



A evolução dos mecanismos de rastreamento e vigilância intrusivos em clientes web

The evolution of the intrusive tracking and surveillance mechanisms on web clients

La evolución de los mecanismos de rastreo y vigilancia intrusivos en los clientes web

RODOLFO DA SILVA AVELINO¹

Resumo

Este trabalho parte de um breve levantamento histórico sobre o surgimento e a evolução dos mecanismos de rastreamento e de vigilância intrusivos em navegadores Web. Atualmente, utilizados na navegação entre páginas na Internet. Busca-se compreender se esta evolução surgiu por necessidades técnicas e/ou por motivações do modelo de negócios de empresas de marketing e publicidade on-line que movimentam a economia do mercado de dados pessoais. Se propõe, ainda, analisar as características dos principais mecanismos de rastreamento de comportamento online intrusivos, considerando, sobretudo, os menos transparentes, mais resilientes a remoção e que contribuem para a formação de uma infraestrutura de coleta de dados mantidas por corporações e plataformas de negócios. Para tal, serão analisados os documentos e as listas de formulação dos padrões das entidades e das organizações responsáveis pela regulamentação e padronização da Internet, bem como a demonstração de exemplos de rastreamento destes mecanismos.

Palavras-chave: Rastreamento de comportamento; Cookies; vigilância; web tracking.

Abstract

This work is based on a brief historical survey on the emergence and evolution of intrusive tracking and surveillance mechanisms in Web browsers. Currently used for navigation between Internet pages. It seeks to understand if this evolution arose due to technical needs and/or motivations of the business model of online marketing and advertising companies that move the economy of the personal data market. It is also proposed to analyze the characteristics of the main intrusive online behavior tracking mechanisms, considering, in particular, the less transparent, more resilient to removal and which contribute to the formation of a data collection infrastructure maintained by corporations and business platforms. To this end, the documents and formulation lists of the entities and organizations responsible for regulating and standardizing the Internet will be analyzed, as well as demonstration examples of the tracking of these mechanisms.

Key words: behavior tracking; web cookies; Surveillance; web tracking.

Resumen

Este trabajo parte de un breve relevamiento histórico sobre el surgimiento y la evolución de los mecanismos de rastreo y de vigilancia intrusivos en navegadores Web. Actualmente, utilizados en la navegación entre páginas de Internet. Se busca comprender si esta evolución surgió por necesidades técnicas y/o por motivaciones del modelo de negocios de empresas de marketing y publicidad en

¹ Doutorando no Programa acadêmico interdisciplinar de Ciências Humanas e Sociais da UFABC, onde concentra sua pesquisa na Linha Cultura, Comunicação e Dinâmica Social, e-mail para contato: rodolfo.avelino@ufabc.edu.br.

línea que mueven la economía del mercado de datos personales. Se propone, además, analizar las características de los principales mecanismos de rastreo de comportamiento online intrusivos, considerando, sobre todo, los menos transparentes, más resilientes a la remoción y que contribuyen a la formación de una infraestructura de recolección de datos mantenidas por corporaciones y plataformas de negocios. Para esto, se analizarán los documentos y las listas de formulación de las normas de las entidades y de las organizaciones responsables por la reglamentación y la estandarización de Internet, así como la demostración de ejemplos de rastreo de estos mecanismos.

Palabras clave: ratreamiento comportamental; cookies; vigilância; seguimiento web.

Introdução

Em 1996, o economista Hal R. Varian, hoje Chief Economist na Google, observou que a redução de custo das Tecnologias e da manipulação das informações, novos modelos de negócio surgiram (VARIAN: 101). No paper *Aspectos Econômicos da Privacidade Pessoal*², Varian identifica a possibilidade de expansão de um mercado de dados pessoais que poderia reduzir custos de transação dos empreendimentos tecnológicos caso as pessoas pudessem vender ou “alugar” seus dados (1996: 11).

Já em 1999, em seu livro *A Economia da Informação: como os princípios econômicos se aplicam à era da Internet*, Carl Shapiro e Hal R. Varian, alinhados com ideais neoliberais, defendem a tese de que os princípios econômicos duráveis podem orientar o frenético ambiente empresarial de hoje, discordando que a nova configuração econômica seja distinta da economia tradicional e afirmando que a tecnologia muda, porém as leis da economia não. Ainda, afirmaram que “há uma diferença essencial entre a velha e a nova economia: a velha economia industrial era movida pelas economias de escala; a nova economia da informação é movida pela economia de redes” (SHAPIRO; VARIAN, 1999: 204).

A teoria da economia informação também é observada por Manuel Castells no primeiro volume de sua trilogia que busca esclarecer a dinâmica econômica e social da nova era da informação, *A Era da Informação*, compreende que este novo cenário influenciado pelas novas tecnologias da informação e sua interligação por meio das redes computacionais, possam configurar uma “nova economia”, sendo esta informacional e global (CASTELLS, 2011: 119).

Esta economia estruturada sobre a infraestrutura da rede mundial de computadores, procura despertar as atenções dos usuários para estimular seus desejos de consumo por intermédio das ações de publicidade e marketing.

