# Suscetibilidade através da forja de fidedignidade: uma abordagem sobre ataques de phishing

**Carlo M. R. da Silva[1,2], Lucas C. Teixeira[2], Júlio C. G. de Barros[2], Eduardo L. Feitosa[3], Vinícius C. Garcia[1]**

[1] CIn – Universidade Federal de Pernambuco (UFPE)

[2]Campus Garanhuns – Universidade de Pernambuco (UPE)

[3]IComp – Universidade Federal do Amazonas (UFAM)

{cmrs,vcg}@cin.ufpe.br, efeitosa@icomp.ufam.edu.br

***Abstract.*** *In the fight against phishing attacks and related incidents, numerous solutions have been proposed in order to minimize them. However, these attacks continue to grow today, making it necessary to reflect on the accuracy of these solutions. This article explores phishing that is based on a set of characteristics that aim to take advantage of the susceptibility of the end user. As a result, in addition to quantitative data, the study also performed a qualitative analysis of behavior, identifying aspects such as relevance, relationships, and similarities among the characteristics. It is expected that the results obtained will provoke reflection regarding new approaches or greater robustness in existing ones.*

***Resumo.*** *No combate aos incidentes de segurança relacionados a ataques de phishing, inúmeras são as soluções propostas no intuito de minimizar a incidência desses ataques. Contudo, esses continuam crescendo nos dias de hoje, fazendo refletir sobre a precisão dessas soluções. Este artigo enfoca a exploração de phishing baseada em conjunto de características que visam abusar da suscetibilidade do usuário final. Como resultado, além dos dados quantitativos, o estudo também realizou uma análise qualitativa dos comportamentos, conseguindo identificar aspectos como relevância, relações e similaridades entre as características. Diante disso, é esperado que os resultados obtidos tragam uma reflexão sobre as novas abordagens ou maior robustez nas existentes.*

## 1. Introdução

No que tange a detecção de *phishing* na Web, geralmente, os navegadores atuam como a primeira camada defensiva, concedendo aos desenvolvedores de navegadores a árdua responsabilidade em oferecer uma proteção capaz de identificar uma página fraudulenta no momento em que a mesma é acessada [AlEroud and Zhou 2017].

Uma estratégia comumente utilizada pelos desenvolvedores é fazer uso de **plataformas de denúncias**, cujo propósito é atuar como um serviço externo para fornecer periodicamente listas negras (*black lists)* a serem utilizadas pelos mecanismos de proteção dos navegadores. A alimentação dessas listas pode ser realizada por diversas maneiras, mas, em sua maioria, é baseada em denúncias voluntárias da comunidade [OpenDNS 2019]. Atualmente, *SafeBrowsing* [Google 2019], *PhishTank* [OpenDNS 2019] e *SmartScreen* [Windows 2019] são as principais plataformas de denúncias. O *SafeBrowsing* é mantido

pela *Google* e atua nos navegadores *Chrome*, *Firefox* e *Safari*. Já o *PhishTank* é mantida pela *OpenDNS*[1] e atuante no navegador *Opera*. Por fim, o *SmartScreen* é a solução da Microsoft para proteger o *Internet Explorer* e *Edge*.

Um problema de soluções baseadas em lista negra é o combate de **phishing recém-criados**, comumente denominados como *phishing zero-day*, já que propiciam a ocorrência de **falsos negativos** [Srinivasa et al. 2019]. No contexto, a problemática se enviesa no momento em que o *phishing* é disponibilizado na Web até o instante em que o mesmo é registrado na lista negra. Tal intervalo representa uma **janela de vulnerabilidade**. Além disso, existe um considerável esforço na **manutenabilidade** das listas negras. Primeiramente, é preciso considerar o tempo de vida curto do *phishing*, a exemplo dos que atuam sobre redes *fast-flux* [Almomani 2018]. Em decorrência, os mantenedores de lista negra acabam por armazenar um amontoado de *phishing offline* em suas listas.

Considerando esse contexto, os ataques de *phishing* acabam adotando uma estratégia com foco na **propagação** e **volatilidade**, ou seja, um grande número de páginas com um baixo tempo de atividade [Goel and Jain 2018]. Diante disso, para acompanhar o ritmo frenético e iminente da atividade do *phishing*, os atacantes precisam aperfeiçoar a exploração da **suscetibilidade** de suas vítimas, para tanto, adotam recursos que visam aumentar a fidedignidade da página falsa, a exemplo de um registro de domínio ou a presença do cadeado de segurança [Parsons et al. 2019]. Tais investidas consideram três aspectos, a saber: (i) o **conteúdo**, a exemplo de código-fonte malicioso; o **contexto**, como a exploração de recursos do navegador ou a sensibilidade com dispositivos móveis; (iii) e a **URL**, com a exploração por combinação de palavras-chave ou caracteres homógrafos.

Esse estudo apresenta um *survey* para **investigar** 7 características que exploram a suscetibilidade do usuário considerando **conteúdo, contexto e URL** do *phishing* no intuito de observar padrões de comportamento que visam forjar a fidedignidade da fraude com a página genuína. O estudo analisou as características em *phishing* reais para testar hipóteses. Foi adotado o *PhishTank* como plataforma de denuncia e base de dados para a extração de dados e definição das amostras. Por fim, também é apresentada a coleta e interpretação, além das limitações, conclusões e perspectivas futuras do estudo.

## 2. Contextualização

De acordo com o estudo de Amiri et al. [Amiri et al. 2014], *phishing* é um ataque caracterizado por tentativas fraudulentas contra usuários da Internet. Nesse cerne, o atacante desenvolve uma página falsa que apresenta-se como um ambiente confiável, induzindo suas vítimas a submeterem dados sensíveis, através, por exemplo, de formulários com credencias de acesso a um determinado serviço genuíno, explorando a suscetibilidade.

