REVERSING PRIVACY RISKS:

STRICT LIMITATIONS ON THE USE OF COMMUNICATIONS METADATA AND TELEMETRY INFORMATION

SUSAN LANDAU AND PATRICIA VARGAS LEON*

* Susan Landau, Bridge Professor in Cyber Security and Policy, The Fletcher School and School of Engineering, Department of Computer Science, Tufts University. Patricia Vargas Leon, Cybersecurity Policy Postdoctoral Research Fellow, The Fletcher School, Tufts University. We would like to particularly thank Fred Cate, Steven M. Bellovin, Jim Dempsey, Paul Ohm, Stephanie Pell, Sarah Radway, and John Treichler for a number of insightful and valuable comments. We also wish to thank participants of the 2022 Annual Privacy Law Scholars Conference, whose suggestions on an earlier draft were immensely useful. This research was supported in part by funding from the National Science Foundation grants CNS 1923528 and CNS 1955805 and the William and Flora Hewlett Foundation under grants 2018-7277 and 2020-1484.

INTRODUCTION

When Assistant U.S. Attorney Patrick Fitzgerald was investigating the 1993 World Trade Center bombing in the mid-1990s, he sought to connect the actions of various suspects. "I would find defense attorneys who would say Mr. X and Mr. Y didn't know each other or never talked… and I would spend my evenings with phone bills trying to prove they must have talked, they must have talked lots of times."[1]

Though the phone records did not provide call content, they nonetheless provided invaluable information on the workings of the conspiracy. Fitzgerald later described, "You figured out that the one person that everyone agrees is a terrorist—maybe he's already convicted (he's not on trial)—bought urea nitrate. And you show that just before he went to buy the urea nitrate, he called the defendant. Then you said he went and bought explosive detonators, and you showed a phone record; he called the defendant. Then he went and did a surveillance; then he called the defendant."[2]

Uncovering such details required Fitzgerald to examine voluminous piles of paper phone records and required hundreds of hours. When Fitzgerald conducted this investigation, he used three different sets of phone records that the FBI supplied. The investigator spent his evenings comparing the different records by hand.[3]

A quarter of a century later, these records and the tools for comparing them are digital. The months of evenings that Fitzgerald

1. Palantir, *The Evolving Role of Technology in the Work of Leading Investigators and Prosecutors*, YOUTUBE (June 12, 2013), https://www.youtube.com/watch?feature=player_embedded&v=Nd2fZZhxuzQ [https://perma.cc/VK36-SMBH] (interview with Patrick Fitzgerald, 4:48–5:00).

2. *Id.* at 7:52–8:12.

3. *See generally* SUSAN LANDAU, LISTENING IN: CYBERSECURITY IN AN INSECURE AGE 200–01 (2017).

spent studying the records can be reduced to seconds using a simple computer command. Indeed, the type of work that Fitzgerald did forms the backbone of the surveillance reports provided by any number of private companies.[4] The ease of examination and the amount of data to study have increased exponentially, greatly changing the way law enforcement and national security investigators conduct their work.

Dropping costs in storage, search, and the amount of non-content data available illuminates one aspect of the increasing value of searching such data.[5] There are many other changes surrounding the use of the "non-content" part of communications and related device usage.

Cellphones caused the first change. At the time Fitzgerald began his investigations, fewer than 2 percent of the world's population had a cellular phone subscription[6] (numbers were higher in Europe—in Germany, it was 4.6 percent,[7] in the U.K, just under 10 percent[8]—while in the U.S., the number was just under 13 percent).[9] When a person walked around a major U.S. city in the early 1990s, they found public pay phones everywhere.[10] Very few such public pay phones remain.[11] They have been replaced by individuals' mobile phones and, occasionally, free broadband.[12] Our

4. For an example of backlash to the ease by which industry could create these reports even as early as 2005, *see e.g.*, Chris Jay Hoofnagle, Elec. Priv. Info. Ctr., Petition of the Electronic Privacy Information Center for Rulemaking to Enhance Security and Authentication Standards for Access to Customer Proprietary Network Information 1 (Aug. 30, 2005), https://archive.epic.org/privacy/iei/cpnipet.html [https://perma.cc/RJY4-4MXW].

5. *See generally Historical Cost of Computer Memory and Storage*, OUR WORLD IN DATA, https://ourworldindata.org/grapher/historical-cost-of-computer-memory-and-storage?country=~OWID_WRL [https://perma.cc/X47P-UN5Y] (dropping costs in storage mean lower costs for search).

6. *Mobile Cellular Subscriptions (per 100 People)*, WORLD BANK, https://data.worldbank.org/indicator/IT.CEL.SETS.P2 [https://perma.cc/YV6E-9JTH].

7. *Id.* (click into the top search bar, type "Germany," then press enter).

8. *Id.* (click into the top search bar, type "United Kingdom," then press enter).

9. *Id.* (click into the top search bar, type "United States," then press enter).

10. Ann Chen & Aaron Reiss, *Looking Back at the Pay Phone's New York Heyday*, N.Y. TIMES (May 27, 2022), (describing how some pay phones were arrayed in "pay phone banks": a row of phones with little privacy between them. These were common in such places as Manhattan's Penn Station and busy street corners.).

11. Ann Chen & Aaron Reiss, *The Only Living Pay Phones in New York*, N.Y. TIMES (May 27, 2022), https://www.nytimes.com/2022/05/27/arts/pay-phones-nyc.html [https://perma.cc/TF7B-YSZT].

12. Some cities have replaced the pay phones with a "free" broadband service. However, the "free" service requires registration; use of pay phones did not. The broadband alternative in New York city is LinkNYC. *See* Keith Williams, *What Are These Tall Kiosks That Have Replaced Pay Phones in New York?*, N.Y. TIMES (Jan. 11, 2018), https://www.nytimes.com/2018/01/11/nyregion/what-are-those-tall-kiosks-that-have-replaced-pay-phones-in-new-york.html [https://perma.cc/3FAP-H53H]. However, the

embrace of the Internet, smartphones,[13] and apps has had a profound effect on society. Local businesses lost out to Internet providers, hollowing out Main Street; social media replaced the press as a source of news; and advertising became highly targeted.

One of the biggest impacts of new communication technologies was on user privacy. Four innovations have driven this change.

The first, digitization of phone records, made searching communications metadata—who called whom when and for how long—easy and fast. This change happened in the 1960s.[14]

The second occurred through the U.S. public's move to cellphones. Suddenly, telephone service providers automatically received information about users' locations whenever a user's phone was turned on.[15] Cell Site Location Information (CSLI) provides a rough approximation of a phone's location.[16] The Federal Communications Commission adopted a requirement that carriers be able to locate Enhanced 911 (E911) callers within 410 feet of the user's location.[17] Although there are various ways to determine location, including through the network, this requirement ended up driving adoption of chips for reading Global Positioning System (GPS) signals,[18] which quickly became standard issue for mobile phones. The result was precise location data being available on the phones themselves.

rollout of LinkNYC is minimal in other cities. *See, e.g.*, Victor Fiorillo, *Here's Where to Get Free Wi-Fi in Philly*, PHILADELPHIA (Feb. 18, 2020, 10:10 AM), https://www.phillymag.com/news/2020/02/18/free-wifi-in-philly/ [https://perma.cc/6MSZ-C34F].

13. We use the term "cellphones" to denote mobile phones that function purely as phones and have no computing capability, "smartphones" to denote mobile phones that function both as a mobile phone operating over the Public Switched Telephone Network but that also have computing capabilities, and "mobile phone" to encompass both cellphones and smartphones.

14. A.E. JOEL, JR. ET AL., A HISTORY OF ENGINEERING AND SCIENCE IN THE BELL SYSTEM: SWITCHING TECHNOLOGY (1925-1975) 141–42 (G. E. Schindler, Jr., ed., 1982).

15. *See, e.g.*, *ECPA Reform and the Revolution in Location Based Technologies and Services, Hearing Before the Subcomm. on the Const., C.R., and C.L. of the H. Comm. on the Judiciary*, 111th Cong. 13–15 (2010) (statement of Matt Blaze) (as the number of mobile phones increased, so did the number of cell towers and, thus, the capability for determining location from cell towers). In the U.S., cellphones typically belong to a single individual.

16. *Id.*

17. Revision of the Commission's Rules to Ensure Compatibility with Enhanced 911 Emergency Calling Systems, Third Report and Order, 14 FCC Rcd. 17388, 17391 (Sept. 15, 1999) (previously codified at 47 C.F.R. §§ 20.1–20.23).

18. *See, e.g.*, Kevin E. McCarthy, Conn. Gen. Assemb. Off. Legis. Rsch., 2004-R-0461, *Searching for Callers Using GPS* (2004), https://www.cga.ct.gov/2004/rpt/2004-R-0461.htm [https://perma.cc/LUR7-XGWS] (originally, Nextel, Sprint, and Verizon chose GPS, while AT&T, T-Mobile, and Cingular opted for a network solution).

The third change came from the public's shift to Internet communications,[19] which began during the same period as the shift to cellphones.[20] This shift was greatly augmented by a consumer move to smartphones. The use of smartphones, which are computers as well as phones, rose from 35 percent of the U.S. population in 2011 to 85 percent by 2021.[21] The smartphone ad industry arose alongside these devices. Relying on "behavioral advertising,"[22] this industry targets advertisements based on user behavior.[23]

In the early days of the web, advertisers relied on cookies, short text files that reside on a user's device that contain a record of the user.[24] Smartphones made far more relevant information about users available to advertisers. The website or app could learn user location through the phone's current IP address or GPS coordinates and discern information about user income from a combination of learning where the user most likely lived and what device—e.g., iPhone or Android—they have.

The fourth change arose from telemetry, streaming data from user devices providing measurements of device functioning and user activity.[25] Measuring the performance of device software provides operating system (OS) and device manufacturers insight

19. *See* discussion *infra* Section II.C.

20. *See Share of the Population Using the Internet*, OUR WORLD IN DATA, https://ourworldindata.org/grapher/share-of-individuals-using-the-internet [https://perma.cc/2EUW-R2CL] (the North Americans public's use of the Internet went from near 0 percent in 1990 to 44 percent by 2000); *see also Historical Timeline of Cellular Telecommunications*, UNDISTRACTED DRIVING ADVOC., http://undistracteddrivingadvocacy.net/historical-timeline-of-cellular-telecommunications/ [https://perma.cc/3H4A-JFNQ] (the U.S. use of mobile phones went from 5 million in 1990 to over 100 million in 2000).

21. *See Mobile Fact Sheet*, PEW RSCH. CTR. (Apr. 7, 2021), https://www.pewresearch.org/internet/fact-sheet/mobile/ [https://perma.cc/453V-FK7C] (U.S. smartphone ownership jumped from 35 percent in 2011 to 85 percent by 2021).

22. *See* discussion *infra* Section III (earlier definitions of behavioral advertising state that is based on web browsing behavior, but, as we shall see, behavioral advertising relies on other signals as well).

23. *See generally* JOSEPH TUROW, THE AISLES HAVE EYES: HOW RETAILERS TRACK YOUR SHOPPING, STRIP YOUR PRIVACY, AND DEFINE YOUR POWER (2017); *see generally* SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

24. Cookies may contain records of a user's activities at a site, a user's authentication information, etc.

25. According to the Oxford English Dictionary, telemetry is "the science or practice of obtaining physical measurements or other data at one place and transmitting them (chiefly by means of electrical signals or, in later use, radio waves) to another place for display or recording" (definition 2a) and "data obtained using telemetry" (definition 2c). *Telemetry*, OXFORD ENG. DICTIONARY (3rd ed. 2016).); *see* discussion *infra* Section II.D.

on how well the devices are working.[26] An OS manufacturer wants to know, for example, what actions a user took just before the device crashed and whether users are paying attention to device notifications. OS providers and app developers also seek to understand how users interact with tools in order to improve user experience.[27]

Because smartphones are not just phones, but also computers, the devices furnish a rich source of information about device operations.[28] Smartphone telemetry provides information beyond how the software is functioning, and device sensors also play a large role. Smartphone sensors are intended to improve the device's functionality (accelerometers, gyroscopes), preserve its battery life (power sensors, ambient light sensors), and handle input (touch sensors, proximity sensors).[29] Thus, gyroscopes enable correct orientation of a webpage or photo, magnetometers enable mapping capabilities, proximity sensors prevent a user from inadvertently pressing the touch pad when holding a phone to their ear, etc.[30] While sensors serve purposes on the device, the information they collect can also be employed off the device.

Such information is often used in other ways that differ substantially from the data's original purpose.[31] A website or app might learn how a user employs a device, what mode of transportation they are using, and mouse movements.[32] Even if users turn off providing GPS location information, device telemetry can reveal that two users spent the night in each other's company.[33] Telemetry can reveal whether a user is on a bike ride or at a concert—and allow a provider to determine whether to post a text to the user—or wait until later.[34] Some people may find this useful,

26. *See e.g.*, Titus Barik et al., *The Bones of the System: A Case Study of Logging and Telemetry at Microsoft*, 2016 IEEE/ACM38TH INT'L CONF. ON SOFTWARE ENG'G COMPANION at 92.

27. *See id.* at 98 (Are they using a tool? Are they abandoning it for another one?).

28. *See* discussion *infra* Section IV (this is through the revelatory data provided by user searches, location, activities on the phone, etc.).

29. David Nield, *All the Sensors in Your Smartphone, and How They Work*, GIZMODO (June 29, 2020), https://gizmodo.com/all-the-sensors-in-your-smartphone-and-how-they-work-1797121002 [https://perma.cc/F7QX-S3RZ].

30. *See id.*

31. *See* discussion *infra* Section IV.C.

32. *See* discussion *infra* Sections II.D, III.

33. *See* discussion *infra* Sections III.A, IV.B, IV.C (this can be done by knowing the initial location of a user and then their movement based on information provided by the accelerometer, gyroscope, and magnetometer).

34. *See generally* Methods and Systems for Providing Notifications Based on User Activity Data, U.S. Patent No. 2016/0037482 A1 (filed July 29, 2014) (issued Jan. 24, 2017).

others, highly intrusive. Some sites even do key logging.[35] The Mayo Clinic, a premier U.S. medical center, tracks people's symptoms as they fill in an appointment form—and the Clinic has this highly personal data even if the person decides not to submit a form.[36] From the user's point of view, such uses can be quite unexpected.

This data is typically transmitted to other sites; these could be data processors that do processing for the original site (and do not otherwise interact with the data) and data controllers, which may also use the data for their own purposes, such as ad marketing or processing the data and selling it to other third parties. Data can be shared through explicit data-sharing agreements between websites, apps, data processors, and controllers or as the result of the use of Software Developer Kits (SDKs) to build apps.[37] SDKs simplify app development, but part of the arrangement is that they also transfer data to the SDK provider.[38] One striking aspect of SDKs is that, even if a user does not have an account at a particular company, say Facebook, if a developer uses a Facebook SDK to build an app, the user's data will be sent to Facebook.[39]

Researchers found that, despite the strict restrictions of the European Union's (EU) General Data Protection Regulation (GDPR), 40 percent of apps sent out personal data such as the Android Advertising ID, an email address, user location, and mobile device identification number (IMEI) without first obtaining user consent.[40] Providing such information enables the data controllers to conduct invasive user tracking.

35. To be clear, we are discussing key logging in the situation the user is knowingly on a site and the site itself is doing the logging but doing so before the user has pressed a "submit" button; we do not mean the situation where a different entity is conducting the logging—and thus hacking.

36. Aaron Sankin & Surya Mattu, *The High Privacy Cost of a "Free" Website*, MARKUP (Sept. 22, 2020, 6:00 AM), https://themarkup.org/blacklight/2020/09/22/blacklight-tracking-advertisers-digital-privacy-sensitive-websites [https://perma.cc/HGQ2-6NGW]. According to Blacklight, the tool developed by Surya Mattu, Weight Watchers uses "a session recorder, which tracks user mouse movement, clicks, taps, scrolls, or even network activity." *Id.* (in the search bar, type in "weightwatchers.com/us" and click "scan site"; once the analysis is complete, click the dropdown icon next to "This website could be monitoring your keystrokes and mouse clicks").

37. *See, e.g.*, Charlie Warzel, *The Loophole that Turns Your Apps into Spies*, N.Y. TIMES OP. (Sept. 24, 2019), https://www.nytimes.com/2019/09/24/opinion/facebook-google-apps-data.html [https://perma.cc/7QK6-HJ6S] (discussing the sharing of data as a result of using SDKs).

38. *See* Sankin & Mattu, *supra* note 36; *see also* Warzel, *supra* note 37.

39. *See* Warzel, *supra* note 37.

40. Trung Tin Nguyen et al., *Share First, Ask Later (or Never?) Studying Violations of GDPR's Explicit Consent in Android Apps*, PROC. 30TH USENIX SEC. SYMP. 3667, 3673

For a long time, U.S. laws and policy failed to address the issue of the personal nature of non-content data. This is now changing. In 2016, Federal Trade Commission (FTC) Chairwoman Edith Ramirez announced:

> As a result of the increased ability to identify consumers, we now regard data as personally identifiable when it can *be reasonably linked* to a particular person, computer, or device. In many cases, persistent identifiers, such as device identifiers, MAC addresses, static IP addresses, and retail loyalty card numbers meet this test.[41]

In 2018, in *Carpenter v. United States* ruled the Supreme Court that warrantless collection of seven days of CSLI violated Fourth Amendment rights.[42] Five states, California,[43] Colorado,[44] Connecticut,[45] Utah,[46] and Virginia,[47] have adopted privacy laws with a broad definition of personal information that includes metadata. Yet, legal controls on the use of metadata lags.[48] The situation with telemetry is even worse, with essentially no legal attention yet paid to its use. In this paper, we examine private-

(combining the Android Advertising ID (AAID) with persistent IDs, such as the device identification number, is in violation of both GDPR *and* Google's policies).

41. Edith Ramirez, Chairwoman, FTC, Keynote Address at the Technology Policy Institute Aspen Forum: Protecting Consumer Privacy in the Digital Age: Reaffirming the Role of Consumer Control 3–4 (Aug. 22, 2016) https://www.ftc.gov/system/files/documents/public_statements/980623/ramirez_-_protecting_consumer_privacy_in_digital_age_aspen_8-22-16.pdf [https://perma.cc/EKZ6-PXGR].

42. Carpenter v. United States, 138 S. Ct. 2206, 2221, 2223 (2018).

43. California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.140 (West 2023) (subsection (*o*)(1) states the following definition: "'Personal information' means information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household.").

44. Colorado Privacy Act, COLO. REV. STAT. § 6-1-1303 (2023) (subsection (17) states the following definition: "'Personal data': (a) Means information that is linked or reasonably linkable to an identified or identifiable individual").

45. Connecticut Data Privacy Act, Conn. Legis. Serv. P.A. 22-15, § 1 (subsection (18) states the following definition: "'Personal data' means any information that is linked or reasonably linkable to an identified or identifiable individual.").

46. Utah Consumer Privacy Act, UTAH CODE ANN. § 13-61-101 (West 2022) (subsection (24)(a) states the following definition: "'Personal data' means any information that is linked or reasonably linkable to an identified or identifiable individual.").

47. Virginia Consumer Data Protection Act, VA. CODE ANN. § 59.1-575 (2023) (States the following definition: "'Personal data' means any information that is linked or reasonably linkable to an identified or identifiable natural person.").

48. *See generally* American Data Privacy and Protection Act, H.R. 8152, 117th Cong. (2022) (as reported by H. Comm. on Energy and Com., Dec. 30, 2022) (proposing legislation that would remedy this problem).

sector collection and use of metadata and telemetry information and provide three main contributions:

First, we lay out the extent to which "non-content"—the hidden parts of Internet communications (aspects the user does not explicitly enter) and telemetry—are highly revelatory of personal behavior. We show that, privacy policies notwithstanding, users rarely know that the metadata and telemetry information is being collected and almost never know the uses to which it is being put.

Second, we show that consumers, even if they knew the uses to which this type of personal information were being put, lack effective means to control the use of this type of data. The standard tool of notice-and-choice has well known problems, including the user's lack of information with which to make a choice; and then, even if the user had sufficient information, doing so is not practical.[49] These are greatly exacerbated by the nature of the interchanges for communications metadata and telemetry information. Each new transmission—each click on an internal link on a webpage, for example—may carry different implications for a user in terms of privacy. The current regimen, notice-and-choice, presents a completely unworkable set of requests for a user, who could well be responding many times a minute regarding whether to allow the use of metadata beyond the purposes of content delivery and display. This is especially the case for telemetry, where the ability to understand both present and future use of the data provided from the sensors requires a deeper understanding of what information these devices can provide than anyone but a trained engineer would know.

Third, while there has been academic and industry research on telemetry's use, there has been little exploration of the policy and legal implications stemming from that use. We provide this factor, while at the same time addressing the closely related issues raised by industry's use of communications metadata to track user interests and behavior.

We begin in Section I by exploring the changing history of the collection of metadata and telemetry information. We carefully examine how and why the public switched telephone network (PSTN) acquired this data and the purposes to which the information was put. To set the stage for how the private sector handles metadata and telemetry, we take a brief look at the legal

49. *See, e.g.*, Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: J.L. & POL'Y FOR INFO. SOC'Y 543, 565 (2008) (using mathematical modeling to conclude that it would cost the average user the equivalent of $3,534 per year to read every privacy policy they encountered for a total yearly economic loss of $781 billion).

arguments behind the different jurisprudence governing U.S. Government acquisition and use of content versus communications metadata for law enforcement and national-security investigations. We follow that overview with a concise discussion of how Internet Protocol (IP)-based communications are effectively eviscerating the content/non-content distinction. Our focus, based on the work by Bellovin et al.,[50] is from a technical vantage point of how the complexities of IP-based communications rendered the content/non-content "functionally meaningless."[51]

We then turn to software and device telemetry. Software telemetry can help an app or OS provider debug the system when problems occur; they can also inform the app or OS provider exactly how a user is interacting with the application (e.g., did she continue to checkout?). Sensors can, for example, reveal where the user is going even if she has shut off GPS.[52] They can show who is traveling with her by revealing the changing SSIDs in her immediate locale. Sensors can even indicate how concerned the user is about getting an Uber ride (her battery is getting low).[53] That such information—both software and sensor data—is shared is relatively new, as indeed is the fact that the information exists at all.

With that background, we are ready to explore the different types of personal information that metadata and telemetry can reveal. This, we do in Section II, dividing the data into three categories: location, revelatory data about population groups, and revelatory data about individuals.

Section III shows how far current practices have strayed from the original intent of the Fair Information Practices (FIPs). We begin by studying the original intent of the U.S. privacy principles and their development in practice by exploring how that vision was

50. Steven M. Bellovin et al., *It's Too Complicated: How the Internet Upends Katz, Smith, and Electronic Surveillance Law*, 30 HARV. J.L. & TECH. 1, 5 (2016).

51. *Id.*; *see* Orin S. Kerr, *Internet Surveillance Law After the USA Patriot Act: The Big Brother That Isn't*, 97 NW. U.L. REV. 607, 630–31 (2003); David McPhie, *Almost Private: Pen Registers, Packet Sniffers, and Privacy at the Margin*, 2005 STAN. TECH. L. REV. 1, 4–5 (2005); Matthew J. Tokson, *The Content/Envelope Distinction in Internet Law*, 50 WM. & MARY L. REV. 2105, 2124–25 (2009).

52. *See, e.g.*, Sashank Narain et al., *Inferring User Routes and Locations Using Zero-Permission Mobile Sensors*, 2016 IEEE SYMP. ON SEC. & PRIV. 397, 398.

53. Shankar Vedantam & Maggie Penman, *This Is Your Brain on Uber*, NPR (May 17, 2016, 12:01 AM), https://www.npr.org/2016/05/17/478266839/this-is-your-brain-on-uber [https://perma.cc/7FWB-FZ3Y] (summarizing interview with Keith Chen, Uber's Head of Economic Research, who noted that the Uber app had access to users' battery level data and that riders were more likely to accept surge pricing when their batteries were low—although he maintained that Uber does not exploit this data. There have been claims otherwise; *see* Jessica Lindsay, *Does Uber Charge More if Your Battery Is Lower*, METRO (Sept. 27, 2019, 1:19 PM), https://metro.co.uk/2019/09/27/uber-charge-battery-lower-10778303/ [https://perma.cc/KA7H-HTXE]).

operationalized through the FIPs.[54] We use Julie Cohen's work on the centrality of choice[55] to frame our thinking about how to put privacy principles into practice. We then consider Paul Ohm's insightful 2009 critique of ISP's deep packet inspection[56] and contrast it with current practices of Internet companies' use of metadata and telemetry information. These can include minute-by-minute indications of a person's location,[57] activities,[58] and companions.[59] Due to vertical integration and consolidation in the telecommunications industry, current usage of metadata and telemetry[60] is notably more invasive than the situation described by Ohm over a decade ago.

We, next, turn to examining the intent of privacy protections. We start with Alan Westin's powerful analysis of the role of privacy within a liberal democratic society,[61] taking particular note of Lisa Austin's research demonstrating that Westin's vision meant enabling meaningful privacy choices.[62] We end by showing the impossibility of (1) meaningful privacy choices and (2) providing informed consent regarding the use of communications metadata and software and device telemetry information.

Section IV proposes possible directions for change. We begin with the failure of current legal remedies for protecting users against invasive use of metadata and telemetry. Communications

54. *Fair Information Practice Principles*, INT'L ASS'N PRIV. PROS., https://iapp.org/resources/article/fair-information-practices/ [https://perma.cc/XA8C-LJ7W].

55. *See generally* Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373 (2000).

56. *See generally* Paul Ohm, *The Rise and Fall of Invasive ISP Surveillance*, 5 U. ILL. L. REV. 1417, 1462–74 (2009).

57. Yves-Alexandre de Montjoye et al., *Unique in the Crowd: The Privacy Bounds of Human Mobility*, SCI. REPS., 3:1376, Mar. 25, 2013, at 1, 1; *see* discussion *infra* Section IV.A.

58. Noah Apthorpe et al., *Keeping the Smart Home Private with Smart(er) IoT Traffic Shaping*, 2019 PROC. ON PRIV. ENHANCING TECH. 128, 128, 143; Jonathan Mayer et al., *Evaluating the Privacy Properties of Telephone Metadata*, 113 PROC. NAT'L ACAD. SCIS. 5536, 5540–41 (2016); Routine Deviation Notification, U.S. Patent No. US 2016/0316341 A1, at [57] (filed July 6, 2016) (issued Oct. 27, 2016); *see* discussion *infra* Section IV A.

59. Identifying and Locating Users on a Mobile Network, U.S. Patent No. 2017/0026796 A1, at [57] (filed July 25, 2016) (issued Jan. 26, 2017); *see also* discussion *infra* Section IV.A.

60. FTC, A LOOK AT WHAT ISPS KNOW ABOUT YOU: EXAMINING THE PRIVACY PRACTICES OF SIX MAJOR INTERNET PROVIDERS iv (2021), https://www.ftc.gov/system/files/documents/reports/look-what-isps-know-about-you-examining-privacy-practices-six-major-internet-service-providers/p195402_isp_6b_staff_report.pdf [https://perma.cc/3AZW-BAVR].

61. *See generally* ALAN F. WESTIN, PRIVACY AND FREEDOM (1967).

62. Lisa Austin, *Re-Reading Westin*, 20 THEORETICAL INQUIRIES L. 53, 73 (2019).

metadata and telemetry information are necessary for the delivery and/or display of requested information and the systems' proper functioning, but the technology sector's use of metadata and telemetry for purposes outside of these is what Daniel Solove and Woodrow Hartzog have aptly called "broken expectations" of privacy.[63] Current laws and policies have failed to handle the problems posed by these broken expectations.

We consider possible remedies, starting with Solove's proposals for handling the failure of consent—proposals of nudges, partial self-management, adjusting timing and focus, and "moving towards substance."[64] Because metadata and telemetry requests for use are likely to be frequent, numerous, and with ultimate use of the information opaque to all but a highly expert user, Solove recommendations are unlikely to solve the problem.

We propose a modification of Fred Cate's work on controlling use.[65] We propose a strict Purpose Limitation Principle for metadata and telemetry information in which such data is to be used exclusively for the explicit purposes for which it was collected: delivery and/or display of requested information, ensuring the system is working properly, investigating fraud, and projecting future customer needs. To that, we add exceptions to use aggregated data in cases of public health emergencies and to conduct peer-reviewed studies in the public interest.[66] No other uses would be permitted.

We propose two potential routes for enacting these protections. One is a regulatory approach. Use of metadata and telemetry

63. Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 667–69 (2014).

64. Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1800, 1900–03 (2013).

65. *See generally* Fred H. Cate, *The Failure of Fair Information Practice Principles*, *in* CONSUMER PROTECTION IN THE AGE OF THE 'INFORMATION ECONOMY' 341 (Jane K. Winn & Geraint Howells eds., 2006), *and* Fred H. Cate & Viktor Mayer-Schönberger, *Tomorrow's Privacy: Notice and Consent in a World of Big Data*, 3 INT'L DATA PRIV. L. 67 (2013); *but see* Susan Landau, *Control Use to Protect Privacy*, 347 SCIENCE 504, 504 [hereinafter *Control Use*] (commenting that notice and choice is not always functional in practice), *and* Joel R. Reidenberg et al., *Privacy Harms and the Effectiveness of the Notice and Choice Framework*, 11 I/S: J. L. & POL'Y FOR INFO. AGE 485, 488 (2015); *see also* Steven M. Bellovin, Comment Letter on Developing the Administration's Approach to Consumer Privacy, Privacy Docket No. 180821780-8780-01 (Nov. 7, 2018), https://www.ntia.doc.gov/files/ntia/publications/ntia-privacy.pdf [https://perma.cc/ET2A-6F66], *and* Susan Landau, Comments on Developing the Administration's Approach to Consumer Privacy, Privacy Docket No. 180821780-8780-01 (Oct. 24, 2018) [hereinafter *Landau Privacy Comments*], https://ntia.gov/sites/default/files/publications/comments_to_ntia_0.pdf [https://perma.cc/C8JA-P5ZQ].

66. Our last proposed exception is modeled on subsection 101(b)(10)(A) of the proposed "American Data Privacy and Protection Act." *See* H.R. 8152 § 101(b)(10)(A).

beyond the purpose of delivering content and ensuring quality of service constitutes a deceptive practice and unfair use, providing a potential for Federal Trade Commission action. The other is a legislative solution, which would result in stronger protections. We briefly summarize our arguments in the conclusion.

I. THE PAST IS PROLOGUE: FROM CDRS TO PACKET HEADERS AND TELEMETRY

As communications technologies changed over the last three decades, the transactional data to enable these services became more extensive. This change proved of great benefit to law-enforcement and national-security investigations. In recent years, the fact that people's real-time locations were trackable through the device in their pocket—their cellphone—was a particular boon to government investigators. For a long time, such "non-content" data was not deemed worthy of strong legal protections; investigators could learn an individual's location without needing a search warrant.[67] In the 1979 case *Smith v. Maryland*, the Supreme Court ruled that installation of a pen register, which collects all the numbers dialed from a particular phone, did not violate a person's "reasonable expectation of privacy."[68]

In 2018, the U.S. Supreme Court redefined *Smith's* third-party doctrine in a rather striking fashion. On the surface, the Court ruled in *Carpenter v. United States* that law enforcement acquisition of seven days of Cell Site Location Information (CSLI), data routinely collected by service providers, required a search warrant.[69] The Court's described the "deeply revealing nature of CSLI, its depth, breadth, and comprehensive reach, and the inescapable and automatic nature of its collection, the fact that such information is gathered by a third party does not make it any less deserving of Fourth Amendment protection."[70] Thus, the *Carpenter* decision carries a deeper significance than the simple holding that seven days of CSLI records requires a search warrant. As Paul Ohm put it, "[u]ntil now, the Supreme Court has tended to pay more attention to the nature of the police intrusion required to obtain information than to the nature of the information obtained."[71] That focus changed with *Carpenter*. Ohm presented a strong argument that in future, records of websites visited and

67. *See* discussion *infra* Section II.B.i.

68. Smith v. Maryland, 442 U.S. 735, 745–46 (1979).

69. Carpenter v. United States, 138 S. Ct. 2206, 2221, 2223 (2018).

70. *Id.* at 2223.

71. Paul Ohm, *The Many Revolutions of Carpenter*, 32 HARV. J.L. & TECH. 358, 362 (2019).

massive collections of bank and phone records will fall under the same warrant requirements as CSLI.[72] Many more types of data collected by third parties—or by the government—are also likely to fall under the same protections *Carpenter* provides to seven-days' worth of CSLI records.[73]

As momentous as *Carpenter* is, the decision addresses the tip of a very large iceberg. The private sector has also taken advantage of the great wealth of information provided by metadata and telemetry information. This is not new; the telecommunications industry used call detail records (CDRs) for over a half century for such purposes as discovering individual preferences for pricing, customization, recommendation services, in addition to retaining customers and obtaining new ones.[74] But, over the last two decades, there has been a massive change in scale in the collection and use of metadata and telemetry information—data over which users have almost no knowledge of collection and even less control over use.

Consider, for example, that from its start, Google logged user data to improve the company's services.[75] In 2005, data collected included originating IP address, browser type, browser language, time and date of the query, search terms, and how long a user spent on each site.[76] Clocking how long users spent on a search link before trying another allowed the company to determine which responses users found most useful.[77] IP addresses helped in providing the right responses, which could vary by location[78] (e.g., for providing

72. *Id.* at 378–81.

73. *Id.* at 392 (noting the records collected by state-owned Automated License Plate Readers could fall under the protections as well).

