyla ilgili tüm alan adlarını bulmada %100 başarı sağlanamamaktadır. Bu yüzden sadece mevcut whois verileri arasında değil geçmiş whois bilgileri arasında da tarama yapılmaktadır.
3. RDN, kayıtlı 350 milyon alan adı arasında isim sunucusu taraması yapar. Burada, bir önceki aşamada bulunan organizasyonun alan adlarının yönlendirildiği isim sunucuları dikkate alınır. Bunlar genel isim sunucusu değilse (ns1.domaincontrol.com gibi), yani isim sunucusu organizasyona özel ise (ns1.organizasyon.com gibi), bu isim sunucuları ile aynı olan diğer tüm alan adları bulunur ve listelenir.
4. İsim sunucusu taraması ile bulunan alan adları arasında farklı bir email adresi ile kayıtlı alan adı varsa, bu email adresi ile 2 numaralı aşama tekrarlanır. Sonra ihtiyaç duyulması halinde 3 numaralı aşama tekrarlanır.
5. İlgili bulunan tüm alan adları listelenir.

Sonraki aşamada her bir organizasyonun sahip olduğu ana alan adı ve diğer alan adları analiz edilmiş ve FDD'de kullanılacak kelimeler çıkartılmıştır. Hazırlanan kelime listesi FDD algoritmalarında kullanılmak üzere FDD sürecine verilmiştir.

Çizelge 4.2 : FDD için takip edilen kurum ve kuruluşlar listesi.

KURULUŞ	İNTERNET SİTESİ	TÜR
AKBANK T.A.Ş.	akbank.com.tr	Banka
ALTERNATİFBANK A.Ş.	abank.com.tr	Banka
ANADOLUBANK A.Ş.	anadolubank.com.tr	Banka
ARAP TÜRK BANKASI A.Ş.	atbank.com.tr	Banka
BANK MELLAT	mellatbank.com	Banka
BANK OF CHINA TURKEY A.Ş.	bankofchina.com.tr	Banka
BURGAN BANK A.Ş.	burgan.com.tr	Banka
CITIBANK A.Ş.	citibank.com.tr	Banka
DENİZBANK A.Ş.	denizbank.com	Banka
DEUTSCHE BANK A.Ş.	deutschebank.com.tr	Banka
FİBABANKA A.Ş.	fibabanka.com.tr	Banka
HABİB BANK LİMİTED	hbl.com.tr	Banka
HSBC BANK A.Ş.	hsbc.com.tr	Banka
ICBC TURKEY BANK A.Ş.	icbc.com.tr	Banka
ING BANK A.Ş.	ingbank.com.tr	Banka
INTESA SANPAOLO S.P.A.	intesasanpaolo.com	Banka
JP MORGAN CHASE BANK NATIONAL ASSOCIATION	jpmorganchase.com	Banka
MUFG BANK TURKEY A.Ş.	-	Banka
ODEA BANK A.Ş.	odeabank.com.tr	Banka
QNB FİNANSBANK A.Ş.	qnbfinansbank.com.tr	Banka
RABOBANK A.Ş.	rabobank.com.tr	Banka
SOCIETE GENERALE S.A.	societegenerale.com.tr	Banka
ŞEKERBANK T.A.Ş.	sekerbank.com.tr	Banka
T.C. ZİRAAT BANKASI A.Ş.	ziraat.com.tr	Banka
TURKISH BANK A.Ş.	turkishbank.com	Banka
TURKLAND BANK A.Ş.	tbank.com.tr	Banka
TÜRK EKONOMİ BANKASI A.Ş.	teb.com.tr	Banka
TÜRKİYE GARANTİ BANKASI A.Ş.	garanti.com.tr	Banka
TÜRKİYE HALK BANKASI A.Ş.	halkbank.com.tr	Banka
TÜRKİYE İŞ BANKASI A.Ş.	isbank.com.tr	Banka
TÜRKİYE VAKIFLAR BANKASI T.A.O.	vakifbank.com.tr	Banka
YAPI VE KREDİ BANKASI A.Ş.	yapikredi.com.tr	Banka
AKTİF YATIRIM BANKASI A.Ş.	aktifbank.com.tr	Banka
BANKPOZİTİF KREDİ VE KALKINMA BANKASI A.Ş.	bankpozitif.com.tr	Banka
DİLER YATIRIM BANKASI A.Ş.	dilerbank.com.tr	Banka
GSD YATIRIM BANKASI A.Ş.	gsdbank.com.tr	Banka
İLLER BANKASI A.Ş.	ilbank.gov.tr	Banka

Çizelge 4.2 (Devam): FDD İçin Takip Edilen Kurumlar Listesi.

KURULUŞ	İNTERNET SİTESİ	TÜR
İSTANBUL TAKAS VE SAKLAMA BANKASI A.Ş.	takasbank.com.tr	Banka
MERRILL LYNCH YATIRIM BANK A.Ş.	ml.com.tr	Banka
NUROL YATIRIM BANKASI A.Ş.	nurolbank.com.tr	Banka
PASHA YATIRIM BANKASI A.Ş.	pashabank.com.tr	Banka
STANDARD CHARTERED YATIRIM BANKASI T.A.Ş.	standardchartered.com.tr	Banka
TÜRKİYE İHRACAT KREDİ BANKASI A.Ş.	eximbank.gov.tr	Banka
TÜRKİYE KALKINMA VE YATIRIM BANKASI A.Ş.	kalkinma.com.tr	Banka
TÜRKİYE SINAİ KALKINMA BANKASI A.Ş.	tskb.com.tr	Banka
ALBARAKA TÜRK KATILIM BANKASI A.Ş.	albarakaturk.com.tr	Banka
KUVEYT TÜRK KATILIM BANKASI A.Ş.	kuveytturk.com.tr	Banka
TÜRKİYE EMLAK KATILIM BANKASI A.Ş.	emlakbank.com.tr	Banka
TÜRKİYE FİNANS KATILIM BANKASI A.Ş.	turkiyefinans.com.tr	Banka
VAKIF KATILIM BANKASI A.Ş.	vakifkatilim.com.tr	Banka
ZİRAAT KATILIM BANKASI A.Ş.	ziraatkatilim.com.tr	Banka
ADABANK A.Ş.	adabank.com.tr	Banka
BİRLEŞİK FON BANKASI A.Ş.	fonbank.com.tr	Banka
AKÖDE EP VE ÖH A.Ş.	-	Elektronik Para
BELBİM EP VE ÖH A.Ş.	belbim.com.tr	Elektronik Para
BİRLEŞİK ÖH VE EP A.Ş.	birlesikodeme.com	Elektronik Para
CEMETE EP VE ÖH A.Ş.	cemete.com.tr	Elektronik Para
D ÖDEME EP VE ÖH A.Ş.	hepsipay.com	Elektronik Para
HIZLIPARA ÖH VE EP A.Ş.	payporter.com	Elektronik Para
İNİNAL Ö VE EPH A.Ş.	ininal.com	Elektronik Para
İYZi Ö VE EPH A.Ş.	iyzico.com	Elektronik Para
PALADYUM EP VE ÖH A.Ş.	peppara.com	Elektronik Para
PAPARA EP VE ÖH A.Ş.	papara.com	Elektronik Para
TURK ELEKTRONİK PARA A.Ş.	turkpara.com.tr	Elektronik Para
TURKCELL Ö VE EPH A.Ş.	-	Elektronik Para
VODAFONE EP VE ÖH A.Ş.	eparahizmetleri.com	Elektronik Para
WİRECARD Ö VE EPH A.Ş.	wirecard.com.tr	Elektronik Para
AYPARA ÖDEME KURULUŞU A.Ş.	ipara.com	Ödeme Kuruluşu
BPN ÖDEME KURULUŞU A.Ş.	bpn.com.tr	Ödeme Kuruluşu
BURADAÖDE ÖDEME KURULUŞU A.Ş.	buradaode.com.tr	Ödeme Kuruluşu
CEO ÖDEME HİZMETLERİ A.Ş.	faturatim.com.tr	Ödeme Kuruluşu
EFİX ÖDEME HİZMETLERİ A.Ş.	efixfatura.com.tr	Ödeme Kuruluşu
ELEKSE YETKİLİ VEZNE ÖDEME KURULUŞU A.Ş.	elekse.com	Ödeme Kuruluşu
FATURAKOM ÖDEME HİZMETLERİ A.Ş.	faturakom.com	Ödeme Kuruluşu
FATURAMATİK ÖDEME KURULUŞU A.Ş.	faturamatik.com.tr	Ödeme Kuruluşu
FÖY FATURA ÖDEME KURULUŞU A.Ş.	faturaodemeyeri.com.tr	Ödeme Kuruluşu
GLOBAL ÖDEME HİZMETLERİ A.Ş.	getmoneyglobal.com	Ödeme K.

