MARKETING DIGITAL E A CLICK FRAUD:

Conceitos Computacionais para a Defesa da Fraude com CAPTCHA Clicáveis

Rodrigo Alves Costa

Atena Editora

Ano 2018

Rodrigo Alves Costa

MARKETING DIGITAL E A CLICK FRAUD: Conceitos Computacionais para a Defesa da Fraude com CAPTCHA Clicáveis

Atena Editora
2018

2018 *by Atena Editora*

Editora Chefe: Profª Drª Antonella Carvalho de Oliveira
Diagramação e Edição de Arte: Geraldo Alves e Natália Sandrini
Revisão: Os autores

Dados Internacionais de Catalogação na Publicação (CIP)
(eDOC BRASIL, Belo Horizonte/MG)

C837m

Costa, Rodrigo Alves.
Marketing digital e a click fraud [recurso eletrônico]: conceitos computacionais para a defesa de fraude com captcha clicáveis / Rodrigo Alves Costa. – Ponta Grossa (PR): Atena Editora, 2018.
217 p. : il. ; 14 x 21 cm

Formato: PDF
Requisitos de sistema: Adobe Acrobat Reader
Modo de acesso: World Wide Web
Inclui bibliografia
ISBN 978-85-455090-6-6
DOI 10.22533/at.ed.066181110

1. Marketing na internet. 2. Fraude – Prevenção. I. Título.
CDD 658.800285

Elaborado por Maurício Amormino Júnior – CRB6/2422

Dedico este trabalho ao tesouro da sabedoria e do conhecimento, sem o qual ninguém verá a Deus, mas diante de quem todo joelho se dobrará, Jesus.

AGRADECIMENTOS

A sensação de dever cumprido é uma recompensa significante e gratificante, no fim de uma batalha. Sinto-me um privilegiado por ter tido a oportunidade e a possibilidade de realizar este trabalho e concluir tão importante etapa em minha vida e carreira. O que parecia impossível, Deus tornou possível e por isto estamos alegres.

Acima de tudo, agradeço a Ele, o Senhor Deus de toda a terra, o Soberano, mas que amou o mundo de tal maneira que deu Seu Filho unigênito, para que todo aquele que crer neste Filho não pereça, mas tenha a vida eterna. Quem entende este mistério, por mais que seja visto como insensato, verdadeiramente não vive de qualquer maneira. Assim sendo, a Deus toda a glória.

Agradeço aos meus pais, Hianto (*in memoriam*) e Núbia, por terem tantas vezes abdicado de suas vidas para que minhas irmãs e eu pudéssemos ter a melhor educação possível. Tenho certeza que o galardão deles está no Senhor.

Agradeço a minha noiva Karine, pelo incentivo e por encher minha vida de amor e companheirismo, e também por ser uma pessoa tão maravilhosa e apoiadora em todos os instantes.

Agradeço de maneira muito especial ao professor Ruy Guerra, por sua disponibilidade e receptividade constante, que me orientou com tanta dedicação e apoio, sempre com muita atenção, me apresentando os melhores caminhos.

Agradeço à editora Atena, pela confiança e interesse no meu trabaho.

Agradeço ao representante da coordenação da pós-graduação e a todos os professores do Centro de Informática da UFPE, pela dedicação e estímulo à pesquisa.

Por fim, um agradecimento particular para todos aqueles que me incentivaram com uma palavra de ânimo, ou com algo que colaborou para a realização deste trabalho.

PREFACIO

Um estudo da PwC de 2009 intitulado “Measuring the effectiveness of online advertising” já dizia que nos cinco anos anteriores os recursos alocados à mídia de Internet haviam crescido espetacularmente, prevendo que em 2010 a Internet representaria 16% do investimento total mundial com anúncios, e que os números poderiam atingir 21% nos quatro anos seguintes. Uma das constatações foi a de que esse crescimento estaria sendo significativamente alimentado por ferramentas de busca e de “performance” tais como marketing de afiliados, e-mail, websites de comparação, etc., muito embora o chamado anúncio de display (em inglês, “display advertising”) continuasse a representar uma grande parcela dos orçamentos de propaganda online.

Entre as tendências que estariam alimentando esse “boom” estariam o aumento no uso da web reforçando o papel da Internet não apenas na recomendação como também na preparação dos consumidores para realizar compras e recomendações, e os desenvolvimentos nos formatos e técnicas de propaganda dirigida que acabam ajudando a tornar as campanhas publicitárias mais relevantes e mais comunicativas.

Entre as formas de marketing digital estão os anúncios contextuais em páginas de resultados de engenhos de busca, os banners, anúncios em mídia interativa, anúncios em redes sociais, anúncios intersticiais, anúncios em videogames, anúncios classificados online, redes de anunciantes e marketing por correio eletrônico. Uma das grandes vantagens do marketing digital é sua publicação imediata, além do fato de que o conteúdo não fica limitado por geografia ou por tempo. Por outro lado, a eficiência do anúncio atinge níveis bem superiores às formas mais tradicionais de propaganda pois permite a customização e a interatividade de anúncios e a medição mais precisa do seu impacto.

Embora não esteja imune às intempéries da economia mundial, o negócio do marketing digital ainda é um dos que mais crescem no momento: para o ano de 2012 a JupiterResearch estima que somente nos EUA o investimento em propaganda online deve atingir a casa dos 34,5 bilhões de dólares.

O fato concreto é que, hoje em dia, anunciar pela Internet é uma das formas mais rentáveis, tanto para pequenas quanto para grandes empresas, de realizar campanhas de marketing com o objetivo de atingir diversos tipos de clientes. Um advertiser de Internet (por exemplo, o eBay) provê os seus anúncios a um representante (por exemplo, o ValueClick), reserva uma determinada quantia de dinheiro e se compromete a pagar uma comissão para determinadas ações dos usuários do serviço do representante, como, por exemplo, clicar em um anúncio, realizar uma compra, ou dar um lance em um leilão. Os publishers de Internet (por exemplo, o MySpace.com), motivados pela comissão paga pelos advertisers, procuram os representantes com o objetivo de contratá-los para exibir os anúncios nas suas páginas Web, e obterem assim uma

parte da comissão. O ponto principal desta relação são os representantes, que atuam como mediadores entre os publishers e os advertisers.

Sempre que um usuário da Internet visita a página de um publisher, este usuário é associado a um dos servidores do representante. Este servidor escolhe um conjunto de anúncios e exibe este conjunto na página de publisher, que está sendo exibida para o usuário em seu navegador. Se o usuário clicar no anúncio no site do publisher, esta ação será associada ao servidor do representante (que exibe o anúncio na página do publisher), que salva o clique em um histórico, para cobrança posterior, e direciona o usuário à página do advertiser cujo anúncio foi clicado.

Uma vez que os publishers lucram com os eventos de clique nos anúncios dos advertisers, é possível observar um incentivo para que publishers desonestos aumentem o número de cliques que seus sites geram. Além disso, advertisers desonestos simulam cliques nos anúncios de seus concorrentes com o objetivo de esgotar seus recursos destinados a anúncios e marketing. O mundo da publicidade online rejeita estas práticas, conhecidas como fraude de clique. A fraude de clique resulta em má reputação para os representantes, e há diversos casos de pagamento de multas para advertisers. O fato é que esse tipo de fraude põe em risco toda a indústria de anúncios pela Internet.

A fraude de clique tem sido uma preocupação para representantes de anúncios desde a sua concepção. Os números são difíceis de quantificar; existem diversas formas de se estimar a proporção de cliques falsos, que variam de 10% a 50%. Um estudo amplamente citado da MarketingExperiments.com, uma ferramenta de pesquisa em marketing online, relatou que 29,5% dos cliques em três campanhas experimentais do Google eram fraudulentos.

Mesmo com números potencialmente tão expressivos, as empresas de busca e muitos dos seus clientes vêm defendendo que o problema em suas redes está sob controle. Entretanto, alguns observadores do mercado de cliques online, como a Holcomb, acreditam que a fraude de clique traz prejuízos da ordem de bilhões de dólares e, como dito anteriormente, possuem o potencial de causar danos importantesà indústria cimo um todo. Independentemente do número exato, a fraude de clique hoje está impregnada no negócio de anúncios pela Internet, e, muito embora as ferramentas de busca procurem se defender de diferentes maneiras, os fraudadores tornam-se cada vez mais sofisticados e os programas utilizados para automatização da fraude são cada vez mais complexos, disfarçando, inclusive, a origem dos cliques.

As pesquisas realizadas nesta área, em sua grande maioria, investigam a fraude do publisher, já que ela pode ser generalizada para a fraude do advertiser. Além disso, tais pesquisas invariavelmente discutem a detecção da fraude através de diversos métodos, tais como: a abordagem criptográfica, técnicas de análise de dados, ferramentas para detecção de fraude, análise de tráfego, e algoritmos de força bruta. Entretanto, todos estes métodos são técnicas de detecção, e tratam a fraude depois que ela ocorreu. Como já dito anteriormente, os programas têm se tornado

cada vez mais complexos e a detecção da fraude tem se tornado um problema de difícil resolução. Por estas razões, é preciso desenvolver uma metodologia focada na prevenção da fraude de clique, de maneira que a detecção se faça desnecessária, ou no mínimo secundária.

Com efeito, encontra-se em pleno desenvolvimento no contexto do grupo de Segurança Computacional do Centro de Informática da UFPE, e é um dos resultados apresentados neste livro, um esquema que envolve uma nova entidade, o comprovador, que por seu turno provê credenciais a clientes que respondam a um teste. Tais credenciais têm o papel de permitirque o representante seja capaz de distinguir os cliques válidos, realizados por humanos, de cliques originados do tráfego em geral. O comprovador é uma forma de classificar o clique, de maneira fortalecer a heurística de isolamento de cliques fraudulentos. A ideia da técnica inovadora é complementar a técnicas existentes para filtragem de cliques para validação.

No final das contas, trata-se de um método de prevenção de fraude de clique, o que vai ao encontro da grande maioria dos métodos de combate àfraude de clique atuais, que tratam a fraude após a ocorrência, por meio da detecção de cliques fraudulentos. Contrariamente aos métodos atuais de filtragem, o esquema proposto neste livro se baseia no uso de testes de diferenciação entre humanos e computadores, através de CAPTCHA. A resposta destes têm a função de servir de atestado de validade dos cliques, que após considerados “bons”, são contabilizados.

Prof Dr. Ruy José Guerra Barreto de Queiroz, professor associado, Centro de Informática da UFPE

SUMÁRIO

CAPÍTULO 1

A ECONOMIA GLOBAL E OS ANÚNCIOS PELA INTERNET

Este capítulo discute as principais motivações para a realização deste trabalho, apresenta a finalidade de nossa pesquisa, o sumário da contribuição, o escopo do livro e, finalmente, a estrutura do livro.

1.1 Fatores Motivadores do Estudo

No atual clima desafiador da economia global, anunciantes e suas agências encontram-se em contínua busca por oportunidades de negócios que os levem a aumentar significamente – e efetivamente – a exposição de suas marcas e de seus produtos a custos cada vez menores (ARRINGTON, 2010).

Nesse cenário, recursos anteriormente destinados a campanhas de *marketing* clássicas, que normalmente têm sua estrutura de pagamento associada ao número de pessoas expostas aos anúncios, têm sido redirecionados a campanhas de *marketing* na Internet do tipo "pagamento por clique" ou "pagamento por ação", cujos pagamentos estão associados à realização de alguma ação, por parte dos potenciais consumidores que seja considerada benéfica para o anunciante – como o clique em um anúncio.

Dados de diversos relatórios acerca da disposição competitiva das formas de marketing digital (BROWSER MEDIA, 2010) mostram que o total de recursos associados com marketing digital aumentaram até 15% na primeira metade do ano.

Em contrapartida, fundos associados à forma de anúncios por exibição foram reduzidos em 6% pelo mesmo período. Estes dados claramente significam que os recursos de marketing das empresas têm sido cada vez mais deslocados das formas de anúncios clássicas para programas do tipo pagamento por clique (PPC).

De acordo com Anupam et al (1999), o modelo de PPC, que será explicado em maiores detalhes no capítulo 2 deste livro, envolve um anunciante, que contrata uma entidade especializada para a exibição de seus anúncios (como uma rede de anúncios), que por sua vez distribui conteúdo textual ou gráfico, abrigando os anúncios, para diferentes publicadores (como portais na Internet) deste conteúdo.

Quando um usuário clica no banner exibido na página do publicador, é direcionado para o site associado ao anúncio. As entidades atualmente mais populares, que funcionam como grandes repositórios de anúncios, são ferramentas de busca como o Google ou o Yahoo, e controlam a maior porção do tráfego relacionado a PPC na Internet. Essas entidades tanto mostram os anúncios em seus próprios

sites de pesquisa (quando funcionam também como publicadores), quanto terceirizam os anúncios para outros publicadores. Cada clique de usuário de Internet em um determinado anúncio gera uma cobrança, por parte da entidade, para o anunciante.

A entidade, por sua vez, repassa uma parte dos pagamentos para os publicadores. Na Figura 1.1, podemos ver um exemplo que contém anúncios do tipo pagamento por clique, no qual o Google realiza papel duplo, funcionando tanto como publicador quanto como entidade de anúncios. Pode-se ver as seguintes entidades nessa figura:

1) Palavras-chave que foram buscadas na consulta;

2) Anúncios do tipo PPC (que foram exibidos baseados na busca); e

3) Resultados da busca.

Os servidores que hospedam as páginas do repositório de anúncio e dos publicadores observam cada clique como uma requisição de uma URL que está associada a um determinado anúncio e, assim, não tem, atualmente, a capacidade de determinar se um humano iniciou a requisição ou, se um humano estiver envolvido, se o mesmo agiu consciente e honestamente.

O ato de gerar cliques ilícitos como uma tentativa de fraudar a disposição deste modelo de negócios é denominado de *click fraud*, e também é explorado no capítulo 2.

Figura 1.1 – Exemplo de anúncios do tipo pagamento por clique

Conforme descrito por Mann (2006), a *click fraud* pode beneficiar os fraudadores de diversas formas. Antes de mais nada, um fraudador pode se utilizar deste meio para aumentar os lucros de um publicador. Posteriormente, a *click fraud* pode ser empregada por concorrentes para prejudicar uma determinada companhia através da geração de custos indevidos.

Finalmente, um fraudador pode modificar o ranking de exibição de anúncios por meio do emprego de uma combinação de impressões e cliques – uma impressão é a visualização de um banner de anúncios sem cliques, o que causa o ranking dos anúncios associados diminuir. Esta terceira forma de *click fraud* pode ser utilizada para beneficiar campanhas próprias de marketing em detrimento de concorrentes.

Todas as partes envolvidas no negócio de marketing digital podem ser prejudicadas direta ou indiretamente pelas fraudes de clique. Por exemplo, embora vejamos um benefício inicial para as entidades de anúncio – pois as mesmas lucram sempre que um clique é realizado –, a longo prazo, quando os seus clientes passam a se tornar sensíveis a perdas, a *click fraud* pode colocar em risco a relação entre essas entidades e os seus anunciantes. Para garantir a credibilidade do negócio de marketing digital , toda e qualquer forma de fraude precisa ser combatida (STONE, 2004).

Conforme descrito na metodologia DETECTIVES (METWALLY; AGRAWAL; ABBADI, 2007), tipicamente, as entidades adotam uma postura de detecção da *click fraud* para impedir que os valores de cobrança indevidamente gerados sejam de fato computados junto aos anunciantes.

Essas entidades realizam buscas nos seus históricos de navegação por cliques fraudulentos, baseados nos tipos de anúncios requisitados, nos custos das palavras-chave que foram buscadas, nos endereços IP que originaram as requisições e no número de requisições desses endereços.

Embora filtros heurísticos sejam eficientes contra a *click fraud* menos especializada, estes possuem limitações contra fraudadores mais sofisticados, que possuem a tendência de “driblar” tais filtros de maneira a não terem seus cliques detectados por estes.

Em paralelo, esta abordagem não impede que a *click fraud* aconteça de fato, mas toma medidas caso a mesma seja detectada. Entretanto, tomando por base a premissa de que a melhor forma de segurança neste modelo de negócios seria a prevenção de cliques fraudulentos, neste trabalho é proposta uma abordagem alternativa – ao invés de realizar buscas para detectar cliques inválidos, são consideradas formas de impedir que a fraude ocorra, através da realização de testes para:

1. Identificar que os usuários clicando nos anúncios são, de fato, humanos, e não programas automatizados;
2. Autenticar cliques válidos de clientes comprovadamente reais, isolando, desta forma, riscos de fraude.

Impedir a *click fraud* parece ser de fundamental importância para a continuidade e o crescimento do mercado de marketing digital . Anunciantes potencialmente mais conservadores acabam se afastando desta forma de *marketing* por sua natural aversão ao risco (ZELLER, 2004) e, assim, a possibilidade de se desenvolver novos negócios acaba impactada. Além disso, a forma de anunciar na web descrita anteriormente, no modelo de redes de anúncios, é extremamente benéfica para o consumidor de internet, e a *click fraud* coloca este tipo de anúncio em risco.

Para se alcançar o mundo ideal, no qual a *click fraud* ou inexiste ou possui influência mínima no faturamento das companhias, é necessário o desenvolvimento de diversos aspectos arquiteturais e o surgimento de técnicas de segurança que suportem o crescimento deste mercado.

Neste cenário, a academia desempenha papel fundamental, por se tratar da forma de pesquisa e desenvolvimento mais indicada para a predileção de cenários econômicos e para a exploração – e resolução – de situações tecnicamente desafiadoras.

Este trabalho tem por objetivo primordial a proposição de uma abordagem cuja hipótese é que a eliminação da fraude automática de clique em taxas muito significativas é possível e que o compartilhamento de informações entre participantes de uma federação de anunciantes, observando padrões de privacidade, pode ser a solução para a erradicação da *click fraud* não-automática por tornar a sua realização extremamente custosa.

Nestes cenários, pode-se observar a necessidade de uma mudança na forma como a sistemática de negócios online é conduzida atualmente. Assim, é necessário um ambiente de colaboração entre as diversas empresas participantes de uma rede de anúncios. Além disso, métodos para diferenciação entre usuários humanos e não-humanos precisam ser continuamente aprimorados.

Esta pesquisa sugere um método que estende os testes tradicionais desta natureza para aumentar a sua eficiência, embora se torne necessário, para isso, a colaboração mútua e o alinhamento de interesses.

O pagamento por clique tem importância significativa para a forma atual da realização de negócios pela Internet. Pode ser visto como a principal vitrine da forma de anúncios através de entidades de anúncio, que são cada vez mais populares. Sua capacidade de penetração junto aos consumidores é notória, em termos de conhecimento da marca e da realização de negócios em si, e todos os envolvidos possuem interesses que precisam ser atendidos para garantir o sucesso desta forma de negócios.

Neste sentido, este trabalho apresenta também uma contribuição para anunciantes na internet, que possuem o natural interesse de elevar a segurança e a garantia de negócios legítimos nos serviços atualmente existentes de pagamento por clique. Esta pesquisa, por meio de um livro acessível, explora o estado atual da arte e descreve uma proposta de arquitetura que leva em consideração elementos-chave deste estado da arte, com o objetivo de avançar a tecnologia atual.

Paralelamente, este trabalho apresenta contribuições para os portais que publicam anúncios na medida em que, ao se exporem a este livro, estes eventualmente passarão a dispor de uma ferramenta escrita que pode auxiliá-los no processo de tomada de decisão acerca dos serviços de anúncio atualmente disponíveis, incluindo as diferentes redes de anúncios, de maneira a melhor entenderem os riscos aos quais estão expostos e garantir a confiabilidade do conteúdo que está sendo exibido em suas páginas.

Também é objetivo deste trabalho que os participantes fundamentais deste negócio de anúncios na Web entendam a necessidade de realizar negócios de maneira ética e socialmente responsável e como atingir altos níveis de confiabilidade influencia positivamente a sustentabilidade desta forma de *marketing*.

Quanto às redes de anúncios atualmente existentes, a possíveis entrantes e a outras formas de entidades que funcionam como repositório de anúncios para a Internet, este livro é importante no sentido de motivá-los, enquanto grandes interessados no assunto em termos comerciais, a buscar sempre estimular os fatores acima entre seus clientes e usuários e a melhorar a confiabilidade através de uma análise criteriosa da segurança de seus serviços.

Por fim, este livro é recomendada também a usuários de Internet em geral, que são consumidores em potencial dos serviços disponibilizados pelas redes de anúncios e publicadores. Ela os auxiliará a entender a disposição atual do negócio de pagamento por clique na Internet e a buscar, de maneira pró-ativa, dentre as diversas opções disponíveis, a utilização de serviços que disponibilizam níveis aceitáveis de segurança. Estes serviços devem contribuir, de maneira ética, para a evolução contínua da utilização de negócios online.

1.2 O Problema de Negócios do PPC

O problema principal desta pesquisa está primariamente fundamentado no relatório de Tuzhilin (TUZHILIN, 2006), que foi produzido como parte do acordo judicial entre as empresas Google e The Lane's Gift, resultado de uma ação judicial motivada pela *click fraud*. Neste relatório, pode-se encontrar uma discussão detalhada e abrangente dos problemas decorrentes da *click fraud*. Entre outros conceitos, este relatório define "o problema fundamental da *click fraud*":

1- Com a excessão de casos óbvios, é impossível haver uma definição conceitual que possa ser operacionalizada do que venham a ser "cliques inválidos";

2- Ainda que houvesse, uma definição operacional não poderia ser disponibilizada para o grande público por causa das possibilidades de usuários sem ética tentarem obter alguma forma de vantagem sobre essa definição, o que poderia causar uma onda de *click fraud* maciça.

Entretanto, ao passo que não é disponibilizada, os anunciantes encontram-se em uma situação muito delicada, pois não podem nem verificar nem contestar a razão pela qual foram cobrados por determinados cliques.

De acordo com Metwally, Agrawal e Abbadi (2007), esses aspectos devem ser vistos como um dos fatores principais da sustentabilidade do pagamento por clique, relacionando confiabilidade, qualidade e custos.

Esta pesquisa tem a finalidade de, levando em consideração o acima exposto, propôr uma mudança de paradigma na forma como a *click fraud* é combatida atualmente no mercado e na academia, saindo de uma abordagem de detecção de cliques através da análise de históricos (ou seja, após os mesmos terem anteriormente ocorrido) e se aproximando de uma abordagem de prevenção, através da utilização de um mecanismo de verificação de validade dos usuários, que não seja custosa para as companhias e nem prejudique a experiência do usuário na utilização dos serviços.

Levando em consideração esta finalidade, é de fundamental importância a realização de um estudo extensivo do estado atual da arte da sistemática de marketing digital , quais as formas de anúncios e metodologias utilizadas para buscar o entendimento de quão expostas estas se encontram à *click fraud*.

Assim, este trabalho pretende investigar e delimitar o problema da *click fraud*, existente na disposição atual dos mercados, em razão de ser este um questionamento relevante, proeminentemente identificado na literatura relacionada de marketing digital , como mencionado em (MANN, 2006), (SAGAR; SEO, 2010) e (LIEDTKE, 2006).

Um passo natural desta investigação está na análise de *benchmarks* que poderiam ser potencialmente utilizados como componentes da abordagem a ser proposta, e uma análise destes *benchmarks* em função do modelo de negócios de marketing digital , de maneira a selecionar uma solução, ou um conjunto de soluções, que possa ser utilizada sem modificar drasticamente este modelo.

Levando em consideração cada um dos estudos realizados acima, este livro objetiva a proposição de uma abordagem que possa ser posteriormente desenvolvida para garantia de que a *click fraud* foi contornada a níveis aceitáveis em uma rede de marketing digital .

1.3 Resumo das Contribuições

Este livro produz várias contribuições para a área de desenvolvimento de soluções para *Anúncios na Web* e, especificamente, para o negócio denominado *pagamento por clique*.

As contribuições estão relacionadas ao estudo extensivo do estado atual da arte de diferentes abordagens da área de marketing na Internet e à proposição de uma arquitetura que explora os principais aspectos positivos de cada uma destas abordagens, buscando auxiliar a percepção e a comunicação entre os diversos participantes de uma estrutura de *pagamento por clique*.

No final das contas, objetivamos melhorar a aquisição de conhecimento do domínio do negócio de anúncios na Web para apoiar na identificação de requisitos e necessidades de sistemas deste domínio. Mais especificamente, enumeramos as

contribuições da seguinte forma:

- Estudo da sistemática do marketing digital e de sua atual disposição de negócios, com o objetivo de garantir que a abordagem aqui proposta se encaixa neste modelo de negócios;
- Análise das formas de monetarização das empresas envolvidas nesta sistemática, dando foco especial ao pagamento por clique, e estudo de como esta forma de obtenção de receitas se relaciona com a *click fraud*;
- Estudo de um conjunto de métodos de fraude neste nicho de negócios, dando especial destaque à fraude do clique. Estudo de casos legais da fraude do clique, em especial do relatório de Tuzhilin (TUZHILIN, 2006)*;*
- Estudo de CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart), histórico, classificação, análise de *benchmarks*, OCR (Optical Character Recognition), e a introdução de CAPTCHA clicáveis como uma forma para prevenir a automatização da *click fraud*;
- Proposta da abordagem C_2FAC_2A (Combatendo a *Click Fraud* Através de CAPTCHA Clicáveis e Autenticação), baseada na utilização de CAPTCHA clicáveis e autenticação por cupons originários de *cookie*s, que se propõe a ser de prevenção de *click fraud*, de acordo com a abordagem proposta, que possa servir de referência para desenvolvimento futuro;
- Proposta de uma arquitetura de rede de anúncios que implementa esta abordagem. Em termos desta arquitetura, faz parte do escopo desta pesquisa:
    - Apresentação da necessidade de um banco de dados como um componente desta arquitetura;
    - Apresentação das diversas entidades e serviços envolvidos com o esquema da rede de anúncios;
    - Proposição do tipo de CAPTCHA a ser utilizado nesta abordagem e proposta de um esquema de desenvolvimento do fluxo de comunicações associado à resolução dos desafios;
    - Proposição da sistemática de comunicação entre os usuários e a rede de anúncios, que irá incluir troca de cupons, respostas aos desafios e fluxo de comunicações;
    - Proposta de autenticação por cupons baseados em *cookie*s;
    - Análise de segurança da abordagem

Por se tratar de uma estrutura extremamente complexa, com muitas entidades participantes, é importante detalhar também os aspectos que não fazem parte do escopo desta pesquisa:

- Estudo dos bancos de dados disponíveis que poderiam ser utilizados como componente desta abordagem;
- Proposição de um modelo de dados para a arquitetura como um todo;
- Estudo de heurísticas para construção de anúncios a serem exibidos nos

sites de Internet;

- Estudo e escolha de algoritmos para a criação de desafios CAPTCHA;
- Estudo extensivo do histórico e da tecnologia de *cookie*s e de alternativas ao seu uso. Em compensação, serão referenciados trabalhos que realizam esses estudos e será oferecida uma proposta detalhada da utilização de *cookie*s nesta abordagem.

1.4 Escopo do livro

Este livro enfoca a definição de uma abordagem para implementação de sistemas para pagamento por clique, com o objetivo de explorar extensivamente o estado atual da arte neste domínio da informação

A Figura 1.2 demonstra que o escopo deste livro reúne a área de pesquisa de Arquitetura em projetos de desenvolvimento de Sistemas de Informação para a Web sob a perspectiva de segurança da informação.

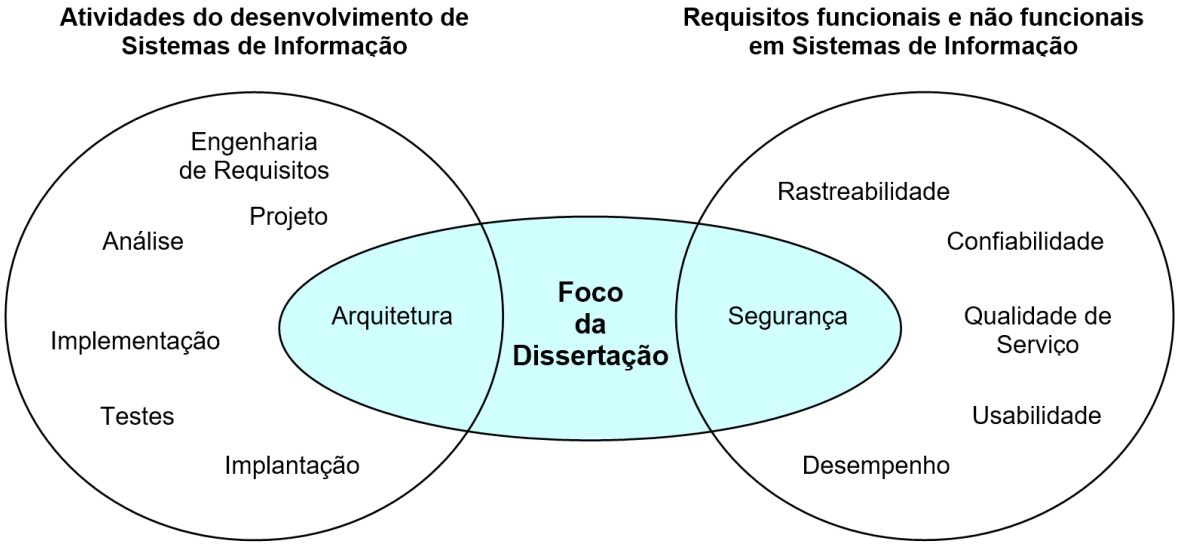

Figura 1.2 – Foco do livro

A abordagem proposta pretende favorecer a melhoria contínua da segurança na sequência de atividades envolvidas na sistemática de marketing digital que utilizam como forma de realização de negócios o modelo de pagamento por clique, majoritariamente presente em redes de marketing digital .

Na condição de priorizar a segurança deste processo de comunicação, e o entendimento, através da descrição do modelo proposto, das diversas ameaças ao funcionamento legítimo deste nicho de negócios, nossa abordagem faz uso da representação de conceitos e argumentos para acoplar objetivos e processos envolvidos nesta sistemática de comunicação.

1.5 Estrutura deste Livro

Além deste capítulo introdutório, este trabalho consiste de mais quatro capítulos, que estão organizados como apresentado a seguir:

Capítulo 2 – Marketing digital e a *Click Fraud*

Este capítulo discute em mais profundidade a sistemática de marketing digital , sua evolução e as disposições e tendências atuais deste mercado. Além disso, um estudo aprofundado da *click fraud* é apresentado, com o objetivo de explicar o seu impacto no mercado anteriormente abordado.

Capítulo 3 – Garantindo que Usuários Humanos Utilizam Serviços na Web

Este capítulo explora metódos atualmente conhecidos para diferenciação entre usuários humanos e programas automatizados. A diferenciação é, naturalmente, de fundamental importância para o combate da *click fraud* e um componente importante na abordagem a ser proposta.

Capítulo 4 – A Abordagem C_2FAC_2A

Neste capítulo, a abordagem C_2FAC_2A (Combatendo a *Click Fraud* Através de CAPTCHA Clicáveis e Autenticação) é apresentanda. Serão discutidas as formas de interação dessa abordagem com o usuário, a frequência dessa interação, a forma como os CAPTCHA clicáveis seriam alimentados e uma análise de segurança da abordagem proposta. O diálogo com potenciais implementações é inevitável, e é também parte integrante deste capítulo.