74. A.E. JOEL, JR. ET AL., *supra* note 14, at 92, 106–7, 400.

75. Sergey Brin & Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, 30 COMPUT. NETWORKS & ISDN SYS. 107, 115 (1998); *Google Privacy FAQ*, GOOGLE (Oct. 15, 2005), https://web.archive.org/web/20051015025545/http://www.google.com/intl/en/privacy_faq.html [https://perma.cc/MM2F-2LG3] ("[O]ur servers automatically record the page requests made when users visit our sites. These 'server logs' typically include your web request, Internet Protocol address, browser type, browser language, the date and time of your request and one or more cookies that may uniquely identify your browser.").

76. GOOGLE, *supra* note 75.

77. *See, e.g.*, *Session Duration, Avg*, GOOGLE, https://support.google.com/analytics/answer/1006253?hl=en [https://perma.cc/HD7D-BK7G].

78. *What Is IP Address Geolocation and How to Change It?*, HEFICED, (May 29, 2019), https://www.heficed.com/blog/articles/what-is-ip-address-geolocation-and-how-to-change-it/#:~:text=IP%20geolocation%20is%20the%20mapping,area%20code%2C%20and%20other%20information [https://perma.cc/52G8-F7NA] (IP addresses provide a reasonable, though not always accurate, determination of location. Internet service providers allocate IP addresses to "points of presence" serving particular geographic areas. This

store location for chain stores, for language selection. For example, when a U.K. user searches for biscuit recipes, she would be looking for a cookie recipe, while a U.S. user would want to a recipe for a scone-size quick bread, etc.). As early as 2004, Google's holy grail was personalized responses.[79] Storing IP addresses helps in providing personalization; if yesterday a user searched for "hatchbacks with four-wheel drive," while today her query is for "four-wheel drive cars," then hatchback cars might appear higher in Google's search response than otherwise.

Sometimes, Internet companies' use of metadata and telemetry has clear benefits for the user. For example, on seeing a login to an account from a new device, Google sends a message to a user asking, "Did you just sign in?"[80] Such responses help prevent fraudulent access to accounts. Other benefits can include studying how customer use systems (e.g., a phone's home screen) to provide better functional configurations for the user.

Apps complicate what constitutes communications content. As Albert Gidari observed during 2010 Congressional testimony: "In the case of many location-based services ("LBS"), some logging of a user's location may occur and be retained. In many such applications, the user is conveying his or her location to another user essentially as a communication—'here I am.'"[81] The communication is functioning much like a text message— "I'm five minutes away." But it is not explicitly presented to the user as a conversation.

Use of the metadata and telemetry information can result in release of information that users would prefer to keep private. Researchers have shown that communications metadata can disclose whether a user has participated in a particular political protest,[82] expose a user's sleeping patterns,[83] or reveal their gender, marital status, income, and education. Categories of

form of location identification can be off by some distance—and can also be incorrect for other reasons (e.g., the owner may have sold a sub-block of addresses and this change has not yet shown up in the registry, a user might be employing a VPN service, etc.).

79. *Google Introduces Personalized Search Services; Site Enhancements Emphasize Efficiency*, GOOGLE (Mar. 29, 2004), http://googlepress.blogspot.com/2004/03/google-introduces-personalized-search.html [https://perma.cc/PPR7-NPJE].

80. *Protect Your Account if There Is Unfamiliar Activity*, GOOGLE, https://support.google.com/accounts/answer/7305876?hl=en [https://perma.cc/P4U4-E8VW].

81. *Electronic Communications Privacy Act Reform: Hearing Before the H. Subcomm. on the Const., C.R., and C.L. of the H. Comm. on The Judiciary*, 111th Cong. 30 (2010) (written testimony of Albert Gidari, Partner, Perkins Coie LLP).

82. Adrian Dobra et al., *Spatiotemporal Detection of Unusual Human Population Behavior Using Mobile Phone Data*, 10 PLOS ONE, Mar. 2015, at 14.

83. *Beddit Sleep Monitor*, BEDDIT, http://www.beddit.com [https://perma.cc/G9E4-WR37].

information that metadata and telemetry information can reveal include information about social characteristics of groups in which the user may be a member—order of battle of a military organization, "community of interest" for a terrorist, organizational structure of a corporation, community characteristics such as religion and income level, societal movement such as commuting patterns or response to a natural disaster—and about individuals—identifying devices, device activity, device or application user, and personality characteristics about the user.[84] Device telemetry data—data that monitors "Global Positioning System (GPS), Wi-Fi signal, radio signal modulation, or geo-tagging"—might indicate that two people are in close proximity, while "data from at least one of a gyroscope, an accelerometer, or a motion processor of the computing system" might reveal that the users are in the same bus, car, or other vehicle.[85]

There have been some attempts to regulate the collection and use of metadata. One early effort involved the EU stopping search engine retention of user IP addresses. After data commissioners became aware of how easy it was to identify users from linking their search queries,[86] in 2006, data commissioners passed a resolution that search engines should not enable such linking without explicit user permission.[87] A more official response occurred in 2008, when an EU advisory board on data protection, the Article 29 Working Group, stated, "[i]f personal data are stored, the retention period should be no longer than necessary for the specific purposes of the

84. Susan Landau, *Categorizing Uses of Communications Metadata: Systematizing Knowledge and Providing a Path for Privacy*, NEW SEC. PARADIGMS WORKSHOP 2020, Oct. 2020, at 8.

85. Systems and Methods for Utilizing Wireless Communications to Suggest Connections for a User, U.S. Patent No. 2017/0359711 A1, col. 1 l. 18–19 (filed Aug. 28, 2017).

86. Michael Barbaro & Tom Zeller Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES (Aug. 9, 2006), https://www.nytimes.com/2006/08/09/technology/09aol.html [https://perma.cc/E7N8-9CAF] (This discovery came about because in 2006 AOL had released three months of "anonymized" search queries, but these anonymized searches were quickly deanonymized. AOL had provided an anonymous ID number for each user, but the searches themselves were so personally revelatory that in many instances the user's identity could nonetheless be determined. That was not the only incident that precipitated concern over retention of IP addresses and other user information in search logs.).

87. Resolution, 28th Conf. of Int'l Data Prot. and Priv. Comm'rs, Resolution on Privacy Protection and Search Engines 2 (Nov. 2–3, 2006), http://globalprivacyassembly.org/wp-content/uploads/2015/02/Resolution-on-Privacy-Protection-and-Search-Engines.pdf [https://perma.cc/4MEB-WE23].

processing."[88] The Article 29 opinion was limited to the retention period and did not discuss further use of the collected metadata.

Such a policy ran contrary to the techniques companies used to provide services. Google's initial proposal in 2007 was that it would anonymize search logs after 18-24 months, but the company provided no details on what anonymization would mean.[89] Later, the company agreed to remove the last byte of the IP address after nine months and to protect the privacy of users by discarding part—but not all—of every IP address.[90] Microsoft proposed removing user identifiers immediately and full IP address at eighteen months,[91] while Yahoo proposed to remove user names and the last byte of an IP address after 90 days.[92]

In the U.S., government efforts to limit the private sector's use of communications metadata or telemetry lagged. Five states have now passed privacy laws defining personal information as information that is linked or reasonably linkable to an identified or identifiable individual.[93] And, since 2016, the Federal Trade Commission (FTC) has taken a broad view of what constitutes

88. *Opinion 1/2008 of the Article 29 Data Protection Working Party on "Data Protection Issues Related to Search Engines"* 19 (Apr. 4, 2019), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2008/wp148_en.pdf [https://perma.cc/LE4R-QTAL].

89. Miguel Helft, *Google Adds a Safeguard on Privacy for Searchers*, N.Y. TIMES (Mar. 15, 2007), https://www.nytimes.com/2007/03/15/technology/15googles.html [https://perma.cc/6LWY-ZSNS].

90. Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1774 (2010).

91. *Search Privacy Comparison Chart*, MICROSOFT, https://public.dm.files.1drv.com/y4pDFCYtVl2jN7OnK2jvAtHuw6cxmPpjg35Hz5tSaa4jlDOu8dtTPUmxSRnnAhLaxlejbQxEu4gqLHYBg4rydKl4x94h_p-SK8te5W2RRNkTxwQvI4MTqcjy1WcNrp75rCtersyqrWNkQRXKxb_Ti4Y9zyJLqpvX_2XFopgqjYcR5RFl6lkF8zpPqIwVO-ZK7gn/chart2.jpg?rdrts=305049457 [https://perma.cc/X3UB-KA2T].

92. Ryan Singel, *Yahoo to Anonymize User Data After 90 Days*, WIRED (Dec. 17, 2008, 11:14 AM), https://www.wired.com/2008/12/yahoo-to-anonym/ [https://perma.cc/4ANW-3L9L] (Note that some ISPs appear to periodically change IP addresses, so Google's—or other companies'—long-term retention of IP addresses is not necessarily useful in such cases.).

93. CAL. CIV. CODE § 1798.140(o)(1) ("'Personal information' means information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household."); COLO. REV. STAT. § 6-1-1303(17) ("'Personal data': (a) Means information that is linked or reasonably linkable to an identified or identifiable individual"); Conn. Legis. Serv. P.A. 22-15, § 1(18) ("'Personal data' means any information that is linked or reasonably linkable to an identified or identifiable individual."); UTAH CODE ANN. § 13-61-101(24)(a) ("'Personal data' means any information that is linked or reasonably linkable to an identified or identifiable individual."); VA. CODE ANN. § 59.1-575 ("'Personal data' means any information that is linked or reasonably linkable to an identified or identifiable natural person.").

"personal information."[94] Yet, despite the revelatory nature of communications metadata and telemetry, there has been little to no regulatory control over its use.

In this section, we lay out the circumstances, both technical and legal, that have led to this privacy-invasive juncture. In Subsection 1.A, we provide the context for how the PSTN first came to collect communications metadata and how this information was handled, which was largely—though not fully—privacy protective. Next, in Subsection 1.B, in order to consider possible controls on private-sector collection and use of metadata and telemetry, we present a brief legal history of government acquisition and use of communications metadata in law-enforcement and national-security cases. In Subsection 1.C, we examine how in the context of IP-based communications, the *Katz*[95] and *Smith*[96] distinction of content/non-content fades away (and include a short primer on IP communications). In Subsection 1.D, we study telemetry information and begin the discussion of privacy impacts of its use. These discussions set up the background for Section II, which presents an understanding how current usage of metadata and telemetry information violates users' privacy expectations.

A. Collecting Call Metadata: How and For What Purpose

It is a well-known secret that successful businesses focus on measurement: corporations know the cost to provide a product, the marginal cost of increasing production, what failure rates are in manufacturing, what failure rates are in use, etc. Such measurement is necessary for retaining customers and for growing market share. This is true even for monopolies, though in some ways, the needs for measurement can arise from different reasons.

In 1913, the Kingsbury Commitment, an arrangement between the U.S. Government and AT&T, effectively made the phone business a monopoly, a situation that lasted until the phone company's break-up in 1984.[97] After the Kingsbury Commitment, there remained small telephone service providers throughout the United States, but these were limited to providing local service. For all intents and purposes, between 1913 and 1984, the United States

94. Ramirez, *supra* note 41, at 3–4.

95. *See generally* Katz v. United States, 389 U.S. 347 (1967).

96. *See generally* Smith v. Maryland, 442 U.S. 735 (1979).

97. Letter from James C. McReynolds, U.S. Att'y Gen., to N.C. Kingsbury, Vice Pres., AT&T (Dec. 19, 1913), *in* ANNUAL REPORT OF THE DIRECTORS OF THE AMERICAN TELEPHONE AND TELEGRAPH COMPANY 26–27 (1914), https://www.bellsystemmemorial.com/pdf/1913ATTar_Complete.pdf [https://perma.cc/7C8X-2CCZ].

had a single telephone company: AT&T. The Kingsbury Commitment required that AT&T rid itself of stock holdings in the telegraph company, Western Union, stop acquiring competing independent telephone exchanges, and open up its long-distance phone lines to independent exchanges. In response, the federal government dropped its antitrust efforts.

Long-distance competition was largely eliminated by this agreement, but that did not remove the need for measuring quality of service. Indeed, on the contrary, once the phone company was a regulated monopoly, two parties were involved in setting prices: the company itself and its government regulator. With the government involved in determining AT&T's rates, the company needed to be able to prove it was providing service at the level promised. That meant that the company had to be able to measure those capabilities.

That brings us to the technology. A local telephone exchange—known by AT&T as the telephone company central office—provides switching for subscribers. Phone switching was first done manually though a panel of jacks that corresponded to phone "numbers." A human operator would use cables to connect the jacks after a caller requested a connection. Although automated switches were developed in 1891,[98] AT&T did not install automated phone switches until 1921,[99] somewhat later than its competitors. This delay was because AT&T largely serviced cities with multiple exchanges, where many of the calls went between telephone exchanges; smaller locations did not have multiple exchanges. Operators could connect calls faster than did the automated switches of the time.[100] So, Western Electric, AT&T's equipment supplier, built technology for the phone operators, rather than directly for the callers.[101]

Charging for calls was relatively easy when operators helped set up the call. Operators would ask the caller where they were calling from[102] and record the data on "operator tickets" that included date, calling number, called number, connect time,

98. The first automated switch was developed by Almon Strowger, a Kansas City undertaker, who was concerned that human operators were directing calls to his competitors. Brenda Maddox, *Women and the Switchboard*, *in* THE SOCIAL IMPACT OF THE TELEPHONE 272 (Ithiel de Sola Pool ed., 1977).

99. *See* A.E. JOEL, JR. ET AL, *supra* note 14, at 11.

100. Sheldon Hochheiser, *Electromechanical Telephone-Switching*, ENG'G & TECH. HIST. WIKI, https://ethw.org/Electromechanical_Telephone-Switching [https://perma.cc/G52V-3XJR].

101. A.E. JOEL, JR. ET AL, *supra* note 14, at 7–8.

102. Phil Lapsley, *Extra Goodies - Automatic Message Accounting*, EXPLODING PHONE, http://explodingthephone.com/extras/ama.php [https://perma.cc/7AF7-DPW2].

disconnect time, and number of attempts.[103] By the late 1930s, it was clear that automatic switching would replace operators. Whatever automatic system replaced the operator tickets had to be cheap to operate.[104] For while the phone company was a multi-million-dollar business, the cost of a phone call was but a few cents; a recording system could not add appreciably to the price of a call.

AT&T introduced automatic ticketing in Los Angeles in 1944. This recorded data was the same as what the operator had collected, but without directly querying the caller.[105] The automated message ticketer also recorded which telephone central office was dialed and which long-distance trunk line was used.[106]

Measurement was always part of AT&T's system.[107] The company wanted to know the volume of traffic served and the volume of traffic denied (i.e., calls that did not go through).[108] That was not all AT&T sought to learn. How long it took to obtain a dial tone was particularly important during World War II when some circuits were overloaded.[109]

Engineers needed to know what percentage of the time the trunk lines were being used.[110] AT&T implemented the Trunk Usage Recorder, which checked the fullness of each trunk line every 100 seconds.[111] With this information, AT&T could determine whether any particular trunk group had sufficient capacity; for the company, sufficient capacity meant being able to handle the load at "the busiest hour of the busiest season."[112]

An overload on one part of the network could cause overloads elsewhere, so even this information was not enough for providing

103. A.E. JOEL, JR. ET AL, *supra* note 14, at 134 (Fig. 6-12).

104. *See id.* at 92.

105. *Id.* at 133.

106. *Id.*

107. *See* Venus Green, *Goodbye Central: Automation and the Decline of "Personal Service" in the Bell System, 1878-1921*, 36 TECH. AND CULTURE 912, 932 (1995). (For example, at the turn of the 20th century, the company measured how busy trunk lines were using "peg counts," implemented electromechanically. These peg counts, like Netflow described in Section II.C, *infra*, do not capture communications content.). *Glossary of Telecommunications Terms*, TELECOMMS. INDUS. ASS'N, http://standards.tiaonline.org/market_intelligence_/glossary/index.cfm?term=%26%23TCZQB%3BO%0A#:~:text=1.,From%20Weik%20'89%5D%202 [https://perma.cc/XUP9-42F6].

108. A.E. JOEL, JR. ET AL., *supra* note 14, at 400.

109. *Id.* at 105–06.

110. *Id.* at 400.

111. *Id.*

112. *Id.*

quality service.[113] Research at Bell Labs in 1963 modeled the cascading problem, enabling better monitoring and network management[114]—and fewer instances of "your call did not go through; please hang up and try again." This solution came just in time. The U.S. population's mass migration to the suburbs in the 1960s created a new calling pattern: long-distance calls increased across a wider array of telephone exchanges.

This snapshot of changing requirements between the 1930s and 1960s—and it is just a snapshot—captures the essence of what the phone company sought to learn: which equipment was being used how much of the time, what call failure rates were and what the causes were for these failures.[115] And like any business, the company also sought to understand its customers: how much and what kind of services they were using. Such information could be used to determine the development of new services as well as future pricing.

The questions technology could answer—and at what cost—limited AT&T's quest for data. For example, in the 1930s, AT&T could measure trunk usage, but computing message ticketers for individual trunks was quite expensive.[116] The 1960s is when technology—electronic switches and magnetic tape, which had become sufficiently reliable—made recording and storing call charge information easy.[117] Electronic Switching System #1 (No. 1 ESS), deployed in Succasunna, New Jersey in 1965, was the first fully electronic—as opposed to electro-mechanical—switching system.[118] It included an Automatic Message Accounting system that collected and recorded the data needed to charge customers for the call (e.g., caller number, number dialed, connection time, disconnect time).[119] With that development, AT&T began recording

113. *Id.* at 108. (This was particularly so during the war years, but also at other times, especially as AT&T provided telephone service for the U.S. military.); *id.* at 106, 577.

114. *Id.* at 108.

115. *See id.* at 108.

116. *Id.* at 135–41. (Engineers at AT&T developed an elaborate system of paper tape recording.); *see* THE NATION AT YOUR FINGERTIPS (Audio Production Inc. 1951) (on file with the Internet Archive), https://archive.org/details/the-nation-at-your-fingertips-1951/the-nation-at-your-fingertips-1951-10mbps.mp4.7:41-8:40 [https://perma.cc/LLD7-R5P9]. Local calls were largely flat rate system of so-much per month, so this system was used primarily for long-distance billing. A.E. JOEL, JR. ET AL., *supra* note 14, at 142.

117. *See id.* at 141–42.

118. *Id.* at 253, 258. (The move to electronic switching even merited a front-page story in the New York Times.); Robert Alden, *A Shift to All-Electronic Phones Begun in Biggest Step Since Dial*, N.Y. TIMES, Sept. 20, 1964, at 1.

119. J.G. Ferguson et al., *No. 1 ESS Apparatus and Equipment*, 43 BELL SYS. TECH. J. 2355, 2357 (1964).

connection data for local calls,[120] creating the Call Detail Records that are so familiar today. Because AT&T largely did flat-rate billing for local calls,[121] there was no immediate direct use for those records—but one came with the 1984 divestiture.[122]

The breakup of AT&T turned the company into a long-distance carrier with no local services. This change resulted in negative consequences for the company; supplying long-distance access was not, in itself, a viable business, especially as competition surfaced.[123] But AT&T had one thing going for it: data—and lots of it. Studying vast data sets—numbering in the millions daily and billions annually—whether credit card transactions, traffic routes, or calls carried by a phone company,[124] gave the company an edge over its competitors. In the late 1990s, AT&T was servicing approximately 275 million phone calls daily[125] and could use its massive data sets to its advantage.

By studying calling patterns, AT&T researchers learned how to determine the "bizocity" of a phone number (i.e., the extent to which a phone number follows "business-like" patterns).[126] Times of day, days of the week, and length of calls were all significant predictors of whether a number was likely to belong to a business.[127] This differentiation benefitted AT&T, which charged

120. *Id.* at 2309.

121. A.E. JOEL, JR. ET AL., *supra* note 14, at 142 (In the late 1970s, customers preferred that model even when alternatives were available; the preference may simply have been custom, since that was the model that AT&T had used for decades); *see* J.G. Cosgrove & P.B. Linhart, *Customer Choices Under Local Measured Telephone Service*, PUB. UTIL. FORT., Aug. 30, 1979, at 27; *see also* L. Garfinkel & P.B. Linhart, *The Transition to Local Measured Telephone Service*, PUB. UTIL. FORT., Aug. 16, 1979, at 17; L. Garfinkel & P.B. Linhart, *The Revenue Analysis of Local Measured Telephone Service*, PUB. UTIL. FORT., Oct. 9, 1980, at 15.

122. The company currently known as AT&T is not the same company it was in 1984. *See* Andrew Beattie, *AT&T's Successful Spinoffs*, INVESTOPEDIA (Dec. 6, 2022), https://www.investopedia.com/ask/answers/09/att-breakup-spinoff.asp#:~:text=In%201984%2C%20AT%26T's%20local%20telephone,internet%20service%20for%20many%20consumers [https://perma.cc/J4EW-AXFP]. (After the 1984 divestiture, AT&T only served long distance. *Id.* That changed when the company was acquired by SBC in 2005; although SBC was bigger, the merged company became AT&T). Ken Belson, *SBC Agrees to Acquire AT&T for $16 Billion*, N.Y. TIMES (Jan. 31, 2005), https://www.nytimes.com/2005/01/31/business/sbc-agrees-to-acquire-att-for-16-billion.html [https://perma.cc/S33D-7BAP]; *The Historical Brands of AT&T*, AT&T, https://about.att.com/innovation/ip/brands/history [https://perma.cc/2F2V-DPNU].

123. Joseph H. Weber, *The Bell System Divestiture: Background, Implementation, and Outcome*, 61 FED. COMMC'NS L.J., 21, 28 (2008).

124. Corinna Cortes & Daryl Pregibon, *Giga-Mining*, *in* PROCEEDINGS OF THE FOURTH INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 174, 174 (Rakesh Agrawal & Paul Stolorz eds., 1998).

125. *Id.*

126. *Id.*

127. *Id.* at 176.

business and residential customers differently.[128] It also provided fraud protection. As residences were a much likelier setting for making long-distance calls at night than businesses, a sudden flood of such calls to or from a business line would generate a red alert warning of potential fraudulent use.[129]

Studying the CDRs provided ways of uncovering various types of telephone fraud.[130] In the 1990s, hustlers at New York City's Port Authority Bus Station were "selling" cheap international phone calls to China, India, and elsewhere from public pay phones.[131] CDRs revealed that the callers were accessing companies' Private Branch Exchange[132] systems and using stolen access codes to place these calls.[133]

By 1997, fraudulent calls, largely wireless and international, were estimated to be costing the U.S. telecommunications industry $1 billion annually.[134] Finding anomalies in the CDRs—different times of day for calling and different days of the week (e.g., weekend versus weekday) for international calls and a sudden increase in the number of international calls—was useful for uncovering fraud.[135] Additionally, link analysis—seeing how "close" a particular telephone number is to those of known fraudsters—helped uncover other defrauders.[136]

Link analysis was also useful in other ways. Using this data, companies were able to learn "who their customers are, where they are, what their needs are, how they use existing services and products, what makes customers stop using or buying the offered

128. *Id.* at 175 (describing the two sets of customers receiving different customer care, fraud prevention and detection services).

129. *Id.*

130. Richard A. Becker et al., *Fraud Detection in Telecommunications: History and Lessons Learned*, 52 TECHNOMETRICS 20, 29–31 (2010). There are multiple types of fraud in telecommunications, including subscription fraud (i.e., signing up for an account with no intention to pay), intrusion fraud (i.e., intruding into a valid user's account with intent to use services for free), fraud based on vulnerable technology, and masquerading as another user. *Id.* at 21–22.

131. Gisela Bichler & Ronald V. Clarke, *Eliminating Pay Phone Toll Fraud at the Port Authority Bus Terminal in Manhattan*, *in* CRIME PREVENTION STUDIES, VOLUME 6: PREVENTING MASS TRANSIT CRIME 93, 96 (Ronald V. Clarke ed., 1996).

132. *Id.* at 101. A Private Branch Exchange is a small version of a telephone switch located within a company or organization; it supports the private network. *Id.* at 112.

133. *Id.* at 101, 112.

134. Kenneth C. Cox et al., *Visual Data Mining: Recognizing Telephone Calling Fraud*, 1 DATA MINING AND KNOWLEDGE DISCOVERY 225, 226 (1997).

135. Kathleen Fisher et al., *An Application-Specific Database*, *in* DATABASE PROGRAMMING LANGUAGES 213, 214 (Gosta Grahne ed., 2001); Cortes & Pregibon, *supra* note 124, at 178.

136. Corinna Cortes et al., *Communities of Interest*, *in* ADVANCES IN INTELLIGENT DATA ANALYSIS 105, 110–11 (Frank Hoffman et al. eds., 2001).

services, and what offers could attract new customers."[137] Thus, if a customer moved—or if a customer called to complain about their charges—a service representative would offer that customer a more attractive service plan based on their calling patterns.

Other uses of such data were possible. Beginning in 1991, the carrier Microwave Communications, Inc. (MCI) offered customers a "Friends and Family" plan that would discount calls to a set of numbers they called most often, on the condition that those they called were also MCI customers.[138] This plan turned out to be a great marketing strategy for the company, which was an upstart competing against "Ma Bell" (AT&T).[139] MCI recruited new customers by using information about those whom MCI customers called most often. In what may seem to be a quaint point by today's standards, an AT&T spokesman expressed privacy concerns about the MCI plan because customers were "sharing what is essentially private information with a private corporation."[140]

AT&T also did research into "communities of interest," closed sets of a few people who communicated only amongst themselves. The company's interest was in finding fraudsters who have changed their network identity—their phone number—but continued their criminal behavior. CDRs provide ways to determine this information due to the criminals maintaining similar calling patterns despite changing their telephone identifiers.[141]

AT&T was not the only organization to find CDR data useful. Law enforcement investigators who use CDR data generally considered it to be an invaluable tool for identifying criminals about whom little is known aside from their location during the commission of a crime. Intelligence organizations employ CDRs in this[142] and other ways. One pattern of interest is a very small,

137. Cortes & Pregibon, *supra* note 124, at 174.

138. *The History of MCI (Microwave Communications, Inc.)*, TEL. WORLD, https://telephoneworld.org/long-distance-companies/the-history-of-mci/ [https://perma.cc/L76D-QU9A]; MCI "Friends and Family" was initially limited to twelve numbers. Bart Ziegler, *MCI Announces "Friends & Family" Discount Plan Graphic*, AP NEWS (Mar. 18, 1991), https://apnews.com/article/d9a32c48ef6369df5556582a50b25bbb [https://perma.cc/D85D-BJBF]]; It would later expand to twenty. *The History of MCI (Microwave Communications, Inc.)*.

139. Ziegler, *supra* note 138. "Ma Bell" was the colloquial name for the phone company. Richard L. Ottinger, *MaBellOpoly*, N.Y. TIMES, Apr. 19, 1975, at 27; Tim Bajarin, *The History of AT&T's Long Evolution into a Technology Powerhouse*, TIME (Oct. 24, 2016, 10:53 AM), https://time.com/4542446/att-time-warner-history/ [https://perma.cc/JQ89-72X2].

140. Ziegler, *supra* note 138.

141. Cortes et al., *supra* note 136, at 110–11.

142. *See, e.g.*, Stephen Gray & Don van Natta, *Thirteen with the C.I.A. Sought by Italy in a Kidnapping*, N.Y. TIMES (June 25, 2005), https://www.nytimes.com/2005/06/25/world/europe/thirteen-with-the-cia-sought-by-

closed circle of callers; this can indicate a group of spies or a terrorist organization.[143]

As a legacy of its role in the long-distance business, AT&T owned more than three-quarters of the landline switches in the United States and the second largest number of cell towers in the country.[144] Those numbers, plus telephone records the company had going back to 1987, meant AT&T's CDR collection vastly swamped whatever connection information other telecommunications providers had.[145] AT&T combined the information with the various analysis tools (e.g., the ability to find a criminal's new phone number from his calling pattern (so-called "dropped phones") and additional phones the suspect is using)[146] that researchers had developed. AT&T marketed this new product, Hemisphere—a proprietary CDR database—to law enforcement.[147] The tool provided coverage no other company could match.[148]

There was no law enforcement agent peering into the database and seeing what was there, and there appeared to be nothing illegal

italy-in-a-kidnapping.html [https://www.nytimes.com/2005/06/25/world/europe/thirteen-with-the-cia-sought-by-italy-in-a-kidnapping.html [https://perma.cc/AF4N-J3LJ]; Ronen Bergman, *The Hezbollah Connection*, N.Y. TIMES MAG. (Feb. 10, 2015), https://www.nytimes.com/2015/02/15/magazine/the-hezbollah-connection.html [https://perma.cc/QJJ4-FTE9].

143. Kashmir Hill, *How Israeli Spies Were Betrayed by Their Cell Phones*, FORBES (Nov. 11, 2011, 3:11 PM), https://www.forbes.com/sites/kashmirhill/2011/11/21/how-israeli-spies-were-betrayed-by-their-cell-phones/?sh=e6bd0e341bf4 [https://perma.cc/83PH-5BAA]; Black Hat, *Black Hat USA 2013 - OPSEC Failures of Spies*, YOUTUBE (Dec. 3, 2013), https://www.youtube.com/watch?v=BwGsr3SzCZchttps://www.youtube.com/watch?v=BwGsr3SzCZc [https://perma.cc/UD6U-X2A3] (22:45–25:10).

144. Kenneth Lipp, *AT&T Is Spying on Americans for Profit*, DAILY BEAST (Apr. 13, 2017, 2:36 PM), https://www.thedailybeast.com/atandt-is-spying-on-americans-for-profit [https://perma.cc/UD6U-X2A3].

145. Scott Shane & Colin Moynihan, *Drug Agents Use Vast Phone Trove, Eclipsing N.S.A.'s*, N.Y. TIMES (Sept. 1, 2013), https://www.nytimes.com/2013/09/02/us/drug-agents-use-vast-phone-trove-eclipsing-nsas.html [https://perma.cc/37TR-Y8ZT].

146. *Hemisphere: Law Enforcement's Secret Call Records Deal with AT&T*, ELEC. FRONTIER FOUND., https://www.eff.org/cases/hemisphere [https://perma.cc/2YFR-E2BL].

147. *See generally, e.g.*, Off. Nat'l Drug Control Pol'y, Los Angeles Hemisphere (on file with the Electronic Frontier Foundation), https://www.eff.org/files/2015/07/07/nyt_hemisphere_powerpoint.pdf [https://perma.cc/8HHU-TUDN] (slide 2).

148. Letter from Adrian Garcia, Sheriff, Harris Cnty. Sheriff's Off., to the Hon. Ed Emmett, Judge, Harris Cnty., Ct., Tex., and Comm'rs Lee, Morman, Radack, & Cagle, Members, Harris Cnty. Comm'rs Ct. (Aug. 18, 2014) (on file with the Electronic Frontier Foundation Foundation), https://www.eff.org/files/2015/07/07/2014_documents_-_5_pages_.pdf [https://perma.cc/EU4S-BA3Y] (describing the richness of the AT&T data by pointing to the language of a contract from the Sheriff's Department in Harris County, Texas: "Project Hemisphere is a crucial service, only available through AT&T Government Services;" and "[t]here are no alternatives to this provider").

either in the set-up or in the running of the program. Members of law enforcement agencies could access Hemisphere only upon production of a subpoena[149] requesting CDRs linked to a particular number or set of numbers.[150] Yet, although Hemisphere was an unclassified system,[151] it operated under high secrecy. Users were warned to "never refer to Hemisphere in any official document" and if information did need to refer to a Hemisphere request, refer to "information obtained from an AT&T subpoena."[152] The high level of secrecy for the unclassified program was not centered on protecting the production of the CDRs, whose results appeared in court cases; its purpose was to hide the fact that AT&T was marketing such a privacy-invasive program.

B. Government Access to Telecommunications Metadata: A Brief Overview

Although our focus in this paper is on private-sector collection and use of communications metadata, studying how U.S. law on the government's use of communications metadata evolved in response to changing communications technologies provides valuable context for discussing private-sector use of metadata and telemetry. Thus, we take a brief detour to examine this issue considering (1) U.S. jurisprudence regarding access to communications metadata, including the special case of location data, and (2) the more recent development of purchasing such information from private vendors.

1. The Law Regarding U.S. Government Access to Communications Metadata

We do not reprise the history of U.S. wiretap law, which has been written about extensively. The salient point is that in 1967 the Court ruled in *Katz v. United States*: "[T]he Fourth Amendment protects people, not places."[153] Thus, a warrant was needed for government wiretapping. Collection of call detail records was not addressed in *Katz*, nor was it addressed in the 1968 *Omnibus Crime Control and Safe Streets Act*, Title III of which laid out a warrant procedure for wiretapping in criminal cases.[154]

149. *See, e.g.*, Off. Nat'l Drug Control Pol'y, *supra* note 147 (slide 9); *see also* discussion of requirements for accessing CDRs and telephone dialing, routing, addressing, or signaling information *infra* Section II.B.i.

150. Off. Nat'l Drug Control Pol'y, *supra* note 147 (slide 10).

151. *Id.* (slide 2).

152. *Id.* (slide 12).

153. Katz v. United States, 389 U.S. 347, 351 (1967).

154. Omnibus Crime Control and Safe Streets Act of 1968 § 801 et seq., Pub. L. No. 90-351, 82 Stat. 197.

This omission was not surprising. The first No. 1 ESS and its associated accounting system—the CDRs—had only been introduced in 1965.[155] Use of CDRs was unlikely to have become a standard police tool in 1968, the time Title III was passed. But, as the Court had observed, "[t]ime works changes, brings into existence new conditions and purposes"[156]—and so it was with telephone metadata.