Çizelge 4.2 Devam: FDD İçin Takip Edilen Kurumlar Listesi

KURULUŞ	İNTERNET SİTESİ	TÜR
GÖNDERAL ÖDEME HİZMETLERİ A.Ş.	gonder-al.com	Ödeme Kuruluşu
İSTANBUL ÖDEME KURULUŞU A.Ş.	istanbulodeme.com	Ödeme Kuruluşu
KLON ÖDEME KURULUŞU A.Ş.	payby.me	Ödeme Kuruluşu
MOKA ÖDEME KURULUŞU A.Ş.	moka.com	Ödeme Kuruluşu
MONEYGRAM TURKEY ÖDEME HİZMETLERİ A.Ş.	global.moneygram.com	Ödeme Kuruluşu
N KOLAY ÖDEME KURULUŞU A.Ş.	nkolayislem.com.tr	Ödeme Kuruluşu
NESTPAY ÖDEME HİZMETLERİ A.Ş.	paratika.com.tr	Ödeme Kuruluşu
OCTET EXPRESS ÖDEME KURULUŞU A.Ş.	octet.com.tr	Ödeme Kuruluşu
ÖDEAL ÖDEME KURULUŞU A.Ş.	ode.al	Ödeme Kuruluşu
PAY FİX ÖDEME HİZMETLERİ A.Ş.	-	Ödeme Kuruluşu
PAYNET ÖDEME HİZMETLERİ A.Ş.	paynet.com.tr	Ödeme Kuruluşu
PAYTR ÖDEME HİZMETLERİ A.Ş.	paytr.com	Ödeme Kuruluşu
PAYTREK ÖDEME KURULUŞU HİZMETLERİ A.Ş.	paytrek.com	Ödeme Kuruluşu
PAYU ÖDEME KURULUŞU A.Ş.	payu.com.tr	Ödeme Kuruluşu
PRATİK İŞLEM ÖDEME KURULUŞU A.Ş.	pratikislem.com.tr	Ödeme Kuruluşu
RIA TURKEY ÖDEME KURULUŞU A.Ş.	riafinancial.com	Ödeme Kuruluşu
SENDER ÖDEME HİZMETLERİ A.Ş.	send-r.com	Ödeme Kuruluşu
TAM FATURA ÖDEME HİZMETLERİ A.Ş.	tamfatura.com	Ödeme Kuruluşu
TREND ÖDEME KURULUŞU A.Ş.	payguru.com	Ödeme Kuruluşu
TT ÖDEME HİZMETLERİ A.Ş.	turktelekomodemehizmetleri.com	Ödeme Kuruluşu
UPT ÖDEME HİZMETLERİ A.Ş.	upt.com.tr	Ödeme Kuruluşu
VEZNE24 TAHSİLAT SİSTEMLERİ VE ÖH A.Ş.	v24.com.tr	Ödeme Kuruluşu
VİZYON TAHSİLAT SİSTEMLERİ VE ÖH A.Ş.	faturavizyon.com	Ödeme Kuruluşu

4.2 Kötü Niyetli Alan Adı Tespiti (FDD) Algoritması

4.2.1 Kötü niyetli alan adı tespitinde benzerlik etiketlemeleri

Oltalama amacıyla kullanılabilecek kötü niyetli alan adlarının tespitinde temel çıkış noktası, kötü niyetli alan adının hedef alınan alan adına benzer olmasıdır.

Hedef alınan alan adına benzer olan alan adlarını tespit etmek için FDD adı verilen bir algoritma geliştirilmiştir. Bu algoritmada yedi farklı benzerlik etiketlemesi yapılmaktadır:

1. Tam Eşleşen (Exact)
2. Karıştırılabilir Tam Eşleşen (Confusable Exact)
3. İçeren (Contains)
4. Bulanık (Fuzzy)
5. Bulanık İçeren (Fuzzy Contains)
6. Karıştırılabilir Bulanık (Confusable Fuzzy)
7. Karıştırılabilir Bulanık İçeren (Confusable Fuzzy Contains)

Söz konusu benzerlik tanımlamaları yapılırken mesafe algoritmalarındaki ekleme, çıkarma ve yer değiştirme bilgilerine bağlı olarak etiketleme yapılmaktadır. Yer değiştirme karıştırılabilir ve normal olmak üzere ikiye ayrılmaktadır. Karıştırılabilir olanlar yer değiştirme olarak sayılmamaktadır ve bunların bulanıklık mesafesine etkisi yoktur.

Etiketleme işlemi yapılırken alan adı uzantısı hariç bırakılmaktadır. Sadece ikinci derece alan adı kısmına bakılmaktadır. Aşağıda her bir benzerlik etiketlemesi örneklerle açıklanacaktır. Örnek için ***Bank of Example*** adında farazi bir banka ve bu farazi bankaya ait ***bankofexample.com*** farazi alan adı kullanılacaktır.

