Dispositivo de vigilância algorítmica: algoritmos rastreadores e coleta de dados

Algorithmic surveillance device: tracking algorithms and data collection

Dispositivo de vigilancia algorítmica: seguimiento de algoritmos y recopilación de datos

Recebido em 02-09-2021
Modificado em 12-11-2021
Aceito para publicação em 11-12-2021

https://doi.org/10.47456/simbitica.v8i4.37348

Maria Rita Pereira Xavier

Doutora em Ciências Sociais pelo Programa de pós-graduação em Ciências Sociais da Universidade Federal do Rio Grande do Norte. Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Graduada e Mestre em Ciências Sociais pela Universidade Federal do Rio Grande do Norte. E-mail: mariaritax_@hotmail.com

Alexsandro Galeno Araújo Dantas

Pós-Doutor pela Universidade de São Paulo (2015); Doutor em Ciências Sociais pela Pontifícia Universidade Católica de São Paulo (2002). Professor do Instituto Humanitas da Universidade Federal do Rio Grande do Norte. E-mail: alexgalenno@gmail.com

Resumo

Este artigo, parte componente de pesquisa de doutorado, intenta analisar o uso de algoritmos rastreadores por meio de smartphones como uma forma corporativa/empresarial de vigilância. O papel do smartphone seria o de abrir caminho para a anuência da entrega de dados apoiada tanto na cooperação voluntária dos usuários quanto no uso de algoritmos rastreadores de dados, numa prática conhecida como *tracking*. O rastreamento é um mecanismo para registrar padrões de navegação de um usuário na internet. A metodologia empregada é o uso de algoritmos de contravigilância desenvolvidos por pesquisadores como Karaj *et al.* (2019) e Macbeth *et al.* (2016). Foram pesquisados algoritmos rastreadores em aplicativos e sites, além de identificar as diferentes técnicas pelas quais se dá o rastreio de dados. Utilizou-se o algoritmo "*Who tracks me*", disponibilizado por Karaj *et al.* (2019), nos 4 sites mais acessados do Brasil segundo o ranking Alexa, sendo possível identificar uma grande prevalência de rastreadores em todos eles.

Palavras-chave: dispositivo; rastreadores, smartphones; algoritmos; vigilância.

Introdução

Em 1665, ao examinar um pedaço de cortiça num microscópio composto, Robert Hooke observou uma estrutura semelhante a uma cela, era a primeira vez que a estrutura de células humanas estava sendo vista. O responsável por batizar a célula biológica era um agrimensor da cidade de Londres, que auxiliou na medição e divisão dos terrenos da cidade após o grande incêndio de 1666. Não por acaso, em inglês, o termo *surveyor* (agrimensor) vem de *survey* (inspecionar, avaliar, questionar, fiscalizar), a mesma origem do termo *surveillance* (vigilância/fiscalização). Hooke usou um termo de vigilância comum à sua profissão para batizar a descoberta. Contudo, a este trabalho, cabe indagar por que chamamos os aparelhos móveis de celular? De onde vem essa aplicação do termo? Na verdade, o que há de celular no telefone móvel é a rede. Na figura 1, é possível ver o diagrama que V.H. MacDonald usou em seu artigo, em 1979, para o *The Bell System Technical Journal*, intitulado de "*The Cellular Concept*" (Madrigal, 2011).

Figura 1 - Rede de telefonia: layout celular que ilustra a reutilização de frequência

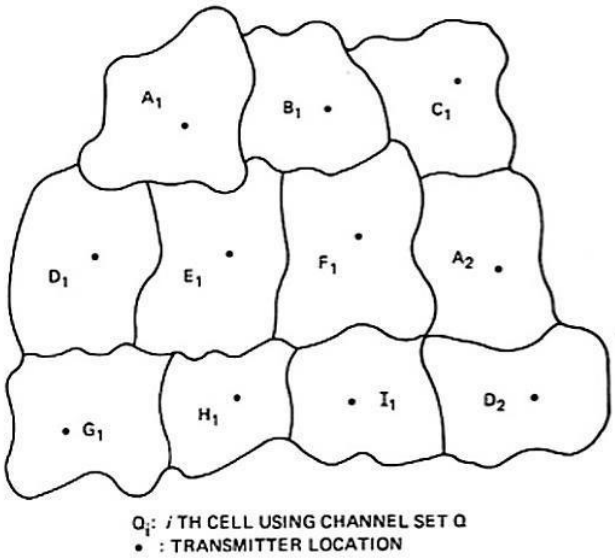


Fig. 1—Cellular layout illustrating frequency reuse.

Fonte: Madrigal (2011)

Criado nos anos 1970, o celular só se tornou um item de consumo no início dos anos 1990, contudo, foi a introdução do smartphone no mercado, em 2007 que inseriu esse artefato de vez no dia a dia das pessoas. E com a arquitetura de anúncios do Google, baseada nos históricos de busca, surgiu um lugar privilegiado de aquisição de dados dos usuários. O que proporcionou

uma infraestrutura de vigilância corporativa baseada em técnicas de persuasão, interpretação, coleta e monitoramento de dados, também conhecidas como *crowdforcing*, *profiling* e *tracking*.

Os grandes conjuntos de dados obtidos com a internet estão cada vez mais difíceis de serem acessados pelos métodos tradicionais das Ciências Sociais. O estudo da sociedade por meio das tecnologias da informação e computação são chamados de Ciências Sociais Computacionais (Conte *et al.*, 2013). Levando em consideração a abordagem interdisciplinar das Ciências Sociais Computacionais e o seu entendimento de que "a complexidade social é resultado da adaptação humana aos desafios do meio ambiente (princípio de Simon)" (Conte *et al.*, 2013:15), o presente artigo tem o intuito de demonstrar como funciona a obtenção de dados por meio de algoritmos rastreadores em aplicativos e sites, além de identificar diferentes técnicas pelas quais se dá esse rastreio. Para tal, utiliza-se o algoritmo de contravigilância "*Who tracks me*", disponibilizado por Karaj *et al.* (2019), nos 4 sites mais acessados do Brasil segundo o ranking Alexa, sendo possível identificar uma grande prevalência de rastreadores em todos eles.

O objetivo geral deste artigo é analisar o uso de algoritmos rastreadores por intermédio de smartphones como uma forma corporativa/empresarial de vigilância. A proposta teórica é utilizar o conceito de dispositivo (Foucault, 1996) para analisar a formação e atuação de um dispositivo de vigilância algorítmica, mediante o uso de algoritmos rastreadores em navegadores da internet, especialmente de smartphones. Parte-se da ideia de que o dispositivo vigente ainda é disciplinar, mas que está em vias de atualização, visto que se utiliza das formas de comunicação em rede como fonte de alimentação de suas configurações de saber. Desse modo, as vigilâncias algorítmicas corporativas e estatais são aqui entendidas como preenchimentos estratégicos do dispositivo, mediadas por tecnologias e incorporadas à vida cotidiana. Presume-se que essas técnicas foram se imbricando na vida social, tanto por meio de infraestruturas de informação quanto pela dependência digital nas relações do dia a dia (Lyon, 2018).

O conceito de dispositivo

A etimologia da palavra "celular" vem mesmo de "cela", contudo, apesar de a técnica de detenção em celas ser disciplinar, a sua história remonta à era jurídico-legal e à ordem religiosa. Ao retomar a história dos deslocamentos e da utilização da técnica celular de detenção, Foucault (2008b) defende que seria possível delimitar o momento em que o dispositivo de disciplina celular passa a ser empregado no sistema penal comum, bem como que conflitos ele suscita e até como regride. O dispositivo envolve jogos de poder e formação de saber, já que a sua noção evidencia a relação íntima entre poder e saber ao mostrar que não existe a formação de um

campo de saber neutro. Isso porque as intenções, interesses e estratégias de poder se apropriam do campo segundo seus próprios fins, desviando-o do que seriam os seus propósitos iniciais, essenciais ou autênticos (Bruno, 2013).

Na "*Microfísica do poder"*, Foucault (1996) define o dispositivo ao demarcar três características. Primeiro, o dispositivo é tido como um conjunto heterogêneo "que engloba discursos, instituições, organizações arquitetônicas, decisões regulamentares, leis, medidas administrativas, enunciados científicos, proposições filosóficas, morais, filantrópicas" (Foucault, 1996:244). Segundo, há uma delimitação do caráter da relação estabelecida entre esses elementos heterogêneos. A heterogeneidade dos elementos comporta tanto o dito quanto o não dito, não exatamente o não implícito, o oculto, mas sim o que se expressa em técnicas, procedimentos, ordenações espaciais, arquiteturais etc. O dispositivo consiste menos nos elementos e mais na rede que se estabelece entre eles. A terceira característica é o entendimento do dispositivo como um tipo de formação que tem uma função estratégica dominante de ser resposta a uma urgência de determinado momento histórico.

Um dispositivo pode ser definido tanto como uma estrutura de elementos heterogêneos quanto por uma determinada espécie de gênese, a qual possui dois momentos essenciais: a predominância de um objetivo estratégico, seguida da constituição do dispositivo como tal, e a sua continuação como dispositivo, uma vez que engloba um duplo processo. Um é constituído por um processo de *sobredeterminação funcional*, no qual os efeitos, positivos ou negativos, desejados ou não, constituem uma relação ou de ressonância ou de contradição uns com os outros; havendo uma demanda por uma rearticulação, um reajuste dos elementos heterogêneos que surgiram de modo disperso. O outro é o denominado processo de perpétuo *preenchimento estratégico*. Esse processo seria uma espécie de readequação engendrada pelo dispositivo para lidar com efeitos não previstos, involuntários e negativos, de modo a transformá-los em novas estratégias, que são destinadas a ocupar espaços vazios ou a transformar efeitos negativos em aspectos positivos a seu favor. A função estratégica é uma resposta a uma urgência determinada pelo momento histórico.

O surgimento do covid-19 é um bom exemplo da função estratégica em ação. Ao induzir a formação de um ordenamento em caráter de urgência totalmente distinto do que vinha sendo aplicado nas últimas décadas, a pandemia obriga o dispositivo de poder a se reorganizar, visto que as urgências requerem diagramas outros para lidar com o ordenamento de multiplicidades não previstas. É estabelecida uma espécie de novo modelo, que modula elementos da disciplina da peste – como a quarentena para todos, o isolamento total para os doentes e as medidas de

mitigação para o comércio – e os associa a técnicas da sociedade de controle[1], como o uso de dados de GPS (Sistema de Posicionamento Global)e de torres de celular para rastrear a circulação de pessoas nas cidades e de algoritmos para calcular as curvas de contágio. Agostinho (2017:9) nos explica que "a forma organiza tanto matérias quanto funções, ela organiza a prisão, o hospital, a escola, a fábrica e administra os corpos dos doentes, dos prisioneiros, dos estudantes. Assim, punir, educar, fazer trabalhar, são funções formalizadas". A questão foucaultiana seria pensar a dissociação entre os enunciados de poder, é por meio da forma que as relações de conteúdo e expressão são tecidas, produzindo enunciados e o regime de visibilidade. O panóptico é tanto um agenciamento concreto (ótico, um regime de visibilidade) quanto uma máquina abstrata que atravessa todas as funções enunciáveis (Agostinho, 2017).

Foucault (1975:254) diz que os dispositivos são "técnicas que asseguram o ordenamento de multiplicidades humanas". O que corrobora o raciocínio de que essas técnicas precisam ser atualizadas à medida que as multiplicidades vão se alterando, não só em decorrência das funções estratégicas impostas pelo poder ou das urgências, mas também porque as suas próprias brechas e linhas de fuga vão elaborando novas multiplicidades de resistência a serem apreendidas. É possível afirmar que as técnicas de ordenamento do dispositivo acompanham as mudanças pelas quais as multiplicidades passam.

O dispositivo de vigilância algorítmica

A importância do big data está na potência de conhecimento gerada pela quantidade de dados acumulada. A vigilância digital segue o caminho da classificação com vistas a governar condutas, mediante uma taxonomia distinta dos procedimentos disciplinares. Os processos algorítmicos de vigilância seguem uma lógica própria que é iniciada na navegação online, de modo que a maior quantidade possível de vestígios e traços de dados seja coletada. O passo seguinte é uma mineração para distinguir que tipo de informações foram recolhidas. Bruno (2013) elabora sobre esse processo ao dizer que,

> Sob o gigantesco fluxo de rastros pessoais em plataformas participativas, apresenta-se processos como *dataveillance* (vigilância de dados), *data mining* (mineração de dados) e *profiling* (perfilagem), que monitoram e classificam os dados, construindo saberes que sustentam uma vigilância proativa sobre indivíduos e populações (Bruno, 2013:127).

[1] Deleuze (1992) traz à tona o conceito de sociedade de controle ao alertar que foi o próprio Foucault quem anteviu esse processo: "Foucault é com frequência considerado como o pensador das sociedades de disciplina, e de sua técnica principal, o confinamento. Porém, de fato, ele é um dos primeiros a dizer que as sociedades disciplinares são aquilo que estamos deixando para trás, o que já não somos. Estamos entrando nas sociedades de controle, que funcionam não mais por confinamento, mas por controle contínuo e comunicação instantânea [...]. O que está sendo implantado, às cegas, são novos tipos de sanções, de educação, de tratamento [...]. Num regime de controle que nunca termina" (Deleuze, 1992:220).