Shortly after Baltimore resident Patricia McDonough was robbed in 1976, she began receiving threatening phone calls.[157] During one of these McDonough was asked to step onto her porch; when she did so, a car passed by that appeared to be the same as the automobile present during the robbery.[158] McDonough told the Baltimore police, who traced the auto's license plate; the car belonged to a Michael Lee Smith.[159] The police arranged for a pen register, which records the phone number of outgoing calls, to be placed on Smith's phone line.[160]

Though the pen register installation was done sans warrant or court order, in *Smith v. Maryland* the Supreme Court ruled that the tool's use was legal.[161] Basing its judgement on the third-party rule, the Court stated: "Petitioner in all probability entertained no actual expectation of privacy in the phone numbers he dialed, and even if he did, his expectation was not 'legitimate.'"[162] The Court wrote:

> Telephone users . . . typically know that they must convey numerical information to the phone company; that the phone company has facilities for recording this information; and that the phone company does in fact record this information for a variety of legitimate business purposes. Although subjective expectations cannot be scientifically gauged, it is too much to believe that telephone subscribers, under these circumstances, harbor any general expectation that the numbers they dial will remain secret.[163]

155. A.E. JOEL, JR. ET AL., *supra* note 14, at 199.
156. Weems v. United States, 217 U.S. 349, 373 (1910).
157. Smith v. Maryland, 442 U.S. 735, 737 (1979).
158. *Id.*
159. *Id.*
160. *Id.*
161. *Id.* at 745–46.
162. *Id.* at 745.
163. *Id.* at 743.

The Court reasoned that the information obtained from a pen register[164]—number called, date, time, and duration—did not fall under Fourth Amendment protection.[165] Although the *Smith* opinion explicitly states, "that a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties,"[166] as law professor Susan Freiwald noted, "[F]ar from establishing a broad "non-contents" rule, Smith covered only the telephone numbers the target dialed and limited its reasoning to that data."[167]

Law-enforcement use of pen registers and trap-and-trace devices, which capture the numbers of incoming calls, were later codified in the Electronic Communications Privacy Act of 1986 (ECPA).[168] Through one of its subsidiary acts—the Pen Register Act—ECPA established legal requirements for the use of pen registers and trap-and-trace devices (the latter capture all numbers calling a specific number).[169] A court order was needed to install pen registers and trap-and-trace devices.[170] Such an order could be applied for by a federal lawyer and granted by a federal district judge or magistrate or, unless prohibited by a state, applied for by a state government lawyer and granted by a court of "competent jurisdiction."[171] For pen register grants, the order required a showing that the request was "relevant to an ongoing criminal investigation"; by contrast, granting the a request for other types of "wire or electronic communication" data required only "reason to believe the . . . information sought [was] relevant to a legitimate law enforcement inquiry."[172]

In 1986, pen registers and trap-and-trace devices provided information on the calls made and received by a landline phone, including the numbers called or calling and the time and disconnect time. But, within a few years of ECPA's passage, the "richness" of the transactional data of Internet communications and mobile

164. *Id.* at 736 n.1 (citing United States v. New York Tel. Co., 434 U.S. 159, 161 n.1 (1977)).

165. *Id.* at 745–46.

166. *Id.* at 743–44.

167. Susan Freiwald, *Cell Phone Location Data and the Fourth Amendment: A Question of Law, Not Fact*, 70 MD. L. REV. 681, 733 (2011).

168. The Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848.

169. Tit. III, § 301, 100 Stat. at 1869–70, 1871 (originally codified as 18 U.S.C. §§ 3122, 3123, 3126(3), -(4)).

170. Tit. III, § 301, 101 Stat. at 1869 (codified at 18 U.S.C. § 3122(a)).

171. *Id.*

172. Tits. II, § 201, III, § 301, 100 Stat. at 1862–63, 1869 (originally codified at 18 U.S.C. §§ 2703(d), 3122(a)).

phones' exposure of location information argued for adjustments to the law, a history that legal scholar Jim Dempsey described in detail.[173] Thus, in 1994, the Communications Assistance for Law Enforcement Act (CALEA) amended ECPA to prevent collection of cellphone location purely on the pen-register standard[174]—but failed to specify what the standard should be.[175]

Dempsey recounted that, during the late 1990s and into 2000, Congress considered various bills that would have strengthened the requirements for obtaining pen/trap information, including cellphone location;[176] such attempts at reform died after the 9/11 attacks.[177] The attacks on the U.S. created remarkable changes in surveillance law. We will start with what did not happen—strengthening requirements for obtaining non content—then look at what did.

Tracking location data is extremely valuable for criminal investigations, but the requirements regarding government acquisition of such information were unclear. In the absence of new laws governing the acquisition of location data, to a certain extent, courts floundered. Did people have a reasonable expectation of privacy in that information, which would imply there was a warrant requirement for location information? Law enforcement tried a go-round instead, with a so-called "hybrid order."[178] This order combines the *retrospective* collection of "records or other information" under the Stored Communications Act with the *prospective* pen register and trap-and-trace information obtainable under ECPA.[179] In August 2005, Magistrate Judge James Orenstein ruled that the hybrid order was insufficient for obtaining real-time CSLI, stating that law enforcement needed a warrant to

173. James X. Dempsey, *Keynote Address: The Path to ECPA Reform and the Implications for United States v. Jones*, 47 U. SAN FRANCISCO L. REV. 225, 227 (2012).

174. Communications Assistance for Law Enforcement Act of 1994, Pub. L. No. 103-414, tit. I, § 103, 108 Stat. 4280, 4281 (codified at 47 U.S.C. § 1002(a)(2)(B)).

175. Dempsey, *supra* note 173, at 227–28.

176. *Id.* at 227–29.

177. *Id.* at 229*; see* Stephen Wm. Smith, *Gagged, Sealed, and Delivered: Reforming ECPA's Secret Docket*, 6 HARV. L. & POL'Y REV. 313, 324–26 (2012).

178. BHAIRAV ACHARYA ET AL., SAMUELSON L., TECH. & PUB. POL'Y CLINIC, CELL PHONE LOCATION TRACKING: A NATIONAL ASSOCIATION OF CRIMINAL DEFENSE LAWYERS (NACDL) PRIMER (2016), https://www.law.berkeley.edu/wp-content/uploads/2015/04/2016-06-07_Cell-Tracking-Primer_Final.pdf [https://perma.cc/F3A8-KM4K] (describing the courts' acceptance of a "hybrid" authorization in the "Judicial Authorization" table); *see* United States v. Graham, 824 F.3d 421, 437–38 (4th Cir. 2016), *abrogated by* Carpenter v. United States, 138 S. Ct. 2206 (2018) (holding that the collection of CSLI does not require a warrant).

179. ACHARYA ET AL., *supra* note 178; Pen Register Act, 18 U.S.C. §§ 3121–3127.

seek such information.[180] A month later, a Texas magistrate, Judge Stephen Smith, also rejected the hybrid order for obtaining CSLI.[181]

Meanwhile Americans were now increasingly using cellphones and, after the introduction of the iPhone in 2006, smartphones.[182] This trend led to an increase in the number of base stations, potentially making CSLI even more useful to law enforcement. By 2010, while some cellphone sectors in rural areas could stretch over several miles, in densely populated areas, some cell sites served areas as small as a train station waiting room.[183] However, neither the law nor the courts kept up with the changes of technology. They were struggling with making sense of ECPA, CALEA, and the provision of location information. Testifying before Congress on cellphone tracking, Judge Smith said, "For nearly a quarter-century, magistrate judges have been issuing tens of thousands of these orders under a fiendishly complex statute without any substantial guidance from a higher court."[184] The result was a thicket of inconsistent magistrate rulings.[185]

Freiwald explained some of this phenomenon. There was the usual reluctance of the courts to handle a broad issue if a narrow ruling would solve the problem, but other issues also created clouds. Freiwald pointed out that ECPA's lack of an exclusionary remedy meant there was little reason for defendants to bring statutory claims for law enforcement acquisition.[186] At the same time, government litigators preferred not to contest cases that might raise Fourth Amendment issues regarding such acquisition.[187]

Freiwald observed that few appellate courts had dealt with the issue.[188] One case in particular was striking. Freiwald described

180. *In re* Authorizing the Use of a Pen Register, 384 F. Supp. 2d 562, 564 (E.D.N.Y. 2005).

181. *In re* Application for Pen Register and Trap/Trace Device with Cell Site Location Auth., 396 F. Supp. 2d 747, 764–65 (S.D. Tex. 2005).

182. Russell Heimlech, *Americans and Their Cell Phones*, PEW RSCH. CTR. (Aug. 27, 2011), https://www.pewresearch.org/fact-tank/2011/08/29/americans-and-their-cell-phones/ [https://perma.cc/UW9D-GCEY].

183. *ECPA Reform and the Revolution in Location Based Technologies and Services*: *Hearing Before the Subcomm. on the Const., C.R., and C.L. of the H. Comm. on the Judiciary*, 111th Cong. 16 (2010) (statement of Matt Blaze) ("But the latest technology has trended toward what are called variously microcells, picocells and femtocells that are designed not to serve an area of miles in diameter, but rather to serve a very, very specific location, such as a floor of a building or even an individual room in a building such as a train station waiting room or an office complex or hotel or even a private home").

184. *Id.* at 76 (statement of Judge Stephen Wm. Smith).

185. Dempsey, *supra* note 173, at 231.

186. Freiwald, *supra* note 167, at 681–82.

187. *Id.* at 682.

188. *Id.*

how, in 2010, the Third Circuit Court of Appeals was faced with the question of whether the government could compel a service provider to disclose stored location data under a court order (this was a "D order" under 18 U.S.C. § 2703(d)).[189] Lower courts denied the collection on statutory and constitutional grounds,[190] but the Third Circuit chose not to rule on the Fourth Amendment issue on reasonable expectation of privacy.[191] Instead, the Third Circuit remanded the case to the magistrate judge to make a determination on whether to impose a warrant requirement.[192]

Congress did not react to this absence of law. For a time, neither did the courts. Even in *United States v. Jones*, in which law enforcement tracked a suspect's car through a GPS device that had been placed on the vehicle without a valid warrant, the Supreme Court did not address the reasonable expectation of privacy (REP) in public spaces.[193] The majority opinion was that the warrantless attachment of the GPS device while in the defendant's driveway constituted trespass, thus avoiding ruling on the REP issue.[194] But just as Justice Brandeis's dissent in *Olmstead*,[195] served as a harbinger of future Court rulings on privacy, so also might Justice Sotomayer's concurring opinion in Jones do so:

> More fundamentally, it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties. *E.g., Smith*, 442 U. S., at 742; *United States* v. *Miller*, 425 U.S. 435, 443 (1976). This approach is ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks. People disclose the phone numbers that they dial or text to their cellular providers; the URLs that they visit and the e-mail addresses with which they correspond to

189. *Id.* at 682–84; *In re* Application of U.S. for an Order Directing a Provider of Elec. Commc'n Servs. to Disclose Records to the Gov't, 620 F.3d 304, 305–06 (3d Cir. 2010).

190. In re Application of U.S. for an Order Directing a Provider of Elec. Commc'n Servs. to Disclose Records to the Gov't, 534 F. Supp. 2d 585, 586, 612, 616 (W.D. Pa. 2008), *affirmed*, No. 07-524M, 2008 WL 4191511 (W.D. Pa. Sept. 10, 2008), *vacated*, 620 F.3d 304 (3d Cir. 2010).

191. Freiwald, *supra* note 167, at 683–84.

192. *In re* Application of U.S. for an Order Directing a Provider of Elec. Commc'n Servs. to Disclose Records to the Gov't, 620 F.3d at 319.

193. *See* U.S. v. Jones, 565 U.S. 400, 413 (2012).

194. *Id.* at 409–11.

195. Olmstead v. United States, 277 U.S. 438, 478 (1928) (Brandeis, J., dissenting).

> their Internet service providers; and the books, groceries, and medications they purchase to online retailers.[196]

Such a response took half-dozen years. It was only with the 2018 *Carpenter* decision that the Court addressed government acquisition of non-content. Even then, the Court's opinion was limited to the collection of location data. However, as Paul Ohm adroitly described, many other types of records, including banking, phone, websites visited, etc. of websites visited, are likely to fall under the same warrant requirements as CSLI.[197] Others are likely to fall under this umbrella as well.[198]

With this brief summary of U.S. jurisprudence on communications metadata, let us consider the laws enacted after the 9/11 attacks. The USA PATRIOT Act (PATRIOT Act) modified law governing pen-register and trap-and-trace orders in several ways,[199] extending the purview of pen/trap orders from the location in which they were filed to a nationwide order[200] and providing consistency with how electronic surveillance orders were handled.[201] The law required such orders be issued only by a court "having jurisdiction over the offense being investigated," a change intended to prevent "forum shopping."[202]

Those aspects of the PATRIOT Act were small wins for privacy; other aspects were not. Because traditional pen register and trap-and-trace devices captured all transmitted electronic impulses, and not just those related to call signaling, they were capable of collecting communications content such as the electronic banking transactions, a prescription order, etc.[203] At the request of Senate Judiciary Committee chair Senator Patrick Leahy, the PATRIOT Act expressly prohibited the use of pen/trap devices to collect content.[204] The pen/trap statute was also amended to add "routing" and "addressing" information to describe information that can be

196. *Jones*, 565 U.S. at 417.

197. Ohm, *supra* note 71, at 378–81.

198. *Id.* at 392–93.

199. *See* Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT Act) Act of 2001, Pub. L. No. 107-56, tit. II, §§ 214, 216, 115 Stat. 272.

200. Tit. II, § 216, 115 Stat. at 288–90.

201. Beryl A. Howell, *Seven Weeks: The Making of the USA Patriot Act*, 72 GEO. WASH. L. REV. 1145, 1195 (2004).

202. *Id.* at 1196.

203. 147 CONG. REC. S11,000 (daily ed. Oct. 25, 2001) (statement of Sen. Patrick Leahy).

204. Tit. II, § 216(a)(3), 115 Stat. at 288; *See* Howell *supra* note 201, at 1198.

intercepted under the pen/trap statute,[205] leaving open the question of where to draw the line between content and non-content. After the act passed, the Department of Justice acknowledged that "reasonable minds may differ as to whether, and at what stage, URL information might be construed as content."[206]

The language generated some confusion around "To" and "From" of email addresses, which are roughly akin to addresses on a card inside a package;[207] these were viewed by the Department of Justice as addressing information, not content.[208] Despite Senator Leahy's effort to prevent the collection of content under pen/trap standards, a combination of haste in the passage of the PATRIOT Act (passed just one month after the 9/11 attacks), an arguably deliberate lack of clarity on the part of the Department of Justice regarding what could be collected, and the complexity of IP-based communications,[209] led to a situation in which pen register and trap-and-trace devices could and did[210] collect content.

It was not until Edward Snowden's disclosures in June 2013 that the public learned of the U.S. Government's bulk collection of domestic CDRs. Shortly after the 9/11 attacks, President George W. Bush authorized bulk collection of the metadata of American's telephone and email communications through a program run by the National Security Agency (NSA).[211] This authorization continued to be renewed, until a December 2005 *New York Times* article revealed that the government was conducting warrantless wiretapping.[212] At that point, a telecommunications operator

205. Howell, *supra* note 201, at 1197.

206. *Id.* (quoting Letter from Daniel A. Bryant, Assistant Att'y General, to Patrick J. Leahy, Chairman, Comm. on the Judiciary (Nov. 29, 2001) (on file with The Geo. Wash. L. Rev.) (response to questions to Attorney General Ashcroft in letter dated Nov. 1, 2001) (answer to question number 5).

207. *See* Bellovin et al., *supra* note 50, at 57–64.

208. U.S. Dep't of Just., Electronic Surveillance Manual: Procedures And Case Law Forms 39 (2005); Bellovin et al., *supra* note 50, at 61–64.

209. *See, e.g.*, Bellovin et al., *supra* note 50, at 5.

210. This was not simply a "could." The 2009 Department of Justice Electronic Surveillance Manual conflates the "To" and "From" in email headers with actual address information. *See* Bellovin, *supra* note 50, at 61–64.

211. OFFS. OF INSPECTORS GEN. OF THE U.S. DEP'T OF DEF., U.S. DEP'T OF JUST., CIA, NSA, AND OFF. DIR. NAT'L INTEL., NO. 2009-0013-A REPORT ON THE PRESIDENT'S SURVEILLANCE PROGRAM 1 (2009), https://oig.justice.gov/reports/2015/PSP-09-18-15-full.pdf [https://perma.cc/5N9V-CFC4].

212. James Risen & Eric Lichtblau, *Bush Lets U.S. Spy on Callers Without Courts*, N.Y. TIMES (Dec. 16, 2005), https://www.nytimes.com/2005/12/16/politics/bush-lets-us-spy-on-callers-without-courts.html [https://perma.cc/S3CM-ND3W].

requested that the bulk metadata collection be moved to operate under court order rather than "Presidential authorization."[213]

The PATRIOT Act authorized the "production of any tangible things (including books, records, papers, documents, and other items) for an investigation to protect against international terrorism or clandestine intelligence activities,"[214] but in 2006, an amended version of the law added a requirement that the tangible things be relevant to the investigation.[215] Broadly interpreting the meaning of "relevant," the Foreign Intelligence Surveillance Court (FISC) authorized the bulk collection program, which operated unknown to the public until the disclosures in 2013.[216]

Originally, there were concerns that the NSA was data mining the bulk collection for "interesting patterns" (e.g., whether there were very small groups of people who communicated only with each other, which was potentially indicative of a terrorist group). In fact, the database could only be searched by starting with a "seed" (a telephone number or other selector) approved by a senior NSA official affirming that "there is reasonable, articulable suspicion" that the seed "is associated" with a terrorist organization identified by the Foreign Intelligence Surveillance Court.[217] An NSA analyst could search for numbers in the database that were directly in contact with the seed (one hop), or within one additional step (two hops)—at one point, three hops was permissible, but this practice was eventually limited.[218]

President Obama tried to downplay the impact of the collection—"[I]f you're a U.S. person, then NSA is not listening to your phone calls and it's not targeting your emails unless it's getting an individualized court order,"[219] but public response to the

213. OFF. OF THE INSPECTOR GEN., NSA, ST-09-002, REPORT ON THE PRESIDENT'S SURVEILLANCE PROGRAM 39–40 (2009), [https://perma.cc/JN85-9FGV] (working draft) (noting comments by NSA General Counsel Vito Potenza that the decision to transition the telephony metadata program to the Business Records provision was due to a private sector company reacting to the *N.Y. Times* story).

214. Tit. II, § 501(a)(1), 115 Stat. at 287.

215. USA PATRIOT Act Improvement and Reauthorization Act of 2005, Pub. L. No. 109-177, tit. I, § 106, 120 Stat. 192, 196 (codified as amended at 50 U.S.C. § 1861(b)(2)(A)).

216. The first court order was issued in May 2006. *See In re* Application of the Federal Bureau of Investigation for an Order Requiring the Production of Tangible Things from [REDACTED], Order, Docket No. BR 06-05 (FISA Ct. May 24, 2006) (approving FBI request to collect mass telephone metadata).

217. PRIV. AND C.L. OVERSIGHT BD., REPORT ON THE GOVERNMENT'S USE OF THE CALL DETAILS RECORDS PROGRAM UNDER THE USA FREEDOM ACT 60 (2020).

218. *Id.* at 8–9.

219. *President Obama Defends NSA Spying*, BUZZFEED NEWS (June 17, 2013, 1:45 PM), https://www.buzzfeednews.com/article/buzzfeedpolitics/president-obama-defends-nsa-spying [https://perma.cc/P84G-L24J].

bulk communications metadata collection was extremely negative. To begin with, the broad interpretation of the law was unknown even to some of the bill's sponsors until the Snowden disclosures. Furthermore, the information collected through the program proved useful in only a few cases.[220] By 2013, the program's value was largely nil due to changes in the organization of terrorist groups and the development of new communications technologies.[221] But the Snowden disclosures complicated discussion of ending the program.

In 2014 the White House announced that the NSA would be limited to querying only within two hops of a selection term.[222] There was strong Congressional and public pressure for further limits on the program. A year later, Congress passed the USA FREEDOM Act of 2015 (UFA), which fundamentally transformed the metadata surveillance program.[223] The NSA would no longer directly collect the communications metadata; instead, the agency would have access to the data held under constrained rules.[224] The data itself would be held by the providers.[225] At first, it appeared that the program was functioning well,[226] but in June 2018, the NSA announced it was purging three years of records.[227] The NSA later suspended the program.[228] Although the Trump

220. PRIV. & C.L. OVERSIGHT BD., REPORT ON THE TELEPHONE RECORDS PROGRAM CONDUCTED UNDER SECTION 215 OF THE USA PATRIOT ACT AND ON THE OPERATIONS OF THE FOREIGN INTELLIGENCE SURVEILLANCE COURT 152–53 (Jan. 23, 2014), https://irp.fas.org/offdocs/pclob-215.pdf [https://perma.cc/4AR4-NSW7].

221. Susan Landau & Asaf Lubin, *Examining the Anomalies, Explaining the Value: Should the USA FREEDOM Act's Metadata Program Be Extended?*, 11 HARV. NAT'L SEC. J. 308, 311 (2020).

222. Press Release, White House Press Off., FACT SHEET: The Administration's Proposal for Ending the Section 215 Bulk Telephony Metadata Program (Mar. 27, 2014), https://obamawhitehouse.archives.gov/the-press-office/2014/03/27/fact-sheet-administration-s-proposal-ending-section-215-bulk-telephony-m [https://perma.cc/UH65-B8V7].

223. Landau & Lubin, *supra* note 221, at 317.

224. USA FREEDOM Act of 2015, Pub. L. No. 114-23, tit. I, § 103, 129 Stat. 268, 272.

225. *See, e.g.*, Bart Forsyth, *Banning Bulk: USA FREEDOM Act and Ending Bulk Collection*, 72 WASH. & LEE L. REV. 1307, 1334–39 (2015).

226. On April 19, 2019, former NSA Deputy Director Rick Ledgett explained that UFA "transferred the compliance burden from NSA, which had to maintain the universe of call data, to the telecommunications providers, who only had to give NSA those contacts responsive to an authorized query" in personal communication with Susan Landau. *See* Landau & Lubin, *supra* note 221, at 321.

227. Charlie Savage, *NSA Purges Hundreds of Millions of Call and Text Records*, N.Y. TIMES (June 29, 2018), https://www.nytimes.com/2018/06/29/us/politics/nsa-call-records-purged.html [https://perma.cc/N89W-KEP5].

228. Charlie Savage, *Trump Administration Asks Congress to Reauthorize NSA's Deactivated Call Records Program*, N.Y. TIMES (Aug. 15, 2019), https://www.nytimes.com/2019/08/15/us/politics/trump-nsa-call-records-program.html [https://perma.cc/D7Z3-D4GB].

administration supported extending the law, it was allowed to expire in 2020.[229]

Our focus in this paper is on private-sector use of metadata and telemetry, and here, we end our discussion of the law's evolution of government use of communications metadata. But, the government's use of such data is not limited to subpoenas and warrants; it also buys such information from the private sector.

2. The Loophole: Private Sector Collection and Sale of Communications Metadata

In the wake of 9/11, the U.S. Government had pushed collection to the limits of the law—and perhaps beyond—but there was never an indication that the bulk collection was used for purposes other than terrorism investigations.[230] There were other "bulk collections"; AT&T had vast records of call data, but government access to such records was restricted by law.[231] ECPA not only provided a legal process for government use of pen registers and trap and trace; it also governed access to subscriber information (e.g., name, address, and phone number).[232] In addition, ECPA included requirements for government access to information held by an electronic computing service (ECS), a service enabling wire or electronic communications,[233] or a remote computing service (RCS), a service providing the public with computer storage or processing using an ECS.[234]

Nonetheless, the 1986 ECPA, written well before the rise of the public Internet and the "surveillance capitalism"[235] economy that followed, left a gaping loophole. While government agents needed a court order to access subscriber information, call detail records, and other non-content data from ECSs and RCSs, there was no

229. Devlin Barrett, *Surveillance Program that Gathered Americans' Phone Data Was Illegal, Court Finds*, WASH. POST (Sept. 4, 2020, 10:09 PM), https://www.washingtonpost.com/national-security/phone-records-surveillance-edward-snowden/2020/09/02/97f26498-ed67-11ea-99a1-71343d03bc29_story.html [https://perma.cc/69SJ-UX9U].

230. *See, e.g.*, Letter from George Ellard, Inspector Gen., NSA, to Sen. Charles Grassley (Sept. 11, 2013), https://www.nsa.gov/public_info/press_room/2013/grassley_letter.pdf [https://perma.cc/EMQ4-J8YR] (discussing how, in a few instances, some NSA employees searched the database to spy on significant others, and how they were uncovered through internal investigations).

231. Off. Nat'l Drug Control Pol'y, *supra* note 147 (slide 9).

232. 18 U.S.C. § 2703(c)(2) (2019).

233. 18 U.S.C. § 2703(a); *see* 18 U.S.C. § 2510(15) (defining "electronic communications service").

234. 18 U.S.C. § 2511(2)(i).

235. *See* ZUBOFF, *supra* note 23, at 10.

restriction on how those services shared this data with other parties. ECSs and RCSs could sell the data to third parties; then the third parties could sell such information to others, including government agents. In short, government investigators could either obtain a court order for subscriber information, communications metadata, and the like—or they could just buy it from third-party data brokers.[236]

Despite an increasing number of data brokers, there is no generally agreed upon legal definition of the term.[237] The U.S. Federal Trade Commission describes data brokers as "companies that collect consumers' personal information and resell or share that information with others."[238] A definition provided by the Norwegian Consumer Council: "companies that aggregate, combine and trade massive amounts of data about consumers from a wide variety of sources, largely without consumers' knowledge" more accurately captures how data brokers operate.[239]

Data brokers provide a way for government investigators to avoid judicial oversight while obtaining information that might require a court order if requested directly. In 2020, the *Wall Street Journal* reported that the data broker Venntel sells location information from individuals' mobile devices to multiple federal agencies,[240] including to the Department of Homeland Security in

236. *See generally* Carey Shenkman et al., Ctr. for Democracy & Tech., Legal Loopholes and Data for Dollars: How Law Enforcement and Intelligence Agencies Are Buying Your Data from Brokers 5 (2021), https://cdt.org/wp-content/uploads/2021/12/2021-12-08-Legal-Loopholes-and-Data-for-Dollars-Report-final.pdf [https://perma.cc/QL7G-JZ74].

237. Aaron Rieke et al., Upturn, Data Brokers in an Open Society 3 (2016), https://www.opensocietyfoundations.org/uploads/42d529c7-a351-412e-a065-53770cf1d35e/data-brokers-in-an-open-society-20161121.pdf [https://perma.cc/4QXF-GLWR].

238. FTC, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY i (2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf [https://perma.cc/D2D5-HENR].

239. ANDREAS CLAESSON & TOR E. BJØRSTAD, NORWEGIAN CONSUMER COUNCIL, "OUT OF CONTROL"—A REVIEW OF DATA SHARING BY POPULAR MOBILE APPS (2020), https://fil.forbrukerradet.no/wp-content/uploads/2020/01/mnemonic-security-test-report-v1.0.pdfNorwegian [https://perma.cc/U2XM-KJBB]; Consumer Council, Out of Control 19 (2020), https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf [https://perma.cc/85UR-V2SF].

240. Byron Tau & Michelle Hackman, *Federal Agencies Use Cellphone Data for Immigration Enforcement*, WALL ST. J. (Feb. 7, 2020), https://www.wsj.com/articles/federal-agencies-use-cellphone-location-data-for-immigration-enforcement-11581078600 [https://perma.cc/EK26-W69N]. A December 2021 report from the Center for Democracy and Technology details substantial use of data brokers, such as to obtain location information. *See* CAREY SHENKMAN ET AL., *supra* note 236; *see also* Privacy Policy, VENNTEL, https://www.venntel.com/privacy-policy [https://perma.cc/ZWS7-KCYA].

2017, the U.S. Immigration and Customs Enforcement in 2018, and the U.S. Customs and Border Patrol in 2019.[241] Babel Street, another data broker, offers "access to a Data Feed [with] historical digital device location data."[242] Babel Street says the location data is anonymized, with an identifier tied to a phone app.[243] Such "anonymization" is unlikely to be useful; often as few as four such data points suffice to identify an individual.[244] If other data is also available—which is often the case—it is even easier to re-identify the user. Paul Rosenzweig, who served as the first Deputy Assistant Secretary for Policy at the Department of Homeland Security, stated that, "[T]he government is a commercial purchaser like anybody else. Carpenter is not relevant."[245] Rosenzweig's statement may be accurate as a point of law. But it lays out the fundamental contradiction in warrantless government acquisition of such private information, some of which the Court observed has notable "depth, breadth, and comprehensive reach."[246]

Much of the personal data that data brokers assemble has similar "depth, breadth, and comprehensive reach" that the Court describes in *Carpenter*. Data brokers largely collect and operate on user information without having a direct relationship with the user[247]—a situation that typically prevents a user from being able to control a data broker's exploitation of her personal information.[248]

241. Tau & Hackman, *supra* note 240.

242. BABEL STREET, BABEL X SERVICE DEFINITION 1–2, https://assets.digitalmarketplace.service.gov.uk/g-cloud-12/documents/712032/750265271566952-service-definition-document-2020-07-20-1357.pdf [https://perma.cc/CC4K-ALHK].

243. Charles Levinson, *Through Apps, Not Warrants, 'Locate X' Allows Federal Law Enforcement to Track Phones*, PROTOCOL (Mar. 5, 2020), https://www.protocol.com/government-buying-location-data [https://perma.cc/AH99-HQFY].

244. de Montjoye et al., *supra* note 57, at 1.

245. Tau & Hackman, *supra* note 240; *but see*, Matthew Tokson, *Government Purchases of Sensitive Private Data*, DORF ON L. (Mar. 29, 2021, 8:00 AM), http://www.dorfonlaw.org/2021/03/government-purchases-of-sensitive.html [https://perma.cc/W2HQ-9MAP] (expressing a more privacy protective viewpoint); *see also* Orin Kerr, *Buying Data and the Fourth Amendment*, HOOVER INST.: AEGIS SERIES PAPERS, No. 2109, 2021, at 1, 1, https://www.hoover.org/sites/default/files/research/docs/kerr_webreadypdf.pdf [https://perma.cc/W3B4-N7XS] (agreeing that in some circumstances a more restrictive approach may be appropriate).

246. Carpenter v. United States, 138 S. Ct. 2206, 2223 (2018).

247. MAJORITY STAFF OF S. COMM. ON COM., SCI., AND TRANSP., 112TH CONG., A REVIEW OF THE DATA BROKER INDUSTRY: COLLECTION, USE, AND SALE OF CONSUMER DATA FOR MARKETING PURPOSES 2 (2013).

248. *Id.* at 3. The situation can be even worse: for a user of an app employing the LeanPlum adtech company, the user's acceptance of the app's privacy policy is an implicit

C. In IP-Based Communications, the Content/Non-Content Distinction Has Disappeared

In most of the world, the public telephone network relies on a protocol called Signaling System 7, first developed in 1975.[249] Call set-up and disconnect information travels on one channel—the Call Data Channel (CDC)—while the actual voice communication—is transmitted on the Call Content Channel.[250] The PSTN is a "circuit-switched network"; the CDC enables the establishment of a fixed circuit that is used for the duration of the communication.[251] Such an architecture makes a great deal of sense for a phone call, where people expect minimal time to elapse between the end of one person speaking and the beginning of the other person's response.[252] Thus, it is important that the communications channel always be open to transmit. This architecture, which consumes resources even when there is a pause in communication, is much less appropriate if the communication does not need to arrive in a particularly timely manner, such as for a large data file that does not need to be read immediately.[253]

In the 1960s, researcher Donald Davies sought to solve the problem of how to efficiently transmit such relatively time-insensitive communications.[254] His solution was packet-switching: splitting the communication into many small "packets" that can be routed independently over the network, then reassembled at the other end. This idea was the genesis of the Internet, which still, of

acceptance of LeanPlum's privacy policy. *Terms of Service*, LEANPLUM, [https://perma.cc/FUZ6-QW7R] (Dec. 3, 2017). *See* discussion *infra* Sections IV.B, IV.C.

249. T.J. Cieslak et al., *No.4 ESS: Software Organization and Basic Call Handling*, 56 BELL SYS. TECH. J. 1113 (Sept. 1977).

250. Micah Sherr et al., *Can They Hear Me Now? A Security Analysis of Law Enforcement Wiretaps*, 2009 ACM CONF. ON COMPUT. & COMMC'N SEC. 512, 514 (discussing how the separation of these two channels enabled faster call establishment, better network management, more complex billing systems (e.g., 800-number "free" phone calling) and text messaging).

251. TECH. ADVISORY COUNCIL, FCC, POST-PSTN PUBLIC COMMUNICATIONS RESILIENCY (2013), https://transition.fcc.gov/oet/tac/tacdocs/reports/2013/Resiliency_White_Paper-FCC_TAC-2013-FINAL_working_group_paper.pdf [https://perma.cc/A9FP-YQTQ].

252. *See* Tanya Stivers et al., *Universals and Cultural Variation in Turn-Taking in Conversation,* 106 PROC. NAT'L ACAD. SCIS. 10587, 10589 (2009) (establishing that the typical gap between speakers in conversation is 200 milliseconds).

253. This could be NASA shots of the moon, X-rays to be read remotely several hours later, etc.

254. Cade Metz, *Why Do We Call Them Internet Packets? His Name Was Donald Davies*, WIRED (Sept. 10, 2012, 6:38 PM), https://www.wired.com/2012/09/donald-davies/ [https://www.wired.com/2012/09/donald-davies]. Paul Baran of the RAND Corporation also came with the idea of packet-switched networks, although his motivation was somewhat different; Leonard Kleinrock, a graduate student at UCLA, independently studied related problems that were also critical for the development of the Internet. *Id*.

course, needed the development of various technologies to enable such transmissions. Between 1900 and the 1980s, the cost of maintaining a single voice circuit for a minute along a mile of a long-distance phone circuit dropped by a factor of a million;[255] fiber-optic cables were introduced in the 1980s, which reduced transmission costs significantly more. Meanwhile, the development of low-cost home computers provided an economic incentive since they could be widely available at the network's endpoints. These changes were essential to the development of the Internet.

