Uma avaliação de comportamentos homográficos em ataques de *phishing* direcionados que exploram a suscetibilidade pela fidedignidade e sazonalidade

Lucas C. Teixeira[1], Carlo M. R. da Silva[1], Bruno J. T. Fernandes[1], João Fausto Lorenzato de Oliveira[1], Eduardo L. Feitosa[2], Gerson D. de C. Filho[3], Henrique F. Arcoverde[3], Vinícius C. Garcia[4]

[1]Escola Politécnica (POLI) – Universidade de Pernambuco (UPE)

{lct, cmrs, bjtf, jflo}@ecomp.poli.br

[2]Instituto de Computação (IComp) – Universidade Federal do Amazonas (UFAM)

efeitosa@icomp.ufam.edu.br

[3]Tempest Security Intelligence – https://www.tempest.com.br/

{gerson.castro, henrique.arcoverde}@tempest.com.br

[4]Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE)

vcg@cin.ufpe.br

Abstract. *The advance of phishing attacks is characterized not only in propagation, but also in rigor in detail, making fraud increasingly convincing in the eyes of the end user. Given this scenario, this study presents an approach to the homographic behaviors commonly present in phishing attacks associated with a specific target brand, either in the URL or in the page content. Through the application of 16 triggering questions, one can observe the cunning application of genuine terms, maliciously applied, in approximately 79% of the pages. It was also identified the application of homographic terms in just over 20% of the URLs. In addition to these, a predilection for the application of terms in the URL was revealed.*

Resumo. *O avanço de ataques de phishing não se caracteriza apenas na propagação, mas também no rigor em detalhes, tornando a fraude cada vez mais convincente ao crivo do usuário final. Diante deste cenário, o presente estudo apresenta uma abordagem sobre os comportamentos homográficos comumente presentes em ataques de phishing associados a uma determinada marca-alvo, seja na URL ou no conteúdo da página. Mediante a aplicação de 16 perguntas disparadoras, pode-se observar a aplicação ardilosa de termos genuínos, aplicados de forma maliciosa, em aproximadamente 79% das páginas. Também foi identifica a aplicação de termos homográficas pouco mais de 20% das URLs. Além destas, foi revelada uma predileção pela aplicação dos termos na URL.*

1. Introdução

De acordo com a Konduto[1], metade dos golpes relacionados a cartões de crédito são aplicados através de ataques de *phishing*. Muitas dessas fraudes exploram a suscetibilidade do

[1]Terceira edição do raio-x da fraude: https://bit.ly/2mv2uvN (último acesso: 01/07/2021)

usuário final em não perceber ataques homográficos, aqueles caracterizados por utilizar uma palavra (no caso uma URL ou domínio) que compartilha a mesma forma escrita de outra, mas que não remete para o endereço esperado. Casos muito comuns são domínios registrados com o emprego de termos muito similares e com variações sutis àqueles amplamente conhecidos, tais como,"go0gle" ou "amaazon", fazendo o usuário crer que a página fraudulenta trata-se de um ambiente legítimo da marca com alta reputação [Piredda et al. 2017]. Com esse uso ilícito e proposital da marca, a fraude visa explorar a suscetibilidade por fidedignidade.

A primeira vista, o ataque *phishing* aparenta ser de fácil detecção, mas os números de investidas bem sucedidas sugerem tal compreensão pouco perspicaz. Segundo relatório da *Kaspersky*[2], o número de ataques de *phishing* dobrou durante o mês de março de 2020, na quarentena causada pela pandemia do coronavirus (COVID-19). Uma das justificativas é o fato do atacante visar momentos sazonais, períodos oportunos do ano-calendário, para explorar sentimentos como a euforia e o impeto, despertados por promoções aparentemente vantajosas e de curto tempo [Hijji and Alam 2021]. Um bom exemplo são os golpes que prometiam gratuitamente, mediante dados pessoais, contas da *netflix* e álcool em gel durante o distanciamento social [3], ou o aumento de golpes que supostamente sugeriam disponibilizar o auxilio emergencial[4].

Não obstante, o estudo da KnowBe4[5] revela que 37.9% dos funcionários de diversas empresas são propensos a cair em ataques dessa natureza, evidenciando a possibilidade de ataques direcionados. Embora não existam evidências concretas que comprovem a responsabilidade dos ataques homográficos em tal feito, diante sua frequente e expressiva utilização, é sensato assumir que eles possivelmente impulsionam os crimes causados por *phishing*. Além disso, é notório o crescente número de domínios maliciosos registrados alavancados pelos ataques homográficos [Chiba et al. 2018, Liu et al. 2016, Piredda et al. 2017]. Assim, a fraude apresenta maior fidedignidade e, consequentemente, é menos suscetível ao crivo humano.

Diante o cenário, soluções anti-*phishing* que visam atenuar ataques homográficos vem sendo propostas na literatura [Chiba et al. 2018, Husain and Iqbal 2017, Le Pochat et al. 2019]. Contudo, existem alguns desafios nessa mitigação. Primeiramente, o esforço em estabelecer um **controle dos termos** que sugere fidedignidade, já que novos sinônimos e palavras-chave surgem com frequência. Com isso, o *phishing* adota um dinamismo na representação textual e, possivelmente, impulsionado pela sazonalidade. Outro desafio é que os ataques possuem uma **estratégia de termos variados**, seja pela composição de palavras-chaves genuínas ou a engenharia ardilosa de termos, como ausência, troca, duplicação de letras, entre outras. A consequência é o surgimento de variações que resultam em falsos positivos ou negativos, aumentando o esforço em estabelecer um padrão léxico.

O presente artigo apresenta um estudo empírico para a detecção de marcas em *phishing* impulsionados por termos homográficos, visando atenuar a problemática apresentada. Para tanto, uma metodologia de detecção foi elaborada e teve seus impactos

[2]https://bit.ly/3dTjQsR
[3]https://bit.ly/33ET1D7
[4]https://glo.bo/3jnvqkb
[5]https://bit.ly/37ef31h

e desafios na aplicabilidade discutida. Como diferencial, a abordagem proposta, por considerar aspectos da fidedignidade e sazonalidade, resulta em maior sensibilidade no contexto de atuação do *phishing*, o que sugere maior precisão na detecção. O estudo tem como propósito detectar a marca-alvo, sem necessariamente que a fraude faça uso explícito e literal da mesma. Concomitantemente, detectar a marca-alvo facilita identificar as intenções e reputação da página em questão.