Capítulo 5 – Conclusões e Trabalhos Futuros

Este capítulo aponta as principais contribuições desta pesquisa para a área de marketing digital , especificamente para a subárea de pagamento por clique. Traz, ainda, as principais lições aprendidas durante o processo de estudo da tecnologia atual que diz respeito a esta área do conhecimento. Serão mostrados alguns trabalhos relacionados, assim como será feito um direcionamento para futuras pesquisas nesta área.

CAPÍTULO 2

MARKETING DIGITAL E A *CLICK FRAUD*

Este capítulo fornece uma visão geral das diversas possibilidades existentes para se anunciar na Internet, destacando o método de pagamento por clique. Em seguida, examina esquemas ilegais que foram criados e desenvolvidos para a obtenção de ganhos pessoais ou para prejudicar outros, e aborda algumas técnicas fraudulentas atualmente utilizadas, incluindo a *click fraud*, foco deste trabalho.

Dentro do universo da *click fraud*, é realizada uma análise de como as diferentes formas de *click fraud* estão relacionadas, listando casos legais. Por fim, procede-se a uma consolidação da literatura existente por trás da *click fraud*.

2.1 Considerações Iniciais

Os mercados e a economia global, na qual operam e competem os negócios, têm sido alvos de constantes mudanças na última década. O desenvolvimento de *sites* de Internet comerciais acompanhou a popularização da própria Internet, desde meados de 1997. Anunciar online tem se tornado comum e os anúncios são cada vez mais chamativos: em janelas de *pop-up*, que tocam músicas e trilhas sonoras, que nadam através da tela, entre outros.

Esta tendência é verificada em praticamente todos os sites comerciais na Internet. Há diversas formas de anunciar *online*, e o seu uso tem se tornado cada vez mais óbvio. Muitos usuários da Web têm diversas questões sobre estes novos tipos de anúncio, tais quais:

- Por que há tantos anúncios nos *Websites* hoje em dia?
- Por que os *Websites* permitem anúncios em janelas de *pop-up* que abrem novas janelas?
- Por que os *Websites* permitem anúncios flutuantes, que cobrem conteúdo do próprio *site*?
- Como fazer todos estes anúncios desaparecerem?

Neste capítulo, diferentes formas de anunciar online serão estudadas, assim como a motivação econômica por trás de cada uma delas.

2.2 Marketing digital

Hoje em dia, anunciar pela Internet é uma das formas mais rentáveis de realizar campanhas de marketing com o objetivo de atingir diversos tipos diferentes de clientes, tanto para pequenas quanto para grandes empresas. Um dos grandes benefícios de anunciar *online* é a possibilidade de se publicar informação e

conteúdo sem fronteiras geográficas ou fuso horário. Com este fim, a área emergente de anúncios interativos apresenta novos desafios para anunciantes.

Outro benefício é a eficiência do investimento do anunciante. Anunciar online permite a customização de anúncios, incluindo o conteúdo dos mesmos e os lugares onde serão exibidos. Por exemplo, o AdWords, o AdSense (GOODMAN, 2008) e o Yahoo! Search Marketing (O'REILLY, 2007) permitem que anúncios sejam mostrados tanto em páginas Web relevantes quanto em resultados de pesquisas de palavras-chave relacionadas.

2.2.2 No Princípio, Anúncios em Banners

Quando a Internet começou a ser utilizada efetivamente para fins comerciais, em meados de 1997, milhares de novos *sites* nasceram e bilhões de dólares em capital de ventura fluíram através deles. Essa foi a primeira vez que os *Websites* se dividiram em categorias diferentes (SAGAR, 2009):

- *Sites* de e-commerce: *sites* cujos objetivos eram a venda de produtos. Tais sites faturam por meio dos produtos que vendem.
- *Sites* de conteúdo: *sites* que criam ou colecionam conteúdo (palavras, imagens, vídeos, entre outros) para que pessoas interessadas e leitores os acessem. Esses *sites* faturam primariamente através dos anúncios contidos neles, similarmente a canais de TV, estações de rádio e jornais.

Nesse período, anunciar na Internet significava elaborar anúncios em *banners*, que são imagens de de 728x90 pixels exibidas em quase todas as páginas na Internet de hoje. No fim dos anos 90, esta forma de anunciar na Web era extremamente lucrativa, e sites populares como o Yahoo cobravam entre 30 e 100 dólares a cada mil exibições dos anúncios em *banners* em suas páginas.

Essas taxas de anúncio representavam uma boa parte do capital de ventura disponível na Web (por exemplo, se um *site* fosse capaz de gerar 100 milhões de exibições diferentes por mês, a uma taxa de 30 dólares a cada mil exibições, poderia lucrar 3 milhões de dólares em um mês apenas com *banners*), e haviam sido herdadas dos modelos e das taxas cobradas por revistas de grande circulação para imprimir anúncios em suas páginas.

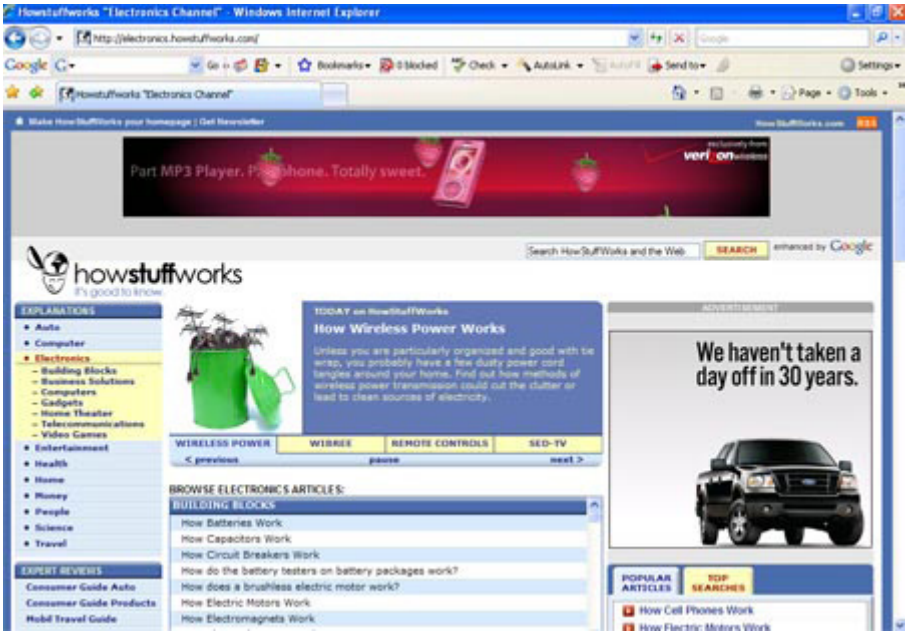

Figura 2.1 – Anúncio em *banner* padrão no topo da página

Entretanto, a partir de um determinado ponto, anunciantes chegaram à conclusão que anúncios em *banners* não eram tão eficientes quanto outras formas de anúncio, como comerciais de trinta segundos na TV, ou até mesmo páginas impressas em revistas de grande circulação. Ao mesmo tempo, havia na Internet uma natural disposição para anúncios – milhares de *sites* com milhões de acessos por mês.

Desta forma, empresas como a DoubleClick (O'REILLY, 2007) começaram a armazenar esses sites em repositórios de anúncios de *banner*. O princípio econômico da "oferta e procura" começou a ser aplicado na Internet, de maneira que as taxas pagas por anúncios em *banners* começaram a flutuar.

Para determinar o valor de um *banner* em um site é necessário analisar a razão pela qual uma empresa compra espaços de anúncio. Tradicionalmente, isto acontece por duas razões (HOFFMAN; NOVAK, 2000):

- Desenvolvimento da marca (*branding*)
- Vendas diretas

O desenvolvimento da marca (*branding*) refere-se ao processo de estabelecimento do nome de uma empresa ou produto na memória coletiva da sociedade. O processo de desenvolver uma marca acontece tanto para produtos novos quanto para produtos que já existem, e não tem o objetivo primário de vender os produtos associados à marca imediatamente – o anunciante quer apenas ser lembrado na consciência de uma população.

Em contrapartida, o anúncio de venda direta tem o objetivo de despertar uma ação imediata no usuário. O anunciante quer que o usuário, ao se deparar com o anúncio, imediatamente clique nele, ligue para um número de telefone (gratuito ou não), dirija-se para uma loja, realize uma compra, baixe algum arquivo da Internet ou ainda se inscreva em algum formulário.

O anunciante contabiliza as respostas diretas ao anúncio e mede a eficiência do anúncio através de tais respostas. Sendo assim, a idéia daqueles que buscavam anunciar com o objetivo de desenvolver uma marca passou a ter sobre anúncios em *banner* é que estes não eram o veículo mais eficiente para *branding*. Se comparados a anúncios em jornais, revistas ou na TV, anúncios em *banner* são bem menores e facilmente ignorados.

Da mesma forma, as taxas de resposta a anúncios que tinham por objetivo a realização de vendas diretas se mostraram baixas. Para a maioria dos anúncios em *banner*, a média da indústria oscilava entre dois e cinco cliques a cada mil exibições do anúncio. Uma taxa tão baixa quanto esta não tem valor significativo para os anunciantes, especialmente se levamos em consideração que nem todos os cliques em *banner* são obrigatoriamente traduzidos em vendas.

Assim, as taxas cobradas por grandes *websites* para se colocar anúncios em *banners* passaram a declinar. Para a maioria dos *sites*, há hoje em dia muito pouco dinheiro a ser obtido de anúncios em *banner*. As grandes empresas de anúncio da Internet passaram então a procurar soluções para que seus anúncios passassem a:

1. Possuir mais influência em termos de *branding*
2. Atingir uma taxa significativa de cliques em anúncios

Fruto destas necessidades, diversos formatos de anúncio na Internet foram desenvolvidos.

2.2.2 Anúncios em Barra Lateral

Um anúncio em barra lateral (também conhecido como *sidebar* ad ou *skyscraper ad*) é similar a um anúncio em *banner*, mas é orientado verticalmente. Como é vertical, a altura da barra lateral pode atingir 600 pixels ou mais, com uma largura de normalmente 120 pixels (KANG; LEE, 2003).

Um anúncio em barra lateral tem mais impacto do que um anúncio em banner por duas razões:

- Uma barra lateral alta é duas ou três vezes maior que o banner.
- O anúncio desta forma também é mais longo. Em um anúncio de *banner* clássico, um *scroll down* de meros 60 pixels é suficiente para remover o anúncio da tela do usuário. O tamanho do anúncio em barras laterais é cerca de dez vezes mais longo.

Por causa deste aumento significativo no impacto dos anúncios, anúncios em barras laterais têm um poder de *branding* maior e também uma maior taxa de cliques por anúncio. Um anúncio em barra lateral típico tem uma taxa de 1% (ou dez cliques a cada mil exibições), o que o torna cerca de três vezes mais eficiente que um anúncio de *banner*, tornando também um anúncio desta natureza mais caro para anunciantes.

2.2.3 Formatos e Formas Variadas

Enquanto anúncios em banner e em barras laterais têm tamanhos padronizados, nos últimos anos, tem-se visto na internet uma proliferação enorme de diferentes tipos de marketing digital , incluindo tamanhos e locais de exibição próprios. Alguns exemplos podem ser vistos nas figuras abaixo.

Figura 2.2 – Anúncio com 250x250 pixels de área.

Na Figura 2.2, pode-se ver o exemplo de um anúncio extremamente grande, na cor laranja e na parte superior direita, que chega a ocupar cerca de 20% da área visível. Anúncios deste tamanho podem ser encontrados dentro do texto de artigos *online*. O objetivo destes anúncios é causar o mesmo efeito que anúncios em páginas de revista, que “quebram” o texto para obter mais atenção.

Figura 2.3 – Diversos anúncios diferentes na mesma página

Na Figura 2.3, é possível ver uma tira fina de anúncio da Netscape no topo, além de um anúncio em *banner* padrão, um quadro do AOL no centro direito da página, e quatro anúncios menores na parte de baixo da página.

É possível observar também algumas outras táticas interessantes utilizadas por redes de anúncios como descritas por Goodman (2008). Pode-se citar a tática de colocar mais de um anúncio da mesma marca na mesma página, com o objetivo de aumentar a probabilidade de cliques em um destes anúncios. Redes de anúncios são estudadas mais detalhadamente na seção 2.3.

2.2.4 Pop-up e Pop-Under

Um anúncio em janelas de *pop-up* é tal que aparece em uma janela não requisitada quando um usuário acessa uma determinada página. Este tipo de anúncio normalmente sobrepõe a página que o usuário está tentando ler, de maneira que o mesmo tenha que fechar a janela contendo o anúncio ou movê-la do seu caminho.

Anúncios *pop-under* são similares, mas se localizam dentro do conteúdo que o usuário está tentando ler sendo, assim, menos intrusivas. Nas figuras 2.4 e 2.5 podem-se ver exemplos de anúncios deste tipo.

Figura 2.4 – Pop-up típico

Figura 2.5 – Anúncios *pop-up* na frente da página principal

Anúncios *pop-up* e *pop-under* são irritantes para a maioria dos usuários porque enchem a área de trabalho e demandam tempo para serem fechadas. Entretanto, em termos de *marketing*, eles são muito mais eficientes que anúncios em *banner*.

Enquanto anúncios em *banner* podem ter cinco cliques em até mil exibições, um anúncio de *pop-up* pode ter até trinta cliques (BEIGHTON, 2010). Assim, anunciantes acabam aceitando pagar mais caro para anunciarem por meio de *pop-up* e *pop-under* e é esta a razão pela qual tantos anúncios desta categoria podem ser vistos na Web nos dias de hoje.

2.2.5 Anúncios Flutuantes

Estes anúncios aparecem tipicamente quando o usuário realiza o primeiro acesso a uma página Web e eles parecem “flutuar”, ou “voar”, por cima da página sem ter uma direção específica durante cinco a trinta segundos. Quando estão na tela, eles sobrepõem a visão da página que o usuário busca ler. Muitos deles são, ainda, configurados para ignorar os comandos que o usuário dá pelo *mouse*.

A Figura 2.6 mostra um exemplo de um anúncio flutuante completamente animado, com quatro partes que se movem independentemente. O anúncio é exibido por cerca de vinte segundos. Note que há um botão de “Close” neste anúncio, que fornece a possibilidade de o usuário fechá-lo – infelizmente, muitos anúncios flutuantes não possuem esta facilidade.

Muitos outros exemplos de anúncios flutuantes podem ser vistos em campanhas deste tipo de anúncio, especialmente nos sites UnitedVirtualities.com e no EyeBlaster.com.

Anúncios flutuantes têm se tornado cada vez mais freqüentes por várias razões:

- Eles definitivamente conseguem chamar a atenção dos usuários: são animados, muitos têm sons associados. Como os anúncios da TV, eles interrompem o programa e forçam os usuários a assisti-los. Além disso, podem tomar toda a tela, assim:
- Sob uma perspectiva de desenvolvimento da marca, eles são muito mais poderosos do que anúncios em *banner* ou em barras laterais. Eles simplesmente não podem ser ignorados.
- Eles acabam tendo também uma taxa de clique alta, tendo uma média de 3% (isto quer dizer que trinta pessoas irão clicar no anúncio para cada mil exibições a páginas que o contenham).

As altas taxas de cliques associadas ao maior poder de *branding* destes anúncios são naturalmente traduzidas nos preços que os anunciantes precisam pagar para ter um anúncio flutuante. E como eles podem valer muito dinheiro, os *Websites* encontram-se cada vez mais abertos a disponibilizar este tipo de anúncio em suas páginas.

O único problema de anúncios flutuantes é que eles acabam irritando as pessoas. De acordo com Beighton (2010), as taxas de reclamação deste tipo de anúncio são bem maiores do que de outros tipos, sendo esta a razão pela qual eles ainda não se proliferaram de maneira descontrolada na Internet.

Entretanto, o problema da irritação acaba indicando algo interessante sobre *marketing*. Quando os anúncios em janelas de *pop-up* apareceram pela primeira vez, eles eram considerados muito chatos e por isso não eram vistos em muitos sites.

Criaram-se até políticas de bloqueadores de *pop-up* que eram implementadas pelos navegadores disponíveis no mercado. Entretanto, após um determinado tempo, as pessoas se acostumaram com eles e pararam de reclamar, e agora este tipo de anúncio está presente em incontáveis localidades na Internet.

Figura 2.6 – Anúncio flutuante para um produto da Norton

A televisão deixa também outro exemplo útil. Se programas de televisão de hoje em dia fossem livres de anunciantes e, de repente, uma estação de TV começasse a utilizá-los, certamente as reclamações dos usuários aumentariam de maneira exacerbada. Entretanto, como todos os telespectadores já estão familiarizados com anúncios na TV, eles não se mostram verdadeiramente chateados. Um exemplo disso é que, nos Estados Unidos, durante a exibição do *Super Bowl*, os anúncios são considerados uma parte do show.

A expectativa é de que, quando as pessoas passarem a se acostumar com anúncios flutuantes, eles se tornarão mais comuns.

2.2.6 Unicast

Este tipo de anúncio traz uma mudança de perspectiva de marketing digital muito significativa e tem se tornado cada vez mais populares na Internet, com taxas de clique extremamente satisfatórias. Funcionam como um comercial de TV que roda em uma janela de *pop-up* ou *pop-under*. É animado, possui som e pode durar até mesmo minutos. O site Unicast.com traz vários exemplos desse tipo de anúncio.

Estima-se que um anúncio dessa natureza tem a capacidade de desenvolvimento de marca equiparada a de um comercial de TV. Entretanto, ele fornece uma possibilidade que a publicidade televisiva dificilmente oferece: a capacidade de clicar no anúncio para obter mais informações. Números iniciais acerca da taxas de clique nesse tipo de anúncio chegam a 5% (ou cinqüenta cliques para cada mil exibições).

Obviamente, por possuirem números tão satisfatórios em termos de taxas de clique e por terem um poder de *branding* tão significativo, a tendência é que anúncios do tipo Unicast se espalhem rapidamente pela Internet e que custem cada vez mais. Hoje em dia, paga-se até 30 dólares para cada mil exibições desse tipo de anúncio.

2.2.7 Outras Variações

Com o passar do tempo, será possível ver cada vez mais variações de marketing digital . Já podemos citar algumas destas abaixo:

- O site HowStuffWorks, em 2014 , ofereceu o que foi denominado "*takeover campaign*" (tradução livre: campanha pública de aquisição). Na primeira visita diária ao site, os visitantes viam um enorme anúncio, e a mensagem era reiterada em praticamente todo o site através de *banners* e barras laterais. Essencialmente, o anunciante "adquiriu" o site por um ou mais dias. A abordagem funcionou de maneira interessante, pois, em termos de desenvolvimento de marca, o anúncio era visível em todo o site. As taxas de clique também foram muito altas. Os anunciantes se mostraram extremamente satisfeitos com os resultados, e a reação negativa dos leitores foi mínima por causa dos níveis de familiaridade com anúncios em *banner* e em barras laterais.
- A CNN.com tem experimentado o que chama de anúncios de vídeo em barras laterais, como mostrado na Figura 2.7. Um pequeno anúncio de vídeo aparece na barra lateral, com som, e toca por trinta segundos. O visitante do site pode controlar o vídeo com botões de *play, pause* e *stop* disponíveis dele

Figura 2.7 – Anúncio de vídeo em barra lateral

- Anúncios em *banners* suspensos apareceram em alguns sites. Sua operação varia, dependendo do site. Em alguns, ao se colocar o mouse em cima do anúncio, o mesmo expande para preencher praticamente toda a página. Em outros, o *banner* é expandido por alguns segundos, e depois retorna ao tamanho normal.

Nas figuras 2.8 e 2.9 vemos um exemplo no qual o banner é exibido por alguns segundos e depois é “encolhido” para o tamanho normal. Note que o site em questão utiliza *banners* grandes, com área de 725 x 70 pixels, e a largura do site é determinada pelos anúncios. Botões no anúncio possibilitam aos usuários re-expandi-lo caso desejem.

Figura 2.8 – Anúncio suspenso exibido por alguns segundos

Figura 2.9 – Anúncio suspenso após retornar para o tamanho original

- Anúncios por *e-mail*: também chamado de *e-mail marketing* ou *e-mail advertising* (DIRECT MARKETING ASSOCIATION, 2006), anunciar por *e-mail* é uma forma de marketing direto (STONE, 2004) que usa *e-mails* como uma forma de comunicar mensagens comerciais para uma determinada audiência. No sentido mais genérico desta forma de anúncio, cada *e-mail* enviado para clientes em potencial ou para clientes atuais pode ser considerado uma forma de *e-mail advertising*. Entretanto, o termo normalmente refere-se a:

    - Envio de ***e-mails*** com o objetivo de melhorar a relação de uma empresa com seus clientes atuais ou anteriores de maneira a encorajar a relação de lealdade com o consumidor e também estimular a continuidade do negócio.

    - Envio de ***e-mails*** com o objetivo de adquirir novos clientes ou de convencer clientes atuais a comprar algo imediatamente.

    - Adição de anúncios a ***e-mails*** regulares, enviados por empresas a seus clientes.

De acordo com a DMA (2006), empresas norte-americanas investiram cerca de 400 milhões de dólares com *e-mail advertising* em 2006.

- Marketing de afiliados: esta é uma prática de marketing em que um negócio compensa um ou mais afiliados por cada novo visitante ou cliente referenciado ao negócio devido aos esforços de marketing do afiliado. Exemplos desta disposição incluem alguns serviços online de recompensa, nos quais usuários são compensados financeiramente ou com algum presente ao completarem uma ordem de serviço ou indicarem novos usuários para o serviço.

Conforme a Figura 2.10, esta indústria tem quatro participantes: a marca ou serviço, a rede de comunicação, o afiliado (também referenciado como publicador) e o cliente. De certa forma, esta categoria de marketing digital sobrepõe-se a outros métodos de marketing para Internet, uma vez que afiliados normalmente usam algum dos métodos de anúncio descritos anteriormente. Em resumo, o método de marketing de afiliados consiste em direcionar o tráfico de um *website* para outro.

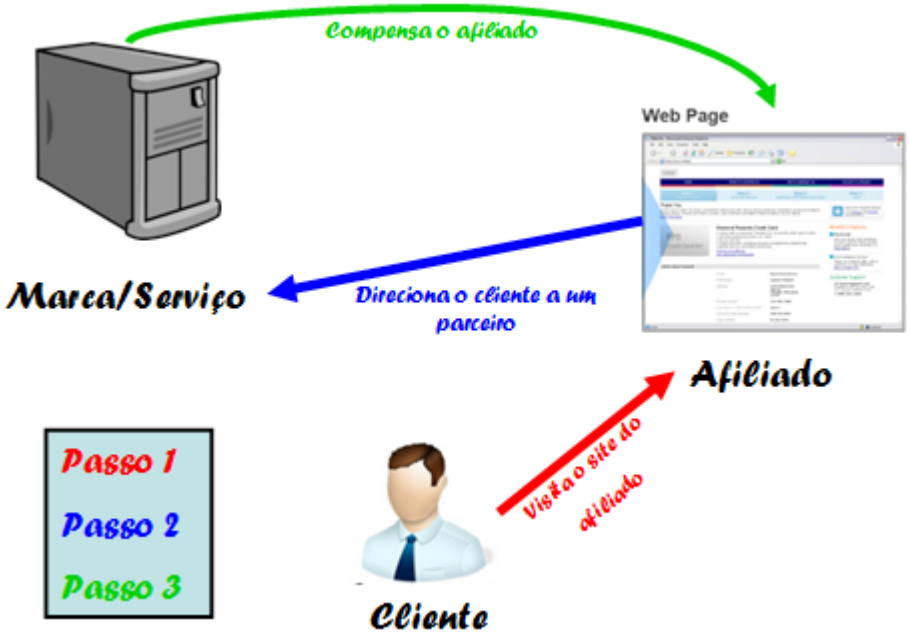

Figura 2.10 – Marketing de afiliados

- Anúncios contextuais: muitas redes mostram determinados anúncios – gráficos ou textuais – motivados por busca de conteúdo na Internet ou baseados no conteúdo exibido nas páginas onde o anúncio é mostrado. Acredita-se que esses anúncios possuam uma probabilidade maior de atrair os usuários, pois têm a tendência de compartilhar um contexto similar ao seu interesse. Por exemplo, uma busca por "flores" na Internet normalmente retorna um anúncio para o site de um florista.

Uma nova variação dessa técnica é incluir links de internet em artigos que são patrocinados por um anunciante. Quando o usuário clica no link, é enviado para o site de um dos patrocinadores.

- Anúncios semânticos: aplica tecnologias de análise semântica a soluções de marketing digital . A função dessa tecnologia é analisar semanticamente a página em exibição, para interpretar o seu significado, e garantir que ela contenha o anúncio mais apropriado. Esse tipo de anúncio aumenta a chance de que o usuário acabe clicando nele, pois o fato de estar visitando uma página com determinado conteúdo presume interesse prévio pelo mesmo.

A Web Semântica é uma área de pesquisa em contínuo crescimento na academia e mais informações sobre anúncios semânticos podem ser encontradas no site Peer 39 (http://www.peer39.com/advertisers.html).

Todos esses diversos diferentes formatos de anunciar na Internet são tentativas de se achar a combinação ideal para fornecer aos anunciantes o que eles desejam – altas taxas de clique e poder de desenvolvimento de marca. Em retorno, os anunciantes possuem consciência de que o marketing na Web funciona e, assim, propõem-se a pagar aos sites que rodam anúncios.

2.2.8 Ética

Anunciar na Web envolve um número enorme de diferentes tipos de anúncio. Infelizmente, apenas alguns deles são empregados de maneira ética. Por exemplo, pode-se comumente ver anúncios em *banners* brilhantes e coloridos que piscam e distraem os usuários, e outros que contêm imagens cujos objetivos são os de criar confusão para os usuários, e que parecem na verdade mensagens de erro do sistema operacional.

Sites de Internet que usam, de maneira não-ética, marketing *digital* para obter receita dificilmente monitoram quais anúncios estão sendo exibidos em suas páginas, e acabam permitindo alguns que levam a sites com softwares maliciosos ou de conteúdo adulto.

Em contrapartida, *Webmasters* que usam os anúncios de Internet de maneira ética normalmente exibem uma pequena quantidade de anúncios que nunca têm o objetivo de irritar ou distrair o usuário, e não alteram o design de seus *websites*. Os proprietários destes tipos de site normalmente fazem acordos com as empresas que querem exibir anúncios, aumentando os níveis de confiabilidade e legitimidade dos anúncios exibidos.

Anúncios legítimos são normalmente claros ou possuem uma forma clara de serem fechados, o que os diferencia de *spam* (JAKOBSSON; RAZMAN, 2008).

2.2.8.1 Malware

Há também uma classe de métodos de anúncio que são considerados não-éticos e até mesmo ilegais. Eles incluem aplicações externas que alteram as configurações do sistema (como a página inicial dos navegadores), abrem inúmeras janelas de *pop-up* e inserem anúncios para páginas não afiliadas. Tais aplicações são denominadas de *spyware* e *adware* (URBACH; KIBEL, 2004).

Esses tipos de aplicação mascaram suas atividades questionáveis por meio da realização de um serviço simples, como mostrar a previsão do tempo ou adicionar uma barra de busca. São programas desenhados para iludir os usuários, atuando como Cavalos de Tróia (TOWNSEND, 2003). Normalmente, remover ou desinstalar essas aplicações do sistema é difícil, e a popularização do uso dos computadores facilitou a proliferação desse tipo de programa.

2.2.8.2 Privacidade

O uso de marketing digital tem implicações naturais na privacidade dos usuários e em sua capacidade de se manterem anônimos. Por exemplo, se uma empresa de anúncios publica *banners* em dois sites diferentes, por meio da administração das imagens que estão armazenadas nos servidores da empresa e por meio do uso de *cookies* (DOYLE, 2003), essa empresa pode mapear estatísticas de navegação dos usuários dos dois sites.

Cookies podem ser bloqueados pela maioria dos navegadores para aumentar a privacidade e diminuir a capacidade de mapeamento de perfil sem afetar negativamente a experiência de navegação dos usuários. Além disso, muitas redes de anúncios fornecem a opção de desabilitar o anúncio baseado em perfis de navegação.

2.3 Redes de Anúncios

Uma rede de marketing digital (ou simplesmente rede de anúncios) é uma empresa que conecta anunciantes para sites de Internet que querem publicar anúncios em suas páginas. Por conseqüência, as redes de anúncios pagam aos sites quando os usuários destes usam os softwares da rede de anúncios localizados nos mesmos.

O mercado das redes de anúncios cresce de maneira exponencial. De acordo com Kahn et al. (2007), que apresenta números de 2007, as 20 principais empresas deste mercado lucraram nada menos que dois bilhões de dólares no ano. Isto representa aproximadamente 13% do mercado total de anúncios, cuja previsão de crescimento beira 18% em 2010.

Tal crescimento resultou em diversos novos participantes desse mercado e

encorajou aquisições de redes de anúncios por grandes empresas entrantes neste mercado. Por exemplo, em 2007, o Google comprou uma rede de anúncios denominada DoubleClick por 3,1 bilhões de dólares.

O principal negócio das redes de anúncios é vender espaço para que um subconjunto do seu inventário completo de marketing digital apareça. Esse subconjunto pode ser visualizado de formas diferentes, incluindo espaço em *websites*, em RSS *feeds* (RAMASUBRAMANIAN, 2005), em blogs, em aplicações de mensagem instantânea, em adware, *e-mails* e em outras fontes.

A forma dominante de exibição de anúncios ainda é *websites* de terceiros, que trabalham com redes de anúncios por uma taxa, ou recebem uma porcentagem em cima das receitas obtidas pelo anúncio.

A rede roda um serviço de anúncios a partir de seus servidores internos. Cada anúncio está associado a um determinado site na Internet. Um trecho de código é invocado do servidor de anúncios pelo site que publica o anúncio, e a resposta a essa requisição é, normalmente, um *banner* contendo um ou mais anúncios (Figura 2.11).

Figura 2.11 – Destaque para um banner de rede de anúncios em um site

Grandes sites de internet que publicam anúncios vendem apenas parte do seu inventário de anúncios através dessas redes. Já sites menores normalmente associam todo o seu inventário através delas.

2.3.1 Tipos de Redes de Anúncios

Há basicamente duas formas de se classificar redes de marketing digital . A primeira classificação, que diz respeito à distribuição dos anúncios, separa as redes em três tipos:

1. **Redes verticais**: redes que atuam como representantes dos locais onde o seu portfólio de anúncios será publicado. Neste caso, há transparência total do anunciante a respeito dos locais onde seus anúncios irão aparecer. Essas redes visam a qualidade

do serviço, oferecendo tráfego diferencial de dados a preços acessíveis, e são usadas extensivamente por empresas interessadas em *branding*. Seu modelo de negócio é normalmente baseado no compartilhamento pré-acordado das receitas obtidas com os anúncios.

2. **Redes cegas (*blind networks*)**: estas empresas oferecem bons preços para empresas interessadas em vendas diretas que não necessitam de ter controle do local onde os seus anúncios irão aparecer. Redes cegas são bem mais baratas do que os outros tipos de redes de anúncios e conseguem isso através da compra no atacado de locais de anúncio combinadas com a utilização de tecnologia de orientação de anúncios e otimização de conversão (KING, 2008). Desta forma, o modelo de negócios desse tipo de rede não segue um padrão bem definido (DE JONG; ROSENTHAL; VAN DIJK, 2009).