The two networks, PSTN and the Internet, share many features. They use similar transmission facilities; they use electronic switches and routers; they are digitally switched, and they seek to maximize the number of customers served while minimizing the cost to do so.[256] They have, however, different approaches to transmission and connection. The PSTN uses expensive, highly reliable switches, while the Internet historically used less expensive routers and focused on "best-effort" delivery (the latter has changed as the Internet has become the communication mode of choice for industry).[257] The other distinction is the nature of the connection. The PSTN establishes a circuit for a phone call which is used until the call ends. Internet Protocol (IP) based communications operate differently. At least in theory, different packets of an IP communication may take different routes to the destination. Then the communication—the webpage, the email, a voice call sent over the Internet (VoIP for Voice over IP)—is reassembled. Packet headers read by different layers of the Internet "stack" contain information that enable these different layers to send and reassemble packets so that the applications—webpages, email, and the like—can be successfully used at the receiving end.

Post the *Katz/Smith* decisions, there has been significant legal scholarship on the content/non-content distinction; this issue became especially important after mobile phones introduced the ability to track a user's location and IP-based communications introduced richer types of "non-content" information enabling discovery of attributes of a user's actions. Though we focus on private-sector use of metadata and telemetry information, the

255. AT&T BELL LABORATORIES, A HISTORY OF ENGINEERING AND SCIENCE IN THE BELL SYSTEM: TRANSMISSION TECHNOLOGY, 1925-1975 779 (E.F. O'Neill ed., 1985) (Fig. 24-4).

256. Steven Bellovin et al., *Security Implications of Applying the Communications Assistance to Law Enforcement Act to Voice over IP*, ITAA, June 13, 2006, at 1, 9, https://privacyink.org/pdf/CALEAVOIPreport.pdf [https://perma.cc/5JV2-EZT8].

257. *Id.*

discussion regarding whether certain types of data are content or not is relevant, regardless of who accesses and uses the data.

In 1879, in *Ex Parte Jackson*, the Supreme Court ruled that "The constitutional guaranty of the right of the people to be secure in their papers against unreasonable searches and seizures extends to their papers, thus closed against inspection, wherever they may be."[258] The Court, then, noted, "Whilst in the mail, they can only be opened and examined under like warrant."[259] The outside of the package, in which the addressing was written, was open to "examination and inspection" searching inside required a warrant.[260]

Telephone communications presented a new challenge to the courts. The Wiretap Act originally stated that content, "includes any information concerning the identity of the parties to such communication or the existence, substance, purport, or meaning of that communication";[261] this was amended by ECPA to "'[C]ontents', when used with respect to any wire, oral, or electronic communication, includes any information concerning the substance, purport, or meaning of that communication."[262] While the voice communication was clearly content, was anything else also?

As we saw in Fitzgerald's investigations of the first World Trade Center bombing, it was possible to infer the "substance, purport, and meaning of a communication" simply from the timing of communications, a situation that was somewhat different than when communications transited by post—and, therefore, did not arrive with the speed of a conversation.[263] Even before the advent of richer modalities, such as provided by mobile phones and IP-based communications, the actual meaning of the Wiretap Act's definition of content presented problems.

Orin Kerr noted that the addressing and signaling information conveyed to the phone company during dialing are content; they are content conveyed to the telephone company.[264] David McPhie pointed out gaps in the meaning of content, writing that "the exact relationship between the positive and negative definitions of

258. *Ex Parte* Jackson, 96 U.S. 727, 733 (1878).

259. *Id.*

260. *Id.*

261. Omnibus Crime Control and Safe Streets Act of 1968, Pub. L. No. 90-351, tit. III, §2510(8), 82 Stat. 197, 213.

262. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, tit. I, § 101(a)(5), 100 Stat. at 1848.

263. For more examples, *see supra* Section IV.A.

264. Orin S. Kerr, *Internet Surveillance Law After the USA PATRIOT ACT: The Big Brother That Isn't*, 97 NW. U.L. REV. 607, 646 (2003).

'content' (substance and meaning versus addressing or signaling data) is unclear."[265] Does it include *all* data that is not addressing or signaling information?[266] Does some data fall in both categories?[267] Does some fall in neither?[268] McPhie also discussed the situation in which the mere existence of a call provided by dialing information can, in fact, disclose the content of the communication (e.g., if a 900 number is called).[269]

Matthew Tokson left the issue of whether the content/non-content distinction made in *Ex Parte Jackson* sufficiently protects the privacy of Internet communications as "a question for another day"[270] and, instead, focused on developing "a legal framework for distinguishing content from [non-content] envelope information"[271] for two types of Internet communications: the email to/from addresses and URLs. He proposed that "electronic information that can reveal the underlying text or subject matter of an Internet communication must be classified as content."[272]

In considering how wiretap law is affected by the shift from the PSTN to IP-networks, Steven Bellovin, Matt Blaze, Susan Landau, and Stephanie Pell showed that the Dialing, Routing, Addressing, and Signaling (DRAS) information of the PSTN—the language used in the PATRIOT Act to handle pen/trap collection[273]—fails to map onto the architecture of the Internet. The PSTN and the Internet use different communication protocols to transmit information; these protocols differ in what parts of the network see what aspects of the communication metadata.

To show how some of the information used for transmitting content over the Internet may reveal the content of the communication, we provide a brief explanation of Internet routing and packet headers.[274]

265. McPhie, *supra* note 51, at 9.
266. *Id.*
267. *Id.*
268. *Id.*
269. *Id.* at 10.
270. Tokson, *supra* note 51, at 2112.
271. *Id.* at 2105.
272. *Id.*
273. 18 U.S.C. § 3127(3) (2012); Bellovin et al., *supra* note 50, at 46–52 (2016). The authors also have a longer discussion of the legislative history involving Dialing, Routing, Addressing, and Signaling in the context of pen/trap. *Id.* at 12–19.
274. *Id.* at 36–44 (providing a deeper explanation of this and its implications for U.S. surveillance law).

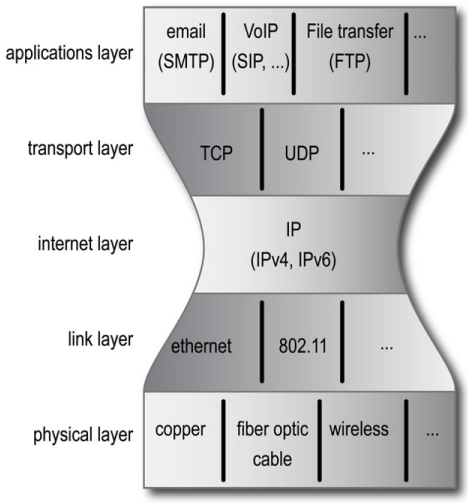

Figure 1[275]

The standard way to describe Internet communications is as a protocol stack. At the bottom is the physical layer, which consists of wires, fiber optic cables, etc.; at the top is the set of applications such as email, VoIP, file transfer, web access, and the like. Each of these layers provides services to the layer above and receives services from the layer below. The physical layer is the set of wires or radio signals that transmit the content; the communications metadata is not part of the set of issues in this layer and so we do not discuss it further. The link layer operates within a single network (e.g., a Local Area Network, a Wi-Fi network, etc.) and, thus, largely does not access communications metadata of the sender or receiver.[276] Our interest is in access and use of user communications metadata, and thus our focus is on the remaining layers: Internet (IP), transport, and application.[277] IP's role is to transmit packets from the source computer to the destination machine, while the transport layer's job is to reassemble the packets for use by the application.

275. Nancy Snyder, Figure 2.2: Internet Protocol Stack, *in* SUSAN LANDAU, SURVEILLANCE OR SECURITY? THE RISKS POSED BY NEW WIRETAPPING TECHNOLOGIES 23 (2011) (illustration) (reprinted with permission from Susan Landau).

276. *Id.* at 40 (describing some exceptions to the lack of access to communications metadata of the sender or receiver).

277. *Id.* at 37–39.

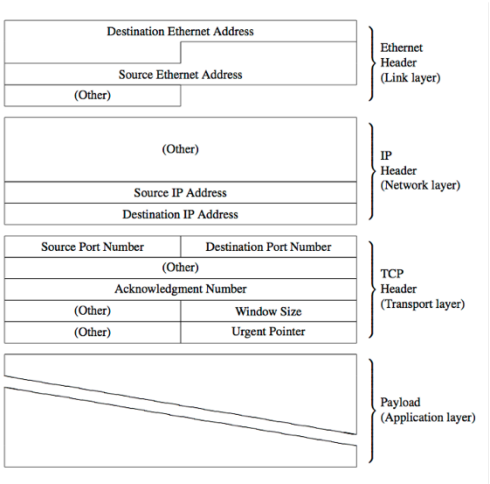

Figure 2[278]

Bellovin et al. noted that, "All layers except the physical and application layers consist of a 'header' and a 'payload.'"[279] We are interested in the information in the headers, for that is processed by that layer in the stack, while the packet payload is passed on to the next higher layer on the protocol stack for processing. Because we focus on what ISPs and applications can learn from metadata and telemetry information, we consider what is learned in the network and transport layers.

The Transport Communication Protocol (TCP), the most common of the transport layer protocols, is an "end-to-end" protocol; only the recipient receives the TCP header information. The TCP header contains port numbers, which are addresses within the source and destination machine. Certain port numbers are standardly used for certain types of content, and so knowing the port number provides insight into the type of content being transmitted (email, webpage, etc.). Other fields in the TCP packet header can be used to digitally fingerprint an operating system.[280]

The transport layer is not the only layer which accesses the port number. Though the Internet Protocol is simply responsible for delivering the packet to its destination machine—and thus IP headers contains the source and destination IP addresses, but not port numbers—as Bellovin et al. observe, ISPs access port

278. First published in *id.* at 39; *see* Jon Postel, INTERNET PROTOCOL (RFC 791) (1981) at 11, https://datatracker.ietf.org/doc/rfc791/ [https://perma.cc/X685-N6GN]; *see also* JON POSTEL, TRANSMISSION CONTROL PROTOCOL (RFC 793) (1981) at 15.

279. Bellovin et al., *supra* note 50, at 38.

280. *See, e.g.*, Toby Miller, *Passive OS Fingerprinting: Details and Techniques*, [https://perma.cc/2498-4MXH].

numbers.[281] One example of this is Cisco's NetFlow product.[282] Use of this product enables ISPs to better manage their traffic—and, through knowing the port numbers, to learn what applications users are employing, information to which ISPs should not necessarily have access.

Bellovin et al. distinguish "between two types of content, 'communicative content' and 'architectural content.'"[283] The former is the familiar type of "communicative content. . . predicated upon the semantic meaning of the communication itself," while the latter is content that enables different layers of the Internet to transport application data.[284]

With that paradigm in hand, Bellovin et al. examined various types of IP-based applications to determine what parts of a packet header were effectively content. Their answer was, "It's complicated." Whether an email address—as opposed to the IP address to which the mail was delivered—is addressing information depends on whether the recipient is using a third-party mail server, shared mail server, or private mail server.[285] The URLs https://maps.google.com and https://google.com/maps effectively function the same from the user's point of view, but the former is an address (DRAS), while the latter includes a communication to the Google server ("serve the user content from the 'maps' server") that is not.[286] From this and a variety of other examples, Bellovin et al. concluded that the ability to distinguish when information is shared with a third party is effectively impossible.[287] The complexity of IP-based communication makes the content/non-content "functionally meaningless."[288] Though Bellovin et al. were writing in the context of the *Katz/Smith* distinction, the failure of a meaningful distinction between content and non-content means that a user, who has no control over what is contained in metadata, supplies content without awareness of having done so. With this background in hand, we now turn to look at a new type of data largely unknowingly transmitted by users: software and device telemetry.

281. Bellovin et al., *supra* note 50, at 48.

282. *Netflow Services Solutions Guide,* CISCO SYS. INC. (July 31, 2001), [https://perma.cc/WX3T-EMX7].

283. Bellovin et al., *supra* note 50, at 32.

284. *Id.* (discussing how the paradigm of communications content/architectural content generalizes the example of the telephone operator provided by Kerr).

285. *Id.* at 57–64.

286. *Id.* at 69–73.

287. *Id.*

288. *Id.* at 5.

D. Newer Forms of Non-Content Collection: Telemetry Information

The early PSTN concentrated the intelligence of the communication system into the phone switches; the phones themselves were devices that a customer used to make a call or answer one—that was it.[289] This focus continued through the time of *Smith*, though by then there were small changes in the model and some "intelligent" capabilities in the phones themselves.[290] Until smartphones changed telephone functionality, a phone remained largely a device for making and receiving calls. Other changes were needed before mobile phones could become the multi-purpose devices they are today. Early mobile phones were analogue; by the 1990s second generation (2G) mobile wireless carriers built their systems based on digital technologies.[291] This system provided the necessary infrastructure for smartphones, which Apple introduced a little over a decade later.

In Section I.A, we explored how AT&T started collecting communications metadata to measure performance. Fully electronic telephone switches and reliable magnetic tape storage enabled the company to inexpensively measure and record data about individual calls;[292] aggregate data was used to determine quality of service. At first, AT&T used the individualized CDRs for billing and forecasting future services. With time, however, the company found other purposes for the data, including fraud detection.[293] Later, AT&T offered additional types of telephone services for customers based on their patterns of use.[294] All uses, whether measuring systems performance, customer billing, or proffering new forms of telephone service, involved providing core communication services for which the customer has contracted—or for improving those services.[295] Mobile devices provide far more opportunity for data collection than the landline systems that preceded them.

Predicting and running tests on how someone will use a product is one thing but observing it in use at scale is quite another. As the world transitioned to network-based systems—think "software as a service"—engineers began collecting performance

289. *Id.* at 34.

290. Bellovin et al., *supra* note 50, at 34.

291. JONATHAN E. NUECHTERLEIN & PHILIP J. WEISER, DIGITAL CROSSROADS: TELECOMMUNICATIONS LAW AND POLICY IN THE INTERNET AGE 134 (2013).

292. A.E. JOEL, JR. ET AL., *supra* note 14, at 141, 260.

293. *See* discussion *supra* Section II.A.

294. *See* discussion *supra* Section II.

295. FTC, *supra* note 60, at ii.

data from remote-running applications.[296] Studying use of early versions of products "in the wild" revealed performance bottlenecks, uncovered software errors, and disclosed how customers were actually employing the technology; it was thus invaluable in multiple ways. Collecting and analyzing such data, both user and system events, became the standard stock in trade for companies providing networked software.[297]

Telemetry enables developers to learn about customer use of the product. Is the phone set up the right way? Is the home screen set up to serve users well? Or do users almost always immediately track elsewhere? What was the user doing when a device crashed? What sequence of inputs led the system to operate slower than expected? Are the techniques the OS put in to prevent fraud succeeding? Might another technique be better?

Telemetry enables online businesses to do much of the same. They use metrics (e.g., time taken for a response, how many requests were made, value of responses, etc.), events (did a user click a button or go to checkout?), logs (the steps the software performed), and tracing, otherwise known as a user's "journey" through an application (e.g., from clicking on "add to cart" to the order being completed), to learn how well an application is functioning and whether users are receiving the service they expect.[298]

Smartphones are full of sensors. That provides another aspect of telemetry. Smartphones have a battery sensor to ensure a user knows when the device battery is low in power.[299] They have an accelerometer that measures how quickly your phone is accelerating in three dimensions and a gyroscope to measure the angular velocity (the speed in which it is tilting in all three directions) to keep the picture steady on the screen as a user moves her device to show something on the screen to another person. Smartphones have a GPS receiver, enabling mapping applications to inform users of their location. Many smartphones have fingerprint sensors for use in authenticating a user. Smartphones have magnetometers, allowing the phones to point north on maps regardless of the phone's orientation. They have ambient light sensors, which dim the screen if it is dark out, making it easier to

296. Robert Musson et al., *Leveraging the Crowd: How 48,000 Users Helped Improve Lync Performance*, 30 IEEE SOFTWARE 38, 38 (2013).

297. *See, e.g.*, Barik et al., *supra* note 26, at 92.

298. Erwan Paccard, *What Is OpenTelemetry and Why Should You Care?*, APPDYNAMICS (Feb. 16, 2022), https://www.appdynamics.com/blog/product/what-is-opentelemetry/ [https://perma.cc/5LDR-8WXL].

299. Ahmad Rahmati & Lin Zhong, *Human-Battery Interaction on Mobile Phones,* 5 PERVASIVE AND MOBILE COMPUTING 465, 465 (Oct. 1, 2009).

read while simultaneously saving power. Smartphones have proximity sensors that shut off a smartphone's display when the device is close to a person's face; this allows the device to ignore unintentional taps on the screen. Most importantly, smartphones have touch sensors.[300]

These sensors enable phones to work correctly, provide more capabilities, and conserve battery usage. They allow a user to show a picture on her phone to a friend even while her movement changes the device's orientation, and they simplify a user's ability to read a map while traveling down a bumpy road. They are an integral part of making a smartphone "smart."

Increasingly, however, sensor information is collected off the phone, almost always for different purposes than they are used for on the phone. Accelerometers and gyroscopes do more than allow the proper display of content; that same ability to track orientation when displaying a photo also allows a user's actions in gaming applications to display properly on the device as the phone moves. That means developers must be able to access information from accelerometers and gyroscopes.[301] Information from a battery sensor can also be used by a video application to lower battery usage when power is low.[302]

The industry has been active in discovering what kind of information can be distilled from telemetry data. It seems that intelligence services have also explored this area,[303] though there is nothing public about U.S. Government efforts (nor does there appear to be public information about other nations' efforts). But if, in general, the collection and use of telemetry information is well known to software engineers and developers, it is made notably less clear to consumers. Public knowledge of this form of data collection is complicated by the fact that named different providers call the data by different names. Microsoft uses "records that capture system and user events" as a working definition of telemetry data.[304] Google sometimes names the information telemetry[305] and

300. David Nield, *All the Sensors in Your Smartphone, and How They Work*, GIZMODO (June 29, 2020), https://gizmodo.com/all-the-sensors-in-your-smartphone-and-how-they-work-1797121002 [https://perma.cc/GYC9-NVG3].

301. *See* discussion *infra* Section III.B.

302. Portable Device with Priority Based Power Savings Controls and Method Thereof, US Patent No. 8,135,443 B2 fig. 9 (filed Aug. 21, 2006) (granted Mar. 13, 2012).

303. Private communications between Susan Landau and three anonymous sources (June 2, 2022; June 16, 2022; and July 1, 2022).

304. *See, e.g.*, Barik et al., *supra* note 26, at 92.

305. *Google API for Exposure Notifications*, *Note on General Android Platform Telemetry*, GOOGLE, https://developers.google.com/android/exposure-notifications/telemetry-design, [https://perma.cc/TV2E-QBJ4].

sometimes usage and diagnostics information,[306] while Apple calls this data analytics.

Google tells developers how to collect information that will inform about an "app's stability, performance, battery usage, and more."[307] So do Apple and Microsoft. Google does not publicly state what telemetry information it collects. But as a company that measures everything,[308] its telemetry collection is bound to be extensive. Google participates in so many parts of the ecosystem: as a device developer (Chromebooks), OS developer (Android), a browser developer (Chrome), and provider of multiple widely used apps (search, Gmail, Maps, Docs, Sheets, YouTube, Calendar, Drive, and others). The company collects telemetry information on all of these. Some of that information is used for ensuring the particular system is working properly, some that the set of systems are doing so,[309] and sometimes the information is undoubtedly used to improve products.

Apple, whose products are also focused on the consumer market, is far more explicit than Google about explaining privacy controls, stating, "None of the collected information identifies you personally . . . Personal data is either not logged at all, is subject to privacy preserving techniques such as differential privacy, or is removed from any reports before they're sent to Apple."[310]

Telemetry enables app developers to debug and improve their products, but it also has serious impacts on privacy. Telemetry can give developers deep insight into how customers are responding to an app, information they may combine with other data to target particular customers to personalize ads. In the late 2010s, for example, Microsoft ran afoul of the European Union's General Data Protection Directive for data it was collecting from Windows users.

306. *Google Account Help*, GOOGLE, https://support.google.com/accounts/answer/6078260 [https://perma.cc/A3KS-FNSL].

307. *Monitor Your App's Technical Quality with Android Vitals*, GOOGLE: PLAY CONSOLE HELP, https://support.google.com/googleplay/android-developer/answer/9844486?hl=en#zippy=%2Creview-the-overview-dashboard-and-detailed-metric-pages [https://perma.cc/W6LX-JCKR].

308. *How Google's Opaque Packaging Resulted in 3.1 Million Fewer Calories Consumed Over 7 Weeks*, DECISION LAB, https://thedecisionlab.com/intervention/how-googles-opaque-packaging-resulted-in-3-1-million-fewer-calories-consumed-over-7-weeks [https://perma.cc/VVX5-EQKY].

309. Google uses its knowledge about a user's behavior to track unusual logins and possible account compromise. *See generally Protect Your Account if There's Unfamiliar Activity*, GOOGLE: GOOGLE ACCOUNT HELP, https://support.google.com/accounts/answer/7305876?hl=en [https://perma.cc/3VQS-UF7J].

310. *Device Analytics and Privacy*, APPLE: LEGAL, https://www.apple.com/legal/privacy/data/en/device-analytics/ [https://perma.cc/N8S2-RD4S].

This misstep included information on "use of an app of an online casino, of a Turkish newspaper, of a magazine targeted at gay people, an app that indicates Islamite prayer times, an app collection details about a woman's pregnancy and an app targeted at diabetes patients."[311] The German Federal Office for Information Security produced a report on how to disable sending Microsoft data from Office and other applications.[312] Recently published Microsoft webpages on the company's use of telemetry emphasized the value of the data for finding errors, determining compatibility of a user's device for updates, and understanding performance.[313] There was little discussion of privacy on those pages.

It is hard for the average consumer to know that these large companies are collecting telemetry information, let alone what they do with it. In 2021, by looking at the telemetry communications of iPhones and Androids, Douglas Leith provided some insight on such collection by Apple and Google.[314] Leith observed that when a SIM card is inserted into an iPhone, iOS sends the MAC address of nearby routers; this is likely to often include the MAC address of a home gateway router.[315] Leith observed that, as long as any device in the area has GPS location turned on, the home gateway router will be tagged with a physical location; and thus, so will any other

311. AUTORITEIT PERSOONSGEGEVENS (DUTCH DPA), SUMMARY OF INVESTIGATION REPORT (PUBLIC VERSION): MICROSOFT WINDOWS 10 HOME AND PRO INVESTIGATION BY THE AUTORITEIT PERSOONSGEGEVENS (DUTCH DPA) 2 (2017), https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/public_version_dutch_dpa_informal_translation_summary_of_investigation_report.pdf [https://perma.cc/95Z4-Q7FR].

312. *See generally* ALEKSANDER MILENKOSKI, ERNW & FED. OFF. FOR INFO. SEC., MICROSOFT OFFICE TELEMETRY: ANALYSIS REPORT, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Studien/Office_Telemetrie/Office_Telemetrie.pdf?__blob=publicationFile&v=5 [https://perma.cc/LL9V-VS78].

313. *Diagnostics, Privacy, and Feedback in Windows*, MICROSOFT: MICROSOFT SUPPORT, https://support.microsoft.com/en-us/windows/diagnostics-feedback-and-privacy-in-windows-28808a2b-a31b-dd73-dcd3-4559a5199319 [https://perma.cc/6C69-Q64F]. Note that this information does not appear to have been provided until 2021.

314. *See generally* Douglas J. Leith, *Mobile Handset Privacy: Measuring the Data iOS and Android Send to Apple and Google*, 399 LECTURE NOTES INST. FOR COMPUT. SCIS., SOC. INFORMATICS, & TELECOMMS. ENG'G 231 (2021) [hereinafter *Mobile Handset Privacy*]; Leith also studied the telemetry communications of six major browsers—Chrome, Firefox, Safari, Brave, Edge and Yandex—and observed that some browser identifiers persist over unexpectedly long timespans, enabling tracking a device over a long period. Leith characterized the identifiers as being (i) ephemeral; (ii) session identifiers, reset on browser restart; (iii) browser instance identifiers set on installation, and (iv) device identifiers. He found that Brave used only ephemeral identifiers, while Chrome, Firefox, and Safari used session and browser instance, and Yandex, device identifiers. *See* Douglas J. Leith, *Web Browser Privacy: What Do Browsers Say When They Phone Home?*, 9 IEEE ACCESS 41615 (2021), https://ieeexplore.ieee.org/document/9374407 [https://perma.cc/948U-SWZH].

315. *Mobile Handset Privacy*, *supra* note 314, at 231–32.

device using that router.[316] Leith further found that idle iPhones connect to Apple back-end servers every 264 seconds on average, while idle Androids do so every 255 seconds.[317] When the phones do so, they reveal their IP address, a proxy for location.[318] The phones also reveal their IMEI, SIM serial number, and IMSI to the OS manufacturer.[319]

Leith observed that Google Messages and Google Dialer on Android report to Google when phone messages and texts are received and sent; the communications to Google also include time and duration of a call and hash of the message.[320] All but the last enables Google to determine who is communicating with whom, a record of which, were the communications over the PSTN, would be stored at the telecommunications carrier; this collection and storage is not surprising. In some instances, however, Google maintains more information about a communication's content than its predecessor, AT&T, did. The storage of a message hash raises concerns, for it provides the potential to leak data about a message's content.

When a manufacturer collects product usage information to improve the product, the manufacturers would not be acting differently than AT&T did when the company monitored the networks to determine quality of service.[321] But, it appears that software and device telemetry information is also used for other, less user-friendly, purposes, an issue we tackle in Section II.

II. WHAT METADATA AND TELEMETRY CAN REVEAL[322]

Governments were the first to exploit the information provided by communications traffic. A century before the Snowden disclosures, the military used communications metadata to track the enemy. In 1904, during the Russo-Japanese war, Japanese warships used intercepted radio messages from the czar's fleet to determine enemy whereabouts—and defeat them.[323] The Japanese

316. *Id.* at 246.
317. *Id.* at 233.
318. *Id.*
319. *Id.*
320. Douglas J. Leith, *What Data Do the Google Dialer and Messages App Send to Google?*, 462 LECTURE NOTES INST. FOR COMPUT. SCIS., SOC. INFORMATICS, & TELECOMMS. ENG'G 549, 549 (2022).
321. *See* discussion *supra* Section II.A.
322. *See generally* Susan Landau, *supra* note 84.
323. Bartholomew Lee, *Wireless–Its Evolution from Mysterious Wonder to Weapon of War, 1902 to 1905*, 25 ANTIQUE WIRELESS ASS'N 1, 1–2 (2012), https://www.californiahistoricalradio.com/wp-

were not decrypting the communications; instead, they used metadata to determine location. During World War I, the French used signal strength to map locations of German military radio stations.[324] Because traffic analysis provides the broadest picture of an adversary's activities, it became the backbone of communications intelligence agencies' work.[325]

In the aftermath of the 2013 Snowden disclosures, former NSA General Counsel Stewart Baker, described the revelatory nature of metadata. "Metadata absolutely tells you everything about somebody's life," he said. "If you have enough metadata you don't really need content … [It's] sort of embarrassing how predictable we are as human beings."[326] Such a statement casts doubt on the reasoning behind *Smith*; if the NSA clearly saw communications metadata as providing "the substance, purport, or meaning,"[327] then surely the data deserved protection as content. As one of the two authors of this paper noted in 2008, "[t]ransactional information is remarkably revelatory."[328] But, until *Carpenter*, the courts did not see it that way. And, at present, *Carpenter* applies to seven-days' worth of collection of CSLI.[329]

In many instances, it is not necessary to look at content to discern people's activities—or their interests.[330] After the Snowden disclosures, three Stanford researchers, Jonathan Mayer, Patrick Mutchler, and John Mitchell, used an Android app to collect eight-months' worth of telephone metadata of 823 volunteer participants.[331] The dataset included 62,229 unique phone

content/uploads/2013/01/BartWirelessWar190205Lee.pdf [https://perma.cc/ZR8G-HUBD].

324. The French collected "callsigns"; these denoted which station was sending the communication. From this the French were able to determine the deployment of the German troops. *See* DAVID KAHN, THE CODEBREAKERS: THE COMPREHENSIVE HISTORY OF SECRET COMMUNICATION FROM ANCIENT TIMES TO THE INTERNET 300 (1996).

325. WHITFIELD DIFFIE & SUSAN LANDAU, PRIVACY ON THE LINE: THE POLITICS OF WIRETAPPING AND ENCRYPTION 102 (rev. and updated ed., 2007).

326. Alan Rusbridger, *The Snowden Leaks and the Public*, N.Y. REV. BOOKS (Nov. 21, 2013), https://www.nybooks.com/articles/2013/11/21/snowden-leaks-and-public/ [https://perma.cc/3SPT-76WR]. A few months after Baker's remark, former NSA Director Michael Hayden said, "[w]e kill people because of metadata." Johns Hopkins University, *The Johns Hopkins Foreign Affairs Symposium Presents: The Price of Privacy: Re-Evaluating the NSA*, YOUTUBE (Apr. 7, 2014), https://www.youtube.com/watch?v=kV2HDM86XgI [https://perma.cc/UVW5-DDRQ].

327. 18 U.S.C. § 2510(8).

328. Siobhan Gorman, *NSA's Domestic Spying Grows as Agency Sweeps Up Data*, WALL ST. J. (Mar. 10, 2008, 12:01 AM), https://www.wsj.com/articles/SB120511973377523845 [https://perma.cc/J8Z5-FY62].

329. Carpenter v. United States, 138 S. Ct. 2206, n.3 (2018).

330. Gorman, *supra* note 328.

331. Jonathan Mayer et al., *Evaluating the Privacy Properties of Telephone Metadata*, 113 PROC. NAT'L ACAD. SCIS. 5536, 5536 (May 16, 2016) [hereinafter

numbers, based on a quarter of a million calls and 1.2 million text messages.[332] Using thirty thousand numbers from their dataset and querying public interfaces at Yelp, Google, and Facebook, the researchers were able to reidentify 32 percent of the users.[333] In many cases, Mayer et al. uncovered quite personal information about individual participants. They discovered, for instance, that "Participant B received a long phone call from the cardiology group at a regional medical center, talked briefly with a medical laboratory, answered several short calls from a local drugstore, and made brief calls to a self-reporting hotline for a cardiac arrhythmia monitoring device [while] Participant D placed calls to a hardware outlet, locksmiths, a hydroponics store, and a head shop in under 3 weeks."[334] From such information, it is easy to surmise personal activities occurring in Participants B and D's lives. Similarly, the apps you use—readily identifiable through their IP addresses—are likely to reveal your gender, age, race (white versus non-white) and marital status.[335]

Used in aggregate, communications metadata can provide real-time information about social characteristics of groups, including religion, economic status, and organizational structure (formal and informal). Communications metadata can monitor social movements and expose the social fracturing that occurs during political upheaval.[336] Communication metadata can also reveal information about an individual, including the device[337] and person using it,[338] as well as substantial personal and intimate

Evaluating Privacy Properties], https://www.pnas.org/doi/epdf/10.1073/pnas.1508081113 [https://perma.cc/9ECX-F6VN]; *see* Jonathan Mayer et al., *Supporting Information* (May 16, 2016) [hereinafter *Supporting Information*], https://www.pnas.org/action/downloadSupplement?doi=10.1073%2Fpnas.1508081113&file=pnas.201508081SI.pdf[https://perma.cc/N2W8-RGCD].

332. *Supporting Information*, *supra* note 331, at 1.

333. *Evaluating Privacy Properties*, *supra* note 331, at 5538.

334. *Id.* at 5540.

335. Eric Malmi & Ingmar Weber, *You Are What Apps You Use: Demographic Prediction Based on User's Apps*, TENTH INT'L AAAI CONF. WEB & SOC. MEDIA, Feb. 29, 2016, at 2, https://arxiv.org/pdf/1603.00059.pdf [https://perma.cc/TLR4-QPA9].

336. Cellphone metadata revealed that a year after the highly divisive 2016 presidential election people who were in "opposite" precincts (that is, precincts that had voted differently for president) spent fifty minutes less time together at the Thanksgiving meal than they had done the year before; *see* M. Keith Chen & Ryne Rohla, *The Effect of Partisanship and Political Advertising on Close Family Ties*, 360 SCIENCE 1020, 1020 (June 1, 2018), https://www.science.org/doi/10.1126/science.aaq1433 [https://perma.cc/8TGJ-PDDE].

337. Irene Amerini et al., *Smartpone Fingerprinting Combining Features of On-Board Sensors*, 12 IEEE TRANSACTIONS INFO. FORENSICS & SEC. 2457, 2457 (2017).

338. Yves-Alexandre de Montjoye et al., *Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata*, 347 SCIENCE 536, 536 (2015).

information about that user, from their sleeping patterns,[339] sexual orientation,[340] to their gender, race, and marital status.[341]

Telemetry enables tracking of users' activities and attributes in unexpected ways. For example, data from a smartphone magnetometer, gyroscope, and accelerometer make it possible to track a user's location.[342] Uber, indeed, has a patent to use sensor data from an accelerometer, altimeter, GPS, or gyroscope from a passenger's phone to track the driver's style.[343] This, of course, could be determined from data collected from the driver's phone, obviating the necessity of obtaining the information elsewhere. Given our knowledge of the company's business practices,[344] there is reason to wonder whether Uber might collect or use the data for other, less user-friendly, purposes as well. Various Internet

339. Landau, *supra* note 84, at 13.

340. *See* Min Joo Kim, *Tracing South Korea's Latest Virus Outbreak Shoves LGBTQ Community into Unwelcome Spotlight*, WASH. POST (May 11, 2020), https://www.washingtonpost.com/world/asia_pacific/tracing-south-koreas-latest-virus-outbreak-shoves-lgbtq-community-into-unwelcome-spotlight/2020/05/11/0da09036-9343-11ea-87a3-22d324235636_story.html [https://perma.cc/XYS6-W4BJ] (discussing how, in May 2020, a twenty-nine-year-old man infected with COVID-19 caused an outbreak of the disease in Itaewon, Seoul, a neighborhood known for its gay venues. South Korea was publishing the locations where infected people had been; a positive SARS-CoV-2 test result shortly after this incident was often seen as revealing previously unknown sexual identity).