Tam eşleşen benzerlik (exatch)

Herhangi bir ekleme, çıkarma ya da yer değiştirmenin söz konusu olmadığı benzerlikler tam eşleşen benzerlik olarak etiketlenmektedir.

Örnek

Orijinal alan adı: ***bankofexample****.com*

Tam eşleşen benzerlik: ***bankofexample****.club*

Tam eşleşen benzerlik: ***bankofexample****.ml*

Karıştırılabilir tam eşleşen benzerlik (confusable exatch match)

Herhangi bir ekleme ya da çıkarmanın söz konusu olmadığı ancak karıştırılabilir karakterlerle yer değiştirmenin mevcut olduğu benzerlikler karıştırılabilir tam eşleşen benzerlik olarak etiketlenmektedir.

Karıştırılabilir karakterler Unicode Consortium tarafından hazırlanan Karıştırılabilir Karakterler (Confusable Characters) listesinde yer alan karakterlerdir [52].

Örneğin a harfi için karıştırılabilir karakterlerden bazıları şunlardır:

ɑ, a, 𝛂 a ａ 𝒶 𝐚 𝗮 𝙖 a 𝞪 a ɑ ɑ ɑ 𝜶 𝘢 𝛼 ɑ 𝙖 𝝰 𝛼 a

Örnek

Orijinal alan adı: ***bankofexample****.com*

Karıştırılabilir tam eşleşen benzerlik: ***bankofexamplê****.com*

Karıştırılabilir tam eşleşen benzerlik: ***bankofexample****.com*

İçeren benzerlik (contains)

Herhangi bir çıkarma ya da yer değiştirmenin söz konusu olmadığı ancak eklemenin mevcut olduğu benzerlikler içeren benzerlik olarak etiketlenmektedir.

Örnek

Orijinal alan adı: ***bankofexample****.com*

İçeren benzerlik: ***bankofexample1****.com*

İçeren benzerlik: ***bbankofexample****.com*

Bulanık benzerlik (fuzzy)

Ekleme, çıkarma ya da yer değiştirmenin söz konusu olduğu ancak bunların toplamının tanımlı en fazla bulanıklık mesafesinden büyük olmadığı durumlar bulanık benzerlik olarak etiketlenmektedir.

En fazla bulanıklık mesafesi, karşılaştırılan iki öğenin arasındaki değişim miktarını belirten bir tanımdır. Örneğin bankofexample.com ve bankofexampla.com arasındaki bulanıklık mesafesi 1'dir çünkü sadece bir karakter değişimi vardır.

Örnek

Orijinal alan adı: ***bankofexample****.com*

Bulanık benzerlik: ***bankofexamp1e****.com*

Bulanık benzerlik: ***bankof-example****.com*

Bulanık benzerlik: ***banofexample****.com*

Bulanık içeren benzerlik (fuzzy contains)

Ekleme, çıkarma ya da yer değiştirmenin söz konusu olduğu ve bunların toplamının tanımlı en fazla bulanıklık mesafesinden büyük olduğu durumlar bulanık içeren benzerlik olarak etiketlenmektedir.

Örnek

Orijinal alan adı: ***bankofexample****.com*

Bulanık içeren benzerlik: ***bank-off-exanb1e****.com*

Karıştırılabilir bulanık benzerlik (confusable fuzzy)

Ekleme, çıkarma ya da yer değiştirmenin söz konusu olduğu, bunların toplamının tanımlı en fazla bulanıklık mesafesinden büyük olmadığı ve en az bir tane karıştırılabilir değişiklik olduğu durumlar karıştırılabilir bulanık benzerlik olarak etiketlenmektedir.

Örnek

Orijinal alan adı: ***bankofexample****.com*

Karıştırılabilir bulanık benzerlik: ***bank-of-examplê****.com*

Karıştırılabilir bulanık içeren benzerlik (confusable fuzzy contains)

Ekleme, çıkarma ya da yer değiştirmenin söz konusu olduğu, bunların toplamının tanımlı en fazla bulanıklık mesafesinden büyük olduğu ve en az bir tane karıştırılabilir değişiklik olduğu durumlar karıştırılabilir bulanık içeren benzerlik olarak etiketlenmektedir.

Örnek

Orijinal alan adı: ***bankofexample****.com*

Karıştırılabilir bulanık içeren benzerlik: ***bank-off-exanb1ê****.com*

Benzerlik tespitinde yardımcı kelimeler (helper words)

Tez çalışması kapsamında geliştirilen kötü niyetli alan adı tespitinde takip edilen asıl kelimelere ek olarak yardımcı kelimelere de ihtiyaç duyulmuştur. Bu yardımcı kelimeler tespit sürecinde kesinliği artırıcı kelimelerdir.

Örnek

Takip edilen kelime: ***example***

Pozitif yardımcı kelimeler: ***bank, login, creditcard, secure, update***

5. DEĞERLENDİRME

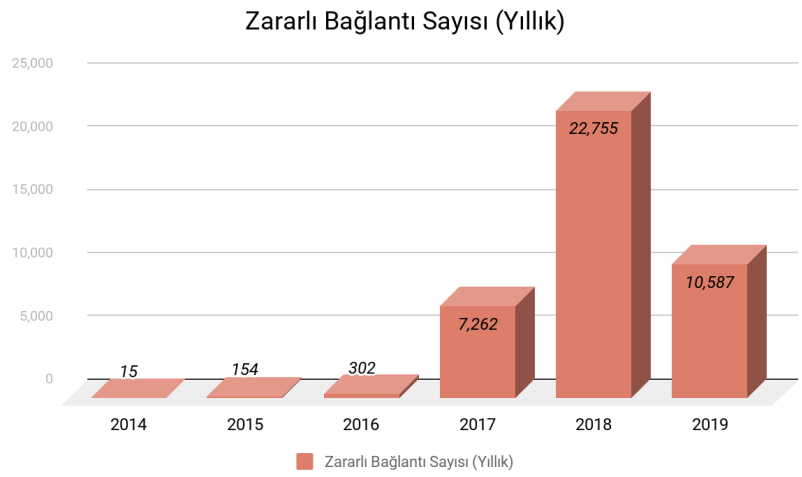
Bu bölümde, bir önceki bölümde önerilen kötü niyetli alan adı tespit yöntemi FDD'nin gerçek verilerle test edilmesi sonucunda ortaya çıkan sonuçlara yer verilmiştir.