Como base na problemática descrita por alguns estudos [Gupta et al. 2016, Sonowal and Kuppusamy 2017, Vayansky and Kumar 2018], a suscetibilidade, por ter como alvo o usuário final, geralmente é explorada fazendo uso de sentimentos que resultam em insuspeição sobre a oportunidade em questão. Exemplos dessa natureza seriam euforia (vantagens, benefícios), pânico (ameaças, boatos) e confiança (fidedignidade). Todavia, por ser direcionada a aspectos humanos, existe o caráter subjetivo que impossibilita qualificar quando o alvo será suscetível a fraude. Ou seja, não há como assumir que um

[1] https://www.opendns.com/

indivíduo precisa ser condicionado a esses sentimentos para cair em golpes. Na mesma linha, não há como quantificar o nível do sentimento para a suscetibilidade ser explorada.

Em contrapartida, é possível considerar que certos elementos ofereçam maior fidedignidade, no que tange a confiança, aumentando assim as chances de exploração. Diante disso, através da engenharia social, o atacante observa aspectos visuais do conteúdo, contexto e URL da página. No caso do conteúdo, os atacantes se esforçam em copiar o *template* da página de forma fiel ao da página genuína [Goel and Jain 2018]. Já o contexto remete a aspectos que fazem parte do ecossistema, ou seja, observa que o serviço mantenedor define senhas de "n" dígitos, cadeado de segurança (HTTPS) e códigos maliciosos que interfiram no comportamento do navegador Web. Não obstante, a URL também é um elemento chave, uma vez que proporciona o uso de palavras-chave em sua composição no intuito de simular algum serviço popular, visando a fidedignidade [Parsons et al. 2019].

Com isso, o atacante extrai perfis de cada alvo envolvido, que se traduzem como um conjunto de comportamentos que serve de subsídio para a elaboração da página maliciosa [Khonji et al. 2013]. Em tese, quanto maior a qualidade do perfil, maior a fidedignidade. Além dessas, o atacante também pode realizar outras atividades, como registrar ou sequestrar um domínio, no intuito de atribuir combinações arbitrárias através de palavras-chave. Por fim, o registro de domínio encurta a URL o que, indiretamente, melhora seu *rank* no *Search Engine Optimization* (SEO) e, consequentemente, sua reputação.

Devido as lacunas mencionadas na Seção 1 sobre as soluções baseadas em lista negra, tornou-se uma tendência na literatura a presença de soluções que consideram características presentes no *phishing* com o propósito de realizar uma predição para classificar se uma página é maliciosa ou não [Amiri et al. 2014]. Todavia, o cenário de atuação do *phishing*, por ser dinâmico, acaba por ser suscetível a diversas mudanças em suas características. No contexto de predição de *phishing* com base em características, tais mudanças são conhecidas como *concept drift* [Elwell and Polikar 2011]. Na literatura existem abordagens *anti-phishing* suscetíveis a esses problemas [Amiri et al. 2014], que serão apresentas em subseções.

### 2.1. Abordagens de detecção de phishing através da URL

São abordagens que não necessitam analisar o conteúdo da página suspeita para classificá-la como legítima ou fraudulenta. Em geral, tais soluções empregam: filtros com base em listas negras ou brancas, análise de padrões léxicos na URL, ou ambas as abordagens. Diante das limitações já apresentadas sobre o uso de lista negra, as estratégias de padrões léxicos ganharam força. Exemplos de padrões dessa natureza seriam verificar o tamanho da URL, a quantidade de caracteres como separadores, a reputação ou localização geográfica do *host* da página, entre outros.

### 2.2. Abordagens de detecção de phishing através do conteúdo

São abordagens que baseiam a predição em elementos constantes na página suspeita. Elementos dessa natureza incluem campos de senha, erros ortográficos, formatação do *CSS*, além de código *HTML* ou *JavaScript*. Apesar da possibilidade de descartar o uso de listas negras, devido ao ambiente do *phishing* ser dinâmico, esse tipo de solução precisa realizar revisões constantes para se adequar aos problemas de *concept drift*.

### 2.3. Abordagens de detecção de phishing através do contexto

Como já dito anteriormente, o contexto remete a recursos disponibilizados e projetados para trabalhar em conjunto com a página em questão, constituindo assim em um ecossistema. O navegador Web com ou sem plug-ins e extensões, o usuário final em questão, a plataforma de denúncia e o dispositivo móvel utilizado pela vítima são exemplos de recursos do contexto do *phishing*. Não é incomum os casos de ataques de *phishing* que são direcionados para um determinado contexto, como uma organização, a exemplo do *Spear phishing*. Não obstante, existem casos de ataques que são direcionados para dispostivos móveis, como o *SMiShing*, que apesar de utilizar o serviço de SMS, muitas vezes a mensagem possui um link que direciona o usuário para um ambiente malicioso.

## 3. Trabalhos Relacionados

Nesta seção, serão descritos trabalhos da literatura que possuem soluções correlatas à proposta deste estudo, mais precisamente, estudos que abordam as características definidas na Tabela 1 da Seção 5. Em Khonji et al. [Khonji et al. 2013], são apresentadas técnicas para detecção de *phishing* considerando o fator humano, além de oferecer uma visão geral sobre técnica de detecção. As técnicas consideram abordagens de defesa ofensiva, correção e prevenção dos ataques. No estudo de AlEroud and Zhou [AlEroud and Zhou 2017], o foco principal é analisar ataques de *phishing* resultantes de uma revisão da literatura. O diferencial declarado pelo estudo é que o modelo proposto considera técnicas, ambientes específicos e contramedidas para atenuar tipos de *phishing*.