341. *See* Malmi & Weber, *supra* note 335, at 2–3 (using a sample of 3,760 users, researchers were able to determine user gender based on app usage with 82.3%, white versus non-white with 72.7% accuracy, and marital status with 72.5% accuracy).

342. Narain et al., *supra* note 52, at 397.

343. Augmenting Transport Services Using Driver Profiling, U.S. Patent No. 2019/0139450 A1 col. 1 [0033] (filed Jan. 3, 2019) (issued Feb. 12, 2019).

344. *See* B. Voytek, *Rides of Glory*, UBER BLOG (Mar. 26, 2012), https://web.archive.org/web/20140827195715/http:/blog.uber.com/ridesofglory [https://perma.cc/5CS8-CJTG]; *see also Takeaways from the Uber Files Investigation,* WASH. POST (July 11, 2022), https://www.washingtonpost.com/business/2022/07/10/uber-files-explained [https://perma.cc/D6YG-A575]; *see also* Lily Hay Newman, *Uber Didn't Track Users Who Deleted The App, But It Still Broke the Rules*, WIRED (Apr. 24, 2017, 6:58 PM), https://www.wired.com/2017/04/uber-didnt-track-users-deleted-app-still-broke-rules/ [https://perma.cc/ML5L-C72E].

companies, including Apple,[345] AT&T,[346] Facebook,[347] Google,[348] Microsoft,[349] and Uber[350] have received multiple patents for use of metadata and telemetry. Acquisition of patents do not prove that the companies are using the information in this way, but it is an indicator of possible intent.[351]

The driver of this data collection is the online ad industry. Collectors of the information include websites, operating systems such as Android and iOS, platforms such as Facebook, Google, Instagram, and YouTube, and apps. These all collect personal information about users—where they are, what they are doing, who they are—and use this information to target ads. Often, they share the information they've collected with other sites, including data brokers. The fact that a user was at a gay bar may be in the databases of many Internet companies—even if the user has shut off location tracking at the time of his visit.[352]

While various scholars have previously surveyed the use of communications metadata,[353] and others have looked at what telemetry could reveal,[354] our focus is on controlling use of this

345. Identifying and Locating Users on a Mobile Network, U.S. Patent No. 2017/0026796 A1 (filed Jan. 3, 2019) (issued Jan. 25, 2017).

346. *See, e.g.*, Interactive Community of Interest Profile, U.S. Patent No. 7,970,111 B2 (filed Sept. 1, 2006) (issued June 28, 2011); Using App Location Data and Mobile Application Data to Assess Product Competition, U.S. Patent No. 2021/0035123 A1 (filed Aug. 1, 2019) (issued Mar.18, 2021).

347. Systems and Methods for Utilizing Wireless Communications to Suggest Connections for a User, U.S. Patent No. 2016/0014677 A1 (filed July 10, 2014) (issued Jan.14, 2016); Predicting Household Demographics Based on Image Data, U.S. Patent No. 10,277,714 (filed May 10, 2017) (issued Apr.30, 2019); Offline Trajectories, U.S. Patent No. 10,149,111 (filed May 30, 2017) (issued Dec. 4, 2018).

348. Advertising Based on Environmental Conditions, U.S. Patent No. 8,138,930 (filed Jan. 22, 2008) (issued Mar. 20, 2012).

349. User Activity Detection on a Device, U.S. Patent No. 7,711,815 (filed Oct. 10, 2006) (issued May 4, 2010).

350. Enabling a User to Verify a Price Change for an On-Demand Service, U.S. Patent No. 2013/0268406 A1 (filed Mar. 14, 2013) (issued Oct. 10, 2013).

351. Note that patents do not provide the right or legal ability to do something; instead, they are way to exercise to prevent someone else from implementing a particular invention.

352. *See* discussion *infra* Section IV.C.

353. *See generally, e.g.*, Francesco Calabrese et al., *Urban Sensing Using Mobile Phone Network Data: A Survey of Research*, 47 ACM COMPUTING SURVS., Nov. 2014, at 1; Vincent Blondel et al., *A Survey on Results of Mobile Phone Analysis*, 4 EPJ DATA SCIENCE, 2015, at 1; Jorg Daubert et al., *A View on Privacy & Trust in IoT*, 2015 INT'L CONF. COMMC'N WORKSHOP, at 1; Landau, *supra* note 84.

354. *See generally, e.g.*, Michalis Diamantris et al., *This Sneaky Piggy Went to the Android Ad Market: Misusing Mobile Sensors for Stealthy Data Exfiltration*, PROC. 2021 ACM SIGSAC CONF. COMPUT. AND COMMC'NS SEC. 1065; ANDREAS CLAESSON & TOR E. BJØRSTAD, *supra* note 239; Rahmadi Trimananda et al., *OVRseen: Auditing Network Traffic and Privacy Policies in Oculus VR*, USENIX SEC. SYMP. (2022).

privacy-invasive information. Thus, we categorize the types of personal information metadata and telemetry can reveal.

We start in Section A with location, which is both a category as well as the basis for determining many other social characteristics of groups and individuals. Next, in Section B, we turn to categories of data about groups, while in Section C we consider categories of data about individuals. We use the categories provided by Landau[355] to discuss personal information that can be discerned through metadata, telemetry, and publicly available sources of information. We do not attempt to describe all the ways this can be done, but simply to be illustrative.

A. Determining User Location

Mobile phones move matters not just to their users, but also to advertisers, police, public health experts, and many others. Legal scholars and civil liberties organizations honed in on the issue of warrantless collection of location data immediately. For over a decade, they argued that under a Fourth Amendment analysis, CSLI collection needed a warrant.[356] In *Carpenter*, the Court agreed.[357]

Computer scientists had previously shown how knowing CSLI information readily identified an individual. In 2009, Philippe Golle and Kurt Partridge demonstrated that knowing home and work locations through the approximation provided by CSLI and IP addresses would uniquely identify individuals.[358] In 2013, Yves-Alexandre de Montjoye et al. refined this analysis by demonstrating that four CSLI identifications sufficed to identify 95 percent of individuals out of a population of 1.5 million.[359]

There are multiple ways of determining an individual's location without involving CSLI or IP addresses. These include GPS data, which is commonly requested by various apps, and mobile device sensors, for example tracking auto travel using data from a phone's accelerometer.[360] In 2012, determining an individual's location could be done at precision then equivalent to GPS

355. *See generally* Landau, *supra* note 84.

356. *See, e.g.*, Freiwald, *supra* note 167, at 745.

357. Carpenter v. United States, 138 U.S. 2206, n.3 (2018) (finding that seven days of collection required a warrant, but declining to address whether fewer days of collections would be subject to the same requirement).

358. Philippe Golle & Kurt Partridge, *On the Anonymity of Home/Work Location Pairs*, 5538 LECTURE NOTES COMPUT. SCI. 390, 390 (2009) (noting that this was pre-pandemic, thus occurring at a time when few people worked from home).

359. de Montjoye, *supra* note 57, at 2.

360. Jun Han et al., *ACComplice: Location Inference Using Accelerometers on Smartphones*, 2012 FOURTH INT'L CONF. ON COMMC'N SYS. & NETWORKS, at 1, 12012.

tracking[361] (GPS tracking has since improved). The accelerometer calculation relied, in part, on the fact that roads constrain where automobiles can drive. Subsequent research showed that data from the accelerometer, gyroscope, and magnetometer can reveal people's paths to a high degree of precision, even enabling the ability to determine someone's path inside a building[362] (GPS does not work in such situations). These sensors can also be used to determine whether two people were using the same mode of transportation.[363]

A device's location can be determined from the set of nearby WiFi networks. When a device searches for nearby WiFi, the device discovers a number of network names; the more precise term for these network names is Service Set IDentifiers or, more commonly, SSIDs.[364] Each SSID has a Basic SSID (BSSID) that is the access point to that WiFi network.[365] An app that collects the local BSSIDs can check a public database—of which there are a number—to locate the user.

Some data collectors combine these data sources to reach almost pinpoint accuracy of a user's location. In 2020, the Norwegian Consumer Council commissioned a cybersecurity company, Mnemonic, to track data flowing from ten popular apps used on Android devices.[366] The apps were: Grindr, Perfect365, MyDays, OkCupid, My Talking Tom 2, Muslim: Qibla Finder, Tinder, Clue, Happn, and Wave Keyboard.[367] There were, of course, expected flows of advertising IDs from the apps, but perhaps the most interesting data point is that one of the apps shared additional user data with Placed,[368] a data collector owned by FourSquare.

361. *Id.*

362. Catia Real Ehrlich & Jörg Blackenbach, *Indoor Localization for Pedestrians with Real-Time Capability Using Multi-Sensor Smartphones*, 22 GEO-SPATIAL INFO. SCI. 73, 73 (2019).

363. Systems and Methods for Utilizing Wireless Communications to Suggest Connections for a User, U.S. Patent No. 2016/0014677, at [57] (filed July 10, 2014) (issued Jan. 14, 2016).

364. *Network Director User Guide: Configuring WLAN Service (SSIDs): Understanding the Network Terms SSID, BSSID, and ESSID*, JUNIPER NETWORKS (Oct. 5, 2018), https://www.juniper.net/documentation/en_US/junos-space-apps/network-director4.0/topics/concept/wireless-ssid-bssid-essid.html [https://perma.cc/84JQ-U7S8].

365. *Id.*

366. CLAESSON & BJØRSTAD, *supra* note 239, at 2.

367. *Id.* at 9.

368. Placed is now a part of Foursquare. *See* Jeff Glueck, *Foursquare to Acquire Placed from Snap, Inc. to Deepen Its Location Technology Platform*, FOURSQUARE BLOG (May 30, 2019), https://location.foursquare.com/resources/blog/news/foursquare-to-acquire-placed-from-snap-inc-to-deepen-its-location-technology-platform/#:~:text=Today%2C%20we%20are%20excited%20to,bundle%20of%20location%2Dbased%20offerings [https://perma.cc/5AZC-CRTA].

MyDays also provided GPS data to Placer, which claims to provide "unprecedented visibility into consumer foot-traffic."[369] "Get insights for any place" says the company's webpage, which then lists such potential locations as retailers, commercial real estate, shopping malls, hospitality, and municipalities.[370] Mnemonic found that Placer received "detailed GPS location data, WiFi access point data, cell tower data, and Bluetooth properties from the MyDays app."[371] This enabled Placer to track a device's location to a particular floor within a building.[372]

In short, the only way to keep one's location private is not to carry a mobile phone. If a user must carry a phone, perhaps because of an absence of public ones, then a "dumb" phone not tied to an account and with no Wi-Fi or apps will provide a modicum of privacy. Even then, CSLI will provide a rough guide of where a user has been, what her daily patterns are, and when she deviates from them.[373]

B. Revelatory Information About Populations

Communications are ultimately to and from an individual, but there are also patterns that emerge from larger groups of people. These actions of populations can be revealed through the presence of physical signals, such as radio communications, analysis of the actual metadata, such as CDRs or IP-packet headers, or data from sensors. We have classified them here into different types of information that can be learned through metadata and telemetry.

Order of battle: Radio enables command, control, and communications at a distance, making it an invaluable technology for the military, but radio communications, even if encrypted, leak valuable information. Studying the presence of signals between units can help an adversary determine the composition of troops and equipment on the battlefield—the so-called "order of battle." Such traffic analysis was deployed first during the Russo-Japanese War and has been used ever since. Traffic analysis not only uses radio signals, but also relies on information like callsigns, radio frequencies, schedules and timing of messages, and the location and characteristics of the transmitter.[374] Because this information

369. PLACER, https://www.placer.ai/ [https://perma.cc/VKB8-JBCL].

370. *Id.*

371. CLAESSON & BJØRSTAD, *supra* note 239, at 45.

372. *Id*.

373. One could go one step further and shut the phone off except when making outgoing calls; that, of course, limits the functionality of the phone as a "mobile" one.

374. DONALD A. BORRMANN ET AL., CTR. FOR CRYPTOLOGICAL HIST., NSA, THE HISTORY OF TRAFFIC ANALYSIS: WORLD WAR I—VIETNAM 3 (2013).

largely travels through the air, it is available to anyone with a radio antennae, friend or foe alike.

Use of consumer-grade equipment by military personnel has created new ways to leak such data.[375] While military purchases of such equipment can be adjusted to mitigate this, sometimes information can leak because of personal use of equipment by members of the military. An example was how Strava's fitness app exposed a secret military base: the app, which learns everyone's running routes, publicly shares that information on a website. The path of soldiers who went out for a jog was capable of revealing the base's location.[376] No fancy traffic analysis or deployment of special equipment was required to discover the information because the information was available on a public website.

Telemetry also has the potential to leak data. If it is known which military units have particular types of mobile devices (say from a purchase order or public bidding) and then members of the unit access a site or app,[377] the site can learn where the soldiers are located through the users' IP address. App providers can also learn considerably more from GPS information, including the movement of the troops; combined with knowing what type of device is being used, the provider can know which unit of the military is being spied upon.[378] The Internet and smart devices provide a wide range of capabilities for an adversary to exploit.

Communities of Interest: As previously discussed, AT&T used communities of interest (COI) to uncover fraudsters who had changed their phone number but continued their modus operandi.[379] Because the criminals continued their communications patterns, AT&T was able to discover them through analyzing call detail records.[380] Analysis enabled investigators to track down the terrorists responsible for the assassination of former

375. Susan Landau, *Under the Radar: NSA's Efforts to Secure Private-Sector Telecommunications Infrastructure*, 7 J. NAT'L SEC. L. & POL'Y. 411, 434–36 (2014).

376. Alex Hern, *Fitness Tracking App Strava Gives Away Location of Secret US Army Bases*, GUARDIAN, (Jan. 28, 2018, 4:51 PM), https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases [https://perma.cc/C77A-MNBP].

377. This could be a site or app created by an adversary; such efforts are not uncommon. *See, e.g.*, *Israeli Soldiers 'Caught in Hamas Online Honey Trap'*, BBC (Jan. 12, 2017), https://www.bbc.com/news/world-middle-east-38594669 [https://perma.cc/7KJG-WNLM].

378. This presumes that the user is not accessing the site through a VPN, which would obscure the IP address.

379. See discussion *supra* Section II.A.

380. Corinna Cortes et al., *supra* note 136, at 112–13.

Prime Minister Rafik Harari in Beirut in 2005.[381] Exactly the same type of analysis (and access to the CDRs) allowed Hezbollah to uncover and publicly expose CIA agents and their contacts in Beirut.[382]

Another way to determine communities of interest is to see the common SSIDs user devices have stored. As smartphones seek nearby networks, they expose the ones known to the device; similar probes reveal common membership in a household, workplace, or other group to receivers in the area, the OS, which has access to this data, and applications that collect this information.[383]

Organizational structure: As one might expect, communication patterns reveal corporate structure. Studies of a half million emails sent and received from 151 of the Enron employees from 2000 to 2002[384] revealed the "social hierarchy" of the organization.[385] By autumn 2001, some of the information that had been hidden by the company's dishonest accounting system became public;[386] the email patterns of the company's top leadership changed, with notably decreased upward reporting.[387] Communications metadata illuminates more than the formal networks of an organization. Indeed, because communications metadata reveals key influencers, internet companies rely on it for social media products.[388]

381. Ronen Bergman, *The Hezbollah Connection*, N.Y. TIMES MAG. (Feb. 10, 2015), at 35; *see* Stephen Grey & Don van Natta, *Thirteen with the CIA Sought by Italy in a Kidnapping*, N.Y. TIMES (June 25, 2005), at A1 (demonstrating that similar efforts were used to track down thirteen CIA agents who had kidnapped an Egyptian cleric in Milan).

382. Black Hat, *supra* note 143.

383. Marco V. Barbera et al., *Signals from the Crowd: Uncovering Social Relationships Through Smartphone Probe*, PROC. 2013 CONF. ON INTERNET MEASUREMENT 265, 265 (2013).

384. Jana Diesner et al., *Communication Networks from the Enron Email Corps: "It's Always About the People. Enron is No Different."*, 11 COMPUTATIONAL AND MATHEMATICAL ORG. THEORY 201, 202 (2005).

385. Ryan Rowe et al., *Automated Social Hierarchy Detection Through Email Network Analysis*, PROC. 9TH WEBKDD & 1ST SNA-KDD 2007 WORKSHOP ON WEB MINING & SOC. NETWORK ANALYSIS 109, 113 (2007).

386. Enron systematically eliminated negative balance sheets from its "special purpose entities"; the entities themselves were not included in Enron's primary financial reports. *See, e.g.*, Diesner, *supra* note 384, at 203.

387. *Id.* at 221.

388. *See, e.g.*, Method and System for Identifying a Key Influencer in Social Media Utilizing Topic Modeling and Social Diffusion Analysis, U.S. Patent No. 8,312,056 (filed Sept. 13, 2011).

Communications metadata can also illustrate subtle aspects of group behavior. Studying nine months of Twitter communications from the Occupy Wall Street movement from July 2011 to March 2012, Michael Conover et al. discovered that local communications—ones that did not cross state boundaries (a proxy for communications across a distance)—focused on logistics of local demonstrations while distant communications (ones that did cross state lines) emphasized strategic concerns. Michael D. Conover et al., *The*

In criminal networks, clustering—finding subsets of the criminal enterprise who are closely connected[389]—is particularly useful. Finding highly connected network components,[390] determining who receives the most calls (these turn out to be the "lieutenants" of criminal groups), and those on the "edge" of the criminal activity (potentially easier to investigate and "turn") has proved valuable in investigations.[391] Clustering in such investigation can be done through CDRs, available to the service providers (and law enforcement via subpoenas), plus anyone to whom the companies choose to sell these records.[392]

Telemetry data provides another way to uncover such social networks, including at a fine grain. CSLI will reveal who is in a large group, but telemetry—data from the accelerometer, gyroscope, and magnetometer—will show which people are walking together.[393] In much the same way, this data will reveal which people are in each other's close presence while at a large gathering. This data is available to any of the apps that download this telemetry information and any of the partners with whom they share it. Tracking radio signals can also reveal social structure within a gathering; as noted, probes by the mobile devices for nearby networks can expose family and social networks through the similarity of such probes.[394]

The same telemetry information that can reveal subgroups of connected people within a large gathering can also give away seemingly hidden actions of individuals. "Deep Throat," the FBI source who provided *Washington Post* reporter Bob Woodward with crucial information during the Watergate scandal, would have been tracked by his phone to the same garage as Woodward.[395] Video recordings, now ubiquitous, especially at such places as garages,

Geospatial Characteristics of a Social Movement Communication Network, 8 PLOS ONE, Mar. 2013, at 1, 1–2.

389. Emilio Ferrara et al., *Detecting Criminal Organizations in Mobile Phone Networks*, 41 EXPERT SYS. WITH APPLICATIONS 5733, 5741 (2014).

390. This involves building a graph in which phone numbers are represented by vertices and calls are represented by edges.

391. Ferrara, *supra* note 389, at 5747.

392. CAREY SHENKMAN ET AL., *supra* note 236, at 5.

393. *See, e.g.*, Systems and Methods for Utilizing Wireless Communications to Suggest Communications for a User, U.S. Patent No. 10,111,059 (filed July 10, 2014).

394. Barbera, *supra* note 383.

395. Casey Silvestri, *Bob Woodward Met Deep Throat in a Parking Garage to Report on Watergate*, WASH. POST (June 14, 2022, 10:34 PM), https://www.washingtonpost.com/video/politics/bob-woodward-met-deep-throat-in-a-parking-garage-to-report-on-watergate/2022/06/15/2b2e7f0b-cbb9-4396-879b-c06e267967a5_video.html [https://perma.cc/U9SG-M4WA].

are likely to have shown the two men's cars entering within minutes of each other at 2 am (an odd time to park and thus very noticeable). But telemetry data from the phones would provide proof: it would reveal the men moving in unison as they talked within the garage. Apps with access to accelerometer, gyroscope, and magnetometer data can learn this and similar information.

Demonstrating Community Characteristics: Knowing where someone lives can sometimes provide insight into aspects of personal characteristics, including likely income, religion, and ethnicity and/or race. But such neighborhood information can be dated.[396] Studying the CDRs of a community gives insight into such group attributes at the time of collection, providing real-time community information, rather than demographic information that may be two, ten, or twenty years out of date.

Calling patterns can reveal social capital, the set of networks of relationships within a community that glues the society together,[397] and various types of demographic characteristics, such as the religion of a neighborhood (e.g., no calls from Friday sundown to Saturday sundown may indicate a religious Jewish community).[398]

Telemetry can show the social patterns within a community. Do people stroll in the neighborhood on summer evenings? Do they

396. During the pandemic, certain areas immediately lost large parts of their population; one example was some neighborhoods of Manhattan, where up to 40% of the population temporarily relocated. *See, e.g.*, Kevin Quealy, *The Richest Neighborhoods Emptied Out Most as Coronavirus Hit New York City*, N.Y. TIMES (May 15, 2020), https://www.nytimes.com/interactive/2020/05/15/upshot/who-left-new-york-coronavirus.html [https://perma.cc/KD4R-RTLZ]. Even in less extreme times, neighborhood infrastructure can outlast the population. Centers of worship—synagogues and churches—can be in a community for far longer than there is active usage. *See* Hans Holznagel, *Churches Let Go of Buildings and Find Ways to Keep Worshipping*, UNITED CHURCH CHRIST (Aug. 16, 2022), https://www.ucc.org/churches-let-go-of-buildings-and-find-ways-to-keep-on-worshiping-witnessing-working/ [https://perma.cc/9BJX-NCEC].

397. *See* Marco Mamei et al., *Is Social Capital Associated with Synchronization in Human Communication? An Analysis of Italian Call Records and Measures of Civic Engagement*, 7 EPJ DATA SCI., July 17, 2018, at 1, 1 (using Italian phone records, investigated synchronization within close proximity-based communities and between communities in a larger geographical area to measure social capital. The authors found positive correlation with traditional social capital measures (blood donations, association density) and close proximity-based synchronization; they also found negative correlation between social capital measures and synchronization between communities).

398. Paolo Bajardi et al., *Unveiling Patterns of International Communities in a Global City Using Mobile Phone Data*, 4 EPJ DATA SCI., Apr. 29, 2015, at 1, 2 (determining the national background of a community requires subtlety as it is necessary to distinguish between areas that have tourist and visitor attractions (and hence a high transient population) and neighborhoods that genuinely have a high number of immigrants. The authors showed how to measure this "entropy"—and thus eliminate it—in order to be able to determine the home country of immigrant neighborhoods, again using CDRs).

sit on the brownstone steps and chat? Is the neighborhood liquor store highly frequented? What about the cannabis shop? Which groups of teenagers are spending time together? Is this a neighborhood with lots of dogs? Dog walkers follow a pattern of walks at regular times of the day; these walks typically involve short stops but not at shops. What mode of transportation does the neighborhood use to get to work?

Monitoring People's Movements: Mobile phones provide real-time information on where people are, which means they provide invaluable information during emergencies. Tracking mobile phone data often provides more accurate information than other types of efforts.[399]

Mobile phone data shows, in aggregate, the movement of displaced people, enabling relief to be directed to where the population is, not where it used to be.[400] Studying population movement also facilitates the ability to predict transmission of infectious and contagious diseases, such as malaria[401] and dengue fever,[402] thus enabling the setup of public health capabilities in advance of need.

Public protests are recognizable by slow movement and relatively few calls being made.[403] The same techniques that can find communities of interest by looking for similar network usage work here as well; radio tracking can expose family and social networks through the similarity of such probe groups.[404] Meanwhile, apps would be able to track subgroups within the larger group by their private patterns: where they stop together for water or a discussion, etc.

The Privacy Impact on Individuals of Information about Population Groups: Information about population groups

399. Sibren Issacman et al., Climate Change Induced Migrations from a Cell Phone Perspective, (6th International Conference on the Analysis of Mobile Phone Datasets, 2019); Xin Lu et al., *Unveiling Hidden Migration and Mobility Patterns in Climate Stressed Regions: A Longitudinal Study of Six Million Anonymous Mobile Phone Users in Bangladesh*, GLOB. ENV'T CHANGE, May 2016, at 1, 6 (2016).

400. Xin Lu et al., *Predictability of Population Movements After the Haiti 2010 Earthquake*, 109 PROC. NAT'L ACAD. SCI. 11576. 11576 (2012) (discussing Haiti and Nepal as examples); Robin Wilson et al., *Rapid and Near Real-Time Assessments of Population Displacement Using Mobile Phone Data Following Disasters: The 2015 Nepal Earthquake*, PLOS CURRENTS, Feb. 24, 2016, at 1.

401. *See generally* Andrew Tatem et al., *The Use of Mobile Phone Data for the Estimation of the Travel Patterns and Imported Plasmodium Falciparum Rates Among Zanzibar Residents*, MALARIA J. (2009); Amy Weslowski et al., *Quantifying the Impact of Human Mobility on Malaria*, 338 SCIENCE 267 (2012).

402. *See generally* Amy Wesolowski et al., *Impact of Human Mobility on the Emergence of Dengue Epidemics in Pakistan*, 112 PROC. NAT'L ACAD. SCI., 11887 (2015).

403. Adrian Dobra et al., *supra* note 82, at 1.

404. Barbera, *supra* note 383, at 266.

potentially provides information about individuals. Some communications metadata, such as radio signals, are available for anyone who has an antenna. In such a situation, there is no easy protection against collection, including against foreign adversaries. Roaming mobile phones use a Temporary Mobile Subscriber Identity ("TMSI"),[405] which protects an individual phone subscriber from being tracked through the air interface between their phone and the tower.[406] But, nothing prevents anyone with an antenna from collecting TMSIs en masse—and thus they are able to monitor group movements even if they do not have the ability to track particular individuals over time.

By contrast, other information such as CDRs, which are collected by service providers, are generally not publicly available. If an attacker plants a rogue base station—a so-called "IMSI catcher" masquerading as a genuine cell tower—then this will capture the device ID and phone number (IMEI and IMSI respectively), enabling tracking as well as eavesdropping. U.S. law enforcement has used such devices for conducting surveillance since at least the early 1990s.[407] Such cell site simulators are also used by foreign adversaries to track users and eavesdrop,[408] with Washington, D.C., being one target for this form of spying.[409]

While some information about population groups, e.g., organizational structure, may reveal information about individuals, other data is only revealed in aggregate. A particular individual fitting a group pattern does not provide definitive information about that individual.

C. Revelatory Information About Individuals

The transformation from phones as stationary "dumb endpoints" on a smart network to mobile devices that were both

405. PATRICK TRAYNOR ET AL., SECURITY FOR TELECOMMUNICATIONS NETWORKS 46 (2008).

406. Depending on the carrier, the TMSI can change as frequently as after every call or as infrequently as every several days. This information is usually found within the carriers' private documentation of their practices.

407. Stephanie K. Pell & Christopher Soghoian, *Your Secret StingRay's No Secret Anymore: The Vanishing Government Monopoly over Cell Phone Surveillance and Its Impact on National Security and Consumer Privacy*, 28 HARV. J.L. & TECH. 1, 14 (2014).

408. Jeff Stein, *New Eavesdropping Equipment Sucks All Data Off Your Phone,* NEWSWEEK (June 22, 2014, 8:27 AM), http://www.newsweek.com/2014/07/04/your-phone-just-got-sucked-255790.html [https://perma.cc/2L3Q-9L7F].

409. *See, e.g.*, Letter from Christopher C. Krebs, Senior Off. Performing Duties Under Sec'y, U.S. Dep't Homeland Sec., to Senator Ron Wyden (May 22, 2018), https://www.wyden.senate.gov/imo/media/doc/Krebs%20letter%20to%20Wyden%20after%20May%20meeting.pdf [https://perma.cc/63L8-U5AC].

telephones and computers means that far more information can now be discerned about users. Discovering this information about individuals from metadata and telemetry was novel in the 2000s; by now that trickle of papers has become a torrent. We discuss the types of personal information that can be determined about individuals through combining metadata and telemetry with publicly available information,[410] but do not attempt to fully survey the literature on this work.[411]

Identifying Devices: There are multiple ways to identify devices. For webpages to display properly,[412] browsers send configuration and version information to webservers. This information allows device "fingerprinting." In 2010, Electronic Frontier Foundation researcher Peter Eckersley showed in a population of approximately half million browsers, 94.2 percent of those with Flash or Java installed had distinct fingerprints.[413] In 2016, Pierre Laperdrix et al. showed that browser fingerprinting worked in the more restricted environment of mobile devices.[414] While battery status can be useful for performance metrics—and is apparently used for this purpose by YouTube—it can also be used to profile and track users.[415] Olejnik et al. found this was the case almost half the time.[416]

Internals of a machine can enable fingerprinting; fingerprinted accelerometer chips[417] can be combined with browser and location information to fingerprint a device.[418] Irene Amerini et al. have used individual anomalies within four sensors—accelerometer, gyroscope, magnetometer, and microphone—to fingerprint a

410. *See generally* Mayer et al., *supra* note 331, at 5536 (following the model used by Mayer et al.)

411. Nor could we do so: the research literature is growing so quickly that we would be out of date before we finished putting pen to paper.

412. In the early days of the World Wide Web, it was often the case that webpages that had previously displayed correctly would fail to after a change in browser code. This problem is rare now.

413. Peter Eckersley, *How Unique Is Your Web Browser?*, 6205 LECTURE NOTES COMPUT. SCI. 1, 2 (2010).

414. Pierre Laperdrix, et al., *Beauty and the Beast: Diverting Modern Web Browsers to Build Unique Browser Fingerprints*, 2016 IEEE SYMP. ON SEC. & PRIV. 878, 879.

415. Lukasz Olejnik et al., *Battery Status Not Included: Assessing Privacy in Web Standards*, 2017 INT'L WORKSHOP ON PRIV. ENG'G 5, https://lukaszolejnik.com/AssessingPrivacyWebStandardsIWPE17.pdf [https://perma.cc/AF52-UQP7].

416. *Id.*

417. Sanorita Dey et al., AccelPrint: Imperfections of Accelerometers Make Smartphones Trackable 1 (Network and Distributed System Symposium, 2014).

418. Thomas Hupperich et al., *On the Robustness of Mobile Device Fingerprinting: Can Mobile Users Escape Modern Web-Tracking Mechanisms?*, PROC. 31ST ANN. COMPUT. SEC. APPLICATIONS CONF. 191, 193–94 (2015).

device.[419] This fingerprint could be used to authenticate a device—or to track the device owners.

There are multiple other ways to identify a device. Mnemonic discovered that data shared included "device model, information about the hardware display, device configuration, battery levels, locale settings, carrier, nearby wireless networks."[420] In many cases, the information supplied would be sufficient to identify the device.[421] For a period of time, Apple allowed developers to track users through a Unique Device Identifier (UDID), but this tracking enabled developers to identify users even after users removed the app. With iOS 5, Apple changed its policy and no longer permitted developer use of the UDID once an app was removed.[422]

There are many legitimate reasons to fingerprint a device, including fighting fraud. Wired reported that Uber did so to prevent drivers from "gaming a promotion rewarding them for maximizing ride volume."[423] There are also illegitimate reasons, which include tracking users. Uber apparently did so even after the users removed the app from their phones.[424]

Identifying the device can play a role in user authentication, an issue we discuss in *Identifying the User*.

Identifying Device Activity: One way to recognize device activity is through characteristics of network traffic. It used to be that port numbers were quite accurate for identifying most applications,[425] but, due to various changes in delivery and transport, including dynamic port negotiation, tunneling, and efforts to obfuscate traffic, port numbers have become less useful for such identification.[426] Instead, network flow features can classify some types of network traffic including, in some cases, what applications or web services are being used.[427]

In 2007 Laurent Bernaille and Renata Teixeira showed how an intermediate ISP could recognize an application from the size of its

419. Amerini et al., *supra* note 337, at 1.

420. CLAESSON & BJØRSTAD, *supra* note 239, at 16.

421. *Id.* at 21 (explaining how data shared can sufficiently identify specific devices).

422. Erick Schonfeld, *Apple Sneaks a Big Change into iOS 5: Phasing Out Developer Access to the UDID*, TECHCRUNCH (Aug. 19, 2011, 2:24 PM), https://techcrunch.com/2011/08/19/apple-ios-5-phasing-out-udid/ [https://perma.cc/FJM9-WBBV].

423. Newman, *supra* note 344.

424. *Id.*

425. Hyunchul Kim et al., *Internet Traffic Classification Demystified: Myths, Caveats, and the Best Practices*, PROC. 2008 ACM CONEXT CONF. 1.

426. Raouf Boutaba et al., *A Comprehensive Survey and Machine Learning for Networking: Evolution, Applications and Research Opportunities*, J. OF INTERNET SERVS. & APPLICATIONS, June 18, 2018, at 1, 32.

427. *Id.* at 20–23.

first few packet headers of an SSL connection.[428] That was at the time when the Internet supported a large variety of applications, but far from the millions of mobile apps we currently see. In 2015, Qinglong Wang et al. used the fact that certain aspects of network flow, including bursts of traffic, size of packets, pattern of the communications, are relatively unique for different mobile apps, enabling ISPs to recognize apps from packet header and flow information.[429] A year later, Mauro Conti et al. showed how to use information from packet headers, including IP address and port number along with packet flow information to recognize seven popular mobile apps (Facebook Gmail, Twitter, Tumblr, Dropbox, Google+, and Evernote).[430]

The ability to do such app recognition continues to improve. In a literature survey in 2021, Eva Papadogiannaki and Sotiris Ioannidis found that machine learning techniques are increasingly improving the ability of ISPs to identify the use of particular apps simply from packet header information and/or flow data.[431] Wi-Fi networks in the user's vicinity will be able to determine much about device activity simply from the packet headers.