5.1 USOM Zararlı Bağlantılar Listesi

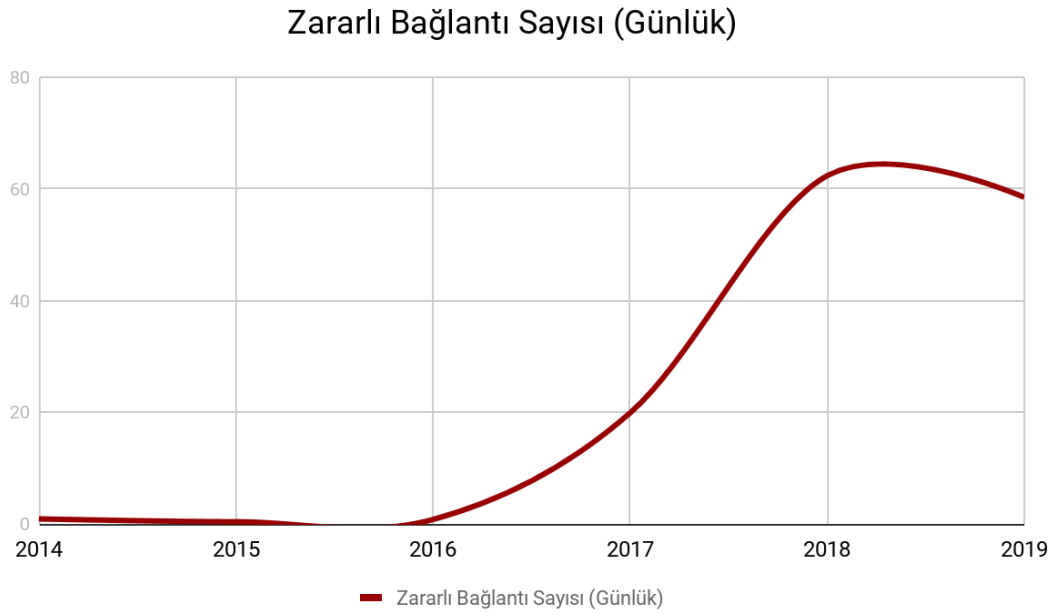
Bilgi Teknolojileri ve İletişim Kurumu bünyesinde çalışmalar yürüten Ulusal Siber Olaylara Müdahale Merkezi (USOM), 15 Aralık 2014'den itibaren, internet sitesi usom.gov.tr aracılığıyla zararlı bağlantıların listesini (zararlı yazılım veya kod içeren bağlantılar) yayınlamaktadır [53].

5.2 USOM Zararlı Bağlantılar Listesine Ait İstatistikler

Tez çalışması kapsamında, zararlı bağlantıların yayınlanmaya başlandığı 15 Aralık 2014 ile 30 Haziran 2019 arasında yayınlanan toplam 41,075 zararlı bağlantı analiz edilmiştir. Buna göre; 2014-2016 yılları arasında eklenen zararlı bağlantı sayısı birkaç yüz ile sınırlı kalmış, 2017 yılında günlük ortalama yirmi bağlantı eklenmiş, 2018 itibarıyla bu sayı günlük altmışa kadar yükselmiştir.



Şekil 5.1 : USOM ZB Listesi'ne yıllara göre eklenen bağlantı sayısı.



Şekil 5.2 : USOM ZB Listesi'ne eklenen günlük ortalama bağlantı sayısı.

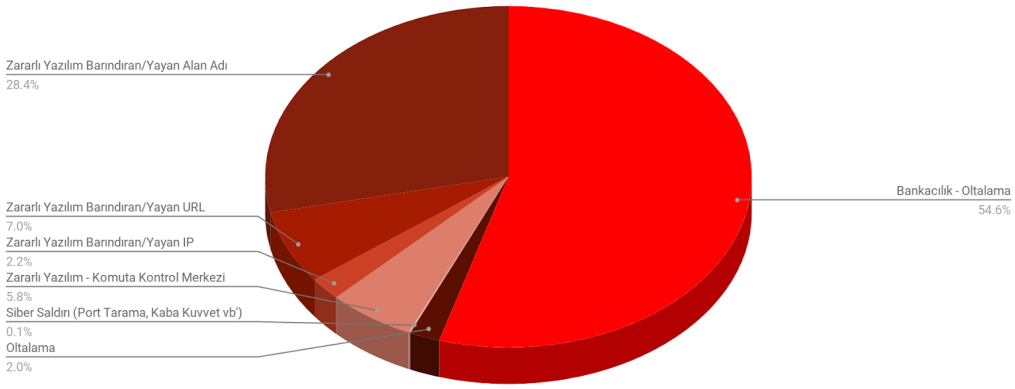
Çizelge 5.1 : USOM ZB Listesi'ne yıllara göre eklenen bağlantı sayısı.

Yıl	Eklenen Zararlı Bağlantı Sayısı	Eklenen Zararlı Bağlantı Sayısı (Günlük Ortalama)
2014	15	0.9
2015	154	0.4
2016	302	0.8
2017	7,262	19.9
2018	22,755	62.3
2019	10,587	58.5

Zararlı Bağlantı Kategorileri

USOM Zararlı Bağlantılar Listesi yedi ana kategori altında listelenmektedir:

1. Bankacılık – Oltalama
2. Oltalama
3. Zararlı Yazılım - Komuta Kontrol Merkezi
4. Zararlı Yazılım Barındıran/Yayan Alan Adı
5. Zararlı Yazılım Barındıran/Yayan URL
6. Zararlı Yazılım Barındıran/Yayan IP
7. Siber Saldırı (Port Tarama, Kaba Kuvvet vb.)



Şekil 5.3 : USOM Zararlı Bağlantılar Listesi'nin kategorilere dağılımı.

Tez çalışması kapsamında sadece ***Bankacılık – Oltalama*** kategorisi değerlendirmeye alınmıştır.

Zararlı Bağlantı Tespit Kaynakları

USOM Zararlı Bağlantılar Listesi üç ana kaynaktan beslenmektedir:

1. USOM
2. SOME
3. İHBAR

Çizelge 5.2 : USOM ZB Listesi'ne kaynak türüne göre eklenen bağlantı sayısı.

Kaynak	Eklenen Zararlı Bağlantı Sayısı*
USOM	13,158
SOME	8,401
İHBAR	19,516

**30 Haziran 2019'a kadar.*

Buna göre zararlı bağlantıların ***19,516***'sı ihbar yoluyla, ***8,401***'i SOME'ler aracılığıyla, ***13,158***'i ise USOM'un kendi çalışmaları ile tespit edilmiştir.

5.3 Kötü Niyetli Alan Adı Tespiti'nin (FDD) USOM Zararlı Bağlantılar Listesinde Uygulanması

Tez çalışması kapsamında ortaya konulan kötü niyetli alan adlarının tespit edilmesi (FDD) yöntemi USOM Zararlı Bağlantılar Listesi üzerinde çalıştırılarak önerilen yöntemin doğruluğu ve verimi test edilmiştir.

1 Ocak 2018 ile 30 Haziran 2019 tarihleri arasında kayıt edilen tüm alan adları FDD sistemi ile analiz edilmiş ve USOM Zararlı Bağlantılar Listesi'nde 1 Ocak 2018 ile 30 Haziran 2019 tarihleri arasında listeye eklenmiş olan, Bankacılık - Oltalama kategorisindeki toplam ***17,613*** adet zararlı bağlantı ile karşılaştırılmıştır.

