

Comments from
THE FUTURE OF PRIVACY FORUM



to

FEDERAL COMMUNICATIONS COMMISSION
Washington, D.C. 20554

WC Docket No. 16-106:

*Protecting the Privacy of Customers of Broadband
and Other Telecommunications Services*

Jules Polonetsky, CEO

John Verdi, Vice President of Policy

Stacey Gray, Legal & Policy Fellow

THE FUTURE OF PRIVACY FORUM*†

1400 I St. NW Ste. 450

Washington, DC 20005

www.fpf.org

May 27, 2016

* The Future of Privacy Forum (FPF) is a Washington, DC based think tank that seeks to advance responsible data practices by promoting privacy thought leadership and building consensus among privacy advocates, industry leaders, regulators, legislators and international representatives.

† The views herein do not necessarily reflect those of our members or our Advisory Board.

Table of Contents

Executive Summary	1
I. Data Exists on a Spectrum of Identifiability	3
Pseudonymous Data.....	4
Not Readily Identifiable Data	5
Techniques for Practical De-Identification.....	6
II. Explanation of the Market	8
Multi-Site Tracking Occurs Throughout the Internet Ecosystem.....	9
Why They Track: Understanding Ad Effectiveness	12
The Challenging Prognosis for Publishers.....	13
The Democratization of Data.....	14
Cross-Device: Bridging the Disconnect between Devices, Browsers, and Apps.....	16
Geo-Location Data.....	20
(1) <i>Location Services</i>	20
(2) Cell Tower Locations	22
(3) Carrier Triangulation.....	22
(4) Wi-Fi Based Location	23
(5) Mobile Location Analytics.....	24
(6) Beacons	25
III. Benefits and Limitations of Current Industry Rules	26
Opt In Required for Sensitive Data and Precise Geo-Location	27
“Conspicuous” and “Easy to Use” Opt Outs Required for Non-Sensitive Data.....	27
The Value of Uniformity	28
The Multitude of Consumer Controls	28
IV. Call for Reasonable Standards that will Elevate Industry Norms	29
The FCC’s Proposed Rules will Exclude ISPs from the Market	30
With Appropriate Controls, the Use of “Pseudonymous” or “Not Readily Identifiable” Data Should be Permitted for Cross-Device State Management Subject to Meaningful Opt Outs and Other Safeguards.....	31
The FCC Should Establish a Multi-Stakeholder Process to Develop Privacy Rules for Sensitive ISP Consumer Data and Out of Context Uses of Such Data.....	32
Conclusion	34

Executive Summary

The Federal Communications Commission’s (hereinafter FCC, or Commission) March 31, 2016 Notice of Proposed Rulemaking (NPRM or Notice) states that responsible data practices protect important consumer interests. We wholeheartedly agree.

The NPRM states that the FCC “is empowered to protect the private information collected by telecommunications, cable, and satellite companies”¹ and that Section 222 of the Communications Act “is a sector-specific statute that includes detailed requirements that Congress requires be applied to the provision of telecommunications services, but not to the provision of other services by broadband providers nor to information providers at the edge of the network.”²

When crafting privacy protections for data used by Internet Service Providers (ISPs),³ it is important to understand that ISPs, online service providers, and others in the digital ecosystem engage in business practices that implicate consumers’ privacy interests. Many entities have access to personal data, and from a consumer’s perspective, navigating the Internet often involves interacting with a complex network of entities. For example, consumers’ experiences with online advertising typically include engagement with an intertwined group of publishers, advertisers, advertising networks, and others.

In the online advertising space, edge providers protect consumers’ privacy by complying with laws, rules, and robust self-regulatory standards, as well as heeding best practices and norms. The Federal Trade Commission (FTC) plays a key enforcement role, as do leading self-regulatory organizations. Leading edge providers employ practices that provide substantial transparency, control, and security for consumers. **The FCC has an opportunity to issue rules for ISPs that are consistent with the best practices currently used by edge providers. Such rules would protect broadband consumers, be workable for companies, provide incentives for all providers to adopt the best practices that have been identified by leading companies, and increase competition in the online advertising market.**

We urge the FCC to:

- Issue a rule that recognizes that de-identification is not a black and white binary, but that **data exists on a spectrum of identifiability**, taking particular note of the FTC’s extensive guidance regarding de-identification;

¹ Protecting the Privacy of Customers of Broadband and Other Telecommunications Services, 81 Fed. Reg. 23359 (proposed April 20, 2016) (to be codified at 47 CFR 64), at para. 7 (hereinafter, *Notice*).

² *Notice* at para. 13.

³ The Commission proposed rules that apply to Broadband Internet Access (BIAS) providers. Although BIAS providers comprise a subset of ISPs, in these comments we refer to ISPs for the sake of convenience and clarity.

- Specifically recognize that **non-aggregate data can be de-identified** in a manner that makes it not reasonably linkable to a specific individual;
- Establish a framework that treats like data alike and allows ISPs to use data that are **pseudonymous or not readily identifiable** for limited purposes subject to a meaningful and uniform Opt Out mechanism, strict retention periods, regulations for the use of appended (offline) data, and ethics oversight; and
- Establish a **multi-stakeholder process** to enable advocates, companies, technical experts and others to determine the best way to approach ISPs' use of data that are **sensitive or out of context**, taking into account degrees of identifiability.

I. Data Exists on a Spectrum of Identifiability

Perhaps the most central question facing the FCC in this rulemaking is how best to define customer proprietary information (PI), including personally identifiable information (PII). The current proposed rules define proprietary information *very* broadly⁴—excluding all but the most high-level aggregate data⁵—and then apply a single rigid framework to that information. This structure reflects a rigid binary understanding of personal information that does not align with the spectrum of intermediate stages that exist between explicitly personal and wholly anonymous information. As a result, it is simultaneously too narrow and too broad, both excluding and including data uses that should be permitted subject to reasonable controls and safeguards.

The FCC’s binary approach stands in sharp contrast to leading government and industry guidelines with respect to de-identified data. According to the Federal Trade Commission (FTC), data are not “reasonably linkable” to individual identity to the extent that a company: (1) takes reasonable measures to ensure that the data are de-identified; (2) publicly commits not to try to re-identify the data; and (3) contractually prohibits downstream recipients from trying to re-identify the data (the “Three-Part Test”).⁶

Industry guidelines correspondingly define de-identified data, with the Digital Advertising Alliance stating that data are de-identified “when an entity has taken reasonable steps to ensure that the data cannot reasonably be re-associated or connected to an individual or connected to or be associated with a particular computer or device.”⁷ Similarly, the Network Advertising Initiative distinguishes between personally identifiable information (PII), defined as “data that is used, or intended to be used, to identify a particular individual,” non-PII, defined as “data that is not linked, or reasonably linkable, to an individual, but is linked or reasonably linkable to a particular computer or device,” and de-identified data, defined as “data that is not linkable to either an individual or a device.”⁸

⁴ Notice para. 56 et seq.

⁵ Notice para.154 et seq. (proposing to allow ISPs to use, disclose, and permit access to “aggregate customer PI” if the provider: (1) determines that the aggregated customer PI is not reasonably linkable to a specific individual or device; (2) publicly commits to maintain and use the aggregate data in a non-individually identifiable fashion and to not attempt to re-identify the data; (3) contractually prohibits any entity to which it discloses or permits access to the aggregate data from attempting to re-identify the data; and (4) exercises reasonable monitoring to ensure that those contracts are not violated).

⁶ FEDERAL TRADE COMMISSION, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE (2012), at 21, <https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf>.

⁷ DIGITAL ADVERTISING ALLIANCE, SELF-REGULATORY PRINCIPLES FOR MULTI-SITE DATA (Nov 2011), at 8, *available at* <http://www.aboutads.info/resource/download/Multi-Site-Data-Principles.pdf>.

⁸ NETWORK ADVERTISING INITIATIVE, 2015 UPDATE TO THE NAI CODE OF CONDUCT (2015), at 5, *available at* https://www.networkadvertising.org/sites/default/files/NAI_Code15encr.pdf.

Thus, as a preliminary matter, we urge the FCC to recognize that de-identification is not a black and white binary, but that data exists on a **spectrum of identifiability**. In a forthcoming academic paper for US and EU audiences,⁹ we describe data on a spectrum of identifiability, from explicitly personal data, to pseudonymous data, de-identified data, and finally, to fully anonymous and aggregated data such as high-level statistical data. Data is not either “personal” or “non-personal.” Instead, it falls on a spectrum; with each step towards “very highly aggregated,” both the utility of the data and the risk of re-identification are reduced. *See* Figure 1 (“A Visual Guide to Practical De-Identification”).

Pseudonymous Data

In *A Visual Guide to Practical De-Identification*, *see* Figure 1, we describe pseudonymous data as information from which direct identifiers have been eliminated or transformed, but indirect identifiers remain intact. This can include key-coded information, including research datasets where only the curator retains the key, or unique, artificial pseudonyms in place of the original identifiers (e.g. John Doe = 5L7TLX619Z) that are not used anywhere else. When this data is not shared publicly and is protected by technical and legal, contractual safeguards, it is considered “protected pseudonymous.” As explained *infra*, Part IV, this data can often be used safely and securely to permit cross-device tracking and the ability to target advertisements without posing privacy risks.

A number of publicized de-identification attacks have led some critics to believe that de-identification is rarely feasible. It is important to realize that every one of those attacks was on information that was actually **unprotected** pseudonymous data containing **well-recognized** indirect identifiers, such as multiple location points and similar data. These data sets were all made public, and thus subject to any and all possible attacks.¹⁰

“Pseudonymous” is, admittedly, a highly contentious term in the de-identification literature. Technologists regard pseudonymization as a process for removing direct identifiers and replacing them with pseudonyms, that is, a “particular type of anonymization.” In contrast, the Article 29 Working Party stated that “pseudonymisation is not a method of anonymisation,” but rather merely reduces the linkability of a dataset to the original identity of a data subject, and is therefore merely a “useful security measure.”

⁹ Jules Polonetsky, Omer Tene, & Kelsey Finch, *Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification*, SANTA CLARA L. REV. (forthcoming 2016).

¹⁰ *See, e.g.*, Khaled El Emam, Elizabeth Jonker, Luk Arbuckle, & Bradley Malin, *A Systematic Review of Re-Identification Attacks on Health Data*, PLoS ONE 6(12) (2011); Daniel C. Barth-Jones, *The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now* (July 2012), available at SSRN: <http://ssrn.com/abstract=2076397> or <http://dx.doi.org/10.2139/ssrn.2076397>.

For data that will not be made public, the FCC should rely on de-identification standards that take reasonable safeguards and controls into account, and should rely on risk-based assessments of whether a set of data is de-identified. This mirrors the recently finalized GDPR, which split the difference and defined pseudonymization as “the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person.”¹¹ Aggregation is by no means the only tool for such de-identification process.

Not Readily Identifiable Data

Data which is considered “Not Readily Identifiable” includes unique identifiers (such as cookies, or IP addresses) that have been protected by safeguards and other controls, such as hashing and contractual restrictions. In practice, policymakers should recognize the need to use and exchange such device identifiers for various purposes, as well as their being less explicit than direct identifiers such as name and address.