² Título original: “Economic Aspects of Personal Privacy”.

Deleuze (1992) antes da popularização da Internet, ao caracterizar a passagem da sociedade disciplinar para uma sociedade de controle reconheceu que o marketing é a alma das empresas e um grande instrumento de controle social. Argumentou que este controle é de curto prazo e de rápida rotação, contudo, contínuo e ilimitado. Atualmente é possível observar a grande influência do marketing no funcionamento da Internet. Controle de curto prazo e rotação rápida ecoam nas estratégias de marketing pelos meios digitais.

Estas estratégias estão condicionadas as atuais técnicas de rastreamento de comportamento online, onde a principal questão é compreender diariamente as necessidades dos usuários. Em uma sociedade já envolvida pelas TIC, Lazzarato (2006), ressalta a força dos investimentos de marketing por parte das empresas, disponibilizando até 40% de seu capital de giro em publicidade, modelagem e design. Estes investimentos podem ser superiores aos de produção de trabalho e meios de produção (LAZZARATO, 2006: 102).

Assim na Internet cria-se um ambiente concebido para estimular desejos, no qual os usuários são envolvidos pelo ímpeto das estratégias do marketing. Ainda, estes ambientes são caracterizados pela multiplicação de “mundos” de consumo, onde a relação entre a oferta e a demanda ficam invertidas segundo Lazzarato (LAZZARATO, 2006: 100).

Nesta inversão de cenário, Varian e Shapiro ponderam que “Conhecer Teu Cliente” em uma economia da informação é tão ou mais importante se comparada a economia industrial. Ressaltam que a ampliação da captação de informações dos clientes permite seu uso para a projeção de produtos e valores que melhor se adaptem às suas necessidades. Para eles, este mecanismo é essencial para maximizar ganhos do negócio. Criam-se ambientes para a coleta de dados com o objetivo de estimular desejos e necessidades, onde os usuários são alvo das criativas forças estratégicas do marketing. Assim, economia e sociedade evoluem na velocidade de bilhões de bits por segundos, por meio de conexões digitais em um mundo em que as interações são principalmente digitais.

Lazzarato reconhecendo as afirmações de Deleuze acerca do marketing na economia contemporânea, afirma que as sociedades de controle “caracterizam-se assim pela multiplicação da oferta de “mundos” (de consumo de informações, de trabalho, de lazer)” (2006: 101). Ainda ratifica, que estes mundos são padronizados e oferecidos a partir de interesses, sendo que nossa liberdade está condicionada exclusivamente na escolha destes, sendo assim privados de suas construções. O autor ainda ressalta o papel da publicidade por meio de empresas e agências na invenção de

acontecimentos que devem ser encarnados nos corpos, ditando maneiras de viver, de comer, de se vestir, entre outras.

Ambientes controlados também são analisados por Eli Pariser que os definem como bolha de filtros. Para ele um universo particular é criado para cada um de nós por mecanismos de previsão capazes de criar e refinar constantemente uma teoria sobre quem somos e sobre o que vamos fazer ou desejar (PARISER, 2012: 14). As características da bolha de Pariser, vem ao encontro da descrição do controle social descrito por Deleuze como uma ferramenta de controle de curto prazo e de rotação rápida e contínua.

A Internet se tornou o mais importante dispositivo de marketing para as empresas. David Lyon descreveu o momento em que os bancos de dados offline de empresas de marketing começaram a ser enriquecidos com base de dados online geradas na Internet. Esta combinação permitiu aos profissionais de marketing mesclarem informações individuais identificáveis relacionadas a dados postais, com evidências de hábitos de compra ou interesses adquiridos pelo rastreamento do uso da Internet.

Os modelos de negócios criados a partir do rastreamento de comportamento online de usuários cresceram paralelamente à explosão da web. Os rastreadores também conhecidos como *web trackings* não eram percebidos pela grande economia no lançamento comercial da Internet em 1995. Apesar disso, já eram vistos com bons olhos pelas empresas de mídias digitais como uma potencial ferramenta para estreitar o relacionamento entre usuários e consumidores.

Nos primeiros *papers* publicados por Varian, em 1996, as orientações e as perspectivas para uma economia de dados já eram ponderadas e, aproximadamente uma década após, foi possível perceber que o acúmulo de dados comportamentais se tornou cada vez mais massivo, trazendo reflexos e implicações nas relações sociais e políticas, e gradualmente foram constituindo um ambiente favorável ao capitalismo de informação, definido por Shoshana Zuboff como capitalismo de vigilância (ZUBOFF, 2015: 75).

A definição do capitalismo de vigilância de Zuboff, fundamenta-se em tecnologias como o big data, na coleta massiva de dados pessoais nas redes cibermediadas, em sua extração e análise para a exploração do mercado. Ela ainda afirma que “essa nova forma de capitalismo informacional pretende prever e modificar o comportamento humano como meio para a produção de lucros e o controle do mercado” (ZUBOFF, 2015: 75).