Bu bağlantılar, tez çalışması kapsamında geliştirilen Alan Adı Doğrulayıcı (DNV) adlı programa verilmiş, IP adresleri ve farklı türdeki URL'lerin elenmesi ile ***16,953*** alan adından oluşan yeni bir liste hazırlanmıştır.

USOM Zararlı Bağlantılar Listesi'nde FDD uygulanarak temelde iki sorunun cevabı aranmıştır:

1. USOM Zararlı Bağlantılar Listesi'nde yer aldığı halde tez çalışması kapsamında önerilen kötü niyetli alan adı tespit sistemi FDD tarafından bulunamayan alan adı var mıdır?
2. Alan adlarının USOM Zararlı Bağlantılar Listesi'ne eklenme tarih ve saati ile kötü niyetli alan adı tespit sisteminin bulduğu tarih ve saat arasında ortalama ne kadar fark bulunmaktadır?

İlk sorunun cevabı ile önerilen kötü niyetli alan adı tespit sistemi FDD'deki eksikler bulunabilir, tespit edilemeyen alan adlarının neden tespit edilemediği görülebilir ve bunları tespit etmek için FDD'de ne gibi geliştirmeler yapılması gerektiği ortaya konulabilir.

İkinci sorunun cevabı ile, önerilen kötü niyetli alan adı tespit sistemi FDD'nin ne kadar hızlı olduğu ve gerçek hayatta uygulandığı takdirde ne gibi sonuçlar alınabileceği görülebilir.

Buna göre; ***16,953*** alan adından oluşan zararlı bağlantılar listesinde FDD uygulandığında bunların ***15,088***'inin tespit edildiği, ***1,865***'inin ise tespit edilemediği görülmüştür. Yani doğru pozitif oranı ***%89***'dur.

FDD tarafından tespit edilemeyen alan adları incelendiğinde bunların ***%43***'ünün rastgele (anlamsız) alan adları olduğu görülmüştür. Anlamsız (rastgele) alan adları tez çalışmasında önerilen yöntemin kapsamı dışında tutulduğu için kalan ***%57***'lik yani toplamdaki ***%6***'lık kısım dikkate alınmış ve tespit edilememe sebepleri araştırılmıştır. Buna göre tespit edilemeyen alan adlarının ***profilime-kimbaktitr-com.ga***, ***trafikcezaodemefiyatlar.com, anketteklifi.com, yazgunleriguzel2019.com trcarpente.com*** ve ***sarayrentacar.tk*** gibi genel alan adları olduğu ve takip edilen finans kurumlarını hedef almadığı görülmüştür. Bunun yanı sıra kötü niyetli alan adı takip sistemi FDD'nin daha iyi sonuçlar vermesi için sadece finans kurumlarını değil kötü niyetli olarak kullanılabilecek diğer genel kelimelerin de tek başına takip edilmesi gerektiği görülmüştür. Örneğin, ***bireysel***, ***musteri***, ***ozel***, ***firsat***, ***yaz***, ***hediye, duyuru, giris, basvuru, iade, kampanya*** gibi kelimeler, (herhangi bir finans kurumunun ismini içermeden) oltalama saldırılarında kullanılan alan adlarında sıklıkça görülmüştür. Bunun sonucunda kötü niyetli alan adı tespit sistemi FDD'nin tespit oranını artırmak için, bu tür kelimelerden oluşan bir liste hazırlanmıştır.

Kötü niyetli alan adı tespit sistemi FDD'nin tespit ettiği alan adların, USOM Zararlı Bağlantılar Listesi'ndeki kaynaklarına göre incelendiğinde şu görülmüştür:

Kaynağı ***SOME***'ler olan zararlı bağlantıların ***%92***'si, kaynağı ***USOM*** olan zararlı bağlantıların ***%85***'i, kaynağı ***İHBAR*** olan zararlı bağlantıların ise ***%87***'si FDD tarafından tespit edilmiştir.

Çizelge 5.3 : USOM ZB Listesi – FDD doğru tespit etme oranları.

Kaynak	**FDD'nin Doğru Tespit Etme Oranı**
TOPLAM	***%89***
USOM	%85
SOME	%92
İHBAR	%87

Alan adlarının USOM Zararlı Bağlantılar Listesi'ne eklenme tarih ve saati ile kötü niyetli alan adı tespit sistemi FDD'nin tespit ettiği tarih ve saat arasındaki fark incelendiğinde FDD'nin her bir zararlı bağlantıyı ortalama ***2.67 gün*** yani ***64 saat*** daha erken tespit ettiği görülmüştür. Bu fark; kaynağı ***SOME***'ler olan zararlı bağlantılarda ***2.21 gün*** yani ***53 saat***, kaynağı ***USOM*** olan zararlı bağlantılarda ***3.85 gün*** yani ***92 saat***, kaynağı ***İHBAR*** olan zararlı bağlantılarda ise ***2.96 gün*** yani ***71 saat*** olarak görülmüştür.

Çizelge 5.4 : USOM ZB Listesi – FDD tespit süresi farkı.

Kaynak	FDD Tespit Süresi Farkı (saat)
GENEL ORTALAMA	***64***
USOM	92
SOME	53
İHBAR	71

Bu farkın temel sebebi FDD'nin zararlı bağlantıları henüz daha alan adı yeni kayıt edilmişken tespit etmesidir. Bilindiği üzere, USOM herhangi bir alan adında aktif oltalama saldırısı yoksa o alan adını zararlı bağlantı olarak nitelendirmemektedir. Bu durum oltalama saldırılarının henüz başlamadan engellenmesinin önüne geçmekte ve önleyici bir hizmet verilmemektedir.

Bir oltalama saldırısında, zararlı bağlantının aktif olduğu her saat sadece 1 kişinin zarar gördüğü varsayılsa bile, tez çalışmasının yapıldığı ve önerilen yöntemin test edildiği 18 ay boyunca erken tespit edilen 15,080 alan adına erişim engellendiği taktirde, ***dokuz yüz altmış beş bin*** kişinin bu saldırılardan zarar görmesinin engellenebileceği görülmektedir. Bu durum kötü niyetli alan adlarının erken tespitinin önemini ortaya koymaktadır.

6. SONUÇ VE ÖNERİLER

Kötü niyetli alan adı tespitinde önerilen yöntemin gerçek veriler üzerinde test edildiğinde %89'un üzerinde başarı göstermesi oldukça önemlidir. Bu tez çalışması kapsamında önerilen yöntemin geliştirilerek gerçek bir çözüm olarak kullanılması mümkündür.

Tez çalışması, Türkiye'deki kullanıcıları hedef alan saldırıları önlemeye yönelik bir tespit çalışması olmasına rağmen aynı yöntem diğer ülkeler ve diller için de kullanılabilir ve bütüncül bir çözüm meydana getirilebilir.