Regulators should take advantage of these gradations of identifiability to impose more nuanced use restrictions, similar to self-regulatory frameworks in the U.S. The NAI Code of Conduct, for example, applies obligations for notice, choice, opt-out, and non-discrimination to datasets defined as “non-personal”—that is, neither anonymous nor obviously personally identifiable.¹² The DAA Self-Regulatory Principles also set forth protections for these kinds of identifiers, determining that “data is not considered PII under the Principles if the data is not used in an identifiable manner.”¹³ Here, collection in isolation of an IP address, for example, is not considered processing of PII, and thus does not require consent or transparency even if used for online behavioral advertising, but is considered PII subject to the full set of Principles when it is “in fact linked to an individual in its collection and use.”¹⁴

To be sure, much turns on where the borders are drawn between Explicitly Personal, Potentially Identifiable, and Not Readily Identifiable data, as well as on the safeguards and controls that apply to the various categories of data. But clearly, a more nuanced approach will provide organizations with an incentive to enhance privacy protection by pushing data down the identifiability spectrum.

¹¹ GDPR, ART. 4(3b).

¹² NETWORK ADVERTISING INITIATIVE, 2015 UPDATE TO THE NAI CODE OF CONDUCT (2015), at 3, *available at* https://www.networkadvertising.org/sites/default/files/NAI_Code15encr.pdf.

¹³ DIGITAL ADVERTISING ALLIANCE, SELF REGULATORY PRINCIPLES FOR ONLINE BEHAVIORAL ADVERTISING (July 2009), at 25, *available at* <http://www.aboutads.info/resource/download/seven-principles-07-01-09.pdf>.

¹⁴ *Id.*

Techniques for Practical De-Identification

We urge the FCC to specifically recognize that *non-aggregate* data can be de-identified in a manner that makes it not reasonably linkable to a specific individual. The FCC’s suggestion that data must be aggregated to be de-identified ignores the range of de-identification tools that are available to make it difficult or impossible to re-identify data as pertaining to a specific individual. Such measures include the following:¹⁵

- **Blurring:** reducing the precision of disclosed data to minimize the certainty of individual identification. For example, converting continuous data elements into “categorical” elements that subsume unique cases.
- **Perturbation:** making small changes to the data to prevent identification of individuals from unique or rare population groups. For example, swapping data among individual cells to introduce uncertainty
- **Suppression:** removing data, for example, from a cell or row, to prevent the identification of individuals in small groups, or those with “unique characteristics. This usually requires suppression of “non-sensitive” data.

¹⁵ Simson L. Garfinkel, NISTIR 8053, *De-Identification of Personal Information* (Oct 2015), at 2, <http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf>.

Fig. 1. A Visual Guide to Practical De-Identification

A VISUAL GUIDE TO PRACTICAL DATA DE-IDENTIFICATION

Produced by
**FUTURE OF
PRIVACY
FORUM**
FFP-DRG

In collaboration with
EY

What do scientists, regulators and lawyers mean when they talk about de-identification? How does anonymous data differ from pseudonymous or de-identified information? Data identifiability is not binary. Data lies on a spectrum with multiple shades of identifiability.

This is a primer on how to distinguish different categories of data.

DEGREES OF IDENTIFIABILITY

Information containing direct and indirect identifiers.

PSEUDONYMOUS DATA

Information from which direct identifiers have been eliminated or transformed, but indirect identifiers remain intact.

DE-IDENTIFIED DATA

Direct and known indirect identifiers have been removed or manipulated to break the linkage to real world identities.

ANONYMOUS DATA

Direct and indirect identifiers have been removed or manipulated together with mathematical and technical guarantees to prevent re-identification.

DIRECT IDENTIFIERS

Data that identifies a person without additional information or by linking to information in the public domain (e.g., name, SSN)

INDIRECT IDENTIFIERS

Data that identifies an individual indirectly, helps connect pieces of information until an individual can be singled out (e.g., DOB, gender)

SAFEGUARDS and CONTROLS

Technical, organizational and legal controls preventing employees, researchers or other third parties from re-identifying individuals

	EXPLICITLY PERSONAL	POTENTIALLY IDENTIFIABLE	NOT READILY IDENTIFIABLE	KEY CODED	PSEUDONYMOUS	PROTECTED PSEUDONYMOUS	DE-IDENTIFIED	PROTECTED DE-IDENTIFIED	ANONYMOUS	AGGREGATED ANONYMOUS
NOT RELEVANT <small>due to nature of data</small>	NOT RELEVANT	NOT RELEVANT	NOT RELEVANT	NOT RELEVANT	NOT RELEVANT	NOT RELEVANT	NOT RELEVANT	NOT RELEVANT	NOT RELEVANT	NOT RELEVANT
SELECTED EXAMPLES	Name, address, phone number, SSN, government-issued ID (e.g., John Smith, 123 Main Street, 555-555-5553)	Unique device ID, license plate, medical record number, address (e.g., 66A8-6D0-555-65003)	Same as Potentially identifiable except data are also protected by safeguard controls (addresses to legal representations)	Clinical or research datasets where only curator retains key (e.g., Jane Smith, 123 Main Street, 555-555-5553)	Unique, artificial pseudonyms replace direct identifiers (e.g., John12345678901, unique sequence not used anywhere else)	Same as Pseudonymous, except data are also protected by safeguards and controls	Data are suppressed, generalized, perturbed, swapped, etc. (e.g., OPR, female = Gender: male)	Same as De-identified, except data are also protected by safeguards and controls	For example, noise is collated to data set to hide whether an individual is present or not (differential privacy)	Very highly aggregated data (e.g., statistical data, census data, or population data that DC residents are women)

II. Explanation of the Market

In order to craft rules that will be relevant and influential for consumers, it is imperative that the FCC first understand that individualized multi-site and cross-device tracking occurs throughout the Internet ecosystem, in order to provide a wide range of services, and subject to broad enforcement authority from the Federal Trade Commission (FTC).

While the FCC lacks statutory authority to regulate the rest of the Internet ecosystem, it is nonetheless important for the FCC to appreciate the effects that its proposed rules will have on that ecosystem. In the Notice of Proposed Rulemaking, the FCC specifically requested comment on this issue, asking “what effect, if any, [the] proposed opt-in approval framework will have on marketing in the broadband ecosystem, over-the-top providers of competing services, the larger Internet ecosystem, and the digital advertising industry.”¹⁶

In this section, we describe the variety of methods by which multi-site and cross-device tracking occurs in the online advertising ecosystem for purposes of measuring ad effectiveness. These activities are subject to the authority of the FTC under its broad Section 5 authority to bring civil enforcement actions against companies engaging in unfair or deceptive practices.¹⁷ Many leading companies have been subject to consent decrees in recent years,¹⁸ even in cases where no direct tangible consumer harm was identified.¹⁹ In creating rules which govern similar or identical uses of data by ISPs, the FCC should look to the FTC’s standards and enforcement authority as an effective model of regulating online privacy.

Furthermore, increasingly it has been seen that state attorneys general are emerging as regulators in the sphere of online privacy.²⁰ As Professor Danielle Citron recently stated:

State attorneys general have played a critical role in U.S. privacy law. Much as Justice Louis Brandeis imagined states as laboratories of the law, offices of state attorneys general have been laboratories of privacy enforcement. . . . state attorneys general have been privacy pioneers, setting baseline norms that have been emulated by federal agencies. They have entrenched existing privacy and security norms and, in the process, sharpened them.²¹

¹⁶ Notice para. 132.

¹⁷ 15 U.S.C. §45(a) (“FTC Act”).

¹⁸ See, e.g., *In the Matter of Snapchat, Inc.*, File No. 132 3078, (Dec. 31, 2014); *In the Matter of Google, Inc.*, File No. 122 3237 (Dec. 5, 2014); *In the Matter of Twitter, Inc.*, File No. 092 3093, (Mar. 11, 2011).

¹⁹ See *In the Matter of Nomi Technologies, Inc.*, File No. 132 3251 (Sept. 3, 2015).

²⁰ See Danielle Keats Citron, *Privacy Enforcement Pioneers: The Role of State Attorneys General in the Development of Privacy Law*, NOTRE DAME L. REV. (forthcoming 2016).

²¹ *Id.* at 35, 65.

With localized authority and the ability to adapt quickly to emerging technologies, the attorneys general are likely to play a growing role in online privacy in coming decades.

Multi-Site Tracking Occurs Throughout the Internet Ecosystem

“When I go to Google [. . .] that is a decision that I am making. [. . .] I go to WebMD, and WebMD collects information on me. I go to Weather.com and Weather.com collects information on me. I go to Facebook and Facebook collects information on me. But only one entity connects all of that information, that I’m going to all those different sites, and can turn around and monetize it.”²²

*“[E]dge providers only have direct access to the information that customers choose to share with them **by virtue of engaging their services**; in contrast, broadband providers have direct access to potentially all customer information, including such information that is not directed at the broadband provider itself to enable use of the service. We seek comment on these expectations.”²³*

The framing of the issue in the two quotes above, if taken at face value, reflects a fundamental misunderstanding of the current online ecosystem. As a preliminary matter, if the FCC seeks to create rules that will be relevant and influential across the Internet ecosystem, providing consumers with a uniform set of privacy expectations, it must first understand that ecosystem. Currently, by working through third party ad exchanges and data brokers, it is not only possible but common for Internet actors to create profiles of consumers’ Web browsing behavior across the Internet as well as between devices.

The third party tracking industry is inter-woven and provides comprehensive detail that goes beyond the information that consumers provide to an edge provider “by virtue of engaging their services.” Using Mozilla’s *Lightbeam for Firefox*²⁴ browser extension, it can quickly be seen that after visiting only one website (WebMD.com), a typical user has connected with 29 third party sites. *See Fig. 2.*

²² *Examining the Proposed FCC Privacy Rules*, Hearing Before the S. Comm. on the Judiciary, Subcomm. on Privacy, Tech., and the Law, 114th Cong. (2016) (statement of Chairman Thomas Wheeler, Federal Communications Commission), available at <http://www.judiciary.senate.gov/meetings/examining-the-proposed-fcc-privacy-rules> (1:30:52) (last visited May 24, 2016).

²³ *Id.*

²⁴ *Lightbeam for Firefox*, MOZILLA, <https://www.mozilla.org/en-us/lightbeam/> (last visited May 24, 2016).