3. **Redes orientadas**: consideradas a nova geração de redes de anúncio, estas redes focam em tecnologias de orientação e adequação de buscas, tais como análise contextual de navegação e comportamental de usuários. Redes orientadas são especializadas em utilizar dados do consumidor para aumentar o valor do inventário de anúncios aos quais eles serão submetidos.

A segunda forma de classificação redes de marketing digital diz respeito à proximidade do relacionamento comercial da rede com os anunciantes, e são divididas em redes de primeiro nível e redes de segundo nível.

As redes de primeiro nível possuem uma quantidade maior de anunciantes cadastrados como seus clientes diretos e possibilitam tráfego de rede de alta qualidade, servindo anúncios e tráfego de rede para grandes portais de internet e para redes de anúncios de segundo nível. Exemplos de redes de anúncios de primeiro nível são as grandes ferramentas de busca online.

Em contrapartida, redes de anúncios de segundo nível podem até ter alguns clientes cadastrados como anunciantes diretos, mas sua principal fonte de receita vem do repasse de anúncios de outras redes.

2.3.2 Servidores de Anúncios

As empresas que desenvolvem tecnologias para veiculação de marketing digital fornecem software para que redes possam servir os anúncios, realizar contabilidade, escolher que anúncios devem ser exibidos (e que vão gerar mais receitas para os anunciantes e publicadores) e monitorar o progresso de diferentes campanhas de anúncio.

Um servidor de anúncios é um servidor de computadores, especificamente um servidor web (TEIXEIRA, 2004), que armazena anúncios que serão exibidos em sites de Internet como parte do mecanismo de anunciar na Internet. O conteúdo do servidor de anúncios precisa ser atualizado constantemente, de maneira que o site nos quais os anúncios são exibidos tenha a capacidade de absorver novos anúncios quando a

página é visitada ou atualizada por um usuário.

Além disso, o servidor de anúncios também realiza diversas outras atividades como contar o número de cliques (ou de exibições) para uma campanha de marketing e a geração de relatórios, que é útil na determinação do retorno sobre investimento (também conhecido como ROI, *return over investment*) (KASSAI, 1996) para um anunciante de um serviço particular.

O primeiro servidor de anúncios foi desenvolvido pela empresa FocaLink Media Services. Introduzido em 17 de Julho de 1995, o servidor controlava marketing digital em *banners*. A empresa, com sede em Palo Alto, Califórnia, foi fundada por Dave Zinman e Jason Strober. Em 1998, a empresa foi renomeada para AdKnowledge e foi adquirida pela empresa CMGI em 1999.

Quanto à localização, servidores de anúncio podem ser classificados em duas categorias: servidores de anúncio locais e servidores de anúncios remotos (de uma terceira parte). Servidores locais são tipicamente utilizados por publicadores que querem servir anúncios em seus próprios domínios, permitindo uma customização maior para beneficiar seus anunciantes e também uma maior taxa de controle de conteúdo.

Servidores remotos podem servir anúncios em diversos domínios diferentes, que são propriedades de diferentes publicadores. Estes servidores entregam anúncios de uma fonte central, de modo que os anunciantes e os publicadores possam acompanhar a distribuição de seus anúncios dessa fonte, e possuam também uma localidade para controlar a rotação e a distribuição de seus anúncios em toda a Web.

Não nos deteremos mais neste assunto, pois o estudo aprofundado da tecnologia e serviços envolvidos na disponibilização de anúncios em sites de Internet não fazem parte do escopo deste livro, conforme descrito na seção 1. Mais detalhes sobre estes tópicos podem ser encontrados na matéria de veiculação de anúncios (ANUPAM et al, 1999), (JAKOBSSON; MACKENZIE; STERN, 1999).

2.3.3 Funcionalidades Existentes em Servidores de Anúncios

As funcionalidades típicas de servidores de anúncio incluem:

- Envio de novos anúncios
- Distribuir anúncios de acordo com regras de negócio pré-definidas.
- Orientar anúncios por usuário ou por conteúdo.
- Atualizar configurações de otimização baseadas em resultados.
- Relatar exibições, cliques, atividades realizadas após exibições e após cliques, e métricas de interação.

Pode-se também citar algumas funcionalidades avançadas:

- Freqüência de anúncios: configurações acerca do número e da duração dos anúncios recebidos pelos usuários (anunciantes podem também limitar a

freqüência dos anúncios baseados na incorrência de gastos).

- Seqüenciamento de anúncios, de maneira que os usuários recebam as mensagens em uma ordem específica.
- Exclusão da concorrência, de maneira que os usuários não vejam os anúncios dos concorrentes um após o outro.
- Anúncios que são mostrados de forma a que um anunciante tenha exclusividade em uma determinada página.
- Anúncios orientados, exibidos de acordo com o comportamento anterior dos usuários.

2.3.4 Métodos para Orientação de Anúncios e Otimização

Um aspecto da tecnologia de veiculação de anúncios é a utilização de métodos automáticos e semi-automáticos para otimizar os preços dos anúncios, sua localização e outras características. Podemos citar os métodos a seguir:

- Orientação comportamental de anúncios: consiste em usar um perfil de comportamento anterior do usuário para determinar qual anúncio deve ser exibido durante uma dada visita. Por exemplo, a exibição de anúncios de venda de carros em um portal de notícias para um usuário que tenha anteriormente visitado a seção automotiva de um site.
- Orientação contextual de anúncios: consiste em inferir o melhor local onde se deve colocar um determinado anúncio a partir da informação contida na página onde o anúncio será exibido. Por exemplo, exibir um anúncio de bicicletas em uma página que contenha um artigo sobre ciclismo.
- Orientação por geolocalização: consiste na utilização de técnicas para identificar a localização real dos usuários de um publicador e, a partir dessa informação, exibir anúncios que sejam interessantes para tal região. Um exemplo deste tipo de anúncio é mostrado na Figura 2.12, quando a busca textual no Google por “Personal Injury Lawyers” (tradução livre: advogados especializados em danos físicos), realizada da cidade de Dallas, Texas, retorna um número de escritórios de advocacia daquela região.

Existe uma boa quantidade de empresas especializadas em geolocalização na Internet e uma lista interessante pode ser encontrada em Smith (2007). Na verdade, através de uma consulta ao *American Registry for Internet Numbers* (ARIN) (SMITH, 2007), qualquer usuário de Internet pode obter informações (cidade, estado, país e, até mesmo, CEP), associadas no momento a um determinado IP.

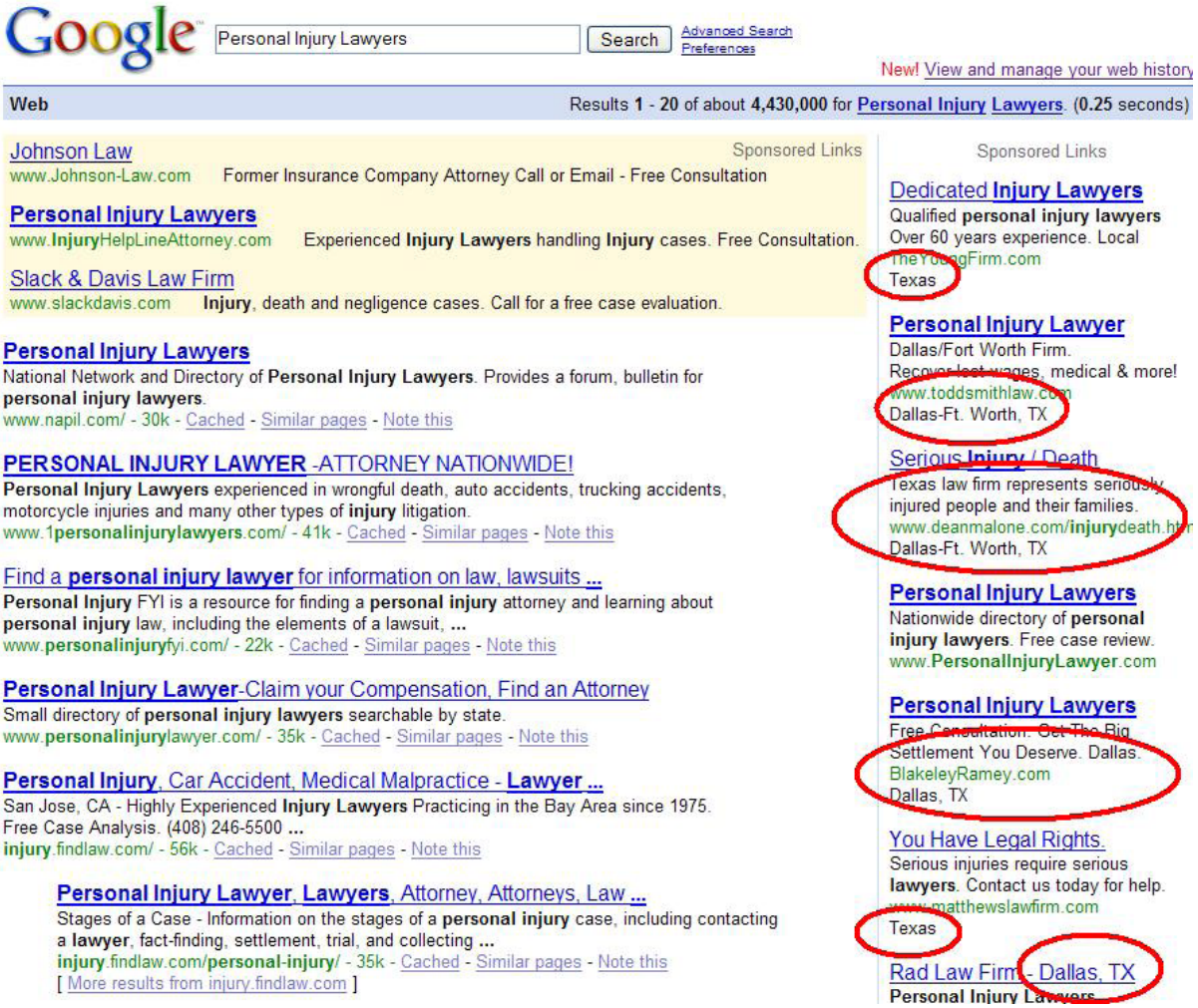

Figura 2.12 – Anúncios orientados por geolocalização

- Otimização criativa: consiste em usar métodos experimentais e preditivos para explorar a melhor forma de se exibir um dado anúncio, e explorar o que foi determinado em exibições posteriores.

2.4 Modelos para Obtenção de Receita em Marketing digital

Existem diversos modelos de mercado através dos quais as empresas participantes do negócio de marketing digital se organizam para obter receitas e estruturar o negócio. Os mais comuns deles são o Pagamento por Clique, o Pagamento por Exibição e o Pagamento por Ação, mas existem ainda outros modelos usados com menos frequência. Vale salientar que esses modelos têm evoluído ao longo dos anos e a lista que aqui apresentamos está sujeita constantes a atualizações.

2.4.1 Pagamento por Clique (Pay per Click ou Cost per Click)

Também conhecido como *PPC advertising*, o pagamento por clique é, em síntese, um acordo entre empresas. O primeiro grupo de empresas, os publicadores, exibe em seus sites *links* clicáveis, que são anúncios do segundo grupo, os anunciantes. Em troca, há uma cobrança por cada clique.

Com o crescimento desta indústria, desenvolveu-se uma terceira entidade, denominada rede de anúncios, que passou a agir como mediadora entre os dois grupos (publicadores e anunciantes), e que cresceu de maneira muito rápida. Neste novo modelo, sempre que um usuário web válido clica em um anúncio, o anunciante paga um valor à rede de anúncios, que repassaria uma parte deste pagamento ao publicador.

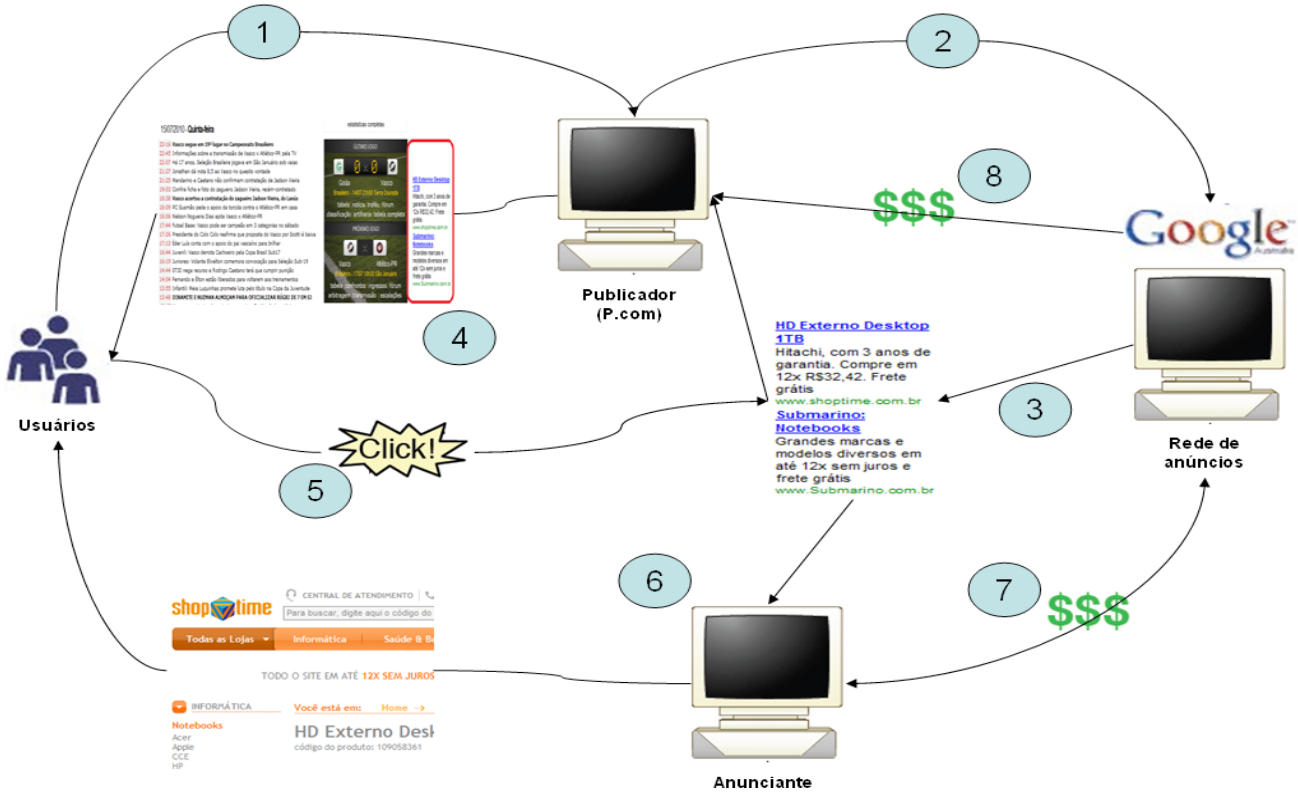

Figura 2.13 – Sequência de ações de uma solicitação PPC clássica

Conforme esquematizado na Figura 2.13, sempre que um usuário da Internet visita a página de um publicador, esse usuário é associado a um dos servidores da rede de anúncios. O servidor escolhe um conjunto de anúncios e o repassa ao publicador, que exibirá os anúncios juntamente com o conteúdo da página no navegador do usuário.

Se o usuário clicar em um anúncio no site do publicador, esta ação será associada ao servidor do representante (que exibe o anúncio na página do publicador), que salva o clique em um histórico, para cobrança posterior, e direciona o usuário à página do anunciante cujo anúncio foi clicado. Esse esquema pode ser detalhado da seguinte forma:

1) Um usuário comum de Internet acessa o portal de um publicador de anúncios;

2) O publicador possui na página a ser exibida uma requisição por anúncios, que é feita à rede de anúncios devida;

3) A rede de anúncios retorna o conjunto de anúncios para o publicador;

4) O publicador exibe o conteúdo de seu site contendo o conjunto de anúncios retornado pela rede;

5) O usuário clica em um desses anúncios;

6) O banner contém links que redirecionam o usuário ao site de um dos anúnciantes;

7) Ao verificar o clique, a rede de anúncios registra a cobrança junto ao anunciante;

8) Parte da verba arrecadada junto ao anunciante é repassada para o publicador.

O sistema de compartilhamento de divisas entre a rede de anúncios e os publicadores é, naturalmente, visto como um incentivo para a *click fraud*. A razão para isto é que as redes de anúncios têm o incentivo de agir em um papel duplo, uma vez que elas próprias também são publicadores, através de suas ferramentas de busca. De acordo com críticos, essa relação complexa pode acabar gerando um conflito de interesses: ao mesmo tempo em que a rede de anúncios perde dinheiro quando não consegue detectar *click fraud*, pois tem que pagar o publicador, ela também ganha dinheiro ao realizar as cobranças originárias de cliques fraudulentos juntamente ao anunciante.

Devido ao fato de as taxas pagas pelos anunciantes serem compartilhadas entre a rede de relacionamentos e o publicador (afinal, a rede de anúncios só repassa ao publicador uma parte do que ganha), a rede de anúncios acaba ganhando dinheiro com a *click fraud*.

Um anunciante de Internet (por exemplo, o eBay) provê os seus anúncios a um representante (por exemplo, o Google Adsense), reserva uma determinada quantia de dinheiro e se compromete a pagar uma comissão por determinadas ações dos usuários do serviço do representante, como, por exemplo, clicar em um anúncio, realizar uma compra ou dar um lance em um leilão.

Os publicadores de Internet (por exemplo, o MySpace.com), motivados pela comissão paga pelos anunciante*s*, procuram os representantes com o objetivo de estabelecerem com estes contratos para exibir os anúncios nas suas páginas Web e obterem, assim, uma parte da comissão. O ponto principal desta relação são os representantes, que atuam como mediadores entre os publicadores e os anunciante*s*.

2.4.2 Pagamento por Exibição (Cost per Impression ou Cost per Mille)

O pagamento por exibição acontece quando anunciantes pagam pela exposição de seus anúncios para uma audiência específica. É também chamado de pagamento por milha (*cost per mille*), porque o pagamento da quantidade de anúncios é normalmente medida em milhares de exibições. Sendo assim, quando um publicador disser “nosso CPM é de $5”, está querendo dizer que o custo por exibição é de $0,005.

Uma exibição é uma aparição simples de um anúncio em uma página web. Cada vez que um anúncio é carregado na tela de um usuário, o servidor de anúncios conta este carregamento como uma exibição. Entretanto, o servidor de anúncios normalmente é configurado para ignorar certas categorias de atividade, como recarregamento da

página, ações de usuários internos e outros eventos que a rede de anúncios concordou em não contar.

O pagamento por exibição é a opção preferida pelas redes de anúncios e pelos publicadores, pois a taxa de pagamento é proporcional à quantidade de tráfego de rede realizado. Desta forma, hoje em dia, é muito comum que grandes publicadores realizem esse tipo de cobrança no seu inventário de anúncios.

2.4.3 Pagamento por Ação (Cost per Action ou Cost per Acquisition)

O pagamento por ação é baseado em desempenho e é comumente usado no modelo de anúncios denominado marketing de afiliados. Nesse esquema de pagamento, os publicadores assumem todos os riscos por trás da exibição do anúncio, de modo que o anunciante pague somente pelas ações dos usuários que completem uma transação, como uma inscrição em serviço ou uma compra. Esta é a melhor forma de pagamento por anúncios em *banner* e o pior tipo de relação de pagamento para as redes de anúncios e para os publicadores.

2.4.4 Outros Modelos

Existem ainda outros modelos nos quais marketing digital são adquiridos. Dentre os quais, podemos citar:

- *Custo por prospecção (cost per lead)* – pode ser identificado como uma forma de pagamento por ação. O pagamento é baseado no preenchimento de um formulário online, no registro para receber uma *newsletter* ou na execução de qualquer ação identificada pelo anunciante como tal que torna o usuário um prospecto para uma possível venda futura.
- *Custo por ordem (cost per order)* – pagamento baseado em cada vez que uma ordem de compra ou de serviço é realizada.
- *Custo por interação (cost per engagement)* – as exibições dos anúncios são gratuitas e que os anunciantes pagam apenas quando um usuário interage com um anúncio específico. Também é uma forma de custo por ação.
- *Custo por conversão (cost per conversion)* – descreve o custo de adquirir um cliente, tipicamente calculado pela divisão do custo total de uma campanha de marketing pelo número de conversões. A definição de “conversão” varia, dependendo da situação: pode ser considerada uma prospecção, venda ou aquisição.

A diferença essencial entre os modelos de pagamento acima e o pagamento por ação tradicional é que, em campanhas do tipo pagamento por ação, o anunciante paga apenas quando uma venda é efetivamente processada (normalmente por meio de pagamento por cartão de crédito).

2.4.5 Divisão do Mercado de Redes de Anúncios

Abaixo podemos ver a lista das principais empresas que vendem anúncios (publicadores e redes de anúncio) na Internet em 2008. Os números estão em milhões de usuários que visualizaram anúncios. É importante frisar que o Google adquiriu a DoubleClick em 2007 por US$ 3,1 bilhões. A pesquisa abaixo, realizada pela Browser Media (BROWSER MEDIA, 2008), baseou-se em uma amostra de 68 milhões de domínios. Desde 2008, o Google controla cerca de 69% do mercado de marketing digital :

Empresa	Visualizações (em milhões)
Google	1,118
DoubleClick (Google)	1,079
Yahoo!	362
MSN (Microsoft)	309
AOL	156
Adbrite	73
Total	**3,087**

Tabela 2.1 – Divisão de mercado das principais redes de marketing digital

2.4.5.1 A Barra de Ferramentas Alexa

A Alexa Internet Inc. é uma empresa subsidiária da Amazon.com conhecida pela sua barra de ferramentas em navegadores. Uma vez instalada, essa barra de ferramentas coleta dados acerca do comportamento de navegação do usuário, que são transmitidos aos sites vão sendo visitados por ele. A barra de ferramentas também oferece ao usuário sugestões sobre que site visitar em seguida, baseadas nos padrões de navegação de sua comunidade de usuários. Essa barra de ferramentas oferece, ainda, o contexto de cada site visitado: quem é o proprietário do site, quantas páginas ele possui, quantos outros sites na rede apontam para ele e quão freqüentemente ele é atualizado.

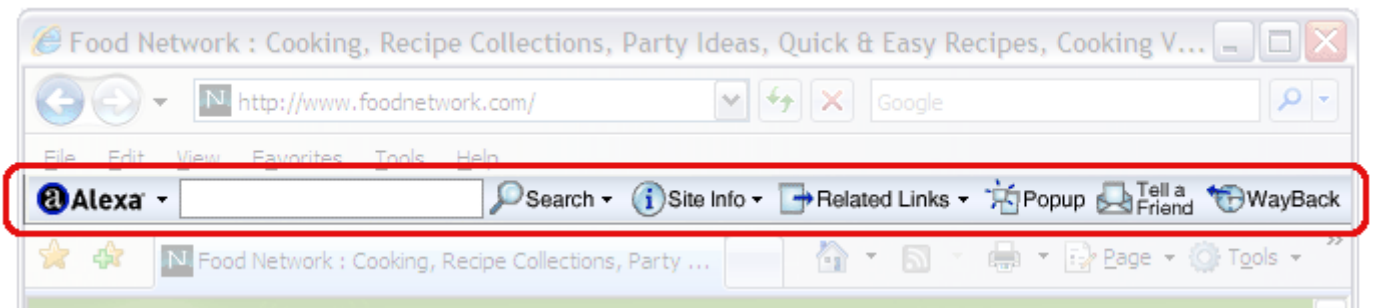

Figura 2.14 – A barra de ferramentas Alexa

A relação entre a barra de ferramentas Alexa e a sistemática de redes de marketing digital está no fato de a primeira manter um ranking de classificação dessas redes, baseado em informações de acompanhamento dos usuários da barra de ferramentas. Existem diversas controvérsias acerca de quão representativa de fato é a base de usuários da Alexa em termos de comportamento de navegação da Internet como um todo (CROLL; POWER, 2009), especialmente ao se levar em consideração sites que não são tão populares. Michael Arrington (2007) mostrou alguns exemplos de contradição em alguns dados publicados no ranking de Alexa, incluindo uma edição do ranking acerca da importância e tamanho dos sites, que listou o Youtube na frente do Google.

Mesmo com questionamentos relevantes como esses, o ranking da Alexa ainda é considerado como referência global para o posicionamento e análise do mercado de marketing digital . A partir do ranking global e do estado de operação da Alexa, podemos destacar na tabela 2.2 uma lista de redes de anúncios notáveis, ordenadas por nome:

Nome	Localidade	Ano fundação	Alexa global page rank	Status
AdBrite	San Francisco	2002	896	Active
AdChina	China	2007	66.070	Active
AdWords	Michigan	2000	1 (Google.com)	Active
Azoogle	New York	2000	3.272	Active
BidClix	New York	2001	N/A	Defunct
Bidvertiser		2003	1.922	Active
Blogads	North Carolina	2003	15.575	Active
BlueLithium	San Jose, California	2004	7.285	Unknown
Casale Media	Toronto	2003	8.651	Active
Chitika	Westborough, Massachusetts	2003	3.043	Active
Clickbooth	Sarasota, Florida	2002	3.317	Active
ClickDiario Network	Guatemala	2000	797.875	Active
DoubleClick	Chicago	1998	48	Active
Hiro-Media	Tel Aviv, Israel	2004	1.174.163	Active
Hydra Network	Beverly Hills, California	2003	2.256	Active
Infolinks	Mountain View, California	2007	845	Active
Kontera	San Francisco	2003	3.860	Active
Neverblue	Victoria, British Columbia	2004	3.014	Active
Nuffnang	Singapore, Malaysia, Philippines, Austria	2007	20.712 (Nuffnang MY)	Active
RealTechNetwork	Congers, New York	2004	206.594	Active
Right Media	New York	2004	45.234	Active
Tribal Fusion	Emeryville, California	2003	343	Active
ValueClick	Westlake Village, California	1998	24.286	Active
VoloMedia	Sunnyvale, California	2005	862.577	Active
Yahoo! Publicador Network	Sunnyvale, California	2005	4 (Yahoo.com)	Active
Zedo	San Francisco	1999	139	Active

Tabela 2.2 – Lista de redes de marketing digital notáveis, segundo Alexa

2.5 Fraudes em Marketing digital

Uma rede de marketing digital pode sofrer abusos de diferentes formas, e algumas delas são descritas nesta seção. Cada um dos modelos de receita descritos na seção 2.4 (pagamento por clique, por exibição e por ação) está sujeito a fraudes correspondentes. Nesta seção, algumas fraudes que podem ocorrer em marketing digital serão discutidas.

Fraudes online têm como principal conseqüência a redução do ROI dos anunciantes, embora existam muitas outras razões pelas quais o ROI de um anunciante pode ser afetado – anúncios de baixa qualidade, altas taxas de latência no site do anunciante, preços pouco competitivos, e muitos outros fatores.

O foco desta seção não é discutir esses fatores, mas analisar as fraudes online como uma das muitas variáveis que podem ocasionar prejuízos a anunciantes. Desta forma, nesta seção, uma taxonomia abreviada dos tipos diferentes de fraude que podem impactar uma rede de marketing digital é apresentada.

De acordo com Jakobsson e Razman (2008), *spam* pode ser definido como um artefato que não produz nenhum valor, utilidade ou benefício conforme esperado por um ou mais participantes associados ao artefato. Desta forma, um *e-mail* é considerado spam quando o receptor não consegue derivar nenhum valor apreciável, útil ou benéfico do *e-mail*. Sob esta perspectiva, podemos classificar cada uma das fraudes em marketing digital como formas diferentes de spam. Nas seções subseqüentes, vamos discutir as seguintes formas de fraude: fraude de exibição (Seção 2.5.1), fraude de conversão (Seção 2.5.2) e *click fraud* (Seção 2.5.3).

2.5.1 Fraude de Exibição (Spam de Exibição)

A fraude de exibição resulta de requisições HTTP para páginas na Web que um usuário nunca chega a ver e assim não produz nenhum valor significativo para o dinheiro que os anunciantes estão pagando pela exibição. A fraude de exibição afeta campanhas de anúncio baseadas em pagamento por exibição, pois, na prática, os anunciantes não deveriam ser cobrados por tais requisições HTTP.

Mais adiante, conforme explorado por Immorlica et al. (2005), a fraude de exibição afeta cálculos do tipo *clickthrough rate* (CTR), também conhecida como “taxa de cliques por impressões”, que é uma forma medir o sucesso de uma campanha de marketing digital pela razão entre o número de usuários que clicou em um anúncio em uma página web e o número de vezes que o anúncio foi exibido (também conhecido como o número de impressões), uma vez que o denominador comum no cálculo do valor é o número de exibições da página.

Como o ranking entre os diferentes anúncios de uma rede pode depender do CTR, a manipulação desta taxa pode ser de interesse de um anunciante malicioso

como um método indireto para alterar a saída de um leilão de anúncios.

Uma conseqüência negativa para os fraudadores é que o aumento anormal do CTR de determinados anúncios pode ser uma evidência de fraude. Para dificultar a detecção, fraudadores experientes fazem uso de técnicas que evidenciam a fraude em anúncios de concorrentes, para assim manipular o CTR da campanha e realizar a fraude em seus anúncios de maneira mascarada.

Uma vez que as requisições de HTTP classificadas como fraude sejam filtradas em um histórico no servidor web, o número que deve ser considerado para cálculos de CTR e para a contabilidade da campanha deve ser o número total de visualizações de página menos o número daqueles que provavelmente são classificados como fraude de exibição

2.5.2 Fraude de Conversão

Uma conversão ocorre quando uma requisição HTTP de uma URL pré-determinada pelo anunciante é submetida. Por exemplo, a requisição da URL pode significar uma visualização de página, a submissão de um formulário, o início do download de algum arquivo ou até mesmo a finalização de uma venda online.

A fraude de conversão ocorre quando requisições HTTP são submetidas com o objetivo de produzir conversões que não são geradas por usuários reais. Da mesma forma como há um incentivo para a fraude de exibição em anúncios do tipo CPM, existe um incentivo para a fraude de conversão em anúncios do tipo pagamento por clique (PPC), de maneira a inchar a quantidade de cliques relevantes.

Depois de clicar em um anúncio e ser redirecionado para o site do anunciante, um usuário mal-intencionado pode baixar um arquivo (como uma versão de avaliação de um pacote de software) deste site em uma tentativa de simular o comportamento de um usuário legítimo.

Anúncios cujo modelo de receita é baseado em ação também são suscetíveis à fraude de conversão, pois publicadores fraudulentos podem tentar obter receita a partir de ações produzidas artificialmente.

Por exemplo, se um publicador mal-intencionado recebe parte da receita de uma campanha de anúncios com pagamento baseado em ação que contabiliza os downloads de um pacote de avaliação de software, tal publicador pode iniciar downloads depois de clicar em anúncios como uma tentativa de obter receita sem o objetivo real de testar o software.

2.5.3 Click Fraud (Click Spam)

A *click fraud* ocorre quando requisições HTTP pelos endereços dos anúncios em exibição são realizadas sem intenção legítima por diferentes fontes de usuários considerados legítimos até o momento da detecção da fraude. Tais requisições devem

ser consideradas inválidas pelas redes de anúncios, de maneira que podemos definir *cliques inválidos* como quaisquer cliques pelos quais uma rede de anúncios escolha não cobrar.

Quando cliques são classificados como inválidos, o usuário que fez a requisição ainda é redirecionado para o site do anunciante, mas este não é cobrado pelo clique.