Para a máquina algorítmica, o conteúdo não é o principal ponto, mas sim a movimentação do usuário. A indiferença formal do algoritmo (Haraway, 1992) não se importa com o que os sujeitos pensam, somente com tentar acumular a maior quantidade possível de dados. Esse é um ponto que difere o dispositivo atual do da disciplina, que fazia um investimento sobre o que o sujeito falava sobre si, para entender quem ele era e o que ele pensava sobre si (os seus gostos, escolhas, preferências e desejos). Zuboff (2015) explica esse processo:

> The extractive processes that make big data possible typically occur in the absence of dialogue or consent, despite the fact that they signal both facts and subjectivities of individual lives. These subjectivities travel a hidden path to aggregation and decontextualization, even though they are produced as intimate and immediate, tied to individual projects and contexts (Nissembaum, 2011 *apud* Zuboff, 2015:79)[2].

O que torna esses dados tão valiosos para os anunciantes são possíveis sinais de subjetividade que possam ter sido capturados. Na vigilância disciplinar, os dados eram mais populacionais, em um contexto de consentimento, presença física, biopolítica e a nível estatal. Na vigilância algorítmica, há um efeito performativo muito acentuado, um modelo de controle sem consentimento dos indivíduos, em que a correlação entre os dados é a maior responsável por gerar um padrão comportamental.

Os dados passaram a ser transacionais, o *small data* se apoia na extração, análise, personalização e experimento contínuo. São coletados em transações econômicas; em sensores acoplados a objetos, a pessoas e a lugares; por meio de bancos de dados governamentais e corporativos. O big data se institui a partir dos vestígios/rastros e da escala gigantesca decorrente da quantidade massiva de dados coletados (Zuboff, 2015). A principal intenção é captar os rastros de subjetividade, as formas de expressão e comunicação em ambientes informacionais, a mina de ouro está na riqueza de detalhes.

> Individual needs for self-expression, voice, influence, information, learning, empowerment, and connection summoned all sorts of new capabilities into existence in just a few years: Google's searches, iPod's music, Facebook's pages, YouTube's videos, blogs, networks, communities of friends, strangers, and colleagues, all reaching out beyond the old institutional and geographical boundaries in a kind of exultation of hunting and gathering and sharing information for every purpose or none at all. It was mine, and I could do with it what I wished! These subjectivities of self-determination found expression in a new networked individual sphere characterized by what Benkler (2006) aptly summarized as non-market forms of 'social production (Zuboff, 2015:79)[3].

[2] "Os processos extrativos que tornam o "*big data*" possível normalmente ocorrem na ausência de diálogo ou de consentimento, apesar de indicarem tanto fatos quanto subjetividades de vidas individuais. Essas subjetividades percorrem caminhos ocultos para agregação e descontextualização, apesar de serem produzidas como íntimas e imediatas, ligadas a projetos e contextos individuais" (Zuboff, 2015:79, tradução nossa).

[3] "As necessidades individuais de auto-expressão, voz, influência, informação, aprendizagem, empoderamento e conexão reuniram em poucos anos uma ampla gama de novas capacidades: pesquisas do Google, música do iPod, páginas do Facebook, vídeos do YouTube, blogs, redes, comunidades de amigos, estranhos e colegas, todos ultrapassando as antigas fronteiras institucionais e geográficas em uma espécie de exultação de caça, coleta e compartilhamento de informações para todos os propósitos, ou mesmo para nenhum. Isso era meu, e eu posso fazer

Nada é trivial para a coleta, captura-se todo o possível, porque os dados nunca são excessivos nem são limitados como o são matérias-primas naturais. A lógica é colete primeiro, analise depois. A correlação se faz suficiente porque não há necessidade de entender os porquês ou os motivos das ações, não há perguntas sobre as causas quando só o padrão importa. O nexo é o de que é possível produzir, ou induzir, comportamentos sem precisar entendê-los completamente, os mecanismos de *profiling* não funcionam com distinções entre verdadeiro/falso, a produção de comportamentos importa mais do que o erro ou o acerto. Essa indiferença formal pode ser vista no sistema de recomendação do *spotify*[4], quando uma música é recomendada e o usuário a escuta, o objetivo do algoritmo já foi atingido, não importando se a pessoa gostou ou não do que lhe foi recomendado.

Já que as origens do big data remontam a um projeto de extração baseado na indiferença formal em relação às populações, que são tanto a fonte de dados quantos os alvos finais desse processo, o big data se torna a peça fundamental para entender o quebra-cabeças da vigilância na sociedade de controle. É possível dizer que, se as coletas de dados estão cada vez menos interessadas em características identitárias, é porque estão mais centradas nos aspectos subjetivos, que antes eram de difícil acesso e que agora se encontram nos rastros digitais imersos na inumerável quantidade de dados disponíveis. De acordo com Zuboff (2015:76, tradução nossa), "a automação gera simultaneamente informação que proporciona um nível mais profundo de transparência a atividades que pareciam parcial ou totalmente opacas"[5].

O imperativo de compartilhamento está muito presente na cultura de vigilância. (Lyon, 2018). A conexão com o setor corporativo é ponto chave, já que "sob uma perspectiva corporativa, *prosumption*[6] e compartilhamento são a origem dos fluxos e inundações de dados sobre preferências, hábitos, opiniões e compromissos de usuários de tecnologia digital que podem ser usados para publicidade ou construção de sujeitos consumidores" (Lyon, 2018:163).

> O compartilhamento também pode ser pensado como um aspecto da exposição, para Kirstie Ball a exposição é explorada em termos de "economia política de interioridade" – em que as instituições associadas a tecnologia, mídia, emprego e consumo criam uma demanda ou mobilizam recursos para focar estados psicológicos ou comportamentos íntimos. A principal preocupação de Ball (2009) é que **a subjetividade tende a ser subestimada na literatura sobre vigilância, na qual muitas vezes é vista, principalmente em termos de opressão, coerção, ambivalência ou ignorância.** Contra isso, ela propõe que – no mínimo – a reflexividade, a performatividade, a

com isso o que eu quiser! Essas subjetividades de autodeterminação encontraram expressão em uma nova esfera individual em rede caracterizada pelo que Benkler (2006) resumiu adequadamente como formas não-mercantis de 'produção social'" (Zuboff, 2015:79, tradução nossa).

[4] Spotify é um serviço de streaming de música mais popular do mundo.

[5] "Automation simultaneously generates information that provides a deeper level of transparency to activities that had been either partially or completely opaque" (Zuboff, 2018:75).

[6] *Prosumption*: termo que se refere à articulação ou fusão entre produção (production) e consumo (consumption) de mídia, relativo, por exemplo, ao fenômeno recente da capitalização sobre conteúdos web produzidos por usuários.

> corporificação e o psicanalítico sejam trazidos mais claramente ao cenário. (Ball, 2009 *apud* Lyon, 2018:163-164).

Para a autora, é preciso relembrar que existem dimensões psicanalíticas da vigilância que podem ser esclarecidas, é preciso levar em conta que os sujeitos da vigilância fazem escolhas, mesmo que fugazes. Ainda que exista, de fato, partes negativas na exposição, como vulnerabilidade e abandono, há também um outro aspecto que é a busca por prazer e satisfação decorrente da exposição que o compartilhamento proporciona. De fato, quando a vigilância é suave e não muito intrusiva, os sujeitos têm uma tendência maior a entregar dados pessoais com facilidade. No entanto, tendo em vista que "as instituições incitam a diferentes tipos de reação à vigilância, é crucial não reduzir a experiência da vigilância a um formato unidimensional ou binário de 'aquiescência ou resistência'" (Lyon, 2018:166).

Crowdforcing e profiling

Ao considerar o fenômeno acelerado da exposição como uma atitude que pode ser feita de forma deliberada, reconhece-se que "existe mais nos sujeitos que disponibilizam seus dados do que a posição reducionista e passiva em que eles muitas vezes se encontram, a aceitação cega ou suave dos usuários não deve ser presumida por analistas" (Lyon, 2018:165). O que também não significa dizer que todas as ações provenientes de vigilância, especialmente a suave, são resultado de atitudes conscientes dos sujeitos. Se por um lado não se pode reduzir todos os sujeitos a um tipo de subjetividade resultante apenas de vigília e coerção, por outro lado, é inegável a amplitude que a vigilância algorítmica tomou; tanto por meio do consumo de dispositivos sociotécnicos quanto da inserção generalizada de sistemas de informação em diversos setores da vida social.

Nesse sentido, não adianta ser simplista e disseminar a ideia de que, para se manter livre de vigilância, basta não criar perfis em redes sociais ou não usar os serviços de grandes corporações. No atual cenário, não há mais fora, não é possível se ausentar dos sistemas de informação e comunicação, todos estão incluídos independentemente de serem usuários de serviços ou não (Kanashiro, 2016). Por oportuno, cabe elucidar a crítica ferrenha de Agamben (2009) ao identificar esse movimento:

> A futilidade daqueles discursos bem-intencionados sobre a tecnologia, que afirmam que o problema dos dispositivos se reduz àquele de seu uso correto. Esses discursos parecem ignorar que, se todo dispositivo corresponde a um determinado processo de subjetivação (ou, neste caso, de dessubjetivação), é de tudo impossível que o sujeito do dispositivo o use "de modo justo". Aqueles que têm discursos similares são, de resto, a seu tempo, o resultado do dispositivo midiático no qual estão capturados (Agamben, 2009:7).

Nessa perspectiva, Golumbia (2015) apresenta a ideia de *crowdforcing*, um mapeamento de parte da população que acaba envolvendo mesmo os que não tiveram dados mapeados. É uma prática que pode guiar valores de serviços, por exemplo, ao premiar os que disponibilizam os dados com descontos e "punir" os não mapeados com preços mais altos.

No Brasil, essa prática acontece quando as seguradoras de carro fazem distinção entre usuários que aceitam colocar geolocalizadores em seus carros e aqueles que não aceitam (Kanashiro, 2016). Um fenômeno ainda mais comum no país envolve muitas redes de farmácia, que cobram um preço menor para os clientes que realizam cadastro. Essa prática se espalhou não só em farmácias, mas em lojas, supermercados etc., o compartilhamento desses dados entre lojistas permite que estabelecimentos comerciais tenham dados de pessoas que nunca foram seus clientes. Isso porque, ao fazer o cadastro, as pessoas não estão cientes de que há essa prerrogativa de compartilhamento das informações com outras redes. O grupo de supermercados "Pão de Açúcar" criou um programa em que trocava dados e preferências de consumo, mediante um aplicativo, por descontos personalizados para os clientes.

> Os hábitos de consumo dos quase 12 milhões de membros de seus programas de fidelidade, o Pão de Açúcar Mais e o Clube Extra. A moeda de troca do Pão de Açúcar era um tesouro que estava enterrado debaixo de uma camada de algoritmos: o grupo abriu para a indústria toda a base de dados de seus programas de fidelidade. Os fornecedores têm acesso ao perfil de quem consome (e de quem ignora) seus produtos e podem fazer ofertas "nichadas" (Viri, 2017:1).

Ao privilegiar o grupo que fornece informações, a prática segue a lógica do *crowdforcing*, termo que reitera uma "pressão coletiva que torna falsa a opção de estar fora de um determinado sistema" (Kanashiro, 2016:23). De acordo com esses exemplos, é possível inferir o quanto a atuação do dispositivo de vigilância se dá pelo monitoramento sistemático, automatizado e à distância de ações e informações de indivíduos por meio da coleta digital de dados. Ainda que, tal qual as técnicas disciplinares, a finalidade ainda seja conhecer para intervir nas condutas, o que se dá mediante mecanismos de monitoramento e rastreamento de ações, informações e comunicações. Especialmente, com a elaboração de bancos de dados, a partir dos quais se estabelecem perfis computacionais por meio de uma prática conhecida como *profiling* (Bruno, 2013). Os rastros digitais podem ser organizados de forma infraindividual ordenados segundo o modelo *top-down*, que utiliza parâmetros como idade, gênero e profissão de modo preestabelecido ou segundo o *bottom-up*, que gera classes como:

> Frequentadores do site Y que clicam nos links tipo X". Essa categoria é submetida [...] à mineração de dados, técnica estatística aplicada que consiste num mecanismo automatizado de processamento de grandes volumes de dados cuja função central é a extração de padrões que geram conhecimento, é um procedimento conhecido como "descoberta de conhecimento em bases de dados" (Bruno, 2013:158).

Esse conhecimento segue processos indutivos que se baseiam em algoritmos programados para extrair padrões e regras de correlação entre elementos. Os mecanismos mais comuns são os de tipo associativo, como o *profiling*, que, a partir de similaridade, vizinhança ou afinidade, associam no mínimo dois elementos e depois diferenciam tipos de indivíduos ou grupos. Desse modo, características e padrões podem ser relacionados a certos tipos de comportamento (Bruno, 2013).

Trata-se de uma nova racionalidade estatística que cria força a partir do tratamento automatizado de informações com aspecto massivo, tal qual o big data,mas que não busca causas para os fenômenos. De fato, "ancora-se na observação puramente estatística das correlações (independente de toda lógica) entre dados coletados de uma maneira absolutamente não seletiva numa variedade de contextos heterogêneos" (Bruno, 2013:159). Os perfis são projeções algorítmicas, são menos sobre indivíduos identificáveis e mais sobre "ações, condutas, escolhas de modo que podem ser suscitadas, desviadas, orientadas e conjuradas. Esse conhecimento é mais da ordem futuro, das regras de similaridade e da exterioridade" (Bruno, 2013:163).