6.1 Gelecek Çalışmalar

Kötü niyetli alan adı tespit sisteminde kural bazlı bir yöntem önerilmiş ve önerilen yöntemin gerçek verilerden oluşan USOM Zararlı Bağlantılar Listesi'nde uygulanması ile %89'un üzerinde başarı sağladığı görülmüştür. Tez çalışması kapsamındaki tüm çalışmalar kötü niyetli (fraudulent) alan adı tespitine yöneliktir. En başta belirtildiği üzere oltalama saldırıları sadece orijinal alan adına benzeyen kötü niyetli alan adları ile yapılmamakta, farklı türde alan adları ve ele geçirilmiş gerçek alan adları da kullanılmaktadır. Yapılacak yeni çalışmalarla tüm bunları kapsayan bir sistem oluşturulabilir.

Kötü niyetli alan adlarını tespit için önerilen yöntem geliştirilmeye çok açıktır. Geliştirilebilecek özelliklerden bazıları şunlardır:

1. Doğal Dil İşleme (Natural Language Processing (NLP)) ile desteklenmesi.
2. Klavye yazım benzerlik kontrolü (QWERTY klavye ve diğerleri).
3. Rastgele (anlamsız) alan adlarının kontrol edilmesi.
4. Tüm yeni kaydedilen alan adları için içerik kontrolü.
5. Her bir alan adı için bilinen URL yollarının kontrol edilmesi.
6. Alan adı harici özelliklerin kullanılması (whois, URL, IP, içerik, logo, kaynak kod, CSS, JS gibi).

7. Makine öğrenmesi yöntemlerinin kullanılması.

KAYNAKLAR

[1] **URL-1** <*https://dofo.com/symbolics.com*>, alındığı tarih: 01.05.2019.

[2] **URL-2** <*https://dofo.com/* >, alındığı tarih: 01.06.2019.

[3] **URL-3** <*https://newgtlds.icann.org/en/program-status/delegated-strings*>, alındığı tarih: 01.05.2019.

[4] **URL-4** <*https://en.wikipedia.org/wiki/List_of_ISO_3166_country_codes*>, alındığı tarih: 01.05.2019.

[5] **URL-5** <*https://www.verisign.com/en_US/domain-names/dnib/index.xhtml*>, alındığı tarih: 01.05.2019.

[6] **URL-6** <*https://www.verisign.com/en_US/domain-names/com-domain-names/index.xhtml*>, alındığı tarih: 01.05.2019.

[7] **URL-7** <*https://www.icann.org/registrar-reports/accredited-list.html*>, alındığı tarih: 01.05.2019.

[8] **URL-8** <*https://www.iana.org/assignments/character-sets/character-sets.xhtml*>, alındığı tarih: 01.05.2019.

[9] **URL-9** <*https://home.unicode.org/basic-info/overview/*>, alındığı tarih: 01.05.2019.

[10] **URL-10** <*https://www.icann.org/resources/pages/idn-guidelines-2011-09-02-en*>, alındığı tarih: 01.05.2019.

[11] **URL-11** <*https://gdpr.eu/*>, alındığı tarih: 01.05.2019.

[12] **URL-12** <*https://www.icann.org/get-started*>, alındığı tarih: 01.05.2019.

[13] **URL-13** <*https://www.iana.org/about*>, alındığı tarih: 01.05.2019.

[14] **URL-14** <*https://btk.gov.tr*>, alındığı tarih: 01.05.2019.

[15] **URL-15** <*https://www.usom.gov.tr/hakkimizda.html*>, alındığı tarih: 01.05.2019.

[16] **URL-16** <*https://www.phishing.org/what-is-phishing*>, alındığı tarih: 01.05.2019.

[17] **URL-17** <*https://apwg.org*>, alındığı tarih: 01.05.2019.

[18] **URL-18** <*https://docs.apwg.org/reports/apwg_trends_report_q1_2019.pdf*>, alındığı tarih: 01.05.2019.

[19] **URL-19** <*https://www.phishlabs.com/*>, alındığı tarih: 01.05.2019.

[20] **URL-20** <*https://info.phishlabs.com/blog/2019-phishing-trends-intelligence-report-the-evolving-threat* >, alındığı tarih: 01.05.2019.

[21] **URL-21** <*https://www.egm.gov.tr/siber*>, alındığı tarih: 15.01.2019.

[22] **URL-22** <*https://www.opendns.com*>, alındığı tarih: 01.05.2019.

[23] **URL-23** <*http://phishtank.com/*>, alındığı tarih: 01.05.2019.

[24] **URL-24** *https://openphish.com/*>, alındığı tarih: 01.05.2019.

[25] **URL-25** <*https://www.comodo.com/lab/index.php*>, alındığı tarih: 01.05.2019.

[26] **URL-26** <*https://phishbank.org/*>, alındığı tarih: 01.05.2019.

[27] **URL-27** <*https://urlscan.io/*>, alındığı tarih: 01.05.2019.

[28] **URL-28** <*https://www.google.com/*>, alındığı tarih: 01.05.2019.

[29] **URL-29** <*https://www.virustotal.com/*>, alındığı tarih: 01.05.2019.

[30] **URL-30** <*https://safebrowsing.google.com/*>, alındığı tarih: 01.05.2019.

[31] **URL-31** <*https://www.netcraft.com/*>, alındığı tarih: 01.05.2019.

[32] **URL-32** <*https://www.phishlabs.com/digital-risk-protection/domain-monitoring/*>, alındığı tarih: 01.05.2019.

[33] **URL-33** <*https://www.domaintools.com/products/*>, alındığı tarih: 01.05.2019.

[34] **URL-34** <*https://www.csisgroup.com/prevent-phishdb/*>, alındığı tarih: 01.05.2019.

[35] **URL-35** <*https://www.markmonitor.com/solutions/defend-your-brand/antifraud-protection-against-phishing/*>, alındığı tarih: 01.05.2019.

[36] **URL-36** <*https://www.rsa.com/en-us/products/fraud-prevention/phishing-protection*>, alındığı tarih: 01.05.2019.

[37] **URL-37** <*https://www.proofpoint.com/us/products/digital-risk-protection/web-domain-fraud-monitoring*>, alındığı tarih: 01.05.2019.

[38] **URL-38** <*https://www.lookingglasscyber.com/products/cyber-threat-intelligence/technical-threat-indicators/*>, alındığı tarih: 01.05.2019.

[39] **URL-39** <*https://www.redmarlin.ai/*>, alındığı tarih: 01.05.2019.

[40] **URL-40** <*https://segasec.com/*>, alındığı tarih: 01.05.2019.

