The Struggle over Data Collection from IoT Devices in American Households

A Sociotechnical Research Paper
presented to the faculty of the
School of Engineering and Applied Science
University of Virginia

by

Eric Sakmyster

April 5, 2021

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Eric Sakmyster

Sociotechnical advisor: Peter Norton, Department of Engineering and Society

The Struggle over Data Collection from IoT Devices in American Households

Connected electronic devices have been proliferating in American homes. The Internet of Things (IoT) is the network of devices that supports device-to-human and device-to-device communication. IoT has diversified from computers and smartphones to kitchen appliances, thermostats and many other devices. Through IoT, homeowners can monitor home systems and improve their cost efficiency, but IoT can also expose homeowners to hackers. Netscout (2018) found that on average, an IoT device is attacked five minutes after connection to the internet; in 24 hours it is targeted by specific exploits. Many small IoT-capable devices are vulnerable because they have no cryptographic capacity (Mckay et al., 2017). Such vulnerabilities can expose all other connected devices in a network to attack (Rehman et al., 2016). This makes small IoT devices attractive hacking targets.

Tech companies can collect sensitive data from IoT systems, compromising residents' privacy. While consumers are typically alert to privacy threats from devices that record sound or video, few recognize other vulnerabilities (Zheng et al., 2018). According to Day et al. (2019), Amazon employees have listened to stored user recordings to improve the artificial intelligence of Alexa, its smart speaker. With machine learning, tech companies can gather sensitive information from homes even from simple IoT devices (Zheng et al., 2018). User data is the ultimate good that tech companies want because it gets them revenue and our preferences give them an edge over other tech companies (Verzun, 2020). Such data collection can go unnoticed, and terms of service may authorize it. Tech companies producing IoT devices for households have succeeded in retaining almost unlimited access to user data, despite the fact that everyone else agrees that user data should be far more restricted. They have done this by reframing the data collection argument to be about improving security to protect from hackers breaching user

privacy, being deceptive and not transparent about data being collected, and taking advantage of users who lack knowledge of their devices.

Review of Research

IoT device makers need user data in order for their systems to work properly and increase the user experience. According to Tschider (2018), "data analyses conducted on big data stores improve functionality of IoT devices and identify device upgrades or changes needed for more efficient or effective devices." Timer series analytics and prescriptive analysis are useful forms of data collection that help household IoT devices draw trends and make better conclusions (Joseph, 2018). Zheng et al. (2018) conducted a survey of home IoT users, with ten out of eleven of them saying, "they did not mind the manufacturers of their devices collecting and analyzing data, acknowledging that it is necessary to improve the product and user experiences." However, it is noted that the participants' responses were based on the belief that manufacturers would be responsible with their data (Zheng et al., 2018). While data collection is often framed in a bad context by the media, it is necessary for tech companies to do, with a lot of trust being placed on them by consumers.

Collecting large amounts of data provides tech companies with better chances of revenue. The selling of the combination of IoT device data with other forms of data "such as: buying habits, web browsing history, demographic data, or other codified behaviors" can be "highly lucrative both for manufacturers selling data and for organizations purchasing or using data for targeted marketing activities" (Tschider, 2018). The good that is provided to tech companies from user data must be recognized to understand their motives for wanting increased data collection.

While consumers expect tech companies to produce secure IoT devices, research shows laws and production methods promote less security. Agile software development methods, which can encourage quick production of software and reuse of existing code, can lead to security being left to the latter stages of development, or not even being integrated until release (Sadler, 2017). The Digital Millennium Copyright Act has also been a way for security measures by tech companies to go unchecked (Schneier, 2017). When IoT makers push for greater security to protect data, research often shows little change is made when security is not a priority.

Examining existing privacy laws shows a lack of needed detail about data collection methods. California's Online Privacy Protection Act of 2003 provides some protection to IoT consumers by requiring that any company who "collects personally identifiable information through the Internet about individual consumers residing in California who use or visit its commercial Web site or online service shall conspicuously post its privacy policy on its Web site." This law gives tremendous power to IoT makers because it doesn't encapsulate all definitions of "personally identifiable information," and it assumes only personal data warrants privacy concerns. Research also shows that tech companies try to obtain user data using convoluted privacy policies. According to Peppet (2014), these policies "are often quite unclear about whether collected sensor data count as 'personal information'—and therefore ambiguous as to what rights and obligations apply to such data." Lack of consumer knowledge of IoT devices can give an edge to tech companies. Elvy (2018) argues, "the level of consumers' awareness and understanding of data-trade agreements may to some extent be context dependent…a consumer that uses an IoT refridgerator or toaster may not be aware of the types of, and extent to which, data are being collected and what can subsequently be done with such data."