Web cookies a origem dos mecanismos de estado e de rastreamento na web

Toda a comunicação cibermediada é capaz de produzir registros de todas as interações realizadas entre as partes comunicantes. Estes rastros são registrados em arquivos digitais, conhecidos tecnicamente como *logs*. Em um servidor de email por exemplo, um arquivo de log é criado e mantido automaticamente registrando as atividades realizadas por uma conta. Geralmente os logs registram o endereço da origem e destino, horário da interação e algum outro dado específico da aplicação ou protocolo responsável pela comunicação.

Posso classificar este método de rastreamento como “sem estado”, ou seja, não existe no lado do usuário nenhum tipo de armazenamento ou registro inserido em seu computador sem o seu consentimento que possibilite o servidor rastrear o comportamento. Dentro de suas limitações, este método permite somente a identificação do dispositivo que interaja com o serviço de rede.

Embora os *logs* e as tecnologias adjacentes de controle de estado nasceram por necessidades técnicas, em um curto espaço de tempo elas foram apropriadas pelas corporações baseadas em dados pessoais e comportamentais da economia informacional.

Com a disseminação do HTML e as tecnologias da camada de aplicação da Internet, os métodos de identificação de dispositivos passam para o estágio de compreender o comportamento do usuário. Muitas destas técnicas surgem a partir da necessidade do comércio e publicidades online. Bujlow *et al*, em seu paper aponta cinco principais grupos de mecanismos usados para o rastreamento de usuários por meio de sessões de acesso, armazenamento no cliente, cache³ do cliente, fingerprinting (impressão digital) e outras abordagens. (BUJLOW *et al*, 2017: 1)

Bujlow *et al.*, enfatiza seu trabalho em três métodos que utilizam mecanismos rastreamento de comportamento visto que são ricos em termos de uso de diversas tecnologias. A seguir uma breve descrição destes:

- **Rastreamento por mecanismo de armazenamento:** depende do armazenamento explícito dos dados no computador do usuário. São mais avançados que os métodos “sem estado”, e são capazes de reconhecer características do computador.
- **Rastreamento por mecanismo de cache:** Funciona de maneira semelhante às técnicas de rastreamento por mecanismo de armazenamento, entretanto o identificador não é armazenado no computador do usuário. Neste tipo de rastreamento um usuário pode ser identificado pela disponibilidade de arquivos em cache ou por informações armazenadas em

³ Memória cache é um tipo de memória computacional temporária.

metadados de arquivos.

- **Rastreamento por mecanismos de impressões digitais (fingerprinting):** Este mecanismo abrange todas as técnicas que tendem a reconhecer um usuário. A identificação do usuário é realizada por intermédio de informações geradas ou entregues pelo próprio dispositivo do usuário. Várias técnicas de rastreamento e identificação são combinadas para garantir que exista entropia suficiente.

As atuais técnicas de rastreamento nem sempre podem ser facilmente categorizadas em um dos três métodos, pois as vezes possuem características de vários tipos. Ademais, para suprir a deficiência de algumas destas técnicas é comum que elas possam ser agregadas para ampliar o poder e a precisão do rastreamento.

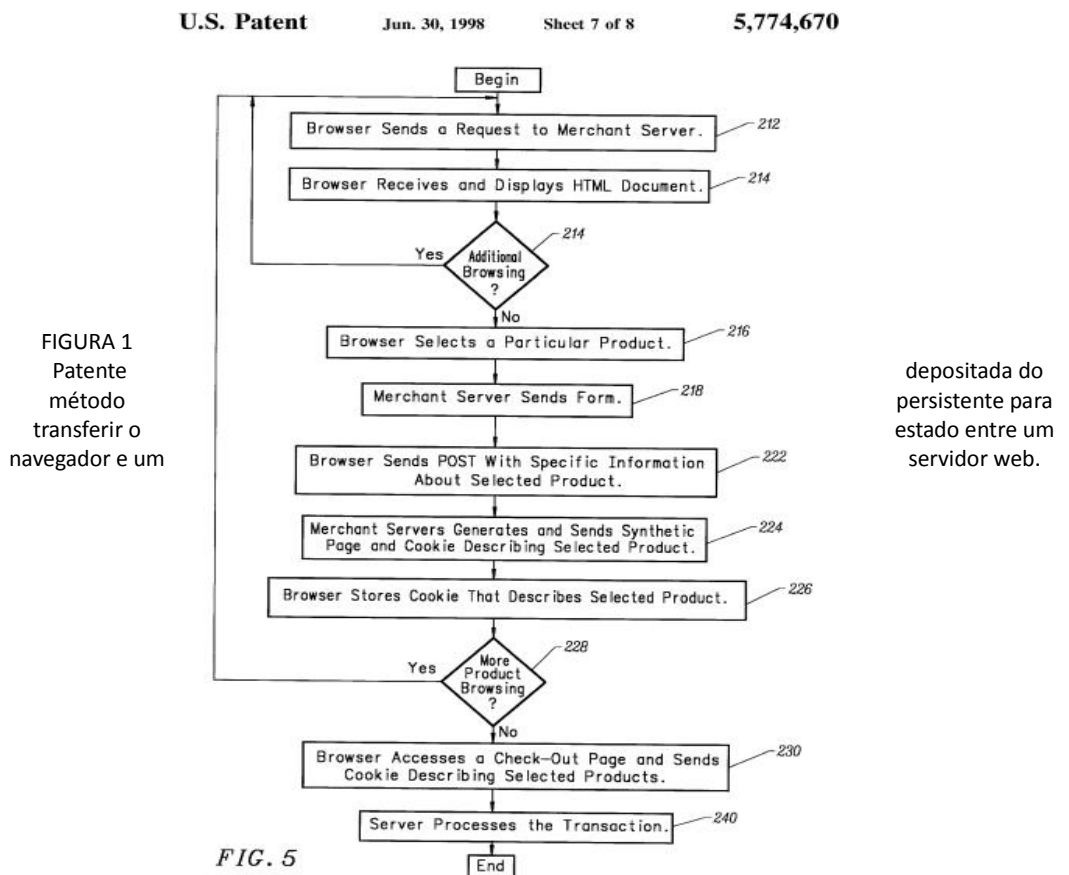
Ao longo de anos os navegadores apresentam-se como um grande campo fértil para o desenvolvimento de mecanismos de rastreamento de comportamento online. Uma das primeiras ferramentas de rastreamento computacional invasiva em camada de aplicação utilizada para fins comerciais na Internet, são os *web cookies*, ou simplesmente cookies. Segundo Tanenbaum (2003), este nome derivou de uma antiga gíria dos programadores, onde um programa solicita algo a um servidor e recebe de volta algo que provavelmente precise apresentar mais tarde para conseguir realizar alguma tarefa (TANENBAUM: 667).