Todavia, ainda que os dados sejam coletados de modo massivo e aleatório, Tufekci (2017:1) nos lembra que os algoritmos podem realizar inferências aleatórias e facilmente distinguir "etnia, posição religiosa e política, traços de personalidade, inteligência, felicidade, uso de substâncias viciantes, separação dos pais, idade e gênero, só a partir das curtidas no Facebook". E mais:

> Além de serem capazes de identificar manifestantes mesmo que seus rostos estejam parcialmente ocultos. Esses algoritmos podem detectar a orientação sexual das pessoas só pelas fotos de perfil de seus relacionamentos. Essas são inferências probabilísticas, então não estarão 100% corretas, mas não vejo os poderosos resistindo à tentação de usar essas tecnologias só porque há alguns falsos positivos. [...] E a tragédia é a seguinte: **estamos construindo uma infraestrutura de vigilância autoritária só para que as pessoas cliquem em anúncios**. E esse não será o autoritarismo do Orwell. Essas estruturas estão organizando o modo como funcionamos e estão controlando o que podemos ou não fazer. E muitas dessas plataformas financiadas por anúncios se vangloriam de serem gratuitas. Nesse contexto, o produto que está sendo vendido somos nós (Tufekci, 2017:1).

Essa infraestrutura de vigilância tem como base uma taxonomia dos bancos de dados, que pode ser interpretada como uma máquina epistêmica e individualizante. O perfil é uma categoria correspondente à probabilidade de manifestação de um fator – comportamento, interesse, traço psicológico – de acordo com um quadro de variáveis. O intuito é a categorização de conduta para simular comportamentos futuros, de modo que não se aplica à divisão norma/desvio, pois as regularidades expressam tendências e potencialidades e não refletem uma natureza ou uma lei. O que é inadequado não é corrigido, mas sim incorporado aos cálculos futuros de definição de

perfil (Bruno, 2013). O desvio aparece não como erro, mas como possibilidade de um acerto cada vez mais preciso no futuro.

> Nesse cenário, a privacidade é um fator que se encontra em momento de disputa, no qual os discursos, forças e práticas estão num embate por determinar o sentido, o valor e a experiência da privacidade. Os dados publicados de forma voluntária geram uma segunda camada de dados. Por meio de bancos de dados e *profiling*, é possível gerar mapas e perfis de consumo, interesse, comportamento, sociabilidade, preferências políticas que podem ser usados para marketing, administração pública, indústria do entretenimento, indústria da segurança etc. O que fica evidente é que o controle do indivíduo sobre os próprios dados é muito restrito e a noção jurídica de privacidade não dá conta da complexidade das questões sociais, políticas e cognitivas envolvidas (Bruno, 2013:129-130).

Esse é um processo de vigilância que não se concentra fortemente na identificação dos indivíduos, o rastreamento de dados opera em níveis menos visíveis – do rastro digital, do plano infraindividual ou supraindividual –, de modo que o interesse está mais concentrado em desenvolver uma forma de exercício do poder sem que haja prioridade de identificação identitária, que, contudo, pode acender uma discussão sobre que tipo de dado pode ser considerado como informação pessoal e violação de privacidade (Bruno, 2013). O que nos leva à discussão a respeito do *tracking* e da inserção de algoritmos rastreadores em aplicativos, de modo que um mesmo algoritmo pode ter, ao mesmo tempo, uma função de detecção de erros e de rastreamento, por exemplo.

Tracking

A vigilância digital opera em várias camadas, as informações pessoais e as publicações divulgadas voluntariamente, por exemplo, encontram-se em um nível mais superficial e explícito da coleta de dados. Níveis mais avançados também vigiam navegação, busca, cliques em links, downloads, produção ou reprodução de conteúdo, que deixam vestígios mais ou menos explícitos, suscetíveis de serem capturados (Bruno, 2013). Já a camada mais profunda fica por conta dos algoritmos rastreadores, cuja maior parte funciona derivando um código de identificação do dispositivo móvel ou navegador da web, que depois é compartilhado com terceiros para traçar o perfil do usuário com maior precisão. Nessa prática, são recolhidos dados variados sobre os usos de aplicativos, geolocalização, preferências, métricas de performance etc. (Grauer, 2017).

De acordo com Bruno (2013), em 2010, 68% dos rastreadores já atuavam no campo do marketing online e da publicidade direcionada, embora o monitoramento de rastros pessoais na Internet também despertasse o interesse de domínios variados, como “segurança, entretenimento, saúde, gestão do trabalho e recrutamento pessoal, consultoria e propaganda política,

desenvolvimento de produtos e serviços, vigilância e controle, inspeção policial e estatal, etc." (Bruno, 2013:124). Sobre a forma como as grandes corporações de internet atuam no dispositivo de vigilância, a autora disserta:

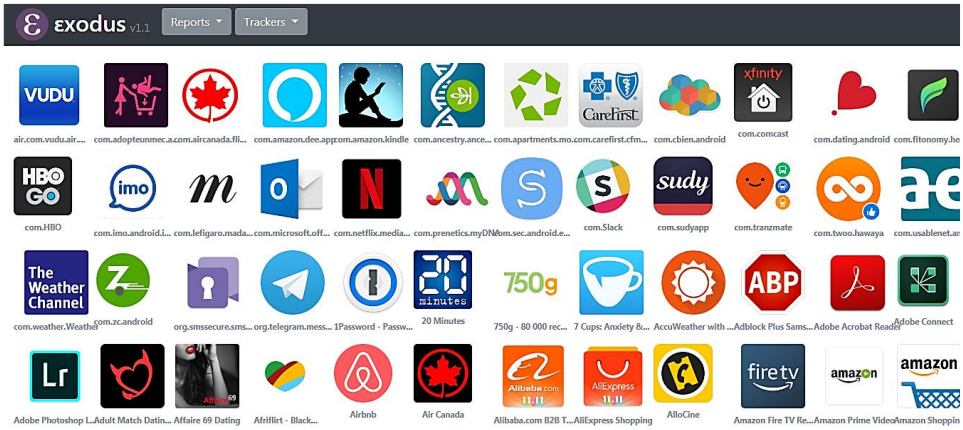
> Não existem redes sociais, por exemplo, isentas de práticas de vigilância ainda que essa não seja sua função potencial, os sistemas de monitoramento são parte integrante tanto da eficiência dessas plataformas que rastreiam, arquivam e analisam as informações disponibilizadas pelos usuários, que encontram na vigilância mútua e consentida, com pitadas de voyeurismo, um dos motores desta sociabilidade. Não existem sistemas de busca, tal qual o Google, já que com sua maquinaria estritamente informacional, os algoritmos de monitoramento das informações e ações dos indivíduos no ciberespaço são constituintes dos parâmetros de eficiência de qualquer motor de busca (Bruno, 2013:31).

A arquitetura da persuasão que o Facebook construiu é a responsável por sua capitalização de mercado perto de 500 bilhões de dólares. Todavia, é importante ressaltar que a estrutura para quem vende sapatos também é a mesma para quem vende política. O algoritmo não faz diferenciação, a lógica que serve para nos tornar mais influenciáveis a anúncios, também serve para organizar o fluxo de informações políticas, pessoais e sociais. (Tufekci, 2017). De acordo com Grauer (2017), o Google também tem interesse particular em permitir o livre uso dos rastreadores nos aplicativos distribuídos pelo Google Play[7]. Um dos rastreadores mais difundidos é desenvolvido por sua plataforma de publicidade DoubleClick, feita para direcionar os anúncios por localização e em diferentes aparelhos e canais. A plataforma também segmenta usuários com base no comportamento online, vinculando-os a informações pessoalmente identificáveis, além de oferecer compartilhamento de dados e integração com vários sistemas de publicidade. Assim, o rastreador da DoubleClick é facilmente encontrado em vários aplicativos populares.

A Exodus é uma plataforma privada de audição para aplicativos Android, que detecta comportamentos potencialmente perigosos para a privacidade do usuário, tais como publicidades, rastreadores e estatísticas. Tem sido usada em parceria com o laboratório de privacidade da Universidade Yale para detectar práticas de vigilância algorítmicas. Segundo Grauer (2017:1), as funções dos rastreadores descobertos pela Exodus incluem "segmentar usuários com base em dados de terceiros, identificar atividade offline por meio de aprendizagem de máquina, rastrear comportamento em diversos aparelhos; identificá-los e correlacioná-los; e segmentar os que abandonam carrinhos de compras". No modelo da plataforma, ao clicar no ícone do aplicativo, ele será escaneado quanto à presença de rastreadores (*trackers*), conforme Figura (2).

[7] Google Play é um serviço de distribuição digital de aplicativos, jogos, filmes, programas de televisão, músicas e livros, desenvolvido e operado pelo Google.

Figura 2 - Layout da plataforma Exodus



Com o auxílio da Exodus, é possível analisar diversos aplicativos, no que concerne à presença de *trackers*, bem como o relatório de qual seria sua função oficial e quais são suas funções extras de vigilância. Na tabela abaixo, explicitam-se os modos como os rastreadores coletam e monitoram os dados. As informações são da pesquisa realizada pelo Laboratório de privacidade da Faculdade de Direito de Yale, durante o desenvolvimento da plataforma Exodus[8].

Tabela 1 - Rastreadores encontrados em aplicativos pelos pesquisadores do Privacy Lab da Universidade de Yale mediante a plataforma Exodus Privacy

RASTREADOR (*TRACKER*)	APLICATIVO	FUNÇÕES DE VIGILÂNCIA
CRASHLYTICS (GOOGLE)	TINDER; OKCUPID; SPOTIFY UBER; SUPERBRIGHT LED E LED LIGHT (APLICATIVOS DE LANTERNA);	VINCULA USUÁRIOS VALENDO-SE DE MÚLTIPLOS COOKIES E DISPOSITIVOS.
HOCKEYAPP (MICROSOFT)	MICROSOFT OUTLOOK; SKYPE. WEATHER CHANNEL;	RASTREIA OS USUÁRIOS ATIVOS DIARIAMENTE E MENSALMENTE, O NÚMERO LÍQUIDO DE NOVOS USUÁRIOS E AS CONTAGENS DE SESSÕES.
APPFLYER	TINDER. SUPERBRIGHT LED. WEATHER CHANNEL;	IDENTIFICA DISPOSITIVOS POR SEUS IDS, RASTREIA USUÁRIOS EM DIFERENTES CONJUNTOS DE DADOS E QUAIS USUÁRIOS INSTALAM QUAIS APLICATIVOS.
BRAZE	OKCUPID; LYFT;	RASTREIA OS USUÁRIOS POR LOCAL, SEGMENTA-OS EM DIFERENTES DISPOSITIVOS E CANAIS, E VEICULA PUBLICIDADE DIRECIONADA COM BASE NAS AÇÕES DOS CONSUMIDORES
SALESFORCE DMP	OKCUPID;	PERMITE AOS VENDEDORES UTILIZAR APRENDIZAGEM DE MÁQUINA (*machine learning*) PARA REVELAR PERSONAS (*profiling*), USA UMA ID ENTRE DIFERENTES DISPOSITIVOS. ANALISA COMPORTAMENTOS PARA ADIVINHAR QUANDO O USUÁRIO ESTÁ DORMINDO. USA ALGORITMO DE COMBINAÇÃO PROBABILÍSTICA PARA CORRESPONDÊNCIA DE IDENTIDADES ENTRE DISPOSITIVOS.

[8] Disponível em https://reports.exodus-privacy.eu.org/en/. Acesso em 22 jun. 2020.

SCORECARDRESEARCH	ACCUWEATHER; WEATHER CHANNEL. SPOTIFY;	RASTREIA DADOS DE UTILIZAÇÃO, INFORMAÇÕES SOBRE NAVEGAÇÃO WEB E COMPORTAMENTO DE USO DE APLICATIVOS; ESTABELECE RELAÇÕES ENTRE NAVEGADORES E DISPOSITIVOS.
FLURRY	MICROSOFT OUTLOOK. WEATHER CHANNEL. SUPERBRIGHT LED E LED LIGHT;	RASTREIA AS MÉTRICAS DE PERFORMANCE DE DISPOSITIVOS E APLICATIVOS, ANALISA AS INTERAÇÕES DOS USUÁRIOS, IDENTIFICA INTERESSES, ARMAZENA PERFIS DE DADOS COMO PERSONAS, AGRUPA E CORRELACIONA DADOS DE USUÁRIOS E INJETA ANÚNCIOS, INCLUSIVE DE VÍDEO.
TUNE	FOCA EM USUÁRIOS QUE UTILIZAM O COMPARTILHAMENTO DE CARONAS;	SEGUE O COMPORTAMENTO ONLINE E OFFLINE DOS USUÁRIOS EM DIFERENTES DISPOSITIVOS E TAMBÉM RASTREIA O SEU COMPORTAMENTO DENTRO DOS APLICATIVOS, IDENTIFICA USUÁRIOS DE FORMA ESPECÍFICA E RASTREIA SUA LOCALIZAÇÃO.
APPNEXUS	SUPERBRIGHT LED, ENTRE OUTROS APLICATIVOS;	UTILIZA APRENDIZAGEM DE MÁQUINA PARA FAZER PUBLICIDADE DIRECIONADA.
DOUBLECLICK, TEEMO, BRAZE, SALESFORCE	TINDER; OKCUPID; LYFT, UBER, SPOTIFY; WEATHER CHANNEL E ACCUWEATER; SUPERBRIGHT LED E LED LIGHT;	COLETAM GRANDES VOLUMES DE DADOS.

Fonte: GRAUER (2017)

Segundo Grauer (2017), foram identificados 44 tipos de rastreadores em mais de 300 aplicativos para smartphones com Android[9].

> Para encontrar os scripts de rastreamento, os pesquisadores da Exodus desenvolveram uma plataforma de auditoria customizada para o ecossistema Android, que pesquisou os aplicativos em busca de "assinaturas" digitais extraídas de rastreadores já conhecidos. Uma "assinatura" pode ser um conjunto de palavras-chave sinalizadoras ou uma cadeia de bytes, encontrados num arquivo de aplicativo, ou uma representação matematicamente derivada do próprio arquivo (um "*hash*") (Grauer, 2017:1).