Um protótipo do processo de avaliação foi desenvolvido e testado com 30 marcas previamente definidas. Foram analisadas 57.356 amostras reais de *phishing* reportados no ano de 2020. Diante disso, o estudo traz evidências sobre a utilização e impactos que impulsionam os expressivos números de ataque por *phishing*. Adicionalmente, também são apresentados limitações e ameaças do estudo diante os desafios na aplicabilidade e investigação do processo.

2. Conceitos Básicos

Esta seção trata da fundamentação teórica, apresentando os conceitos básicos para seu entendimento. Primeiramente, é feita a apresentação dos conceitos de ataque homográfico, bem como seu funcionamento e estratégias de atuação. Depois são apresentadas as abordagens existentes para sua mitigação. Por fim, são apresentados os trabalhos relacionados que consideram as estratégias de atuação apresentadas.

2.1. Ataque homográfico

O ataque homográfico é uma técnica de exploração da engenharia social que visa manipular os caracteres exibidos ao usuário final [Spaulding et al. 2017b]. Comumente, o fraudador manipula termos de forma maliciosa em partes da URL como o domínio, subdomínio ou *path* da URL. Apesar de mais atuante na URL, é interessante considerar seções oportunas do conteúdo da página, como o <*body*>, <*title*> e <*meta-description*>.

O objetivo é convencer o usuário final que a página em questão é um ambiente convincente e confiável e com isso fazer com que as vitimas forneçam dados sensíveis como senhas do cartão de crédito ou credencias de acesso, o que caracteriza o ataque como *phishing* [Tahir et al. 2018]. A ideia de tornar a página falsa mais confiável aumenta a medida que se assemelha com a genuína, ou seja, a fidedignidade sugere que o ataque irá investir em detalhes visuais, como domínio registrado e cadeado HTTPS [Husain and Iqbal 2017], visando ganhar a confiança de suas vítimas.

Deste modo, é possível afirmar que a fidedignidade é embasada pela marca-alvo, ou seja, a fraude estabelece uma marca-alvo e a alta riqueza em detalhes caracteriza o *phishing* como direcionado. Esse nível de sofisticação é visto no ataque conhecido como *Spear Phishing*, onde fraudador estuda a fundo o alvo no intuito de enganar os principais executivos de uma organização (por isso o nome *Spear Phishing*. Contudo, a riqueza pode ser acrescida a outras técnicas como, por exemplo, *SMiShing* que explora o público de dispositivos móveis [Mishra and Soni 2019].

2.2. Técnicas de exploração por ataque homográfico

Diante do cenário descrito, os fraudadores investem em técnicas cada vez mais refinadas para realizar os ataques homográficos. Basicamente, os atacantes se baseiam em duas estratégias. A primeiro é explorar as palavras-chave que fazem menção a uma marca através

de termos genuinamente precedentes e de grafia correta, rotulado como **termos em plain-text** neste trabalho. A segunda é elaborar termos através de engenharias ardilosas, visando propor um entendimento semântico através de termos que fogem do convencional, o que este estudo rotula como **termos homográficos**.

Na estratégia **termos em plain-text**, é comum a prática de *cybersquatting*, visando registrar um domínio que expresse a identidade textual de uma marca. Embora não seja considerada crime, tal prática permite que fraudadores a empregue como oportunidade para realizar seus crimes. Por exemplo, um endereço como *http://fgts.caixa.acesso-seguro.com* sugere um serviço ligado ao FGTS e sob responsabilidade da Caixa Econômica Federal.

Na estratégia de **termos homográficos** ocorrem as práticas de *typosquatting*, e a aplicação deturpada de *punycode*. *Typosquantting* é a prática de registrar deliberadamente domínios com erros tipográficos, em um esforço para redirecionar o usuário para uma página fraudulenta através da fidedignidade conferida por domínio levemente deturpado, a exemplo de *"netfliix.com"* e *"facbook.com"* [Spaulding et al. 2017a]. Já *punycode*, conforme o RFC 3492 [Costello 2003], é um protocolo que faz conversão de caracteres com *unicode* específicos, a exemplo do chinês ou russo, em uma versão compatível para nomes de domínios registrados. A presença do prefixo *"xn–"* no endereço é a características de *punycode*. Uma vez que é o navegador *Web* o responsável por "converter" os caracteres, se um fraudador registrar um domínio como *http://xn–netflx-7va.com*, o navegador o converterá para *http://Netflíx.com* (com o i acentuado). Além disso, a prática permite a combinação de *unicode* de idioma distintos, tornando possível registrar um endereço como *gooGle.com*, sendo que o carácter G sublinhado seria do idioma cirílico e os demais seriam do *unicode* latino. Em suma, o protocolo dá margens a investidas maliciosas e convincentes.

2.3. Mitigando ataques homográficos

Diversas práticas vêm sendo propostas na literatura para atenuar ataques homográficos. Um exemplo é o registro de possíveis domínios fraudulentos pelas próprias organizações, como no caso do domínio *http://magalu.com.br*, já registrado pela empresa "Magazine Luíza". O mesmo foi feito pelas empresas *Facebook* e *Netflix*, em providenciar a apropriação de domínios como "facbook.com" e "fcebook.com", bem como "netfiix.com", "netflix.com" e "netfliix.com".

Todavia, é notório observar que fazer uso apenas dessa prática é uma decisão temerária, visto que, dependendo da grafia da marca, as possibilidades podem ser inúmeras, o que pode inviabilizar a mitigação, seja de forma probabilística ou econômica. Diante disso, novas propostas, geralmente baseadas em Inteligência Artificial (IA) através do Processamento de Linguagem Natural (PLN), vêm sendo debatidas. A exemplo do mecanismo *TypoEval*, uma abordagem baseada em Redes Neurais Siamesas, capaz de identificar ataques homográficos em domínios com rapidez e eficácia [Ya et al. 2018].

2.4. Trabalhos Relacionados

No estudo de Liu et al. [Liu et al. 2016] é proposta uma abordagem para o combate de *typosquatting* baseada na distância *Levenshtein*, que mede a semelhança entre domínios maliciosos e legítimos. Basicamente, essa distância contabiliza o número de mudanças

necessárias que uma cadeia de caracteres precisaria para ser exatamente idêntica a outra. O estudo desempenha essa análise quantitativa baseada em colisão de *hash*. Semelhantemente, no estudo de Moubayed et al. [Moubayed et al. 2018], é proposto um modelo de classificação que analisa a reputação do domínio registrado baseando-se em aspectos textuais como tamanho e tipos de caracteres. Além disso, o estudo também visa apresentar os comportamentos comumente adotados em ataques de *typosquatting*.