Sharma et al. [Sharma et al. 2017] propõem a comparação de 8 ferramentas para detecção de *phishing*, submetendo as mesmas a uma amostra que avalia a eficiência. A amostra foi extraída de bases como *APWG* e *PhishTank*. A análise também compara amostras entre páginas genuínas e *phishing* legítimos. Já o estudo de Goel and Jain [Goel and Jain 2018] avalia alguns mecanismos *anti-phishing* para dispositivos móveis, dividindo em 4 etapas, a saber: (i) detalhar sobre o contexto do ataque para dispositivos móveis; (ii) analisar os tipos de ataques envolvidos; (iii) contramedidas dos ataques apresentados; por fim (iv), são debatidos os desafios sobre o combate de *phishing*.

Em Chiew et al. [Leng Chiew et al. 2018], apresenta uma discussão sobre as abordagens de execução de ataques de *phishing*. O intuito é apresentar uma revisão que melhore a compreensão das características utilizadas. Por fim, o estudo de Qabajeh et al. [Qabajeh et al. 2018] analisa estratégias anti-phishing com base em características que consideram aspectos legais, treinamento, conscientização e abordagens inteligentes. Além disso, também evidencia aspectos positivos e negativos no desempenho.

## 4. Metodologia

Esse estudo tem o intuito de investigar padrões de comportamentos existentes em *phishing* em ambientes reais. Segundo Wohlin et al.[Wohlin et al. 2000], por *survey* entende-se um estudo primário, ou seja, uma abordagem da engenharia de software experimental que observa comportamentos em um contexto específico. Tais comportamentos podem ser traduzidos como as características que o estudo visa extrair em páginas maliciosas.

Como as amostras e variáveis (dependentes e independentes) não podem ser controladas e o ambiente a ser observado trata-se de um ambiente real, o estudo adotou a

metodologia *survey*, conforme debatidos na obra de Molleri et. al [Molleri et al. 2016]. Geralmente, um *survey* apresenta-se como uma investigação em retrospecto, analisando comportamentos através de dados coletados em uma amostra representativa da população do contexto. Com os resultados dessa coleta, extrai-se conclusões que possam ser generalizadas a população da qual a amostra foi derivada. Durante as etapas do estudo algumas terminologias serão utilizadas e descritas nessa seção.

### 4.1. População e Amostras

Para o processo de amostras, o estudo adotou a plataforma *PhishTank* como população para extração de dados para a definição das amostras de *phishing* reais. Apesar da possibilidade de redução na precisão, em relação a uma população, ainda sim, fazer uso de amostras viabiliza a obtenção dos objetivos de um *survey* dessa natureza. E embora existam outros tipos de serviços, como o OpenPhish[2] e o SafeBrowsing[3], a opção pelo *PhishTank* teve como critério o fato da base ser aberta e com maior volume.

### 4.2. Definição e extração das amostras

Para a definição das amostras, foi necessário obter uma quantidade significativa de *phishing*. Como alternativa, a plataforma disponibiliza um *web service* que fornece um **arquivo JSON**[4]. O mesmo é atualizado a cada 1 hora e tem em média 15.000 registros, conquanto, em média, cerca de 90% das URL ainda persistem nos arquivos subsequentes. Além da **URL**, **status** (*online* ou *offline*), **confirmação** (se a denúnica é procedente ou não, classificando como "válido" ou "inválido") e **data de publicação** (momento da denúncia), também fornece a **data de confirmação** (momento da confirmação) e a **marca alvo**. O processo teve como objetivo extrair registros durante todo o ano de 2018.

Foi possível definir a **Amostra #1**, resultante da coleta dos JSON periódicos, com 49.089 registros válidos, confirmados e com conteúdo preservado. Essa amostra atendeu satisfatoriamente as extrações de dados com caráter quantitativo e realizadas através de um algoritmo de forma objetiva. Ainda sim, por ter um número significativo de registros, a mesma impossibilitou contemplar características de caráter qualitativo, ou seja, que consideram o conteúdo e necessitam de uma inspeção manual subjetiva. Diante disso, conforme as Equações 1 e 2, foi necessário definir a **Amostra #2**, aleatória [Singh and Mangat 1996] considerando a variabilidade máxima na população, com confiança de 95% e precisão de 5%, resultando em 382 registros.

$$\frac{z^2 \times p(1-p)}{e^2} \tag{1}$$

$$1 + \left( \frac{z^2 \times p(1-p)}{e^2 N} \right) \tag{2}$$

O cálculo possui duas etapas, na (1) o $z$ representa o grau de confiança em desvios padrões (95% ou 1.96 na escala Z), o $e$ representa a margem de erro (5% ou 0.05),

[2]https://openphish.com/

[3]https://safebrowsing.google.com

[4]http://data.phishtank.com/data/online-valid.json.bz2

o $N$ representa a população total de 49.089 URL. Por fim, o $p$ é a verdadeira probabilidade do evento, ou seja, a chance probabilística individual de uma URL ser escolhida, representada por uma constante 0.50 (50% de probabilidade).

## 5. Apresentação e Interpretação dos Resultados

Essa seção descreve os dados obtidos através dos métodos de extração, bem como a interpretação dos mesmos sobre uma análise de seus comportamentos. Conforme descrito na Tabela 1, são descritas as características com seus respectivos contextos de extração e amostras utilizadas para a obtenção dos dados.