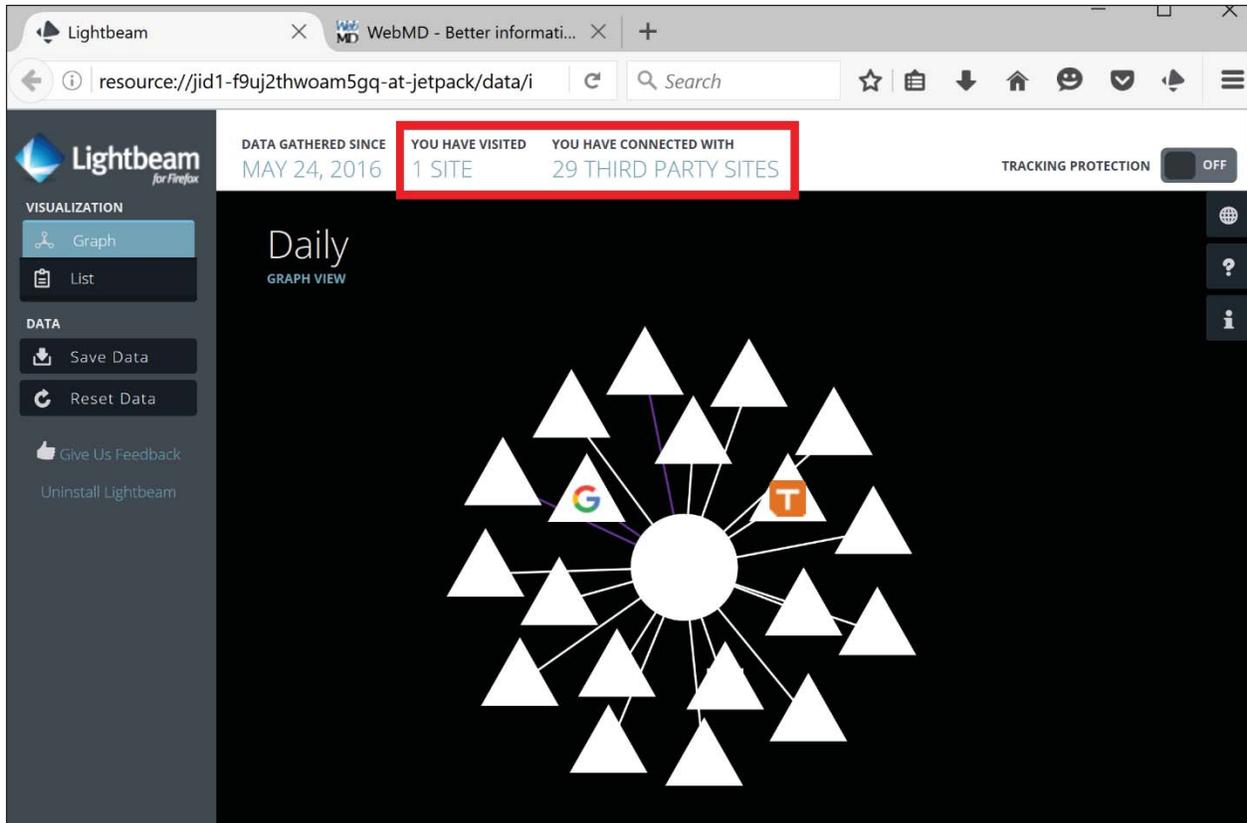


Fig. 2. Lightbeam for Firefox demonstrates that a visit to one website generated 29 third party connections. Circular nodes represent websites visited, and triangular nodes are third party sites. Purple lines identify when a site has stored data on the browser (cookies). Site accessed May 24, 2016 12:21PM Eastern Time.

After visiting additional sites—for a total of five websites—a typical user has connected with 133 third party entities. See Fig. 3. Each “connection” means that entity can identify the web page the consumer is visiting and can share that data with other parties to which they are interconnected. Some parties are linked up to many web sites, but even for those that are not directly connected to a particular site, third party entities who are linked are capable of buying and selling this data at third party data exchanges. These data exchanges, by linking and compiling data from hundreds of different online and offline sources, can match up consumer behavior across the Internet, creating comprehensive and detailed individual profiles.

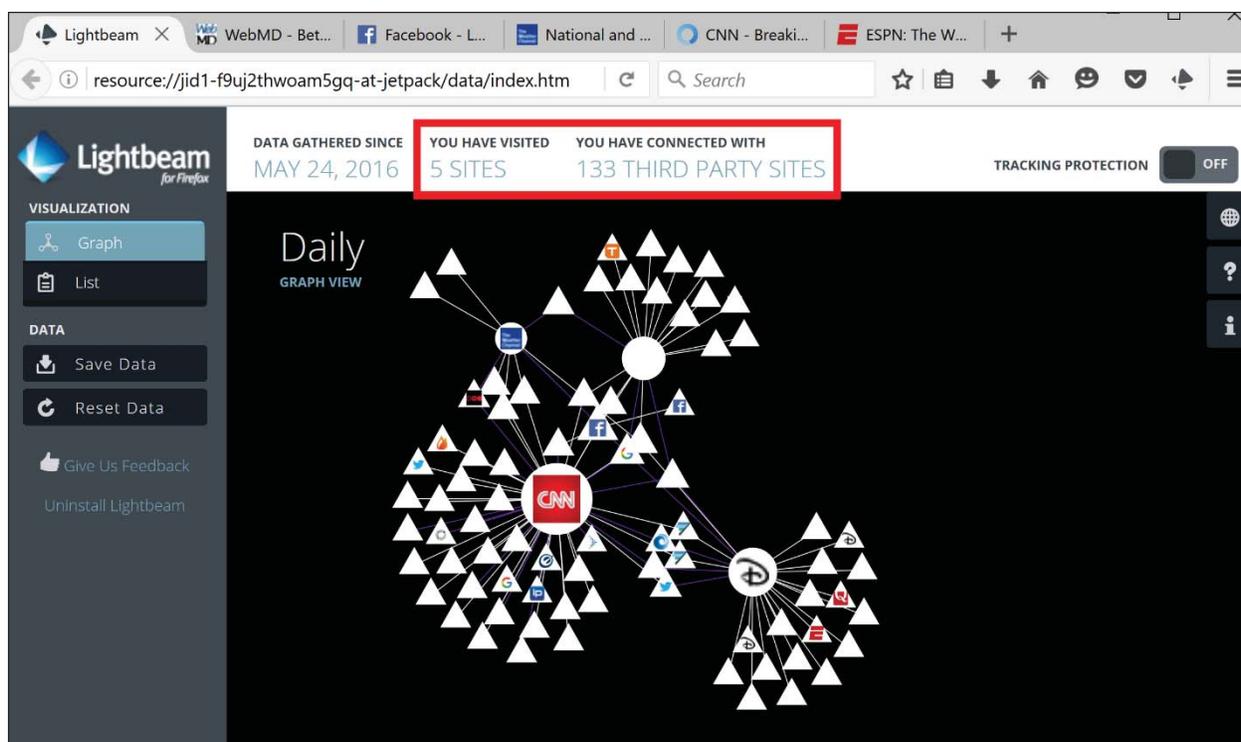


Fig. 3. Lightbeam for Firefox visualization after visiting five websites. Sites accessed May 24, 2016, between 12:21PM and 12:24 PM Eastern Time.

The third party advertising networks and data partners visualized above use a variety of methods designed to create comprehensive profiles of a user’s web browsing history. This includes persistent identifiers (cookies), IP addresses, device identifiers, direct authentication (such as an email address), or probabilistic methods (such as browser fingerprinting).²⁵ Furthermore, this information can be combined with offline data (appended data), such as a user’s in-store purchase history, for an even more comprehensive consumer profile.

Many of the leading online platforms also correlate data across websites. For example, many websites (including WebMD) carry social media plug-ins that allow those social media platforms to compile browsing histories of individuals across the Internet and link that browsing activity to the same user’s social media behavior. If a consumer browsing the Web sees a Twitter button on a website they visit, that indicates that data on their browsing pattern on that Website is going to Twitter to help serve them with targeted advertisements on Twitter.

Mobile apps often collect even more granular information, such as information about the user’s in-app behavior, and other mobile data such as the Calendar or Contacts. Access to some mobile data (such as *Location Services*) requires the user’s Opt In permission, but access to other mobile

²⁵ See generally, Jules Polonetsky & Stacey Gray, FUTURE OF PRIVACY FORUM, *Cross-Device: Understanding the State of State Management* (Nov. 2015), available at https://fpf.org/wp-content/uploads/2015/11/FPF_FTC_CrossDevice_F_20pg-3.pdf.

information (such as the nearby Wi-Fi networks, from which location can be inferred) often does not. Some leading apps serve ads based on knowing what other apps are installed on a user’s device. WebMD, for example, in addition to tracking and sharing the data visualized above, has a mobile app that enables it to track users across desktop and mobile platforms.²⁶

These processes usually occur without directly sharing explicitly personal information—rather, for security and privacy reasons, industry players typically match up individual behavior using “hashed” identifiers. And many, including Commissioner Ajit Pai,²⁷ have pointed out that online tracking has generated benefits for consumers, including the availability of free and reduced-cost online content subsidized by online advertising that can be made more efficient and relevant through information about online audiences.

There have been many industry efforts in recent years to self-regulate the market in order to alleviate these privacy concerns and build consumer trust, and there are benefits to these efforts that we urge the FCC to recognize. **By acting in accordance with the leading best practices of the online ecosystem, the FCC can create a standard that will be universally relevant and provide a consistent and comprehensible level of protection for consumers.** *See infra*, Parts III-IV.

The FCC should first understand that its rulemaking covers data that is already being used, shared, and traded by a wide range of companies other than ISPs for the identical purposes covered by this NPRM subject to the FTC’s broad regulatory authority.

Why They Track: Understanding Ad Effectiveness

Most of the online tracking that occurs—whether via cookies or other unique identifiers—is done with the purpose of measuring ad effectiveness. As a user browses a first party website (say, [newyorktimes.com](http://www.nytimes.com)), third party ad networks typically dispatch cookies that can uniquely identify her as the person who viewed specific ads, and whether or not she has clicked on them. This user’s data—including, sometimes, a log-in authenticator, such as an email address associated with that website—is sent off to a series of intermediate data management providers and partners.

A data management provider (“data vault” intermediate such as BlueKai) may serve simply to match up the user’s data with data about that same user, from other sources—such as from other

²⁶ *Medscape Goes Mobile with New CME & Education App*, PR NEWSWIRE (Jul. 7, 2015), <http://www.prnewswire.com/news-releases/medscape-goes-mobile-with-new-cme--education-app-300109628.html> (last visited May 24, 2016).

²⁷ *Examining the Proposed FCC Privacy Rules*, Hearing Before the S. Comm. on the Judiciary, Subcomm. on Privacy, Tech., and the Law, 114th Cong. (2016), Testimony of Commissioner Ajit Pai, at 1, 3, <https://www.judiciary.senate.gov/imo/media/doc/05-11-16%20Pai%20Testimony.pdf>. *See also Notice* (accompanying statement of Commissioner Ajit Pai).

online ad networks, or from offline purchasing behavior. Typically, this user data, which usually contains PII, is “hashed” during this matching process, a technique which involves transforming the data into a shorter, fixed-length string of characters.²⁸ After matching data on individual users from a broad range of sources, the data provider often has comprehensive information of online and offline behavior. This can be used by partner companies who measure effectiveness of an ad campaign by comparing purchases from the matched group to a control group. *See* Fig. 4.