Por sua vez, um *clique fraudulento* é tal que foi realizado com intenção maliciosa, e a *click fraud* pode ser simplesmente definida como a prática de realizar cliques fraudulentos.

Como a intenção por trás de um clique em uma campanha de marketing digital é algo pessoal, que só pode ser defendida pelo usuário realizando o clique – ou pelo autor do software que realiza cliques de maneira automática –, nos sistemas de redes de anúncios atuais, é impossível ter certeza absoluta de que qualquer clique seja é fraudulento ou não (TUZHILIN, 2006).

Desta forma, os sistemas atuais de redes de anúncios buscam armazenar sinais e evidências que serão posteriormente avaliadas, automaticamente ou não, caso haja suspeita de uma ação maliciosa. Assim, cliques suspeitos podem acabar sendo marcados como inválidos após esta análise, mas não há uma garantia probabilística acerca da completude do número de cliques fraudulentos que serão detectados e classificados como inválidos.

Quando cliques suspeitos são marcados como inválidos, o usuário que executou o clique ainda é redirecionado ao site do anunciante. Esta abordagem de prosseguir com o redirecionamento oferece dois benefícios:

1. Um fraudador não recebe nenhum retorno se foi detectado como tal;
2. Se um clique suspeito na verdade for legítimo (ou seja, a avaliação retornou um falso positivo), então a experiência do usuário com a rede de anúncios e com o anúncio em si não é impactada negativamente.

Como se pode imaginar, um número alto de falsos positivos é algo extremamente ruim para o publicador, uma vez que os falsos positivos podem acarretar receitas para os anunciantes. Assim, é de vital importância que uma rede de anúncios realize todos os esforços possíveis para minimizar o número de falsos positivos e assim balancear o aumento do ROI dos anunciantes com a prospecção de novos publicadores e relacionamento de qualidade com eles.

É importante notar que nem todos os cliques marcados como inválidos são necessariamente fraudulentos. Cliques podem ser marcados inválidos pelo simples interesse legítimo de uma rede de anúncios em aumentar o ROI de seus anunciantes.

Por exemplo, muitos cliques são marcados como inválidos devido a ações de usuário – como cliques duplos ocasionados por um usuário que clicou duas vezes no mesmo anúncio – ou por outras razões técnicas.

Duas fontes de cliques inválidos que ocorrem por intenção maliciosa são os cliques originados por concorrentes de anunciantes e os cliques de publicadores desonestos.

Uma vez que os publicadores lucram com os eventos de clique nos anúncios exibidos em seus sites, é possível observar um incentivo para que publicadores desonestos aumentem o número de cliques que seus sites geram (ANUPAM et al, 1999), (JAKOBSSON; MACKENZIE; STERN, 1999), de maneira a gerar receita através do pagamento por clique.

Além disso, concorrentes de anunciantes também acabam se sentindo incentivados a simularem cliques nos anúncios de seus concorrentes com o objetivo de esgotar o orçamento dos departamentos de marketing destes últimos (MANN, 2006).

Click fraud resulta em má reputação para os representantes, e existem diversos casos de pagamento de multas para anunciante*s* (LIEDTKE, 2006). *Click fraud* coloca em risco toda a indústria de anúncios pela Internet.

A *click fraud* tem sido uma preocupação para representantes de anúncios desde a sua concepção (ZELLER JR, 2004). Os números envolvidos em *click fraud* são difíceis de quantificar; existem diversas formas de se estimar a proporção de cliques falsos, que variam de 10% a 50%.

Um estudo largamente citado da MarketingExperiments.com, uma ferramenta de pesquisa sobre marketing online, relatou que 29,5% dos cliques em três campanhas experimentais do Google eram fraudulentos. Mesmo com números tão expressivos, as empresas de busca e muitos dos seus clientes vêm argumentando que o problema em suas redes está sob controle.

Entretanto, alguns observadores do mercado de cliques *online*, como a Holcomb, acreditam que a *click fraud* traz prejuízos da ordem de bilhões de dólares e, como dito anteriormente, possuem potencial para destruir a indústria inteira.

Independentemente do número exato, a *click fraud*, hoje, está impregnada no negócio de anúncios pela Internet, e, muito embora as ferramentas de busca procurem se defender de diferentes maneiras, os fraudadores tornam-se cada vez mais sofisticados e os programas utilizados para automatização da fraude são cada vez mais complexos, disfarçando, inclusive, a origem dos cliques.

2.5.3.1 Click Fraud por Partes Não Contratadas

Uma forma alternativa de realização da *click fraud* é através de partes não contratadas, que não se envolveram de nenhuma forma com o acordo de *pagamento por clique*. Definir regras para este tipo de fraude é ainda mais difícil, pois o criminoso, na maioria das vezes, não pode ser processado por quebra de contrato nem criminalmente acusado de fraude. Exemplos de partes não contratadas são:

- Concorrentes de anunciantes: conforme já descrito anteriormente, essas partes podem querer prejudicar um concorrente do mesmo mercado, ao clicar nos seus anúncios. Os criminosos, nesses casos, não lucram diretamente, mas forçam seus concorrentes a pagar por cliques irrelevantes com o objetivo de enfraquecê-los ou eliminá-los economicamente.

- Concorrentes de publicadores: essas partes podem querer difamar um publicador. Esta fraude é feita de maneira a parecer que o publicador está clicando nos seus próprios anúncios – e a rede de anúncios pode querer terminar a relação. Muitos publicadores têm apenas os anúncios como fonte de renda, e um ataque como esse pode simplesmente retirar uma empresa do mercado.

- Outras partes maliciosas: além de vandalismo, existe uma infinidade de razões para se prejudicar tanto um anunciante como um publicador, mesmo que o criminoso não tenha nada a ganhar com o ato. Os motivos podem variar de vinganças políticas a pessoais. Estes casos são, normalmente, os mais difíceis de lidar, uma vez que é complicado rastrear o acusado e, caso encontrado, existem poucas ações, legalmente falando, que possam ser tomadas contra ele.

- "Amigos" dos publicadores: já aprendemos que um publicador ganha dinheiro quando os anúncios são clicados do seu *site*. Então, um "amigo" do publicador, como um fã, familiar, ou amigo pessoal acaba clicando em diversos anúncios para "ajudar". Entretanto, isto pode acabar sendo ruim para o publicador e o mesmo (ao invés do "amigo") pode acabar sendo acusado de *click fraud*.

As redes de anúncios tentam combater todas as fraudes possíveis, mas às vezes é impossível saber quais cliques são legítimos. Além disso, com a exceção das fraudes cometidas pelos publicadores, é difícil definir quem deve pagar quando uma fraude cometida no passado é descoberta.

Em termos de relação comercial e credibilidade de mercado, é muito prejudicial para os publicadores ter que reembolsar cliques cujas fraudes, muitas vezes, não são culpa sua. Entretanto, os anunciantes são, obviamente, irredutíveis quando o assunto é (não) pagar por cliques falsos.

2.5.3.2 Organização

A forma mais simples de *click fraud* é quando um usuário, ao iniciar um pequeno empreendimento na web, se torna um publicador de anúncios e passa a clicar os endereços que aparecem em seu site para gerar receita. Normalmente, o número de cliques e o seu valor são tão pequenos que a fraude não é detectada.

Muitas vezes, os publicadores argumentam que algumas quantidades de tais cliques foram acidentais e, algumas vezes, isso é verdade. Entretanto, também acontece fraude em larga escala.

Os envolvidos em *click fraud* em larga escala normalmente rodam *scripts* para simular que um usuário humano está clicando nos anúncios. Obviamente, uma quantidade enorme de cliques originados de um único computador ou de um pequeno grupo de computadores, ou ainda de uma única região geográfica, parecerá extremamente suspeito para a rede de anúncios e para os anunciantes.

Cliques que possuem origem em um computador que reconhecidamente pertence ao publicador também parecerão suspeitos àqueles que vigiam contra *click*

fraud. Desta forma, uma pessoa que tentar realizar fraude em larga escala sozinha, em casa, certamente estará correndo grande risco de ser descoberta.

A fraude que transforma o tráfego real de usuários em cliques inválidos dificilmente é detectada, mesmo quando aplicados métodos de filtragem de padrões repetidos de endereços IP (ANUPAM et al, 1999). Tal ataque pode ser escondido dos usuários ao se usar *iframes* (TEIXEIRA, 2004) de tamanho 0 para acessar anúncios por meio de JavaScript. Pode também ser camuflado de anunciantes e portais que possuem *Web Crawlers* (KOSTER, 2008) ao se certificar que os mesmos acessam uma pagina web legítima, enquanto visitantes humanos recebem uma página que realiza a fraude.

A técnica acima e outras técnicas que usam visitantes reais podem ser combinadas com o chamado tráfego incentivado, em que membros de determinados *sites* (*Paid to read*, ou simplesmente PTR) recebem pequenas quantidades de dinheiro para, centenas de vezes durante um dia, simplesmente visitar um site (concorrente ou não), clicar em palavras-chave ou resultados de pesquisa (MANN, 2006).

Alguns donos de sites PTR também são membros de ferramentas PPC e podem enviar muitos anúncios por *e-mail* para usuários que realizam essas pesquisas, ao mesmo tempo em que enviam pequenos anúncios para os que não realizam. Isso acontece porque, muitas vezes, a cobrança por cliques é a única fonte de renda do site – conhecido como “busca forçada”, uma prática mal vista na indústria.

O crime organizado também pode fazer uso de *click fraud*, ao possuir diversos computadores com suas próprias conexões a Internet em diferentes locais (geograficamente falando).

Muitas vezes, os *scripts* falham ao imitar o comportamento humano, então as redes de crime organizado usam códigos maliciosos em cavalos de tróia (*trojans)* (TOWNSEND, 2003) para transformar o computador de uma pessoa comum em uma espécie “computador zumbi”, para, esporadicamente, realizar ações que lhe beneficiam. Lidar com casos envolvendo redes de pessoas espalhadas em diferentes paises é muito difícil para anunciantes, redes de anúncios e autoridades.

Existe ainda outra forma de fraude, chamada de *Impression Fraud* (IMMORLICA, 2005), que acontece quando impressões de anúncio geradas de maneira maliciosa afetam a conta de um anunciante. Por exemplo, existem redes que utilizam modelos onde o anunciante pode ser penalizado se tiver um nível de aceitação (cliques) muito baixo em uma determinada palavra-chave.

A fraude consite em realizar inúmeras pesquisas sobre uma mesma palavra-chave, sem nunca se clicar no anúncio. Tal anúncio é desabilitado automaticamente, fazendo com que os anúncios mais caros (aqueles que aparecem nas primeiras páginas de pesquisa) não sejam exibidos, em detrimento do anúncio mais barato, do fraudador, que passa a aparecer nas primeiras páginas quando a mesma palavra-chave for pesquisada.

2.6 Casos Legais de Click Fraud

Disputas sobre *click fraud* resultaram em um grande número de processos. Em um determinado caso, descrito por Davis (2005), o Google (que agia tanto como anunciante quanto como rede de anúncios) venceu um processo contra uma empresa do Texas chamada de Auction Experts (que agia como publicador). O Google acusava a Auction Experts de pagar pessoas para clicar nos anúncios que apareciam no próprio *site*, causando um prejuízo de cinquenta mil dólares aos anunciantes.

Mesmo com os esforços das redes para parar este tipo de fraude, a verdade é que os publicadores desconfiam dos reais motivos dessas redes, já que elas também lucram com a fraude.

Em julho de 2005, o Yahoo! entrou em acordo em um processo de queixa no qual era acusado de não ter tomado precauções suficientes para prevenir a *click fraud*. O Yahoo! teve que pagar 4,5 milhões de dólares em taxas legais para os queixosos, e concordou em datar os valores do acordo para 2004.

Este caso é melhor explicado por Ryan (2006). Em março de 2006, o Google acertou, por 90 milhões de dólares, um acordo similar com a Lane's Gifts & Collectibles, caso detalhado por Tuzhilin (2006).

Em 2004, um morador da Califórnia, chamado Michael Anthony Bradley, criou o chamado Google Clique (NARAINE, 2004), um programa que, de acordo com o criador, tornaria possível que o Google fosse fraudado em milhões de dólares.

De acordo com autoridades, Michael Anthony Bradley foi preso e declarado culpado por chantagem, condenado a pagar cento e quintenta mil dólares e forçado a entregar o programa à empresa. Acredita-se que esta foi a primeira prisão por *click fraud*.

As acusações foram retiradas, sem explicação, em 22 de Novembro de 2006. Tanto o Google quanto o escritório de procuradores dos Estados Unidos se recusaram a comentar o caso. De acordo com Elgin (2006), em artigo publicado na revista Business Week, o Google não quis cooperar com o processo, uma vez que seria obrigado a:

i) expor publicamente suas técnicas de detecção de *click fraud*;

ii) admitir publicamente que lucra (através dessas técnicas ou não) com cliques fraudulentos.

2.7 Soluções para a Click Fraud

Provar a *click fraud* pode ser muito complexo, uma vez que é difícil saber quem está por trás de um computador e quais as suas intenções. O que a rede de anúncios pode fazer é identificar quais cliques são potencialmente fraudulentos e não cobrar os anunciantes por esses cliques. Existem diversos métodos sofisticados de detecção, mas nenhum é livre de falhas.

O relatório de Alexander Tuzhilin (TUZHILIN, 2006), produzido como parte do acordo entre Google e The Lane`s Gifts, possui uma discussão detalhada sobre estes problemas. Em particular, o relatório define que o "problema fundamental de cliques inválidos (fraudulentos)" é:

- "Não há uma definição conceitual de cliques inválidos que possa ser operacionalizada (com a exceção de casos óbvios)";

- "Uma definição operacional não pode ser totalmente revelada ao público em geral, uma vez que possibilitará que usuários maliciosos obtenham vantagem através do uso maciço de *click fraud*. Entretanto, se não for revelada, anunciantes não poderão verificar a razão e, assim, opor-se à cobrança de determinados cliques".

Existe atualmente, nos Estados Unidos, um considerável *lobby* da indústria de PPC para que leis mais rígidas sejam definidas para lidar com esse problema. A esperança é que estas leis vão descrever casos que não podem ser especificados em contratos.

Um grande número de empresas está desenvolvendo soluções viáveis para a identificação de *click fraud* através de relações intermediárias com redes de anúncio. Tais soluções subdividem-se em duas categorias:

1. Análise judicial dos arquivos de log originários dos servidores web de anunciantes

Essa análise de dados requer uma investigação profunda da fonte do tráfego de dados e do seu comportamento. A idéia é desenvolver arquivos de log "padrão", para análise, e comparar os dados disponíveis nos servidores com esses arquivos. O problema com esta abordagem é que ela confia na idoneidade das ferramentas de busca que têm a responsabilidade de identificar a fraude.

2. Confirmação de terceiros

Imaginemos a seguinte situação: um *site* externo oferece soluções para "etiquetar" os anúncios através da colocação de imagens, ou de Javascript, nas páginas web dos anunciantes para as quais os visitantes são direcionados. O visitante recebe um *cookie* ao visitar tais páginas.

A informação do visitante é coletada, armazenada em um banco de dados e disponibilizada para *download*. Desta forma, ao se analisar as melhores ofertas, é possível identificar um conjunto de cliques suspeitos, e, por causa da "etiqueta", expôr as razões para a desconfiança.

Essas informações, ao serem associadas com os arquivos de log dos anunciantes, formam um conjunto de evidências mais convincente, que pode ser apresentado à rede de anúncios para verificação.

O problema com as soluções baseadas na colaboração de terceiros está no fato de que tais soluções podem enxergar apenas uma parte do tráfego na rede inteira. Consequentemente, elas dificilmente identificarão a amplitude total de padrões de *click fraud* afetando muitos anunciantes ao mesmo tempo.

Além disso, devido a limitações na parte do tráfego ao qual têm acesso

(comparadas com ferramentas de busca), os terceiros podem ter dificuldades para julgar um determinado tráfego como fraudulento.

2.8 Click Fraud na Academia

Conforme descrito anteriormente, o fato de as ferramentas de busca serem mais confiáveis na definição de cliques inválidos é a razão básica do conflito de interesses entre os anunciantes, publicadores e as redes de anúncio. Isso é detalhado no relatório de Tuzhilin (TUZHILIN, 2006), o qual conclui ser impossível, atualmente, descrever de maneira detalhada as definições operacionais por trás da *click fraud* e, assim, definir um conceito público de cliques inválidos.

Em contrapartida, o relatório torna possível obter uma abstração em um sistema de detecção de fraudes. Tuzhilin (2006) também classifica como “razoável” que a ferramenta de busca seja colocada sob investigação em casos de *click fraud*.

Um dos objetivos do relatório é a preservação da privacidade do sistema de detecção de fraudes, de maneira a manter a sua efetividade. Isso acabou por motivar alguns pesquisadores a analisarem, por meio de pesquisas públicas e científicas, como as ferramentas de busca podem combater a *click fraud*.

Considerando que tais pesquisas não sejam corrompidas pelas forças de mercado, há uma esperança de que elas possam ser adotadas para avaliar, em casos jurídicos no futuro, quão eficientes ferramentas de busca podem ser na identificação da *click fraud* as.

Obviamente, ainda existe um certo temor acerca da necessidade de se expôr sistemas internos de detecção de fraude das redes de anúncios. Entretanto, se as pesquisas identificarem métodos de controle de fraude para cada técnica de fraude publicada, essa exposição se tornará menos crítica.

Um exemplo de uma dessas pesquisas pode ser encontrada no trabalho de Metwally, Agrawal e Abbadi (2005), que sugere a utilização de um histórico de ações e posterior busca em profundidade para identificação de comportamento suspeito, tornando necessário, assim, o controle de navegação dos usuários de uma rede de anúncios em diversos níveis (desde o acesso ao site do publicador, as ações que foram realizadas nesse site que causaram a exibição, a seleção e o clique no anúncio, e ainda ações posteriores ao redirecionamento).

Trabalhos recentes realizados por Majumdar, Kulkarni e Ravishankar (2007) propõem o uso de protocolos de comunicação para a identificação de comportamento fraudulento por parte de intermediários em redes de disseminação de conteúdo.

2.9 Considerações Finais

Existem várias abordagens para o combate de fraudes em negócios de marketing digital . Neste capítulo, foram analisadas algumas técnicas para a sistemática de

anúncios na Internet e exploradas, de maneira geral, algumas das diversas formas existentes de fraudes e abusos neste domínio de mercado, algo raro na literatura acadêmica em língua portuguesa.

Um foco especial foi dado à chamada *click fraud*, tendo-se estudado um apanhado acadêmico que incluiu diversas formas para a sua realização e métodos para o seu combate. Essa fraude foi escolhida por apresentar dificuldades técnicas no que diz respeito a sua prevenção. Obviamente, é impossível impedir que alguém realize cliques sem a intenção de compra em um anúncio específico, mas é possível adotar alguns passos para impedir o acontecimento da *click fraud* em larga escala e automatizada.

Este livro está fundamentado na realização de um estudo extensivo do domínio de negócios de marketing digital , de maneira a entender o que se pode fazer para prevenir a *click fraud*, a qual tem sido largamente discutida na academia e no âmbito de mercado, e representa um verdadeiro desafio para este negócio.

Conforme já mencionado anteriormente (vide capítulo 1 – seção 1.2), uma das finalidades principais deste trabalho está relacionada com a proposição de uma abordagem, a ser apresentada no capítulo 4, para o combate efetivo à *click fraud*.

Para que a construção da abordagem que seja realizável no modelo de negócios existente, ela adotará um conjunto de métodos já propostos anteriormente e utilizados em organizações. Tais métodos foram implementados em projetos de grande porte, tendo apresentado resultados satisfatórios sob a perspectiva dos participantes do negócio e dos demais envolvidos em pesquisas nesta área, no âmbito da academia.

Após a proposição desta abordagem, a comunidade contará com uma ferramenta poderosa para trabalhos futuros, que podem conter a especialização desta abordagem no nível de arquitetura e desenvolvimento.

CAPÍTULO 3

UTILIZANDO CAPTCHA PARA DISTINGUIR HUMANOS DE COMPUTADORES

Este capítulo estuda a tecnologia de CAPTCHA como uma forma de garantir a utilização de serviços disponíveis na Internet por usuários humanos. Inicialmente, o leitor será introduzido à tecnologia e veremos alguns exemplos, de maneira a apresentar o conceito por trás da utilização de CAPTCHA. Posteriormente, será apresentado um histórico da tecnologia, as suas diversas aplicações e uma classificação para os CAPTCHA.

Após esta exposição, serão estudados alguns requisitos que, se atendidos, caracterizam um bom CAPTCHA. Esses requisitos serão utilizados posteriormente para uma análise de *benchmarks* de alguns CAPTCHA cuja representatividade seja considerada satisfatória. Por fim, apresentamos os CAPTCHA clicáveis, uma solução mais intuitiva e segura para garantir a distinção entre humanos e computadores.

3.1 Introdução a CAPTCHA

Nos últimos anos, um número crescente de serviços públicos na web tentou prevenir o abuso de programas automáticos ao requerer que seus usuários resolvessem um desafio de diferenciação entre usuários humanos e máquinas, inspirado no teste proposto por Turing (1950) (hoje conhecido como CAPTCHA, “Completely Automated Public Turing test to tell Computers and Humans Apart”) antes de começarem a usar o serviço.

Em teoria, os testes são fáceis de serem gerados, e difíceis, para usuários não humanos, de serem resolvidos. Todos os CAPTCHA possuem alguma informação secreta que é conhecida pelo desafiante mas não pelo agente desafiado.

O termo CAPTCHA foi inventado em 2000 por Luis von Ahn, Manuel Blum, Nicholas J. Hopper (todos da universidade Carnegie-Mellon) e por John Langford (IBM) (AHN et al., 2003) e envolve, basicamente, um computador (servidor) que pede que um cliente realize uma interação com o objetivo de terminar um teste. O objetivo do teste é determinar que todo cliente que aponte uma solução correta seja presumidamente humano.

CAPTCHA pode ser visto também como um programa que faz um teste do tipo “desafio-resposta” com o objetivo de determinar se seu usuário é um humano ou um computador. Um tipo comum de CAPTCHA requer que o usuário identifique as letras de uma imagem colorida e distorcida, às vezes com a adição de uma seqüência obscurecida das letras e/ou dos dígitos que

aparecem na tela. CAPTCHA são usados por muitos *websites* como uma forma de prevenir o abuso de "*bots*" ou de outros programas automáticos normalmente escritos para gerar *spam*.

Um CAPTCHA também é descrito como um "teste reverso de Turing", já que o teste é administrado por um computador, em contraste ao teste padrão de Turing, que é administrado por um ser humano. Em teoria, *bots* não conseguem ler textos distorcidos tão bem quanto humanos, de maneira que CAPTCHA podem impedir que *bots* naveguem em *sites* protegidos. Na prática, um CAPTCHA introduz um problema cuja solução só pode ser automatizada utilizando técnicas de inteligência artificial. Um sistema de CAPTCHA é uma forma de gerar, de maneira automatizada, novos desafios que:

- Os computadores atuais sejam incapazes de resolver;
- A maioria dos humanos possa resolver;
- Não garante que o atacante desconhece o desafio. Por exemplo, embora um checkbox do tipo "clique aqui se você não é um bot" possa servir para distinguir um humano de computadores, não pode ser caracterizado como um CAPTCHA porque confia no fato de que um potencial fraudador não precisou de grandes esforços para acertar aquela pergunta específica.

3.1.1 Origem e Aplicações

A dificuldade em distinguir humanos de computadores fingindo ser humanos foi inicialmente discutida em 1950, quando Alan Turing descreveu o seu famoso "teste de Turing". A primeira discussão sobre usar testes automáticos que realizam esta diferenciação com o objetivo de controlar acesso a serviços web que se tem registro é um manuscrito de Moni Naor (1996) do Instituto de Ciência de Weizmann, entitulado "*Verification of a Human in the Loop, or Identification via the Turing Test*".

Os primeiros CAPTCHA parecem ter sido desenvolvidos em 1997, para o site de busca Altavista, com o objetivo de impedir que bots adicionassem páginas web para a ferramenta de busca. De maneira a fazer com que as imagens não fossem detectadas por um OCR (*Optical Character Recognition*), o time simulou situações que manuais de aparelhos de *scanner* diziam resultar em mal funcionamento de OCRs. Eles inventaram múltiplos exemplos de CAPTCHA, incluindo os primeiros a serem usados publicamente, que foram os adotados inicialmente no Yahoo!.

CAPTCHA são usados para previnir que *softwares* automatizados realizem ações que degradam a qualidade do serviço de um dado sistema por causa do abuso no uso de recursos. Na maioria das vezes, os CAPTCHA são usados como uma resposta para invasões por interesses comerciais, mas eles podem ser utilizados em diversos outros aspectos, como a proteção de sistemas vulneráveis a spam, como *webmail*, ou de forma a evitar postagem automática em *blogs* ou fóruns, de maneira a impedir os anúncios comerciais não-autorizados, vandalismo ou outras formas de perturbação.

Além disso, CAPTCHA podem ser usados em sites de estatísticas, para limitar o uso automatizado de um serviço em detrimento dos usuários humanos quando uma determinada taxa de uso for atingida.

Em tais casos, os CAPTCHA serão aplicados para garantir políticas de uso de aplicações automáticas, definidas por um administrador, quando métricas de uso excederem um determinado limiar. Um exemplo de aplicação na qual tais vulnerabilidades existem, e no qual tais vulnerabilidades poderiam ser facilmente anuladas usando um CAPTCHA, pode ser encontrado em (ARORA, 2007).

Algumas das aplicações de CAPTCHA, utilizadas em segurança, podem ser encontradas abaixo:

- Prevenção de *comment spam* em *blogs*: a maioria dos donos de *blogs* conhece programas para submeter comentários aleatórios em *blogs* (como “compre online no site x”), normalmente com o objetivo de aumentar a pontuação em *page ranks* de algum *website*. Isto é chamado de *comment spam*. Ao usar um CAPTCHA, apenas humanos conseguem comentar em um blog, de modo que não é necessário que os usuários se registrem para poderem comentar e os comentários legítimos ficam guardados de maneira satisfatória.

- Proteção de registro em *websites*: muitas empresas (Yahoo!, Microsoft, etc.) permitem o cadastro de *e-mails* de maneira gratuita. Até alguns anos atrás, a maioria destes serviços sofreram um tipo específico de ataque: *bots* que registravam milhares de contas a cada minuto. A solução para este problema era o uso de CAPTCHA para garantir que apenas humanos conseguissem contas gratuitas. De maneira geral, todo tipo de serviço gratuito *online* deve ser protegido por CAPTCHA para prevenir o uso abusivo de programas automáticos.

- Proteção de pesquisas *online*: em novembro de 1999, o site http://www.slashdot.org fez uma pesquisa que perguntava qual era a melhor escola de graduação de ciência da computação nos Estados Unidos. A segurança era definida da seguinte forma: os endereços IP dos votantes eram armazenados de maneira a prevenir que usuários votassem mais de uma vez.

Os alunos da Carnegie Mellon University (CMU) escreveram um script e usaram diferentes *proxies* para votar na sua universidade milhares de vezes. A pontuação da CMU cresceu de maneira muito rápida. No dia seguinte, os alunos do Massachusetts Institute of Technology (MIT) escreveram seu programa, e a disputa se tornou uma luta entre “*bots*”.

O MIT terminou com 21.156 votos, e a Carnegie Mellon com 21.032. Todas as outras escolas terminaram com menos de 1.000 votos. O resultado de uma pesquisa *online* é confiável? A resposta: apenas se garantir que apenas humanos votarão.

- Prevenção de ataques de dicionário (JAKOBSSON; RAZMAN, 2008): CAPTCHA podem também ser usados para prevenir ataques de dicionário em sistemas de senha. A idéia é simples: impedir a interação de um computador solicitando que, após uma determinada quantidade de *logins* sem sucesso, o usuário resolva um CAPTCHA.

- *Bots* de busca: algumas vezes, é interessante manter algumas páginas *web* ocultas (ou seja, não indexadas) para impedir que sejam localizadas de maneira fácil por outros. Existe uma tag HTML para requisitar que os *bots* não leiam a página, mas a mesma não passa de um "pedido" – de maneira que não impede que o *bot* leia, de fato, a página. Os *bots* de busca automáticos de grandes empresas normalmente respeitam estas tags, mas, para realmente garantir que os bots não acessarão um web site, CAPTCHA são necessários.

- *Worms* e *spam*: CAPTCHA normalmente oferecem uma maneira plausível contra *worms* de *e-mail* e contra *spam*: a idéia é que o servidor (ou o usuário, por meio de definições de preferência) só aceite *e-mails* de humanos.

3.1.2 Classificação

De acordo com Elson et al. (2007), podem-se definir duas classes distintas de CAPTCHA. Os CAPTCHA de Classe I, nos quais a informação secreta é apenas um número aleatório, que é usado por um algoritmo público para alimentar e gerar um desafio (análogo a um cripto-sistema de chave pública). Os CAPTCHA de Classe II, além de um número aleatório, usam um banco de dados secreto com alto grau de entropia (análogo a um cripto-sistema do tipo *one-time-pad*).

Os CAPTCHA de Classe I possuem muitas vantagens. Algoritmos para a implementação deste tipo de CATPCHA podem ser desenvolvidos em poucas linhas de código, não tornam necessário o armazenamento de nenhuma informação secreta e podem gerar um conjunto praticamente ilimitado de desafios únicos. Entretanto, a sua grande realização – um desafio para reconhecer texto distorcido – evidencia uma diferença muito pequena nas taxas de sucesso de humanos e máquinas.

Algoritmos de *Optical Character Recognition* (OCR) são competitivos com humanos em reconhecer caracteres distintos, o que fez com que os pesquisadores se motivassem a aumentar a dificuldade de segmentar os caracteres de uma imagem em regiões distintas (SIMARD; STEINKRAUS; PLATT, 2003).

Entretanto, esse aumento de dificuldade também afeta humanos e, embora experimentos de laboratório relatem que humanos podem segmentar os caracteres de maneira acurada (CHELLAPILLA et al, 2005), CAPTCHA em páginas comerciais da Internet continuam a usar desafios claramente segmentados (exemplo: Figura 3.1).

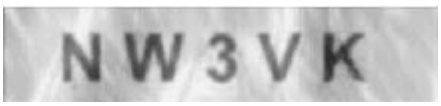

Figura 3.1 – CAPTCHA simples (register.com)

É importante observar que vários proprietários de páginas web atualmente

disponíveis decidiram que o sucesso de um usuário ao resolver um CAPTCHA não depende somente do fato de ele ser capaz de resolver o desafio, mas também de sua disposição de se esforçar para tal, pois CAPTCHA relativamente simples são suficientes para afastar um número substancial de clientes.

Os CAPTCHA de Classe II são capazes de superar as fraquezas descritas acima. Uma vez que não são restritos a desafios gerados por algoritmos de baixo nível de entropia, eles podem explorar uma quantidade bem maior de habilidades humanas, como o reconhecimento de detalhes em imagens fotográficas capturadas do mundo real.

Tais desafios evidenciam taxas de sucesso muito distintas entre humanos e não humanos, devido basicamente a dois motivos: além de ser muito mais difícil para um computador interpretar uma imagem do que resolver desafios baseados em texto, é possível elaborar desafios baseados em imagem bem menos tediosas para humanos sem perder sua eficiência no bloqueio de programas.