Os níveis de invasão dos rastreadores variam, embora todos capturem mais informação do que anunciam. Segundo Bruno (2013), o valor econômico, estratégico e heurístico dos rastros digitais está concentrado no tipo de conhecimento gerado a partir deles. A importância do big data vai além da capacidade de armazenamento de dados. Para a vigilância, seu valor está na potência de conhecimento gerada pela sua quantidade de dados. Os sistemas de classificação da vigilância digital seguem o caminho da classificação com vistas a governar condutas, lançando mão de uma taxonomia específica que é distinta dos procedimentos modernos. "Sob o gigantesco fluxo de rastros pessoais em plataformas participativas, apresenta-se processos como *dataveillance* (vigilância de dados), *data mining* (mineração de dados) e *profiling* (perfilização), que monitoram e classificam os dados, construindo saberes que sustentam uma vigilância proativa sobre indivíduos e populações" (Bruno, 2013:127).

[9] Sistema operacional móvel do Google.

Rastreadores deixam rastros: resultados da aplicação de algoritmos de contravigilância

A web atual conta com um ecossistema complexo e dinâmico de publicidade e análise para otimizar a monetização dos dados, pois, em praticamente toda página visitada, as ações são monitoradas por scripts de terceiros que coletam e agregam dados sobre as atividades e ações dos usuários, tendo em vista que, em média, 77% das páginas que o usuário mediano visita contêm rastreadores (Karaj *et al.*, 2019).

Rastreamento, do inglês *tracking*, é um mecanismo para registrar padrões de navegação de um usuário na internet. A priori, não haveria nada de errado com isso, já que, se usuários visitarem sites de viagens e depois começarem a receber propaganda sobre hotéis nos seus sites de notícias favoritos, pode-se argumentar que seria um benefício, uma economia de tempo. Esse *re-targeting*[10] descrito não implica necessariamente numa perda de privacidade, normalmente a perda de privacidade aparece como resultado da forma pela qual o *tracking* é implementado (Macbeth *et al.*, 2016:2).

Uma típica implementação é o exemplo da relação entre o site de viagens *https://www.kayak.de/* e o site de notícias *https://www.huffingtonpost.co.uk/*, que compartilham um rastreador de propriedade de Bluekai. Há um pedaço de código javascript[11] (tags.bkrtx.com/js/bk-coretag.js) que é executado no navegador dos usuários toda vez que eles visitam qualquer página de kayak.de ou hungtonpost.co.uk. Esse pedaço de javascript normalmente envia à página de origem *S* (a página que está sendo visitada) no remetente HTTP, as seguintes informações:

> bklc=**55f6ad4d** l=https://www.kayak.de/
> ua=f82610bef1d54776cde605b90b0c7949
> t=1444203542439 m=020810a3483fc8307caa483fd192bc02
> lang=07ef608d8a7e9677f0b83775f0b83775
> sr=1440x900x24 cpu=4b4e4ecaab1f1c93ab1f1c93ab1f1c93
> platform=6d44fad93929d59b3929d59b3929d59b
> plugins=d4de4a68c91685d0ff4838ce3714359a
> cn=df62ddfcfa96f717f2ee5a7d912e7102 (Macbeth *et al.*, 2016:2)

[10] O *re-targeting*, também conhecido como *remarketing*, funciona por meio de um *cookie* que se instala no navegador do usuário quando ele acessa o site de um anunciante, a partir disso é possível localizá-lo e oferecer-lhe anúncios bem segmentados ao acessar outros sites na web.

[11] Javascript é uma das linguagens de programação mais populares do desenvolvimento web, executada do lado do cliente (esta expressão significa que os scripts ou programas são executados no navegador do usuário); funciona como uma extensão do código HTML (acrônimo de *Hypertext Marking Language*, é um tipo de linguagem usada para programar e desenvolver websites).

No entanto, entre esses dados, podemos ver a sequência 55f6ad4d, que provavelmente identifica exclusivamente o usuário, agindo como um *user unique identifier* (UID)[12]. A partir dos dados, o rastreador Bluekai tem a capacidade de aprender a relação (*u; s*), ou seja, o usuário "u" visitou a página "s". Além de kayak.de e hungtonpost.co.uk, o Bluekai está em quase 4.000 outros sites. O que significa que com base em um único acesso, o Bluekai pode apreender um pedaço considerável do histórico de navegação de um determinado usuário, provavelmente sem seu conhecimento (Macbeth *et al.*, 2016:2).

O rastreamento é a coleta de pontos de dados em várias páginas e sites diferentes, que podem ser vinculados a usuários individuais por meio de um identificador de usuário exclusivo. A geração desses identificadores pode ser *stateful*[13] (protocolo com estado), na qual o navegador cliente salva um identificador localmente que pode ser recuperado posteriormente, ou *stateless*[14] (protocolo sem estado), em que as informações sobre o navegador e/ou rede são usadas para criar uma impressão digital única (Karaj *et al.*, 2019:3). O rastreamento *stateful* usa mecanismos nas APIs[15] de protocolo e navegador para que o navegador salve um identificador da escolha do servidor de rastreamento, que pode ser recuperado e que sabe quando uma solicitação subsequente é feita ao mesmo rastreador. O método mais comum seria usar cookies, mas como esse mecanismo é implementado pelo próprio navegador, é uma decisão do lado do usuário se ele vai manter os cookies ou não. Já o rastreamento stateless combina informações sobre o sistema mediante as APIs do navegador e da rede de informação, que, quando combinadas, criam uma identificação única e persistente para o dispositivo eletrônico ou para o navegador. A diferença é que esse valor é um produto do sistema *host*[16], em vez de um estado salvo, e não pode ser excluído ou limpo pelo usuário. Em geral, o método *stateless* exigirá execução de código, seja via *JavaScript* ou *Flash*[17], que está habilitado a coletar os dados de APIs que

[12] Sistemas operacionais baseados em Unix identificam os usuários dentro do kernel por um valor inteiro sem sinal chamado de *user identifier* (em português, identificador de usuário, ou ainda, número de identificação do usuário), muitas vezes abreviado para UID ou User ID.

[13] Um firewall *stateful* é um firewall de rede que monitora o estado de funcionamento e características de conexões de rede que atravessam ele. O firewall está configurado para distinguir os pacotes legítimos para diferentes tipos de ligações. Apenas os pacotes que correspondem a uma ligação ativa conhecida são autorizados a passar o firewall.

[14] Um protocolo sem estado (do inglês *stateless)* é um protocolo de comunicação que considera cada requisição como uma transação independente que não está relacionada a qualquer requisição anterior, de forma que a comunicação consista em pares de requisição e resposta independentes. Um protocolo sem estado não requer que o servidor retenha informação ou estado de sessão sobre cada parceiro de comunicação para a duração de várias requisições.

[15] Uma interface de programação de aplicativos (API) é uma interface de computação que define interações entre vários intermediários de software. Ela define os tipos de chamadas ou solicitações que podem ser feitas, como fazê-las, os formatos de dados que devem ser usados, as convenções a seguir, etc.

[16] Um *host* de rede é um computador ou outro dispositivo conectado a uma rede de computadores. Um host pode funcionar como um servidor oferecendo recursos de informação, serviços e aplicativos para usuários ou outros hosts na rede. Os *hosts* recebem pelo menos um endereço de rede.

[17] O Adobe Flash é uma plataforma de software multimídia que costumava ser usada para a produção de animações, aplicativos web complexos, aplicativos de desktop, aplicativos móveis, jogos móveis e players de vídeo de

fornecem atributos do aparelho, como a resolução do dispositivo, o tamanho da janela do navegador, as fontes e plugins instalados etc. (Karaj *et al.*, 2019).

Mas por que os proprietários de sites concordariam em colocar esse código em seus sites? Segundo Macbeth *et al.* (2016), a web evoluiu para se tornar um *software* como um serviço *mash-up*[18], no qual os proprietários de sites tendem a terceirizar certas funcionalidades. O efeito colateral disso é que fontes de terceiros (*third-party services*[19]) passam a rodar no navegador dos usuários e rastrear suas informações. Por exemplo, toda vez que alguém visita uma página que possui um componente do Facebook (como o botão de curtir ou a caixa de comentários), o Facebook irá receber a URL[20] via HTTP-*referrer*[21], além dos seguintes dados extras via *Cookie*[22]:

> datr=_zr8VGU5cOvsTE_CjXTxF9
> lu=TTA08XEc9ieLocEDius7A
> fr=0SoRz_o5WZz6ioQ.BV5h.WE.FYS.0. AWZSMd
> **c_user=100002835278978** (Macbeth *et al.*, 2016:2).

Nesse caso, o UID do usuário está explícito e é conhecido, já que usa o parâmetro *c_user* e, quando o usuário deslogar[23], o *c_user* será removido. Entretanto, o restante dos dados permanece, ainda que seja difícil avaliar quão sensíveis esses dados serão. A partir daí, é possível comparar dados entre vários usuários identificados, porque linhas de código do tipo ***datr****=_zr8VGU5cOvsTE_CjXTxF9* são praticamente únicas, só se repetem em uma população muito grande. O que quer dizer que a *datr* pode até não ser uma UID intencional, mas ainda assim pode ser usada como tal. Dessa forma, o elemento de dado é inseguro, independentemente de sua função, pois pode ser associado a um único usuário e, consequentemente, não deveria ser

navegadores da Web incorporados. Flash exibe texto, gráficos vetoriais e gráficos *raster* para fornecer animações, videogames e aplicativos. Ele permite o streaming de áudio e vídeo, e pode capturar entrada de mouse, teclado, microfone e câmera. A plataforma de desenvolvimento relacionada Adobe AIR continua a ser suportada. O Flash foi descontinuado em 31.12.20, apesar de sua importância histórica, tornou-se um estorvo em termos de segurança digital. Sua extrema popularidade e suas vulnerabilidades o transformaram em um vetor perfeito para ataques em grande escala.

[18] Um *mashup* é um site personalizado ou uma aplicação web que usa conteúdo de mais de uma fonte para criar um serviço.

[19] Uma "fonte de terceiros" é um fornecedor de software (ou um acessório de computador) que é independente do site. São ferramentas terceirizadas pelo site, por exemplo, os botões de curtir do facebook ou os botões de compartilhamento do twitter que aparecem em diversos sites.

[20] O *Uniform Resource Locator* (URL), é um termo técnico traduzido para a língua portuguesa como "localizador uniforme de recursos". Um URL se refere ao endereço de rede no qual se encontra algum recurso informático, como por exemplo um arquivo de computador ou um dispositivo periférico (impressora, equipamento multifuncional, unidade de rede etc.). Essa rede pode ser a Internet, uma rede corporativa (como uma intranet) etc.

[21] O campo *referer* é um campo de cabeçalho HTTP que identifica o endereço da página web (i.e. o URI ou IRI) que liga ao recurso sendo solicitado. Pela verificação do *referer*, a nova página web pode ver de onde a requisição se originou. (Em suma: saber de onde o usuário veio, qual página o mandou para determinado site).

[22] Pequenas etiquetas de software que são armazenadas nos equipamentos de acesso por meio do navegador.

[23] Ato de sair de qualquer tipo de sistema onde há o uso de usuário e senha.

enviado ao Facebook. Contudo porque Facebook e Blukai enviam elementos de dados que podem ser usados como uma UID?

> O Bluekai requer conhecimento da intenção ou interesse dos usuários para redirecionar a publicidade de forma eficaz. Para fazer isso, constrói um perfil para o histórico de navegação do usuário, usando um uid como uma chave estrangeira para agrupar dados pelo usuário. Mas a necessidade de um UID pode ser justificada devido a escolhas técnicas não o torna menos problemático no que diz respeito à privacidade (Macbeth *et al.*, 2016:3).

De acordo com Macbeth *et al.* (2016), essa singularidade pode até não ser intencional, mas sim um efeito colateral inesperado de alguma funcionalidade. O que coaduna com o funcionamento do dispositivo foucaultiano, de que há elementos não intencionais que surgem no processo e o reforçam, passando assim a serem incorporados no arranjo dos elementos do dispositivo.

Uma interessante plataforma computacional de contravigilância é a "*Who trackers me*" (Karaj *et al.*, 2019), que monitora as violações de privacidade a partir de fontes de terceiros. Os estudiosos desenvolveram um algoritmo para observar as requisições feitas pelas páginas visitadas por meio dos navegadores Cliqz e Ghostery, o que corresponde a mais de 5 milhões de usuários. Isso significa o monitoramento de requisições "da localização de rede, do provedor de internet (ISP – *Internet service provider*), do sistema operacional, do *software* de navegação, das extensões do navegador e dos *softwares* de fontes de terceiros" (Karaj *et al.*, 2019:2).

Diante das diversas formas de identificar usuários e seus aparelhos eletrônicos, a questão é saber até que ponto esses dados têm sido usados para quantificar o valor da publicidade online. Posto que não existe transparência em torno de quais fontes de terceiros estão presentes nas páginas nem o que é feito com os dados coletados por essas aplicações (KaraJ *et al.*, 2019).