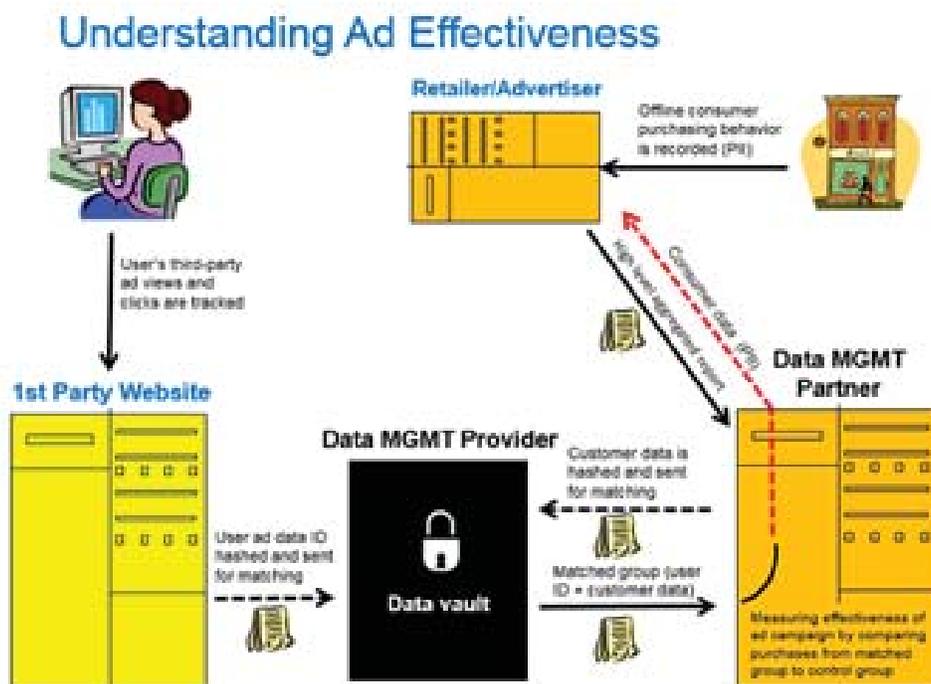


Fig. 4. Understanding Ad Effectiveness

The Challenging Prognosis for Publishers

The exponential increase in third party tracking and data management, described below, has affected online publishers in resonating ways. Although advertising revenue provides the basis for much free and reduced-price content, publishers who depend on third party advertising often express fears over a “race to the bottom” of quality content production and investment.²⁹ Margins of the profitability on digital advertising are slim, causing online publishers to acquiesce to market pressure to produce more pages of content more quickly, and to bend towards “clickbait”—

²⁸ See Margaret Rouse, What is hashing? - Definition from WhatIs.com, SEARCHSQLSERVER, <http://searchsqlserver.techtarget.com/definition/hashing> (last visited May 24, 2016).

²⁹ See, e.g., Don Marti, *Service journalism and the web advertising problem*, DIGITAL CONTENT NEXT (Apr. 27, 2016), <https://digitalcontentnext.org/blog/2016/04/27/service-journalism-and-the-web-advertising-problem/> (last visited May 24, 2016); Jasper Jackson, *Vox, belly fat and why even the cleverest digital publishers can have trouble with automated ads*, THE MEDIA BRIEFING (Apr. 14, 2014), <https://www.themediabriefing.com/article/vox-belly-fat-and-why-even-the-cleverest-digital-publishers-can-have-trouble-with-automated-ads> (last visited May 24, 2016).

sensational or provocative content whose sole purpose is to drive up page-views and generate more advertising revenue.

The FCC should promote competition in the online advertising market, thereby enhancing the opportunities for publishers of every size to succeed. Currently, five companies—Google, Facebook, Microsoft, Yahoo and AOL—lead the market in online advertising, bringing in 61% of total domestic digital ad revenue in 2014.³⁰ In the online services market, Facebook and Google account for 67% of mobile advertising, with social advertising comprising 70% of all of the revenue growth in display advertisements.³¹

Newspapers are particularly challenged to profit online, as they tend to lack external funding sources and have significant reporting costs. For this reason, news websites contain more third party trackers than any other website category,³² and yet they still face competition from many third parties who republish news content at low cost, while capturing ad revenue. In its proposed rules, the FCC should consider the effects those rules will have on the promotion of more competition into this already challenging market, and facilitate the small but growing role of ISPs.

Given the difficult state of the market for quality news media, the FCC should facilitate the entry of ISPs into the online ad market to support competition.

The Democratization of Data

In regulating the uses of data related to online behavior, it is important for the FCC be aware of the rapid pace of change in technology related to online tracking. The FCC has cited, as its primary rationale for the proposed regulation, the fact that ISPs are in a position “to develop highly detailed and comprehensive profiles of their customers.”³³ In particular, the Commission has noted that “a consumer, once signed up for a broadband service, simply cannot avoid that network in the same manner as a consumer can . . . switch search engines . . . surf among competing websites, and select among diverse applications . . .”³⁴

Decades ago, the leaders in the world of ad tracking were those companies who had access to the most data. Leading ad networks boasted about the breadth of their networks, the partners who shared data with them, and the third party data that they had linked to cookies. In 1999-2001, the

³⁰ PEW RESEARCH CENTER, STATE OF THE NEWS MEDIA 2015 (April 2015), *available at* <http://www.journalism.org/2015/04/29/state-of-the-news-media-2015/>.

³¹ INTERACTIVE ADVERTISING BUREAU, INTERNET ADVERTISING REVENUE REPORT (April 2016), *available at* <http://www.iab.com/insights/iab-internet-advertising-revenue-report-conducted-by-pricewaterhousecoopers-pwc-2/>.

³² Princeton Web Census, Steven Englehardt & Arvind Narayanan, *Online tracking: A 1-million-site measurement and analysis*, *available at* <https://webtransparency.cs.princeton.edu/webcensus/>.

³³ *Notice* at para. 4.

³⁴ *Id.*

merger of DoubleClick, a leading advertising network, with Abacus, an offline data collector, was driven by DoubleClick’s desire to expand its network by adding the rich data from the Abacus data co-op to its own data on web surfing patterns. For ad networks, assembling and linking data on a range of demographic, psychographic, and purchase history information was still an expensive and technologically complex endeavor.

Today, in sharp contrast, data has been “**democratized.**” Advances in technology have lowered the costs of storing and managing data, and as a result, individuals and small businesses have direct access to unprecedented amounts of data about themselves and others.

BlueKai, the key data provider in Oracle’s new data division, currently offers more than eighty different data attributes to its business customers for marketing.³⁵ By aggregating data from many sources, the BlueKai Exchange can offer detailed access to “over 350 million global in-market shoppers” for ad targeting.³⁶ Consumers are identified based on their Web browsing as having specific demographic profiles or purchasing intents—e.g. shopping for a car or travel services, or falling into a custom category such as “Back to School Shopper” or “Graduation Gift Buyer.”³⁷



Fig. 5. Branded data aggregators providing data to the BlueKai Exchange³⁸

³⁵ ORACLE CLOUD, ORACLE DATA AS A SERVICE (DAAS) FOR MARKETING (March 2016), at 3-10 available at https://docs.oracle.com/cloud/latest/daasmarketing_gs/DSMKT/DSMKT.pdf.

³⁶ *Id.* at 3-6.

³⁷ *Id.* at 3-10, 3-13.

³⁸ *Id.* at 3-2 to 3-5.

For example, according to the BlueKai Marketplace, more than 72% of people who are considering buying a car use the Internet for research and comparison before making that purchase.³⁹ As a result, the ability to send an advertisement to an individual as she navigates across different auto research sites, even if it is only by way of a hashed cookie or other unique identifier, contains great value:

“Until now, there hasn't been one place to buy this type of data for targeting in-market auto buyers at scale. We are meeting this need by aggregating valuable auto shopping and research activities **across the Internet**, and building the world's largest database of true auto intenders. Unlike ad networks, we do not sell ads or impressions, it simply provides **data a-la-carte** for marketers, ad networks or publishers to boost the quality and scale of ad targeting initiatives.”⁴⁰

Cross-Device: Bridging the Disconnect between Devices, Browsers, and Apps

Increasingly, users now access the Internet from a diverse range of connected devices. However, because cookies are specific to each unique device and browser, web publishers and third parties may not recognize that the same person is behind each device and browser. When a user visits a website from his laptop to check, for example, basketball game statistics, a cookie is delivered onto that laptop. However, if he uses his phone later to check the same content, a new cookie is set, as the publisher does not recognize him as the same visitor. This lack of connectedness restrains web publishers from delivering customized and consistent content, services, and features, as well as from enabling the tracking and customized advertisements that often allow the content to be provided to the user for free or at a reduced cost.

In addition to the divide between devices, there is a lack of communication between mobile web browsers and between mobile apps that constrains tracking even on the same device. While mobile web publishers and content providers may use cookies (to the extent permitted by the mobile web browsers), mobile apps do not. Since mobile platforms do not support the use of cookies by apps, an app cannot place cookies into storage on the device; instead, it must rely on the device's platform-level identifier.

Initially, app developers and other third parties tracked user behavior in apps using a range of operating system identifiers, device identifiers, MAC addresses, and other identifiers assigned by manufacturers or operation systems and permanently linked to the device. This generated privacy concerns from advocates who criticized the use of identifiers that were fixed and that could not be controlled by users. In response to concerns over the use of permanent device identifiers by third parties, mobile platforms such as iOS and Android began using new advertising identifiers (e.g.

³⁹ *Id.* at 3-22.

⁴⁰ *Id.* at 3-7.

the Apple IDFA, and the Android Advertising ID), which can be re-set by the user. As the cookie model becomes increasingly ineffective, third party ad networks are turning to a new range of deterministic and probabilistic methods to associate an individual’s online behavior over time and across devices. For example, a third party advertiser can match the cookie of a user who has provided her email address to a partner website, with the mobile activities (via Mobile Ad ID) of a user who has provided that same email address to any partner app.

Data Matching via Websites and Apps

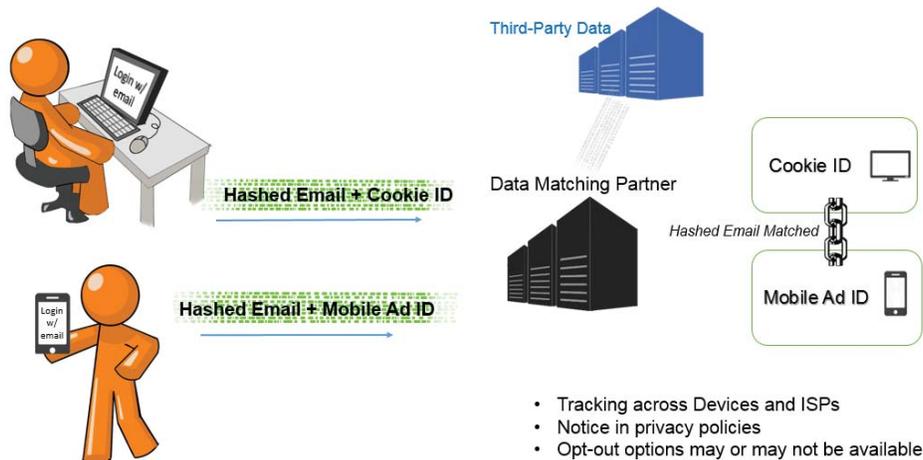


Fig. 6. Data Matching via Websites and Apps (Deterministic).

In the absence of authenticated data, it is often still possible to track individuals, with different measures of accuracy, across devices and platforms on the basis of statistical information gathered from the device, browser, app, and operating system. For example, this information can include the fact that multiple devices (say, a laptop and a phone) consistently use the same home Wi-Fi router, and are turned on at roughly the same time every evening. Using these kinds of rough data points, and many others that may be collected about a device, a system can infer—within ranges of confidence—that those devices are being used by the same person.