O cookie é um pequeno arquivo de texto com tamanho máximo de 4 Kb, muitas vezes incompreensível para os usuários, que é depositado constantemente enquanto uma navegação em páginas da Internet é realizada. O W3C define cookie como “dados enviados por um servidor da web para um cliente da web, para serem armazenados localmente pelo cliente e enviados de volta ao servidor em solicitações subsequentes”⁴. Ele permite que os sites acessados obtenham informações específicas sobre quem está acessando suas páginas. Cada vez que você acessa um site, o seu navegador envia de volta o cookie correspondente a página acessada (servidor web) para que ele possa manter o controle de sua atividade, ou seja, sua experiência neste site.

⁴ Tradução do próprio autor para a definição disponível na página do W3C web Characterization Terminology & Definitions Sheet. “Data sent by a web server to a web client, to be stored locally by the client and sent back to the server on subsequent requests.” Disponível em: <https://www.w3.org/1999/05/WCA-terms/01>. Acesso em 10/07/2018.

O principal propósito do cookie é identificar usuários e possivelmente preparar páginas personalizadas ou para salvar as informações de sessão de um site, entre outras possibilidades. Portanto, ao retornar a um site, em vez de ser exibida uma página de boas-vindas genérica, você terá uma página de boas-vindas já com seu nome. A primeira aplicação dos cookies foi no servidor web da Netscape, sua criadora.

Seu idealizador, o programador Lou Montulli, também um dos criadores do Lynx, o primeiro *browser* ainda em modo texto, depositou a patente US5774670A⁵ em 1995, onde já descrevia um método persistente para transferir o estado entre um navegador e um servidor web conforme registrado na FIGURA 1. Montulli, também foi um dos fundadores da Netscape Communications, empresa responsável pelo desenvolvimento das primeiras versões gráficas dos navegadores web, é considerado como um dos principais idealizadores dos mecanismos de rastreamento de comportamento online.



⁵ Patente US Grant, 5774670A - Persistent client state in a hypertext transfer protocol based client-server system. Disponível em: <https://patents.google.com/patent/US5774670A/en>

O primeiro padrão informal que descreve os cookies, tecnologia que modificaria o comportamento padrão de uma navegação web e criaria modelos de negócios suportados pelo rastreamento de comportamento dos usuários, foi descrito no próprio site da Netscape Communications Corporations⁶.

cookies are a general mechanism which server side connections (such as CGI scripts) can use to both store and retrieve information on the client side of the connection. The addition of a simple, persistent, client-side state significantly extends the capabilities of web-based client/server applications⁷.

Desde sua concepção uma navegação HTTP não possui estado, ou seja, o navegador solicita uma página a um servidor web e quando a confirmação da entrega é feita a conexão é encerrada e o servidor “esquece” que já viu este usuário. Sendo assim, a cada clique que o usuário realiza em uma página, esta é considerada como uma nova conexão sem relacioná-la com conexões anteriores ou futuras.

Quando a web ainda era limitada a uma ferramenta para recuperação documentos publicamente disponíveis, o modelo padrão de navegação HTTP era bem adequado. Contudo, quando foi necessário ela assumir outras funcionalidades como autenticação de usuários, compras online e personalização de portais web, esse modelo causou problemas. Os servidores não conseguiam localizar os usuários apenas observando seus endereços IPs, uma vez que os computadores podem ser compartilhados por vários usuários e principalmente, o endereço IP identifica apenas o computador e não o usuário (TANENBAUM, 2003: 666).