**Tabela 1. Categorias e abordagem de extração nas amostras utilizadas**

| Características | Extração | Amostra #1 | Amostra #2 |
|---|---|---|---|
| C01. Ataque Homográfico | Contexto (navegador) | X | |
| C02. Concatenação de Subdomínios | URL (subdomínio) | X | |
| C03. Domínio com Reputação | Contexto (domínio) | | X |
| C04. Protocolo de Tunelamento | URL (protocolo) | X | |
| C05. Simulação por Código Malicioso | Conteúdo (código-fonte) | | X |
| C06. Redirecionamento da URL | URL (path e querystring) | X | |
| C07. Tipografia Maliciosa | Contexto (palavras-chave) | | X |

Além disso, o estudo busca, com base nos dados obtidos, expor um veredito sobre a **relevância** da característica, em uma escala entre *BAIXA*, *MODERADA* e *ALTA*. O critério para classificar a relevância considera aspectos quantitativos e qualitativos, por exemplo, um resultado quantitativo, sendo discreto ou expressivo, pode influenciar consideravelmente a relevância, apesar disso, comportamentos temporais ou aspectos do contexto, quando aplicáveis, podem equilibrar esse veredito.

### 5.1. C01. Ataque Homográfico

Essa característica avalia as explorações de recursos do navegador que podem ser feitas através da URL da página com o protocolo denominado *punycode*. Conforme o RFC 3492 [Costello 2003], o *punycode* é um protocolo que permite a conversão de caracteres com *unicode* específico, como o chinês ou russo, em uma versão compatível para nomes de domínios DNS. Essa cadeia de caracteres é sempre precedida do prefixo *"xn–"* e a responsabilidade da conversão é atribuída ao navegador *Web*.

Na prática, domínios com caracteres acentuados, por exemplo "Netflíx.com", resultam em "http://xn–netflx-7va.com". Todavia, as opções não se restringem ao uso do *unicode* latino. Ao utilizar o alfabeto cirílico (línguas eslavas), a palavra *apple* em *unicode* cirílico [Costello 2003] convertida em *punycode* resulta em *xn–80ak6aa92e*, ou seja, uma versão ASCII válida para um domínio. Outra maneira de sofisticar o golpe é a possível combinação de *unicode* distintos para resultar em uma cadeia de caracteres homógrafos, ou seja, idêntica ao site genuíno. O fato é que alguns idiomas, a exemplo do cirílico, não possuem caracteres como *g* ou *l* minúsculos, o que dificulta o infrator forjar um conjunto de caracteres que se assemelhem a um domínio como o *google.com*. O mais aproximado seria algo como *GooGle.com*, porém, suscetível à suspeitas.

Contudo, nada impede que o infrator use combinações de *unicode* diferentes, como por exemplo usar o *g* e *l* minúsculos do alfabeto latim básico e os demais caracteres

do alfabeto cirílico, resultando na cadeia de caracteres *xn–ggl-tdd6ba.com*, que convertida pelo *punycode*, resultaria em *google.com*[5]. O *punycode* foi proposto para tornar os registros DNS mais amigáveis, entretanto, acabou se tornando uma oportunidade de fraude. O atacante muitas vezes combina recursos para ficar bem verossímil, como registrar o domínio *xn–80ak6aa92e.com* e adotar HTTPS. Os dados estão ilustrados na Figura 1.

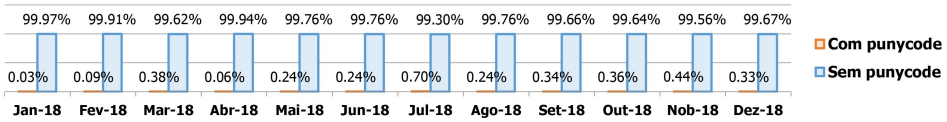


**Figura 1. Resultados de C01. Ataque Homográfico**

Atualmente os navegadores minimizaram o problema, uma vez que no ato da conversão com *punycode* o navegador possibilita que o usuário desative a troca "amigável" na barra de endereço, permanecendo a exibição dos caracteres originais. Outro comportamento defensivo foi alertar casos em que os caracteres fazem uso de *unicode* distintos. No entanto, ainda é possível fazer uso dessas técnicas em elementos HTML ou corpo de *e-mails*, a exemplo do atributo *href* em elementos $< a >$ e o uso de *unicode HTML*, exigindo assim maior acuidade por parte dos usuários-finais em observar tal exploração. Diante disso, a característica foi considerada com relevância *MODERADA*.

## 5.2. C02. Concatenação de Subdomínios

Essa característica avalia padrões em URL que apresentam diversas concatenações de subdomínios para levar a crer, através de uma observação pouco apurada do usuário final, que a URL apresentada no navegador induz um ambiente de execução com domínio legítimo, por exemplo, *facebook.edit.youraccount.com*, quando na verdade o domínio em questão trata-se de um registro com nome *youraccount.com* que foi criado pelo mal intencionado, e o subdomínio edit e facebook foi o efeito visual manipulado pelo fraudador. Os dados extraídos estão ilustrados na Figura 2.

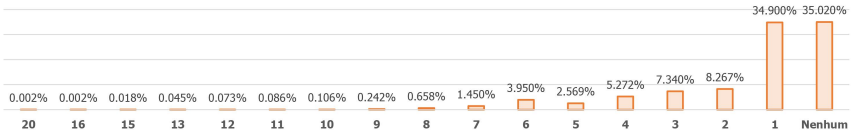


**Figura 2. Resultados de c02. Concatenação de Subdomínios**

Foi possível observar que 30.08% dos registros utilizam 2 ou mais subdomínios, evidenciando que a utilização desse recurso é bem recorrente, justificando a característica com relevância *ALTA*. Foi possível detectar casos de 20 subdomínios em uma única URL e as palavras-chave mais comuns na composição das concatenações era fazendo alusão a serviços consolidados, como Facebook ou Dropbox, em combinação com termos relacionados a segurança. O grande volume de subdomínios não é por acaso, o mesmo resulta em uma URL grande o suficiente para ser suprimida visualmente do usuário da barra de *status* do navegador, no intuito de reduzir as suspeitas.