Identifying the User: There are many ways to identify a user through their device. If a device is used only by one person, then, of course, identifying the device effectively identifies the user. In part because of zero trust architectures in which the user or device must authenticate themselves to access network resources,[432] research on device authentication is proceeding apace. There will increasingly be ways to identify a device and, thus, its user. Once such techniques are developed, their use will undoubtedly expand beyond the purpose of authenticating to zero trust networks.

428. Laurent Bernaille & Renata Teixeira, *Early Recognition of Encrypted Applications*, 4427 LECTURE NOTES COMPUT. SCI. 165, 165 (2007).

429. Qinglong Wang et al., *I Know What You Did on Your Smartphone: Inferring App Usage Over Encrypted Data Traffic*, 2015 IEEE CONF. ON COMMC'NS & NETWORK SEC., 433, 433–34.

430. Mauro Conti et al., *Analyzing Android Encrypted Network Traffic to Identify User Actions*, 11 IEEE TRANSACTIONS ON INFO. FORENSICS & SEC. 114, 115 (2016).

431. Eva Papadogiannaki & Sotiris Ioannidis, *A Survey on Encrypted Network Traffic Analysis*, 54 ACM COMPUTING SURVS., July 2021, at 1, 8 (2022).

432. *See e.g.*, SCOTT ROSE ET AL., NAT'L INST. STANDARDS AND TECH., ZERO TRUST ARCHITECTURE, 4 (2020), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf [https://perma.cc/TM9L-3HAU] (describing a zero trust architecture as a model for an enterprise cybersecurity architecture that trusts no one and no asset. Everything that accesses a resource—both users *and* devices—must be authenticated; reauthentication requests can occur frequently).

We have already discussed how there are multiple ways to track a user's location through communications metadata. Tracking a device's location, also, effectively enables user identification. There are other methods as well to track the user, largely though the physical attributes of user interactions with their device. This includes authentication through patterns of typing,[433] touching,[434] and motion.[435]

The method described above for identifying device activity can also be used to identify the user. It is not hard to determine the apps someone is using;[436] the packet destination addresses provide this information. In 2021, Vedran Sekara et al. showed that, much like location, the set of used apps identifies the individual.[437] Studying smartphone app usage of 3.5 million people, Sekara et al. showed it was possible to uniquely identify an individual simply from a profile of their app usage.[438] Thus, for example, nearby Wi-Fi networks may be able to identify the user simply by eavesdropping on packets and reading the headers.

Profiling the User: Metadata and telemetry information reveal where you are, what you are doing, who is with you, and who you are. Sometimes it can even give away that you are not actually a human being. Traffic metadata can recognize a Twitter bot; people and bots differ on such measures, like who the account is following, number of "friends" and "followers," and number of tweets; this information can be used to find automated accounts.[439] Other uses of metadata and telemetry information to profile the user are potentially less privacy friendly.

433. Prima Chairunnanda et al., *Privacy: Gone with the Typing! Identifying Web Users by Their Typing Patterns*, 2011 IEEE THIRD INT'L CONF. ON PRIV., SEC., RISK & TRUST 974, 974.

434. Rahat Masood et al., *Touch and You're Trapp(ck)ed: Quantifying the Uniqueness of Touch Gestures for Tracking*, 2018 PROC. ON PRIV. ENHANCING TECHS. 122, 122.

435. Anupam Das et al., *Every Move You Make: Exploring Practical Issues in Smartphone Motion Sensor Fingerprinting and Countermeasures*, 2018 PROC. ON PRIV. ENHANCING TECHS. 88, 88.

436. Vincent F. Taylor et al., *AppScanner: Automatic Fingerprinting of Smartphone Apps from Encrypted Network Traffic*, 2016 IEEE EUROPEAN SYMP. ON SEC. AND PRIV. 439, 439.

437. *See* Vedran Sekara et al., *Temporal and Cultural Limits of Privacy in Smartphone App Usage*, 11 SCI. REPS. 1, 3 (2021) (discussing how identification of the individual is done, not just through the smartphone app data, but through cross-references with other data sources).

438. *Id.* at 1.

439. *See, e.g.*, Onur Varol et al., *Online Human-Bot Interactions: Detection, Estimation, and Characterization*, PROC. ELEVENTH INT'L AAAI CONF. ON WEB & SOC. MEDIA 280, 280–81 (2017); Zi Chu et al., *Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?*, 9 IEEE TRANSACTIONS ON DEPENDABLE & SECURE COMPUTING 811, 811–12 (2012); Jytte Klausen et al., *Finding Online Extremists in Social Networks,* 66 OPERATIONS RSCH. 957, 960 (2018).

CDRs demonstrate social ties, the beginning and waning of friendships and relationships, as do many other attributes available to networks, OSes, and applications. In 2014, Suranga Seneviratne et al. showed it was possible to predict the user's "religion, relationship status, spoken languages," and whether or not they had small children from a single snapshot of apps installed on a smartphone.[440] In 2020, by examining six types of metadata collected by smartphones, including duration of gaming app usage, music consumption, communication and social behavior, Clemens Stachl et al. found this metadata could reveal major behavioral traits of the users, including "openness, conscientiousness, extraversion, agreeableness, and emotional stability."[441] AT&T's original use of communities of interest was for fraud prevention, but in 2011, AT&T Mobile received a patent to monitor a user's network accesses—which apps she uses—in order to recommend a social network to the user based on her interests.[442] Nearby Wi-Fi networks will also have access to such information.

The metadata of personal and home IoT devices is, of course, particularly revelatory. It is possible to know when someone sleeps,[443] when and where they exercise, what time in the morning they want their coffee, and when they have guests over (increased use of the dishwasher). Yet, other uses of communications metadata may be even more intrusive; communications metadata reveals not only what people do but can also reveal how they feel about others. M. Keith Chen and Rhyn Rola used communications metadata to explore the social split that developed after the 2016 presidential election.[444] The researchers looked at time spent together by guests from an "opposite" voting precinct, who came for the day (and presumably the meal) at Thanksgiving in 2015 and 2016.[445] The

440. Suranga Seneviratne et al., *Predicting User Traits from a Snapshot of Apps Installed on a Smartphone*, 18 ACM SIGMOBILE MOB. COMPUTING & COMMC'NS REV. 1, 1 (2014).

441. Clemens Stachl et al., *Predicting Personality from Patterns of Behavior Collected with Smartphones*, 117 PROC. NAT'L ACAD. SCIS. 17680, 17681 (2020) (The six types of metadata collected were "1) app usage (e.g., mean duration of gaming app usage), 2) music consumption (e.g., mean valence of played songs), 3) communication and social behavior (e.g., number of outgoing calls per day), 4) mobility behaviors (e.g., mean radius of gyration), 5) overall phone activity (e.g., number of unlock events per day), and 6) a higher-level behavioral class that captured the extent of daytime versus nighttime activity (e.g., outgoing calls at night.").

442. Interactive Community of Interest Profile, U.S. Patent No. 7,970,111 col. 2 l. 28 (filed Sept. 1, 2006) (issued June 28, 2011).

443. *See Beddit Sleep Monitor*, *supra* note 83.

444. Chen & Rohla, *supra* note 336, at 1020.

445. *Id.*

guests spent 30 to 50 minutes less with their hosts in 2016 than they had in 2015.[446]

The Mnemonic researchers carefully examined the traffic from Grindr, "the world's largest social networking app for gay, bi, trans, and queer people."[447] Data from the app is sensitive; indeed, even the fact that someone is using the app can be sensitive. The app used SDKs from several adtech companies; it shares a great deal of information about the user, including with MoPub.[448] MoPub made requests for the Android Advertising ID (AAID), IP address, GPS location, app name, and user gender and age, which, in turn, it shared with a number of third parties.[449] Use of the AAID makes it relatively easy to track a user; the location data provided by MoPub makes it even easier to identify the user.[450]

Social connections can reveal protected traits, such as sexual orientation.[451] Social connections are easily revealed though such telemetry information as SSIDs held in common, data from accelerometers, or data from the combination of accelerometers, gyroscopes, and magnetometers, etc. In some parts of the world, a user's calling patterns are analyzed to see if she would be a good credit risk.[452] It is easy to see how the use of metadata and

446. *Id.*

447. CLAESSON & BJØRSTAD, *supra* note 239, at 23; *see* GRINDR, https://www.grindr.com [https://perma.cc/2KPS-UF3J].

448. CLAESSON & BJØRSTAD, *supra* note 239, at 23–26. MoPub was owned by Twitter between 2013 and early 2022. *See* Felipe Espósito, *Twitter Finalizes Sale of MoPub as Apple's Privacy Policies Affect Advertising Industry*, 9TO5MAC (Jan. 3, 2022), https://9to5mac.com/2022/01/03/twitter-finalizes-sale-of-mopub-as-apples-privacy-policies-affect-advertising-industry/#:~:text=Twitter%20finalizes%20sale%20of%20MoPub%20as%20Apple's%20privacy%20policies%20affect%20advertising%20industry&text=Twitter%20announced%20in%20October%202021,company%20AppLovin%20for%20%241.05%20billion [https://perma.cc/CRD2-KFSZ].

449. CLAESSON & BJØRSTAD, *supra* note 239, at 27.

450. In late 2021, Google updated its systems to enable users to opt of the use of this ID—and thus ad personalization. *See Advertising ID*, GOOGLE, https://support.google.com/googleplay/android-developer/answer/6048248?hl=en [https://perma.cc/JN22-SQL3]).

451. Carter Jernigan & Behram F.T. Mistree, *Gaydar: Facebook Friendships Expose Sexual Orientation,* 14 FIRST MONDAY (2009).

452. Olga Kharif, *No Credit History? No Problem? Lenders Are Looking at Your Phone Data*, BLOOMBERG (Nov. 25, 2016), https://www.bloomberg.com/news/articles/2016-11-25/no-credit-history-no-problem-lenders-now-peering-at-phone-data#xj4y7vzkg [https://perma.cc/6LW2-3KW9]. U.S. companies were loath to detail their use of calling records, but the article cited Cignifi, a startup that "translates customer behavior into financial opportunities. *See* CIGNIFI, https://www.cignifi.com/ [https://www.cignifi.com/]. In 2016, Jonathan Hakim, Cignifi CEO, told Bloomberg News: "The way you use the phone is a proxy for the way you live," he said. "We are capturing a mirror of the customer's life." Bloomberg reported that the company collects phone data, including who the user is calling and how often, from such companies as Airtel Ghana. This data is used to assess a potential borrower's reliability. Kharif, *supra* 452.

telemetry information can reinforce already existing discriminatory patterns.[453]

By revealing the user's activities, telemetry data can also profile the user. Data from an accelerometer can be used to determine whether the user is standing, walking, hopping, jumping, going up or down stairs, or in a car.[454] A person's level of physical activity has been linked to their emotions[455]—and even their personality traits (extraversion, neuroticism, conscientiousness, and openness).[456] Such information can be revealed through smartphone accelerometers. For example, eBay has a patent for monitoring a user's location and activity from a host of sensors—acoustic, temperature, humidity, accelerometer, gyroscope, altitude, and more[457]—in order to know when a user has switched activities and can be interrupted with a notification (perhaps to bid).[458] Knowing when a user has changed activities is of interest for many other reasons as well. Are they off for a long lunch while at work? Are they exhibiting the walking patterns of someone who has imbibed a bit too much?[459] Smartphone accelerometer data can even reveal the text typed on a phone touchscreen, including 4-digit PINs, passwords, and Google touch-screen password patterns.[460]

453. *See, e.g.*, danah boyd et al., *The Networked Nature of Algorithmic* Discrimination, *in* DATA AND DISCRIMINATION: COLLECTED ESSAYS 53 (Seeta Peña Gangadharan ed., 2014).

454. Adil Khan et al., *Exploratory Data Analysis of Acceleration Signals to Select Light-Weight and Accurate Features for Real-Time Activity Recognition on Smartphones*, 13 SENSORS 13099, 13103 (2013); Stephen. A. Antos et al., *Hand, Belt, Pocket or Bag: Practical Activity Tracking with Mobile Phones*, 231 J. NEUROSCI. METHODS 22, 22 (2014); Alvina Anjum & Muhammad U. Ilyas, *Activity Recognition Using Smartphone Sensors*, 2013 IEEE 10TH CONSUMER COMMC'NS & NETWORKING CONF. 914, 914.

455. Neal Lathia et al., *Happier People Live More Active Lives: Using Smartphones to Link Happiness and Physical Activity*, 12 PLOS ONE, Jan. 2017, at 1, 1.

456. Kathryn E. Wilson & Rodney K. Dishman, *Personality and Physical Activity: A Systematic Review and Meta-Analysis*, 72 PERSONALITY & INDIVIDUAL DIFFERENCES 230, 233 (2015).

457. These privacy-invasive sensors include brain wave, perspiration, heart rate, blood pressure, eye tracking, and more; *see* Methods and Systems for Providing Notifications Based on User Activity, U.S. Patent No. 2016/0037482 A1 fig. 3, (filed July 29, 2014) (issued July 24, 2017).

458. *Id.* col. 2 ¶ 13.

459. Jacob Leon Kröger et al., *Privacy Implications of Accelerometer Data: A Review of Possible Inferences*, 2019 PROC. 3RD INT'L CONF. ON CRYPTOGRAPHY, SEC., & PRIV. 81, 82.

460. Adam J. Aviv et al., *Practicality of Accelerometer Side Channels on Smartphones*, 2012 PROC. 28TH ANN. COMPUT. SEC. APPLICATIONS CONF. 41, 41; Emmanuel Owusu et al., *ACCessory: Password Inference using Accelerometers on Smartphones*, PROC. TWELFTH WORKSHOP ON MOBILE COMPUTING SYS., Feb. 28, 2012, at 1, 5.

The phone might easily transform into being a spy in one's pocket. It is not just eBay that seeks to know what a person is doing; credit companies do so to determine a client's suitability for a loan,[461] and so might many other types of firms.[462] Surveying the personal information inferable from data supplied by smartphone accelerometers, Jacob Kröger et al. concluded that the devices can cause significant privacy intrusions.[463] We could not agree more.

Having seen what may be inferred from metadata and telemetry and for us to adequately comprehend the seriousness of the privacy threat, we need to look at who has access to this data. CDRs are collected by the telecommunications provider as well as intermediaries.[464] Packet headers are similarly available to any entity that transmits or receives the packet. That means, of course, the ISPs transmitting the communication as well as the packet recipient. Both are subject to eavesdropping through IMSI catchers or packet sniffers. Telemetry information is typically collected by the phone manufacturer, oftentimes by the developer of the phone OS, and apps. From there, as we have seen, many others can access this information. The ad infrastructure that has developed over the last two decades has provided a great economic incentive to share this information widely. It is time to consider privacy protections, both real and imagined.

III. THE FAILURE OF NOTICE AND CHOICE

Though users nominally provide consent through companies' privacy policies, as a result of interlocking issues, companies' use of metadata and telemetry to discover user characteristics and behavior effectively occurs without informed consent. This effect is what Solove and Hartzog call "broken expectations of consumer privacy"[465] as users provide metadata and telemetry with one set of expectations as to how it will be used, but the reality of how this information is actually used is way beyond users' expectations.

461. Kharif, *supra* note 452.

462. For example, there appears to be interest from health insurance firms. *See* Marshall Allen, *Health Insurers Are Vacuuming Up Details About You—and It Could Raise Your Rates*, PROPUBLICA (July 17, 2018, 5:00 AM) https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates [https://perma.cc/2E4P-E4N4]. *See also* Predictive Data Analytics with Automatic Feature Extraction, U.S Patent No. 2020/0175314 A1 (filed Dec. 4, 2018). This patent, filed by Optum (a health insurer) employees, notes that "the metadata information can include information about timing of a latest question retrieval by the user computing device." *Id.* col. 2 ¶ 0041:113.

463. Kröger, *supra* note 459, at 85.

464. *See, e.g.*, Landau & Lubin, *supra* note 221, at 309. (These records may not always be fully correct due to temporary numbers provided when a caller is roaming).

465. Solove & Hartzog, *supra* note 63, at 667 (emphasis omitted).

There are multiple reasons why users are misled in the case of metadata and telemetry. To receive content, users must supply communications metadata but lack a realistic way to prevent further use of the metadata they provide. This situation is exacerbated by the fact that communications metadata and telemetry information largely consist of microdata. Consider querying the user with such issues as: May Android collect this phone swipe? May Apple store the length of time you spent on the home screen of the updated iOS system? May Facebook collect SSIDs as you move about the city? May Uber collect accelerometer data from the phone? The number of queries that would result precludes presenting notice of collection and seeking consent for each such use. Such requests would very quickly overwhelm users.

Even worse, the user is unlikely to understand what information metadata and telemetry information provides, let alone how such microdata may ultimately be used. Thus, despite the invasive uses to which metadata and telemetry may be put, not only is informed consent not part of the current situation for users, in fact, informed consent of the use of metadata and telemetry is unworkable. The complex set of privacy controls—the Fair Information Practice Principles in their various instantiations—that society has built since the 1970s fail to handle the determined assault that Internet companies have made on users' privacy.[466] The controls fail because the protective structures fail to match the situations in which users find themselves. This situation is particularly acute for communications metadata and telemetry information.

In Subsection A, we explore the original U.S. Government efforts to protect privacy, then show how these privacy protections came to be operationalized through the Fair Information Practice Principles (FIPPs). In Subsection B, we focus on the critical role that user choice has in protecting privacy, and how that role may be inadequate. We examine how, even though Paul Ohm raised the issue regarding ISP use of communications metadata over a decade ago, privacy practices have failed to implement controls on the collection and use of metadata and telemetry. In Subsection C, we return to metadata and telemetry information and the complex issue of consent. We analyze whether consent for the use of communications metadata and telemetry information can genuinely be given. Is consent—informed consent with the user

466. *See* Robert Gellman, *Fair Information Practices: A Basic History Version 2.21* (Apr. 6, 2022) (working paper at 25), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2415020 [https://perma.cc/9JR7-GCK7]; *see* discussion *infra* Section IV.A.

understanding the implications of consent—possible? We see how the use of communications metadata outside the delivery and display of content is a departure from what legal scholar Lisa Austin describes as "meaningful privacy protections."[467]

A. U.S. Privacy Protections: Philosophy to Policy

In 1962, Alan Westin was asked to organize a study for the Association of the Bar of New York's Special Committee on Science and the Law on the impact of modern technology on privacy. To answer it, Westin realized he needed an understanding of what privacy meant in a liberal democratic state. This realization led to his seminal work of the role of privacy in modern American society, *Privacy and Freedom*.[468]

Westin described privacy as the ability of individuals, groups, and institutions to control when and how information about themselves is shared.[469] He viewed "strong citadels of group and individual privacy" as an absolute essential for a democratic society.[470] It is in such safe spaces that ideas, including ones that appear radical when first proposed, can percolate. Westin described privacy as manifesting itself as solitude, or the absence of being observed; as intimacy, or solitude for a small set of people such as a couple or a family; as anonymity, the ability to be unknown while in a public space; and as reserve, the ability to have barriers that enable an individual to be private, that is, to be able to not fully expose themselves.[471] This last, the most subtle of Westin's delineations, includes the discretion not to intrude by those closest to a person.

Privacy provides space for the individual to be a different person to different people: an exuberant singer at karaoke night, or a quiet daughter at her parents' anniversary celebration, or a whip-smart prosecutor in court, and even a joke-cracking bridge player on a night out with friends. Privacy permits people to let their hair down, slip an extra scoop of hot fudge onto the ice cream at dinner, or do jumping jacks and chin-ups while the camera is off during a colleague's Zoom presentation.

Privacy is more than protection of individuals; it provides protection for society. As sociologist Edward Shils put it, to achieve political civility, a democratic society must provide sufficient

467. Austin, *supra* note 62, at 53.
468. Oscar M. Ruebhausen, *Foreword* to WESTIN, *supra* note 61, at vi–xii.
469. WESTIN, *supra* note 61, at 7.
470. *Id.* at 24.
471. *Id.* at 31–32.

privacy to enable "individual creativity and group expression."[472] By enabling the exploration and development of ideas in solitude, or, with a small, trusted set of people, privacy is essential to handling complex conflict. Whether in labor union negotiations or in nation-state parleys, privacy is essential for success. Privacy allows an idea to be safely assessed, providing a way for developing concepts that may seem extreme at first—Social Security, government air and water purity standards, universal health care—and, thus, bringing them into political discourse and, sometimes, into law and government policy. Hence, within a liberal democracy, privacy serves not only to protect individuals but is, in a deeply fundamental way, essential for a nation's stability.[473]

Westin proposed some now familiar forms of privacy protection, including consent to collection and limitations on repurposing the information.[474] At the time that Westin wrote *Freedom and Privacy*, he was not, of course, the only one thinking about responses to new threats to privacy. Various efforts led to the FIPs. It is not our intent to reprise the history of the FIPs—others have done so quite effectively[475]—but in the context of understanding how to provide users with control over their communications metadata and telemetry information, it is important to lay out how the U.S. arrived at a situation in which a user has effectively no control over the uses of metadata collected by the communications industry. We present the history, albeit in very shortened form.

In the early 1970s, the U.K. government chartered a committee to study the impact of private organizations on privacy;[476] its recommendations included what we now know as notice, choice, and purpose limitation.[477] The U.S. Department of Health, Education, and Welfare (HEW) chartered an Advisory Committee on Automated Personal Data Systems, which issued its report in 1973.[478] The HEW report put forward five principles—a "code of

472. Edward Shils, *Privacy: Its Constitution and Vicissitudes*, 31 L. & CONTEMP. PROBS. 281, 284 (1966).

473. *See* WESTIN, *supra* note 61, at 35–36, 42–51.

474. *Id.* at 373–75.

475. Gellman, *supra* note 466 (working paper at 25).

476. This was known as the "Younger" report after its chair, the Rt. Hon. Kenneth Younger. *See id.*

477. SEC'Y'S ADVISORY COMM. ON AUTOMATED PERS. DATA SYS., U.S. DEP'T HEALTH, EDUC., & WELFARE, REPORT OF THE SECRETARY'S ADVISORY COMMITTEE ON AUTOMATED PERSONAL DATA SYSTEMS, RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS 173 (1973).

478. *Id.* at vi.

fair information practices"—regarding the collection of information on individuals:

1. There must be no personal data record-keeping systems whose very existence is secret.
2. There must be a way for a person to find out what information about the person is in a record and how it is used.
3. There must be a way for a person to prevent information about the person that was obtained for one purpose from being used or made available for other purposes without the person's consent.
4. There must be a way for a person to correct or amend a record of identifiable information about the person.
5. Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuses of the data.[479]

These principles are now more commonly called the Fair Information Practice Principles (FIPPs).[480] Many aspects of the HEW report made its way into the Privacy Act of 1974.[481] European efforts on privacy applied to both the public and private sector.[482] However, the Privacy Act of 1974 only governs Federal agency collection, dissemination, and use of citizens' information;[483] U.S. privacy law for private-sector collection and use is sectoral.

The Communications Act of 1934 gave the Federal Communications Commission (FCC) rulemaking abilities over telephone service providers;[484] this was augmented by the Telecommunications Act of 1996.[485] Since public access to the Internet was originally done through dial-up modems, regulatory authority, including privacy rules, over Internet service fell, at first, to the FCC. By 1996, access to the Internet was changing. The 1996 Act distinguished information services from telecommunications services, the former was unregulated, leaving the question of whether the Internet was an information service or a

479. *Id.* at xx-xxi.

480. *See* Gellman, *supra* note 466, at 27 (noting that the Department of Homeland Security was the first to introduce the Fair Information Practice Principles (FIPPs) as opposed to the FIPs).

481. *Id.* at 8.

482. *Id.* at 14.

483. Privacy Act of 1974 § 2(a)(1), Pub. L. No. 93-579, 88 Stat. 1896 (codified as amended at 5 U.S.C. § 552a).

484. Communications Act of 1934, Pub. L. No. 73-416, 48 Stat. 1064.

485. Telecommunications Act of 1996, Pub. L. No. 104-104, 110 Stat. 56.

telecommunications service undecided. A game of government ping-pong then ensued.

There has been an ever-changing set of decisions regarding this issue,[486] the result of which has often left only the Federal Trade Commission (FTC), an agency with limited authority, acting in the role of privacy regulator. In 2002, the FCC ruled that cable broadband companies were "interstate information services," thus giving itself little regulatory authority over them.[487] The Supreme Court concurred with the FCC's ruling.[488] The FCC then followed up by ruling that wireless Internet access was also an information service.[489] Under the Obama administration, the situation shifted, classifying broadband internet access services as telecommunications services,[490] thus allowing FCC regulatory authority over their privacy practices. The FCC developed and finalized privacy rules governing broadband internet service providers collection, use, and dissemination of customer private data.[491] Yet, in 2017 Congress repealed those rules through a resolution of disapproval.[492] Later that year, under the Trump administration, the FCC reversed the classification of Internet broadband as a telecommunications service.[493]

Although the FTC has broad jurisdiction to investigate "unfair or deceptive acts or practices in or affecting commerce,"[494] the FTC lacks jurisdiction over common carrier functions; that role belongs to the FCC.[495] Thus, the FCC, by changing the definition of what is and isn't a common carrier function, can limit the FTC's authority over common carrier aspects of information services.[496]

486. *See* discussion of the treatment of the internet *infra* Section V.A.

487. Press Release, FCC, FCC Classifies Cable Modem Service as "Information Service" (Mar. 14, 2002), https://transition.fcc.gov/Bureaus/Cable/News_Releases/2002/nrcb0201.html [https://perma.cc/J3AU-WC9Z].

488. Nat'l Cable & Telecomms. Ass'n v. Brand X Internet Servs., 545 U.S. 967, 1003 (2005).

489. *See* Appropriate Regulatory Treatment for Broadband Access to the Internet of Wireless Networks, Declaratory Ruling, 22 FCC Rcd. 5901 (Mar. 23, 2007).

490. Protecting and Promoting the Open Internet, Report and Order, Declaratory Ruling, and Order, 30 FCC Rcd. 5601 (Mar. 12, 2015).

491. Protecting the Privacy of Customers of Broadband and Other Telecommunication Services, Report and Order, 31 FCC Rcd. 13911 (Nov. 2, 2016).

492. S.J. Res. 34, 115th Cong. (2017).

493. Restoring Internet Freedom, Declaratory Ruling, Report and Order, and Order, 33 FCC Rcd. 311 (Jan. 4, 2018).

494. 15 U.S.C § 45(a)(1) (2006).

495. 47 U.S.C. § 201 et seq.

496. JULIE COHEN, COLUM. UNIV. KNIGHT FIRST AMEND. INST., HOW (NOT) TO WRITE A PRIVACY LAW 9 (2021), https://s3.amazonaws.com/kfai-

In practice, however, the FCC and FTC have sought to work out protecting consumer privacy. In 2015, the two agencies signed a Memorandum of Understanding (MOU) on their respective roles protecting consumer privacy rights stating, "the scope of the common carrier exemption in the FTC Act does not preclude the FTC from addressing non-common carrier activities engaged in by common carriers."[497] After the 2017 FCC decision to classify Internet broadband as a telecommunications service, the two agencies signed a new MOU that the FTC described as returning "jurisdiction to the FTC to police the conduct of ISPs, including with respect to their privacy practices."[498]

The courts have concurred; in 2018, for example, the Ninth Circuit ruled that common-carrier exemption of the FTC Act was "activity based," not "status based."[499] That is, a company "may be an interstate carrier in some instances but not in others"; aspects of a carrier not directly related to its role as a common carrier are legitimately subject to FTC regulatory action.[500]

Jurisdictional issues hampered government abilities to act on Internet privacy, but the real restrictions on FTC actions stem from the agency's lack of explicit authority to act on privacy violations. The FTC's rule-making process is complex and burdensome, forcing the agency to work on a case-by-case basis.[501] That may now be changing. In September 2019, FTC Commissioner Rebecca Kelly Slaughter gave a speech in which she explained that the FTC's case-by-case approach to privacy was ineffective;[502] in August 2022, the FTC announced it sought public comment on what rules to

documents/documents/306f33954a/3.23.2021-Cohen.pdf [https://perma.cc/PKP7-AHW8].

497. Press Release, FTC, FTC, FCC Outline Agreement to Coordinate Online Consumer Protection Efforts Following Adoption of The Restoring Internet Freedom Order (Dec. 11, 2017), https://www.ftc.gov/news-events/news/press-releases/2017/12/ftc-fcc-outline-agreement-coordinate-online-consumer-protection-efforts-following-adoption-restoring [https://perma.cc/HBK9-Q3HN].

498. *Id.*; *see* Restoring Internet Freedom: FCC-FTC Memorandum of Understanding (Dec. 4, 2017), https://www.ftc.gov/system/files/documents/cooperation_agreements/fcc_fcc_mou_internet_freedom_order_1214_final_0.pdf [https://perma.cc/4R6E-7UGK].

499. FTC. v. AT&T Mobility LLC, 883 F.3d 848, 850 (9th Cir. 2018).

500. *Id*. at 860.

501. CHRIS JAY HOOFNAGLE, FEDERAL TRADE COMMISSION LAW AND POLICY 146 (2016).

502. Rebecca Kelly Slaughter, Comm'r, FTC, Remarks at Silicon Flatirons: The Near Future of U.S. Privacy Law (Sept. 6, 2019) [hereinafter *Slaughter SFC Remarks*]; *see* Rebecca Kelly Slaughter, Comm'r, FTC, Statement of Commissioner Rebecca Kelly Slaughter Regarding the Commercial Surveillance and Data Security Advance Notice of Proposed Rulemaking (Aug. 11, 2022) [hereinafter *Slaughter ANPRM Statement*].

prevent harmful commercial surveillance should look like.[503] Depending on how this plays out, the impact could be a striking step forward for privacy.

Currently, an FTC privacy case can only be brought because of deceptive or unfair actions. "Deceptive" means that the company's actions did not follow its stated privacy policy, while "unfair" means that the company's practice "causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and is not outweighed by countervailing benefits to consumers or competition."[504] Because unfairness has historically been difficult to prove, the FTC has more frequently brought cases based on deception.[505] Meanwhile, the FTC was compelled to operate with one arm tied behind its back. Even as the Internet developed, the anti-regulatory approach of the U.S. Government was in force; its approach to industry collection of personal data was that industry should self-regulate.[506]

To call the U.S. policy a privacy failure is to vastly minimize the damage that has ensued. A 1998 FTC report on commercial websites to Congress observed, the "[I]ndustry's efforts to encourage voluntary adoption of the most basic fair information practice principle—notice—have fallen far short of what is needed to protect consumers."[507] Notice and choice was never intended to be the sole way to protect users' privacy.[508] Indeed, this operational viewpoint somewhat misses the point. Recall Westin's view of privacy included the ability to enjoy solitude (solely or as a small

503. *FTC Explores Rules Cracking Down on Commercial Surveillance and Lax Data Security Practices*, FTC (Aug. 22, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/08/ftc-explores-rules-cracking-down-commercial-surveillance-lax-data-security-practices [https://perma.cc/5LEN-Q3FD].

504. 15 U.S.C. § 45(n).

505. *See* J. Howard Beales III, *The FTC's Use of Unfairness Authority: Its Rise, Fall, and Resurrection*, 22 J. PUB. POL'Y & MKTG. 192, 195 (2003). Note that deception would allow a company to exploit a user's data so long as the company was explicit and its policy about doing so (this provided some protection to companies that chose not to publish a privacy policy).

506. MARTHA K. LANDESBERG ET AL., FTC, PRIVACY ONLINE: A REPORT TO CONGRESS iii (1998).

507. *Id*. at ii–iii. This was based on a survey of 1400 websites. Even with this occurrence, the FTC took a narrowly proscribed view of fair information practices: "core principles [that] require that consumers be given *notice* of an entity's information practices; that consumers be given *choice* with respect to the use and dissemination of information collected from or about them; that consumers be given *access* to information about them collected and stored by an entity; and that the data collector take appropriate steps to ensure the *security* and integrity of any information collected."

508. Fred H. Cate & Viktor Mayer-Schönberger, *supra* note 65, at 67.

unit), anonymity, and reserve.[509] The FTC's structure puts the user in the position of having to actively protect her solitude, anonymity, and reserve rather than having these be the default. That is not a "meaningful" privacy choice.

Despite the early signs of failure of industry self-regulation, the FTC continued to pursue the policy.[510] But the agency also began a small, but robust enforcement effort.[511] Though its jurisdiction was limited, the FTC's efforts caused a certain amount of fear, the result of which was that companies upped their game lest they, too, become a target.[512]

In 2000, the FTC submitted another report to Congress on online privacy. The Commission concluded that online privacy was a challenge, commended the private sector for its efforts on self-regulation, and recommended legislation requiring that websites collecting information from consumers follow four fair information practices: Notice, Choice, Access, and Security.[513] As privacy scholar Bob Gellman observed, the FTC's approach to the FIPS was quite incomplete, "The FTC's 2000 set of privacy standards restates, waters down, and leaves out some FIPs elements."[514]

A decade later, the agency itself acknowledged as much, stating, "Additionally, the emphasis on notice and choice alone has not sufficiently accounted for other widely recognized fair

509. Peter Swire, *Alan Westin's Legacy of Privacy and Freedom*, INT'L ASS'N PRIV. PROS. (Mar. 7, 2013), https://iapp.org/news/a/alan-westins-legacy-of-privacy-and-freedom/ [https://perma.cc/J328-NW2R].