Já no estudo proposto por Spaulding et al [Spaulding et al. 2017b], as contribuições se enviesam em apresentar os comportamentos dinâmicos e imprevisíveis da exploração por ataques homográficos, evidenciando que as engenharias no ataque vem se aperfeiçoando. Concomitantemente, em Tahir et al. [Tahir et al. 2018] o estudo vai mais além e revela um experimento controlado que analisa a suscetibilidade do usuário final considerando o layout do teclado, sendo proposto um modelo de probabilidade baseado em erros propositais com base no layout *QWERTY*. Na mesma linha, o estudo de Le Pochat et al. [Le Pochat et al. 2019] realiza um levantamento que também se baseia no layout do teclado, mas é direcionado para o público alemão. Os autores evidenciam que os atacantes se aproveitam de particularidades do idioma em questão, como dar ênfase ao uso de tremas na realização de seus golpes.

Já este artigo propõe um estudo empírico sobre práticas para atenuar as investidas de *phishing* que se sustentam por ataques homográficos. Apesar da proposta apresentar resultados, em seu estado inicial, o principal objetivo é analisar as abordagens de mitigação e seus respectivos desafios na aplicabilidade do cenário proposto.

Uma característica muito comum entre os estudos apresentados é seu enfoque em analisar a cadeia de caracteres resultante, além de julgar a reputação do mesmo com base em um quantitativo mínimo de diferenças entre o domínio malicioso e legítimo. Apesar dessa abordagem ser eficiente em casos com poucas diferenças, em cenários com maior abrangência na variações de sinônimos ou palavras-chave específicas, as propostas encontram maiores desafios. Essa limitação justifica a decisão do presente estudo em assumir um conhecimento prévio sobre as marcas, visando detectar as semelhanças e baseando-se na ideia principal do ataque homográfico, ou seja, considerar que a representação de uma marca será apoiada por riqueza em detalhes.

3. Metodologia

Essa seção descreve a metodologia proposta neste estudo, que é composta por três (3) atividades: (1) definição das marcas alvo; (2) obtenção das amostras de *phishing*; (3) Avaliação, com base em ataques *plain-text* e homográficos.

A primeira etapa da metodologia consiste na **definição das marcas** alvo, aquelas que serão verificadas sobre a presença em página de *phishing*. Por se tratar apenas do nome, a definição recebe como entrada as marcas que se deseja analisar, podendo ser um simples arquivo texto. Quais marcas serão analisadas é critério de quem utiliza a metodologia. Por exemplo, pode-se querer averiguar se todas as marcas, de um conglomerado de empresas, estão sendo associadas a *phishing*, ou se as marcas de empresas da região estão sofrendo com essas fraudes.

Neste estudo, a definição das marcas foi extraída de uma amostra de páginas

phishing. Para tanto, adotou-se o *PhishTank*[6] como repositório fonte. Além de ser o repositório mais empregado em trabalhos acadêmicos, ele disponibiliza seus registros através de um arquivo JSON (com, em média, 14.000 registros confirmados) ou através de busca no site ("*phish search*"[7]). Para este estudo, as 30 marcas mais recorrentes e atuantes no Brasil, contidas na amostra, foram escolhidas para avaliação da metodologia. Detalhes sobre o processo de definição das marcas são apresentados na seção 4.1.

A segunda etapa da metodologia é a **obtenção de amostras** de *phishing*. Neste estudo, optou-se por usar a mesma a amostra empregada para definição das marcas (obtida do *PhishTank*), mas a metodologia é flexível para permitir o uso de qualquer relação de URLs relacionadas a *phishing* obtidas, por exemplo, em outros sites ou plataformas. Dentre os vários exemplos, pode-se citar o *OpenPhish* [8] e *PhishStat*[9].

De posse das marcas e da amostra, inicia-se a etapa de **avaliação** da metodologia. Para tanto, ao analisar alguns trabalhos da literatura e ao realizar várias experimentações com páginas de *phishing*, visando entender melhor os ataques, foram formuladas 16 perguntas (Figura 1), com objetivo de identificar a existência do *phishing* relacionado a marca e o(s) tipo(s) de ataque(s) que estão sendo realizados. Vale destacar que essas perguntas também surgiram de um estudo empírico anterior, que elaborou uma heurística para predição de *phishing* [da Silva et al. 2020].

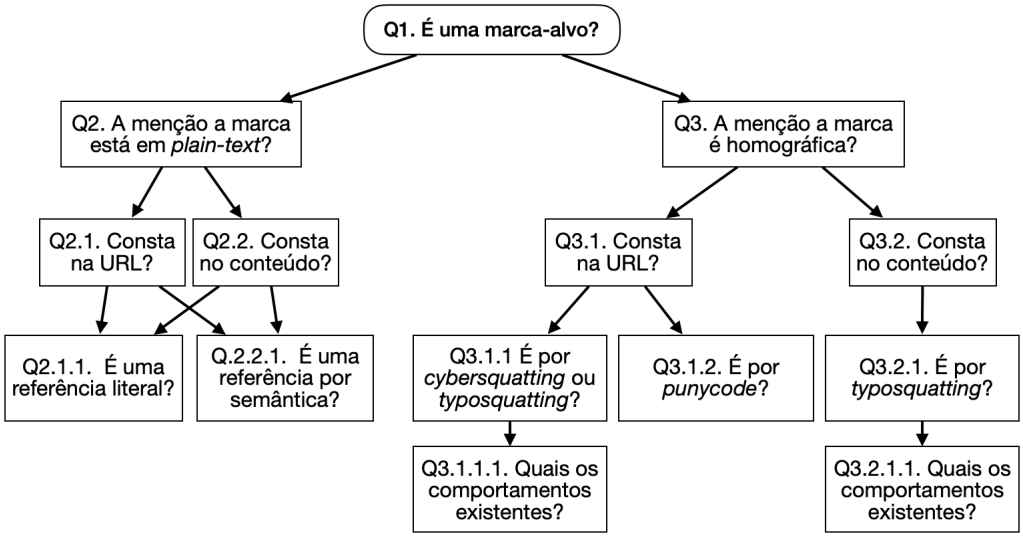

Figura 1. Questões para identificação de ataques *phishing*

Os questionamentos (16 perguntas) iniciam-se pela determinação da marca como alvo ou não (**Q1**), momento em que são descartadas as páginas com menções às marcas não participantes do estudo (definidas na primeira etapa da metodologia). Uma vez identificadas as páginas tidas como alvos, as mesmas são fracionadas com base em seu mecanismo de ataque, *plain-text* (**Q2**) ou homográfico (**Q3**), e posteriormente divididas através de sua localização, URL (**Q2.1 e Q3.1**) ou corpo da página (**Q2.2 e Q3.2**).