3.1.3 reCAPTCHA

Alguns inventores do sistema de CAPTCHA desenvolveram uma forma de aproveitar o esforço e o tempo de pessoas que respondem aos desafios CAPTCHA em uma espécie de sistema distribuído. Este sistema funciona incluindo elementos “resolvidos” e “não-resolvidos” (imagens que não foram reconhecidas satisfatoriamente por um OCR) em cada desafio.

O usuário responde a ambos os elementos – metade do seu trabalho valida o desafio e a outra metade é capturada como trabalho para o sistema distribuído. Tal sistema, denominado reCAPTCHA (AHN et al., 2008), é atualmente usado na conversão de trabalhos impressos (imagens escaneadas) em textos digitais.

Essa abordagem é similar a uma das técnicas através das quais os sistemas de CAPTCHA podem ser fraudados (discutido na seção 1.5.3, quando recursos humanos são utilizados em pequenas partes do trabalho de maneira altamente distribuída).

Os mantenedores do reCAPTCHA estimam que os sistemas de CAPTCHA atuais representam aproximadamente 150.000 horas de utilização por dia, que poderia ser, de maneira transparente, utilizada para revisão no seu sistema distribuído. Isso equivaleria a 19.000 pessoas trabalhando 8 horas por dia para melhorar OCRs que podemser usados para prevenir ataques.

3.2 Requisitos para o Desenvolvimento de CAPTCHA

O desenvolvimento de CAPTCHA, assim como todos os projetos desta natureza, está associado com a satisfação de diversos requisitos, sendo eles técnicos, de segurança, funcionais e não-funcionais.

Entretanto, existem requisitos básicos para o desenvolvimento de CAPTCHA

que precisam ser considerados para que a própria utilização dessa ferramenta possa ser justificada junto aos usuários e potenciais usuários dos serviços online.

Nesta seção, serão listados alguns destes requisitos cuja observação é imprescindível para que a utilização dos sistemas de CAPTCHA nos serviços online seja minimamente aceitável para qualquer usuário que utilize tais serviços.

3.2.1 Usabilidade e Acessibilidade

Uma característica importante na qual desenvolvedores de sistemas de CAPTCHA necessitam ter atenção especial é a acessibilidade. A resolução de testes muito difíceis pode tornar a experiência de uso extremamente penosa, e acabar afastando os usuários dos serviços relacionados a um site. Em contrapartida, CAPTCHA de fácil resolução para humanos, em teoria, também serão de fácil resolução para máquinas. Isso decorre do fato de o uso de CAPTCHA necessitar de "percepção".

Acessibilidade é especialmente importante nesse sentido, pois usuários que sejam incapazes de entender um CAPTCHA (por exemplo, devido a uma deficiência ou ainda por dificuldade de leitura) serão também incapazes de realizar uma tarefa protegida por um CAPTCHA.

O site oficial sobre CAPTCHA, http://www.captcha.net, recomenda que, por motivos de acessibilidade, os sites que os utilizam disponibilizem também uma versão de áudio do CAPTCHA em questão. Entretanto, mesmo um CAPTCHA visual e em áudio necessita de intervenção manual para alguns usuários, como pessoas que são, ao mesmo tempo, surdos e cegos.

Existem inúmeras tentativas de tornar CAPTCHA mais acessíveis, incluindo o uso de Javascript, questões matemáticas (quanto é 10+10?), ou "questões de senso comum" (qual é a cor do céu?). Claramente, essas tentativas violam no mínimo uma das características básicas dos CAPTCHA: ou não podem ser automaticamente gerados ou podem ser facilmente quebrados dado o estado atual da inteligência artificial.

De acordo com Ahn et al. (2003), da maneira anterior, a única segurança destes CAPTCHA é a chamada "segurança por obscurecimento" (MAJUMDAR; KULKARNI; RAVISHANKAR, 2007): um atacante que não encontrou a formula de quebrar o CAPTCHA muito facilmente, provavelmente não vai achar que vale a pena empregar todo o esforço que será necessário para automatizar a resolução do CAPTCHA de um *site* pequeno.

Devido à ausência de segurança em CAPTCHA textuais, a maioria dos sites escolhem usar um CAPTCHA visual e de áudio como uma forma de balancear segurança e acessibilidade. Além disso, suporte por *e-mail* e/ou telefone é usado para prover acesso manual a usuários que são incapazes de resolver um CAPTCHA.

Existem algumas abordagens para derrotar CAPTCHA: explorar bugs na implementação ou arquitetura que permitem ao atacante simplesmente "pular" o CAPTCHA, melhorar *softwares* de reconhecimento de caracteres, ou usar recursos

humanos para simplesmente processar os testes.

Um CAPTCHA é criado para separar humanos de programas automatizados. Desta forma, precisam ser compreensíveis por todos os seus potenciais usuários e precisam ser incompreensíveis pelos computadores. Um computador não vai reclamar se não conseguir ler um CAPTCHA – mas um usuário humano certamente reclamará. É necessário saber que a facilidade de leitura de um CAPTCHA opõe-se de maneira proporcional à incapacidade de ser quebrado por um computador.

Assim, os desenvolvedores devem se perguntar e tomar uma decisão: "o que é mais importante, a capacidade de este CAPTCHA ser compreensível ou a incapacidade de um programa de quebrá-lo?".

A cognição humana é naturalmente capaz de separar sinais válidos de ruídos, objetos de imagens em *background*, e utilizar outros aspectos denominados de "abstrações". Ao tentar desenvolver um CAPTCHA seguro, o desenvolvedor precisa utilizar esta capacidade a seu favor, de maneira que as ações necessárias para separar a informação relevante sejam intuitivas.

Um exemplo disso seria a utilização de uma palavra real em lugar de uma sequência aleatória de letras e números: o entendimento de uma palavra completa pode ajudar a resolução correta do teste. Entretanto, este método precisa ser utilizado com cuidado por causa do requisito explicado a seguir.

3.2.2 Universalidade

As pessoas falam línguas diferentes, escrevem utilizando símbolos diferentes e vivem em suas culturas particulares. Um bom *designer* de CAPTCHA deve ter isso em mente como uma de suas prioridades. Para garantir o requisito da universalidade, é necessário responder a perguntas como as seguintes antes de desenvolver um sistema que contenha testes do tipo CATPCHA:

- Quem serão as pessoas que verão o CAPTCHA?
- Como garantir que o CAPTCHA está claro para todos os que o verão?
- Como garantir que o CAPTCHA não será ofensivo para todos os que o verão?

Bons CAPTCHA utilizam apenas símbolos que são intuitivamente claros e compreensíveis para todas as pessoas que poderão eventualmente acessá-lo: pessoas que potencialmente utilizarão um serviço de massa na internet. Esta consideração é um aspecto chave acerca de como desenvolver código para CAPTCHA:

- Evitar símbolos que não sejam comuns a todos os tipos de teclado (por exemplo, números);
- Evitar CAPTCHA que necessitem de explicações no formato texto (ex: CAPTCHA que pedem que o usuário clique em uma determinada área da imagem e esperam que o resultado do teste seja digitado com informações disponíveis nesta explicação).

É importante que os desenvolvedores tenham em mente que os visitantes do site, aqueles que acessarão e – espera-se – resolverão o CAPTCHA, é que representam o maior valor para o serviço, e não o risco de o serviço ser utilizado de forma ilegítima. Desta forma, algumas das seguintes idéias devem ser evitadas:

- Mostrar fotos de animais e pedir que os usuários digitem de qual espécie eles são;
- Mostrar a foto de uma celebridade e pedir que os usuários digitem o seu nome (é possível que uma determinada celebridade não seja conhecida por usuários de uma determinada região, ou ainda por motivos tais como interesse pessoal);
- Mostrar uma imagem que pede aos usuários que cliquem em uma determinada área (não há garantia que todos os usuários saibam ler uma determinada língua e de fato entenderão o que precisa ser feito).

3.2.3 Singularidade

Cada CAPTCHA precisa ser único ou, pelo menos, muito raro, pois a idéia de utilizar CAPTCHA é similar a qualquer outro sistema de automatização: a realização de algo em um tempo ilimitado que funcionará ilimitadamente.

Se o desenvolvedor decidir desenvolver um conjunto de imagens de maneira manual e mostrá-las uma a uma, aleatoriamente, para resolução, o trabalho do fraudador se resumirá ao reconhecimento manual de cada uma das figuras. E é importante notar que implementar o reconhecimento manual é sempre menos custoso do que o *design*, o que introduz um novo problema: o desafio do CAPTCHA se resume na diferença do tempo em que o desenvolvedor conseguirá fazer novas imagens e quão rapidamente o fraudador conseguirá adicionar cada nova imagem ao seu banco de dados.

Assim, a solução para satisfazer este requisito é criar uma “caixa preta” que tem como saída uma quantidade infinita de imagens únicas, baseadas em um conjunto limitado de símbolos, letras, caracteres, fontes, parâmetros aleatórios e valores, além de adicionar filtros, distorções, ruídos, entre outros aspectos à imagem.

3.2.4 Tecnologia

Muitas vezes o problema que é resolvido pela adoção de um CAPTCHA em um determinado site não é tão sério, de maneira que o orçamento dos serviços, produtos ou projetos dificilmente permitirá que se adquira mais um servidor dedicado exclusivamente a gerar CAPTCHA.

Em contrapartida, é comum que os CAPTCHA, por manipularem imagens e construirem figuras a partir de outras, utilizem uma quantidade de memória significativa – que será multiplicada exponencialmente em função de sua utilização em um ambiente de produção.

É importante frisar que algumas idéias para tornarem o CAPTCHA mais seguro,

tais como a utilização de diferentes cores nos diferentes caracteres do desafio ou ainda a adição de ruído e distorções na imagem gerada, podem ser muito custosas.

3.2.5 Implementação Segura

Como em todo sistema de segurança, falhas de design e arquitetura na implementação podem contribuir para a falta de segurança do mesmo. Existem diversas implementações de sistemas de CAPTCHA e muitas delas, especialmente as que não foram revisadas nem avaliadas por especialistas em segurança, são vulneráveis a ataques comuns.

Muitos sistemas de proteção baseados em CAPTCHA podem simplesmente ser quebrados, inclusive sem a utilização de OCRs, apenas re-utilizando a "session ID" de uma dada imagem de CAPTCHA.

Um sistema implementado de maneira correta não permite que várias tentativas de solução sejam realizadas contra um único CAPTCHA, impossibilitando o reuso de uma solução previamente correta – de um insucesso, a tentativa seguinte deve conter uma nova imagem, com identificador de sessão diferente e um texto de resposta diferente.

Outras implementações de CAPTCHA utilizam sugerem a utilização de um *hash* MD5 (AHN et al, 2003) da solução como uma chave privada passada para o cliente, de maneira a validar o CAPTCHA. Entretanto, como normalmente os CAPTCHA são desafios de texto com poucos caracteres, quebrar este nível de criptografia em *hash* é fácil, e na verdade esta disposição pode ajudar um adversário em sua tentativa de utilizar OCR para quebrar o CAPTCHA. Um esquema mais seguro seria a utilização de um HMAC (AHN et al., 2003).

Finalmente, algumas implementações utilizam apenas um pequeno número fixo de imagens para os desafios. Eventualmente, quando um atacante coletar imagens suficientes, o CAPTCHA poderá ser quebrado de maneira simples, ao se procurar as soluções em uma tabela através de uma busca baseada no *hash* das imagens sendo exibidas.

3.3 Optical Character Recognition (Reconhecimento Óptico de Caracteres)

Em Hollywood, especialmente em filmes de ficção científica, pode-se achar diversos exemplares de robôs com capacidades cognitivas e lingüísticas. Os andróides do cinema podem servir de inspiração para um número significativo de criações computacionais. Para tal, faz sentido repetirmos a pergunta: onde estamos hoje? A tecnologia da informação vai, algum dia, satisfazer as previsões dos futuristas? É de praxe tentar responder a essas perguntas com mais previsões, já que o conhecimento atual torna possível a subdivisão de tendências nas diversas áreas da computação.

Esta seção considera um dos grandes problemas interdisciplinares da

computação: a identificação de caracteres em um documento. Essa área é chamada de *Optical Character Recognition* (reconhecimento óptico de caracteres) (MORI; NISHIDA; YAMADA, 1999) ou, simplesmente, OCR.

Um documento pode conter caracteres impressos por uma máquina (como esta página), gerados artesanalmente ou simplesmente ser composto de manuscritos cursivos (MORI; NISHIDA; YAMADA, 1999). Sistemas para reconhecimento de texto impresso por máquina são originários nos fins dos anos 50 e têm sido largamente utilizados em computadores caseiros desde o início dos anos 90.

Uma das grandes motivações para OCR é o fato de que uma grande parcela de informações históricas arquivadas ainda está na forma não-digital, permanecendo "cativa" em documentos impressos. Os sistemas OCR são uma forma de aumentar a acessibilidade deste tipo de informação, já que convertem o texto escrito em forma eletrônica.

Uma vez digitalizado, sistemas de recuperação de informação podem ser utilizados para a localização do material de interesse, e processadores de texto podem ser utilizados para a edição do mesmo.

Em resumo, OCR pode ser definido como o processo mecânico ou eletrônico de tradução de imagens de textos que foram escritos à mão, digitados ou impressos (normalmente capturados por meio de um scanner) para texto editável em formato de máquina. Entretanto, sistemas de OCR não fazem essa conversão de maneira perfeita. Eles cometem erros, de modo que a versão eletrônica de um documento dificilmente será idêntica à versão em papel.

Um erro comum, por exemplo, em determinados tipos de fonte, acontece quando o sistema não consegue distinguir a letra e da letra c. *Voec podc aeabar eom um tcxto quc sc parcec com isso*. Em tais casos, é mais eficiente digitar o conteúdo inteiro da página no computador do que sair procurando os erros no texto de saída e corrigi-los. A coisa mais importante em OCR é definir parâmetros para que a seguinte condição seja observada: um sistema OCR que cometa muitos erros não é útil.

OCR é um campo de pesquisa em diversas áreas da computação: reconhecimento de padrões (MORI; NISHIDA; YAMADA, 1999), inteligência artificial (AHN et al., 2003) e computação gráfica (AZEREDO; VELHO, 2003). Muito embora ainda existam pesquisas científicas nesta área, o foco atual de OCR mudou para a implementação de técnicas já provadas pela teoria. O escopo de pesquisa de OCR, inicialmente, não incluia o reconhecimento de caracteres digitais (usando scanners e algoritmos), mas estava concentrado nas técnicas ópticas, como as que fazem uso de espelhos e lentes. Entretanto, devido ao fato de que pouquíssimas aplicações atuais fazem uso dessas técnicas, o termo OCR foi estendido para também incluir a área de processamento de imagens digitais (AZEREDO; VELHO, 2003).

3.3.1 Estado Atual da Tecnologia OCR

A tecnologia OCR avançou a ponto de os sistemas atuais serem confiáveis para processar uma grande variedade de documentos impressos por máquina. O reconhecimento acurado do alfabeto latino na sua forma digitada é considerado um problema resolvido.

O reconhecimento para *scripts* latinos atinge, tipicamente, níveis de precisão de até 99% (embora existam algumas aplicações que demandem níveis de precisão ainda maiores – estas necessitam de revisão humana) (MORI; NISHIDA; YAMADA, 1999).

Entretanto, ainda há muito o que se aperfeiçoar. É bom lembrar que 99% de precisão em uma página de 3000 palavras ainda significa 30 erros na página.

O reconhecimento de textos escritos à mão, incluindo o reconhecimento de imagens impressas de manuscritos e de manuscritos cursivos, é ainda objeto de pesquisa, bem como o reconhecimento de textos impressos em outros alfabetos (em particular os que possuem um grande número de caracteres).

Sistemas para reconhecimento de texto manuscrito digitalizado obtiveram um bom desempenho comercial nos últimos anos. Entre eles, destacam-se os dispositivos de PDA (*personal digital assistant*) (KAMBA et al., 1996), como os que rodam em Palm OS. O pioneiro dessa tecnologia de foi o Apple Newton (KAMBA et al., 1996).

Os algoritmos usados nesses equipamentos de digitalização possuem algumas vantagens, uma vez que a ordem, a velocidade e a direção dos segmentos de linhas individuais na entrada são conhecidos. Além disso, o usuário pode ser limitado a formatos específicos de letra.

Métodos como os descritos acima não podem ser usados, por exemplo, em *softwares* que realizam digitalização e processamento de documentos em papel. O reconhecimento de documentos manuscritos digitalizados ainda é, portanto, um problema: os níveis de precisão variam de 80 a 90% em caracteres limpos e escritos claramente, o que significa dúzias de erro por página.

Isso faz com que a tecnologia de OCR seja útil apenas em um número bastante limitado de aplicações, como indexadores de imagens obtidas de formulários digitalizados ou captura seletiva de informações).

Existe uma espécie de variação do OCR na qual diversos fatores de inteligência artificial são considerados: a *Intelligent Character Recognition* (ICR) (MANTAS, 1986). O reconhecimento de textos cursivos ainda é uma área ativa de pesquisa, com taxas de reconhecimento ainda menores do que de textos manuscritos impressos.

Na verdade, taxas altas de reconhecimento e níveis de precisão maiores só parecem ser possíveis com o uso de informações contextuais e gramaticais. Pode-se perceber isso quando analisarmos as seguintes situações: em um dicionário, é bem mais fácil reconhecer palavras inteiras do que ler palavra por palavra e tentar interpretar cada caractere de maneira individual. Entretanto, ler o valor de um cheque (valor numérico) é um exemplo em que o uso de um dicionário menor aumenta

significativamente as taxas de reconhecimento.

O conhecimento da gramática da linguagem que está sendo processada também pode ajudar a determinar se uma palavra deve ser um verbo ou um substantivo e, desta forma, permitir maior precisão. De maneira individual, os formatos de caracteres cursivos simplesmente não contêm informações suficientes para reconhecimento acurado (ou seja, com uma taxa de reconhecimento maior que 98%) *scripts* cursivos que foram manuscritos.

3.3.2 OCR e CAPTCHA

Uma boa quantidade de projetos de pesquisa tentou (e a maioria foi bem sucedida) fraudar CAPTCHA visuais ao criar programas que realizam os seguintes passos para obter o código exibido:

1. Extração da imagem da página.
2. Remoção de padrões de desordem do fundo da imagem (*background*), utilizando filtros de cor e detectando padrões de linhas finas.
3. Segmentação (dividir a imagem em regiões contendo um único caractere – como “NNNN”).
4. Identificação da letra para cada região (exemplo: respondendo à pergunta, “que letra o conjunto de NNN aparenta formar?”).

Os passos 1, 2 e 4 são muito simples para computadores (MANTAS, 1986). A única parte onde os humanos ainda são melhores que os computadores é na segmentação. Por exemplo, se a desordem no *background* consiste de formatos semelhantes a letras, e se as letras estão conectadas à desordem, segmentar a imagem computacionalmente se torna praticamente impossível com os *softwares* atuais (MORI; MALIK, 2003). Desta forma, um bom CAPTCHA textual deve focar em segmentação.

3.4 Análise e Classificação de CAPTCHA Benchmarks

Levando em consideração o estudado na seção anterior, é possível selecionar um conjunto de parâmetros e utilizá-los para analisar um CAPTCHA, atribuindo-lhe notas para se obter uma idéia da dificuldade que um fraudador encontraria para implementar um sistema que fosse capaz de desvendar os códigos que tal CAPTCHA apresenta.

Pode-se considerar os seguintes padrões para se concluir o nível de dificuldade de um CAPTCHA:

- Conhecimento de forma de desenvolvimento (formas mais difundidas são, consequentemente, mais conhecidas entre os fraudadores e devem receber notas menores). Um exemplo de forma de desenvolvimento é a classe HN CAPTCHA do PHP (TRYCAPTCHA, 2008);
- Nível e quantidade de ruídos (quanto menos ruídos e quanto maior a facili-

dade de identificá-los, menor a nota);

- Esquema de cores e de exibição de código relevante;
- Facilidade de identificação do código por usuários humanos (quanto mais difícil, menor a nota);
- Presença de sobreposição entre os caracteres do CAPTCHA e, se existir, qualidade da mesma (quanto mais difícil diferenciar onde se inicia um caractere e termina outro, melhor a qualidade da sobreposição);
- Legibilidade do CAPTCHA.

Esta seção, ao considerar os parâmetros acima, apresenta uma classificação de diversos exemplos de CAPTCHA presentes na Internet. Além disso, atribuímos uma nota de 1 a 5 para cada um desses CAPTCHA, considerando a dificuldade para potencialmente automatizar a solução deles por meio de OCR. Quanto maior a nota, mais difícil a automatização.

É importante ressaltar que o conjunto de CAPTCHA em questão foi selecionado por ser uma boa representação dos CAPTCHA atualmente disponíveis na Internet. Muitos CAPTCHA conhecidos não foram considerados nesta análise, pois teriam avaliação muito semelhante a uma das análises a seguir.

3.4.1 Friendster

O *Friendster* é uma rede social extremamente popular na Internet. Ao acessar o site de cadastro, os usuários são desafiados a responder um CAPTCHA semelhante aos da Figura 3.2, antes de concluir o processo de registro.

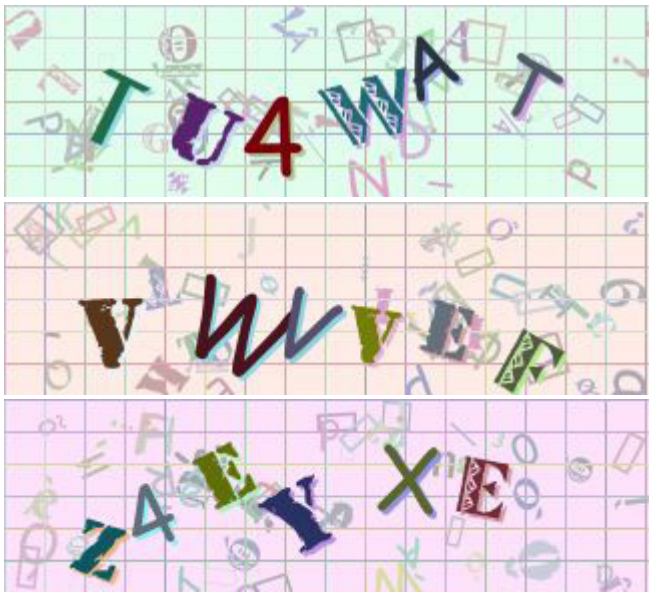

Figura 3.2 – Exemplos de Captcha do Friendster, dificuldade 3

Embora se trate de um CAPTCHA interessante e sem desenvolvimento comum (passando, assim, no requisito “Forma de Implementação”), recebe apenas nota de

dificuldade 3 por favorecer de maneira simples o algoritmo da seção 3.3.2, já que os ruídos são facilmente identificados e as distorções podem ser removidas.

3.4.2 UA.FM

A Figura 3.3 mostra exemplos do CAPTCHA da UA.FM que, embora apresente um nível de ruído fraco, que pode ser facilmente identificado e removido, e devido ao fato de que as fontes escolhidas dificultam a legibilidade do mesmo, se trata de um CAPTCHA forte que requer um esforço considerável para automatização. Dificuldade 3.

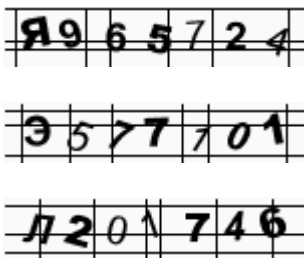

Figura 3.3 – UA.fm, encontrado em http://ua.fm

3.4.3 CAPTCHA HN

Na Figura 3.4 vemos um conjunto de CAPTCHA HN. O CAPTCHA HN é gerado gratuitamente por uma classe PHP que pode ser encontrada no site http://www.nogajski.de/horst/php/captcha/. Embora muitos criadores de CAPTCHA acreditem que adicionar ao fundo da imagem outras fontes no formato ttf, girar letras e as exibir em cores diferentes torne mais difícil a automatização da resolução dos mesmos, na verdade, é relativamente fácil separar os caracteres de CAPTCHA das letras exibidas no plano de fundo, e o modelo de cores deste CAPTCHA diminui significativamente sua usabilidade. Por essas razões, trata-se de um CAPTCHA com dificuldade 2.

Figura 3.4 – CAPTCHA HN, de dificuldade 2.

3.4.4 MySpace.com

Também exibido para efeitos de registro (na página http://signups.myspace.com/index.cfm?fuseaction=signup), o CAPTCHA da Figura 3.5 já apresenta um nível de dificuldade maior. Mesmo não sendo recomendada a utilização exclusiva do modelo de cores para a separação dos códigos, é possível separar as letras de um CAPTCHA através da utilização de seus próprios parâmetros, como posição e formato. Entretanto, é possível observar que se trata de um CAPTCHA bem planejado e desenvolvido por uma equipe que teve acesso a um bom nível de treinamento sobre o assunto, recebendo nota 4 de dificuldade.

Figura 3.5 – MySpace.com, dificuldade 4

3.4.5 Jeans

Jeans é uma empresa provedora de serviços de telefonia celular. Seu CAPTCHA é exibido ao se acessar a página http://www.jeans.com. ua/sms/, onde se pode enviar mensagens SMS. Uma amostra desse de CAPTCHA pode ser encontrado na Figura 3.6. Foi quebrado com facilidade, já que a forma de separação de caracteres através do modelo de cores o torna muito vulnerável. Dificuldade 1.

Figura 3.6 – CAPTCHA do *Jeans* – facilmente quebrado, dificuldade 1

3.4.6 UMC

Na Figura 3.7, pode-se ver O CAPTCHA da UMC, que pode ser acessado na página http://www.umc.ua/ukr/sendsms.php e é muito promissor. Antes de mais nada, possui um conceito interessante para possibilitar a sobreposição entre caracteres. Entretanto, dois pontos tornam este CAPTCHA muito fácil de ser quebrado:

1- A distância entre os caracteres é aleatória, então a sobreposição algumas vezes não acontece;

2- Devido ao fato de os caracteres sobrepostos serem exibidos em cores diferentes, o efeito de ofuscação pretendido pela sobreposição de caracteres torna-se simplesmente irrelevante.

Além disso, a legibilidade deste CAPTCHA não é muito boa. Por causa do exposto acima, recebe dificuldade 2.

Figura 3.7 – CAPTCHA UMC, de dificuldade 2

3.4.7 NNM.RU

Encontrado em http://www.nnm.ru/registration.html, este é um CAPTCHA interessante e bem feito, mas que possui vários pontos negativos. Podemos ver alguns exemplos dele na Figura 3.8. Antes de mais nada, as expressões aritméticas não são realmente um empecilho para um fraudador conseguir automatizar a resolução do CAPTCHA.

A existência de ruído é praticamente irrelevante. Todos os números estão escritos na mesma fonte e na mesma posição, o que facilita muito o trabalho do OCR. Dificuldade 2.

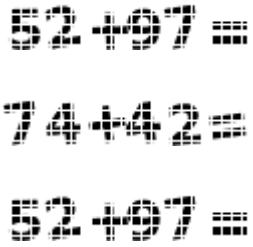

Figura 3.8 – O CAPTCHA do NNM.RU

3.4.8 Paypal.com

O CAPTCHA do Paypal apresenta todas as fraquezas clássicas inerentes aos CAPTCHA na Internet: modelo de cores ruim, ruído fraco e de fácil detecção, e tamanho padronizado de fonte. Por essa razão, dificuldade 2. Exemplos deste CAPTCHA podem ser encontrados na Figura 3.9 a seguir.

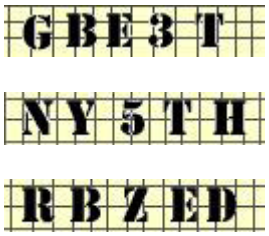

Figura 3.9 – CAPTCHA do Paypal.com

3.4.9 paid4load.de

Trata-se de um CAPTCHA de muito fácil automatização. O ruído, embora presente (observe a Figura 3.10), é relativamente fraco e o esquema de cores ajuda a identificar o código relevante. Assim, recebe a nota de dificuldade 2.

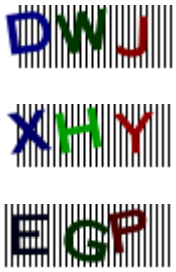

Figura 3.10 – Exemplos de Captcha do paid4load.de, dificuldade 2

3.5 CAPTCHA Clicáveis

Um problema crítico para se construir um CAPTCHA de Classe II é garantir um banco de dados com um conjunto suficientemente grande de informações que possuam certo grau mínimo de entropia. Para resolver este problema, o CAPTCHA ASIRRA (ELSON et al., 2007) propõe um acordo de alinhamento de interesses no qual utiliza um banco de dados externo, que possui uma grande quantidade de imagens, aumenta frequentemente de tamanho e é categorizado manualmente.

Sob essa perspectiva, tudo quanto pode ser armazenado em um banco de dados, de textos a vídeos, é uma fonte em potencial para a formação de CAPTCHA. Analisamos aqui dois tipos que foram explorados mais extensivamente na literatura.

3.5.1 ASIRRA

O ASIRRA é um CAPTCHA que solicita aos usuários que classifiquem fotografias em de cães ou de gatos. A força do ASIRRA vem de uma parceria com o site Petfinder.com (ELSON et al., 2007), o maior serviço de Internet responsável por achar adoções para animais abandonados.

O Petfinder tem um banco de dados de cerca de três milhões de imagens de cães e gatos, cada uma categorizada por voluntários humanos que trabalham em abrigos para animais nos Estados Unidos e Canadá. O alinhamento de interesses ocorre na medida em que o ASIRRA possui acesso ao banco de dados do Petfinder (que aumenta diariamente em cerca de 10.000 novas imagens) e utiliza-o para gerar desafios.

Em troca, o ASIRRA coloca um link com a frase “me adote” próximo a cada foto, promovendo a missão do Petfinder de expor animais para possíveis adoções. Claramente, tal parceria é mutuamente benéfica e promove o duplo papel social de melhorar a segurança em computadores e o bem-estar animal. Um exemplo de desafio ASIRRA pode ser visto na Figura 3.11.