510. *Consumer Privacy on the World Wide Web: Hearing Before the H. Subcomm. on Telecomms., Trade and Consumer Prot. of the H. Comm. on Com.*, 105th Cong. (1998) (written testimony of Robert Pitofsky, Chairman, FTC).

511. Solove & Hartzog, *supra* note 63, at 602.

512. *Id.* at 606; Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247, 274 (2011).

513. *See*, FTC, PRIVACY ONLINE: FAIR INFORMATION PRACTICES IN THE ELECTRONIC MARKETPLACE A REPORT TO CONGRESS iii (2000) ("(1) Notice- Web sites would be required to provide consumers clear and conspicuous notice of their information practices, including what information they collect, how they collect it (e.g., directly or through non-obvious means such as cookies), how they use it, how they provide Choice, Access, and Security to consumers, whether they disclose the information collected to other entities, and whether other entities are collecting information through the site. (2) Choice–Web sites would be required to offer consumers choices as to how their personal identifying information is used beyond the use for which the information was provided (e.g., to consummate a transaction). Such choice would encompass both internal secondary uses (such as marketing back to consumers) and external secondary uses (such as disclosing data to other entities).
(3) Access-Web sites would be required to offer consumers reasonable access to the information a Web site has collected about them, including a reasonable opportunity to review information and to correct inaccuracies or delete information. (4) Security -Web sites would be required to take reasonable steps to protect the security of the information they collect from consumers.").

514. Gellman, *supra* note 466, at 25.

information practices, such as access, collection limitation, purpose specification, and assuring data quality and integrity."[515] Alternatively, as Ella Corren put it, "[R]ather than disciplining the darker side of the market where personal information is exploited, the use of consent *facilitates* it."[516] She observed that, "*[E]mpty* consent (rarely informed nor intentional . . .) continues to be the most prevalent form of consent in markets, rendering consent the leading vehicle that *legitimizes* digital surveillance and other exploitations."[517]

In 2006, Fred Cate analyzed the notice-and-choice regime, observing that people did not read the notices and could not understand them, the combination of which resulted in an "illusion of choice."[518] Arguing that notice-and-choice had failed, Cate proposed that data protection rules should focus on use and "should target information processors that contribute directly and materially to the harmful use of personal information."[519] Cate's proposal of controlling, which has been raised again over the years, did not move forward.[520]

To be fair, the FTC was operating in uncharted waters. From the beginning U.S. law—and, thus, regulatory policy—focused on providing privacy protections against governmental misuse rather than misuse promulgated by the private sector. What federal protections existed against private sector misuse of citizen data were sectoral and many aspects of data collection and use were simply left unregulated.

Despite the lack of federal laws, the FTC had a route for protecting privacy. Since 1938, the agency has had responsibility for protecting against "unfair or deceptive acts or practices in or affecting commerce."[521] Yet, the FTC has been slow to apply that responsibility to Internet commerce. That can be explained by the U.S. Government's interest in Internet commerce. Given the current dominance of the Internet powerhouses and their consequent effect on Wall Street, it may be hard to imagine that in

515. *Id.*; FTC, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: A PROPOSED FRAMEWORK FOR BUSINESSES AND POLICYMAKERS 20 (2010). (Gellman observed that the comment disappeared in the final version of the report).

516. Ella Corren, *The Consent Burden in Consumer and Digital Markets*, 36 HARV. J.L. & TECH. (forthcoming 2023).

517. *Id.*

518. Cate, *supra* note 65, at 360–63, 366.

519. *Id.* at 373.

520. Cate & Mayer-Schönberger, *supra* note 65, at 69; *Control Use*, *supra* note 65, at 504; Joel R. Reidenberg et al., *supra* note 65, at 486, 488; Bellovin, *supra* note 65; *Landau Privacy Comments*, *supra* note 65.

521. 15 U.S.C. § 45(a)(1).

the early days of the public Internet, there was concern that excessive regulation of network activity would stifle innovation.[522] But there was such concern, leading to a policy of allowing the industry to self-regulation.[523]

The lack of privacy regulation extended to the communications arena. While, except for the provision of services, telecommunications carriers may not disclose telephone subscriber information[524] (this is more formally known as customer proprietary network information or CPNI), no such legal restrictions constrained Internet Service Providers (ISPs). As we will see, within a decade, researchers raised concerns about the privacy intrusions of ISP's use of metadata.

B. U.S. Privacy Policy in Practice

From the moment the public Internet arrived, the issue of user privacy and the Internet companies' use of customer data became highly controversial. In some ways, reminiscent of the debates of a century before, when Samuel Warren and Louis Brandeis wrote that "the individual should have full protection in person and in property,"[525] privacy theorists clashed with the new business models that used personal information at a previously unimaginable scale to provide services that had not previously been possible.

This clash between economic opportunity and user privacy was overlaid with issues of choice and freedom of speech. Julie Cohen tackled these four issues in a trenchant article in 2000, concluding that "[P]eople may have legitimate reasons for trading privacy for value in particular cases,"[526] observing that, "[C]onsent [to share information] cannot be meaningful as to unknown uses or unspecified recipients"[527] and that "[T]he quality of consent

522. *See, e.g.*, Solove & Hartzog, *supra* note 63, at 598.

523. This statement is based partially on Susan Landau's experience while at Sun Microsystems from 1999 to 2010; *see id.* at 598–99; KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, PRIVACY ON THE GROUND: DRIVING CORPORATE BEHAVIOR IN THE UNITED STATES AND EUROPE (2015) (Kenneth Bamberger and Deirdre Mulligan have argued that the less prescriptive rules that the FTC used in the 2000s had resulted in stronger privacy protections than were first acknowledged and, at times, in stronger protections than had occurred in Europe. While this issue is out of scope for our paper, we will note that the E.U.'s General Data Protection Regulation appears to have shifted that balance).

524. 47 U.S.C. § 222(c)(1).

525. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy,* 4 HARV. L. REV. 193, 193 (1890).

526. Cohen, *supra* note 55, at 1432.

527. *Id.* at 1433.

attenuates over time."[528] Cohen unpacked the conflict between data privacy and issues of property, and choice, freedom of speech, and knowledge.

Cohen was almost rueful in her reminder that in several domains, including intellectual property (IP), use restrictions exist; they are called licenses. Cohen pointed out that "defining the bounds of a property interest always requires choices between [differing] liberty claims."[529] As such use restrictions function successfully in IP, it is not unreasonable to imagine that putting use restrictions on particular of personal information would provide a useful balance between privacy and value.[530] The issue then becomes: which use restrictions will help achieve this balance?

Cohen argued against a situation in which choice is left purely to the individual. She noted that people have trouble gauging how their private information might be used, especially in secondary and tertiary ways.[531]

In a 2021 report, the FTC reported on the privacy practices of six major U.S. ISPs, observing how the information the providers collect could be shared with "property managers, bail bondsmen, bounty hunters, or those who would use it for discriminatory purposes."[532] Consumers had little awareness of the extent of data collected or how it could be combined. Two decades earlier, Cohen had already observed that users were left with "a bewildering constellation of decisions about which choices to privilege, which to facilitate, and which to restrict."[533] That was just one of the complexities facing users. Website privacy policies govern the use of data on that site but not on sites with which the data is shared.[534] As Daniel Solove noted, "[L]ittle bits of innocuous data can say a lot in combination."[535] Users are unlikely to understand how data aggregation works, at least in practice—or the privacy threats that then result.[536]

Cohen observed that the efficiency that comes from markets having information about customers does not necessarily lead to the best choices for consumers.[537] Letting Amazon know when you get

528. *Id.*
529. *Id.* at 1386.
530. *Id.*
531. *Id.* at 1396.
532. FTC, *supra* note 60, at ii, iv.
533. Cohen, *supra* note 55, at 1401.
534. Lorrie Faith Cranor, *Necessary but Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice*, 10 J. ON TELECOMM. & HIGH TECH. L. 273, 274 (2012).
535. Solove, *supra* note 64, at 1890.
536. *Id.* at 1889–90.
537. Cohen, *supra* note 55, at 1407.

up and go to bed, what movies you watch, and which groceries you buy may not make you a happier or healthier consumer, but it will certainly enable Amazon to exert a great deal of power over your purchases—and your life. And regarding "knowledge," Cohen deftly argued that commercial use of private information is not so much about the publication of private data as it is about the exchange of information as property.[538] The latter is not only regulatable; it is commonly regulated across a variety of industries.[539]

The alternative to no regulation or law was a choice to let the market decide.[540] Placing few restrictions on how industry used personal data, the U.S. pursued a policy of business "self-regulation" during the 1990s and early 2000s.

While Cohen's focus was at a different time—P3P[541] was a prime topic of discussion and Facebook did not yet exist—the point that Cohen made on encumbrance and individual choices about tradeoffs still hold. Here, the situation becomes quite interesting for metadata and telemetry. As Paul Ohm showed in 2009, users had no control over how ISPs handled their metadata.[542]

In 2009, Ohm examined how ISPs were not only delivering their users' communications; they were also examining what was in those communications.[543] Ohm noted many reasons why ISPs might be spying on their customers' activity—including requirements that resulted from the 1994 Communications Assistance for Law Enforcement Act;[544] compliance with the Sarbanes-Oxley Act of 2002;[545] Graham-Leach-Bliley Act;[546] the Health Information Portability and Accountability Act;[547] and movie and music distribution and compliance with copyright law—

538. *Id.* at 1416–17.

539. *Id.*

540. *Id.* at 1401-2.

541. Platform for Privacy Preferences, or "P3P," was a project to enable web browsers to describe how they intended to use collected information. Rigo Wenning, *Platform for Privacy Preferences (P3P) Project*, W3C (Feb. 2, 2018), https://www.w3.org/P3P/ [https://perma.cc/8WT5-63Z7]. The project, started in 2002, was controversial in part because it was seen as too complex to use and thus of little practical benefit. It quietly ended in 2007.

542. Ohm, *supra* note 56, at 1417.

543. *Id.*; *see* discussion of 2021 FTC efforts in this space *supra* Section IV.A and *infra* Section V.B.

544. Communications Assistance for Law Enforcement Act of 1994, Pub. L. No. 103-414, tit. I, § 103, 108 Stat. 4280, 4281 (codified at 47 U.S.C. § 1002(a)(2)(B)).

545. Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745 (codified as amended in scattered sections of 15 U.S.C. & 18 U.S.C.).

546. Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999) (codified as amended in scattered sections of 15 U.S.C.).

547. Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 100 Stat. 1936 (codified as amended in scattered sections of 29 U.S.C. & 42 U.S.C.).

and the possibility of emulating Google by serving ads.[548] Ohm observed that ISPs can see anything[549] that comes across the wires to or from the user. At a time when public Wi-Fi was far from ubiquitous, a user's work or home ISP was the main way of connecting to the Internet. Thus, the ISPs had the potential for tremendous insight into the user's activities—perhaps even more than Google or Facebook (that changed with the arrival of secure connections that encrypt packet content; even so, the ISP knows the address of the site you're visiting, the address of the one you're visiting next, and so on).[550]

Ohm described the following abilities of an ISP:

> [An ISP] can track your ailments, emotions, and the state of your relationships. It can learn your travel plans, big dates, and trips across town to do mundane chores. It can know how often you call your mother, e-mail your sister, or send gifts to your grandfather. It can know what you read, watch, buy, and borrow. And unlike Google, it already has an authoritative record of your home address, because it sends your bill there each month, and very likely your credit card and bank account numbers as well.[551]

The situation is much more extreme now than at the time Ohm wrote. The 2021 FTC report on ISP privacy practices observed that the vertical integration the industry experienced over the last decade provides a much greater tranche of personal information, "without [the companies] having to explain fully their purposes for such collection and use."[552]

Ohm presented three motivations for ISPs to monitor users' content: necessity, convenience, and voyeurism.[553] He observed

548. Ohm, *supra* note 56, at 1426–27.

549. *Id.* at 1438. Note that Ohm was writing in 2009 before the use of https and encrypted Internet content became ubiquitous.

550. In 2022, Vodafone and Deustsche Telekom began experimenting with pseudo-anonymous tokens based on a user's IP address. *Find Out More About TrustPid*, TRUSTPID, https://trustpid.com/findoutmore [https://perma.cc/FX9X-XCH4]. Pseudononmyity comes from the fact that the user receives different tokens for each website participating in the effort. The service provider is, of course, in a position to link the sites and thus fully know the user's activities. *Id.*; *see also*, Chris Stokel-Walker, '*Supercookies' Have Privacy Experts Sounding the Alarm*, WIRED (June 28, 2022, 12:05 PM), https://www.wired.com/story/trustpid-digital-token-supercookie/ [HTTPS://PERMA.CC/Z7GN-HNPG].

551. Ohm, *supra* note 56, at 1445.

552. FTC, *supra* note 60, at i.

553. Ohm, *supra* note 56, at 1462–74; *see* WESTIN, *supra* note 61, at 55 (describing the type of voyeurism "used in the social rather than its clinical sense" as "refer[ring] to the tasteless pursuit and aggressive exposure of the privacies of personal life").

that Cisco's Netflow protocol, which is used for network monitoring, does not provide email addresses, though it did provide port numbers, which reveal what applications are being used.[554] Ohm noted, "[NetFlow] gives network engineers a broad window into the activity on their networks, but it throws away much of the most sensitive data [before doing so]."[555] With that observation, Ohm effectively did away with the ISP's argument of necessity for monitoring user content; they weren't using the data that NetFlow was providing. Such tracking was for ISP convenience—or voyeurism.

Ohm's article, however, was published before smartphones became the main way in which people access the Internet.[556] Now a majority of accesses to websites come from smartphones.[557] A mobile device that provides metadata and telemetry to an Internet company—an app or OS manufacturer—may share information that you're walking or traveling in a car (data derived from the device's accelerometer) or where you are (data derived from GPS, Bluetooth, or Wi-Fi signal).[558] It could share your location—Planned Parenthood[559]—with the same site that provided you with abortion information. It might be able to share the information that every night your phone is in proximity with someone else's. By tracking your movement, it could learn how much you exercise, whether you appear to be stumbling home after time at a bar, or with whom you are spending time at work—and afterwards. Though the information shared by smartphones and other mobile devices with Internet companies is not the full stream of communications metadata going through a home or work ISP, the amount of data can be substantial; it has the potential to reveal even more information than the metadata supplied to the ISP.[560]

Consider, as an example, a 2016 patent issued to Facebook for enabling suggestions of connections to users.[561] The patent is

554. Ohm, *supra* note 56, at 1472–92. Port numbers could potentially be used in tracking network flow issues. *See Id.* at 1492.

555. *Id.* at 1472.

556. In 2020, 61% of web accesses occurred over mobile phones; this percentage has been steadily increasing since smartphones were first introduced. Eric Enge, *Mobile vs. Desktop Usage in 2020*, PERFICIENT (Mar. 23, 2021), https://www.perficient.com/insights/research-hub/mobile-vs-desktop-usage [https://perma.cc/CRL2-VDEZ].

557. *Id.*

558. *See* discussion *supra* Sections III.A–III.B.

559. More precisely, it is your phone's location that is being shared.

560. *See* FTC, *supra* note 60, at 23.

561. Systems and Methods for Utilizing Wireless Communications to Suggest Connections for a User, U.S. Patent No. 10,111,059 (filed Aug. 28, 2017) (issued Oct. 23, 2018).

intended to enable two users who have been in close proximity over several occasions—perhaps on a commuter bus, perhaps at a conference—and who have not exchanged contact information to be able to connect later. The application proposes the following:

> [I]f it is determined that there is a sufficiently high likelihood that [the two users] have met (or that their meeting was sufficiently significant), then the first user can be provided with a suggested connection specifying the second user, and the second user can be provided with a suggested connection specifying the first user. The first user can then choose to add the second user as a connection, or vice versa.[562]

The basis for suggesting a connection includes a:

> plurality of factors . . . includ[ing] at least one of an inferred locational proximity between the first user and the second user, a frequency of inferred meetings between the first user and the second user, a duration of each of the inferred meetings between the first user and the second user, or a pattern of occurrences of inferred meetings between the first user and the second user.[563]

Close proximity would be determined by "locational data of the computing system based on at least one of Global Positioning System (GPS), WiFi, radio signal modulation, or geo-tagging indicates that the computing system is within a specified allowable distance from . . . a source of the second wireless communication."[564] The system would use "data from at least one of a gyroscope, an accelerometer, or a motion processor of the computing system indicates that the computing system is moving in a movement pattern similar to that of a source of the second wireless communication."[565] Facebook would be using communications metadata and telemetry information, information a user is largely unaware is being collected and certainly unaware of the uses to which it is being put.

Imagine Facebook employs this technology to figure out that one of the authors of this paper has ridden a bus several times near someone who commutes at the same time she does. Consider the implications. When one of us rides a city bus to Tufts, she likes to

562. *Id.* col. 5 ll. 11–18.
563. *Id.* col 2 ll. 31, 34–39.
564. *Id.* col. 3 ll. 5–10.
565. *Id.* col. 3 ll. 12–16.

read or perhaps gaze out the window. She might talk with someone else she knows who is also on the bus. She has contact information for that person, but there are others who might ride the bus when she does but whom she does not know—and maybe doesn't wish to. Under the Facebook patent, if she were a Facebook user, she might receive an introduction to one of those passengers—an introduction she did not ask for and almost certainly did not want.

To a city dweller who knows how to establish clear boundaries in public spaces, such an intrusion is highly disturbing. Were either one of us to receive such an introduction, we would find it creepy, and we would feel that Facebook had intruded upon our privacy. If such a person wants to start a conversation with someone they do not know, they should do so deliberately—and not as a result of an app that will increase the number of Facebook connections.

A notable contrast to the Facebook patent is the Exposure Notification (EN) apps developed by researchers in Europe, North America, and Australia during the early days of the COVID-19 pandemic.[566] Researchers focused on developing EN solutions that exchanged user identifiers over Bluetooth without enabling users to track who was nearby.[567] After all, proximity information can reveal activities that people might choose to keep private. Perhaps your spouse would prefer you not meet a former romantic interest, but you ran into each other on the subway and stopped for coffee. Or your boss has asked you not to meet with your friends who work for a competitor, but you play volleyball with them weekly—and you're not about to give up the activity. Who people spend their time with—and who is close by them—is information that people often want to keep private.

EN apps like these are structured to keep peoples' information private from those they have encountered even while still informing users when they've been in close proximity to someone who was contagious with the COVID-19 virus.[568] The apps went one step further. While researchers were developing exposure-notification apps that kept user proximity information private, Google and Apple were building an underlying infrastructure to support these

566. These include the Decentralized Privacy-Preserving Proximity Tracing (DP3T) group and two groups named PACT (one was later renamed "Common Circle); *see* RONALD L. RIVEST ET AL., MASS. INST. TECH., THE PACT PROTOCOL SPECIFICATION 2 (2020), https://pact.mit.edu/wp-content/uploads/2020/11/The-PACT-protocol-specification-2020.pdf [https://perma.cc/3BBT-WSFU].

567. *Id.* at 3.

568. Carmela Troncoso et al., *Decentralized Privacy-Preserving Proximity Tracing*, COMMC'NS ASS'N COMPUTING MACH. 48, 50–51 (2020).

apps.[569] The Google-Apple Exposure Notification (GAEN) infrastructure was designed to prevent exposure-notification apps from collecting location information, another aspect of protecting privacy.[570]

Thus, the Facebook patent regarding introducing two users describes an application that takes essentially the opposite approach to that of the GAEN apps. Users would be less than fully aware of the data collection and its purpose and, thus, would not be in a position to provide informed consent. Insofar as we know, this patent may still be just an idea, not an actual Facebook application. However, the application points to the heart of the concerns in this paper: the use of data from which very private information can be derived and yet over which the user has essentially no control regarding sharing. This is far from providing a user with meaningful privacy choices.

C. The Lack of Meaningful Privacy Choices for Use of Metadata and Telemetry

Ohm's 2009 analysis showed that it was not necessity that drove the ISPs to seek to inspect packet contents. He posited voyeurism might be an aspect of the ISP's interest in "knowing" their customer,[571] voyeurism in the sense of pursuit and exposure of the "privacies of personal life."[572] Decades earlier, Westin had raised such concerns in discussing the invasive technology of the 1960s that were new in the sense of providing methods for easily eavesdropping and spying on people.[573]

The world of parabolic antennae and tiny bugs pales in comparison to our current situation, an environment of constant surveillance by our personal devices and an industry built on encouraging the information learned from these intrusions. Austin has described this as "the creation of an entire social-technical infrastructure that is encouraging new forms of curiosity,"[574] and indeed, the business model of online social networks thrives on the oversharing of private information. The voyeurism that Ohm

569. GOOGLE, EXPOSURE NOTIFICATION: FREQUENTLY ASKED QUESTIONS 2 (2020), https://www.blog.google/documents/63/Exposure_Notification_-_FAQ_v1.0.pdf/ [https://perma.cc/E5JC-WQYJ].

570. *See Google COVID-19 Exposure Notifications Service Additional Terms*, GOOGLE, https://blog.google/documents/72/Exposure_Notifications_Service_Additional_Terms.pdf [https://perma.cc/EWQ6-8PY4].

571. Ohm, *supra* note 56, at 1471. Note that these aspects are content. *See* Bellovin et al., *supra* note 50, at 1.

572. WESTIN, *supra* note 61, at 55.

573. *See* discussion of Westin's work *supra* Section III.A.

574. Austin, *supra* note 62, at 71.

discusses regarding ISP deep packet inspection pales in comparison with that of the Internet companies whose business requires the sharing of such information with so-called friends and contacts.

Our species has had millennia to learn to develop the safety rules of sharing personal information with other people. We know how to express nuanced sharing ("You can let Alice know, but not Eve," "You can tell Bob the first part, but please don't share the second with him"). Yet we have had less than a generation to learn the rules—to the extent that they exist—of how to restrict such sharing when Internet companies are the ones that hold the data. Conversations on how to limit further sharing of personal information fail to be explicit with the consequences that the Internet companies over permit the use of the data. Though the Cambridge Analytica[575] scandal was huge, it was, in many ways, but a small step outside what were vastly over permissive rules regarding the sharing of data about a user.

Austin observed that Westin's term "reserve," the capability of respecting another's personal boundaries and not intruding beyond those, requires a shared sense of civility.[576] What works well when two work colleagues are discussing a third or a pair of friends are dissecting a relationship includes social context and well-understood boundary setting that is lacking when one of the people is instead interacting with Facebook, TikTok, Twitter, or LinkedIn. A shared human culture is replaced by a set of consent rules that are unclear in import and impossible to navigate even for a sophisticated user.[577]

The FIPPs solution of choice fails to be a reasonable solution for controlling the use of metadata and telemetry for three reasons that effectively prevent user choice. First, the user is unaware of information sharing. Second, there is a fundamental information asymmetry between the user, who has little to no understanding of the uses to which the information will be put, and the company that is collecting and using the data (the use might be to share it with a third party). Third, the complexity of the actions impedes effective user control.

A user asks her smartphone to perform a task, perhaps sending an email or a text, starting a VoIP call, or downloading a webpage. In doing so, she transmits information in the packet header to accomplish the task. Unlike the search terms a user might supply

575. *See generally* Joanne Hinds et al., *'It Wouldn't Happen to Me': Privacy Concerns and Perspectives Following the Cambridge Analytica Scandal*, INT'L J. HUMAN-COMPUT. STUD., Nov. 2020, at 1.

576. Austin, *supra* note 62, at 73.

577. *See* Solove, *supra* note 64, at 1900–03.

to Google or Bing, the user has essentially no knowledge of what is in the packet header. This is not the world of *Smith*, in which the user knows they are providing telephone numbers to the phone company in order to complete a call.[578] It is not even the world of *Carpenter*, in which the user knows that the local cell tower is being used for the call's transmission—and that if the user is in motion during the call (on a bus, in a car, on a train), the records of the call may appear on the different cell towers serving it. Rather, it is a world in which along with the IP address of the destination of the user's message is such information as packet length—a packet-header piece of information that users are unlikely to know exists, let alone be transmitted—that may reveal the words spoken during an encrypted VoIP call.[579]

Users are not engineers. Though they may delight in the applications their smartphones provide, users are unaware what mechanisms enable the image on the screen to properly orient as the phone shifts (gyroscope) or allow a mapping app to handle changes in direction (magnetometer). They do not realize that a proximity sensor on their phone prevents the touch screen from registering a touch if their ear happens to touch the screen while they are on a call.

There is no reason for a user to know these things; in our modern world, we drive cars and ride elevators without understanding how internal combustion engines work or elevator counterweights operate. Yet, there is a big difference when user information, such as telemetry data, exits the phone and is used, not to orient the display on the device or inform the user to turn left at the intersection, but instead is employed by a data collector to determine where the user is traveling—and serve her an ad that fits her location. It is as if a Prius were transmitting messages back to Toyota about whether the driver tailgates or the Otis elevator were reporting to headquarters regarding on which floor particular hotel guests exited. Users lack the technical expertise to understand how the computations that orient the screen or direct them on a mapping application actually work. They are, thus, not

578. Susan Sharon, *Telephone Operators of Another Era Gather to Reconnect*, NPR (Dec. 28, 2021, 5:02 AM) https://www.npr.org/2021/12/28/1066402448/telephone-operators-of-another-era-gather-to-reconnect#:~:text=Before%20smartphones%2C%20landline%20telephones%20were,fire%20department%20in%20an%20emergency [https://perma.cc/AE66-VGRW] (Phone bills of the time provided itemized information about long-distance calls, so users were certainly aware of the service provider's collection and knowledge of numbers dialed.).

579. *See* Charles Wright et al., *Language Identification of Encrypted VoIP Traffic: Alejandra y Roberto or Alice and Bob*?, 16TH USENIX SEC. SYMP. 43, 52 (2007), *and* Andrew M. White et al., *Phonotactic Reconstruction of Encrypted VoIP Conversations: Hookt on fon-iks*, 2011 IEEE SYMP. ON SEC. & PRIV. 3, 3.

in a position to make educated decisions about when telemetry information should leave the device.[580]

An engineer might realize that the data from an accelerometer, gyroscope, and magnetometer could reveal travel paths to an app or data collector. Few others would make such a connection. A privacy expert might note that phone swipes could show an OS provider that the phone home screen in not set up in a user-optimal way—or such data could reveal user tension and anxiety. Most users would not anticipate the latter use of the swipe data.

Even worse, the user is not positioned to effectively control the sharing of her communications metadata or telemetry information. The ability of a user to determine in real time which uses she is willing to permit to be used for other purposes is beyond implausible. That is especially the case when one considers the rate at which those queries would occur. Because the packet headers for each type of communication—whether email, VoIP, https, etc.—have multiple fields, a single transaction would produce multiple queries.

The same situation is true of device telemetry. Contemplate, if you will, mapping apps. In 2012, for example, Google Maps was sampling a user's locations as much as ten times a minute;[581] one can safely assume that a decade later, to be competitive mapping apps sample at least that rate. A user is unlikely to be able to determine which minutes of collection are fine, and which are not. Consider the following hypothetical: To avoid traffic jams, a contractor employs such a mapping application. Without much consideration of the issues involved—and, in particular, unaware of what telemetry data is collected or how it may be shared—the contractor has no qualms about having information regarding his morning commute collected. The sampling data reveals a daily ten-minute stop at Dunkin'. Collected by the mapping app, information about those stops is provided to a data broker and then sold to a health insurance company. The driver's health insurance rates rise. Could the contractor have determined that, while sharing the daily commute data was fine, he should shut off the app when nearing Dunkin'? That is hardly likely.

If the user is to exercise choice regarding usage of the communications metadata or device telemetry information, she will

580. The same complexity appears in the issue of when IP communications metadata is being shared with a third party; *see* Bellovin et al., *supra* note 50, at 11.

581. Smartray05, *Location History Sampling Info*, GOOGLE MAPS HELP (July 28, 2022, 1:42 PM), https://support.google.com/maps/forum/AAAAQuUrST8ldKaXij4c_0/?hl=en&gpf=%23!msg%2Fmaps%2FldKaXij4c_0%2FqsnMfunamM4.&msgid=qsnMfunamM4 [https://perma.cc/JKC3-WY4D].

need to respond separately to the release of information of each field of the packet header and type of sensor data. Adding to that complexity is the fact that, for each type of communication, there will be different personal information that may be revealed. Such complexity would overwhelm a user, preventing her from making informed choices. Industry studies estimate that the average user accessed 46 apps per month in the first half of 2021,[582] a fact that only exacerbates the already insurmountable problem of user control of personal data.

It is implausible to expect even an expert to be able to handle such a set of queries—except perhaps by globally answering "no" to all of them. The user would be asked to control a flow of information about which she has neither understanding nor capability of controlling. Although it may appear that the user has choice and can limit the use of her communications metadata and telemetry information, this is false in practice.

In 2012, the FTC recommended that, "For practices requiring choice, companies should offer the choice at a time and in a context in which the consumer is making a decision about his or her data."[583] This recommendation is implausible for metadata and telemetry. As we already noted, there would be simply too many requests; the user would be overwhelmed. Plus, if the metadata or telemetry information is stored—and currently there is no way to prevent this—then the query about usage may occur long after the time of collection, perhaps at a time when the Internet company has determined a new way to use the information it has collected.

Consider, for example, that the company collected accelerometer data from a customer "to provide a better user experience" (we have all seen privacy policies with such claims). But, two years later, the company decides to use accelerometer information to develop its database of user locations. Even if the company were to query the user about whether she gives permission to do so, if the user is queried two years after the data was collected, she is no position to remember whether the collected information was connected to an event she would rather keep private.

582. Stephanie Chan, *U.S. Consumers Used an Average of 46 Apps Each Month in the First Half of 2021*, SENSOR TOWER (Aug. 2021), https://sensortower.com/blog/apps-used-per-us-smartphone [https://perma.cc/NV53-7UDX].

583. FTC, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS 48 (2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf [https://perma.cc/TP9Y-76UG].

In short, users have effectively no control over providing metadata and telemetry, and they are not in a position to understand ultimate usage. This is compounded by the fact that there is too much information coming too quickly for users to be able to make informed choices. In such a situation, user consent is an oxymoron; it is implausible that users are able to make knowledgeable consent choices.

The situation is even worse than that. Recall the intimate information collected by the MyDays app.[584] It is hard to imagine a more intrusive set of knowledge than an ad network system having "actionable insights" into the status of a woman's periods—and possible pregnancies. Who exactly is intended to have access to this shared information?

In the U.S., data brokers operate in the shadows. While the public is largely aware that Google, Facebook, and other Internet companies collect and use consumer data to serve ads and sell products and services, the public is far less aware of data brokers, who collect and operate on user information without having a direct relationship with the user.[585] This means that the user is usually not a position to control the data broker's exploitation of the user's personal information.[586]

Another app, OkCupid, shared accelerometer, gyroscope, and magnetometer data with AppsFlyer,[587] a company that "powers predictable app growth."[588] While the reader has learned that information from a gyroscope and accelerometer of two smartphones can reveal whether the two users are in the same car, bus, or train, the average user will have no idea that the data from her phone's sensors can be that revelatory. That puts the user in the position where her phone sensors reveal extremely private information to an app and its ad networks while leaving the user in the dark both that it has done so and what personal information might be exposed as a result.

We of course, do not know why OkCupid collects magnetometer, gyroscope, and accelerometer information from a user's device nor why the company shares that data with AppsFlyer. But, consider the following hypothetical. Two users arrange a date through OkCupid. At some point, they leave the restaurant and head out in the evening together. Since they know

584. CLAESSON & BJØRSTAD, *supra* note 239, at 40–42.

585. MAJORITY STAFF OF S. COMM. ON COM., SCI., AND TRANSP., *supra* note 248, at ii.

586. Recall that the situation can be even worse. *See* LEANPLUM, *supra* note 248. *See also* discussion *supra* Section IV.B.ii.

587. CLAESSON & BJØRSTAD, *supra* note 239, at 46.

588. *See* APPSFLYER, https://www.appsflyer.com/ [https://perma.cc/Y2TQ-96VJ].

that location outside can be tracked somewhat precisely, the users, wanting to keep their actions private, carefully shut off their apps' access to GPS location data, but they leave their phones on. They know they can be tracked by CSLI, but they also know that cell site location is less precise than GPS tracking; they don't feel there is much privacy risk in the potential use of CSLI data. The users believe they have successfully maintained privacy about where they spend the night.

Unbeknownst to the users, sensor information from their phones—magnetometer, gyroscope, and accelerometer readings—is shared with AppsFlyer even while GPS is off. The ad network is in a position to learn that the two users shared a car ride, then together entered an apartment building, went up a flight of stairs, walked together down a hallway, etc. The users did not—and effectively could not—know this information was being provided or analyzed. They were not in a position to know what sensor information was being shared with AppsFlyer, nor that the sensor information could reveal where the two had spent the night. So imagine their surprise when they open their phones the next morning to see ads that say, "In a new relationship? How about a candlelit dinner at . . .?" or "Is it time for a new form of birth control? Try X, the safe and secure method."

Such a response would be an enormous violation of social norms. It would also be a violation of the users' intents. Yet, there is nothing in the user agreements with OkCupid or in FTC requirements that would prevent AppsFlyer from using the sensor information that OkCupid obtained from the users' phones to determine where they went after their initial rendezvous. How is it that Internet companies are able to collect and use intimate information that allows the tracking of such private activities without the users themselves having any awareness of the possibility? Simply put, it stems from three issues: the inability of users to control either collection or use of this data, the ubiquity of collection of metadata and telemetry, and the potentially invasive nature of information derived from this data.

If users are to have any privacy at all, use—except for the purpose for which the data collection was authorized—must end. In the next section, we suggest two ways to go about doing so.