Sendo assim, os cookies resolveram os problemas do estado de uma conexão HTTP, permitindo um servidor web não apenas identificar o comportamento de um computador por meio de logs, tampouco aguardar as informações obtidas por meio de um formulário de registro disponível em uma página para criar um perfil de visitante teoricamente preciso.

A primeira versão disponível publicamente do Netscape Navigator em setembro de 1994, já suportava o gerenciamento de estado, mesmo considerando que este fato não fosse bem conhecido

⁶ O site da Netscape não está mais disponível na Internet. Uma versão original está disponível em: https://curl.haxx.se/rfc/cookie_spec.html. Acesso em 04/06/2018.

⁷ Tradução livre: Os cookies são um mecanismo que as conexões do lado do servidor (como scripts CGI) podem usar para armazenar e recuperar informações no lado do cliente da conexão. A adição de um estado simples do lado do cliente e persistente estende significativamente os recursos de aplicativos cliente/servidor baseados na web.

na época. Segundo Kristol, o mecanismo foi apresentado a pedido de um dos clientes da Netscape (2001: 158).

A Netscape utilizou os cookies a primeira vez para verificar se os usuários já haviam visitado o seu site. Neste primeiro momento os cookies também apresentaram uma solução útil para carrinhos de compras virtuais, permitindo que os sites de comércio eletrônico lembrassem o que o usuário estava comprando da última vez que visitou o site⁸.

Também conhecido como cookie de sessão este foi o primeiro propósito para este mecanismo de rastreamento. Este cookie é geralmente excluído do computador do usuário assim que ele clicar na função “sair” do site. Na descrição do armazenamento do lado do cliente no site do W3C⁹, os cookies de sessão não coletam informações do computador do usuário e tipicamente sua forma de gerenciar a identificação de sessão não relaciona pessoalmente o usuário.

Ainda na descrição do armazenamento do lado do cliente da W3C, é ressaltado o importante papel dos cookies no rastreamento do comportamento e na criação do perfil de usuário, da mesma maneira que estas informações são importantes para as estratégias de marketing direcionado das empresas de publicidade. A descrição ainda reconhece que habitualmente um usuário não acessaria um site de publicidade para que estes dados pudessem ser analisados por estas corporações, demonstrando um obstáculo para estas empresas¹⁰.

Uma das formas para resolver este obstáculo é a aplicação de outro tipo de cookie, chamado de terceira parte, ou terceiros. Eles são criados a partir de um servidor que não seja o que o usuário acessou originalmente (conhecidos cookies de primeira parte), ou seja, mediante acordos comerciais alguns sites permitem que cookies de outros servidores sejam enviados para o computador do usuário. A maior parte destes cookies são gerenciados por corporações especializadas em marketing digital, publicidade ou grandes bancos de dados privados de usuários. São utilizados pelos sites acessados como fonte para propagandas ou buscas direcionadas.

⁸ Are cookies crumbling our privacy? We asked an expert to find out. Disponível em: <https://www.digitaltrends.com/computing/history-of-cookies-and-effect-on-privacy/>. Acesso em 10/03/2018.

⁹ “They typically store information in the form of a session identification that does not personally identify the user.” W3C – Client-side Storage. Disponível em <https://www.w3.org/2001/tag/2010/09/ClientSideStorage.html>. Acesso em 08/06/2018.

¹⁰ “The most important use of cookies however, and the most controversial, is to use cookies for tracking where you go and what you do there. These are typically used by advertising sites but you do not visit any of the advertising websites, so how can they get their cookies into your local storage?”

Um único site pode depositar em um computador dezenas de cookies de terceira parte. Avelino e Silveira (2016), realizaram uma pesquisa para aferir a quantidade de cookies enviados para um usuário a partir do acesso aos dez sites de notícias mais acessados no Brasil. Em conjunto, os dez sites depositaram 408 (quatrocentos e oito) cookies, sendo que destes 241 (duzentos e quarenta e um) eram de grupos de empresas americanas. Estes cookies sempre foram considerados a grande ameaça a privacidade do usuário. Ao longo do tempo, essas empresas conseguem desenvolver um histórico detalhado dos tipos de sites que o usuário frequentou.

Nesta época a mídia começou a reportar as potenciais ameaças à privacidade. O site cookie Central adverte que o propósito original do cookie foi subvertido por entidades inescrupulosas que semeiam cookies e por meio desta ação encontraram uma maneira de usá-los para rastrear os movimentos de usuário pela web, recuperando-os de tal forma que lhes permitam construir perfis detalhados de seus interesses, hábitos e estilo de vida¹¹.

Os cookies podem permanecer em um computador o período de tempo definido pelo servidor que o concebeu, não sendo excluídos automaticamente após o navegador ser fechado, exceto como já mencionado os cookies de sessão. Esta característica de persistência para rastrear o comportamento de usuários é constantemente usado por anunciantes para registrar informações sobre os hábitos de navegação durante um período determinado de tempo. Ademais, estes cookies, também conhecido como cookie persistente ou de rastreamento, são utilizados para personalizar a experiência do usuário, armazenando suas preferências (como, por exemplo, idioma, temas e últimas pesquisas) de modo que elas possam ser lembradas em uma próxima visita.