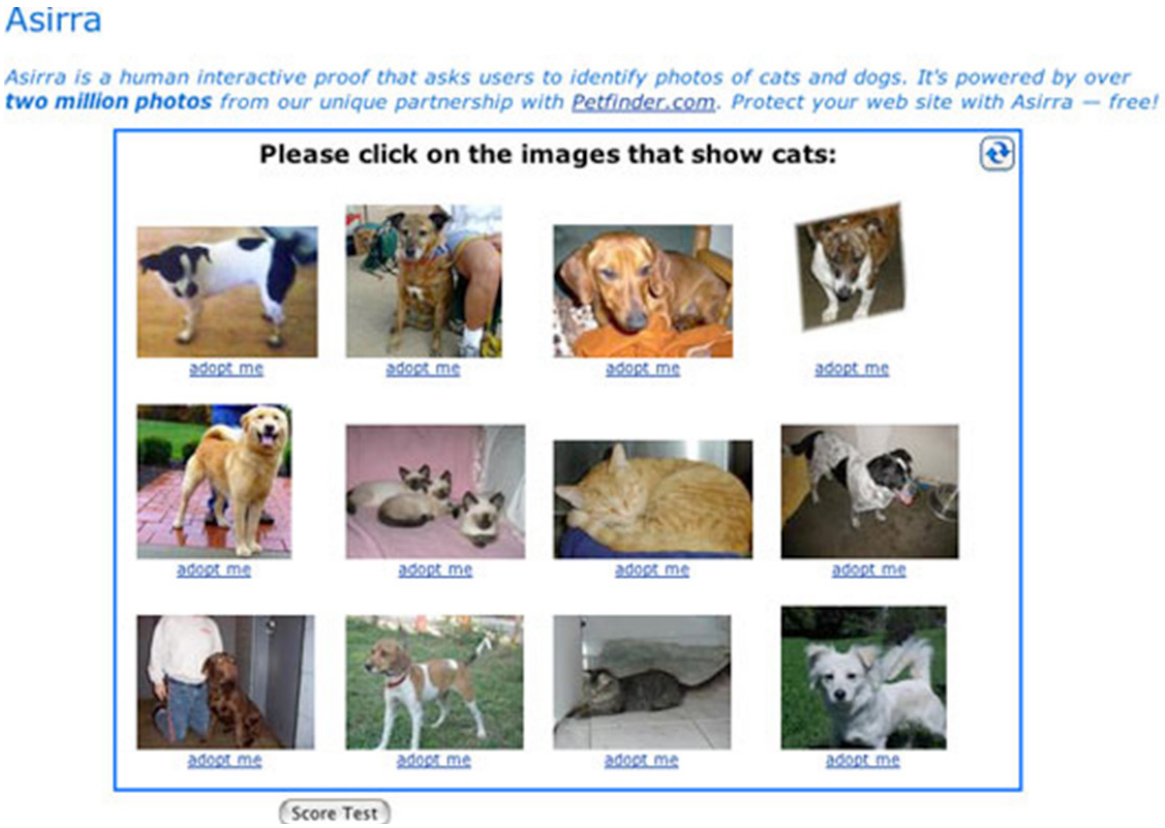

Figura 3.11 – Exemplo de CAPTCHA clicável (selecionar os gatos).

3.5.2 CAPTCHA clicáveis textuais

Basicamente, a idéia aqui é combinar vários CAPTCHA textuais em um *grid* de CAPTCHA clicáveis. A solução do desafio é a determinação, através do clique, dos elementos do *grid* que satisfaçam a um dado requisito. Por exemplo, o usuário pode

ser solicitado a identificar as opções que descrevam palavras na língua inglesa (a premissa é que há palavras de língua inglesa no CAPTCHA clicável, mas nem todas elas fazem sentido).

Enquanto CAPTCHA textuais tradicionais requerem a entrada da palavra a ser descoberta por técnicas de ofuscação, este CAPTCHA requer a seleção de alguns elementos por meio do mouse ou do teclado. Um exemplo de CAPTCHA clicável textual pode ser visto na Figura 3.12 (CHOW et al., 2008), no qual é solicitada a identificação das palavras em inglês (no caso: monster, grass e nation):

Figura 3.12 – Exemplo de CAPTCHA clicável textual

De acordo com Chow et al (2008), a segurança envolvida por trás de um CAPTCHA clicável textual é baseada na mesma premissa de CAPTCHA textuais tradicionais: a dificuldade de reconhecer caracteres distorcidos. Assim, o desenvolvimento de uma solução confiável de CAPTCHA clicáveis textuais depende da existência de uma forma de gerar CAPTCHA textuais fortes.

Ainda assim, CAPTCHA clicáveis textuais podem ser de mais fácil resolução para humanos, pois a tarefa cognitiva de decidir se uma palavra satisfaz a um determinado requisito (por exemplo, ser uma palavra na língua inglesa) normalmente não requer que toda palavra seja decifrada – é, portanto, uma tarefa mais fácil.

3.5.3 Benefícios no Uso de CAPTCHA

CAPTCHA clicáveis são, comumente, de fácil resolução para usuários humanos. Conforme Elson et al (2007), este tipo de CAPTCHA pode ser respondido por humanos em 99,6% das vezes em um tempo inferior a 30 segundos. De acordo com dados sobre os problemas de visão de máquinas, a probabilidade de os programas acertarem o desafio é de apenas 1/54.000 (ELSON et al, 2007). Além disso, apesar de o nível de segurança ser teoricamente o mesmo, os usuários de CAPTCHA clicáveis possuem a tendência de achar a sua experiência de uso bem mais interessante do que os CAPTCHA baseados em texto.

Podemos enumerar diversos benefícios no uso de CAPTCHA clicáveis:

- Humanos podem resolvê-los rapidamente e de maneira acurada;

- Ao contrário de diversos CAPTCHA baseados em imagem, que são abstratos ou subjetivos, os desafios são concretos, inofensivos, e não demandam qualquer tipo de conhecimento especializado ou cultural (no máximo, será necessário o conhecimento da língua na qual os desafios são gerados, em casos de CAPTCHA textuais). Isto torna os CAPTCHA clicáveis menos frustrantes para humanos, havendo até quem os julgue divertidos.

- Promove um benefício social adicional: achar habitação para animais desabrigados.

Podemos listar também algumas desvantagens:

- A maioria dos CAPTCHA é implementada como uma biblioteca de programas que pode ser integrada a um *website* sem introduzir dependências externas. Assim, os CAPTCHA clicáveis devem ser implementados como um serviço web centralizado, que gera e verifica desafios sob demanda para todos os usuários;

- Os CAPTCHA clicáveis têm uma dependência direta da confiabilidade de seu banco de dados – se um adversário, por exemplo, contratasse mão-de-obra barata para classificar todas as imagens de um banco de dados estático de CAPTCHA, seria muito difícil impedir o ataque.

- Um desafio típico de um CAPTCHA clicável como o ASIRRA requer mais espaço em tela e maior quantidade de dados trafegando que os CAPTCHA tradicionais baseados em texto.

- Em teoria, assim como todos os outros desafios deste tipo, são necessárias adaptações para que CAPTCHA clicáveis se tornem acessíveis para deficientes visuais. Uma das formas de fazer isso é adicionar uma versão do desafio em áudio.

Pelas razões descritas acima, o uso de CAPTCHA clicáveis é recomendado por este livro como forma de diferenciação entre usuários humanos e programas automatizados na utilização de serviços de marketing digital que objetivem prevenir a *click fraud*.

3.6 Considerações Finais

Este capítulo apresentou um estudo minucioso da tecnologia de CAPTCHA, algo não muito comum na literatura em língua portuguesa. Foram apresentados conceitos, históricos, aplicações e classificações de diferentes CAPTCHA presentes na Web. Acreditamos que o subconjunto de CAPTCHA analisados na seção 3.4 é uma representação satisfatória dos diversos tipos de CAPTCHA que podem ser atualmente encontrados na Internet.

Foram apresentados também conceitos de fundamental importância para este livro – o reCAPTCHA e os CAPTCHA clicáveis, sendo explicados os motivos pelos quais este livro recomenda utilizar os últimos como solução para garantir a

diferenciação entre humanos e computadores na abordagem que será proposta, no capítulo seguinte, para a prevenção da *click fraud*.

Existem vários outros métodos para a realização desta diferenciação. É de conhecimento público que os CAPTCHA normalmente confundem os usuários, inibem as chamadas "taxas de conversão" (WROBLESKI, 2008) e aumentam erros de navegação. Além disso, não são de fato uma garantia que fraudadores estarão definitivamente afastados dos serviços na Web.

Recentemente, Wrobleski (2010) apresentou em seu blog uma alternativa interessante para CAPTCHA. Ao invés de utilizar textos distorcidos, conforme a maioria dos CAPTCHA comuns, o formulário de registro do site They Make Apps (http://theymakeapps.com/users/add) usa um controle deslizante que pede que os novos usuários confirmem a sua "natureza humana" movendo o controle totalmente para a direita.

O usuário do site automaticamente submete o formulário e aciona os controles de erro, tal qual os típicos botões de "Submit" fariam. Um exemplo do formulário pode ser encontrado na Figura 3.13. Um estudo aprofundado e detalhado desse tipo de controle, comparando sua eficiência com a de utilização de CAPTCHA, é sugerido como um tópico para estudo futuro no capítulo final deste livro.

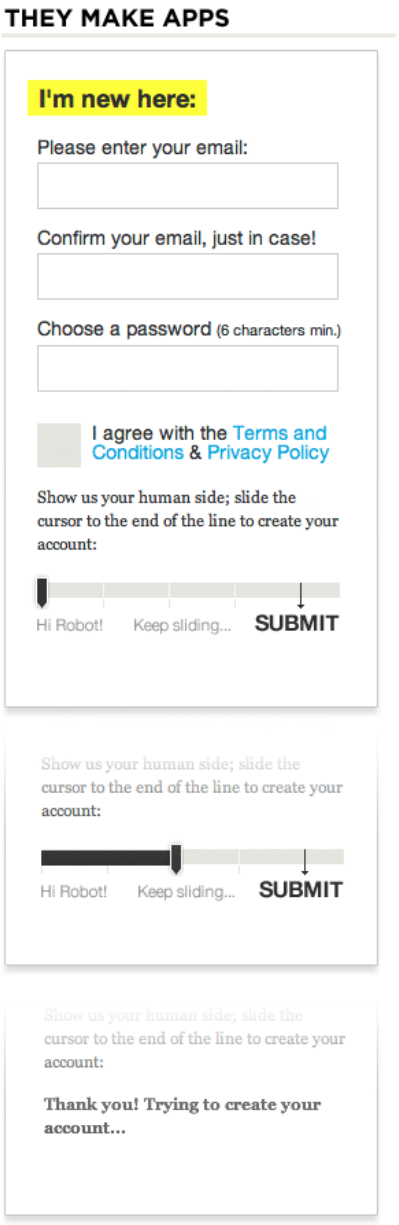

Figura 3.13 – Controle deslizante para mostrar o “lado humano” dos usuários

Embora não seja uma alternativa sem falhas, os números apresentados na seção 3.5 mostram que a utilização de CAPTCHA – mesmo aqueles que não são clicáveis – melhora sensivelmente a segurança dos sites pela diminuição da vulnerabilidade de seus formulários nos programas que os utilizam de forma automatizada.

A abordagem a ser proposta na seção 4 possui o compromisso com a segurança como principal motivação para prevenir a *click fraud* e como o objetivo primordial para a credibilidade da indústria de marketing digital , mas também busca a melhor forma de prover um serviço que não seja penoso para o usuário comum e acabe desestimulando o uso legítimo. Por causa disso, após o exposto nesta seção, a utilização de CAPTCHA clicáveis parece ser a melhor alternativa para satisfazer ambos os objetivos.

CAPÍTULO 4

O FUTURO DO MARKETING E DO MERCHANDISING DIGITAL PELA INTERNET

Este capítulo descreve a principal contribuição deste trabalho, a sugestão da abordagem C_2FAC_2A (Combatendo a *Click Fraud* Através de CAPTCHA Clicáveis e Autenticação). Primeiramente, serão demonstrados os elementos do processo de comunicação entre clientes e redes de anúncio, que constituem o esquema proposto. Depois, o fluxo de comunicação será derivado para a proposta de uma arquitetura de redes de anúncios segundo a abordagem, com foco especial aos serviços web para a comprovação da validade dos usuários. Após ambas as propostas, será realizada uma análise de segurança dos aspectos contidos na abordagem.

4.1 Introdução

Como já explorado anteriormente neste livro, as técnicas de mitigação da *click fraud* utilizam um tipo de combate denominado "defesa em profundidade" (METWALLY, AGRAWAL; ABBADI, 2005). O objetivo desse tipo de combate é adotar um conjunto de medidas tais que a complexidade e o custo de conduzir um ataque de sucesso contra uma rede de anúncios aumente a ponto de não haver benefícios econômicos na realização da fraude.

Entretanto, a abordagem proposta nesta pesquisa é direcionada pela crença de que o objetivo dos métodos de defesa da *click fraud* deve estar centrada na diminuição radical do risco de ocorrência de um ataque de sucesso, e possibilitar isso através de um sistema lucrativo para todos os participantes da rede de anúncios.

Embora possa parecer impossível prevenir que alguém realize um clique manual em um dado anúncio de maneira contínua, é possível adotar um conjunto de passos que evitem a ocorrência da fraude em larga escala, mais sistemática.

Para alcançar este objetivo, este capítulo apresenta a principal contribuição deste trabalho, a abordagem C_2FAC_2A (Combatendo a *Click Fraud* Através de CAPTCHA Clicáveis e Autenticação), no sentido de despertar a percepção dos profissionais e organizações envolvidas com a sistemática de marketing digital de que a prevenção à *click fraud* é um aspecto fundamental para o combate efetivo dessa.

Essa percepção contribui, no momento em que se elabora uma atividade de prevenção, tanto com os objetivos dos negócios de marketing digital quanto com a melhoria da área de Segurança da Informação na Internet.

4.2 Investigações e Objetivos

Em sua grande maioria, as pesquisas

realizadas nesta área investigam a fraude do publicador, já que ela pode ser generalizada para a fraude do anunciante, e discutem a detecção da fraude através de diversos métodos, tais como: a abordagem criptográfica (BLUNDO; CIMATO, 2002), técnicas de análise de dados (METWALLY; AGRAWAL; ABBADI, 2005), ferramentas para detecção de fraude (KLEIN, 1999), análise de tráfego (METWALLY; AGRAWAL; ABBADI, 2007) e algoritmos de força bruta (METWALLY; AGRAWAL; ABBADI, 2007).

Entretanto, todos esses métodos são técnicas de detecção, e tratam a fraude depois que ela ocorreu. Como já dito anteriormente, os programas têm se tornado cada vez mais complexos, e a detecção da fraude tem se tornado um problema de difícil resolução. Por essas razões, deve ser possível propor uma metodologia cujo objetivo é a prevenção de anunciantes, de maneira que a detecção se torne uma atividade complementar: isto é, utilizada apenas se necessária.

Dessa forma, é possível definir como um dos objetivos deste trabalho o preenchimento desta lacuna, através da definição de uma abordagem que seja focada primariamente na prevenção da fraude de clique.

Outro objetivo direto é o estudo minucioso da disposição do mercado de anúncios na internet, para garantir que o desenvolvimento de sistemas seguros possibilite o crescimento da confiabilidade deste mercado. Ainda é importante enfatizar que é também um dos objetivos deste trabalho a proposta de uma abordagem através de uma sólida linha de base técnica, por meio da definição de uma arquitetura que possa ser utilizada no desenvolvimento de sistemas seguindo esta abordagem.

4.3 Esquema Proposto

Em um esquema tradicional, quando um usuário clica em um anúncio localizado no site de um publicador, o anúncio correspondente é obtido do anunciante e a transação é gravada pelo representante. Depois disso, o representante cobra ao anunciante e paga ao publicador.

No modelo que propomos nesta pesquisa, esquematizado na Figura 4.1, há algumas tarefas adicionais a serem realizadas: quando o usuário clica em um anúncio (em um *banner* do representante localizado no site de um publicador, por exemplo) recebe do representante uma página Web que contém um CAPTCHA clicável.

Se o desafio for preenchido erroneamente, um novo desafio é proposto e o anúncio não é exibido. Quando o desafio for preenchido corretamente, um cupom com a comprovação de que se trata de um humano é embutido, pelo representante, no navegador do usuário e, após isto, o anúncio é finalmente exibido.

Sempre que um usuário, previamente autenticado, clicar em algum anúncio, este cupom será liberado para o representante. Essa liberação pode ser iniciada tanto pelo representante quanto pelo anunciante, e ativa outro processo, agora de validação, no representante.

Após essa validação e a confirmação que se trata de um cupom válido, o anúncio

é exibido. Para evitar a situação onde um usuário roda um programa a partir de sua estação com o objetivo de realizar a *click fraud* após uma autenticação manual, os cupons são válidos apenas por um certo período de tempo, após o qual se faz necessária uma nova autenticação.

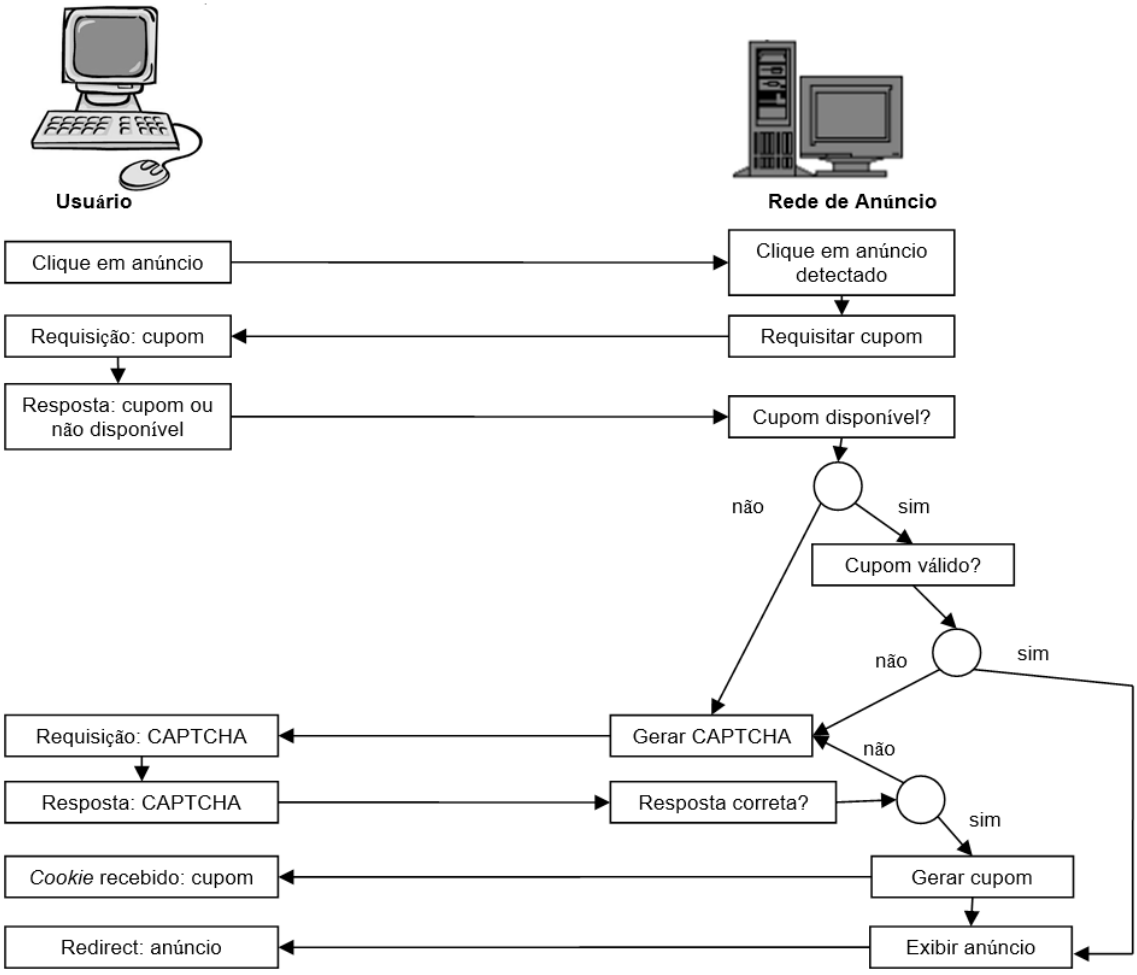

Figura 4.1 – Esquema proposto para autenticação em cupons.

É importante observar que os métodos atuais para filtragem não podem empregar o uso de *cookies* ou cupons na detecção de cliques fraudulentos. Isso acontece porque a filtragem é um processo exclusivo: se um *cookie* fosse usado para marcar e excluir um determinado tipo de usuário malicioso, fraudadores poderiam simplesmente remover os *cookies* dos seus navegadores.

Em contrapartida, a abordagem desta pesquisa é separativista, pois aceita apenas cliques bons e pode, assim, fazer uso de *cookies* ou cupons. Os cupons marcam os bons usuários. Essa é a mesma razão pela qual este método é de prevenção e não de detecção.

O esquema necessário para implementar essa solução é baseado na autenticação de requisições por meio de resposta a CAPTCHA, com o objetivo de distinguir os usuários humanos de usuários automatizados. Uma vez obtido o sucesso no teste, o usuário recebe um cupom, que é armazenado na forma de *cookie*. Por questões

de segurança, o cupom possui um tempo de validade, após o qual o usuário deve responder a um novo teste.

O cupom tanto pode ser armazenado localmente, na sessão dos usuários, quanto pode ser implementado em um mecanismo de autoridade central, sendo essa decisão um tópico para estudo futuro.

O esquema a ser proposto tem dois aspectos distintos:

Comprovador: um elemento arquitetural de extrema importância neste esquema é o serviço Web Comprovador, responsável por identificar e marcar os computadores de usuários humanos, após a resposta correta. A disposição técnica do Comprovador depende da forma como o cupom será armazenado: se a opção de armazenamento local for adotada, uma autenticação simples baseada em *cookie* pode ser utilizada; se o armazenamento em autoridade central for utilizado, será necessária a implementação de uma unidade autenticadora nessa autoridade central, que pode funcionar como um serviço Web em um dos servidores do representante.

Autenticação por cupons: este esquema possibilita a validação de cliques originados por clientes que ainda não produziram um histórico de tráfego de rede. Todo e qualquer cliente que não possua um cupom será considerado "não autenticado", e precisará responder um desafio antes de ter seu clique validado. Uma vez validado, o usuário será redirecionado para o site do anunciante.

Após o preenchimento correto do desafio, o usuário receberá um cupom para que não necessite responder novos desafios em requisições adicionais durante um determinado período de tempo, durante o qual o usuário será considerado "autenticado".

A razão para a proposta de validade se dá devido à necessidade de se evitar um cenário no qual um potencial fraudador autentique manualmente a primeira requisição e posteriormente rode um script local que realize a *click fraud*. É interessante notar também que esta abordagem facilita o método de detecção da *click fraud* por consulta ao histórico de conexão.

Sendo o clique considerado válido apenas se vier acompanhado de um cupom, detectar requisições múltiplas da mesma origem se torna fácil, uma vez que se armazene a apresentação dos cupons. Em abordagens tradicionais, nas quais os cupons não exercem nenhum papel, a detecção de tráfego de origem semelhante é mais difícil, e geralmente depende do mapeamento da origem dos dados, como mapeamento de IP, ou impressões de navegação, como identificadores de sessão.

4.3.1 Arquitetura Proposta

Levando em consideração a abordagem a ser proposta, este esquema de PPC é projetado de maneira a suportar a sua eventual terceirização, ou subcontratação. Ele deve garantir a segurança contra a *click fraud* quando os anúncios são publicados em ferramentas de pesquisa: nesse caso, é necessário tratar a rede de anúncios e os publicadores como a mesma entidade. Sob esta perspectiva, observa-se os seguintes

aspectos:

1. Marcação de cupons: baseado em seu critério para validação de usuários, o Comprovador identifica um vistante como legítimo ou não, e faz isso através da marcação, na cache do navegador do usuário, de um cupom c, que é uma espécie de *token* criptográfico.

2. Liberação de cupons / geração de desafios CAPTCHA: quando o usuário clica em um anúncio no site de um publicador, o navegador do usuário é direcionado para uma URL no site da rede de anúncios. Essa URL carrega informações tais como a identificação do publicador, ID_{pbc} e a identificação do anúncio que foi clicado, ID_{anc}.

A rede de anúncios deve então responder à requisição do usuário, buscando a identificação de um cupom válido existente no seu navegador. Se não existir, o usuário precisará responder um desafio CAPTCHA que será gerado pela entidade "Gerador de CAPTCHA".

Ao responder corretamente, o serviço Comprovador será atividado para liberar um cupom c ao navegador do usuário, simultaneamente com ID_{pbc} e ID_{anc}, além de informações acerca da validade v do cupom. Sendo $C = (c, ID_{pbc}, ID_{anc}, v)$, de agora em diante, C será denominado "cupom", de acordo com o contexto.

3. Verificação de cupons: Ao receber $C = (c, ID_{pbc}, ID_{anc}, v)$, é responsabilidade do Comprovador verificar que c é um cupom criptograficamente bem formado, como será descrito posteriormente.

Além disso, o Comprovador verificará que C ainda não expirou (caso tenha expirado, será necessária a resolução de um novo desafio CAPTCHA) e que o cupom não foi usado excessivamente em um passado recente (a definição de uso "excessivo" e do tempo de validade de cupom é uma decisão de responsabilidade da rede de anúncios).

4. Recompensa: se o Comprovador verificar que C representa um clique válido, então a rede de anúncios irá pagar ao publicador e cobrar ao anunciante adequadamente. É importante observar que uma nova cobrança será disparada cada vez que um cupom for usado.

A partir dessas premissas, podemos gerar a arquitetura detalhada na Figura 4.2 para implementar estes aspectos.

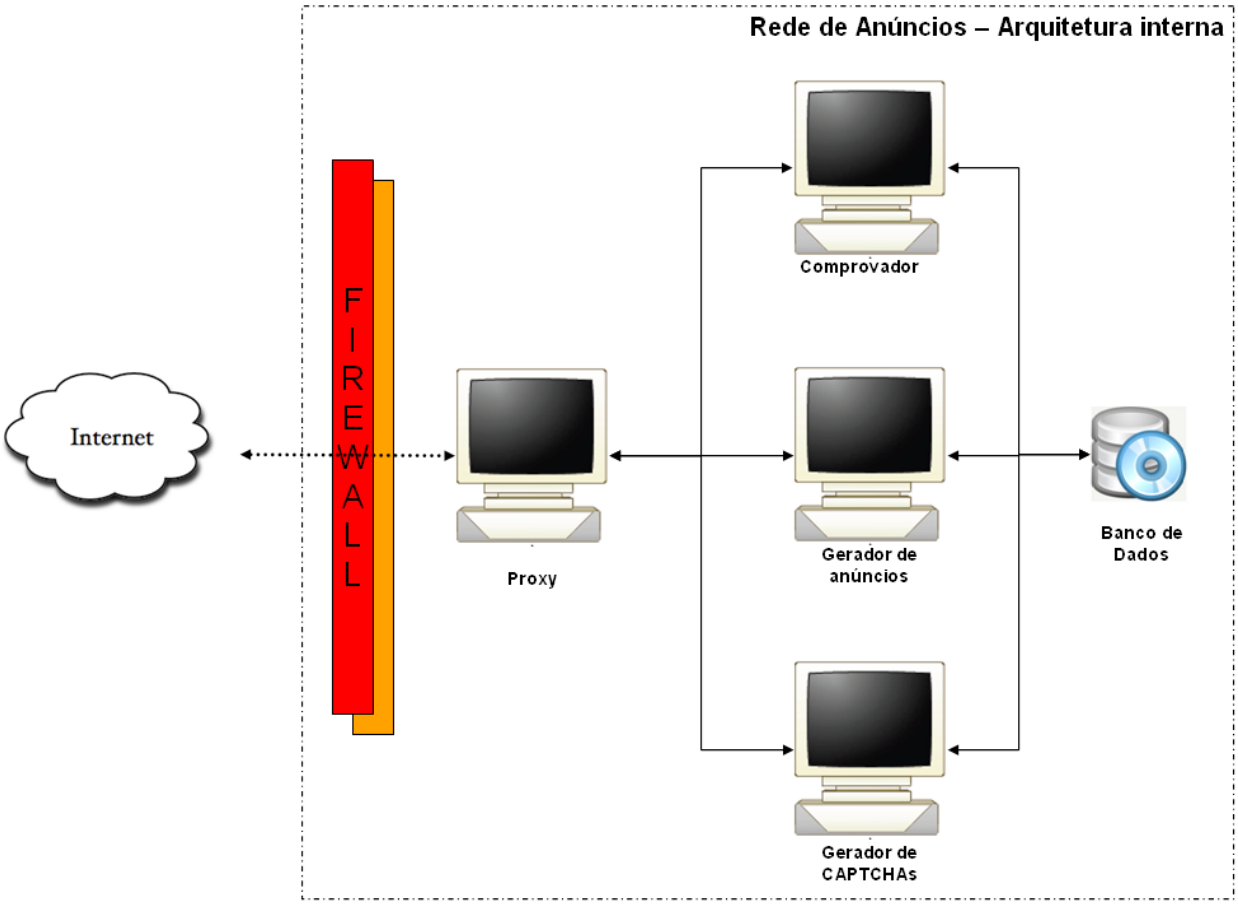

Figura 4.2 – Arquitetura proposta para a abordagem C_2FAC_2A

Obviamente, o publicador possui a opção de incluir informações adicionais em C. A razão principal para isto é que um mesmo computador pode ser usado por várias pessoas. Se cada uma das pessoas acessa a Internet de suas contas pessoais e de suas próprias instâncias de navegadores no computador do cliente, então essas instâncias carregarão cupons específicos. Nesta seção, serão detalhadas as especificidades técnicas da abordagem proposta.

4.3.1.1 Banco de Dados

Um componente chave em qualquer abordagem de pagamento por clique é o banco de dados que armazena informações dos anunciantes.

Em toda abordagem clássica de sistemas que possuam esse fim, é necessário o armazenamento de dados dos anúncios (conteúdo, identificador dos anunciantes e URL), o histórico de navegação dos usuários (ou seja, cada clique ocorrido, que inclui ID_{pbc} e ID_{anc}) e o registro de cupons recebidos. Como, por questões de auditoria, é necessário o armazenamento do histórico dos cupons recebidos e liberados pela rede de anúncios, é recomendado considerar a utilização de um banco de dados confiável na rede de anúncios, conforme detalhado na Figura 4.2.

Este livro propõe ainda a utilização de um banco de dados único (escolha que pode ser modificada por uma política interna) para conter as informações que serão utilizadas na geração dos CAPTCHA clicáveis pela entidade Gerador de Anúncios.

Isto será influenciado pela metodologia de alimentação do CAPTCHA clicável, a ser detalhada na seção 4.3.1.3.

Não é foco deste livro discutir aspectos técnicos da modelagem dos dados ou suas possíveis disposições em uma rede de anúncios, nem realizar uma análise de *benchmarks* a respeito das metodologias de armazenamento de anúncios. Acreditamos que, com o exposto nesta seção, fornecemos requisitos mínimos para que a abordagem seja implementada com sucesso.

Esse aspecto, embora seja extremamente relevante e constitua um co-requisito para a existência desta solução, não faz parte do escopo deste trabalho, de maneira que deixaremos o aprofundamento neste tópico de estudo como sugestão para pesquisas complementares ou para trabalhos posteriores.

4.3.1.2 Gerador de Anúncios

Nesta seção, serão apresentados os aspectos por trás da entidade Gerador de Anúncios. O desenvolvimento deste serviço é uma parte fundamental na existência de qualquer rede de marketing digital , e a sua implementação tem sido estudada extensivamente ao longo dos últimos anos.

Limitar a especificação que se deve seguir para o desenvolvimento do Gerador de Anúncios seria pouco inteligente, já que o desenvolvimento do mesmo precisa ser uma função do modelo de negócios adotado pela rede de anúncios. Por exemplo, há redes de anúncios que “leiloam”, entre os seus anunciantes, as melhores posições nos *banners* de anúncio. Em contrapartida, há redes que preferem ordenar os seus anúncios com base em estatísticas como *Page Ranks*, para dar melhores oportunidades a anunciantes menos visitados, promovendo o desenvolvimento de suas marcas.

Há ainda as redes que tratam todos os anunciantes igualmente, realizando um rodízio nas exibições.

Por causa das inúmeras formas de implementação desse serviço, este livro se limitará a listar as responsabilidades do Gerador de Anúncios para que a realização desta abordagem se torne possível, bem como uma descrição de um modelo de dados simples, mas suficiente para possibilitar uma implementação mínima da nossa proposta.