[5]Conversor *online* de *punycode* para ASCII: http://idna-converter.com/

## 5.3. C03. Domínio com Reputação

Essa característica avalia casos em que o fraudador consegue obter controle de um registro legítimo e usa a reputação do mesmo para persuadir o usuário final, ou seja, um domínio sequestrado. Os casos de sequestros ocorrem comumente por falha de sanitização das entradas das aplicações, por exemplo, sessões de *upload* que permitem injetar página maliciosa. Além desses, casos em que a fraude possui um domínio com registro de acesso mais restrito, a exemplo de domínios ".gov" e ".org", também se enquadram na característica. Diante dos resultados, não foi incomum encontrar domínios governamentais e organizacionais sendo utilizados para o crime, sejam sequestrados ou com registro deliberadamente concedido. Os dados extraídos estão ilustrados na Figura 3.

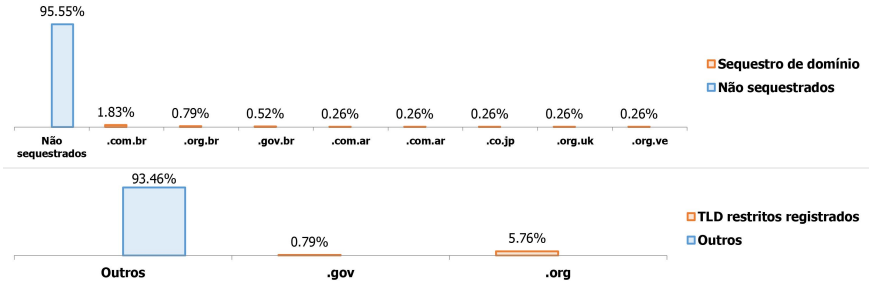


**Figura 3. Resultados de C03. Domínio com Reputação**

Conforme ilustrado na Figura 3, esse tipo de incidente ocorre com uma quantidade considerável e bem disseminado pelo mundo em geral. Nada obstante, o que chama atenção é a quantidade de domínios brasileiros (.com.br) sequestrados, que apresenta-se como o mais explorado, sendo maior que o dobro de ocorrências em comparação ao segundo colocado. Outro ponto que merece destaque são os domínios .org e .gov, que em tese, seriam mais restritos, e com isso, resultam em maior veracidade. Porém, na prática, os serviços de registro de domínio parecem não impor obstáculos para a obtenção de um domínio dessa natureza, justificando a característica com relevância *ALTA*.

## 5.4. C04. Protocolo de Tunelamento

Essa característica avalia os casos em que o fraudador não mede esforços quanto ao investimento em garantir maior fidedignidade em sua fraude, adicionando em suas fraudes recursos de tunelamento e inclusive registrando os mesmos em certificadoras digitais. Os dados extraídos estão ilustrados na Figura 4.

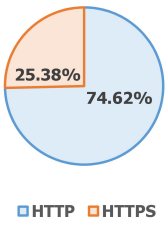


**Figura 4. Resultados de C04. Protocolo de Tunelamento**

Foi possível observar que 74.62% dos *phishing* não utilizam o protocolo HTTPS. Diante disso, inicialmente, a característica foi considerada de alta relevância, todavia, no

ano de 2017 o quantitativo de páginas com cadeado era de apenas 12.82%, ou seja, quase metade das ocorrências em 2018, com 25.38%. Isso significa que no ano anterior era incomum uma fraude com HTTPS, fazendo com que a ausência do cadeado fosse um indicativo mais preciso para julgar a página como genuína. Conquanto, os dados de 2018 demonstraram que a característica perdeu força com o passar do tempo. Uma justificativa seria a consolidação de recursos gratuitos, a exemplo do *Let's Encrypt*[6], aumentando assim a presença dos cadeados. Diante disso, o julgamento da relevância foi questionado sobre o aumento ou diminuição da fidedignidade, tornando a mesma como *MODERADA*.

## 5.5. C05. Simulação por Código Malicioso

Essa característica avalia as explorações através do código-fonte da página no intuito de trazer maior fidedignidade para o usuário final, persuadindo o mesmo a acreditar que a fraude em questão trata-se de um site genuíno. Apesar do código-fonte fazer parte do conteúdo da página, a característica é considerada estática porque descreve uma análise léxica do código-fonte da página, ou seja, o conteúdo analisado não é o mesmo a ser apresentado ao usuário final, portanto, aspectos subjetivos não são considerados.

Padrões dessa natureza são recorrentes através de cabeçalhos HTTP como *User-Agent* e *Referer*, tipicamente utilizados para ataques de *SMiShing*, ou seja, direcionados para um dispositivo móvel ou determinado contexto, a exemplo dos ataques de *Spear Phishing*. Por fim, também foi observado páginas que induziam a instalação de *add-ons* (*plug-in* ou extensão) no navegador, evidenciando a utilização de *malware*.