Existem várias observações sobre como diferentes tipos de conteúdo são usados no contexto do rastreamento. Os conteúdos demonstrados na tabela 2 são os mais frequentemente medidos:

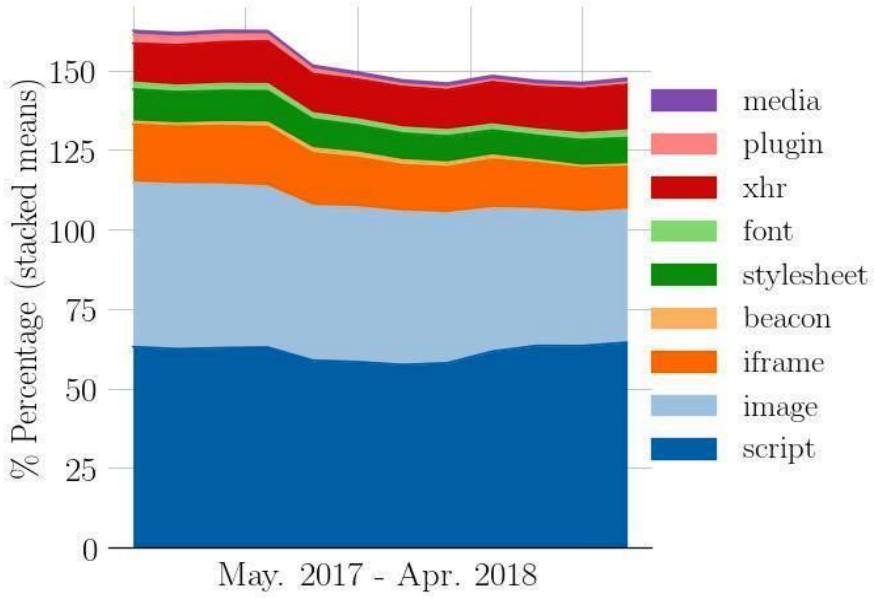
Tabela 2 - Tipos de conteúdo rastreados (Karaj *et al.*, 2019:10)

TIPO DE CONTEÚDO RASTREADO	
SCRIPT	CÓDIGO JAVASCRIPT (VIA <SCRIPT>TAG OU WEB TRABALHADOR).
IFRAME	UM SUBDOCUMENTO (VIA <FRAME>OU<IFRAME>ELEMENTOS).
BEACON	SOLICITAÇÕES MEDIANTE A API DO FAROL.
IMAGEM:	IMAGEM E IMAGENS E RECURSOS.
FOLHA DE ESTILO:	ARQUIVOS CSS.
FONTE	FONTES PERSONALIZADAS

XHR:	SOLICITAÇÕES FEITAS A PARTIR DE SCRIPTS POR MEIO DO XMLHTTPREQUEST OU BUSCA DE APIS.
PLUGIN:	SOLICITAÇÕES DE TIPOS DE OBJETO OU OBJECT_SUBREQUEST, QUE SÃO TIPICAMENTE ASSOCIADOS COM PLUGINS DO NAVEGADOR, COMO O FLASH.
MÍDIA:	PEDIDOS CARREGADOS VIA <VIDEO>OU<AUDIO> ELEMENTOS HTML

A seguir, na Figura (3), é possível ver como se dá a incidência de rastreamento dos conteúdos da tabela 1. Desse modo, é bastante perceptível que os scripts[24] e as imagens são os tipos de conteúdo mais populares para o rastreamento.

Figura 3 - Uso do tipo de conteúdo para terceiros

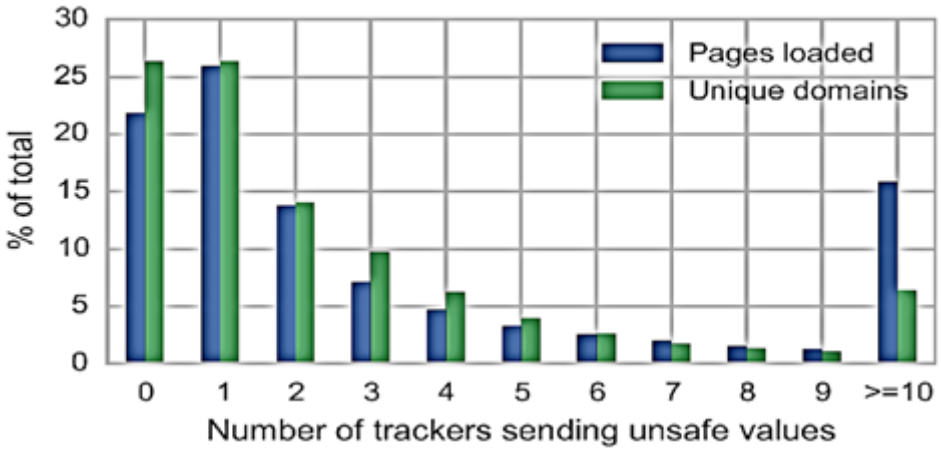


Fonte: Karaj *et al.* (2019:11).

Comumente, os rastreadores apresentam um comportamento dúbio, já que nem todas as solicitações de terceiros contêm dados sensíveis à privacidade. Na figura (4), por exemplo, as simulações do estudo de Macbeth *et al.* (2016) só consideraram os carregamentos de páginas que fizeram uma solicitação a um rastreador em potencial e que solicitaram dados identificadores.

[24] Script é uma linguagem de programação para um ambiente especial de tempo de execução que automatiza a execução de tarefas; as tarefas poderiam ser executadas um a um por um operador humano. As linguagens de script são frequentemente interpretadas, em vez de compiladas.

Figura 4 - Número de rastreadores enviando dados inseguros



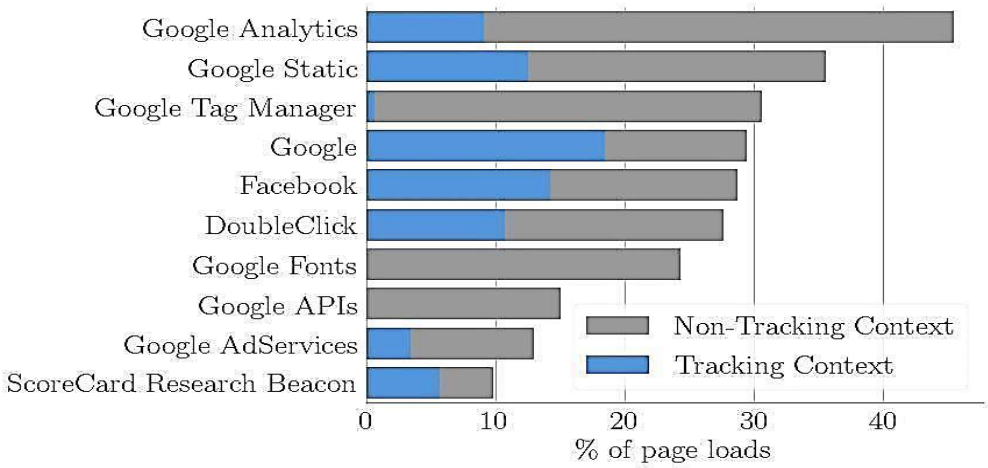
Fonte: Macbeth *et al.* (2016).

Nesse caso, Macbeth *et al.* (2016) consideraram como inseguros os dados que continham elementos capazes de identificar pessoalmente usuários. Com essa restrição, o número de carregamentos de página sem rastreamento envolvido aumenta para 22%, todavia os 78% restantes ainda estavam sujeitos a rastreamento (Macbeth *et al.*, 2016:4)

Nesse ponto, é importante atentar que os próprios estudiosos de ciência da computação não apontaram que os dados sensíveis não são apenas os que levam à identificação pessoal de indivíduos. Em todos os estudos, dos quais foram extraídos esses dados estatísticos, é possível perceber que não há um entendimento de que essa grande coleta de rastros de dados implica na identificação subjetiva de uma população inteira. O que coaduna com o argumento desta tese de que, para o dispositivo de vigilância algorítmica importa menos saber quem são cada um dos 200 mil usuários analisados por Macbeth *et al* (2016), e muito mais ter acesso à navegação de cada um deles e identificar as tendências subjetivas ali presentes.

O dispositivo de vigilância também conta com a presença das corporações de internet, há uma abrangente atuação nessa área, principalmente em associação com a publicidade de nicho. Conforme vemos na figura (5):

Figura 5 - Top 10 Fontes de terceiros, por alcance
Em azul: rastreamento; em cinza: não rastreamento

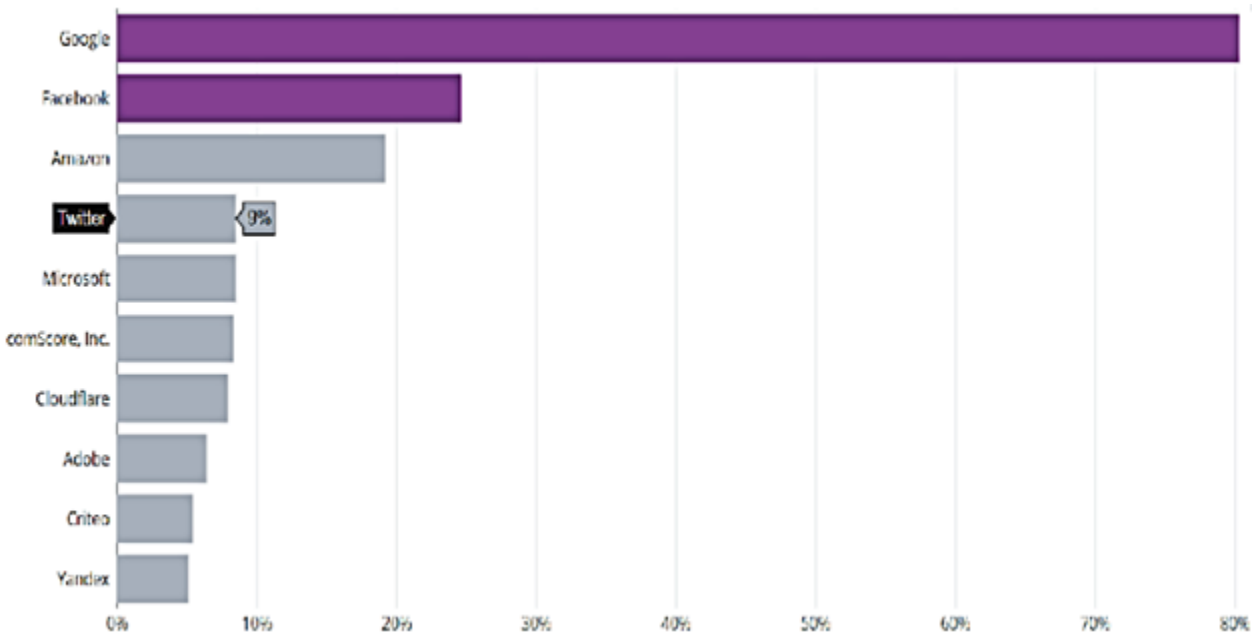


Fonte: Karaj *et al.* (2019:8).

Em abril de 2018, o estudo de Karaj *et al.* (2019) registrou 340 milhões de carregamentos de páginas, dos 1330 sites mais visitados, segundo o ranking Alexa. Conforme visto na figura acima, a medição mostra que 71% do tráfego analisado continha rastreadores, que a média de rastreadores por site é de 8 e que o número médio de requisições por rastreadores é de 17 por página. As fontes de terceiros mais presentes repetem os resultados da análise da plataforma Exodus em aplicativos, os rastreadores do Google são os mais prevalentes.

Na figura (6), fica claro o grande alcance do Google Analytics, com cerca de 46% do tráfego da Web medido e 8 das 10 fontes de terceiros principais a serem operadas pelo Google. Contudo, fontes de terceiros nem sempre operam em um contexto de rastreamento, elas nem sempre enviam identificadores exclusivos de usuários. As APIs do Google são muito usadas para carregar outras partes de terceiros, como fontes do Google e outros scripts estáticos (Karaj *et al.*, 2019:8). Os scripts de terceiros pertencentes ao Google estão presentes em cerca de 82% do tráfego da Web medido por Karaj *et al.* (2019) e operam em um contexto de rastreamento com um pouco menos da metade disso. Facebook e Amazon vêm a seguir. A figura (12), reitera a participação de mercado dos rastreadores, de acordo com a proporção de tráfego rastreado da web por essas empresas.

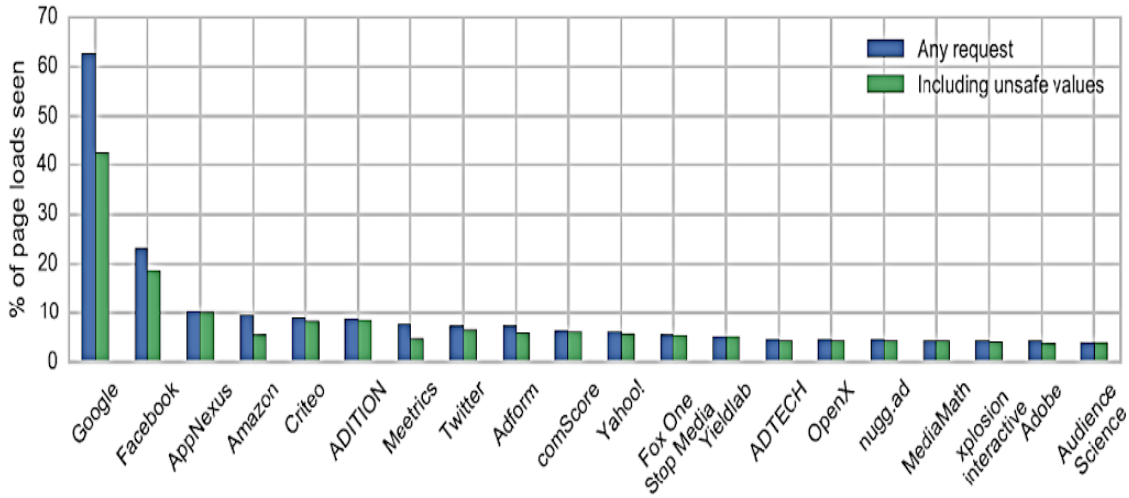
Figura 6 - Proporção do tráfego web rastreado por empresas



Fonte: Site "*Who tracks me*"
Disponível em https://whotracks.me/companies/reach-chart.html - Acesso em 28-09-2020

Já a figura (7) mostra que os resultados do estudo de Macbeth *et al.* (2016), que corroboram os encontrados tanto por esta tese por meio da plataforma Exodus quanto pelas análises de Karaj *et al.* (2019). A coluna azul explica a porcentagem de cargas de página em que uma solicitação ao rastreador potencial é emitida pelo navegador do usuário, a coluna verde é quando a solicitação também contém dados inseguros.