[6]http://phishtank.org/
[7]http://phishtank.org/phish_archive.php
[8]OpenPhish: https://openphish.com/
[9]PhishStat: https://phishstats.info/

Para as páginas contidas nos grupos com referências em *plain-text*, a etapa seguinte resume-se à análise da menção, observando se esta se dá de forma literal (**Q2.1.1**) ou através de referências semânticas (**Q2.2.1**), como a utilização sinônimos ou palavras chave. Já quanto as páginas com referência homográfica, a análise visa identificar e quantificar, tanto na URL quanto no corpo da página, a presença de *cybersquatting* ou *typosquatting* no endereço (**Q3.1.1**) ou conteúdo das páginas (**Q3.2.1**), caracterizando-as com base em clausulas pré-definidas (**Q3.1.1.1 e Q3.2.1.1**). Ainda nas páginas com referência homográfica, verifica-se a existência de *punycodes* nas URLs (**Q.3.1.2**).

Vale ressaltar que os questionamento consideraram sempre critérios de execução do *phishing*, uma vez que podem haver casos da fraude com a marcar ocorrer tanto na URL quanto no conteúdo. Assim, foi estabelecido um critério de priorização, contabilizando apenas um dos comportamentos (estar na URL ou no conteúdo). A avaliação, na metodologia, adota o seguinte fluxo de priorização: 1) se possui *punycode* na URL; 2) se possui *typosquatting* na URL; 3) se possui *plain-text* na URL; 4) se possui *typosquatting* no conteúdo; e 5) se possui *plain-text* no conteúdo.

Além disso, devido as semelhanças, ataques de *cybersquatting* e *typosquatting*, no escopo da URL, são considerados no mesmo montante de ocorrências dos resultados apresentados na seção 5. Por sinal, um aspecto essencial da avaliação é a análise de comportamentos *typosquatting* (**Q3.1.1.1 e Q3.2.1.1**). Através da observação de padrões exibidos durante a manipulação dos caracteres comumente presentes nesse tipo de ataque, pode-se perceber que os padrões se dividem em: inserção, omissão, troca de caracteres e manipulações ardilosas no TLD. Ao todo, 13 comportamentos foram percebidos e analisados por esse estudo, a saber:

- **#01. Inserir letra da tecla vizinha no teclado**: o termo google para goopgle.
- **#02. Inserir pluralidade**: por exemplo, o termo google para googles.
- **#03. Inserir letra de forma repetida**: o termo google para gooogle.
- **#04. Inserir separadores**: por exemplo, o termo "googledrive" para "google-drive" ou "google_drive" ou "google.drive".
- **#05. Omitir letra**: por exemplo, o termo google.com para goole ou googl.com.
- **#06. Omitir SLD**: define-se por omitir o termo referente ao Second Level Domain (SLD), por exemplo, o termo google.com.br para google.br.
- **#07. Trocar por tecla vizinha no teclado**: o termo google para giogle.
- **#08. Trocar por letra que mantém mesma fonética**: por exemplo, o termo google para guogle. Também serve de exemplo nubank para nubenk ou nubanck.
- **#09. Trocar posição de uma letra**: por exemplo, o termo google para ogogle.
- **#10. Trocar carácter por semelhança**: por exemplo, o termo google para goog1e ou go0gle. Também serve de exemplo itau para 1tau ou ltau (trocando a letra i por, respectivamente, um número 1 ou um L minúsculo).
- **#11. Trocar vogal**: por exemplo, o termo google para gaagle.
- **#12. Sequestro de TLD**: por exemplo, o termo google.com para google.org.
- **#13. Simulação de TLD**: por exemplo, o termo google.com para google-com.tk ou o termo google.com para googlecom.tk.

4. Implementação

Essa seção descreve o processo de implementação do protótipo utilizado para realização do estudo, através da aplicação da metodologia, conforme ilustrado na Figura 2.

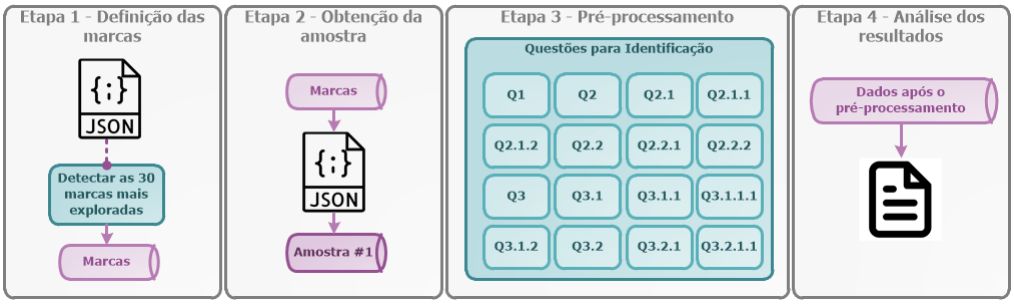

Figura 2. Fluxo da implementação com base na metodologia proposta

4.1. Definição das Marcas e Amostra

Para definição das marcas, inicialmente foi escolhida a análise via JSON, uma vez que o *PhishTank* retorna um campo "*target*" com a marca alvo. Nessa opção, os registros são mais ricos em informação, constando data de confirmação, marca-alvo, entre outros. Contudo, ao avaliar os nomes de marcas retornados, observou-se que diversos registros possuem o valor "*Other*" (algo como não identificado) e vários com nomes que não condiziam com a marca alvo em questão. Assim, extraiu-se as marcas, via "*phish search*". Nessa opção, apesar da abrangência quantitativa, há pouca informação, havendo apenas a URL e data de submissão.

Como mencionado, este estudo avalia as 30 marcas mais recorrentes na amostra do *PhishTank*. Para chegar nessas marcas, foi realizada uma análise textual. Um código foi desenvolvido em *Python*, utilizando a biblioteca *Enchant* (versão 3.2.1)[10], que permite analisar a ortografia das palavras, sugerir correções em eventuais erros ortográficos e suporte a língua portuguesa.

Primeiramente, do total de registros de 2020, foram coletadas 218.523 ocorrências de marcas em páginas relacionadas com *phishing*. Após a análise, com a execução de buscas por entradas com possíveis marcas-alvo, detectadas por PLN, foram encontradas 57.356 ocorrências. A amostra utilizada no estudo está disponível em https://bit.ly/3qNdZvi. Dessas 57.356 ocorrências, foram escolhidas as 30 marcas mais exploradas, conforme apresentado na Tabela 1, que inclui a quantidade de aparições da marca e seu percentual em relação a todas as marcas contidas na amostra. Vale destacar que a escolha de apenas 30 marcas se justifica pela necessidade de demonstrar os nuances sobre os comportamentos homográficos e não avaliar a precisão na detecção de marcas envolvidas.