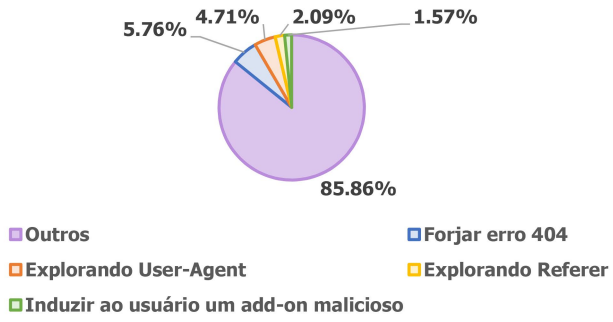


**Figura 5. Resultados de C05. Simulação por Código Malicioso**

Conforme ilustrado na Figura 5, foi possível identificar que boa parte das páginas analisadas realizavam combinação de "User-Agent" com "Forjar erro 404", resultando que uma determinada página era aberta no navegador do *smartphone*, mas ao tentar abrir no navegador de um PC era apresentado um erro forjado de 404, sendo assim um possível *SMiShing*. Diante disso, a característica foi considerada com relevância *ALTA*.

## 5.6. C06. Redirecionamento na URL

Essa característica avalia os casos em que o fraudador explora URL de sites legítimos que possibilitam um redirecionamento através da manipulação de *path* ou *querystring*. Na prática, o protocolo HTTP permite que os valores desses parâmetros sejam modificados arbitrariamente durante as requisições GET. Em muitos casos, a aplicação portadora da

[6]https://letsencrypt.org/

URL não realiza um tratamento dessas entradas, possibilitando que um mal intencionado informe uma URL maliciosa nesses parâmetros. Por ser legítima, visualmente o usuário pode acabar confiando. Ainda sim, a mesma irá redirecionar o usuário para a página que o fraudador informou nos parâmetros GET, representando assim um perigo ao usuário final. Os dados extraídos desse tipo de ataque estão ilustrados na Figura 6.

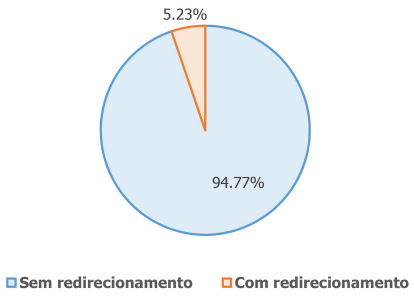


**Figura 6. Resultados de C06. Redirecionamento na URL**

Conforme ilustrado na Figura 6, existe um discreto número de *phishing* com redirecionamento através de um parâmetro da URL, seja armazenado no *path* ou *querystring*. Diante disso, a característica ficou avaliada com relevância *MODERADA*.

## 5.7. C07. Tipografia Maliciosa

Essa característica avalia os casos em que o fraudador faz uso de trocadilhos, que podem ser palavras com erros na grafia, parônimos, homógrafos, homônimos, entre outros, e muitas vezes associados a jogos de palavras para um olhar desatento do usuário final. Importante salientar que essa categoria considera o conteúdo da página e todos elementos da URL, entretanto, explorações no protocolo *punycode* foram isolados na categoria C01 por serem muito específicos e de escopo limitado ao domínio da URL e comportamento do navegador. Os dados extraídos estão na Figura 7.

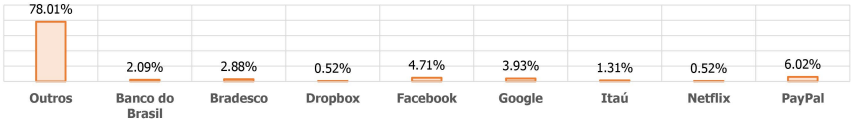


**Figura 7. Resultados de C07. Tipografia Maliciosa**

A Figura 7 descreve marcas exploradas com tipografia maliciosa, ou seja, ocorrências na URL de palavras “faceb00k”, “Netfliiix” ou “dr0pbox”, prática conhecida como *typosquatting* [Stout and McDowell 2012]. No caso do PayPal, foram detectados 37 trocadilhos, totalizando 9.69% da Amostra # 2. Não obstante, destaque para os bancos brasileiros, em especial ao Bradesco com 3.14% de todo o montante mencionado. Não obstante, outro comportamento interessante é a considerável quantidade de domínios que são registrados propositalmente similares a uma marca famosa, porém, criados por terceiros, prática conhecida como *cybersquatting* [Stout and McDowell 2012]. Diante disso, a característica foi considerada com relevância *ALTA*.

É fato que a prática nem sempre é criminosa, já que não é de hoje que costumam realizar registros de domínios com alusão a marcas famosas para posteriormente serem revendidas ao respectivo grupo representante. Contudo, não deixa de ser uma oportunidade

de exploração para os mal intencionados. Um postura interessante de algumas empresas, como Facebook e Netflix, foi a apropriação de domínios com *typosquatting* que redirecionam para a página com a grafia correta, a exemplo de "facbook.com", "fcebook.com", "Netfiix.com", "Netflix.com" ou "Netfliix.com", no intuito de proteger seus usuários.

### 5.8. Análise de Relevância

Diante dos resultados, o estudo ponderou o nível de relevância de cada característica entre *BAIXA*, *MODERADA* e *ALTA*. Essa análise não se limita apenas a aspectos quantitativos, já que, em determinadas situações, aspectos subjetivos, como o conteúdo e contexto, foram determinantes, resultando assim em uma análise objetiva e subjetiva. A exploração de suscetibilidade destaca abusos de recursos do navegador, concatenação de subdomínios, sequestro de domínio, código malicioso e tipografias maliciosas na URL, fatores que podem ser decisivos para o veredito de uma página suspeita.

### 5.9. Análise de Relações

Essa seção descreve as relações observadas entre as características. Tais aspectos podem influenciar diretamente ou indiretamente o resultado de cada característica, além de impactar em uma ou mais características distintas. Não obstante, por existirem relações cruzadas, ou seja, com objetivos distintos, também foi possível observar maior sensibilidade em relação aos aspectos de similaridades entre as características.