Figura 7 - Top 20 organizações pelo alcance combinado do rastreador.
A propriedade de um rastreador é baseada na lista de bloqueio do *Disconnect*[25]



Fonte: Macbeth *et al.* (2016:5).

[25] Disconnect é um serviço de bloqueio de rastreadores. Disponível em https://disconnect.me/. Acesso em 17-09-2020.

Todavia, nem todos os rastreadores em potencial pertencem à mesma empresa, portanto, seria incorreto dizer que uma única organização consegue rastrear 78% da navegação dos usuários, acima foram listadas as 20 empresas que mais aparecem. Na figura 7, os dados de Macbeth *et al.* (2016:4) mostram que, em 2016, os rastreadores pertencentes ao Google estavam presentes em 62% das páginas carregadas. Já na análise do trânsito de dados considerados sensíveis e/ou inseguros (em verde), o alcance do Google representava 42%; seguido pelo Facebook com, aproximadamente, 18% dos dados sensíveis rastreados.

A tabela (3) mostra os 10 sites mais acessados no Brasil, segundo o ranking Alexa. A lista também é encabeçada pelo Google.com, note que os 3 primeiros lugares pertencem a sites dessa empresa. Na lista aparecem 4 sites brasileiros: metropoles.com; uol.com.br; globo.com e mercadolivre.com. São três noticiários e um e-commerce, respectivamente.

Tabela 3 - Os 10 sites mais acessados no Brasil

SITE		TEMPO MÉDIO NO SITE[26]	*PAGEVIEWS* DIÁRIOS POR VISITANTE[27]	% DE TRÁFEGO DE PESQUISA[28]
1º	GOOGLE.COM	15:02	16.46	0.40%
2º	YOUTUBE.COM	16:16	8.79	15.00%
3º	GOOGLE.COM.BR	5:07	6.63	7.40%
4º	**METROPOLES.COM**	3:46	1.90	33.80%
5º	**UOL.COM.BR**	6:26	3.23	26.90%
6º	LIVE.COM	5:16	5.38	10.30%
7º	**GLOBO.COM**	6:22	3.07	20.30%
8º	YAHOO.COM	4:47	4.53	7.90%
9º	**MERCADOLIVRE.COM.BR**	11:24	10.30	28.60%
10º	NETFLIX.COM	4:12	3.17	9.30%

Disponível em https://www.alexa.com/topsites/countries/BR - Acesso em 18-09-2020

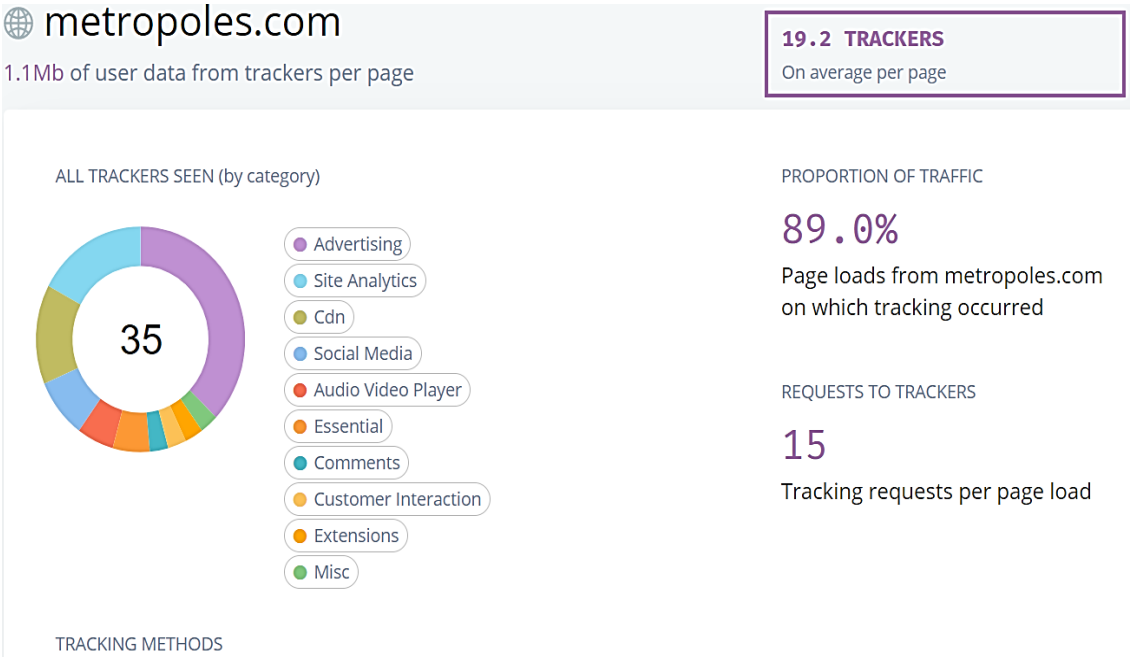
Segundo as análises de Karaj *et al.* (2019), o número de terceiros é mais alto nos sites de notícias, porque a quantidade de rastreadores por página geralmente acompanha as redes de publicidade e a cadeia de suprimentos da *Adtech*[29], que permite que várias partes executem

[26] Tempo diário estimado no local (mm:ss) por visitante ao site.
[27] Estimativa diária de visualizações únicas por visitante no site.
[28] A porcentagem de todos os encaminhamentos que vieram dos mecanismos de busca ao longo do mês. Quanto menor, mais popular é o website.
[29] A Tecnologia de Publicidade (*Advertising Technology* - *Adtech*) é definida como uma gama de softwares e ferramentas que marcas e agências usam para definir estratégias, configurar e gerenciar suas atividades de publicidade digital.

scripts em uma mesma página. Ao utilizar nos 4 sites mais acessados do Brasil o algoritmo "*Who tracks me*", disponibilizado por Karaj *et al.* (2019), é possível perceber a grande prevalência de rastreadores para a publicidade.

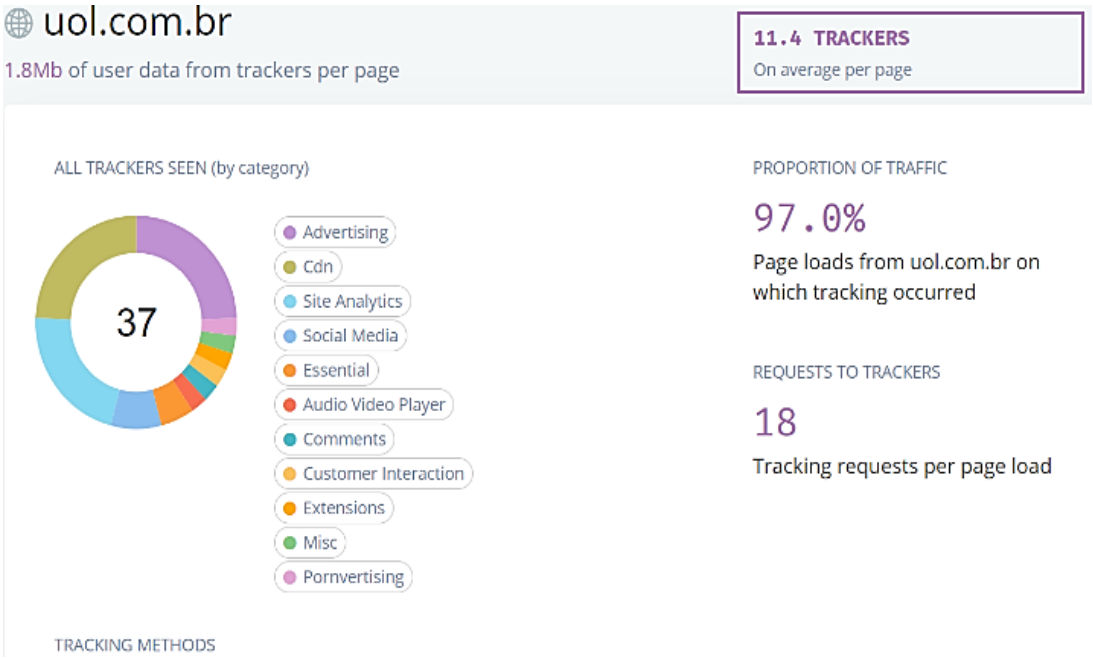
No caso do metrópoles.com, o site não só é a página brasileira mais acessada do país, como também é o campeão em rastreadores entre os quatro citados. A média é de 19.2 rastreadores por página, 89% das páginas carregadas no site possuem rastreamento e, a cada carregamento, são detectadas 15 requisições de informação feitas a partir de rastreadores; além de uma média de 1,1 Mb (megabyte) de dados de usuários rastreados por página. No gráfico, em roxo, é identificada uma maior prevalência de métodos de rastreamento para a publicidade sendo também contabilizado a presença de 35 tipos diferentes de algoritmos rastreadores.

Figura 8 - Presença de trackers no site brasileiro mais acessado em 18 set. 2020, "www.metropoles.com"



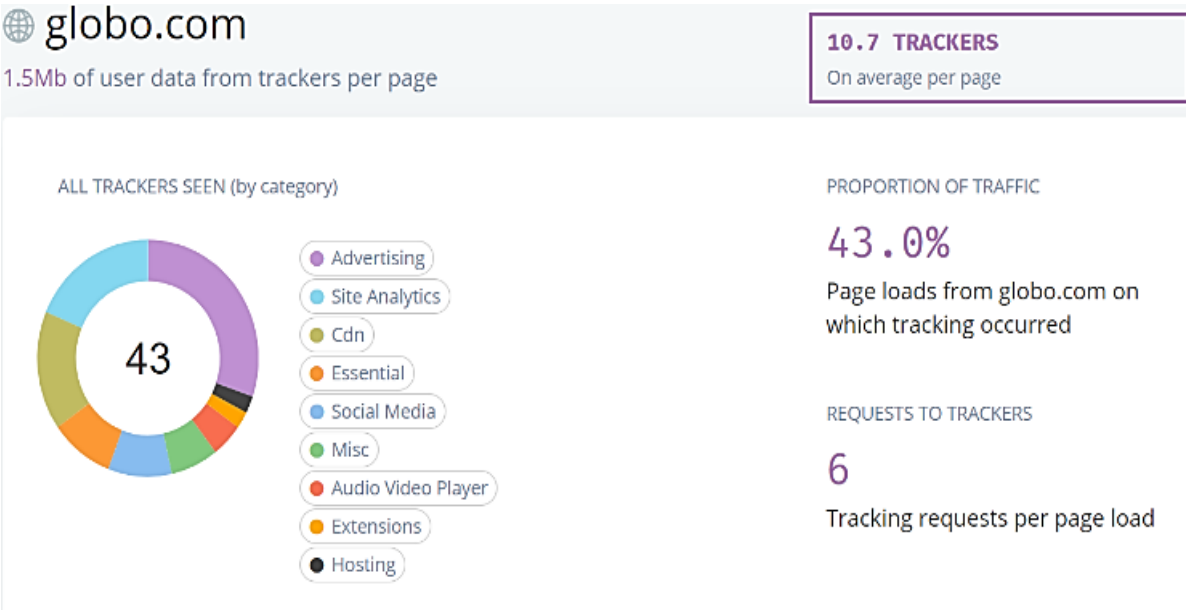
O 2º site em acessos, uol.com.br, é o site com a maior porcentagem de rastreamento, visto que 97% das páginas carregadas são rastreadas. Com um total de 37 tipos de rastreador e uma média de 11,4 rastreadores por página; é o site que apresenta a maior quantidade de dados coletada por usuário: 1,8 megabytes.