4.3.1.2.1 Responsabilidades

A primeira responsabilidade do Gerador de Anúncios é construir o *banner* contendo os anúncios que serão exibidos em um dado momento por um publicador. Essa instância de anúncios poderá ser montada a partir de uma busca, devido a um histórico de navegação, por exemplo. Por causa disso, é recomendável que cada abordagem defina uma entrada comum, que será utilizada para realização da busca por anúncios.

O modelo básico de busca se dá de acordo com palavras-chave. É importante lembrar que o *banner* conterá links para o site da rede de anúncios, contendo identificadores para redirecionamento.

A segunda responsabilidade está associada à seleção dos anúncios que foram retornados pela busca. Isso caracteriza uma busca heurística por dados relevantes e o ranqueamento destes dados. Os anúncios melhor ranqueados serão exibidos, de acordo com a requisição do publicador. Como dito anteriormente, é impossível para esta pesquisa propor uma abordagem para esse fim que satisfaça os inúmeros modelos de negócios conhecidos.

A terceira responsabilidade do Gerador de Anúncios está associada ao fato de essa entidade ser responsável pelo acesso ao banco de dados de anúncios (interface de consulta, atualização, e cadastramento de novos anúncios). É comum, em implementações de redes, que os anunciantes possuam uma interface para editar os seus anúncios, a qual pode rodar como um serviço web no Gerador de Anúncios. Na realidade, a implementação deste serviço é muito simples.

Para possibilitar a realização das atividades listadas, o banco de dados precisa implementar um conjunto de dados que será discutido a seguir.

4.3.1.3 Modelo de Dados

A forma mais simples de catalogar anúncios em um banco de dados que satisfaça esta abordagem é armazenar no banco de dados a informação necessária para suportar a execução do serviço Gerador de Anúncios. Para esta abordagem, isso acontecerá usando-se duas tabelas.

A primeira tabela, denominada AnunciosBanner (Tabela 4.1), necessita conter informações sobre os anúncios, como a localização das imagens, a URL do site do anunciante e o texto a ser exibido pelo anúncio. Os campos associados a essa tabela são:

- IDAnuncio: a chave primária da tabela;
- ImagePath: armazena a localidade da imagem asssociada a este anúncio (que pode ser local ou externa, como /images/ads/bannerXX.gif);
- URL: link para o qual o anúncio redireciona o usuário;
- AltText: texto que o usuário verá ao mover o ponteiro para cima do anúncio.

Nome do Campo	Tipo	Tamanho
IDAnuncio	AutoNumber	
ImagePath	Text	100
URL	Text	100
AltText	Text	255

Tabela 4.1 – AnunciosBanner

A segunda tabela necessária é a ImpressoesBanner, vista na Tabela 4.2, que

contém informações sobre o número de impressões de um determinado anúncio. Sempre que o anúncio é mostrado, um registro é adicionado à tabela. Os campos associados à mesma são:

- IDImpressao: chave primária da tabela.

- IDAnuncio: chave externa que relaciona essa tabela à tabela AnunciosBanner em uma relação “um para muitos”.

- DateClicked: data e hora em que o anúncio apareceu. Este campo pode ser complementado por outros que ordenem os anúncios. Na implementação mais simples possível, a ordenação se dá pela rotação de anúncios.

Nome do Campo	Tipo
IDImpressao	AutoNumber
IDAnuncio	Number
DateClicked	DateTime

Tabela 4.2 – ImpressoesBanner

Como se pode ver, trata-se de um modelo de dados simples, cuja principal intenção é dar uma idéia das informações que serão necessárias para possibilitar a implantação desta abordagem. Certamente, em um ambiente de produção serão necessárias muitas outras nuances que implicarão no uso de tabelas e campos adicionais.

4.3.1.4 CAPTCHA Clicável

Inicialmente, esta abordagem recomendaria a utilização de CAPTCHA clicáveis baseados em imagem para funcionarem como autenticação de usuários para o sistema. Conforme discutido na seção 3.5, tais CAPTCHA tendem a ser bem menos vulneráveis, em termos de automatização da resolução, a ataques de fraudadores.

Entretanto, fatores como a alimentação destas imagens e a necessidade de um banco de dados externo ou extremamente grande para armazená-las, tornam necessária a sugestão de um método alternativo para a exequibilidade desta arquitetura. Nesta seção, exploraremos também a possibilidade da utilização de CAPTCHA clicáveis textuais como parte da solução para esse problema.

A geração de um CAPTCHA é realizada por um código-fonte, que pode ser compartilhado e integrado em sites de Internet sem dependências externas. Entretanto, os CAPTCHA clicáveis em geral, especialmente os baseados em imagem, necessitam da integração com um banco de dados que deve ser mantido em segurança. Para proteger inteiramente o banco de dados, a melhor opção parece ser implementar o serviço Gerador de CAPTCHA como um *web service*.

4.3.1.4.1 Gerador de CAPTCHA como um Web Service

O serviço precisa ser dividido em duas categorias: o serviço confiável, que roda internamente na rede de anúncios, e o componente de cliente não confiável, rodado nos computadores dos usuários. Levando isto em consideração, faz-se necessário o estabelecimento de um passo de validação de integridade do desafio resolvido.

Esse processo pode ser feito de maneira transparente. Uma vez que o usuário visualize, complete e submeta o formulário com a resolução do CAPTCHA, pode-se definir um campo de *input* no formulário HTML denominado Ticket_Verificador. Esse ticket é uma *string* de bits que será validada pelo web service para garantir que o cliente não o forjou nem o está reutilizando.

A validação é simples: o ticket é passado para uma URL de validação. O serviço retorna "Passou" apenas se o ticket for válido, recente e nunca tiver sido validado anteriormente. O JavaScript que gerencia essa interação impede que o formulário seja submetido até que um ticket válido seja recebido. Dessa forma, os webmasters verão falhas de validação apenas se o serviço estiver sob ataque.

4.3.1.4.2 Desenvolvimento

A interação entre o cliente e o serviço web pode ser realizada por meio da utilização de AJAX (requisições HTTP assíncronas executadas dentro do JavaScript). Quando o JavaScript é carregado, cria os elementos visuais do desafio no navegador do usuário. A seguir, manda requisições ao servidor web para criar uma nova sessão, recebe as respostas de um ou mais desafios, e submete as respostas para verificação.

O serviço web receberá a resposta do usuário como uma requisição AJAX (algo como: http://redeanuncios/respostadesafio.php?desafio=z&resp osta=a[1]+a[2]+...+ a[n], sendo a[1...n] o vetor contendo os itens com as respostas assinaladas como corretas) e verificará, em tempo de execução, a corretude da resposta.

Caso a resposta esteja errada, um novo desafio será exibido. Se a resposta for correta, o cliente receberá um ticket de serviço e o Gerador de CAPTCHA iniciará uma requisição, em nome do usuário, junto ao serviço Comprovador, para a definição de um novo *cookie*.

Caso prefira, o usuário pode escolher vizualizar outro desafio. Essa solução de implementação pode utilizar tanto CAPTCHA clicáveis baseados em imagem ou em texto.

4.3.1.4.3 CAPTCHA Clicável Baseado em Imagem

A metodologia para o CAPTCHA clicável baseado em imagem é simples. Semelhantemente ao ASIRRA (ELSON et al., 2007), pode-se ter um *grid* com 15 imagens de animais e a solução do desafio ser a identificação de 3 imagens de gatos. Teremos então um desafio com 3 soluções e 12 erros. O algoritmo para seleção de

imagens é extremamente simples:

1) Selecionar aleatoriamente o tipo das imagens que gerarão desafios corretos (exemplo: imagens de gatos);
2) Selecionar aleatoriamente 3 imagens de gatos da base;
3) Selecionar aleatoriamente 12 imagens que não sejam de gatos da base;
4) Criar um conjunto com as 15 imagens e ordená-las aleatoriamente no *grid* que será exibido ao usuário.
5) Gerar um identificador do desafio.

Como explicado na seção 4.3.1.1, assumimos, para efeitos de descrição desta abordagem, que o banco de dados utilizado para a alimentação do CAPTCHA clicável, caso esta opção seja selecionada, será interno à rede de anúncios. Naturalmente, será necessária a catalogação das imagens que alimentarão os desafios e a criação de um atributo que será utilizado para distinguir a natureza das imagens (gatos, cachorros, pássaros, entre outros).

Entretanto, é de extrema utilidade para esta abordagem que aqueles que optarem pela utilização do CAPTCHA clicável baseado em imagem possam realizar o gerenciamento da catalogação dos dados de maneira externa à rede de anúncios. A sugestão, assim como o ASIRRA (ELSON et al., 2007), é a realização de uma parceria com benefícios mútuos entre a rede de anúncios e um banco de dados externo.

4.3.1.4.4 CAPTCHA Clicável Baseado em Texto

No caso de CAPTCHA baseados em texto, é necessária a criação de um conjunto de CAPTCHA (por exemplo, 12) dos quais um subconjunto (por exemplo, 3) represente algo que faça sentido em um determinado contexto (por exemplo, palavras retiradas de um dicionário). Teremos então um desafio com 3 soluções e 9 erros. O algoritmo para a criação de cada um dos textos distorcidos nos 12 CAPTCHA deve ser desenvolvido e rodar como um serviço no Gerador de CAPTCHA.

A implementação é extremamente semelhante ao pseudo-código disponibilizado anteriormente. Não é foco desta pesquisa entrar em detalhes acerca da implementação do CAPTCHA clicável baseado em texto (Chow (2008) utiliza em sua solução o CAPTCHA do Google), embora, caso essa opção seja escolhida, sua implementação seja de extrema relevância para a existência desta abordagem.

Para resolver o CAPTCHA clicável, o usuário precisará clicar exclusivamente nas 3 células que contenham soluções, sem que importe a ordem. Um usuário que possua familiaridade com a linguagem do dicionário irá resolver o desafio de maneira quase intuitiva. Para impedir que programas automatizados passem a identificar facilmente as soluções dos desafios, pode-se gerar as palavras erradas por meio de transformações de palavras retiradas do dicionário.

4.3.1.5 Cookies

Como já adiantado no capítulo 1, não é foco deste livro o estudo aprofundado da tecnologia de *cookies*, no sentido da análise de suas vantagens e desvantagens, ou de alternativas ao seu uso. Este tipo de estudo pode ser encontrado em trabalhos como (KING, 2008), (SAMAR, 1999), (TEIXEIRA, 2004) e (WROBLEWSKI, 2008), que serão utilizados como base da proposta aqui apresentada.

Assim, o primeiro passo para o prosseguimento de nossa proposta é a definição do tipo de *cookie* recomendado, além da apresentação de como eles serão utilizados pelo mecanismo de autenticação. Por essa razão, esta seção será utilizada para apresentar uma rápida conceituação do que sejam *cookies*, e para discutir a disposição da tecnologia de *cookie*s nesta arquitetura.

Um *cookie* é uma *string* (um pedaço de texto) armazenada no navegador web de um usuário de Internet, e pode ser utilizado para diversos fins: autenticação, armazenamento de preferências de navegação, identificação de sessão ou qualquer outra finalidade para se armazenar dados de navegação no formato texto. De acordo com Samar (1999), todo e qualquer *cookie* precisa ter cinco parâmetros: "nome do *cookie*, valor, data de validade, URL para a qual o *cookie* é válido e se o *cookie* deve ser enviado apenas através de uma conexão segura por SSL". A maioria dos navegadores modernos dá aos usuários a opção de aceitar ou não *cookies*.

O mecanismo de troca de mensagens cliente-servidor de *cookies* se dá da seguinte forma: inicialmente, o *cookie* é enviado como um cabeçalho HTTP pelo servidor web para o navegador, que, durante o período de validade do *cookie*, o envia de volta ao servidor sempre que ele for acessado.

Como são textuais, os *cookies* não podem ser executados e, assim, não podem se auto-replicar como vírus. Entretanto, por causa do mecanismo de navegação, que utiliza funções de atribuir e ler *cookies*, eles podem ser utilizados como *spyware* (DOYLE, 2010), sendo este o motivo pelo qual alguns softwares *anti-spyware* alertam os usuários da presença de *cookies*.

A primeira escolha técnica de nossa abordagem é o meio de transporte dos cupons. Para garantir a associação correta de um *cookie* com o navegador que o criou, o cupom deve ser comunicado na forma de valor armazenado no navegador. Ao mesmo tempo, é importante garantir que os cupons sejam configurados de forma a serem recuperados apenas pela rede de anúncios e a não poderem ser coletados por fraudadores.

Cookies de terceiros são a forma mais óbvia de se instanciar cupons. Entretanto, por terem um histórico de uso abusivo, é muito comum tais *cookies* serem bloqueados por usuários. Uma alternativa para contornar esse problema é a utilização de *cookies* privados.

Uma rede de anúncios pode implantar cupons privados para o seu uso particular, caso o Comprovador use mecanismos de identificação específica por usuários

(ou específica por seções). Entretanto, o uso de *cookies* privados pode gerar uma complexidade indesejada se forem necessários redirecionamentos em múltiplos passos, como no caso do fim da validade de um dado cupom.

Para resolver tanto a questão da complexidade quanto contornar o bloqueio de *cookie*s por parte dos usuários, *cache cookies* (JUELS; JAKOBSSON; JAGATIC, 2006) oferecem uma opção interessante. Nesse caso, um comprovador pode embutir um cupom, que é marcado como de leitura exclusiva pelo servidor proxy da rede de anúncios.

Como podem ser configurados para sites de terceiros, *cache cookies* se assemelham a *cookies* de terceiros, mas possuem uma característica especial: qualquer web site pode causar a liberação do *cache cookie* para o site ao qual está associado (a importância de autenticar o site que inicia a liberação do cupom a partir do navegador do usuário será explorada posteriormente).

Além disso, *cache cookies* funcionam em navegadores que bloqueiam o uso de *cookies* comuns, razão pela qual esta abordagem recomenda sua utilização.

Em síntese, um *cache cookie* funciona da seguinte forma: para definir um *cache cookie* associado ao site www.X.com, o *cache cookie* assume a forma de uma página HTML ABC.html que requer um recurso de www.X.com com o seu valor c associado. Por exemplo, a página ABC.html pode requisitar um arquivo texto na forma http://www.X.com/c.ini.

Observe que qualquer web site pode criar um ABC.html e implantá-lo no navegador de um visitante. Similarmente, qualquer web site que conheça a página/*cache cookie* ABC.html pode referenciá-la, causando o recebimento, por parte de www.X.com, de uma requisição pelo arquivo texto c.ini. Entretanto, apenas www.X.com pode receber o *cache cookie*, isto é, o valor “c.ini”, associado ao *cache cookie*, quando o mesmo for liberado.

4.3.1.6 Autenticação por Cupons

Garantir o combate à criação e ao uso de cupons fraudulentos é um dos desafios de nossa abordagem. Apenas os serviços do tipo Comprovador devem ser capazes de construir cupons válidos. Para possibilitar isto, os cupons precisam conter alguma forma de autenticação criptográfica.

Na medida em que assinaturas digitais podem, a princípio, oferecer uma forma flexível de autenticar cupons, seus custos computacionais são provavelmente proibitivos para o tráfego de massa. Códigos de autenticação de mensagens (*message authentication codes*, ou simplesmente MAC), que são chaves simétricas análogas a assinaturas digitais, podem ser consideradas uma alternativa computacionalmente mais econômica.

Suponhamos que um Comprovador A compartilhe, por meio de canais seguros, uma chave simétrica *k* com um determinado usuário. Sendo MAC_k (m) a representação

de um MAC forte, como o HMAC (BELLARE; CANETTI; KRAWCZYK, 1996), computado a partir de uma mensagem bem formada m, é teoricamente impossível para qualquer participante extra (por exemplo, um adversário) gerar um novo MAC a partir de qualquer mensagem m.

Consequentemente, se um cupom c assume a forma c = m ll MAC_k (m) para uma sequência de bits m que é única a cada visita de um cliente ao site de uma rede de anúncios, então o cupom pode ser copiado, mas não satisfatoriamente modificado pelo participante extra. Nesse cenário, o valor m deve ainda ser sensivelmente longo (por exemplo, 128 bits).

4.3.1.7 Identificação e Autenticação de Publicadores

Além de garantir que o cupom é autêntico, o Comprovador deve determinar também qual publicador causou a sua criação/liberação, de maneira que este receba o pagamento para o clique associado. Como visto anteriormente, um cupom possui a forma C = (c, ID_{pbc}, ID_{anc}, v), onde ID_{pbc} é a identificação do publicador e ID_{anc} identifica o anúncio clicado. Para criar um novo cupom, deve-se adicionar os valores ID_{pbc}, ID_{anc} e v ao valor de c, quando ele for liberado. Para isso, pode-se estender a página HTML que contém o *cookie* para incluir o referenciador do documento e a validade do *cache cookie*. Nesse esquema, a forma de inclusão se dá pela requisição de uma URL na rede de anúncios, onde todas essas informações (c, ID_{pbc}, ID_{anc} e v) são passadas na URL.

Por exemplo, o *cache cookie* pode ter a seguinte forma:

```
<html><body>
<script language="JavaScript">
//Determina a página r referenciada que contém ID_pbc e ID_anc:
var r = escape(document.referrer);
var validade = escape(document.expirationDateTime);
//Escreve o HTML para liberar o cupom c.ini
document.write('<a href="http://www.X.com/'
+ 'c.ini?ref=' + r + '+ '&validade=' + validade + '"/>');
</script> </body> </html>
```

Quando a página da rede de anúncios, com uma URL contendo ID_{pbc} e ID_{anc}, referencia o arquivo HTML, o servidor proxy da rede recebe uma requisição pelo seguinte recurso: **c.ini?ref=www.S.com%3fad%3dhIDadi%26pub%3dhIDpub&validade=86500100** (o valor da *string* ref na requisição é o referenciador do documento, ou a página que requisitou o arquivo HTML, mas codificada para que possa aparecer na URL).

Essencialmente, a rede de anúncios recebe uma requisição pelo arquivo texto c.ini juntamente com uma descrição de parâmetros que contém os identificadores do publicador e do anunciante envolvidos na requisição. Essa *string* passa para o

Comprovador todos os dados necessários ao cupom: C = (c, ID_{pbc}, ID_{anc}, v).

Em casos onde o JavaScript esteja desabilitado pelo usuário, uma abordagem alternativa é possível. O Comprovador pode criar, ao invés de um *cache cookie*, um *array* de *cache cookies* de valores criados de maneira independente c1[0]...ck[0] e c1[1]...ck[1]. Para codificar um publicador de k-bits, ID_{pbc} = b1 ll ... ll bk, o publicador libera um *array* de *cache cookies* correspondente a c1[b1]...c[bk].

Claramente, esse método é mais complicado do que o uso de *strings* contendo os referenciadores do documento, já que requer do Comprovador o recebimento e a correlação um número k de *cache cookies* distintos para uma transação simples.

4.3.1.8 Verificação de Validade

Somente a autenticação não é suficiente para garantir o uso válido de cupons. É necessário garantir que os mesmos sejam válidos, ou seja, que o usuário não o esteja usando mais rapidamente do que o natural. Para garantir a validade dos cupons, vimos que o *cache cookie* carregará uma informação a respeito da sua validade.

Quando um cupom C = (c, ID_{pbc}, ID_{anc}, v) for recebido em uma hora h, o Comprovador deve verificar se v < h e rejeitar C, iniciando uma nova requisição de CAPTCHA juntamente ao Gerador de CAPTCHA.

Além disso, nesta etapa, podemos incluir mais uma validação para cupons: a prevenção da geração de cobranças por cliques realizados repetidamente. Quando um cupom C = (c, ID_{pbc}, ID_{anc}, v) for recebido em um tempo t, o Comprovador pode verificar a base de dados, procurando se uma requisição semelhante já existe, com o tempo t(i). Se t – t(i) < Trepetição, para um determinado parâmetro de sistema “Trepetição”, então o Comprovador pode rejeitar C por causa de uma repetição excessiva.

Obviamente, muitas outras políticas de filtragem são possíveis, não sendo a sua exploração foco desta pesquisa.

4.3.1.9 Serviço Web Comprovador

Na forma mais simples de se implementar um serviço web Comprovador, após responder o CAPTCHA clicável corretamente, o usuário é direcionado a uma página interna que serve um *cache cookie* para o navegador do visitante. Este arquivo HTML simples é transmitido apenas uma vez após a resposta do CAPTCHA.

Enquanto o *cache cookie* for válido, quaisquer requisições do usuário à URL que lhe proveu o *cache cookie* retornará uma mensagem HTTP 304 “*not modified*”. Isso força o navegador a usar a versão armazenada do *cache cookie*, caso exista, e impede que outros navegadores na rede do usuário utilizem a mesma versão armazenada em algum outro navegador a partir de qualquer computador da rede. Impede também que requisições manuais, realizadas fora do processo padrão, consigam realizar parte do processo. Ou seja, cada instância do navegador é tratada como um cliente distinto.

O *cache cookie* servido pelo Comprovador referencia um arquivo texto armazenado no servidor proxy da rede de anúncios, cupom.ini. A URL usada para requisitar o cupom pode ser criada por um JavaScript no momento em que o *cache cookie* é elaborado, e contém uma chave secreta *k* (que é gerada quando o *cache cookie* for definido), a URL referenciada que revela o ID do publicador e o ID do anúncio que foi clicado sempre que o *cache cookie* for carregado juntamente com um clique, e uma informação de data e hora que revela a validade do cupom.

O Comprovador precisa criar e servir o *cache cookie* quando o usuário for autenticado, de maneira que, em um primeiro momento, é necessário um algum processamento adicional no servidor, como veremos a seguir. Entretanto, criar a chave secreta é um processo muito simples, e o *cookie* pode ser servido em um *iframe* escondido e embutido na página retornada. Em termos de experiência do usuário, o resultado não deve ser visível, e os servidores da rede de anúncios precisarão executar um trabalho adicional mínimo.

De todas as entidades, o Comprovador realiza a maior quantidade de trabalho. Ele recebe os cupons liberados pelos *cache cookies* (na forma de requisições textuais), verifica as chaves secretas nos cupons originários de usuários, armazena os cliques no banco de dados e ainda faz verificações de validade dos cupons.

Além disso, cada clique deve passar pelo Comprovador, então ele também serve como um servidor de transferência para direcionar o navegador do usuário para o site do anunciante, e aciona o serviço de geração de CAPTCHA quando necessário. Essa sistemática caracteriza um fluxo de tráfego de dados clássico nos servidores de uma rede de anúncios que, por diversos momentos, delega o controle da rede para o Comprovador, para que ele enderece os cliques.

Em termos de processamento de cliques, quando um anúncio é clicado, o navegador do cliente é direcionado para o site da rede de anúncios, carregando consigo o identificador do anúncio e o identificador do publicador. Algo semelhante a http://redeanuncios/clique.php?anuncio=x&publicador=y.

O Comprovador então armazena o clique (neste ponto, o Comprovador ainda não sabe se o clique é válido – mesmo assim, cliques inválidos são registrados), o marca como “inválido” e responde com uma página web que contém um *iframe* escondido que tem como objetivo carregar o *cache cookie* no navegador do cliente. Essa página web carrega as informações do anúncio e do publicador como variáveis locais geradas dinamicamente em JavaScript. A partir deste ponto, podem-se seguir dois cenários:

1) *Cache cookie* não identificado

Caso nenhum *cache cookie* seja encontrado no navegador do usuário, o código JavaScript embutido na *iframe* irá solicitar uma nova página (semelhante a http://redeanuncios/getCaptcha.php?anuncio=x&publicador=y) ao servidor proxy da rede de anúncios, que direcionará a requição para o Gerador de CAPTCHA.

O Gerador de CAPTCHA então iniciará o processo descrito na seção 4.3.1.3, dependendo da escolha da forma de implementação do CAPTCHA, com o objetivo de

identificar a resposta correta do desafio.

Caso a resposta do CAPTCHA seja correta, o Gerador de CAPTCHA irá então passar uma requisição para o serviço Comprovador, algo como http://redeanuncio/cupom.ini?ref=x, em nome do usuário, que irá então gerar uma chave *k* para o *cache cookie* a ser gerado. A variável ref na URL reflete o identificador do publicador, isto é, ID_{pbc} e o identificador do anúncio clicado, ID_{anc}.

Essas informações, juntamente com a validade do novo cupom (que consistirá na data e na hora em que a requisição foi recebida, adicionada do tempo de validade – por exemplo, 30 minutos) serão armazenadas no *cookie* que será retornado ao cliente.

O Comprovador então armazena o horário, a chave secreta *k* e a referência x no banco de dados, além de atualizar o campo de *status* do clique previamente armazenado para "válido" (para disparar, no futuro, uma cobrança junto ao anunciante e uma compensação para o publicador) e, finalmente, servir o arquivo gerado de volta ao cliente, que o armazenará como um de seus *cookies*.

2) *Cache cookie* identificado

No caso de o *iframe* interno detectar a existência de um *cache cookie* que seja identificado como de origem do Comprovador (o *iframe* faz essa verificação baseado na chave *k*, conforme visto na seção 4.3.1.5), o cupom precisa ser enviado para o Comprovador para verificação de validade.

Os cupons são recebidos na forma de requisições pelo arquivo texto denominado cupom.ini. Como vimos anteriormente, quando tal arquivo for eventualmente requisitado, a requisição se dará por meio de uma URL.

Por exemplo: http://redeanuncio/cupom.ini?chave=k&ref=x&validade =y. A variável ref na URL reflete o ID_{pbc} e o ID_{anc} que foi clicado. A variável validade carregará a informação de data e hora de validade do cupom. Ao receber a requisição, o Comprovador primeiramente analisa a validade do cupom. Se houver expirado, o cliente terá que responder a um novo desafio, de maneira que uma requisição será criada, em nome do cliente, juntamente ao Gerador de CAPTCHA, conforme descrito no parágrafo anterior.

Se o cupom for válido, o Comprovador então armazena o horário da requisição, o chave-secreta *k*, a referência x e a validade y no banco de dados, além de atualizar o campo de *status* do clique para "válido", de maneira que uma cobrança junto ao anunciante possa ser posteriormente disparada.,

Finalmente, o Comprovador serve o *cache cookie* (que conterá novas referências de ID_{pbc} e o ID_{anc}, mas manterá a sua validade, para garantir que a heurística de validação de *cookie*s continue ativa) de volta ao usuário.

4.3.1.10 Auditorias

Como os servidores na rede de anúncios detem a responsabilidade pela decisão final acerca de quais cliques devem ser considerados válidos ou não, e assim ganharem

mais dinheiro quando mais cliques são aceitos como válidos, os publicadores e anunciantes podem acusar as redes de anúncios de inchar a quantidade de cliques válidos propositalmente.

Para resolver esse problema, uma entidade adicional, denominada Auditor, pode ser contratada com a finalidade de monitorar os cupons que são disponibilizados, e assim verificar o processo de decisão do Comprovador acerca da classificação de cliques.

Os *cache cookies* disponibilizados pelos Comprovadores podem ser trabalhados de maneira que, quando um anúncio for clicado, o cupom $C = (k, ID_{pbc}, ID_{anc}, v)$ seja disponibilizado tanto para o usuário quanto para o Auditor, que manterá um banco de dados independente.

Quando os números da rede de anúncios forem contestados, os cupons gravados pelo Auditor poderão ser consultados para recalcular o número de cliques válidos em um dado anúncio ou publicador, e comparados com o cálculo da rede de anúncios.

4.4 Análise de Segurança

Sem conhecer a chave do Comprovador, um adversário não possui meios para, matematicamente, forjar novos cupons, graças ao uso dos MACs. Entretanto, um fraudador ainda pode fraudar este esquema através de algumas outras formas:

- Fraude direta do publicador: usando uma pequena modificação na solução proposta, o publicador pode intermediar a entrega de cupons mesmo quando os usuários não clicam nos anúncios.
- Fraude indireta do publicador: um Web site desonesto pode redirecionar usuários para o site do publicador.
- Cliques direcionados por *malware*: um vírus ou trojan poderia sorrateiramente direcionar o navegador de um usuário e simular um ou mais cliques com o objetivo de roubar um cupom válido, cujo uso seria automatizado em outra plataforma. Para esse ataque ser possível, é necessário contar com a boa vontade do usuário, que resolverá o CAPTCHA clicável, quando exibido. A *click fraud* será realizada durante a validade do cupom. Após a validade, o *malware* repeteria a ação.

Todos esses ataques são possíveis nos esquemas de *click fraud* atualmente existentes. As várias técnicas usadas para endereçá-los podem igualmente ser usadas por esta abordagem. Como exemplo de uma dessas técnicas, se houver alguma desconfiança, a rede de anúncios pode direcionar um de seus computadores para o site do publicador para determinar se o mesmo está gerando cliques fraudulentos.

Certamente, o esquema aqui apresentado torna mais fácil a identificação do comportamento ilegítimo, pois permite “marcar” o cupom de um usuário e, posteriormente, monitorar diretamente o tráfego gerado pelo cliente.

Um adversário pode ainda tentar explorar algumas outras características especiais da abordagem aqui proposta:

- Tentar se comportar como um Comprovador: um adversário que consiga, de uma forma ou de outra, a chave do Comprovador, pode tentar se comportar como tal. Se a rede de anúncios estabelece regras suficientemente boas para identificação da origem dos cupons, isso ficaria bem mais difícil. Além disso, no caso da chave MAC for univocamente identificada como sendo de propriedade de um determinado Comprovador, a rede de anúncios pode monitorar individualmente o tráfego gerado pelo Comprovador, o que torna o combate à fraude mais eficiente.

- Coleta de cupons: Um adversário pode possuir um meio de realizar uma análise de tráfego de comunicação entre os usuários e as redes de anúncios e desenvolver uma estrutura de ataque do tipo "homem no meio", coletando cupons que seriam destinados à rede de anúncios e possibilitando uma posterior *click fraud*. Ao estabelecer políticas de validação da origem dos cupons, a rede de anúncios pode inibir esta prática ao aplicar penas financeiras para esta forma de fraude em detrimento dos ganhos excessivos que um fraudador pode obter a partir de sua prática.

4.5 Considerações Finais

Em um mundo de perfeita transparência, no qual uma rede de anúncios saberia a identidade real de todos os usuários que clicam em seus anúncios, a *click fraud* seria muito mais gerenciável. Em tal mundo, seria muito mais fácil identificar o mau comportamento de um dado usuário – por exemplo, a repetição de cliques em um mesmo anúncio – ou os cliques que foram iniciados por usuários maliciosos ou por *bots*.

Uma rede de anúncios poderia ir ainda mais adiante e referenciar bancos de dados contendo perfis de usuários que clicaram em seus anúncios. Uma rede poderia ainda criar uma estrutura de preços de cliques altamente refinada, baseada no valor que um determinado usuário potencialmente gastaria, ou ainda criar uma compensação diferenciada, baseada em reputação, para publicadores que consigam gerar cliques de determinados usuários. O esquema proposto aqui difere desse mundo ideal em dois sentidos básicos:

1- Conhecimento parcial: dada a natureza fragmentada dos bancos de dados sobre comportamento de usuários, e as políticas de privacidade normalmente difundidas por grupos de regulação, o conhecimento geral dos padrões de cada usuário é modesto, restringindo-se a possibilitar à rede de anúncios o conhecimento de que um determinado clique foi originado por um humano com intenções provavelmente honestas. Não é foco deste trabalho uma diferenciação forte entre usuários, embora os protocolos de comunicação pudessem ser utilizados para suportar este objetivo.