Certas combinações descrevem a utilização de palavras-chave de um idioma para atrair a atenção das vítimas praticando *typosquatting* no conteúdo da página e da URL (C07). Além disso, devido a pouca acuidade dos registradores de domínio, os mal intencionados registram domínios visando algumas vantagens, como a obtenção de registros .gov ou .org (C03) ou a liberdade de explorar a composição da URL com manipulações homográficas, a exemplo da exploração com *punycode* (C01) ou *cybersquatting* (C07). Sem embargo, as explorações não são limitadas ao domínio, como os casos de ataques que exploram através de *typosquatting* em subdomínios (C02).

Demais combinações visam explorar os aspectos da reputação do serviço, com uma engenharia social sobre o perfil alvo aliada a técnicas que aumentam a fidedignidade com a adição de um cadeado na página (C04). Não obstante, ao invés de registrar, os atacantes podem utilizar uma URL legítima que dispara um redirecionamento para uma determina URL que encontra-se como valor de algum parâmetro GET, se aproveitando assim da reputação do domínio da URL legítima em questão (C06).

Todavia, o atacante pode sequestrar um domínio existente (C03) através de alguma vulnerabilidade, a exemplo de injeção de arquivos em uma sessão de *upload*. O atacante utiliza o domínio e se beneficia das mesmas vantagens anteriormente mencionadas mas também se beneficia com outras, como o prestígio. Por fim, demais ameaças remetem a aspectos que exploram os recursos do navegador e que poderiam ser minimizados por seus respectivos mantenedores através de políticas mais assertivas sobre determinados códigos *Javascript* e a interação com o usuário final através de *add-ons* maliciosos (05).

## 6. Ameaças da Avaliação

Essa seção descreve as ameaças e limitações a serem consideradas pelo estudo, as mesmas foram agrupadas por objetivos e fases da metodologia.

### 6.1. Ameaças nos processos de amostragem e extração dos dados

O processo de geração de novos arquivos JSON considerou eventos dos quais a plataforma *PhishTank* **removia** ou **adicionava** registros de forma periodicamente. Não é possível responder com precisão os motivos da adição ou remoção desses registros, contudo, alguns comportamentos podem justificar. Em relação as adições, além do processo natural do surgimento de novos *phishing* na Web, o arquivo JSON pode receber *phishing* remanescentes de votações pendentes anteriormente. Por exemplo, uma URL foi submetida na plataforma, mas a mesma só será considerada no arquivo JSON quando a plataforma tiver um veredito sobre a votação da mesma, o que pode levar horas ou até mesmo dias. Ou seja, o tempo de transição da confirmação "inválido" para "válido" pode variar, ocasionando a presença de novos registros em arquivos futuros gradativamente.

Em relação as remoções, o sistema de votação visa minimizar a ocorrência de falsos positivos, contudo, ainda não isenta a possibilidade, portanto, a plataforma disponibiliza uma seção em que um usuário pode alertar sobre um julgamento inadequado. A própria declara que esse tipo de denúncia é levada muito a sério[7], e caso seja confirmado o equivoco, a URL em questão é mudada de *phishing* válido para inválido, ou seja, o registro acaba por ser **removido** na próxima geração do JSON. O motivo é justificável, já que um falso positivo de um site genuíno pode afetar a credibilidade da plataforma.

### 6.2. Ameaças no processo de interpretação dos dados

Por conter aspectos subjetivos, a escala proposta pelo estudo, *BAIXA*, *MODERADA* e *ALTA*, resultam em vereditos definidos por interpretações oriundas de observações analisadas. O estudo cobriu um contexto pertencente ao seu escopo, como a suscetibilidade por fidedignidade, Ou seja, poderia haver interpretações distintas na existência de outros.

Em relação às características analisadas, algumas outras tiveram resultados iniciais, porém, por questões de escopo, não estão contempladas no estado atual do estudo. Uma delas foi a **política de detecção de *malware* adotada pelo navegador**, observando os elementos da página e as ações do usuário, como os cliques para navegação e *download*. Não obstante, a mesma análise também foi aplicada quanto a **política de instalação de *add-ons***. Muitos serviços, principalmente os bancários, adotam estratégias de *hardening* no lado cliente através de *plug-ins* ou extensões, sejam de autoria própria ou mantida por terceiros, a exemplo dos bancos brasileiros que utilizam a solução *Warsaw*[8]. Nesse contexto, seria interessante analisar a eficiência dessas soluções.

Por fim, outra característica seria a **sincronização entre a base de dados da plataforma e o navegador Web** que utiliza a plataforma como apoio ao mecanismo de proteção. Por exemplo, na página do *PhishTank*, determinados *phishing* estavam confirmados, mas não foram reconhecidos como uma ameaça quando acessados através do navegador Opera. O motivo seria o atraso na sincronização da lista negra do navegador com os registros do respectivo repositório. Considerando que o navegador é a primeira camada defensiva, fica evidenciado um problema crônico que merece maiores investigações, mas que por questão de escopo não foram contemplados nesse estudo.

[7] https://www.phishtank.com/developer_info.php

[8] https://www.dieboldnixdorf.com.br/gas-antifraude

## 7. Conclusões e Trabalhos Futuros

Esse estudo apresentou um *survey* como metodologia para obtenção de evidencias sobre determinados comportamentos a serem analisados em amostras extraídas de ambiente reais de atuação de *phishing*. Considerando que não são poucas as propostas de combate a *phishing*, contudo, o problema continua crônico nos dias atuais, justificando a avaliação através de um estudo primário. Como muitas das soluções são guiadas por um conjunto de características, o presente estudo trouxe como reflexão analisar a relevância de certas características comumente utilizadas no processo de predição na ótica da suscetibilidade.