One common example of probabilistic matching occurs when an individual uses multiple devices on the same home Wi-Fi router. This enables the collection of the user’s IP Address, which can be matched to cookies and mobile advertising identifiers to provide cross-device tracking as well as geographically targeted content. See Fig. 7.

Ad Networks using Home Wi-Fi



- Tracking across Devices and ISPs
- Notice in privacy policies
- Opt-outs may or may not be available

Fig. 7. Ad Networks using Home Wi-Fi (Probabilistic)

Additional methods rely more heavily on probabilistic techniques. For example, browser fingerprinting, in which the user's browser is queried for its agent string, screen color depth, language, installed plug-ins with supported mime types, time zone offset and other capabilities, such as local storage and session storage. See Fig. 8.

Browser Fingerprinting

- Browser is queried for its agent string, screen color depth, language, installed plug-ins with supported mime types, time zone offset and other capabilities, such as local storage and session storage.
- 18 bits of entropy, meaning that only one in 286,777 other browsers will share its fingerprint.
- Industry terminology: probabilistic targeting, server side tracking or device recognition.
- For mobile, time differential latency also used.

```

navigator.userAgent // "Mozilla/5.0 (X11; Linux i686)
AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/27.0.1453.110 Safari/537.36"

navigator.language // "en-US"

var plugins = $.map(navigator.plugins, function(p) { var
mimeTypes = $.map(p, function(mimeType) { return
[mimeType.type, mimeType.suffixes].join("~"); }) join(','); return
[p.name, p.description, mimeTypes].join(':'); }); $.each(plugins,
function(i, p) { // truncate only for blog example if
(p.length > 80) { console.log(p.substring(0, 77) + '...'); } else {
console.log(p); } }); /* Shockwave Flash:Shockwave Flash 11.7
r700:application/x-shockwave-flash~swf.a...

Chrome Remote Desktop Viewer... Widevine Content Decryption
Module:Enables Widevine

licenses for playback of ... Native Client::application/x-nacl~nexe
Chrome PDF Viewer::application/pdf~pdf.application/x-google-
chrome-print-prev... Google Talk Plugin Video Accelerator:Google
Talk Plugin Video Accelerator ver... Google Talk Plugin:Version:
4.0.1.0:application/googletalk~googletalk Google Talk Plugin
Video Renderer:Version: 4.0.1.0:application/o1d~o1d Shockwave
Flash:Shockwave Flash 11.2 r202:application/x-shockwave-
flash~swf.a... /screen.colorDepth // 24 new
Date().getTimezoneOffset() // -240 !!window.localStorage //
true !!

window.sessionStorage // true
    
```

Fig. 8. Browser Fingerprinting

In the current ecosystem, ISPs play a small but growing role in maintaining state with a user across platforms and devices. For example, an ISP may be able to recognize a person who accesses the Internet over the same network via different devices as the same user. This can be done by including a unique identifier for that specific ISP, subject to notice and the ability to opt out. See Fig. 9.

The FCC’s proposed rules would restrict ISPs’ ability to provide state management activities that support ad reporting or delivery by restricting this practice. In the recent Verizon/AOL settlement, for example, Verizon agreed to standards of notice and choice approved by the FCC for its users with respect to the Verizon identifier, allowing users to opt in to its use by third parties, but continuing to use the identifier for the 40% of websites that use AOL’s ad network subject to an Opt Out.⁴¹

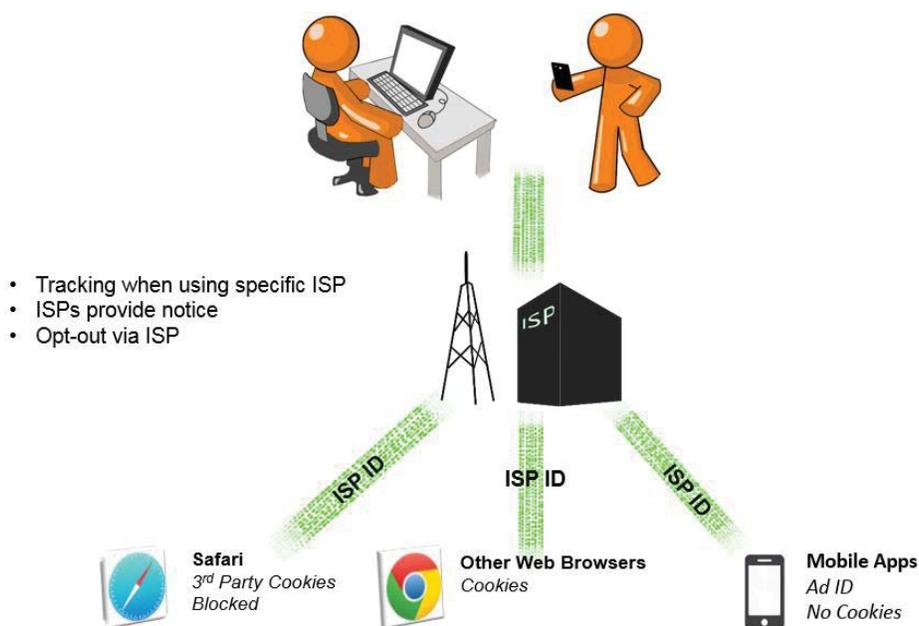


Fig. 9. The role of Internet Service Providers (ISPs) in cross-device tracking.

⁴¹ *In the Matter of Cellco Partnership, d/b/a Verizon Wireless*, File No. EB-TCD-14-00017601, (Mar. 7, 2016). See also Julia Angwin, *Verizon-FCC Settlement Does Not Apply to Verizon’s Tracking of its AOL Customers*, SITE PRO NEWS (Mar. 9, 2016), <http://www.sitepronews.com/2016/03/09/verizon-fcc-settlement-does-not-apply-to-verizons-tracking-of-its-aol-customers/>.

Geo-Location Data

With Smartphone ownership becoming nearly universal—86% ownership among those aged 18-29, and 83% ownership among those aged 30-49⁴²—it has become easier than ever before to track individuals based on their geo-location over time. Location data can be gathered directly by many of the sensors on a mobile device, and may become available to a range of third parties, including the device manufacturer, the operating system, the apps, ad networks, and others. This data can also be gathered indirectly by third parties that interact with the device, or that can detect the device.

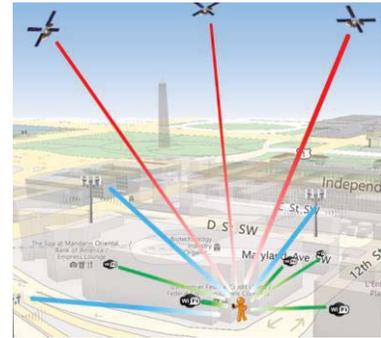


Fig. 10. Location Services

(1) Location Services

Smartphone users are often most familiar with “Location Services,” the service derived and controlled by the mobile operating system (OS). This service aggregates data from different sources—including GPS, cellular triangulation, nearby Wi-Fi signals, and Bluetooth positioning—to pinpoint the device’s location more accurately than any individual system.

The mobile OS requests the user’s permission for its own uses during the initial set-up stage for the device. The consent is a simple pre-checked “NEXT”, making it easy for the mobile OS to get consent and for the consumer to manage the consent. See Fig. 11.

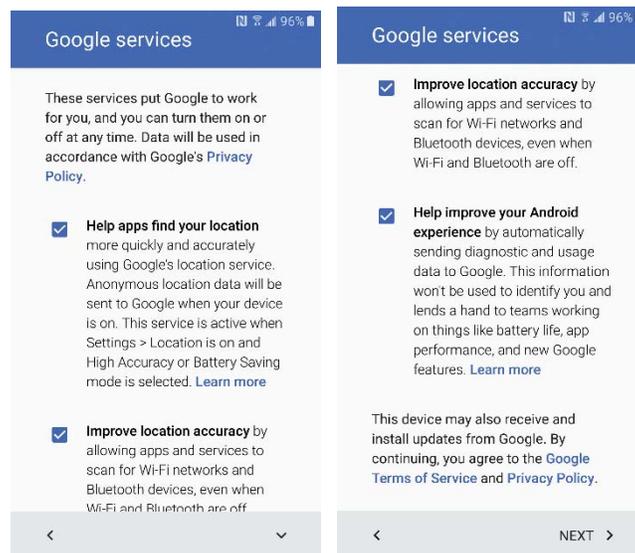


Fig. 11. Consent Screens at Initial Phone Set Up (Samsung Galaxy S7)⁴³

⁴² PEW RESEARCH CENTER, *The Demographics of Device Ownership*, (Oct. 29, 2015), <http://www.pewinternet.org/2015/10/29/the-demographics-of-device-ownership/>.

⁴³ Screen captures on file with Authors (captured May 24, 2016).

Apps and websites must get affirmative permission from the user via the OS to access this data. However, once an app accesses location data, industry standards call for consent for sharing the data only if the precise geo-location is shared. This does not include sharing of 5-digit zip code, city, or location based on the IP address.⁴⁴ First and third parties are required to give clear, meaningful, and prominent notice of precise location data transfer to or collection by third parties. First parties must obtain consent prior to transferring to third parties. There are exceptions for system management, market research, and de-identified data.

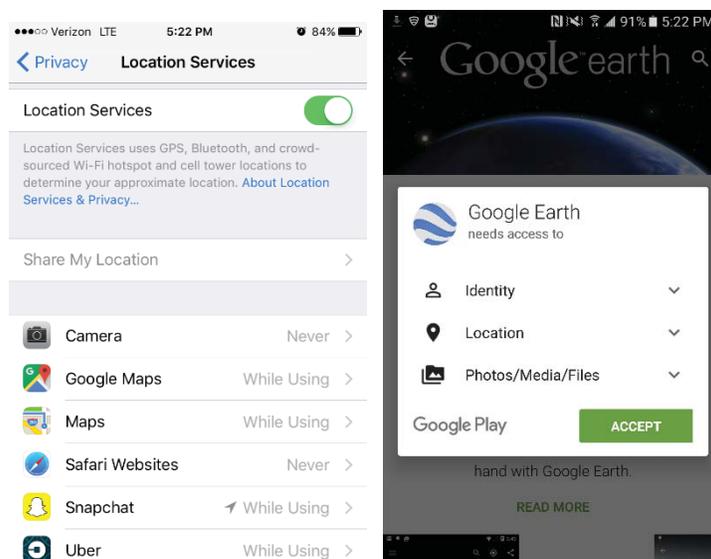


Fig. 12. Location Services permission screens on iOS 9 (left) and Android M (right)⁴⁵

In the current iPhone OS, the user’s location permission is separated into categories of “Never,” “While Using,” or “Always,” with an arrow glyph indicating concurrent app usage. Permission to access Location Services is also separated between foreground and background app use—in other words, if an app has been using *Location Services* in the background (while the app was not in use), the OS will notify the user and re-confirm whether he or she wants to continue granting the app that permission. See Fig. 13.

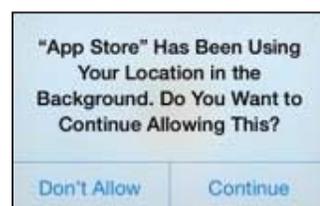


Fig. 13. iOS notification of app background Location Services usage.