[41] **Chiba, D., Akiyama, M., Yagi, T., Hato, K., Mori, T., and Goto, S.,** (2018): DomainChroma: Building actionable threat intelligence from malicious domain names. Computer & Security, vol. 77, no. 2018, pp. 138-161.

[42] **Wang, W., and Shirley, K. E.,** (2015): Breaking Bad: Detecting malicious domains using word segmentation. Arxiv 2015.

[43] **Zhao, H., Chang, Z., Bao, G., and Zeng, X.,** (2019): Malicious Domain Names Detection Algorithm Based on N-Gram. Journal of Computer Networks and Communications, vol. 2019.

[44] **URL-41** <*https://www.alexa.com/topsites/*>, alındığı tarih: 01.05.2019.

[45] **Bilge, L., Kirda, E., Kruegel, C., and Balduzzi, M.,** (2011): EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis. Network and Distributed System Security Symposium, San Diego, USA.

[46] **Agten, P., Joosen, W., Piessens, F., and Nikiforakis, N.,** (2015): Seven Months' Worth of Mistakes: A Longitudinal Study of Typosquatting Abuse. Network and Distributed System Security Symposium, San Diego, USA.

[47] **Vissers, T., Spooren, J., Agten, P., Jumpertz, D., Jannsen, P., Wesemael, M. V., Piessens, F., Joosen, W., and Desmet, L.,** (2017): Exploring the ecosystem of malicious domain registrations in the .eu TLD. International Symposium on Research in Attacks, Intrusions, and Defenses, Atlanta, USA.

[48] **Chiba, D., Yagi, T., Akiyama, M., Shibahara, T., Mori, T., and Goto, S.,** (2017): DomainProfiler: toward accurate and early discovery of domain names abused in future. International Journal of Information Security, vol. 17, no. 6, pp. 661-680.

[49] **Zhang, P., Liu, T., Zhang, Y., Ya, J., Shi, J., and Wang, Y.,** (2017): DomainWatcher: Detecting Malicious Domains Based on Local and Global Textual Features. International Conference on Computational Science, Zurich, Switzerland.

[50] **Hao, S., Kantchelian, A., Miller, B., Paxson, V., and Feamster, N.,** (2016): PREDATOR: Proactive Recognition and Elimination of Domain Abuse at Time-Of-Registration. ACM Conference on Computer and Communications Security, Vienna, Austria.

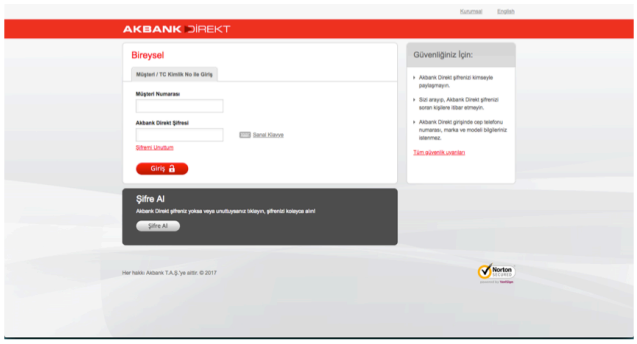
[51] **URL-42** <*https://www.bddk.org.tr/Kuruluslar-Kategori/Bankalar/1/*>, alındığı tarih: 01.05.2017.

[52] **URL-43** <*https://www.unicode.org/Public/security/8.0.0/confusables.txt*>, alındığı tarih: 01.12.2017.

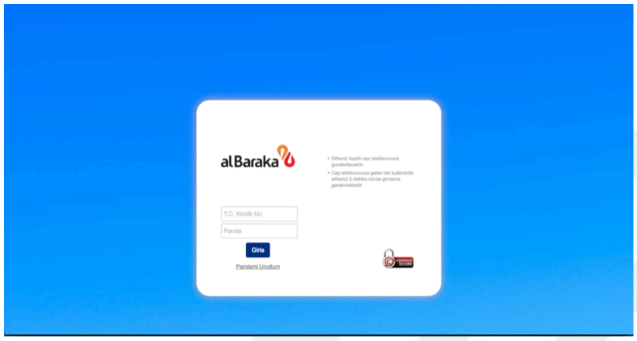
EKLER

EK A.1 : Tez çalışması sırasında karşılaşılan oltalama internet sitelerine ait ekran görüntüleri.

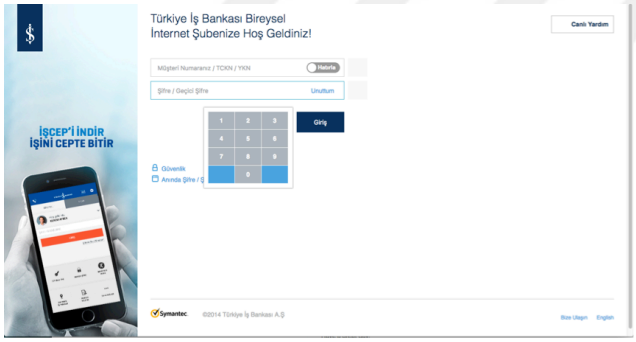
EK A.1

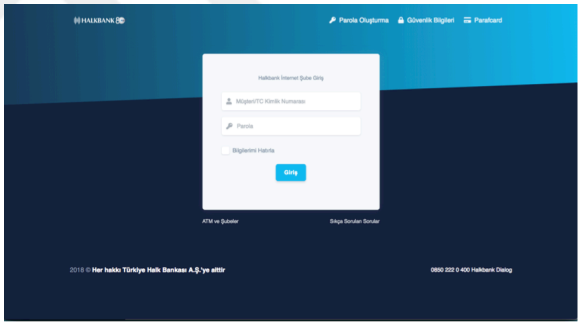

AKBANK DİREKT
Bireysel
Müşteri / TC Kimlik No ile Giriş
Müşteri Numarası
Akbank Direkt Şifresi
Sanal Klavye
Şifremi Unuttum
Giriş
Güvenliğiniz İçin:
Şifre Al
Şifre Al

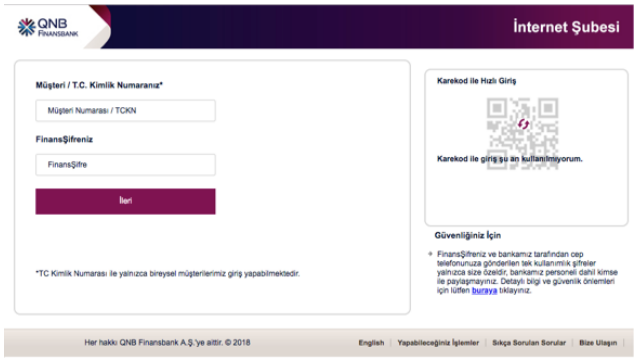

albaraka
SİZİN İÇİN
İŞİNİZ İÇİN
HAKKIMIZDA
Hayatı kolaylaştıran mobil teknolojiler Albaraka'da!
Yepyeni Albaraka Mobil 'i hemen indirin,
KÂR ORANLARI
PİYASALAR
KURLARIMIZ
HESAPLAMA ARAÇLARI
Kapımız herkese açık, masrafa kapalı!
%25 indirim!