A IETF iniciou o processo para desenvolver um padrão para o gerenciamento do estado na Internet em abril de 1995 pressionada pelas preocupações sobre privacidade publicadas na época (KESAN; SHAH, 2004: 301). No final de 1995 três propostas para adicionar estado ao HTTP estavam circulando na comunidade técnica. O IETF cria um “subgrupo de estado” para que uma abordagem única pudesse ser recomendada pelo grupo de trabalho. O subgrupo liderado por David Kristol era composto por oito pessoas, incluindo Lou Montulli (KRISTOL, 2001: 159).

¹¹ “Unfortunately, the original intent of the cookie has been subverted by some unscrupulous entities who have found a way to use this process to actually track your movements across the web. They do this by surreptitiously planting their cookies and then retrieving them in such a way that allows them to build detailed profiles of your interests, spending habits, and lifestyle.” Disponível em: <http://www.cookiecentral.com/cookie5.htm>. Acesso em 10/03/2018.

O grupo de trabalho identifica dois problemas em relação a especificação de cookies da Netscape, na qual um cookie pode ser compartilhado entre vários servidores, com base em seus nomes de domínio. O primeiro os cookies podiam “vazar” para servidores diferentes daqueles destinados pelo servidor de origem. O segundo, um servidor em um domínio pode causar um ataque de negação de serviço, inadvertidamente ou intencionalmente, enviando cookies que interromperão um aplicativo em execução em outro servidor no mesmo domínio (KRISTOL, 2001: 159).

Em fevereiro de 1996 o grupo identificou o que considerou uma ameaça relevante os cookies de terceiros. A preocupação do grupo se dava pelo não consentimento do usuário em receber cookies de terceiros ao acessar um site, ou mesmo saber, que seu navegador irá acessar outro servidor e receberá um cookie a partir dele (KRISTOL, 2001: 159).

Em fevereiro de 1997, o IETF coordenado por David Kristol e Lou Montulli lança a RFC 2109¹² “HTTP State Management”, onde começa a considerar a proposta apresentada por grupos de privacidade, consumidores e educadores para reparar alguns problemas dos cookies. O seu primeiro padrão para o gerenciamento do estado e baseava-se em uma tecnologia diferente dos cookies. Em julho do mesmo ano é lançada sua revisão eventualmente mudando sua proposta para o modelo de cookies da Netscape, sugerindo que os usuários tenham mais controle sobre a intrusão em seus dispositivos (HELLING, 1998). Isso ocorreu principalmente porque a versão Netscape era um modelo de trabalho onipresente que se tornara um modelo padrão de fato (KESAN; SHAH, 2004: 301). Ela especifica que cookies de terceiros não eram permitidos ou pelo menos não habilitados por padrão.

O IETF sugeriu assim algumas soluções para o problema da privacidade segundo Helling (1998):

- O usuário deve ser capaz de rejeitar todos os cookies.
- O usuário deve saber quando uma sessão com estado está ocorrendo.
- O usuário deve ser capaz de gerenciar o cookie com base em sua origem (domínio).

¹² Disponível em <https://www.ietf.org/rfc/rfc2109.txt>. Acesso em 05/06/2018.

O objetivo do IETF era tornar o rascunho de uma página da Netscape em um padrão mais preciso para cookies. No entanto, este processo logo se deparou com problemas. A IETF descobriu que a implementação de cookies pela Netscape estava repleta de problemas de privacidade e segurança (KESAN; SHAH, 2004: 301).

Quando a segunda versão do Navigator foi lançada, uma nova tecnologia se aliou para aprimorar a criação de perfil de um visitante. De acordo com o artigo "Netscape tricks raise security concerns" de James Staten, os recursos de Javascript e cookie demonstraram uma grande preocupação quando a execução do Javascript poderia recuperar o e-mail do usuário entre outras informações privadas por meio do acesso a memória cache do Navigator (HELLING, 1998).

A publicação da RFC 2109 resultou em diversos artigos sobre cookies na imprensa e em protestos nas redes, pois muitos modelos de negócio já estavam operando por meio de cookies de terceira parte para fazerem publicidade segmentada. Por outro lado, as empresas de publicidades online achavam que a obrigação da RFC em desativar os cookies de terceiros por padrão era uma ameaça aos seus negócios. No entanto, o grupo de trabalho levou em consideração que a restrição dos cookies terceiros na RFC afetaria o modelo de negócios dos anunciantes que dependiam do rastreamento de usuários (KRISTOL, 2001: 160-161).

Neste momento, as empresas de publicidade já estavam usando cookies de terceiros. A recomendação sobre cookies de terceiros da RFC 2109 não foi seguida pelo Netscape tampouco pelo Internet Explorer. A IETF levou três anos para revisar a RFC 2109. Novamente, o atraso no desenvolvimento do padrão foi em grande parte resultado de problemas de privacidade com cookies de terceiros. Neste período ainda uma grande força de oposição formada por empresas de publicidade na web procuravam garantir que os consumidores pudessem receber cookies de terceiros. Contudo, os membros do IETF mantiveram seu suporte para desativar os cookies de terceiros por padrão na nova revisão lançada em outubro de 2000, a RFC 2965 (KESAN; SHAH, 2004: 303).