2- O navegador como meio de transporte: ao invés de contar com um repositório central de dados, este esquema confia nos navegadores dos usuários para compartilhar informação entre as entidades participantes. Esta abordagem ajuda eliminar complexidade técnica e protege a privacidade dos usuários.

A princípio, essa abordagem poderia suplantar completamente todos os esquemas

atuais e tradicionais de segurança de pagamento por clique. Entretanto, utilizá-la simultaneamente a um esquema tradicional, como uma forma de complementação, parece ser a opção mais indicada.

Em todo caso, a abordagem aqui proposta certamente poderia ser lançada de forma experimental por alguma rede de anúncios com impacto mínimo nos negócios atuais, e, na medida em que obtivesse garantias de sucesso, ser expandida. Esta abordagem não só é uma nova visão e paradigma para o combate da *click fraud*, mas provê também uma grande oportunidade de controlar essa fraude.

CAPÍTULO 5

O FUTURO DO MARKETING E DO MERCHANDISING ONLINE

Este capítulo discute as principais contribuições do trabalho relacionado com a atividade de marketing digital . Em seguida, descreve alguns trabalhos relacionados com esta área de pesquisa, e, finalmente fornece alguns possíveis direcionamentos para pesquisas adicionais, que poderão ser realizadas a partir deste livro.

5.1 Contribuições do Trabalho

Neste trabalho, é proposto um método para prevenção de *click fraud*, que vai de encontro à grande maioria dos métodos de combate atuais, que são baseados na detecção da fraude após a ocorrência da mesma, por meio da filtragem de cliques mal-intencionados.

Contrariamente a esse método, o esquema apresentado nesta pesquisa propõe o uso de testes para que se realize diferenciação entre humanos e computadores, através de CAPTCHA. A resposta desses testes serão um atestado de validade dos cliques, que, depois de serem considerados “bons”, serão contabilizados.

Em teoria, esse método poderia suplantar os métodos atuais de detecção, muito embora ele seja especialmente atrativo para ser utilizado de maneira complementar.

A área de marketing digital é bastante recente, no entanto inúmeras pesquisas têm sido realizadas, sinalizando grandes perspectivas de crescimento. Destaca-se neste trabalho o foco dado à necessidade de se combater o problema da *click fraud* como requisito fundamental para a continuidade da publicidade na Internet. Este trabalho fornece importantes contribuições nas áreas de pesquisa de interação homem-máquina, segurança da informação e de negócios online, sendo algumas delas:

- Um levantamento de algumas abordagens para a realização da sistemática de marketing digital , analisando sua adequação às necessidades de negócio orientadas pelos pressupostos estabelecidos neste trabalho, incluindo um estudo aprofundado da sistemática das redes de anúncios, algo inédito na literatura acadêmica em língua portuguesa.
- Apresentação da evolução histórica das diversas abordagens de formas de obtenção de receita para o negócio de marketing digital , destacando o pagamento por clique, a partir do qual se realiza a *click fraud*, foco deste trabalho. Além disto, realizou-se uma análise da atual disposição do mercado de negócios online, os seus principais representantes e as formas de avaliação de qualidade e crescimento de serviços.
- Um estudo minucioso acerca das di-

versas fraudes em negócios online, com foco especial na *click fraud*, com a apresentação de números, casos legais, disposição atual do assunto na academia e as diversas soluções atualmente adotadas para o combate à *click fraud*.

- Um estudo detalhado da tecnologia de CAPTCHA, de maneira a ossuísse-la como base para a abordagem proposta por este livro para o combate à *click fraud*. Essa tecnologia deve ser utilizada como camada de diferenciação entre usuários reais – humanos – e máquinas que rodam programas automatizados para realizarem a fraude. Um foco especial foi dado aos CAPTCHA clicáveis, devido a sua maior facilidade de uso para os usuários reais, e à apresentação de números que mostram que este tipo de CAPTCHA é bem mais seguro do que os CAPTCHA clássicos, baseados em texto.
- Uma proposta de abordagem para o fluxo de ações envolvidas na sistemática do pagamento por clique, incluindo os passos necessários para a validação de usuários, de modo a inibir a *click fraud*. É importante notar que a abordagem de prevenção por meio da resolução de CAPTCHA e autenticação por cupons proposta trouxe uma mudança de paradigma na forma como a *click fraud* é tratada, uma vez que as abordagens tradicionais focam em detecção.
- Uma proposta de arquitetura para as redes de anúncios que implementa a abordagem de resolução da *click fraud* proposta neste trabalho. Nessa proposta de arquitetura, tratamos de aspectos tais como banco de dados e entidades que funcionariam como serviços web: Comprovador, Gerador de Anúncios e Gerador de CAPTCHA. Vale salientar que a sistemática de CAPTCHA clicáveis é explorada em detalhes, assim como a proposta para o serviço web Comprovador.

5.2 Trabalhos Relacionados

A utilização de marketing online é uma necessidade real das organizações, sejam elas grandes ou não, no cenário atual, uma vez que a concorrência é cada vez maior e os sistemas de negócio estão cada vez mais complexos, envolvendo vários *stakeholders* com exigências diversas. Esta seção apresenta alguns trabalhos desenvolvidos como metodologias e abordagens para o combate da *click fraud*. Discutiremos também trabalhos que realizam estudos acerca de CAPTCHA e CAPTCHA clicáveis.

Em termos de tratamento da *click fraud*, as abordagens atualmente existentes empregam suas forças em heurísticas para detecção da mesma. Sob essa perspectiva, existem trabalhos extremamente relevantes, como o DETECTIVES (METWALLY; AGRAWAL; ABBADI, 2007), que realiza a proposta do algoritmo *Similarity-Seeker*, que busca identificar coligações de fraudadores localizados em diversos sites na web.

Essa solução é ainda generalizada no mesmo trabalho para coligações de tamanhos arbitrários. O artigo apresenta a implementação de um protótipo do algoritmo que detectou, com sucesso, coligações de fraudadores de diversos tamanhos. O DETECTIVES é inovador, pois, dentro da sistemática de detecção da fraude, apresenta uma proposta de verificação em diversas camadas envolvidas em um sistema de

marketing digital – desde a investigação nas camadas de interface com usuários, até o armazenamento dos cliques em um histórico.

Na medida em que se analisam diferentes camadas de detecção de um histórico de cliques, passa a ser um passo natural o questionamento acerca da prevenção, sendo o algoritmo *Similarity-Seeker* uma verdadeira inspiração para este trabalho.

É exatamente neste questionamento onde está a diferenciação entre os trabalhos. A abordagem aqui proposta não tem o objetivo de ser mais um método de detecção, que, ao analisar o histórico de cliques, irá eliminar os cliques inválidos. Ao contrário, foca na distinção entre cliques válidos e inválidos no momento em que acontecem, utilizando desafios CAPTCHA e autenticação por cupons para validar os cliques.

O passo de validação é realizado em tempo de execução. O algoritmo *Similarity-Seeker* realiza a análise do histórico de cliques em um período de tempo posterior à execução, através de filtros heurísticos. Embora sejam eficientes, a utilidade desses filtros é limitada contra fraudadores sofisticados, e sua performance pode ser gradativamente prejudicada quando esses fraudadores aprendem como ludibriar tais filtros.

Gostaríamos também de citar alguns trabalhos que serviram como referência para o desenvolvimento desta pesquisa. Primeiramente, os trabalhos mais difundidos no tratamento acadêmico da sistemática de marketing digital são o livro *Advertising on the Internet* (ZEFF; ARONSON, 1999) e o artigo *Advertising on the Web* (JAKOBSSON; MACKENZIE; STERN, 1999).

Ambos explicam as formas como os maiores anunciantes na Internet têm maximizado sua presença global, sendo de fundamental importância para todos aqueles que objetivam se beneficiar do marketing na Internet, que é modelo de maior sucesso de publicidade desde o surgimento de anúncios televisivos.

Eles também são de leitura fundamental para todos os interessados no assunto sob uma perspectiva acadêmica, explorando fatores tais como: modelos de sucesso para marketing digital , dados acerca de pesquisa de mercados e tipos de marketing online (direto, promoções, formas de targetting e otimização). Além disso, oferecem um guia para ferramentas de gerenciamento de anúncios e aspectos legais do mercado de anúncios na web.

Em termos de estudos sobre a tecnologia de CAPTCHA, é imprescindível destacar o ASIRRA (ELSON et al., 2007), já citado anteriormente no capítulo 3, como um modelo ideal de CAPTCHA clicáveis que possam ser utilizados para prevenir a *click fraud* de acordo com a abordagem proposta neste livro. Além de introduzir uma metodologia de CAPTCHA significativamente mais confiável (isto é, que reduz as probabilidades de ter sua resolução automatizada por algum software OCR) e de tornar a experiência de utilização de CAPTCHA muito mais fácil e intuitiva para o usuário bem-intencionado, ela utiliza uma conexão com um banco de dados externo (Petfinder.com) para gerar os desafios.

Essa associação garante que os recursos necessários para a alimentação do

CAPTCHA clicável sejam mínimos, mesmo sendo os desafios gerados a partir de um banco de dados dinâmico de mais de três milhões de fotos. Além disso, a parceria com o Petfinder, através da colocação do link “Adopt me” na foto de cada um dos animais, é um dos pontos altos desse CAPTCHA: além de realizar um trabalho de referência técnica, é sempre bom fazer parte de um projeto que se propõe a fazer algum bem para a humanidade.

Além dos trabalhos supracitados, outras pesquisas relacionadas à *online advertising* que são de extrema importância para o estabelecimento deste assunto como um tópico de relevância acadêmica e mercadológica podem ser encontradas em (ANUPAM et al, 1999), (BEIGHTON, 2010), (BROWSER MEDIA, 2010), (DE JONG; ROSENTHAL, 2010), (GOODMAN, 2008), (KANG; LEE, 2003), (KHAN et al, 2010), (KLEIN, 1999), (LIEDTKE, 2006), (MANN, 2006), (METWALLY; AGRAWAL; ABBADI, 2005), (SAGAR, 2010) e (ZELLER, 2004). Todos esses trabalhos ressaltam a necessidade do combate à *click fraud* e exploram esse problema como obstáculo para o crescimento do negócio de marketing digital , especialmente o negócio de pagamento por clique.

Pode-se destacar também os trabalhos que realizam uma pesquisa complementar no assunto de CAPTCHA: (MORI; MALIK, 2003), (NAOR, 1996), (AHN et al, 2003) e e (WROBLEWSKI, 2010). Chow et al (2008) chega a mencionar que o combate à *click fraud* pode ser realizado por CAPTCHA clicáveis.

Com respeito à sistemática de autenticação por *cookie*s e cupons, pode-se mencionar os seguintes trabalhos como de fundamental importância para o aprimoramento neste tópico de estudo: (KING, 2008), (SAMAR, 1999) e (TEIXEIRA, 2004).

5.3 Trabalhos Futuros

De maneira geral, uma página na Internet é um negócio e, como tal, deve cobrir suas despesas para sobreviver. Em termos de continuidade de negócio, pode-se observar que o modelo atualmente utilizado para serviços baseados na Web está submetido a duas opções básicas:

1. A existência de muitos anúncios em *sites* de Internet;
2. A utilização do método de inscrição, paga ou não, de todo serviço ou de parte dele, esperando que milhares de usuários aceitem pagar uma taxa mensal.

Levando isto em consideração, o futuro do mercado de serviços na Web passa pelo amadurecimento da utilização de anúncios em *sites*. Um trabalho futuro para esta pesquisa é estudar e dissertar acerca deste amadurecimento, incluindo uma análise criteriosa dos requisitos de usabilidade de usuários de Internet. Não será uma surpresa se, em breves dias, os *web sites* passarem a utilizar o seguinte conjunto de anúncios em suas páginas:

- Múltiplos anúncios em banner e barras laterais;
- Dois ou três anúncios em janelas internas (*pop-under*);
- Um anúncio flutuante;
- Um anúncio em Unicast na frente disso tudo.

Ainda não se pôde ver uma página na Internet buscar tal variedade de anúncios, mas a verdade é que este é um negócio extremamente lucrativo. Um *site* nesse modelo pode facilmente faturar até 50 doláres cada vez que a página for carregada com essa quantidade de anúncios.

Em contrapartida, como já mencionado, um site como esse tende a afastar usuários. Encontrar a relação idealmente lucrativa entre os requisitos de usabilidade do usuário de Internet comum, a quantidade de anúncios em um site e a disposição desses anúncios ao longo das páginas é um problema multidisciplinar.

Além disso, um trabalho futuro parece ser a implementação de um protótipo das entidades envolvidas nesta arquitetura. O protótipo da entidade Comprovador precisa tão somente seguir as características listadas no capítulo 4. Entretanto, o desenvolvimento dessa entidade é de importância fundamental para o protótipo da rede de anúncios, pois é a entidade que realiza a maior parte do trabalho e possui as principais responsabilidades nesta arquitetura.

Para o desenvolvimento do Gerador de Anúncios, faz-se necessária a definição de uma política de exibição baseada no modelo de negócios adotado para tal, conforme descrito no capítulo 4. Os requisitos mínimos para a modelagem do banco de dados são detalhados na seção 4.3.1.2, e podem ser um ponto de início para orientar o desenvolvimento das diretivas de acesso ao banco e montagem dos banners.

Acerca da questão específica de segurança em redes de anúncio, no futuro, será necessário tomar a decisão acerca do CAPTCHA clicável a ser adotado no projeto. Devido às limitações expostas nos capítulos 3 e 4, propor o desenvolvimento de um novo modelo de implementação para CAPTCHA clicável é uma opção.

Entretanto, para a utilização de um CAPTCHA clicável baseado em imagens, um alinhamento de interesses similar ao do ASIRRA com o Petfinder.com, ou seja, um banco de dados externo que cresça continuamente é o mais indicado. Neste caso, um trabalho futuro seria localizar um parceiro de grande porte para o projeto.

O principal objetivo desta abordagem é propor uma nova implementação para redes de anúncios, na qual *cache cookie*s são atribuídos a usuários válidos e, baseados nestes *cache cookie*s, os usuários se autenticam por meio de cupons.

Nesse sentido, nada impede que tais *cache cookie*s portem mais informações sobre os usuários como, por exemplo, a frequência na qual um determinado usuário realiza uma compra, após clicar em um anúncio. Dessa forma, os cupons poderiam ser classificados, e o modelo de compensação de cliques poderia ser adaptado, de maneira a se basear no valor associado a cada usuário (na verdade, a cada cupom). Isso poderia também estar associado à implementação de um mecanismo de reputação

de cada *cache cookie*, que poderia complementar a solução proposta nesta pesquisa, na funcionalidade relacionada à validade dos cupons.

Quanto melhor a reputação de um dado usuário, mais tempo ele poderia utilizar a rede de anúncios sem ter que responder um novo desafio. Além disso, cliques de usuários que possuíssem uma boa reputação valeriam mais do que cliques dos demais. Isso implicaria em uma mudança no atual modelo de negócios de redes de anúncios, no qual o valor de um clique está muito mais relacionado ao valor de mercado da rede que exibe o anúncio, do que propriamente ao potencial de compra de um dado clique.

A análise dessa possibilidade parece ser extremamente interessante e é definitivamente parte do escopo do projeto em seus estágios subsequentes.

REFERÊNCIAS

AHN, L. v. et al. CAPTCHA: U**sing hard AI problems for security**. In: *Advances in Cryptology – EUROCRYPT 2003.* International Conference on the Theory and Applications of Cryptographic Techniques. Varsóvia, Polônia, 2004. *Proceedings...* Londres: Springer, 2003.

_____. **reCAPTCHA: Human-Based Character Recognition via Web Security Measures**. *Science Express*, v. 321, n. 5895, p. 1456-1468, 2008.

ANUPAM, V. et al. **On the Security of Pay-Per-Click and Other Web Advertising Schemes**. In: *8th WWW International Conference on World Wide Web, 8.* 8th International Conference on World Wide Web. Toronto, Canada, 1999. *Proceedings…* Amsterdã: Elsevier Science, 1999.

ARORA, A.. **Statistics Hacking – Exploiting Vulnerabilities in News Websites**. *International Journal of Computer Science and Network Security*, v. 7 p. 342–347, 2007.

ARRINGTON, M. **Alexa says now Youtube is bigger than Google. Alexa is useless**. Disponível em <http://techcrunch.com/2007/08/13/alexa-says-youtube-is-now-bigger-than-google-theyre-wrong/>. Acesso em 22 de Maio de 2010.

AZEREDO, E.; VELHO, L. **Computação Gráfica: Teoria e Prática**. Rio de Janeiro: Campus, 2003. 368 pgs.

BEIGHTON, C. **Computesystems.com – Web Marketing and Help Log**. Disponível em: <http:// netlogicalsites.com/free_website_marketing_help.php? cnt=internet%20 advertising>. Acesso em 23 de Maio de 2010.

BELLARE, M., CANETTI, R., KRAWCZYK, H. **Keying hash functions for message authentication**. In: *CRYPTO '96: Annual International Cryptology Conference on Advances in Cryptology*. 16th Annual International Cryptology Conference on Advances in Cryptology. Londres, 1996. *Proceedings…* Londres: Springer, 1996.

BLUNDO, C; CLIMATO, S.; SAWM. **A Tool for Secure and Authenticated Web Metering**. *ACM SEKE International Conference on Software Engineering and Knowledge Engineering*. 14th ACM SEKE International Conference on Software Engineering and Knowledge Engineering. Ischia, Itália. *Proceedings…* Nova Iorque: ACM Digital Library, 2002.

BROWSER MEDIA. **DoubleClick deal means Google controls 69% of the online ad market**. Disponível em: <http://www.browsermedia.co.uk/2008/04/01/ doubleclick-deal-means-google-controls-69-of-the-online-ad-market/>. Acesso em: 23 de Maio de 2010.

CHELLAPILLA, K. et al. **Designing human friendly human interaction proofs (HIPs)**. *Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems*. Conference on Human Factors in Computing Systems Portland, Oregon, 2005. *Proceedings…* Nova Iorque: ACM Digital Library, 2005.

CHOW et al. **Making CAPTCHA Clickable**. *Proceedings of HotMobile*. Napa Valley, California, Estados Unidos. *Proceedings…* Nova Iorque: ACM Digital Library, 2008.

CROLL, A.; POWER, S. ***Complete Web Monitoring***. Sebastopool: O'reilly Media, 2009. 38 p.

DAVIS, W. **Google Wins $75,00 in *Click Fraud* Case**. Disponível em: http://www.mediapost.com/

publications/index.cfm?fuseaction=Articles.san&s=31772. Acesso em: 04 de Setembro de 2009.

DE JONG, A.; ROSENTHAL, L.; VAN DIJK, M. **The Risk and Return of Arbitrage in Dual-Listed Companies**. **Review of Finance**, Vol. 13, 495-520, 2009. Disponível em: <http://ssrn.com/abstract=525282>. Acesso em: 17 de Maio de 2010.

DIRECT MARKETING ASSOCIATION. ***The Power of Direct Marketing: ROI, Sales, Expenditures and Employment in the U.S***. Nova Iorque, 2006.

DOYLE, E. **Not All Spyware is as Harmless as *Cookie*s: Block it or Your Business Could Pay Dearly**. Disponível em: <http://www.computerweekly .com/Articles/2003/11/27/198884/Not-all-spyware-is-as-harmless-as-*cookie*s. htm>. Acesso em: 23 de Maio de 2010.

ELGIN, B. **The Vanishing *Click Fraud* Case**. Disponível em: <http://www.businessweek.com/technology/content/dec2006/tc20061204_923336.htm?campaign_id=bier_tcc.g3a.rssd1204f>. Acesso em: 08 de Janeiro de 2010.

ELSON, J. et al. **ASIRRA: a CAPTCHA that exploits interest-aligned manual image categorization.** *ACM conference on Computer and communications security*. 14th ACM conference on Computer and communications security. Alexandria, Estados Unidos. *Proceedings*... Nova Iorque: ACM Digital Library, 2007.

GOODMAN, A. ***Winning Results with Google AdWords***. Nova Iorque: McGraw-Hill, 2008. 376p.

HOFFMAN, D; NOVAK, T. P. **How to Acquire Customers on the Web**. Disponível em: <http://hudding.esmartbiz.com/harvardbusinessreview7.pdf>. Acesso em 23 de Maio de 2010.

IMMORLICA, N. et al. ***Click Fraud* Resistant Methods for Learning Click-Through Rates**. In: *First International Workshop*, WINE 2005, 10., Hong Kong, China. Heidelberg: Springer Berlin, 2005.

JAKOBSSON, M.; MACKENZIE, P.; STERN, J. **Secure and Lightweight Advertising on the Web.** *WWW International Conference on World Wide Web, 8*. 8th WWW International Conference on World Wide Web. Toronto, Canada, 1999. *Proceedings*... Amsterdã, Holanda: Elsevier Science, 1999.

JAKOBSSON, M.; RAZMAN, Z. **Crimeware: *Understanding New Attacks and Defenses*.** Londres: Addison-Wesley Professional, 2008. 608 p.

JUELS, A., JAKOBSSON, M., AND JAGATIC, T. ***Cache cookie*s for browser authentication (extended abstract).** I*EEE Symposium on Privacy and Security.* 2006 IEEE Symposium on Privacy and Security. Oakland, Estados Unidos, 2006. *Proceedings*... Palo Alto: IEEE CS Press, 2006.

KANG, S. R.; LEE, E. **Investigating Elements on the E-Commerce Homepage: Focus on business to customer websites**. *International Symposium on Design Conference*, 6th Asian Design Conference, 2003. Tsukuba City, Japão. *Proceedings...* Tsukuba City: International Conference Proceedings, 2003.

KAMBA, T. et al. **Using Small Screen Space More Efficiently.** *Conference on Human Factors in Computing Systems*. SIGCHI conference on Human Factors in computing systems: common ground. Vancoucer, Canada, 1996. *Proceedings*... Nova Iorque: ACM Digital Library, 1996.

KASSAI, J.R. ***Alguns Aspectos que Contribuem para a Conciliação entre a Taxa Interna de Retorno e o Return On Investment (ROI).*** Livro (Mestrado) – FEA/USP, São Paulo. 1996.

KING, A. ***Website Optimization*.** Sebastopool: O'Reilly Media, Inc. 2008. 43 pgs.

KHAN, I. et al. **The Rise of Ad Networks: An In-Depth Look at Ad Networks**. Disponível em: <http://

www.mediamath.com/docs/JPMorgan.pdf>. Acesso em 20 de Maio de 2010.

KLEIN, D. **Defending Against the Wily Surfer-Web-based Attacks and Defenses.** *Workshop on Intrusion Detection and Network Monitoring.* 1st USENIX ID Workshop on Intrusion Detection and Network Monitoring. Santa Clara, California, Estados Unidos, 1999. *Proceedings*... Berkeley: USENIX Association, 1999.

KOSTER, M. **Guidelines for robot writers.** Disponível em: <http://info.webcrawler.com /mak/projects/robots/guidelines.html>. Acesso em: 12 de Agosto de 2008.

LEYDEN, J. **Miaow to kitten-based authentication.** Disponível em: <http://www.theregister.co.uk/2006/04/12/kittenauth>. Acesso em: 11 de Abril de 2009.

LIEDTKE, M. **Google to Pay $90M in '*Click Fraud*' Case.** *Washington Post Magazine*. Washington, Estados Unidos. 11 mar. 2006.

MAJUMDAR, S.; KULKARNI, D.; RAVISHANKAR, C. **Addressing *Click Fraud* in Content Delivery Systems.** *IEEE Conference on Computer Communications.* 26th IEEE Conference on Computer Communications, Infocom'07. Anchorage, Alaska, Estados Unidos. *Proceedings*... Palo Alto: IEEE CS Press, 2007.

MANN, C. **How *Click Fraud* Could Swallow the Internet.** *Wired Magazine*. São Francisco, v 14, n.1, pp 17-20, 2006.

MANTAS, J. **An Overview of Character Recognition Methodologies**. *Pattern Recognition*, vol. 19, n. 6, pp 425-430, 1986.

METWALLY, A.; AGRAWAL D.; ABBADI, E. **DETECTIVES: Detecting Coalition Hit Inflation Attacks in Advertising Networks Streams**. *International World Wide Web Conference*. 16th WWW International World Wide Web Conference. Alberta, Canada, 2007. *Proceedings*.... Nova Iorque: ACM Digital Library, 2007.

_____. **Duplicate Detection in Click Streams**. *International World Wide Web Conference*. 14th WWW International World Wide Web Conference. Chiba, Japan, 2005 *Proceedings*.... Nova Iorque: ACM Digital Library, 2005.

MORI, G.; MALIK, J. **Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA.** *Conference on Computer Vision and Pattern Recognition*. CVPR '03, Conference on Computer Vision and Pattern Recognition. Madison, Wiscosin, Estados Unidos, 2003. *Proceedings*... Palo Alto: IEEE CS Press, 2003

MORI, S.; NISHIDA, H.; YAMADA, H. **Optical Character Recognition**. Nova Iorque: John Wiley & Sons, 1999. 560 pgs.

NAOR, M. **Verification of a Human in the Loop or Identification via the Turing Test.** Disponível em <http://www.wisdom.weizmann.ac.il/~naor/PAPE RS/human.pdf>. Acesso em: 17 de Março de 2010.

NARAINE, R. **Feds Arrest Google Extortionist**. Disponível em: <http://www.internetnews.com/bus-news/article.php/3329281>. Acesso em 11 de Agosto de 2010.

O'REILLY, T. **What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software**. Disponível em: <http://papers.ssrn.com /sol3/papers.cfm?abstract_id=1008839&download=yes">. Acesso em 23 de Maio de 2010.

RAMASUBRAMANIAN, L. **Client Behavior and Feed Characteristics of RSS**. A PublishSubscribe System for Web Micronews. In*: Internet Measurement Conference 5*. IMC´05, 5, Berkeley, California, Estados Unidos. *Proceedings*... Berkeley: UNENIX Association, 2005.

RYAN, K. M. **Big Yahoo *Click Fraud* Settlement**. Disponível em: http://www.imediaconnection.com/content/10294.asp. Acesso em: 04 de Setembro de 2009.

SAGAR, C. SEO: **A Quick Primer on the Difference Between Ecommerce and Content Sites**. Disponível em: <https://www.openforum.com/idea-hub/topi cs/innovation/article/seo-a-quick-primer-on-the-difference-between-ecommerce-and-content-sites-chaitanyasagar?sorttype=newest&sourcepage=1&postguid= 1f497480-a230-4816-96f4-730a9ea0acc9&pagenumber=1&thumbsup=1f49748 0-a230-4816-96f4-730a9ea0acc9>. Acesso em 23 de Maio de 2010.

SAMAR, V. **Single Sign-On Using *Cookie*s for Web Applications**. *IEEE Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*. 8th IEEE Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises. Stanford, California, Estados Unidos. *Proceedings*... Palo Alto: IEEE CS Press, 1999.

SIMARD, P.; STEINKRAUS, D.; PLATT, J. C. **Best practices for convolutional neural networks applied to visual document analysis**. *International Conference on Document Analysis and Recognition*. International Conference on Document Analysis and Recognition. Edinburg, Escócia, 2003. *Proceedings*... Palo Alto: IEEE CS Press, 2003.

SMITH, C. S. **Geolocation: Core to the Local and Key to Click-Fraud Detection**. Disponível em: <http://searchengineland.com/070813-082025.php>. Acesso em 21 de Maio de 2010.

STONE, B. ***Marketing Direto*.** São Paulo: Nobel, 2004. 570pgs.

TEIXEIRA, M. A. M. ***Suporte a serviços diferenciados em servidores web: modelos e algoritmos***. Tese (Doutorado) – ICMC-USP, São Carlos – SP. 2004.

TOWNSEND, K. **Spyware, Adware and Peer-to-Peer Networks: The Hidden Threat to Corporate Security**. Disponível em <http://www.pestpatrol.com/ Whitepapers/CorporateSecurity0403>. Acesso em: 23 de Maio de 2010.

TRYCAPTCHA. **HN CAPTCHA**. Disponível em <http://www.trycaptcha.com/HNCaptcha>. Acesso em 02 de Outubro de 2010.

TURING, A. M**. Computing Machinery and Intelligence**. Oxford, Inglaterra: Mind, 1950. v. 59.

TUZHILIN, A. **The Lane's Gifts v. Google report, 2006. Independent evaluators assessment of quality of Google`s click-fraud filtering methods**. Disponível em: <http://googleblog.blogspot.com/pdf/Tuzhilin_Report.pdf>. Acesso em: 17 de fev. de 2008.

URBACH, R. R.; KIBEL, G. A. **Adware/Spyware: An Update Regarding Pending Litigation and Legislation**. *Intelectual Property & Technology Law Journal*. Chapel Hill, p. 12, 15 jun. 2004.

WROBLEWSKI, L. **10% Sign up Improvement with one Change**. Disponível em <http://www.lukew.com/ff/entry.asp?706>. Acesso em 18 de Junho de 2010.

_____. **A sliding alternative to CAPTCHA?** Disponível em <http://www.lukew.com/ff/entry.asp?1138>. Acesso em 3 de Julho de 2010.

ZEFF, R. L.; ARONSON, B. ***Advertising on the Internet***. Nova Iorque: John Wiley & Sons Inc, 1999. 436 p.

ZELLER JR, T. **With Each Technology Advance, a Scourge**. *The New York Times*, Nova Iorque, 17 out., 2004.

SOBRE O AUTOR

RODRIGO ALVES COSTA Atualmente (2018) é professor associado do curso de Ciência da Computação da Universidade Estadual da Paraíba (UEPB). Possui doutorado em Ciência da Computação pela UFPE (2016), mestrado pela mesma instituição (2010), MBA em Gerenciamento de Projetos pela Fundação Getúlio Vargas (2007) e graduação em Ciência da Computação pela UFPE (2005). É certificado Project Management Professional (PMP) pelo Project Management Institute (PMI) (2006). Tem vasta atuação profissional no mercado de trabalho, destacando-se em funções como gerente de projetos, analista de sistemas, engenheiro de software e analista de testes e de segurança em empresas como IBM, Siemens, Motorola e C.E.S.A.R. É autor de livros na área de administração empresarial, gerenciamento de projetos e engenharia de software, e atua como consultor e idealizador de qualificações na área de planejamento e governança estratégica em tecnologias da informação em diversas organizações públicas e privadas, incluindo instituições de ensino, sendo um dos fundadores do currículo de Governança de TI da Escola Superior de Redes (ESR), vinculada à Rede Nacional de Pesquisa (RNP), com o livro Gerenciamento de Projetos de TI. Foi bolsista do CNPQ durante o doutorado, mestrado e a graduação (PIBIC), e atualmente está vinculado aos diretórios de pesquisa da entidade, nos grupos de pesquisa em Segurança Computacional, da UFPE, em Gestão, Comportamento e Competências Organizacionais, do IFPB, e no grupo ATLAS - Ação das Tecnologias na Aprendizagem Significativa, da UEPB. Pesquisador na área de Ciência da Computação, com ênfase em segurança computacional e sistemas de informação, e na área de Gerenciamento de Projetos, com ênfase em gerenciamento de projetos de tecnologia da informação, abordagens ágeis de projeto, agilidade organizacional e marketing organizacional por projetos. É membro entusiasta do PMI, da Association for Computing Machinery (ACM) e da Sociedade Brasileira de Computação (SBC).

Agência Brasileira do ISBN
ISBN 978-85-455090-6-6