With respect to what is considered “precise location,” industry guidelines are similar. The DAA Principles define precise location as data “sufficiently precise to locate a specific individual or device,”⁴⁶ and the NAI Code defines precise location as “information that describes the precise geographic location of a device derived through any technology that is capable of determining

⁴⁴ *Id.*

⁴⁵ Screen captures on file with Authors (captured May 24, 2016).

⁴⁶ DIGITAL ADVERTISING ALLIANCE, APPLICATION OF SELF-REGULATORY PRINCIPLES TO THE MOBILE ENVIRONMENT, at 9-10, available at http://www.aboutads.info/DAA_Mobile_Guidance.pdf.

with reasonable specificity the actual physical location of an individual or device.”⁴⁷ Under the NAI’s guidance, the factors relevant to whether location data is precise, and therefore requires Opt In consent, include: the area of the identified location (e.g. how many decimal places were used in the coordinates), the population density of the located area, the accuracy of the data, and the presence and detail of the location’s timestamp (e.g., whether it describes a user’s location at a specific millisecond or a specific month).⁴⁸

Cell Tower Locations

Mobile devices receive limited location information from the locations of nearby base transceiver stations (cell towers). Cell towers broadcast unique Cell IDs, which are compiled in publicly available databases, such as OpenCellID, Mozilla, or MyInkikov GEO.⁴⁹ Similar private databases of Cell IDs, operated by companies such as Combain, LocationAPI.org, Navizon, and WiGLE, are often much larger, with databases of up to over 72 million unique cell towers.⁵⁰

Depending on the density of cell towers in a given geographic area, cell tower location data can often be relatively inaccurate at locating an individual at a single point in time. However, it can be useful when combined with other sources of data, explained below, such as Wi-Fi signals and beacons. Mobile operating systems receive this data without requesting permission from the user.

Table 1. Location databases available based on Cell IDs

Cell Tower Database	Unique Cell Towers	Availability
OpenCellID	> 6 million	Public
Combain	> 72 million	Private
LocationAPI.org	> 72 million	Private
Mozilla	> 26 million	Public
Navizon	> 71 million	Private
Mylnikov GEO	> 15 million	Public
WiGLE	> 6 million	Private

Carrier Triangulation

Uniquely, mobile ISPs can determine geo-location by analyzing signals from multiple surrounding cell towers. This data is less accurate than mobile operating systems’ *Location Services*, because

⁴⁷ NETWORK ADVERTISING INITIATIVE, 2015 Update to the NAI Code of Conduct (2015), at 5, available at http://www.networkadvertising.org/sites/default/files/NAI_Code15encr.pdf.

⁴⁸ NETWORK ADVERTISING INITIATIVE, GUIDANCE FOR NAI MEMBERS: DETERMINING WHETHER LOCATION IS IMPRECISE (July 20, 2015), at 3, available at http://www.networkadvertising.org/sites/default/files/NAI_ImpreciseLocation.pdf.

⁴⁹ See OpenCellID, <http://opencellid.org/> (last accessed May 24, 2016); Mozilla Location Service, *Map*, <https://location.services.mozilla.com/map#2/15.0/10.0> (last accessed May 24, 2016); Alexander Mylnikov, <https://www.mylnikov.org/download> (last accessed May 24, 2016).

⁵⁰ See LocationAPI, *Cell Tower & Wifi Coverage*, <http://locationapi.org/coverage> (last accessed May 24, 2016).

the latter are supplemented by GPS and Wi-Fi (which carriers do not generally receive). Carriers follow a precise code established by CTIA requiring sharing of precise location data only with user permission.⁵¹

Wi-Fi Based Location

Mobile devices can infer slightly more precise geo-location through their routine scanning for nearby Wi-Fi networks. Large databases exist of the unique identifiers (MAC addresses and SSID) of wireless routers and their known locations, which are continuously updated in a variety of ways. Databases can be updated by the mobile operating system itself, which via *Location Services* can periodically check a user's location via GPS, Cell ID, and Wi-Fi, and report back the SSID and MAC address data from nearby publicly broadcasted Wi-Fi access points. The operating system typically provides clear notice of this use during set-up, requesting permission for this use and permitting the user to disable it later if he changes his mind. Mobile apps may or may not have the ability to access the MAC addresses and SSIDs of local Wi-Fi networks from the device's routine scanning—while iOS apps do not have access, Android apps can access it without permission.

The use of publicly broadcasted MAC addresses and SSIDs for geo-location is ubiquitous, with companies such as Mozilla and Comban reporting databases of up to 886 million unique Wi-Fi networks.⁵² Despite the relatively public nature of these identifiers, most (but not all) WPS databases offer an Opt Out mechanism for users who don't wish to be included in the database. For instance, in 2012, Google created a unified approach for opting-out a particular access point from being included in its database, which involves appending the phrase “_nomap” to the end of the wireless router's SSID.⁵³ Mozilla similarly honors the _nomap method, but for most other databases, the user must typically visit their website and enroll their MAC address in order to be opted out.⁵⁴

⁵¹ See CTIA, THE WIRELESS ASSOCIATION, BEST PRACTICES AND GUIDELINES FOR LOCATION BASED SERVICES, Version 2.0 (May 23, 2010), <http://www.ctia.org/docs/default-source/default-document-library/pdf-version.pdf?sfvrsn=0>.

⁵² See Mozilla Location Service, *Statistics*, <https://location.services.mozilla.com/stats> (last accessed May 24, 2016).

⁵³ See Google Official Blog, *Greater choice for wireless access point owners* (Nov. 14, 2011), <https://googleblog.blogspot.com/2011/11/greater-choice-for-wireless-access.html>.

⁵⁴ See, e.g., Skyhook, *End User Opt-Out of Skyhook Products*, <http://www.skyhookwireless.com/opt-out-of-skyhook-products> (last accessed May 24, 2016); Windows Phone, *Opt out of location services*, <https://www.windowsphone.com/en-us/support/location-block-list> (last accessed May 24, 2016).

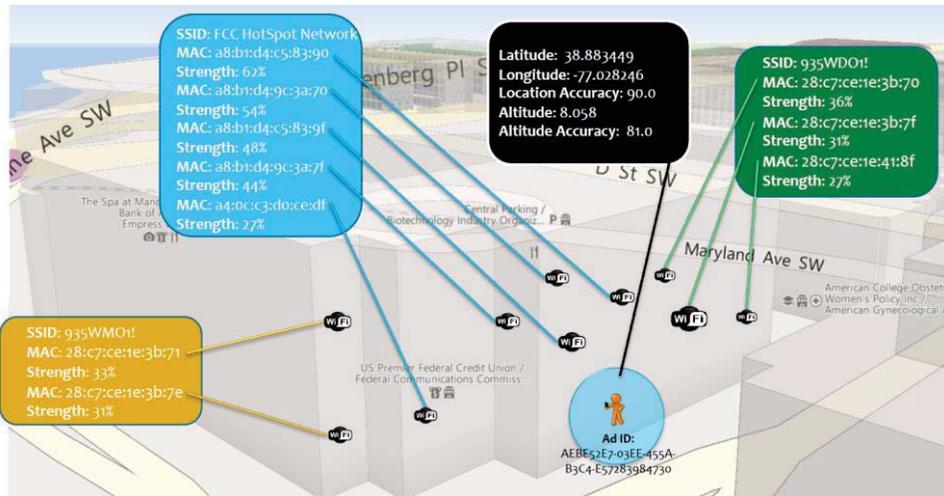


Fig. 14. Mobile devices can infer geo-location by scanning for the MAC addresses and SSIDs of nearby publicly broadcasted Wi-Fi access points.

Advertising networks also routinely use the publicly broadcasted MAC addresses and IP addresses from home Wi-Fi access points. In an effort to recognize an individual user across devices, data algorithm partners can create a link between the browser (cookies) and a mobile device's unique identifier (Ad ID) by recognizing that they are connected to the Internet via the same home Wi-Fi router (MAC address and IP address).

Mobile Location Analytics

Facilities such as airports, stores, and hotels use Mobile Location Analytics (MLA) technology to understand the traffic patterns of people in their venues. By learning and using insights, such as how long customers stand in line and how they generally move around an area, these facilities can enhance operational efficiency and improve user experience.

Most MLA technologies operate by detecting your device's Wi-Fi MAC address or Bluetooth address, a 12 digit string of letters and numbers assigned to your device by its manufacturers. Consumers can opt out of being included in these analytics by entering their Wi-Fi and Bluetooth MAC address at www.smart-places.org for participating companies. Turning off the device's Wi-Fi or Bluetooth will also prevent a consumer's MAC address from being detected.



Fig. 15. Mobile Location Analytics (MLA) signage for participating companies.

Beacons

Another method allowing apps to collect location information is the use of “beacons,” small devices consisting of a chip and other electronic components (e.g. antenna) on a small circuit board. Beacons are essentially radio transmitters that broadcast one-way signals to devices that are equipped to receive them. These devices allow the mobile app to determine (typically via Bluetooth) a user’s location in proximity to the beacons, which may be installed at various places throughout a retail location, such as in front of a special display of products.⁵⁵

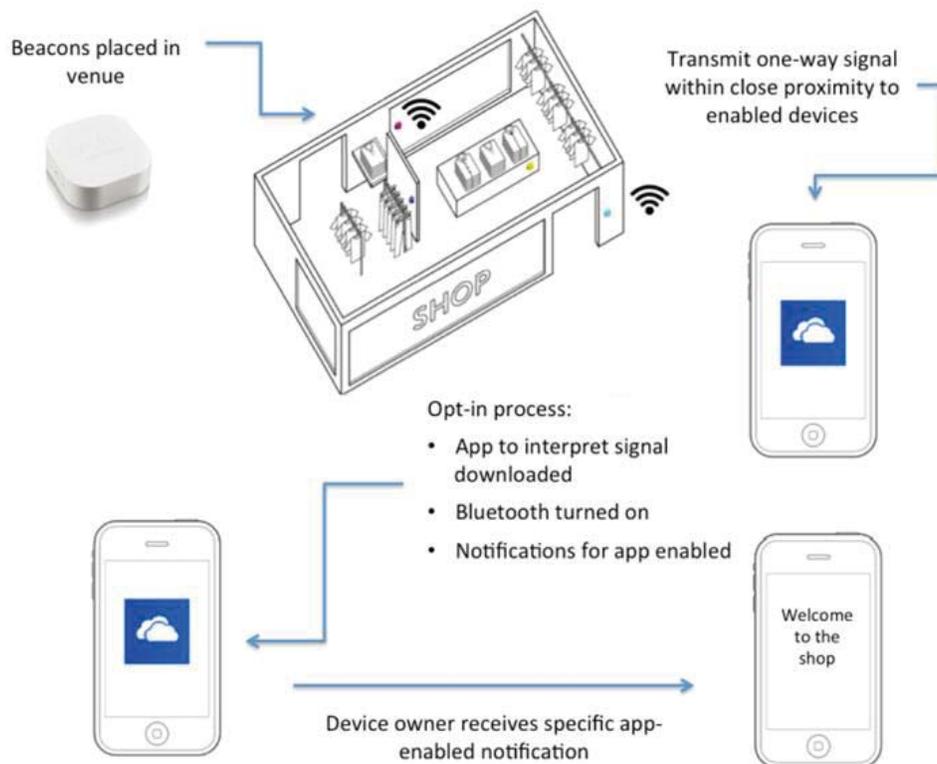


Fig. 16. Beacons transmit one-way location signals within close proximity to enabled devices

Geo-location data is available to a wide range of different companies and business models, many holding information that is much more granular and comprehensive than the data collected by ISPs.