Foi possível observar um grande número de investidas nos recursos do navegador Web que eram disparados pela própria URL. Os comportamentos que mais chamaram atenção foram a exploração em *punycode*, a concatenação de subdomínios, e práticas de *cybersquatting* e *typosquatting* em domínios, aspectos que estão sendo cada vez mais explorados. Em contrapartida, considerar uma página como maliciosa pela ausência ou presença do cadeado acabou perdendo a força com o passar do tempo. E conforme já mencionado, os dados destacam o Brasil como uma região bastante explorada por *phishing*.

Por ter extraído um considerável número de *phishing* reais, a análise realizada por esse estudo considera aspectos quantitativos, a exemplo dos gráficos expostos. Não obstante, por considerar o conteúdo e contexto de cada registro, bem como identificar relevâncias, relações e similaridades, o estudo também oferece resultados qualitativos. O objetivo é que esses dados possam oferecer apoio para o desenvolvimento de um avaliador de modelos de predição de *phishing*, utilizando métricas de avaliação como sensibilidade, especificidade e eficiência, considerando aspectos da suscetibilidade do usuário final.

Como trabalhos futuros, além da continuidade nas limitações das características mencionadas na Seção 6, também seria interessante mensurar ataques com escopo mais fechado, como o *spear phishing*. Devido sua natureza não direcionada a disseminação, esses ataques necessitam de uma abordagem mais direcionada. Contudo, de certa forma, os presentes resultados desse estudo talvez possam apoiar modelos de predição com foco em **proteção da marca**. Uma solução dessa natureza visa monitorar aspectos sobre a **identidade gramatical**, *cybersquatting* e *typosquatting* em domínios, bem como publicações falsas em redes sociais e uso de palavras-chave em motores de busca (SEO). Além disso, visa proteger a **identidade visual**, ou seja, o abuso de elementos que representam visualmente a organização, como *templates* e logomarcas, aspectos debatidos pelo estudo.

## Referências


[AlEroud and Zhou 2017] AlEroud, A. and Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. *Computers & Security*.

[Almomani 2018] Almomani, A. (2018). Fast-flux hunter: A system for filtering online fast-flux botnet. *Neural Comput. Appl.*, 29(7):483–493.

[Amiri et al. 2014] Amiri, I. S., Akanbi, O. A., and Fazeldehkordi, E. (2014). *A Machine-Learning Approach to Phishing Detection and Defense*. Syngress Publishing.

[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*.

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

[Goel and Jain 2018] Goel, D. and Jain, A. K. (2018). Mobile phishing attacks and defence mechanisms: State of art and open research challenges. *Computers & Security*.

[Google 2019] Google (2019). Google safe browsing. *Available at: https://safebrowsing.google.com/*.

[Gupta et al. 2016] Gupta, S., Singhal, A., and Kapoor, A. (2016). A literature survey on social engineering attacks: Phishing attack. In *2016 International Conference on Computing, Communication and Automation (ICCCA)*, pages 537–540.

[Khonji et al. 2013] Khonji, M., Iraqi, Y., and Jones, A. (2013). Phishing detection: A literature survey. *IEEE Communications Surveys and Tutorials*, 15(4):2091–2121.

[Leng Chiew et al. 2018] Leng Chiew, K., Yong, K., and Tan, C. L. (2018). A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications*, 106.

[Molleri et al. 2016] Molleri, J. S., Petersen, K., and Mendes, E. (2016). Survey guidelines in software engineering: An annotated review. In *10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*.

[OpenDNS 2019] OpenDNS (2019). Phishtank. *Available at: https://www.phishtank.com/*.

[Parsons et al. 2019] Parsons, K., Butavicius, M., Delfabbro, P., and Lillie, M. (2019). Predicting susceptibility to social influence in phishing emails. *International Journal of Human-Computer Studies*, 128.

[Qabajeh et al. 2018] Qabajeh, I., Thabtah, F., and Chiclana, F. (2018). A recent review of conventional vs. automated cybersecurity anti-phishing techniques. *Computer Science Review*, 29:44–55.

[Sharma et al. 2017] Sharma, H., Meenakshi, E., and Bhatia, S. K. (2017). A comparative analysis and awareness survey of phishing detection tools. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pages 1437–1442.

[Singh and Mangat 1996] Singh, R. and Mangat, N. S. (1996). *Stratified Sampling*, pages 102–144. Springer Netherlands, Dordrecht.

[Sonowal and Kuppusamy 2017] Sonowal, G. and Kuppusamy, K. S. (2017). Phidma - a phishing detection model with multi-filter approach. *Journal of King Saud University - Computer and Information Sciences*.

[Srinivasa et al. 2019] Srinivasa, R., Alwyn, R., and Pais, R. (2019). Jail-phish: An improved search engine based phishing detection system. *Computers & Security*.

[Stout and McDowell 2012] Stout, B. and McDowell, K. (2012). United states patent. Technical report, Citizenhawk, Inc., Aliso Viejo, CA (US).

[Vayansky and Kumar 2018] Vayansky, I. and Kumar, S. (2018). Phishing – challenges and solutions. *Computer Fraud & Security*, 2018:15–20.

[Windows 2019] Windows (2019). Windows smartscreen. *Available at: https://bit.ly/2ER8yow*.

[Wohlin et al. 2000] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2000). *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers, Norwell, MA, USA.