Ainda após o lançamento da RFC 2109 várias manifestações questionavam a prática do uso dos cookies e as implicações relacionadas a privacidade do usuário. Em junho de 1997, um conjunto de empresas lideradas pela Netscape Communications, a FireFly Network e a Verisign, propõem ao W3C um padrão para coleta de informações sobre usuários por meio de uma base restrita e voluntária chamado Open Profiling Standard (OPS). O OPS permitiria aos usuários criarem seus próprios perfis dando assim o controle sobre seu perfil pessoal e a capacidade de gerenciar quais informações

seriam compartilhadas com os sites¹³. Contudo os documentos não saíram do status de discussão (NOTA), ou seja, não tiveram qualquer endosso do W3C.

A proposta foi organizada em uma série de três especificações relacionadas ao OPS. O primeiro documento "Proposal for an Open Profiling Standard" descreve a estrutura (framework) para troca de perfil, incluindo a organização de dados e primitivas operacionais. O segundo documento, "Implementation of OPS Over HTTP", aborda a implementação do OPS sobre o protocolo HTTP. Já o terceiro, "Standard OPS Practices" aborda os mecanismos de permissão sugeridos para controle de usuários e prestadores de serviços na troca de dados de perfis¹⁴.

Em abril de 2011 a RFC 6265 é publicada como uma nova revisão do padrão de Mecanismo de Gerenciamento de Estado HTTP, tornando obsoleta a RFC 2965. Esta revisão assume que os cookies possuem um histórico que colocam sua segurança e privacidade em suspeita¹⁵.

Para superar estas barreiras algumas soluções de contorno surgiram para potencializar os mecanismos de rastreamento intrusivos. A seguir apresento uma breve exposição desta evolução.

Sincronização de cookies

O fato dos cookies serem específicos de um domínio, ou seja, só podem ser lidos pelo domínio que o criou, gera algumas limitações para as corporações fornecedoras de plataformas de publicidades, segmentação e classificação de usuários, bem como anunciantes. Portanto, para atingir um público-alvo com precisão, as empresas de publicidade online buscam fontes de dados externas para enriquecerem suas bases de dados. Parcerias entre empresas diferentes e contratos de compra de dados de usuários são algumas estratégias.

Uma prática ainda utilizada é construir uma identificação de um usuário (ID) capaz de ser mapeada entre vários sistemas e plataformas de publicidade. Este processo é conhecido como sincronização de cookie e funciona quando duas ou mais plataformas de publicidade online mapeiam os IDs únicos uns dos outros e, subsequentemente, compartilham informações que eles reuniram sobre o mesmo usuário.

¹³ Open Profiling Standard. Disponível em: <https://searchmicroservices.techtarget.com/definition/Open-Profiling-Standard-OPS>. Acesso em 05/07/2019.

¹⁴ Proposal for an Open Profiling Standard. Disponível em: <https://www.w3.org/TR/NOTE-OPS-FrameWork>. Acesso em 10/02/2019.

¹⁵ RFC 6265 disponível em: <https://www.ietf.org/rfc/rfc6265.txt>. Acesso em 07/07/2018.

Flash cookies

Com o objetivo de aumentar a confiabilidade dos métodos de rastreamento na individualização dos acessos Web, em 2005 a empresa de publicidade online United Virtualities, apreensiva com estudos que apontavam a época que mais de 30% dos usuários da Internet já excluíam cookies pelo menos uma vez por mês, anunciou o desenvolvimento de um sistema imune as ferramentas e recursos de exclusão de cookies. Denominado Persistent Identification Element (PIE), ou Elemento de Identificação Permanente, é um marcador exclusivo aplicado no navegador do usuário que não pode ser excluído por nenhum software anti rastreamento (SOLTANI et al., 2009: 1).

O PIE utilizou o recurso Local Shared Object, ou objeto compartilhado local, do Macromedia Flash Player da Adobe. Também conhecido como Flash cookies, este rastreador apresenta mais persistência do que os cookies padrão, pois são armazenados em locais diferentes, ainda não possuem data de expiração e podem conter até 100 KB de informações, capacidade vinte e cinco vezes superior aos 4 KB de um cookie.

Web storage e html 5

Com a finalidade de ampliar a capacidade de armazenamento de dados e aumentar sua persistência, além de superar as restrições do Web cookies, em 2014¹⁶ o W3C publicou a recomendação oficial do HTML5¹⁷, onde foram apresentadas novas especificações para melhor o suporte para aplicações web interativas. Este novo padrão também introduziu novas técnicas de rastreamento de usuários como o armazenamento local e o de sessão.

O funcionamento da técnica de armazenamento local do HTML5, é semelhante ao funcionamento de um web cookie. Os sites podem usar vários objetos para armazenar as informações necessárias e, posteriormente, recuperá-las. Foram inseridos dois objetos semelhante as características dos cookies de sessão e de navegação, o sessionStorage e o localStorage.