⁵⁵ See generally, GREG STERLING, JULES POLONETSKY & STEPHANY FAN, FUTURE OF PRIVACY FORUM, UNDERSTANDING BEACONS: A GUIDE TO BEACON TECHNOLOGIES (Dec. 2014), https://fpf.org/wp-content/uploads/Guide_To_Beacons_Final.pdf.

III. Benefits and Limitations of Current Industry Rules

In the rapidly changing environment described above for a range of global consumer-facing services, the FTC has the broad regulatory authority to ensure that companies engage in fair and non-deceptive practices. Recent enforcement actions demonstrate that the FTC's jurisdiction is an effective model for online privacy. Looking ahead, as well, a range of upcoming FTC Workshops—on drones, SmartTV advertising practices, and malware⁵⁶—show that the FTC, through public education and broad involvement of public and private stakeholders, is increasingly building policies and guidance that are shaping practices in this ecosystem.

The FTC's enforcement is bolstered further when combined with industry efforts to self-regulate the market in order to alleviate privacy concerns and build consumer trust. The National Advertising Initiative (NAI) and Digital Advertising Alliance (DAA) have enforceable codes governing the uses of online data collected and used for the purposes of online behavioral advertising (interest based advertising) and re-targeting, as well as for location based advertising, including when carried out by many ISPs.⁵⁷ The Wireless Association (CTIA) has enforces guidelines around mobile data, especially geo-location.⁵⁸ The Future of Privacy Forum (FPF) operates the Mobile Location Analytics Code, a self-regulatory program for companies tracking mobile phones,⁵⁹ has promulgated, together with Software & Information Industry Association, an enforceable Student Privacy Pledge,⁶⁰ and a Working Group is finalizing best practices for data from consumer wellness and wearables apps.⁶¹

With the right structural controls in place, the NAI Code and DAA Principles can be enforced by trusted non-affiliated industry partners such as the DMA and the BBB. Recently, for example, the BBB brought three enforcement actions against app providers who were not following the DAA Principles with respect to the collection and use of mobile data.⁶² As a result of these enforcement

⁵⁶ See Press Release, *FTC to Host Fall Seminar Series on Emerging Consumer Technology Issues* (March 31, 2016), <https://www.ftc.gov/news-events/press-releases/2016/03/ftc-host-fall-seminar-series-emerging-consumer-technology-issues>.

⁵⁷ See NETWORK ADVERTISING INITIATIVE, 2015 UPDATE TO THE NAI CODE OF CONDUCT (2015), available at http://www.networkadvertising.org/sites/default/files/NAI_Code15encr.pdf; Digital Advertising Alliance, *Digital Advertising Alliance (DAA) Self-Regulatory Program*, <http://www.aboutads.info/>.

⁵⁸ CTIA, *Best Practices and Guidelines for Location Based Services*, <http://www.ctia.org/policy-initiatives/voluntary-guidelines/best-practices-and-guidelines-for-location-based-services>.

⁵⁹ Future of Privacy Forum, *Mobile Location Analytics Code of Conduct*, <https://fpf.org/wp-content/uploads/10.22.13-FINAL-MLA-Code.pdf>.

⁶⁰ Future of Privacy Forum, *Student Privacy Pledge*, www.studentprivacypledge.org (last accessed May 24, 2016).

⁶¹ *Future of Privacy Forum*, www.fpf.org (last accessed May 24, 2016).

⁶² Jedidiah Bracy, *Self-regulatory group takes action against three app publishers*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS (May 5, 2016), <https://iapp.org/news/a/self-regulatory-group-takes-compliance-action-against-three-app-publishers/>.

actions, these providers changed their practices to comply with DAA guidelines.⁶³ The guidelines are also enforceable by the FTC which has used them in important enforcement actions.⁶⁴

Opt In Required for Sensitive Data and Precise Geo-Location

Consistent with the FTC’s 2012 Report,⁶⁵ both NAI and DAA require their members to obtain users’ affirmative Opt In consent before collecting or using certain sensitive data. This includes precise geo-location data, as well as a non-exhaustive list of data considered *per se* sensitive, including Social Security Numbers or other government-issued identifiers, financial account or insurance plan numbers, sexual orientation, and health status (described more below). The NAI and DAA also prohibit certain uses of data entirely, including the use of data for employment eligibility, credit eligibility, health care eligibility, or insurance eligibility, underwriting, and pricing.

“Conspicuous” and “Easy to Use” Opt Out Required for Non-Sensitive Data

Even for data that is not considered sensitive, industry codes require that consumers be provided with the ability to exercise choice with respect to the collection, sharing, and use of that data for purposes of online behavioral advertising (OBA). Under DAA guidelines, this can involve an “enhanced notice link” placed “in” or “around” the advertisement, from the choice mechanism on www.AboutAds.info, or notice via a first party web site operator’s own disclosure.⁶⁶ Notice is often provided by the use of the Ad Choices Icon, *see* Fig. 17, which re-directs users to a website that provides information about OBA and the ability to opt out. In the words of the DAA implementation guidance: “In all cases, the choice mechanism should be *easy to use*.”⁶⁷ Similarly, under the NAI Code of Conduct, members are required to ensure that the websites visited by users that engage in



Fig. 17. AdChoices icon

⁶³ See COUNCIL OF BETTER BUSINESS BUREAUS, Case Number 62-2016 (May 4, 2016), at 9, *available at* <http://www.bbb.org/globalassets/local-bbbs/council-113/media/behaviorial-advertising/bearbit-decision.pdf>; COUNCIL OF BETTER BUSINESS BUREAUS, Case Number 61-2016, at 9, *available at* <http://www.bbb.org/globalassets/local-bbbs/council-113/media/behaviorial-advertising/spinrilla-decision.pdf>; COUNCIL OF BETTER BUSINESS BUREAUS, Case Number 63-2016, at 11, *available at* <http://www.bbb.org/globalassets/local-bbbs/council-113/media/behaviorial-advertising/top-free-games-decision.pdf>.

⁶⁴ See Federal Trade Commission, Press Release, *Google Will Pay \$22.5 Million to Settle FTC Charges it Misrepresented Privacy Assurances to Users of Apple’s Safari Internet Browser* (Aug. 9, 2012), <https://www.ftc.gov/news-events/press-releases/2012/08/google-will-pay-225-million-settle-ftc-charges-it-misrepresented>.

⁶⁵ Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change* <https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf>.

⁶⁶ See DIGITAL ADVERTISING ALLIANCE, *SELF-REGULATORY PRINCIPLES FOR ONLINE BEHAVIORAL ADVERTISING* (July 2009), *available at* <http://www.aboutads.info/resource/download/seven-principles-07-01-09.pdf>.

⁶⁷ *See id.*

OBA/IBA “clearly and conspicuously” post notice, including a “conspicuous link to an Opt-Out mechanism.”⁶⁸

The Value of Uniformity

One important and often overlooked benefit of industry self-regulation in this context is the ability to set uniform standards across the entire Internet ecosystem. The DAA is composed of six trade associations representing website publishers, internet service providers, cell phone carriers, social networks, advertisers, offline data providers, and digital technology companies.⁶⁹ It also encompasses the NAI, which joined the DAA in 2010.⁷⁰ As a result, the DAA’s Principles for OBA and Multi-Site Data collection govern the entire Internet ecosystem and impose obligations not only on ad tech companies such as networks and platforms, but also on website publishers and brand advertisers.⁷¹

NAI Code and DAA Principles, combined with the FTC’s broad regulatory authority, set the baseline for what is considered appropriate and responsible data use for online ad targeting, regardless the source of the data. With the average person visiting ninety-six or more separate domains per month,⁷² and an exponential increase in third party data sharing, it is unreasonable to expect consumers to differentiate between the privacy practices of different platforms and publishers in the Internet ecosystem.

The Multitude of Consumer Controls

Although the NAI Code and DAA Principles provide uniformity with respect to the norms of responsible and appropriate data uses, the mechanisms by which companies comply range from leading examples of best practices to mechanisms which are more confusing, hard to understand, or difficult to find.

For many consumers, deleting cookies is increasingly ineffective in light of the proliferation of non-cookie-based deterministic and probabilistic tracking techniques, discussed *infra* Part II. Central ad industry opt-outs are effective to the extent that they prevent users from receiving targeted advertisements based on their browsing, but opt-out cookies are deleted when users clear cookies, and often cannot be set on the Safari browser which limits third party cookies. Some

⁶⁸ NETWORK ADVERTISING INITIATIVE, 2015 UPDATE TO THE NAI CODE OF CONDUCT (2015), at 7, *available at* http://www.networkadvertising.org/sites/default/files/NAI_Code15encr.pdf.

⁶⁹ *Id.* at 2-4.

⁷⁰ *Id.*

⁷¹ *Id.*

⁷² Nielsen, *January 2013: Top U.S. Entertainment Sites and Web Brands*, (Mar. 22, 2013), <http://www.nielsen.com/us/en/insights/news/2013/january-2013--top-u-s--entertainment-sites-and-web-brands.html> (last accessed May 24, 2015).

companies treat an opt-out as an opt-out of cross device tracking, but many do not.⁷³ Some companies continue to target ads based on third party appended data across sites, but others do not. Some companies stop sending a unique tracking identifier when consumers opt-out, but many continue tracking for analytics and ad effectiveness purposes. Most companies do not respect Do Not Track browser settings, but a few do have strong Do Not Track policies.⁷⁴

Mobile apps, which do not use cookies at all, require users to download a separate opt-out to participate in NAI or DAA opt-out programs. Consumers can enable the “Limit Ad Tracking” setting in iOS or Android, but not every ad network cooperates or respects this setting. Consumers who wish to opt out of having their home Wi-Fi included in geo-location databases can manually add the words “_nomap” to the name of their home router to opt out for Google and Mozilla.⁷⁵ But for Skyhook, Microsoft, and others, it remains necessary to visit those sites and enter their hardware MAC addresses to be opted out of geo-location tracking.

By crafting industry-relevant rules for ISPs, the FCC will have the valuable opportunity to set standards that will be relevant to this larger ecosystem through *meaningful* Opt Out controls, retention limits, ethics oversight to help make decisions relevant to civil rights and fairness, and multi-stakeholder guidance around the uses of sensitive and out-of-context data. *See infra* Part IV.

We note that leading companies provide effective controls that work across devices, that limit data sharing, respect Limit Ad Tracking, provide user-friendly dashboards, and enable users to access and correct data. However, some others do not. The FCC, by acting in a manner that supports the leading practices in the general ecosystem will be providing consumer protection that is relevant broadly.

The FCC has the opportunity to set relevant and influential rules that will raise the bar for the entire industry by focusing on current leading practices.

IV. Call for Reasonable Standards that will Elevate Industry Norms

The FCC’s proposed rules, by deviating markedly from industry norms, will not be relevant to the rest of the Internet advertising and data exchange ecosystem. Other industry players are very likely to retain their current frameworks, because the standards created by the FCC will be seen as unique

⁷³ DIGITAL ADVERTISING ALLIANCE, APPLICATION OF THE SELF-REGULATORY PRINCIPLES OF TRANSPARENCY AND CONTROL TO DATA USED ACROSS DEVICES (Nov. 2015), *available at* http://www.aboutads.info/sites/default/files/DAA_Cross-Device_Guidance-Final.pdf.