Figura 9 - Presença de trackers no 2º site brasileiro mais acessado em 18 set. 2020, "www.uol.com.br"



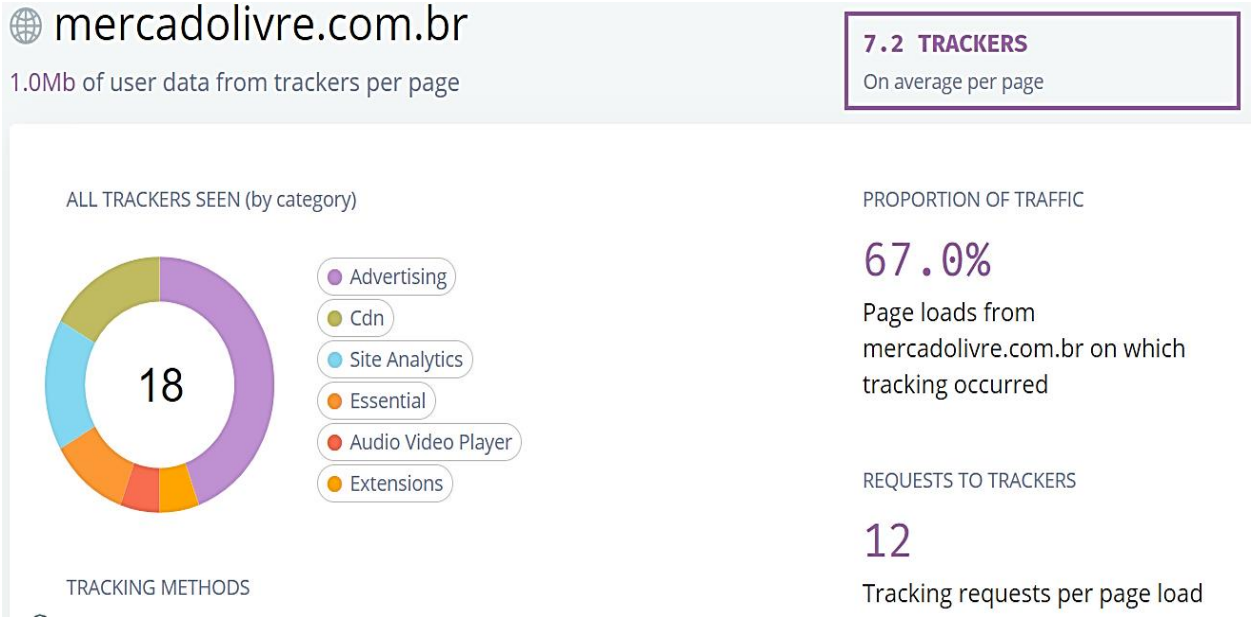
A globo.com apresenta números menores em termos de requisições de rastreadores, um total de 6 por página carregada; além de uma média de 10.7 rastreadores por página e uma proporção de 43% de carregamentos onde ocorrem rastreamentos. No entanto, apesar de apresentar índices menores que os demais, é o site com a maior prevalência de rastreadores, foram contabilizados 43, além de 1,5 Mb de dados coletados por usuário.

Figura 10 - Presença de trackers no 3º site brasileiro mais acessado em 18 set. 2020, "www.globo.com"



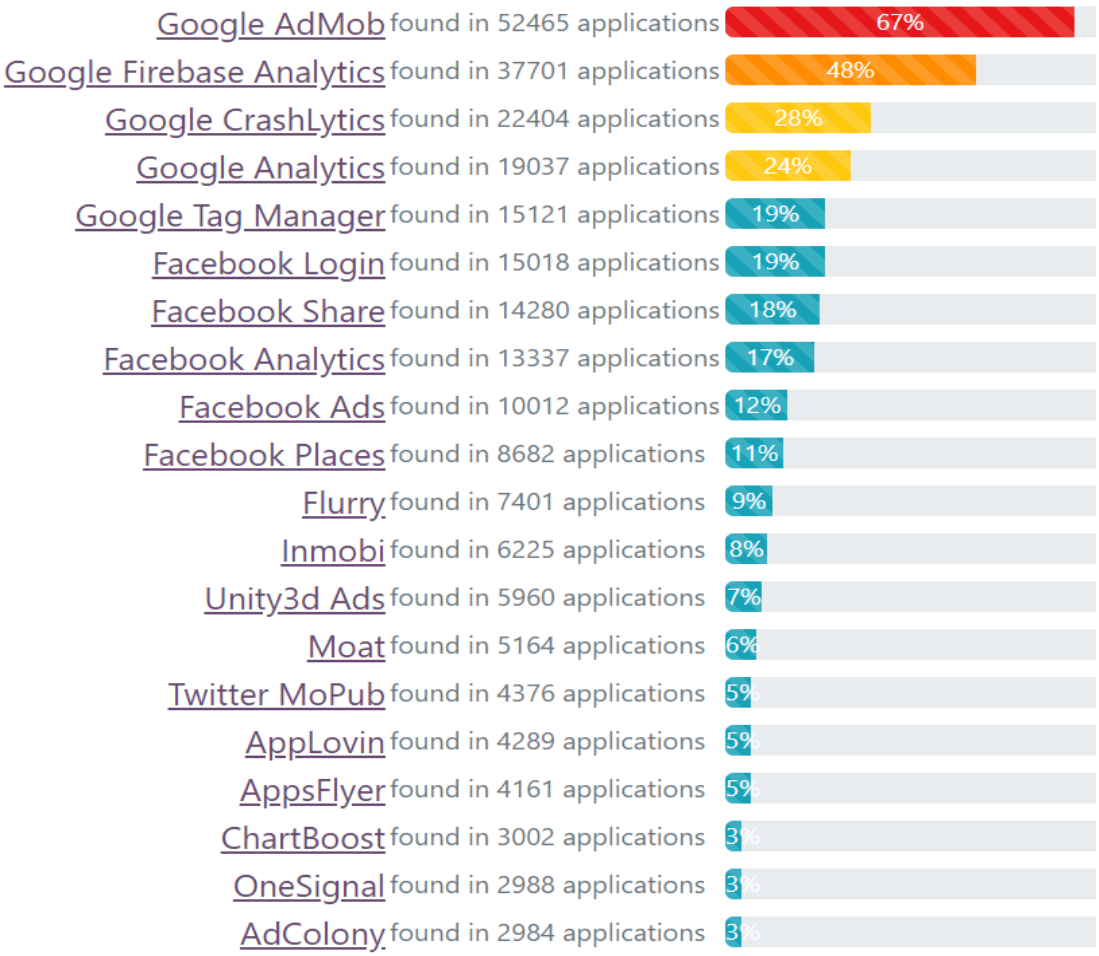
Já o mercadolivre.com é, entre os quatro, o que possui a menor taxa de presença de rastreador por carregamento de página, com um total de 7.2 rastreadores, com uma proporção de tráfego de 67% dos carregamentos de página apresentando rastreamento. O número de requisições feitas por rastreadores é 12 cada vez que uma página é carregada, a quantidade de dados capturada é de 1Mb. Entre os sites, o mercado livre é o único que não é jornalístico.

Figura 11 - Presença de trackers no 4º site brasileiro mais acessado em 18 set. 2020, "www.mercadolivre.com"



Como pudemos constatar, os resultados de Macbeth *et al.* (2016) se repetem ao analisarmos o alcance dos rastreadores nos 4 sites mais acessados do Brasil. O que também acontece no que concerne aos aplicativos para celulares, de modo que Google e Facebook continuam sendo, em 2020, os principais operadores de rastreadores também em termos de acesso móvel. A figura (12) mostra que dos 304 rastreadores reconhecidos pela plataforma Exodus, o Google Admob aparece como o mais recorrente, sendo encontrado em 52.465 aplicativos para o Sistema Android. O que nos incita a investigar melhor a forma como os telefones móveis podem ser rastreados.

Figura 12 - Rastreadores mais frequentes nos aplicativos disponíveis no Google Play



Fonte: Simulação nossa pela plataforma Exodus.
Disponível em https://reports.exodus-privacy.eu.org/en/trackers/stats/ - Acesso em 17-09- 2020

Os sensores de aparelhos móveis e suas impressões digitais

De acordo com Bojinov *et al.* (2014), os navegadores para computadores de mesa (*desktop*) têm características suficientemente diferentes para serem identificados, no entanto, os navegadores móveis, especialmente os do iOS da Apple, são muito semelhantes entre si. Com o aumento consistente do acesso à rede a partir de smartphones, surgiu a necessidade de invenção de identificadores que fossem mais competentes em rastrear dispositivos móveis. Em geral, os métodos padrões de identificação de dispositivos móveis se baseiam na coleta dos números de série presentes no Android e no iOS. Contudo, além de empresas, como a Apple, terem passado

a desabilitar aplicativos que lessem o UDID[30] de seus dispositivos, esses métodos de rastreamento também estavam sujeitos ao apagamento dos dados caso o usuário restaurasse o aparelho para o padrão de fábrica.

O experimento de Bojinov *et al.* (2014) quis mostrar como múltiplos sensores de smartphones passaram a ser usados para rastrear impressões digitais, independentemente do estado do *software* (*Stateful e Stateless*) ou dos números de série. Essa forma bastante sofisticada de identificação dos aparelhos móveis foi comprovada por rastreadores de dois tipos de sensores: o sistema viva-voz/microfone e o acelerômetro.

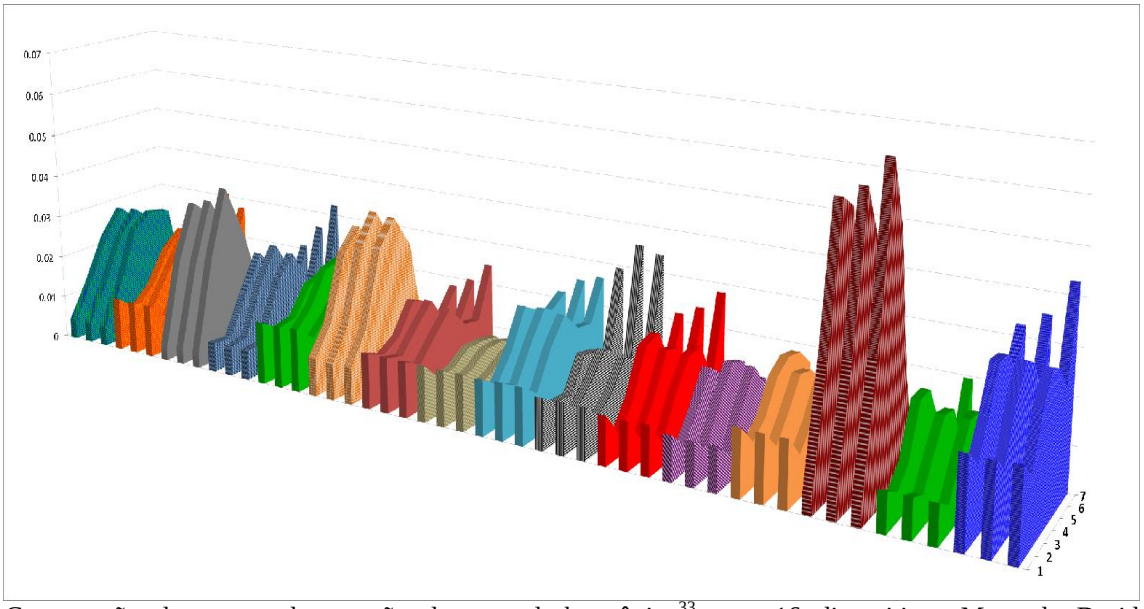
> *The speakerphone-microphone system*: the fingerprinting system uses the speakers to emit a sequence of sounds at different frequencies and records the resulting signals using the microphone. The fingerprint is computed by looking at amplitude and frequency distortions in the recorded signals.
> *The accelerometer*: the accelerometer measures force in each of the three dimensions. Imprecisions in accelerometer calibration result in a device specific scaling and translation of the measured values. By repeatedly querying the accelerometer we estimate these calibration errors by solving an optimization problem and using the resulting six values (two for each dimension) as a fingerprint (Bojinov *et al.*, 2014:1-2[31]).

O mais interessante do estudo de Bojinov *et al.* (2014) é o fato de erros nos sensores terem sido usados como rastreadores. As variações são o ponto de partida para identificação de impressões digitais deixadas pelas frequências de áudio de cada um dos aparelhos, de modo que possam ser identificados. No experimento, os autores registraram três frequências de cada aparelho em diferentes locais (uma mesa de madeira, um armário de arquivo de metal e um parapeito de madeira composta). Depois um algoritmo simples foi projetado para aprender a impressão digital de uma das superfícies e compará-la com o conjunto de dados extraídos das outras duas, em caso de compatibilidade com o aparelho que originou a frequência, a detecção era considerada correta (Bojinov *et al.*, 2014:5).

[30] UDID é uma sigla para "identificador do dispositivo único", uma combinação alfanumérica única de 40 dígitos específica para um gadget Apple, seja ele um iPhone, iPad ou iPod *touch*. Ele é como um número de série, só que muito mais difícil de adivinhar.

[31] **O sistema viva-voz/microfone**: o sistema de impressão digital usa os alto-falantes para emitir uma sequência de sons em diferentes frequências e grava os sinais resultantes usando o microfone. A impressão digital é calculada observando as distorções de amplitude e frequência nos sinais registrados. **O acelerômetro**: o acelerômetro mede as forças em cada uma das três dimensões. Imprecisões na calibração do acelerômetro resultam em um dimensionamento específico do dispositivo e na tradução dos valores medidos. Consultando repetidamente o acelerômetro, estimamos esses erros de calibração resolvendo um problema de otimização e usando os seis valores resultantes (dois para cada dimensão) como uma impressão digital (Bojinov *et al.*, 2014:1-2, tradução nossa).

Figura 13 - Comparação de curvas de geração de segundo-harmônico[32]



Comparação de curvas de geração de segundo-harmônico[33] para 16 dispositivos Motorola Droid idênticos, para avaliar a viabilidade de um esquema de impressão digital baseado na análise de feedback de som. Cada dispositivo é representado por três curvas adjacentes que têm a mesma cor e padrão de preenchimento (Bojinov *et al.*, 2014:5, tradução nossa)

Na figura (13), é possível perceber que cada aparelho apresentou uma frequência e gerou um padrão de forma diferente. A partir disso foi possível estabelecer uma impressão digital para cada um dos 16 Android testados, de modo que pudessem ser sempre identificados pelo aplicativo que dispusesse desses dados. Isso porque as variações nas frequências são codependentes do design do equipamento de áudio, os autores explicam:

> A microphone's frequency response is its normalized output gain over a given frequency range. Conversely, a loudspeaker's frequency response is its normalized output audio intensity over a given frequency range. Ideally for both devices, the frequency response should be the same for all frequencies in the range. However, a typical microphone or loudspeaker has a response curve that varies across different frequencies. These variations are dependent on the design of the audio device (Bojinov *et al.*, 2014:3)[34].