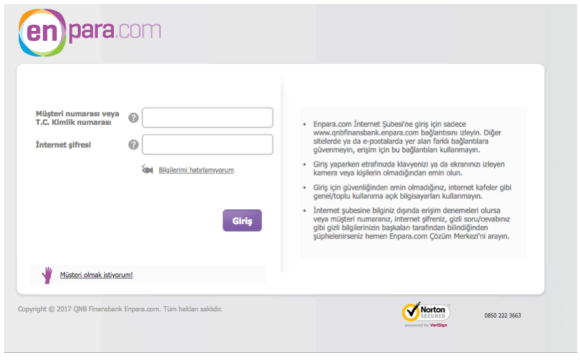

albaraka
T.C. Kimlik No
Parola
Giriş
Parolamı Unuttum

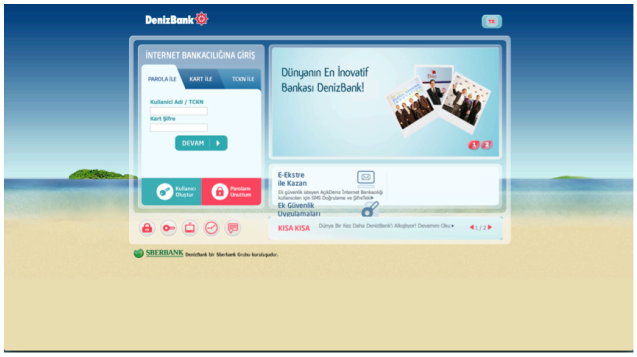

Garanti
pay
Lütfen Bilgileri eksiksiz ve doğru doldurunuz.
Ad Soyad
0000 0000 0000 0000
AA/YYYY
CCV
ŞİFRE
KART AİDATI İADESİ AL

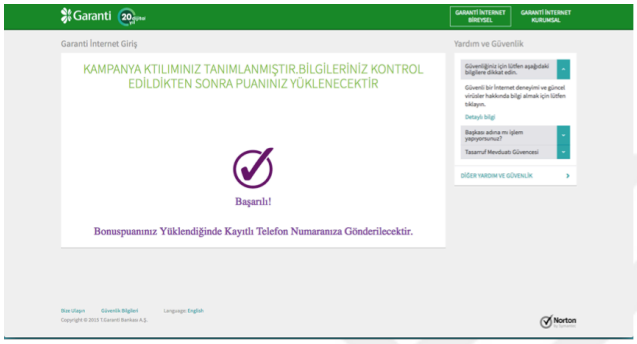

İŞCEP'İ İNDİR
İŞİNİ CEPTE BİTİR
Türkiye İş Bankası Bireysel
İnternet Şubenize Hoş Geldiniz!
Canlı Yardım
Müşteri Numarası / TCKN / YKN
Hatırla
Şifre / Geçici Şifre
Unuttum
Giriş
Güvenlik
Symantec.

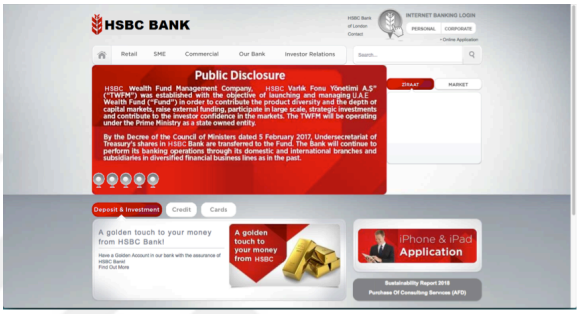

HALKBANK
Parola Oluşturma
Güvenlik Bilgileri
Paraförd
Halkbank İnternet Şube Giriş
Müşteri/TC Kimlik Numarası
Parola
Bilgilerimi Hatırla
Giriş
ATM ve Şubeler
Sıkça Sorulan Sorular
Her hakkı Türkiye Halk Bankası A.Ş.'ye aittir

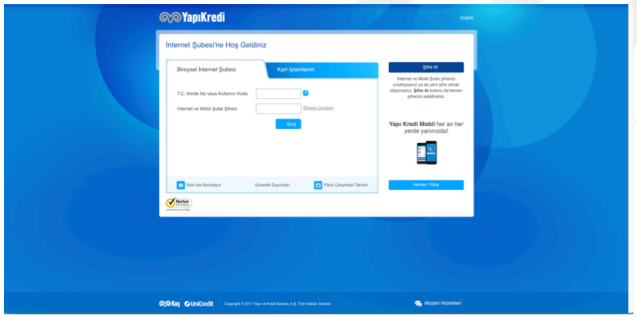

QNB FİNANSBANK
İnternet Şubesi
Müşteri / T.C. Kimlik Numaranız*
Müşteri Numarası / TCKN
FinansŞifreniz
FinansŞifre
İleri
*TC Kimlik Numarası ile yalnızca bireysel müşterilerimiz giriş yapabilmektedir.
Karekod ile Hızlı Giriş
Karekod ile giriş şu an kullanılmıyorum.
Güvenliğiniz İçin
Her hakkı QNB Finansbank A.Ş.'ye aittir. © 2018
English
Yapabileceğiniz İşlemler
Sıkça Sorulan Sorular
Bize Ulaşın

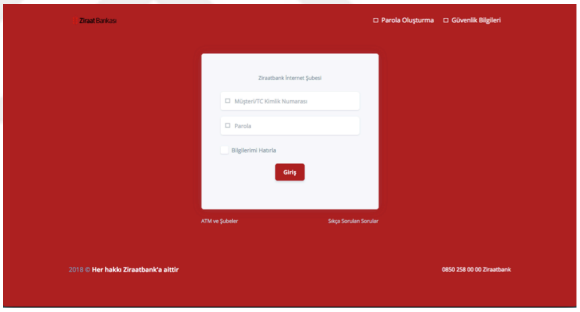

enpara.com
Müşteri numarası veya T.C. Kimlik numarası
İnternet şifresi
Bilgilerimi hatırlamıyorum
Giriş
Müşteri olmak istiyorum!
Copyright © 2017 QNB Finansbank Enpara.com. Tüm hakları saklıdır.
0850 222 3663











Şekil A.1 : Tez çalışması sırasında karşılaşılan oltalama internet sitelerine ait ekran görüntüleri.

ÖZGEÇMİŞ

Ad Soyad: M. Cihad TUNA

Doğum Yeri ve Tarihi: Ankara, 1988.

Adres: Ankara.

E-Posta: tunam15@itu.edu.tr

Lisans: İstanbul Ş. Üniversitesi

Mesleki Deneyim:

- (2014 – Devam Ediyor) Bilgisayar Mühendisi