⁷⁴ *See* Future of Privacy Forum, *Companies that have implemented Do Not Track*, <https://allaboutdnt.com/companies/>.

⁷⁵ *See* Google Support, *Configure access points with Google Location Service*, <https://support.google.com/nexus/answer/1725632?hl=en> (last visited May 24, 2016); Mozilla Location Service, *Opt-Out*, <https://location.services.mozilla.com/optout> (last visited May 24, 2016).

for ISPs. In addition, the rules will be considered irrelevant even for edge providers carrying out identical advertising activities to those captured by the Rule. As a result, consumers will see no change in their online experience or in the extent to which e data about their online activities are collected and used.

We urge the FCC to recognize that with appropriate technical and legal controls in place, the use of pseudonymous or not readily identifiable data for limited purposes, including cross-device tracking, can be permitted subject to meaningful and transparent Opt Out control. First, as a threshold matter, the FCC should issue rules that define “de-identified data,” recognizing that data falls on a **spectrum of identifiability** in line with FTC and NIST standards. We propose that with the right technical and legal controls in place, ISPs’ use of pseudonymous or not readily identifiable data for certain uses, including cross-device tracking, can set a high standard for the rest of the industry while allowing ISPs to participate in the online advertising market on a level playing field. Finally, we propose that the FCC establish a multi-stakeholder process to develop privacy rules for sensitive data and out of context uses of such data.

The FCC’s Proposed Rules will Exclude ISPs from the Market

As a threshold matter, the FCC has proposed an Opt In approach for all customer PI—defined very broadly⁷⁶—even though *most* of the same data is shared broadly throughout the rest of the Internet ecosystem under an Opt Out framework. In doing so, the Commission has expressed that it does not intend to prohibit ISPs from using data or exclude them from innovative data uses. **Yet the proposed Opt In has a cost: here, that cost is the exclusion of ISPs from a much larger market.**

Regulators should be aware that in the question of an Opt In versus an Opt Out regime, the default setting will be determinative of the existence of the data uses being regulated. As a result, when the underlying activity has social value—here, promoting competition and generating online content—the question of default settings is inextricably linked to the question of whether that underlying behavior is a net social good or an activity to be proscribed.⁷⁷

For this reason, we agree with those who have called strict opt-in/opt-out dichotomies “upside down,”⁷⁸ and called instead for contextual privacy controls that provide consumers with transparency and meaningful, granular control over their data.

⁷⁶ Notice at para. 56 et seq.

⁷⁷ See Omer Tene & Jules Polonetsky, *To Track or ‘Do Not Track’: Advancing Transparency and Individual Control in Online Behavioral Advertising*, 13 MINN. J. L. SCI. & TECH. 1 (2012).

⁷⁸ Scott Meyer, *The FCC’s Proposed New Privacy Rules - Is An Opt-In Solution The Only Way?*, GHOSTERY (April 20, 2016), <https://www.ghostery.com/intelligence/business-blog/business/Fcc-new-privacy-rules/>.

The FCC’s proposed rules will exclude one set of industry players (ISPs) from the market, but consumers will continue to see the same number of individually tailored advertisements, subject to inconsistent consumer choices. Furthermore, the FCC will be missing a valuable opportunity to set standards that *will* influence the rest of the ecosystem and substantially raise the bar for transparency, consumer understanding, and consumer control over their data.

With Appropriate Controls, the Use of “Pseudonymous” or “Not Readily Identifiable” Data Should be Permitted for Cross-Device State Management Subject to Meaningful Opt Outs and Other Safeguards

In order to create a uniform, elevated standard for consumers, as well as to support competition in a market that is performing poorly for premium publishers and news media, the FCC should permit ISPs to engage in certain limited uses of pseudonymous and not readily identifiable data.

In order to create a uniform, elevated standard for consumers, as well as to support competition in a market that is performing poorly for premium publishers and news media, the FCC should permit ISPs to use pseudonymous and not readily identifiable data for cross-device tracking, subject to a meaningful Opt Out. In addition, the FCC can raise the bar for the industry by regulating the uses of appended (offline) data, requiring strict retention periods, requiring internal company ethics oversight⁷⁹ to evaluate decisions relevant to civil rights and fairness, and establishing a multi-stakeholder process to develop rules around sensitive data and data uses that are out of context.

Robust and Meaningful Opt Out Mechanism

Many current Opt Out mechanisms only limit future OBA/IBA but continue to permit a range of tracking for other purposes. A strong Opt Out mechanism could more broadly limit tracking when applied and could be a more stable and persistent method of opting out across devices and browsers. For example, unlike other industry Opt Outs, an ISP Opt Out can be uniquely effective across devices and platforms, regardless of the consumer’s browser, because it is not dependent on cookie settings (which can be inadvertently erased by a user who clears cookies).

Appended Data

Current online advertising industry practices with regard to appended data are inconsistent. Some companies respect a user’s choices and limit ad targeting based on third party appended data. Others continue to use the appended data and only limit targeting based on web surfing activity. This often leads to confusion among consumers who do not understand that their offline behavior may become connected to their Web browsing through a variety of methods, including geo-

⁷⁹ See Future of Privacy Forum, *Big Data Ethics*, <https://bigdata.fpf.org/>.

location targeting and direct authentication (e.g. loyalty cards or providing an email address upon purchasing). The FCC has the opportunity here to raise the bar with respect to this standard by requiring ISPs to provide enhanced notice and the ability to opt out of ad targeting based on appending offline data to a customer’s profile. An effective Opt Out could more broadly limit targeting regardless of whether the data was appended or based on web browsing.

Retention Periods

Current industry standards around the retention of consumer data either don’t exist for many, or vary widely for those that do publish retention periods. Limiting the retention of data in a way that is relevant for the business purposes of the data used can be a strong privacy-enhancing step, as well as a way to minimize the risk of re-identification.

Ethics Oversight to Address Questions of Fairness, Discrimination, and Civil Rights

The White House recently called for a greater spotlight on private and public uses of data in ways that avoid algorithmic discrimination and support fairness and accountability.⁸⁰ We recognize that transparency and accountability are key aspects of overall corporate ethics, and have been engaged in efforts to bring together industry and academics to build consensus around issues of ethics in big data and research.⁸¹ The FCC should both permit and encourage ISPs to participate in these industry efforts by encouraging the development of internal ethical review oversight processes to address questions of how online advertising—among other possible uses of ISP data—may affect questions of civil rights, fairness, and discrimination.

The FCC Should Establish a Multi-Stakeholder Process to Develop Privacy Rules for Sensitive ISP Consumer Data and Out of Context Uses of Such Data

The NPRM raises important questions about how the FCC can best protect consumers’ privacy and encourage responsible, practical data practices by ISPs. In addition to our recommended rules for “pseudonymous” or “not readily identifiable” data discussed above, we urge the FCC to establish a multi-stakeholder process, led by the National Telecommunications & Information Association (NTIA), to determine the best way to approach ISPs’ use of data that is sensitive or broadly out of context, taking into account degrees of identifiability. A multi-stakeholder process would enable advocates, companies, technical experts and others to reach consensus on rules that protect consumers and are workable for ISPs.

⁸⁰ See EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS (May 2016), at 24, *available at* https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf (“How big data is used ethically to reduce discrimination and advance opportunity, fairness, and inclusion should inform the development of both private sector standards and public policy making in this space.”).

⁸¹ See Future of Privacy Forum, *Big Data Ethics*, <https://bigdata.fpf.org/>.

As discussed in President Obama’s Consumer Privacy Blueprint, open, transparent multi-stakeholder forums can enable stakeholders who share an interest in specific markets or business contexts to work toward consensus on appropriate, legally enforceable codes of conduct.⁸² It is important that multi-stakeholder processes be driven by the stakeholder community, with the government’s most helpful role as a facilitator and fair broker.⁸³ The NTIA has the necessary authority and expertise to convene multi-stakeholder processes that address consumer data privacy issues.⁸⁴ NTIA has a track record of successfully convening these sorts of initiatives, including successfully concluded engagements regarding mobile privacy notices and unmanned aircraft systems.⁸⁵ The mobile apps process resulted in enhanced privacy notices in apps used by more than 200 million users, and the UAS process recently established consensus best practices to protect privacy during commercial and individual drone operation.⁸⁶

As discussed above, the FTC’s 2012 Report and online advertising self-regulatory frameworks require affirmative Opt In consent before collection or use of certain sensitive data. There is broad agreement that out of context collection and use of data—*i.e.* data practices that are opaque and surprising to consumers—should be subject to more stringent privacy safeguards.⁸⁷ There is also broad consensus that the President’s 2012 Consumer Privacy Bill of Rights (CPBR) articulates high level principles that can usefully guide stakeholder discussion of important privacy issues. These principles—individual control, transparency, respect for context, security, access and accuracy, Focused Collection and Accountability—are consistent with widely accepted norms for responsible data practices. The CBPR calls for a flexible, context-based approach to notice and choice, consistent treatment of similar data, and alternative approaches to top-down, prescriptive regulation. We urge the FCC to establish a multi-stakeholder process, convened by NTIA, to determine the most appropriate privacy rules for ISPs’ collection and use of sensitive and out of context data in the broadband ecosystem.

⁸² THE WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY (February 23, 2012), at 2, <http://www.whitehouse.gov/sites/default/files/privacy-final.pdf>.

⁸³ *Id.* at 24.

⁸⁴ *Id.* at 26.

⁸⁵ See, e.g., Nat’l Telecomm. & Info. Admin., U.S. Dep’t of Commerce, *Multistakeholder Process: Unmanned Aircraft Systems* (May 19, 2016), <https://www.ntia.doc.gov/other-publication/2016/multistakeholder-process-unmanned-aircraft-systems> (last accessed May 24, 2016); Nat’l Telecomm. & Info. Admin., U.S. Dep’t of Commerce, *Privacy Multistakeholder Process: Mobile Application Transparency* (Nov. 12, 2013), <https://www.ntia.doc.gov/other-publication/2013/privacy-multistakeholder-process-mobile-application-transparency> (last accessed May 24, 2016).

⁸⁶ Nat’l Telecomm. & Info. Admin., U.S. Dep’t of Commerce, *Keynote Address of Assistant Secretary Strickling at Silicon Flatirons Conference* (Feb. 01, 2016) (“[E]nhanced privacy notices based on the [NTIA] code are now live in apps used by 200 million consumers and the numbers are growing.”); Press Release, Future of Privacy Forum, *Multi-Stakeholder Group Finalizes Agreement on Best Practices for Drone Use* (May 18, 2016).

⁸⁷ *Id.* at 15.

Conclusion

For the foregoing reasons, we urge the FCC to: (1) issue a rule that recognizes that de-identification is not binary, but that data exists on a spectrum of identifiability; (2) specifically recognize that non-aggregate data can be appropriately de-identified; (3) establish a framework that allows ISPs to use data that are pseudonymous or not readily identifiable for limited purposes; and (4) establish a multi-stakeholder process to determine the best way to approach ISPs' use of data that are sensitive or out of context, taking into account degrees of identifiability.