[32] A geração de segundo harmônico (também chamada duplicação de frequência ou abreviadamente na literatura SHG, do inglês, *second-harmonic generation*) é um processo óptico não linear, no qual fótons com a mesma frequência interagem com um material não-linear são efetivamente "combinados" para gerar novos fótons com o dobro da energia, e, portanto, o dobro da frequência e metade do comprimento de onda dos fótons iniciais. A geração de segundo harmônico, como um (óptica)|efeito óptico de ordem par não linear, só é permitido em meios sem simetria de inversão. É um caso especial de geração de frequência soma e é o inverso da geração de meio-harmônico.

[33] A geração de segundo harmônico (também chamada duplicação de frequência ou abreviadamente na literatura SHG, do inglês, *second-harmonic generation*) é um processo óptico não linear, no qual fótons com a mesma frequência interagem com um material não-linear são efetivamente "combinados" para gerar novos fótons com o dobro da energia, e, portanto, o dobro da frequência e metade do comprimento de onda dos fótons iniciais. A geração de segundo harmônico, como um (óptica)|efeito óptico de ordem par não linear, só é permitido em meios sem simetria de inversão. É um caso especial de geração de frequência soma e é o inverso da geração de meio-harmônico.

[34] A resposta de frequência de um microfone é seu ganho de saída normalizado sobre uma determinada faixa de frequência. Por outro lado, a resposta de frequência de um alto-falante é sua intensidade de áudio de saída

Já a tabela 4, a seguir, mostra o sucesso do experimento em identificar aparelhos móveis com base nas frequências de áudio emitidas por um algoritmo programado com essa finalidade. Quando a primeira superfície (mesa de madeira - *test location* 1) foi comparada com a segunda (arquivo de metal - *test location* 2) e com a terceira (parapeito de madeira - *test location* 3), as performances de segundo-harmônico, representados pela métrica de distância B, foram as que obtiveram maior sucesso de rastreamento do aparelho, apresentando 100% e 75% de correspondência, respectivamente para as superfícies 2 e 3. Esses experimentos mostram claramente a capacidade que algoritmos presentes em aplicativos possuem de identificar aparelhos smartphones e usar esses dados para atividades de rastreamento.

Tabela 4 - Teste de localização de aparelhos móveis baseado na coleta de frequência do sistema viva-voz/microfone

Test Location[1]	Distance Metric			
	A	B	B'	B"
2	68.8%	100%	62.5%	56.3%
3	43.8%	75%	50%	37.5%

Já no acelerômetro[35] foram medidas as imprecisões na calibração que resultam em um dimensionamento específico do dispositivo e tradução dos valores medidos. Ao consultar repetidamente o acelerômetro, foram estimados quais seriam esses erros de calibração, o que resultou em seis valores resultantes como uma impressão digital. Segundo Bojinov *et al.* (2014: 6, tradução nossa), "o acelerômetro é conveniente para a impressão digital porque o usuário muitas vezes deixa o smartphone parado. Quando o dispositivo não está se movendo, a magnitude do vetor de aceleração no dispositivo é igual a gravidade (*g*)[36]". Os autores detalham o método:

normalizada sobre uma determinada faixa de frequência. Idealmente, para ambos, a resposta de frequência deve ser a mesma para todas as frequências da faixa. No entanto, um microfone ou alto-falante típico tem uma curva de resposta que varia entre diferentes frequências. Essas variações dependem do design do dispositivo de áudio. (Bojinov *et al.*, 2014:3, tradução nossa).

[35] Acelerômetro é um dispositivo que mede a vibração ou a aceleração do movimento de uma estrutura. A força causada por uma vibração ou alteração do movimento (aceleração) faz com que a massa "esprema" o material piezoelétrico, produzindo uma carga elétrica proporcional à força exercida sobre ele. O conceito de aceleração própria surgiu em 1915, com os trabalhos de Albert Einstein sobre a Teoria da Relatividade Geral.

[36] The accelerometer is convenient to fingerprint for several fundamental reasons: the user often leaves the device still – for instance on a desk, or in a purse; as noted above, when the device is not moving the magnitude of the acceleration vector on the device equals *g*. (Bojinov *et al.*, 2014:6).

> We can use a well-known acceleration baseline to measure the accelerometer's offset – Earth's gravity (denoted by g) At rest the phone experiences an acceleration with a true magnitude of exactly g. The orientation of that acceleration depends on the relative orientation of the phone to the Earth's surface. […] Detecting the phone is at rest is relatively straightforward: the measured acceleration vector should be static, and its magnitude should roughly be equal to *g* (Bojinov *et al.*, 2014:6)[37].

Finalmente, a aceleração pode ser medida por um aplicativo Android que não requer permissões, além disso, o iOS e navegadores Android expõem essa funcionalidade a sites sem notificar o usuário. Contudo, em contraste com a impressão digital baseada em áudio, não há uma boa maneira de alimentar um sinal no acelerômetro, ou seja, exercer uma força de aceleração conhecida. Em vez disso, Bojinov *et al.* (2014) usaram a abordagem de realizar medições de fundo e esperar até que houvesse dados suficientes para estimar os parâmetros de calibração do acelerômetro.

Assim, por meio da coleta de medições dos sensores de mais de 10.000 dispositivos móveis, o estudo de Bojinov *et al.* (2014) consegue demonstrar que as impressões digitais resultantes dos erros dos sensores são bastante robustas e se mantém mesmo que todos os dados sejam apagados do aparelho com a restauração de padrão de fábrica. Isso porque as instâncias de *hardware* de um sensor específico são bastante diferentes entre si, devido a imperfeições no processo de fabricação e montagem. Bojinov *et al.* (2014) reiteram que essas variações são persistentes ao longo da vida útil do sensor, de modo que a medição das imperfeições permite identificar consistentemente quais aparelhos específicos carregam sensores com aqueles erros específicos.

Em resumo, é um clássico caso de um elemento da categoria de *Funcionamento* especificado pelo dispositivo foucaultiano. Mais uma vez, encontramos um elemento surgente não previsto, mas que gera um reajuste constante no dispositivo como um todo. Aqui são os erros apresentados por cada equipamento que passam a ser integrados ao dispositivo de poder de forma bastante significativa. Afinal, a hipótese de que essa forma de rastreamento a partir dos erros dos sensores já estivesse presente no projeto original dos smartphones é bastante contraproducente. Além do sistema de viva-voz/microfone e do acelerômetro, outros sensores ainda podem apresentar "erros identificáveis", como o áudio, o giroscópio, o magnetômetro, a luz ambiente, o GPS, o *touchscreen* e a câmera.

[37] Podemos usar uma linha de base de aceleração bem conhecida para medir o deslocamento do acelerômetro – a gravidade da Terra (denotada por g) No resto, o telefone experimenta uma aceleração com uma verdadeira magnitude de exatamente g. A orientação dessa aceleração depende da orientação relativa do telefone para a superfície da Terra. [...] Detectar que o telefone está em repouso é relativamente simples: o vetor de aceleração medido deve ser estático e sua magnitude deve ser aproximadamente igual a *g*. (Bojinov *et al.*, 2014:6, tradução nossa).

Diante disso, cabe-nos perguntar como chegamos até aqui? Até ao ponto de usar erros como técnica de vigilância? A resposta mais provável é a de que o dispositivo se adapta às oportunidades de vigília que se apresentam. Se – juntamente com os erros em sensores de aparelhos –, hábitos, interesses e ações também puderem ser investigados e analisados, por que não? Afinal, é preciso conhecer para conquistar e o legado da biopolítica disciplinar segue em ação.

Referências

AGAMBEN, Giorgio (2009), "O que é um dispositivo' in *O que é contemporâneo? E outros ensaios*. Trad. Vinicius Nicastro Honesko. Chapecó, Argos.

AGOSTINHO, Larissa (2017), "Diagrama ou dispositivo? Foucault entre Deleuze e Agamben". *Cadernos de Ética e Filosofia Política*, v. 1, n. 30, pp. 6-19.

BOJINOV, Hristo; BONEH, D.; MICHALEVSKY, Y.; NAKIBLY. G. (2014), *Mobile device identification via sensor fingerprinting*. [Consult. 05-10-2020]. Nova York, Cornell University. Disponível em http://arxiv.org/abs/1408.1416

BORGMANN, Albert (1984), *Technology and the character of contemporary life*. A philosophical inquiry. Chicago/Londres, The University of Chicago Press.

BRUNO, Fernanda (2013), *Máquinas de ver, Modos de ser*: Vigilância, tecnologia e subjetividade. Porto Alegre, Sulina, 196p.

CONTE, Rosaria *et al.* (2013), "Manifesto de Ciência Social Computacional". *Mediações*, Londrina, v. 18, n. 1, pp. 20-54.

CUPANI, Alberto (2004), "A tecnologia como problema filosófico: três enfoques". *Scientiæ zudia*, São Paulo, v. 2, n. 4, pp. 493-518.

DELLEUZE, Gilles (1992), "Post-scriptum sobre as sociedades de controle", *In Conversações 1972-1990*, trad. Peter Pál Pelbart. Rio de Janeiro, Editora 34.

FOUCAULT, Michel (1996), *Microfísica do poder*. Org. Roberto Machado. Rio de Janeiro, Graal.

FOUCAULT, Michel (2008a), *Nascimento da biopolítica*. São Paulo, Martins Fontes.

FOUCAULT, Michel (2008b), *Segurança, território e população*. Curso dado no Collège de France (1977-1978). São Paulo, Martins Fontes.

GRAUER, Yael. (2017), "*Aplicativos populares para celular estão cheios de rastreadores*: *pesquisadores encontraram 44 scripts de rastreamento em 300 apps baixados por milhões de pessoas*". In The Intercept. [Consult. 17-12-2017]. Disponível em https://theintercept.com/2017/12/04/aplicativos-populares-para-celular-estao-cheios-de-rastreadores/

HARAWAY, Donna. (1992), *The Promises of Monsters: a regenerative politics for inappropriate/d others*. Cultural Studies. Nova York, Routledge.

KANASHIRO, Marta M. (2016), "Vigiar e resistir: a constituição de práticas e saberes em torno da informação". *Cienc. Cult.*, v. 68, n. 1, pp. 20-24 [Consult. 03-12-2017]. Disponível em http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252016000100010&lng=en&nrm=iso

KARAJ, Arjaldo. MACBETH, S.; BERSIN, R.; PUJOL, J. M. (2018), *Who Tracks. Me: Shedding light on the opaque world of online tracking*. Nova York, Cornell University.

LYON, David (2018), "Cultura da vigilância: envolvimento, exposição e ética na modernidade digital", in Bruno, Fernanda *et al.* (Orgs.) Tecnopolítica da vigilância: perspectivas da margem. São Paulo, Boitempo, pp. 151-179.

MACBETH, Sam; MODI, K.; PUJOL, J. M.; ZHONGHAO, Y. (2016), "Tracking the trackers", in *Proceedings of the 25th International Conference on World Wide Web, WWW 2016*, Montreal, Canada, April 11-15, 2016, pp. 121-132.

MADRIGAL, Alexis C. (2011), "How the 'Cellular' Phone Got Its Name", in *The Atlantic*. [Consult. 22-06-2020]. Disponível em https://www.theatlantic.com/technology/archive/2011/09/how-the-cellular-phone-got-its-name/245173/

TUFEKCI, Zeynep (2017), "Estamos criando uma distopia só para fazer as pessoas clicarem em anúncios", in *Official TED Conference*. [Consult. 04-12-2017]. Disponível em https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads?language=pt-br#t-842171

ZUBOFF, Shoshana (2015), "Big other: Surveillance Capitalism and the Prospects of an Information Civilization". *Journal of Information Technology*. 30 (1): 75-89.

Abstract

This article, part of doctoral research, aim to analyze the use of tracking algorithms via smartphones as a corporate/enterprise form of surveillance. The role of the smartphone would be to open the way for the consent of data delivery supported both in voluntary cooperation of users and in the use of data tracking algorithms in a practice known as tracking. The principal methodology employed is counter-surveillance algorithms, developed by researchers such as Karaj *et al.* (2019) and Macbeth *et al.* (2016). Tracking algorithms in applications and websites were researched, in addition to identifying different techniques by which data tracking takes place. Tracking, from English tracking, is a mechanism to record a user's browsing patterns on the internet. For example, in the four most accessed sites in Brazil, according to the Alexa ranking, the "Who tracks me" algorithm was used, made available by Karaj *et al.* (2019). It is possible to identify a high prevalence of trackers in all of them.

Keywords: device; trackers; smartphones; algorithms; surveillance.

Resumen

Este artículo, que forma parte de la investigación doctoral, tiene como objetivo analizar el uso de algoritmos de seguimiento a través de teléfonos inteligentes como una forma de vigilancia corporativa / empresarial. El papel del teléfono inteligente sería abrir el camino para el consentimiento de la entrega de datos apoyado tanto en la cooperación voluntaria de los usuarios como en el uso de algoritmos de seguimiento de datos en una práctica conocida como seguimiento. La principal metodología empleada es el uso de algoritmos de contravigilancia, desarrollados por investigadores como Karaj *et al*. (2019) y Macbeth *et al.* (2016). Se investigaron algoritmos de seguimiento en aplicaciones y sitios web, además de identificar diferentes técnicas mediante las cuales se realiza el seguimiento de datos. El seguimiento es un mecanismo para registrar los patrones de navegación de un usuario en Internet. Por ejemplo, en los 4 sitios más visitados en Brasil según el ranking de Alexa, se utilizó el algoritmo "Quién me rastrea", puesto a disposición por Karaj *et al.* (2019), y es posible identificar una alta prevalencia de rastreadores en todos ellos.

Palabras llave: dispositivo; rastreadores; smartphones; algoritmos; vigilancia.