O storagelocal armazena os dados de navegação do usuário sem data de expiração, enquanto o sessionStorage armazena os dados apenas na sessão atual. Enquanto o limite de tamanho de um cookie é 4 Kb, um objeto no HTML5 é até 10 MB.

¹⁶ HTML5 specification finalized, squabbling over specs continues. Disponível em: <https://arstechnica.com/information-technology/2014/10/html5-specification-finalized-squabbling-over-who-writes-the-specs-continues>. Acesso em 10/05/2019.

¹⁷ HTML (Hypertext Markup Language ou em português Linguagem de Marcação de Hipertexto), é a linguagem utilizada para a construção de páginas Web.

Considerações finais

Este trabalho brevemente investigou as dinâmicas de funcionamento dos três principais tipos de mecanismos intrusivos de rastreamento de comportamento online em clientes web que culminou na coleta invasiva de dados e vislumbrou uma sólida economia de dados pessoais, que compõe parte da economia da privacidade. O mecanismo de sincronização de cookies e o armazenamento local padrão do HTML5, são as principais ameaças a privacidade contemporâneas. As tecnologias Web, bem como os clientes web evoluem rapidamente, pesquisas contínuas são necessárias para acompanhar a evolução destes mecanismos.

Referências

- AVELINO, Rodolfo da Silva; SILVEIRA, Sérgio Amadeu da. **A Dependência Do Rastreamento Comportamental Online Para A Economia Globalizada**. In: IV SIMPÓSIO INTERNACIONAL LAVITS, Não use números Romanos ou letras, use somente números Árabicos., 2016, Buenos Aires: Lavits, 2017. p. 1 - 13. Disponível em: <http://lavits.org/wp-content/uploads/2017/08/P5_Avelino1.pdf>. Acesso em: 01 jun. 2018.
- BAUMAN, Zygmunt. **Vigilância líquida: diálogos com David Lyon**. Tradução: Carlos Alberto Medeiros. Rio de Janeiro: Zahar, 2013.
- BUJLOW, T., CARELA-ESPAÑOL, V., SOLÉ-PARETA, J., BARLET-ROS, P. **A survey on web tracking: Mechanisms, implications, and defenses**. Proceedings of the IEEE, vol. PP, no. 99, pp. 1–35, 2017. DOI:10.1109/JPROC.2016.2637878.
- CASTELLS, Manuel. **A sociedade em rede**. Paz e Terra, São Paulo, 2011.
- DELEUZE, Gilles. **“Post-scriptum sobre as Sociedades de Controle”**. In **Conversações**. São Paulo: Editora 34, 1992.
- HELLING, Bill. **Web-site Sensitivity to Privacy Concerns: Collecting Personally Identifiable Information and Passing Persistent cookies**. First Monday. [s.i], p. 1-27. fev. 1998. Disponível em: <<http://journals.uic.edu/ojs/index.php/fm/article/view/574/495>>. Acesso em: 18 jun. 2018. DOI: <https://doi.org/10.5210/fm.v3i2.574>
- KESAN, Jay P., SHAH, Rajiv C. **DECONSTRUCTING CODE**. Yale Journal of Law & Technology. p. 279-389. 2004. Disponível: <https://ssrn.com/abstract=597543>. Acesso em 15/06/2018.
- KRISTOL, David M. HTTP cookies: Standards, Privacy, and Politics
- LYON, David. **Surveillance as Social Sorting. Privacy, risk, and digital discrimination**. London and New York. Routledge, 2003.
- _____. **Identifying Citizens. ID Cards as Surveillance**. Cambridge: Polity Press. 2009.
- LAZZARATO, Maurício. **As revoluções do capitalismo**. Rio de Janeiro: Civilização Brasileira, 2006.
- PARISER, Eli. **O filtro invisível: o que a internet está escondendo de você**. Tradução: Diego Alfaro. Rio de Janeiro. Zahar, 2012.
- SHAPIRO, Carl, VARIAN, Hal. **A economia da informação: como os princípios econômicos se aplicam à era da Internet**. Rio de Janeiro: Campus, 1999.
- Soltani A.,Canty S.,Mayo Q.,Thomas L., and Hoofnagle C. J. Flash Cookies and Privacy. UC Berkeley

survey of flash cookie adoption on the web and related privacy concerns, August 2009.

TANENBUAM, Andrew. **Redes de Computadores**. 4ª ed. Rio de Janeiro, Elsevier, 2003.

VARIAN, Hal R. **Economic Aspects of Personal Privacy**. In: LEHR, William H.; PUPILLO, Lorenzo Maria (Ed.). **Internet Policy and Economics: Challenges and Perspectives**. 2. ed. Massachusetts: Springer, 1996. Cap. 7. p. 101-109.

_____. **Beyond Big Data, Business Economics**. 2014;49:27-31.
<https://doi.org/10.1057/be.2014.1>

ZUBOFF, Shoshana. **Big other: surveillance capitalism and the prospects of an information civilization**. London: Journal of Information Technology, 2015, 75–89. Disponível em: <http://ssrn.com/abstract=2594754>. Acesso em 01 de julho de 2018.