IV. HOW WE CAN ACT

In 2015 Joel Reidenberg, Cameron Russell, Alexander Callen, Sophia Qasir, and Thomas Norton looked at the effectiveness of notice and choice in providing users with the privacy they

sought.[589] Building on the work of Daniel Solove,[590] Lorrie Cranor,[591] and others, Reidenberg et al. found different failures of the notice-and-choice regime: (1) the user lacks adequate information to make an informed choice about whether to share personal information; (2) the system is impractical (the user must read far too many privacy policies, make too many decisions, and lacks the capability to control data flows to third parties); (3) users' cognitive bias causes them to confuse a privacy policy with privacy protections—as opposed to a policy that describes company actions on personal data; and (4) negative externalities mean one user's release of information may impact another user's privacy.[592] These problems occur with ISP, Internet company, and app usage of metadata and telemetry—and in many instances, cause far greater intrusiveness than what the data users knowingly allow.

Individuals have little choice in providing the communications metadata and software and device telemetry that enables surveillance; in many cases, metadata and telemetry must be provided for a user to receive content or for a device to properly display it. Given the nature of the information provided—bits of data, often sent multiple times a minute—consent, let alone informed consent, is not possible. The unfairness of the situation—the inexplicable circumstance in which information that the user did not explicitly share and cannot prevent providing—demands resolution.

We discuss here how we might arrive at such protections. In Subsection A, we briefly examine the failure of U.S. legal tools for protecting against privacy invasions by the private sector. In Subsection B, we consider what tools could be employed to do so, concluding that this is best done through controlling use.

In Subsection C, we propose what might appear to be a radical solution: collection and use of metadata and telemetry information should be limited solely to the purpose for which the data was collected. In other words, use of this data by ISPs, OS providers, platforms, and apps for purposes other than delivery and display of content, debugging, fraud prevention, and provisioning for future services would be off-limits. We also propose exceptions in cases of the use of aggregated data for public health emergencies and for a public or peer-reviewed research project that is in the public interest and adheres to well-established standards for deidentification.

589. Joel R. Reidenberg, et al., *supra* note 65, at 486.

590. *See generally* Solove, *supra* note 64.

591. *See generally* Cranor, *supra* note 534.

592. Reidenberg et al., *supra* note 65, at 490–95.

Our proposal would, thus, enable users to safely and securely employ online services confident that *only the data that they are aware of supplying* could be used for a purpose other than content delivery and display and proper functioning of services. Though this proposal may appear radical, it was the appropriation of metadata and telemetry data for uses outside of delivery and display of content that occurred that is actually the highly radical step. Our recommendation simply returns us to the situation of status quo ante of two decades ago. The proposal's simplicity—no use of metadata and telemetry information for purposes other than delivery and display of content and security protections—is also its strength. We outline the requirements and the avenues for proceeding.

A. Can Current U.S. Law Protect Against Invasive Uses of Metadata and Telemetry?

In 2015, Reidenberg et al. studied roughly fourteen years of federal privacy litigation and FTC actions to understand the extent to which the notice-and-choice regime reflected users' actual privacy choices.[593] Coalescing cases from the same action, the researchers studied 165 class action suits arising from 89 different events and 116 distinct FTC cases, all involving notice and choice.[594] Reidenberg et al. found that the harms considered in the court cases and FTC actions fell into four categories: unauthorized disclosure of personal information, surreptitious collection of personal information, inadequate security for personal information, and wrongful retention of personal information.[595] Note that two of these—surreptitious collection and wrongful retention—inevitably occur with the collection and use of metadata and telemetry information, while the other two may or may not occur as well.

If the intent of notice-and-choice process was, as Reidenberg et al. wrote in 2015, "to enable users to make meaningful, informed decisions regarding their privacy,"[596] it seemed the system failed more often than not. Because the courts sought tangible harms in order to award the plaintiff,[597] rather than accepting that the

593. *Id.* at 496.

594. *Id.* at 498–99, 507–08.

595. *Id.* at 513–17*;* Reidenberg et al. note that these categories are similar to ones found by Daniel Solove and Woodrow Hartzog, who looked at deception and unfairness actions. *See generally* Solove & Hartzog, *supra* note 63.

596. Reidenberg et al., *supra* note 65, at 496.

597. *See*, *e.g.*, Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 B.U. L. REV. 793, 793–94 (2022).

intangible harms that result from a privacy breach are damage in and of themselves,[598] the law was not effectively protecting privacy.

Ryan Calo had taken exception to this "privacy law exceptionalism," arguing that the high bar for privacy harm is suspicious and contending that the fact that the harm is ethereal should not diminish the fact that the harm is real.[599] He challenged scholars to "figure out the nature of this harm."[600]

Citron and Solove responded.[601] They observed that first, privacy harms often involve future uses of data and though the effects of a single privacy harm can be small, their aggregation can have a large effect.[602] Delineating a set of seven privacy harms that result from loss of privacy: physical, economic, reputational, discrimination, relationship, autonomy (including coercion, manipulation, failure to inform, thwarted expectations, lack of control, chilling effects), and psychological harms (including emotional distress and disturbance),[603] they pointed out that privacy litigation has three purposes: compensation, deterrence, and equity. Thus, "courts should require harm [be demonstrated] to the extent that claims are brought to secure compensation . . . [F]or contract cases, courts should enforce the contract. Courts should use remedies, such as specific enforcement, restitution, or recission."[604] But, in the latter set of cases, there should be no need to specifically determine tangible economic or physical harm.

Let us look at hypotheticals described earlier:

- The young person with little to no credit history who seeks a loan; their calling records are assessed to determine if they pose a good credit risk;[605]
- The woman who finds her contact information has been shared with the creep who often sits too close by her on the subway;[606]

598. *See*, *e.g.*, *id.*; *see also* Ryan Calo, *Privacy Harm Exceptionalism*, 12 COLO. TECH. L.J. 361, 361–63 (2014).

599. *Id.* at 363–64.

600. *Id.* at 364.

601. Citron & Solove, *supra* note 597, at 793.

602. *Id.*

603. *Id.* at 831.

604. *Id.* at 823.

605. Olga Kharif, *supra* note 452.

606. See, e.g., *Systems and Methods for Utilizing Wireless Communications to Suggest Connections for a User*, JUSTIA (Aug. 28, 2017), https://patents.justia.com/patent/10111059 [https://perma.cc/9U4W-MBQB]; U.S. Patent No. 2016/0014677 A1, at [16] (filed July 10, 2014) (issued Jan.14, 2016).

- The man whose phone's accelerometer reveals the stumbles of someone who has drunk too much—and who repeatedly receives ads for a sobriety organization on his work computer the next day;
- The couple who diligently shut off location information, but who receive ads indicating that their nighttime spent together is known, at least to some.

These potentially cause tangible economic harm. The first can give rise to a lender's discriminatory action, the second, to physical and psychological harms, the third, potentially to economic and reputational ones, the last, potentially to reputation, relationships, autonomy, and mental health.

In 2021, the FTC started to examine ISP practices in the use of metadata.[607] The agency was deeply troubled by the potential of harms from the consequent user tracking,

> [C]onsumers certainly expect ISPs to collect certain information about the websites they visit . . . [but] they would likely be surprised at the extent of data that is collected and combined for purposes unrelated to providing the service they request . . . More concerning, this data could be used in a way that's harmful to consumers, including by property managers, bail bondsmen, bounty hunters, or those who would use it for discriminatory purposes,[608]

We now propose a remedy to this situation.

B. Preserving Privacy Through Controlling Use

Solove noted that, despite people not being interested in micromanaging their privacy choices,[609] there were multiple cases where privacy self-management works, e.g., some users put great attention into their privacy settings on social media sites.[610] The problem, Solove argued, is not that privacy self-management cannot work, but that better tools are needed for it to do so. He proposed effective user control could be achieved through such tools as employing nudges in the direction of greater self-protection,[611] finding a more global way for users to manage their privacy

607. *See generally, e.g.*, FTC, *supra* note 60.

608. *Id.* at iv.

609. Solove, *supra* note 64, at 1901.

610. *Id.* at 1900.

611. *See* RICHARD K. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 6 (2008).

preferences and querying about use of data at time of use rather at time of collection.[612]

Whether privacy self-management can work is highly debatable; two decades of failures would indicate not. However, that is not the issue here, for it is far easier for a user to understand the consequences of sharing data they explicitly provide—input to a search engine or a mapping application—than data that they implicitly and, typically, unknowingly, supply, as is the case with communications metadata and software and device telemetry. Furthermore, collection might occur multiple times a minute and deciding which data can be shared and with whom would not be a simple issue for a user to determine. Privacy self-management of this data would provide only an illusion of user control. It would not provide the user with the ability to make meaningful, informed choices about the use of their data.

As Christopher Wolf put it in in 2014, we need to "focus on the misuse of data."[613] Aiming at the context in which the data is collected, Wolf explained that information "might be inappropriate to use in a different context" and that examining "how the information is normally used"[614] is important.[615] That is exactly the point about the use of metadata and telemetry.

Earlier, we detailed the uses to which AT&T put communications metadata. These started, of course, with connecting the call and billing. But measurement was also important, including volume of traffic served, volume of traffic denied, delays—e.g., how long it took to get a dial tone, and capacity. Some information was important as aggregated data, e.g., traffic volume served and denied. Yet, other data, such as connecting and billing, was done on a per-customer basis. As technology improved, the company's capabilities for measurement did, and it became easier to discover customer fraud.[616] This detection required understanding patterns of calls, including distinguishing patterns of legitimate callers from illegitimate ones. So here, too, AT&T was also studying the patterns of individuals.

The same model of repurposing holds true for current uses of communications metadata and device and software telemetry. While originally the purpose of communications metadata was

612. Solove, *supra* note 64, at 1901–02.

613. Christopher Wolf, *A Practicing Privacy Lawyer's Perspective on Use Analysis as a Way to Measure and Mitigate Harm*, 12 COLO. TECH. L.J. 353, 357 (2014).

614. *Id.* at 358.

615. This echoes the work of Helen Nissenbaum on contextual integrity. *See* Helen Nissenbaum, *Privacy as Contextual Integrity*, 79 WASH. L. REV. 119, 119 (2004).

616. *See* discussion *supra* Section I A.

delivery of content and smartphone sensors were developed for use *on* the device, both metadata and telemetry were found to be useful off the device for ensuring systems work correctly (e.g., for debugging). Communications metadata can be valuable for investigating fraudulent activities, while in a related application, sensors can provide added verification of user identity. Both metadata and telemetry may sometimes be used for modeling future customer use of services.

These uses should also look familiar, for they are how AT&T used metadata since the 1900s (with steady improvements as technology improved). The uses strike a reasonable balance between the business needs of a communications carrier and the privacy requirements of its customers. While debugging problems, preventing fraud, and anticipating future customer needs were not the purpose of smartphone sensors, using the data they provide to do so is very much in the user's interest.

In other words, using communications metadata and device and software telemetry for delivery and display of content, ensuring the communications network is working properly (e.g., using metadata and telemetry for debugging purposes), conducting fraud investigations and otherwise ensuring security, including device and user identification done for security purposes, and modeling to provision for future use of services are all within long-accepted purposes for an efficient functioning of a communications company.

The shift to mobile phones also created new capabilities, and these should be taken into account. There are two other ways in which the use of such metadata—and possibly telemetry—can be very much in the public interest. The first is to allow government use of aggregated tracking using communications metadata during public health emergencies. We discussed earlier how mobile phone data can track the aggregate movement of people during such situations, enabling aid to quickly be directed to where people are.[617] Thus, we recommend that in times of publicly declared public health emergencies, with such emergencies strictly bounded in time and scope, metadata and telemetry could be used for defined public safety measures.

We carefully scope this provision to public health emergencies and not "public emergencies." There are already multiple other ways to track mass movement of people, and the government already has the power to track an individual's location under a search warrant. Given the power of the technology to track populations, allowing the government to use such data for tracking

617. *See* discussion *supra* Section III.B.

poses an unnecessary and unacceptable risk for public protests and gatherings. Thus, this capability should be used sparingly and only during genuine public health emergencies. The time bounds should be strict. We propose under a week—and not renewable.

The second purpose we propose adding is for gaining understanding of the use of public services for urban and regional planning. Studying how people, in aggregate, use public transit, housing patterns, in which spaces people congregate at which times of the day, migration behavior,[618] etc., is extremely valuable for city and regional planning. We, thus, propose an exception to the use of metadata and telemetry: allow the use of aggregated communications metadata and device and software telemetry for public or peer-reviewed research projects.[619]

Because this type of information—e.g., real-time data on a Black Lives Matter march—can also reveal information useful for other purposes, its use should be strictly limited. The information should be only for public or peer-reviewed research projects in the public interest and should adhere to relevant laws and regulations governing such research.[620] The release of this data to researchers should be done in a way that prevents identification.[621]

Any regulation restricting use of data must take into account the 2011 Supreme Court decision in *Sorrell v. IMS Health, Inc.* The Court held that Vermont's Prescription Confidentiality Law, which prevented one set of speakers—"pharmacies, health insurers, and similar entities"—from using information about doctors' prescription habits for marketing purposes while allowing other speakers to use the data for other purposes, was unconstitutional.[622] In *Sorrell*, the state restricted a single speech use based on the speaker, with the express goal of limiting the speaker's message.[623] By contrast, the prohibition we propose is against all but seven permitted uses and is not viewpoint dependent.

As for tailoring, the proposal to limit use of individuals' metadata and telemetry splits naturally into two categories. The

618. *See, e.g.*, Shengjie Lai et al., *Exploring the Use of Mobile Phone Data for National Migration Statistics*, PALGRAVE COMMC'NS, Mar. 26, 2019, at 1, 1.

619. This exception is based on subsection 101(b)(10)(A) of H.R. 8152.

620. Thus, for example, aggregated data on a Black Lives Matter march should not be available to the Proud Boys.

621. The U.S. Census has considered this issue at length. *See, e.g.*, DANAH BOYD, DATA & SOC., BALANCING DATA UTILITY AND CONFIDENTIALITY IN THE 2020 U.S. CENSUS 6 (2020) https://datasociety.net/wp-content/uploads/2019/12/Differential-Privacy-04_27_20.pdf [https://perma.cc/2BZE-HVBK].

622. Sorrell v. IMS Health, Inc., 564 U.S. 552, 559, 580.

623. *Id.* at 580.

first five purposes limit service providers' use of metadata and telemetry to those functions necessary for ISPs and app providers to deliver user services, including doing future business planning. As already noted, these functions mimic the purposes to which telecommunications providers used communications metadata. The latter two purposes serve the public interest: the first in an emergency public-health situation, the second to enable better planning of services. Both are carefully written to fit the state's policy needs: the improvement of the public's health and welfare in a way that nonetheless protects the privacy interests of private parties. They serve a compelling public interest.

Thus, we make the following recommendation: ISPs, OS providers, platforms, and apps should not be allowed to use communications metadata and software and device telemetry except for the following purposes:

1. delivery and display of content;
2. ensuring the system is working properly (e.g., for debugging purposes);
3. investigating fraud;
4. ensuring security, including device and user identification done for security purposes;
5. modeling to provision for future use of services;
6. during publicly declared public health emergencies, providing information on the movement of people in aggregate; this latter use for a maximum of one week; or
7. providing information in aggregate to a public or peer-reviewed research project that (i) is in the public interest; and (ii) adheres to all relevant laws and regulations governing such research including identification methods.[624]

The first four of these cover the activity-based needs of a common carrier. It is clear how communications metadata is important for one through four. As is the case with communications metadata, software telemetry helps assure correctness and prevent fraud; it is not clear, however, that software telemetry has particular value in projecting future services. Device telemetry can be similarly useful in determining whether the device is working properly, and, due to its application to authentication, is valuable in fraud prevention; its role in projecting future services is unclear. The fourth is important in attack-prone environments such as the Internet; this exception limits use *to solely for security purposes*. The fifth permitted use, "modeling to provision for future use of

624. This list of recommendations is based on subsection 101(b)(10)(A) of H.R. 8152.

services," has some wiggle room. Some companies might try to push this item to gain more information about current customers, but the intent of this item is to model growth of the current service. Any deviation from that narrow purpose would be considered deceptive use.

These five purposes provide clear and direct user benefits. The sixth exception satisfies a public-health objective in addition to providing benefits to affected individuals. The seventh exception is of a different nature; it is focused on providing a societal benefit rather than benefits directly to the individuals. The exception is designed to satisfy community social interests, while preventing a misuse of real-time access of the data. In prohibiting other uses, the requirements eliminate an aspect of discriminatory actions taken as a result of information gleaned from metadata and telemetry.

These proposed permitted uses build off of Cate's approach for handling the failure of notice-and-choice by more carefully controlling use.[625] Cate's proposal was a Consumer Privacy Protection Principle focusing on use: "Data protections laws should target harmful uses of information, rather than mere possession."[626] He proposed that the government not regulate for uses that "present no reasonable likelihood of harm," but prohibit use in cases where it was always harmful.[627] For the middle ground—use neither "per se harmful" nor "per se not harmful"—Cate recommended the government let the user decide in cases where consent "likely would be effective."[628] Since such consent is highly unlikely to be effective in determining whether to allow other uses of metadata or telemetry, we depart from Cate's approach. Thus, we take the approach that those uses of metadata and telemetry that are not explicitly permitted are prohibited.

Note that our proposal does not prevent the online ad industry from collecting information about consumers; instead, it prevents the online ad industry from collecting information about consumers that they are unaware of providing. Information that consumers willingly and consciously provide is unaffected by our recommendations.

This is a strong response with a slightly paternalistic air to it. Yet, such controls are not out of line with policies in the modern world. In the U.S.—and much of the world—regulations require that cars have passenger seatbelts. In a majority of U.S. states, the sale of unpasteurized milk is prohibited (with an exception, in some

625. Cate, *supra* note 65, at 369–71.
626. *Id.* at 370.
627. *Id.*
628. *Id.*

states, permitting sale of "raw milk" at farms).[629] Neither car sales nor milk production has ground to a halt as a result of the government instituting such requirements.

C. Two Ways Forward

There are two routes forward on controlling the use of metadata and telemetry by the private sector: through the FTC and through Congress. We examine both of these.

Recall that when in 2021 the FTC looked at ISP use of metadata, the agency expressed concern over the companies' use of the data the agency stated, "consumers certainly expect ISPs to collect certain information about the websites they visit . . . [but] they would likely be surprised at the extent of data that is collected and combined for purposes unrelated to providing the service they request."[630] Indeed, the agency headlined one section of its report, "Several ISPs in Our Study Gather and Use Data In Ways Consumers Do Not Expect and Could Cause Them Harm."[631]

The FTC's ability to bring a privacy case against ISPs—or OSes, platforms, or apps—over their use of metadata and/or telemetry is limited to considering only non-activity-based actions of common carriers and only situations in which deceptive or unfair actions are present. Permitting use of metadata and telemetry for delivery and display of content, debugging systems, uncovering fraudulent behavior, and planning future services, separates out carrier activity-based usage from status-based and thus fits within the FTC responsibilities. The framework, however, does not loosen the law's grip on limiting FTC privacy investigation to issues of deceptive or unfair practices.

Solove and Hartzog observed that with *In re Sears Holdings Management Corp.*,[632] the FTC shifted its focus from comparing a company's actions to its stated privacy policy and squinted, taking into account customer experience; the authors described a shift from "broken promises" on privacy to "broken expectations of consumer privacy."[633] The question then becomes not, what did the company promise the user (in its remarkably long and difficult-to-

629. *Legal Status of the Sale of Raw Milk and Outbreaks Linked to Raw Milk, by State*, *2013–2018*, CTRS. FOR DISEASE CONTROL & PREVENTION, https://www.cdc.gov/foodsafety/rawmilk/nonpasteurized-outbreaks-maps.htmlhttps://www.cdc.gov/foodsafety/rawmilk/nonpasteurized-outbreaks-maps.html [https://perma.cc/manage/create?folder=22100-177223-187848].

630. FTC, *supra* note 60, at iv (emphasis added).

631. *Id.* at 34.

632. Solove & Hartzog, *supra* note 63, at 624; *see generally* Complaint, *In re* Sears Holding Mgmt. Corp., FTC File No. 082 3099, No. C-4264 (F.T.C. Aug. 31, 2009).

633. Solove & Hartzog, *supra* note 63, at 667–69.

read privacy policy), but what the consumer reasonably expects from the policy? They, then, trace several other cases—including *HTC America Inc.*,[634] *In the Matter of Facebook, Inc.*,[635] *United States of America v. Google Inc.*,[636] and *United States of America v. Path, Inc.*[637]—that show increasing interest by the FTC in pursuing case based on such broken expectations of privacy.

In August 2022, the FTC initiated "Advanced Notice of Proposed Rulemaking," seeking input on implementing U.S. Government regulations regarding the ways in which companies collect, aggregate, protect, use, analyze, and retain consumer data, as well as transfer, share, sell, or otherwise monetize that data in ways that are unfair or deceptive.[638] Such rules would allow the FTC to move from a case-by-case direction to implementing rules with much effect. The announcement was strongly lauded by FTC Commissioner Slaughter.[639]

Days later, the FTC announced a complaint against data broker Kochava, whose rules for access to consumer "precision location" data lacked important controls.[640] Kochava supplied free samples of data to interested purchasers; this was no small set, but data from 61 million devices.[641] This data set provided the ability to precisely track vulnerable consumers, e.g., someone under threat of domestic violence, or to look at an individual's past history (had they been homeless?)[642] The FTC noted that, "Consumers have no insight into how this data is used—they do not, for example, typically know or understand that the information collected about them can be used to track and map their past movements and that inferences about them and their behaviors will be drawn from this

634. *Id.* at 670.

635. *Id.* at 671; *see generally* Complaint, *In re* Facebook, Inc., FTC File No. 092 3184, No. C-4365 (F.T.C. July 27, 2012).

636. Solove & Hartzog, *supra* note 63, at 670; *see generally* U.S. v. Google, Inc., No. CV 12-04177 (N.D. Cal Nov. 20, 2012).

637. Solove & Hartzog, *supra* note 63, at 670; *see generally* U.S. v. Path, Inc., No. 13-cv-00488 (N.D. Cal. Feb. 8, 2013).

638. Press Release, FTC, FTC Explores Rules Cracking Down on Commercial Surveillance and Lax Data Security Practices (Aug. 11, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/08/ftc-explores-rules-cracking-down-commercial-surveillance-lax-data-security-practices [https://perma.cc/2XWG-6942].

639. *Slaughter ANPRM Statement*, *supra* note 502.

640. Complaint ¶ 7, FTC v. Kochava, Inc., Case No. 2:22-cv-00377-DCN (Aug. 29, 2022).

641. *Id.* at ¶ 14.

642. *Id.* at ¶ 20.

information."[643] Thus, "[c]onsumers are therefore unable to take reasonable steps to avoid … injuries."[644]

Exactly. That is the argument we have been making regarding industry's use of communications metadata and telemetry information. If the FTC successfully takes a "broken expectations" approach to the use of communications metadata and telemetry information, that would increase the agency's ability to stop inappropriate uses except for the purposes described above.

Federal Trade Commissioners come and go, and while for stability, policies should remain, that has not always been the case. Thus, a more ambitious and more secure way to instantiate these privacy protections would be legislatively.[645] A narrow law pertaining simply to metadata and telemetry is a patch when the underlying privacy problem is far broader than the singular set of issues on which this paper has focused. Our recommendations should find their way as part of a broader bill on privacy.

A natural vehicle is the 2022 Congressional bill, American Data Privacy and Protection Act,[646] which already includes some aspects of what we propose here. This bill takes the approach of limiting use to specific, delineated ones.[647] Although the bill's list fails to include an emergency public health use, this could easily be added. Other aspects of the bill need greater expansion. While the legislation provides some protections for communications metadata, these are inadequate. The bill deems telephone metadata—"telephone numbers called, telephone numbers from which calls were placed, the time calls were made, call duration, and location information of the parties to the call"[648]—"sensitive covered data" that would require a user's active, affirmative consent before transferring data to a third party. This is insufficient. Communications metadata should include metadata for other forms of electronic communications, including email, texts, etc.

A more serious problem with the bill is that it permits transferring data to third parties with "the affirmative express consent of the individual."[649] As we have explored in some depth in this paper, users are not in position to effectively provide informed

643. *Id.* ¶ 31.

644. *Id.*

645. FTC Commissioner Slaughter has made this point as well. *See Slaughter ANPRM Statement*, *supra* note 502.

646. *See generally* H.R. 8152.

647. H.R. 8152 § 101(b).

648. H.R. 8152 § 2(28)(A)(vii).

649. H.R. 8152 § 102(3)(A).

consent for uses of metadata and telemetry. The bill seeks to remedy this concern by delineating categories of "sensitive covered data"[650] and providing the FTC with rulemaking ability to extend the definition of sensitive covered data to other categories as needed.[651] Each person's sense of what constitutes sensitive data is different, however. One person may seek to keep private their early work history, another, their sexual identity, a third, a family member's bankruptcy. By positively listing what constitutes "sensitive covered data," rather than listing appropriate uses for metadata and telemetry and prohibiting all others, the bill leaves open the likelihood that use of communications metadata by ISPs, OS providers, platforms, and apps will continue to infringe on users' privacy. The "fix" of allowing the FTC additional rulemaking ability does not alleviate the problem.

Finally, the bill fails to address telemetry. As we have shown, protections against misuse of software and device telemetry are essential.

While we would like to see the fundamental idea we espouse here—control use not collection—applied to all manners and forms of data collection, to be realistic, we are making the recommendation only for the use of metadata and telemetry, where the allowed purposes for use of the information can be simply and easily described. More work needs to be done before this idea can be extended further. As Julie Cohen has observed, lasting behavioral change does not occur without specific mandates.[652] We begin here.

To enable effective implementation, an issue also raised by Cohen,[653] we also recommend the implementation of two other of Cate's proposed principles: 4. Transparency, Honesty, and Accountability and 8. Effective and Efficient Enforcement.[654] The latter is particularly important given the entanglement of FCC and FTC responsibilities regarding information services. While the two agencies have thoughtfully worked within current legal frameworks to protect consumers and the public interest, a disentangling that not only gives more authorities to the FTC (viz. the Prevention of Harm Principle) but also simplifies implementation, which would be a real benefit.

Our proposal is bold. However, nothing less than this will protect the public from the increasing encroachment of tech

650. H.R. 8152 § 2(28)(A).
651. H.R. 8152 § 2(28)(B).
652. COHEN, *supra* note 496, at 17.
653. *Id.* at 14.
654. Cate, *supra* note 65, at 372–73.

companies on the public's privacy.[655] Nor is the proposal necessarily harmful to U.S. tech company interests; increasing interest by the European Union on related legislation may force some of these changes on the companies, at least in their operations abroad.[656] In that sense, U.S. action in this direction might prove to be a real benefit for these companies, pushing them to act sooner rather than later and making them more competitive, not less.

In any case, the fundamental unfairness and discriminatory implications stemming from the collection and use of data that users have no capability of withholding, are largely unaware of providing, and do not control via a notice-and-choice regime mean that the problem must be addressed. Proposals weaker than the ones put forth in this section are unlikely to be effective in restoring a modicum of privacy to users. Our proposal is not a radical solution; it is simply a necessary one to provide a rebalancing of privacy rights and consequent social goods including fairness.

CONCLUSION

The invasive use of communications metadata by the private sector began roughly from two decades ago, when the ISPs saw a market opportunity.[657] Since then, ISPs, Internet companies, and app providers have reached a situation where they can completely surveil users not only when users are online, but increasingly during offline times as well.[658] The arrival of Augmented Reality (AR) and Virtual Reality (VR) applications, with their need for rich sensor data about the physical environment, threatens to take us further along the path to total user surveillance by the private sector.[659] Although in *Carpenter*, the Supreme Court stepped in to prevent some of the most invasive uses of such technologies by the

655. *See* COHEN, *supra* note 496, at 19 (describing specifics of serious fines, etc., as a way to effect genuine change).

656. *See Commission Proposal for a Regulation of the European Parliament and of the Council Concerning the Respect for Private Life and the Protection Personal Data and Electronic Communications and Repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)*, at 7, COM (2017) 10 final (Oct. 1, 2017).

657. Ohm, *supra* note 56, at 1424.

658. Thus we have seen, for example, how ISPs, the OS, and sometimes apps can learn which participants in a public protest are from the same household, workplace, are otherwise socially connected. *See* Barbera et al., *supra* note 383; what someone is typing into an online medical form even if they do not submit it; *see* Sankin & Mattu, *supra* note 38; or that two people have spent the night in the same hotel room; this last is possible even if the users have shut off location information; *see supra* Section IV.C.

659. Phone sensors can, for example, recognize agitation onset in people with dementia; *see*, *e.g.*, Christianne Fowler et al., Detecting agitation onset in individuals with dementia using smart phone sensors (HEALTHINFO Conference, Oct. 2017).

government,[660] there is no equivalent Fourth Amendment protection against the surveillance threat posed by the private sector.

Industry has adopted a "notice-and-choice" regimen that has repeatedly been proved to be ineffective for providing users tools to control the dissemination of their private information or protect their privacy. These studies dealt with information users were conscious of providing, such as terms to a search engine, destinations to a mapping location, or photos to a social media site.[661] Yet, the private sector increasingly uses information that users are unaware of providing: communications metadata and software and device telemetry.

In some cases, including delivery or display of content or for certain applications to interact with full user input, such data is necessary for providing user services. In other cases, such as when metadata or telemetry is used to ensure systems are working correctly or for preventing fraud, use of the data in these ways provides direct benefit to the user. By enabling businesses to better predict future customer needs, metadata and telemetry provides indirect, but quite real, value to users. Such uses, much the same as those AT&T employed during its regulated monopoly period,[662] are reasonable and appropriate. They are also not privacy invasive. The problem is that not all uses of metadata and telemetry are similarly benign.

We have detailed multiple examples of patents proposing use of metadata and telemetry for tracking users in various different ways (through common SSiDs, accelerometer, gyroscope, magnetometer data, etc.)[663] to determine which users are in close proximity with each other (through Wi-Fi signals, radio signal modulation, or geo-tagging) and to learn other personal traits and behaviors about users.[664] There are numerous ways in which Internet companies, mobile operating systems, and phone apps can get to "know their customer" using data supplied by users—who frequently have no control over the provision of the data (and thus no ability to control its use). The information revealed can be highly personal[665] or about protected traits, such as religion, sexual

660. Bellovin et al. have addressed some aspects of government collection of metadata; *see generally* Bellovin et al., *supra* note 50.

661. *See* discussion of consumers' expectations *supra* Section III.

662. *See e.g.*, *supra* Section II.A.

663. *See* discussion *supra* Section III.

664. *See* discussion *supra* Section III.

665. *See* discussion *supra* Section IV.C.

orientation,[666] etc. Some could be used in ways that harm the user, e.g., for discriminatory purposes.[667]

Yet consumers have little ability to control either the collection or use of metadata and telemetry. For one thing, users must share information to receive the service; this is the case for communications metadata and is also sometimes true for telemetry information (and will be increasingly the case in the future as AR/VR becomes more popular). Furthermore, because queries about collection and use would involve querying users about minute pieces of data many times a minute, a notice-and-choice regimen is simply implausible. Its use would overwhelm the user, who would quickly click "yes" to all questions in order to use the relevant app.

The cost to users in human dignity, and sometimes in consumers' willingness to use services, is high as a result of this loss of privacy; the cost to society in terms of the inequities this situation creates, may be even higher. The resulting situation endangers the fabric of society itself, for the surveillance capabilities it provides create an imbalance of power that can disenfranchise many, especially the least powerful in society. Resolution of this societally unhealthy situation lies in returning to "meaningful privacy choices" as put forth by Westin.[668] Thus, we recommended adoption of a carefully crafted purpose-limitation principle on the use of metadata and telemetry.

We have proposed a rather strong modification of Cate's proposed Consumer Privacy Protection Principle focusing on use.[669] Because ascertaining harms in telemetry and metadata use cannot be easily determined, we propose that except for expected uses and non-re-identifiable societally beneficial ones, no other uses of this non-content data be permitted. Thus, we recommend permitting the use of communications metadata and software and device telemetry only for the following purposes: (i) delivery and display of content; (ii) ensuring system functionality, including use for debugging; (iii) investigating fraud; (iv) providing security; (v) modeling future services; (vi) only during publicly declared public health emergencies, providing information on the movement of people in aggregate; and (vii) conducting a public or peer-reviewed

666. *See generally, e.g.*, Jernigan & Mistree, *supra* note 451.
667. FTC, *supra* note 60, at 33–44.
668. WESTIN, *supra* note 61, at 66.
669. Cate, *supra* note 65, at 370.

research project, the latter two with requirements that data be protected so as not to be re-identifiable.[670]

Such a policy could be instituted by the FTC using its model of "broken expectations of consumer privacy"[671]—for the current situation on metadata and telemetry use is nothing if not a broken expectation. In addition, reform could be pursued by Congress, which would provide a more permanent protection of privacy rights.

Communications metadata and software and device telemetry are examples of frequent and minute data collection that reveal vast amounts of personal information about consumers and over which the users have effectively no control. This data use represents an instance in which the notice-and-choice regimen will almost certainly fail for the reasons described in this paper. Getting the controls on use right in this case is important not only in and of itself, but also for its implications for other forms of data whose uses outside their intended purpose could cause great harms to individuals and society. We recommend immediate action to strongly limit the uses of metadata and telemetry solely to the purposes we described in this paper. Furthermore, we recommend that the model we provided, which proposes highly constrained use limitations, be applied to types of information over which consumers have effectively no choice of supplying and limited-to-no understanding of the uses to which that information can be put. Only then can users hope to have the privacy that the Fair Information Practice Principles sought to provide.

670. As noted in Section V.B, *supra*, the public-health exemption would be for only a very limited time period; the research exemption is for work in the public interest and with data usage subject to all regulations and laws.

671. Solove & Hartzog, *supra* note 63, at 667–69.