Sobre a amostra, segunda etapa da metologia, ela contém 57.356 URLs de *phishing* que possuem marcas, conforme explicado anteriormente.

4.2. Avaliação

O protótipo da ferramenta de avaliação foi desenvolvido em *Python* (versão 3.8) com biblioteca *Enchant* (versão 3.2.1), que permitiu expandir o dicionário de palavras, adicionando termos referente as marcas. Os termos são categorizados de duas formas: *plain-text* ou *typosquatting*. No caso de *plain-text*, o texto é representado de forma pura e genuína, sem erros propositais na grafia, a exemplo do termo "magazine luiza" ou "magalu". Um

[10]https://pypi.org/project/pyenchant/

Tabela 1. As 30 marcas selecionadas da amostra

#	Marca	Phishes	%	#	Marca	Phishes	%	#	Marca	Phishes	%
1	amazon	4740	8.26%	11	rakuten	1493	2.60%	21	americanas	1063	1.85%
2	runescape	4736	8.26%	12	bank of america	1446	2.52%	22	caixa econômica	820	1.43%
3	google	4668	8.14%	13	wellsfargo	1381	2.41%	23	steam	697	1.22%
4	facebook	4463	7.78%	14	ebay	1350	2.35%	24	dropbox	652	1.14%
5	paypal	4462	7.78%	15	netflix	1308	2.28%	25	bradesco	629	1.10%
6	microsoft	4335	7.56%	16	tsb	1291	2.25%	26	banco do brasil	576	1.00%
7	halifax	3393	5.92%	17	yahoo	1199	2.09%	27	santander	539	0.94%
8	apple	3234	5.64%	18	magazine luiza	1186	2.07%	28	alibaba	455	0.79%
9	itau	2646	4.61%	19	adobe	1093	1.91%	29	hsbc	397	0.69%
10	lloyds	1745	3.04%	20	dhl	1074	1.87%	30	aol	285	0.50%

plain-text ainda pode ser classificado de duas maneiras: através de referência literal da marca, a exemplo de “magazine luiza”, ou de forma semântica, a exemplo de “magalu”, que faz referência a marca através de palavras-chave. Já nos casos de *typosquatting*, a prática leva o texto a ser representado com erros propositais, a exemplo de “goooogle” ou “go0gle”. Contudo, o estudo se preocupa em investigar quantos e quais variações de *typosquatting* ocorrem nos registros, conforme os comportamentos mencionados na seção 3. Além disso, este estudo visou segmentar os resultados considerando o escopo da exploração, isto é, se foi na URL ou no conteúdo da página.

Em relação a análise dos termos *plain-text*, como forma sucinta e objetiva para exemplificar o processo de enriquecimento do vocábulo pela *Enchant*, o processo adiciona uma lista de 30 termos literais referentes a marca. Além disso, também foi adicionada uma lista de termos semânticos (sinônimos ou palavras-chave) que sugerem referência as respectivas marcas. Por exemplo, a Caixa Econômica Federal possui termos semânticos inerentes à fidedignidade e sazonalidade, como: “auxilio emergencial“, “fgts”. Já o banco Bradesco possui termos da fidedignidade como “net empresas” e o Banco do Brasil possui “bbcode”, “gerenciador financeiro”, entre outros, e assim por diante para outras marcas avaliadas no estudo.

Já em relação aos termos *typosquatting*, o próprio *Enchant*, por ser responsável em sugerir correções ortográficas não propositais, tem uma certa eficiência já incorporada, especialmente em casos de troca, duplicação ou substituições de letras (teclado *QWERTY*). Contudo, foi preciso inserir alguns sinônimos de termos com erros propositais e que supostamente fossem explorados por fraudadores, cobrindo o escopo de todos os 13 comportamentos mencionados. Nesse processo, foi preciso se colocar na visão do atacante e praticar a engenharia de composições potencialmente passíveis para exploração. Outro desafio foi estabelecer as formas variadas que os termos podem ser representados, como a atribuição de separadores e concatenação com outros termos.

5. Resultados

Essa seção descreve os resultados obtidos através da aplicação da metodologia, incluindo a análise de comportamentos.

Inicialmente, a Figura 3 ilustra o resultado sintético de todo o processo.

Comportamento															Qtd	%
plain-text	URL	plain-text pela URL	plain-text literais pela URL												29867	52.07%
			plain-text sinônimos pela URL												10994	19.17%
			total plain-text pela URL												40861	71.24%
	Conteúdo da página	plain-text pelo conteúdo	plain-text literais pelo conteúdo												3722	6.49%
			plain-text sinônimos pelo conteúdo												928	1.62%
			total plain-text pelo conteúdo												4649	8.11%
			total plain-text												45510	79.35%
Homográficos	URL	cybersquatting/typosquatting pela URL	#01	#02	#03	#04	#05	#06	#07	#08	#09	#10	#11	#12	#13	-	-
			3648	2286	6789	5678	5445	5576	3567	6621	7543	4221	2367	1897	6098	-	-
			total typosquatting pela URL													9340	16.28%
			total punycode													77	0.13%
	Conteúdo da página	typosquatting pelo conteúdo	#01	#02	#03	#04	#05	#06	#07	#08	#09	#10	#11	#12	#13	-	-
			779	12	1032	345	798	-	582	893	957	1067	154	-	-	-	-
			total typosquatting pelo conteúdo													2429	4.23%
	Total por comportamento		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	-	-
			4427	2298	7821	6023	6243	5576	4149	7514	8500	5288	2521	1897	6098	-	-
			total homográficos													11846	20.65%
Total Geral																57356	100.00%

Figura 3. Resultado geral de todas as extrações realizadas

Nota-se na figura que as menções as marcas avaliadas empregando *plain-text* são o comportamento predominante entre os *phishing* da amostra, representando 79,35% das referências. Dentre estas, também é possível notar que, seja na URL ou no conteúdos das páginas, as menção através de *plain-text* são predominantes, representando mais de 63% de sua aplicação dentre as citações na URL e aproximadamente 75% no conteúdo. Tal comportamento pode ser facilmente justificado pela notória fidedignidade empregada por sua aplicação.

Outro ponto chamam atenção é a grande tendência da marca ser referenciada através da URL, representando quase 80% de todos os registros da amostra (soma *plain-text* pela URL e *plain-text* pelo conteúdo). Também é importante considerar que um pouco mais 20% das ocorrências apresentaram exploração da marca por semântica através de palavras-chaves que referenciam uma marca. Além disso, destes, mais de 20% foram resultantes de ataques homográficos, evidenciando que existe uma expressiva quantidade de ataques que não fazem uso dos termos literais.