Research into discriminatory practices by IoT device makers is necessary to show unfairness and excessiveness in data collection. According to Wachter (2018a), “There are at least three possible ways of monitoring and profiling that offer grounds for discrimination in IoT systems: (a) data collection that leads to inferences about the person (e.g. Internet browsing behaviour); (b) profiling at large through linking IoT datasets (sometimes called ‘sensor fusion’); and (c) profiling that occurs when data is shared with third parties that combine data with other datasets (e.g. employers, insurers).” Discrimination doesn’t only come from personal data, it can also come from neutral data (Wachter, 2018a). While legal issues arise from this treatment of users, it is important to recognize tech companies continue to do this because more data can be gathered as a result.

With the inclusion of trivial devices in IoT, researchers have found consent from users is almost negligible. Peppet (2014) states, “Internet of Things devices are often small, screenless, and lacking an input mechanism such as a keyboard or touch screen” and “at the moment, Internet of Things manufacturers overwhelmingly seem to prefer to only provide privacy- and data-related information in website privacy policies.” Barocas and Nissenbaum (2014) claim consent is necessary, but warns that big data can make inferences about people even if they don’t give consent. Tech companies intentionally make their privacy policies unreasonably long and full of technical terms to make the consumer give consent without understanding what they are agreeing to (Kim & Telman, 2015). Furthermore, Kim and Telman (2015) point out that courts consistently side with tech companies that a click agreement can be considered consent, despite consumers saying in testimony that they don’t remember clicking any button for consent. Consumers can also get “security fatigue” by trying to keep track of security and privacy policies

that all devices in their home IoT system have, potentially leading to thoughtless consent (Stanton et al., 2016). Tschider (2018) suggests most companies have no real consent policy.

Shifting Attention to Security

Tech companies can manipulate users to think their data is safe by claiming their products are secure. In an experiment conducted by Johnson et al. (2020), consumers were much more willing to purchase a device that had a label that it was secure, as opposed to having no label. However, Schneier (2017) states, "According to the [Digital Millennium Copyright Act], it is a crime to bypass security mechanisms that protect copyrighted work, even if that bypassing would otherwise be legal." Because software can be copyrighted, "Vendors of [IoT] devices are more likely to leave them insecure, because no one will notice and they won't be penalized in the market" (Schneier, 2017). This creates a dilemma that tech companies can claim they are working on security upgrades, but nobody outside of the company can test their software. Thus, a distraction of security is always present for tech companies so that consumers don't focus on the data collection tech companies are doing themselves.

Problems with software can be used as an excuse by tech companies to get away with data collection. Following a complaint by the FTC, it was revealed that children using Amazon Echo Dot Kids edition devices "can use an Alexa feature called 'Remember This,' which keeps anything a child says until parents call Amazon customer service to delete the entire profile" (Tung, 2019). Amazon responded to this issue by saying it was just a "bug" in the software written (John, 2019). The problem with this response is that Amazon seems to have a trend of keeping data until a user says they shouldn't. In a 2019 letter written to Senator Christopher Coons answering questions about privacy and data security practices with Alexa, Amazon stated, "We retain customers' voice recordings and transcripts until the consumer chooses to delete

them" (Ng, 2019). This letter came after calls by senators and privacy supporters expressing concern that even if consumers deleted voice data records, their text transcripts were still available on the Amazon cloud, which Amazon says they are working on fixing (Ng, 2019). When its not the consumer to blame, Amazon is quick to say the problem is a software bug, which the public generally accepts.

To promote integrity in data collection, alongside added security, privacy advocates have started to encourage the adoption of blockchain in IoT systems. The tamper-resistant ledger that blockchain has offers a transparent and unchangeable way to ensure that data transactions are not interfered with by hackers or tech companies (Bandara et al., 2021). If implemented correctly, blockchain-based IoT systems become significantly decentralized (Ferreira et al., 2021). This greatly reduces the risk that a hacker can have access to all devices on the network and provides less administrative power to tech companies (Ferreira et al., 2021). Overall, blockchain-based IoT would allow consumers and regulatory agencies to notice bad data practices by tech companies. Because of this, blockchain has been slow to be integrated into IoT because it would significantly reduce tech companies' data collection authority.