Também é possível assumir que o ataque de *punycode* está em declínio de uso pelos fraudadores, observado em apenas 0.13% das ocorrências. Uma justificativa talvez seja pela política dos navegadores em desprezar a conversão em situações suspeitas[11].

Já a Figura 4 apresenta um gráfico os tipos da exploração homográfica mais empregados. Nela é possível assumir que, em relação a ataques de *typosquatting*, os comportamentos #09 (troca de posição), #03 (repetição de letra), #08 (troca por mesma fonética) e #13 (simulação TLD) são os mais explorados. Novamente é preciso lembra que a abordagem deste estudo se sustenta em observar características da fidedignidade e sazonalidade. No caso da fidedignidade, teoricamente, a medida que o ataque de *phishing* possui menor riqueza visual - como, por exemplo, ausência de um domínio registrado, imagens com referências quebradas ou mesmo endereços muito longos, maiores são as chances dos mecanismos *anti-phishing* convencionais detectarem a ameaça. Em contrapartida, a medida que apresenta maiores detalhes, torna-se mais difícil a detecção. No contexto da abordagem do presente estudo, por ser sustentada pela fidedignidade e sazonalidade, quanto maior riqueza em detalhes, maiores são as chances de uma predição bem sucedida.

[11] IDN Display Algorithm : https://wiki.mozilla.org/IDN_Display_Algorithm

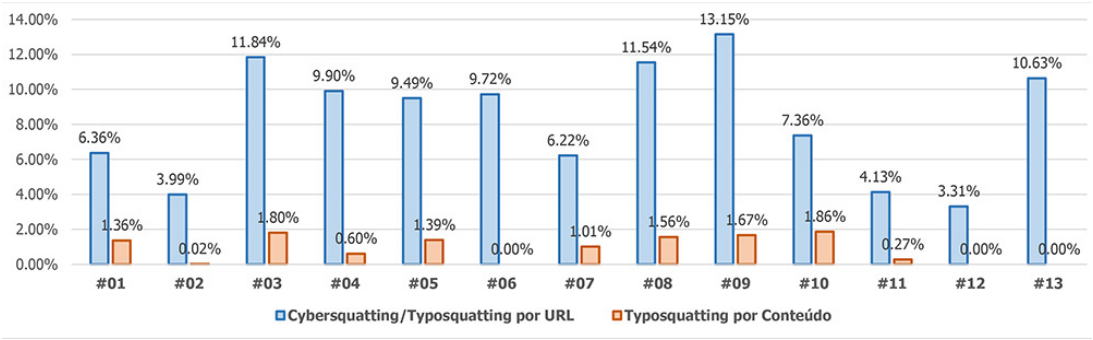

Figura 4. Resultado geral por comportamentos typosquatting

Para ilustrar a distribuição de ataques homográficos por marcas, foi desenvolvida uma tabela, ilustrada na Figura 5. Nela é possível notar e assumir que a marca mais explorada por ataques homográficos é a *Paypal*, seguida da *Amazon* e *Facebook*. Outros grandes marcas, como *Google*, *Apple* e *Microsoft* também se fazem presentes. Um fato curioso é a forte atuação no cenário nacional de marcas como *Netflix*, *Magazine Luíza* e, praticamente, todos os bancos atuantes no Brasil, revelando um comportamento sazonal resultante da pandemia causada pelo Coronavírus. Além disso, considerando que a amostra é oriunda de um repositório global, ou seja, com denúncias de todas as partes do mundo, a elevada presença de marcas atuantes no Brasil evidenciam o cenário brasileiro como muito explorado pela prática de ataques de *phishing*.

Marcas	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	total
	%	%	%	%	%	%	%	%	%	%	%	%	%	%
paypal	4.50%	3.58%	6.59%	16.53%	9.87%	12.24%	13.91%	9.51%	13.58%	11.13%	15.80%	14.94%	14.70%	9.10%
amazon	3.26%	6.21%	6.17%	10.80%	5.96%	9.83%	16.04%	7.49%	8.03%	8.46%	16.12%	18.75%	29.40%	8.37%
facebook	3.70%	4.87%	8.35%	3.04%	9.50%	10.68%	9.43%	9.53%	10.70%	8.41%	18.52%	14.94%	16.21%	7.70%
google	3.12%	3.77%	6.99%	7.12%	8.44%	8.36%	11.68%	8.72%	8.37%	9.02%	10.83%	10.43%	12.33%	6.82%
netflix	3.77%	4.87%	6.21%	3.32%	2.73%	11.21%	3.16%	4.86%	10.21%	6.74%	26.79%	14.45%	7.51%	5.80%
apple	5.67%	4.54%	6.00%	8.90%	8.94%	8.71%	11.96%	4.97%	6.19%	4.00%	5.34%	7.26%	5.62%	5.56%
microsoft	2.51%	3.92%	7.41%	6.85%	8.22%	9.04%	9.61%	7.05%	4.59%	2.44%	4.87%	10.64%	10.35%	5.49%
caixaeconomicafederal	4.70%	4.68%	5.00%	8.08%	4.23%	6.61%	6.95%	4.02%	5.69%	4.93%	4.97%	4.16%	12.67%	4.88%
bradesco	4.82%	4.44%	4.87%	5.96%	4.29%	6.13%	5.67%	3.35%	4.31%	4.55%	2.72%	7.19%	6.73%	4.04%
bancodobrasil	4.67%	3.77%	4.48%	4.69%	5.97%	4.49%	5.57%	3.25%	6.12%	4.18%	1.99%	6.34%	3.90%	3.84%
itau	3.72%	4.20%	5.55%	4.19%	2.28%	4.12%	2.00%	3.72%	3.42%	4.88%	2.20%	2.68%	2.76%	3.06%
santander	3.36%	5.40%	6.03%	2.88%	1.89%	3.59%	2.22%	4.47%	3.69%	5.36%	1.73%	1.27%	2.32%	3.03%
lloyds	3.60%	4.73%	4.14%	2.92%	4.06%	2.50%	4.17%	3.38%	4.76%	4.77%	1.15%	2.33%	1.30%	2.94%
adobe	3.51%	4.63%	5.80%	2.83%	2.06%	3.55%	1.72%	3.10%	3.12%	4.70%	1.73%	1.55%	1.37%	2.70%
wellsfargo	4.43%	4.20%	4.11%	3.83%	3.51%	2.61%	2.98%	2.88%	2.90%	4.34%	0.26%	1.55%	0.90%	2.61%
dropbox	3.36%	3.92%	2.67%	2.77%	2.93%	2.06%	1.72%	4.17%	2.53%	2.10%	1.47%	3.88%	4.44%	2.42%
magazineluiza	4.16%	3.25%	0.88%	2.58%	2.45%	1.09%	1.41%	3.13%	1.72%	7.30%	5.13%	5.29%	4.18%	2.40%
yahoo	3.33%	3.68%	2.91%	2.71%	2.84%	2.36%	2.19%	2.69%	2.69%	2.26%	0.78%	1.06%	1.82%	2.11%
bankofamerica	3.58%	3.11%	3.58%	2.49%	3.02%	1.69%	1.50%	2.46%	2.46%	1.18%	0.89%	0.49%	1.37%	1.95%
rakuten	3.14%	3.68%	4.11%	2.94%	3.17%	0.99%	1.66%	2.66%	1.62%	1.79%	0.58%	0.14%	0.00%	1.87%
steam	3.70%	3.01%	1.05%	2.73%	2.45%	1.34%	1.72%	2.76%	1.90%	1.76%	1.83%	0.14%	2.41%	1.74%
halifax	2.78%	1.77%	2.67%	3.02%	2.82%	0.83%	0.91%	3.08%	1.85%	1.54%	2.35%	0.35%	0.02%	1.67%
americanas	3.09%	3.44%	1.37%	2.83%	2.10%	1.29%	1.60%	2.52%	2.33%	0.86%	1.83%	0.85%	1.13%	1.62%
runescape	4.50%	2.48%	0.79%	2.85%	2.58%	1.53%	1.28%	3.01%	1.40%	2.06%	0.37%	2.68%	0.00%	1.62%
alibaba	2.97%	2.29%	1.60%	2.60%	2.62%	0.48%	1.69%	2.30%	2.30%	2.08%	1.67%	0.35%	0.52%	1.58%
tsb	4.67%	3.87%	0.63%	2.66%	2.30%	1.20%	1.79%	2.24%	1.03%	1.99%	0.00%	0.00%	0.00%	1.44%
ebay	3.92%	2.63%	0.79%	1.95%	1.71%	1.34%	1.32%	1.71%	1.59%	2.51%	0.00%	0.00%	0.00%	1.30%
dhl	3.29%	2.01%	1.12%	1.86%	1.47%	1.05%	1.38%	2.04%	1.55%	1.85%	0.00%	0.00%	0.00%	1.22%
aol	1.92%	2.82%	0.88%	1.40%	1.43%	1.16%	2.69%	1.92%	0.94%	2.37%	0.00%	0.00%	0.12%	1.14%

Figura 5. Resultado geral dos comportamentos typosquatting e as marcas

Por ser sensível à fidedignidade, a abordagem acredita ter tido uma eficiente detecção a medida que a fraude investia em detalhes visuais, fato que vem se tornando

uma tendência nos dias atuais. No caso da sazonalidade, termos do ano-calendário podem ser bastante sugestivos para explorar o ímpeto humano, e os atacantes munem-se disso para obter maior sucesso em suas investidas. Além disso, os termos sazonais acrescentam maior variação e dinamismo na composição dos domínios maliciosos.

Importante destacar que os dados da Figura 5 detectam todas as possíveis combinações, por isso, a soma dos mesmos acabam por ultrapassar o total de ocorrências, já que uma única ocorrência pode conter diversos comportamentos distintos. Diante disso, pode-se assumir que os comportamentos #9, #8, #3, #5 e #13 são, potencialmente, candidatos a uma exploração conjunta em um mesmo incidente. Outra premissa assumida é que no caso do *PayPal*, é comum ocorrer essa exploração por comportamentos combinados.

6. Ameaças e limitação do estudo

Essa seção descreve ameças e limitações do estudo que podem oferecer variações nos resultados apresentados, agrupados por semântica para maior entendimento.

6.1. Ameaças na detecção de marcas

A necessidade de conhecimento prévio faz com que seja necessário que a base de conhecimento(marcas e palavras-chave) seja avaliada periodicamente, uma vez que mudanças podem surgir a todo momento. Por exemplo, o proprietário da marca pode modificar padrões visuais e textuais da identificação de sua marca ou estabelecer novos jargões. Considerando o estado atual do estudo, que se preocupou apenas em oferecer um cenário minimamente operacional, a seleção e caracterização (definição de palavras-chave) preliminar de 30 marcas foi satisfatório, contudo, para um cenário em produção, um esforço maior nesse sentido precisa ser considerado.

Também foi possível evidenciar o problema da heterogeneidade ou falta de detalhes no conteúdo, o que dificulta detectar o direcionamento do ataque para uma marca. A heterogeneidade remete à casos que duas ou mais marcas ficam em evidencia em um mesmo incidente. Além disso, muitos elementos embutidos ou demais referências cruzadas podem surgir e dificultar a detecção da marca, a exemplo de termos como “Google” ou “Twitter”, que são serviços bastante utilizados em conteúdos *mashups*.

6.2. Limitações na abrangência em layout de teclados

Em seu estado atual, os 13 comportamentos abordados sobre *typosquatting* no estudo remetem apenas ao teclado do padrão *QWERTY*. Apesar desse possivelmente ser o mais utilizado no mundo, ainda sim, é sensato considerar que comportamentos dessa natureza possam ocorrer com base em outros layouts, como *AZERTY*, *QWERTZ* ou *DVORAK*.

6.3. Desafios no processo de detecção por semântica

Em relação aos resultados, algumas observações sobre a análise textual precisam ser consideradas. Primeiramente, o esforço em evitar falsos positivos devido a termos muito genéricos como “Caixa”, “Previdência” ou marcas famosas com siglas muito pequenas, como “aol”, “bb”, “dhl” ou “tsb”. Diante disso, foi preciso estabelecer algumas restrições de comportamentos em determinadas marcas. Por exemplo, no caso de “EBay”, não foi considerado os casos de omissão ou substituição da letra inicial “e”. Em outro exemplo, a marca “Steam”, quando se aplicam omissão ou troca de letras, acaba-se resultando em

termos como “team” ou “stream”. No caso da “Caixa Econômica Federal”, não era interessante considerar apenas o termo “caixa”, portanto, o nome completo foi estabelecido com política de separadores. Mesmo caso para “Banco do Brasil” e sua sigla “bb”.

Contudo, foi impossível distinguir algumas situações, a exemplo: comportamentos combinados entre #2 (pluralidade) e #1 (inserção) ou #5 (omissão) ou #7 (troca de letra vizinha), isso quando a respectiva inserção, omissão ou troca resultava em termos terminados com “s”, podendo resultar em ambiguidade, como sugerir casos de pluralidade. Na mesma linha, combinações entre #3 e #9 e #10 poderiam gerar conclusões distintas, portanto, casos como esses podem enviesar os resultados.

7. Conclusão e trabalhos futuros

O presente estudo propôs uma abordagem que visa elucidar soluções para minimizar ataques de *phishing* durante a navegação do usuário na Web. Conforme descrito na seção 2, *phishing direcionados* são fraudes de escopo fechado, a exemplo do *spear phishing* e *SMiShing*. Devido sua riqueza em detalhes, esses ataques sugerem uma abordagem preditiva mais focada em elementos de uma marca-alvo. Diante disso, os resultados dessa pesquisa apresentaram os desafios sobre a **proteção da marca** em ataques homográficos. É importante destacar a elaboração de uma metodologia para realização desse tipo de estudo, algo que até onde os autores conhecem, é a primeira tentativa neste sentido.

A solução monitorou aspectos da **identidade textual** em domínios, subdomínios e uso de suas palavras-chave em um ataque de *phishing*. Além disso, a abordagem defende a ideia de ser sensível a fidedignidade e sazonalidade, ou seja, a medida que a fraude é rica em detalhes textuais, a solução demonstra-se mais eficiente em sua detecção. Na mesma linha, também propõe ser sensível a eventos sazonais, sejam esse programados ou não.

Os estudos futuros se preocupam nas melhorias sobre as abordagens da IA, mais especificamente, em atacar as lacunas apresentadas nas limitações mencionadas na seção 6. Técnicas de *Deep Learning* são cogitadas para uma análise mais minuciosa sobre os aspectos morfológicos e mudanças de conceito que possam ocorrer no cenário de atuação, termo conhecido como *concept drift* [Elwell and Polikar 2011], impulsionados pela sazonalidade. Nesse sentido, técnicas de agregação (*Ensembles*) de classificadores tornam-se opção para maior precisão e robustez nos resultados [Oza and Tumer 2008]. Além da técnica de agregação, é pretendido aplicar práticas da ciência de dados, como as análises prescritivas e de diagnóstico, no intuito de propor modelos prescritivos e preditivos.

Em relação aos problemas de semântica em marcas com siglas muito pequenas e de maior abrangência semântica, é proposto uma prerrogativa que faça análise dos caracteres vizinhos ao termo suspeito, visando decompor os termos de uma frase e realizar um processamento semântico sensível ao contexto.

Referências

[Chiba et al. 2018] Chiba, D., Akiyama, M., Yagi, T., Hato, K., Mori, T., and Goto, S. (2018). Domainchroma: Building actionable threat intelligence from malicious domain names. *Computers & Security*, 77:138–161.

[Costello 2003] Costello, A. M. (2003). Punycode: A bootstring encoding of unicode for internationalized domain names in applications (idna). *Disponível em: https://tools.ietf.org/html/rfc3492*.

[da Silva et al. 2020] da Silva, C. M. R., Feitosa, E. L., and Garcia, V. C. (2020). Heuristic-based strategy for phishing prediction: A survey of url-based approach. *Computers & Security*.

[Elwell and Polikar 2011] Elwell, R. and Polikar, R. (2011). Incremental learning of concept drift in nonstationary environments. *IEEE Transactions on Neural Networks*.

[Hijji and Alam 2021] Hijji, M. and Alam, G. (2021). A multivocal literature review on growing social engineering based cyber-attacks/threats during the covid-19 pandemic: Challenges and prospective solutions. *IEEE Access*, 9:7152–7169.

[Husain and Iqbal 2017] Husain, M. D. and Iqbal, A. (2017). An empirical study on typosquatting abuse in bangladesh. In *Proceedings of 2017 International Conference on Networking, Systems and Security, NSysS 2017*, pages 47 – 54, Dhaka, Bangladesh.

[Le Pochat et al. 2019] Le Pochat, V., Van Goethem, T., and Joosen, W. (2019). A smorgasbord of typos: Exploring international keyboard layout typosquatting. In *Proceedings - 2019 IEEE Symposium on Security and Privacy Workshops, SPW 2019*, pages 187 – 192, San Francisco, CA, United states.

[Liu et al. 2016] Liu, T., Zhang, Y., Shi, J., Ya, J., Li, Q., and Guo, L. (2016). Towards quantifying visual similarity of domain names for combating typosquatting abuse. In *Proceedings - IEEE Military Communications Conference MILCOM*, volume 0, pages 770 – 775, Baltimore, MD, United states.

[Mishra and Soni 2019] Mishra, S. and Soni, D. (2019). Sms phishing and mitigation approaches. In *2019 Twelfth International Conference on Contemporary Computing (IC3)*, pages 1–5.

[Moubayed et al. 2018] Moubayed, A., Injadat, M., Shami, A., and Lutfiyya, H. (2018). Dns typo-squatting domain detection: A data analytics & machine learning based approach. In *2018 IEEE Global Communications Conference, GLOBECOM 2018 - Proceedings*, Abu Dhabi, United arab emirates.

[Oza and Tumer 2008] Oza, N. C. and Tumer, K. (2008). Classifier ensembles: Select real world applications. *Information Fusion*, 9(1):4–20.

[Piredda et al. 2017] Piredda, P., Ariu, D., Biggio, B., Corona, I., Piras, L., Giacinto, G., and Roli, F. (2017). Deepsquatting: Learning-based typosquatting detection at deeper domain levels. In *Lecture Notes in Computer Science*, volume 10640 LNAI, pages 347 – 358, Bari, Italy.

[Spaulding et al. 2017a] Spaulding, J., Nyang, D., and Mohaisen, A. (2017a). Understanding the effectiveness of typosquatting techniques. In *Proceedings of the Fifth ACM/IEEE Workshop on Hot Topics in Web Systems and Technologies*, HotWeb '17, New York, NY, USA. Association for Computing Machinery.

[Spaulding et al. 2017b] Spaulding, J., Upadhyaya, S., and Mohaisen, A. (2017b). You've been tricked! a user study of the effectiveness of typosquatting techniques. In *Proceedings - International Conference on Distributed Computing Systems*, volume 0, pages 2593 – 2596, Atlanta, GA, United states.

[Tahir et al. 2018] Tahir, R., Raza, A., Ahmad, F., Kazi, J., Zaffar, F., Kanich, C., and Caesar, M. (2018). It's all in the name: Why some urls are more vulnerable to typosquatting. In *Proceedings - IEEE INFOCOM*, volume 2018-April, pages 2618 – 2626, Honolulu, HI, United states.

[Ya et al. 2018] Ya, J., Liu, T., Li, Q., Lv, P., Shi, J., and Guo, L. (2018). Fast and accurate typosquatting domains evaluation with siamese networks. In *MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM)*, pages 58–63.