Ethical Issues in Home IoT Data Collection

Permissible use of home IoT data must be regulated. According to Wachter (2018b), "linkage of existing but previously unconnected datasets, can offer new opportunities for data analytics that were not envisioned when the data were collected." However, "even if organizations pursue robust anonymization or de-identification programs, the more data collected, the more likely individuals could be re-identified and have their private, personal information exposed (Tschider, 2018). Current regulation, such as California's Online Privacy Protection Act of 2003, fails to provide protection for combining of data and advanced

algorithms. If regulation is not actively being drafted as new forms of data collection are being used by tech companies, then IoT makers will continue to have leverage on this issue.

Tech companies can collect data without explicitly informing the user. A survey conducted by Peppet (2014) on "twenty popular Internet of Things consumer devices, including Fitbit and Nike+ Fuelband fitness trackers, the Nest Thermostat, the Breathometer, and others" resulted in finding no information regarding privacy or data collection in any of their containers. While this information could be found on the website of these devices, no mention of this could be found on any of the manuals in the packaging (Peppet, 2014). To avert such deception, Hirsch (2019) recommends that at system installation, policy should require "an in-home visit not from an installer but a well-trained data privacy specialist who can take the home owner through all of the data usage options in real time face to face," or that company websites have "privacy bots that explain all of the options and risks." Privacy advocates believe if data collection is going to occur, it must be presented to the user in some fashion.

Personal data is the most sensitive data, but it can also be the most valuable. Tech companies may therefore mislead users to acquire rights to it. California's Online Privacy Protection Act of 2003, known for being one of the strictest personal privacy laws in the U.S., classifies "personally identifiable information" as:

> (1) A first and last name; (2) A home or other physical address, including street name and name of a city or town; (3) An e-mail address; (4) A telephone number; (5) A social security number; (6) Any other identifier that permits the physical or online contacting of a specific individual; (7) Information concerning a user that the Web site or online service collects online from the user and maintains in personally identifiable form in combination with an identifier.

This definition does not include data that is not "stored" in personal identifiable form, or can be combined with other data to profile someone. This gives tech companies power to have personal data, but not give notice of its collection. Peppet (2014) states, "most privacy policies permit manufacturers to share or sell non-personal information far more broadly than personal information." This leads tech companies to either not define whether certain data is personally identifiable, or have policy that "creates conflict between its definition of 'personal information' and 'non-personal information'" (Peppet, 2014). In the U.S., privacy advocates argue terms of use should emulate those of Europe's General Data Protection Regulation: "If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language" (Regulation 2016/679). Since most customer data agreements merge all data usages into one, tech companies should ask for consent each time personal data will be used. However, this cannot be done effectively until proper definition of personal data is put into law.

Discriminatory algorithms and data collection techniques have been used by tech companies with little restriction. According to Rose et al. (2015), "Big data algorithms can examine massive quantities of IoT data and look for statistical and semantic correlations to determine groupings or clusters of related characteristics among users." This categorization of data can be used in discriminatory ways by offering certain product options or recommendations based on the user's race, ethnicity, or cultural habits. Privacy advocates argue that the level of detail in data collection that allows for discrimination needs to be prohibited. The Internet Society (2019) proposes that legislation needs to be passed that doesn't allow any personal data to be used in calculations that could be beneficial or not against a user, such as "preventing

insurance companies from using IoT-derived data as a factor in insurance rates, unless explicit, informed consent has been freely given." However, distinguishing whether data obtained by tech companies is discriminatory is shown to be a "non-trivial task" and approaches that exclude discriminatory characteristics fail because of other background data present (Pedreschi et al. 2008). Nonetheless, privacy advocates at least want some form of recognition from US privacy agencies that discrimination is occurring, similar to the GDPR in the EU anticipating "direct, economic, and moral types of discrimination" in IoT devices (Tschider, 2018).

Lack of User Knowledge

For trivial IoT devices, tech companies can take advantage of consumers who wouldn't be aware data can even be collected. Elvy (2018) provides one example of this by stating, "the Roomba robotic vacuum not only self-cleans a consumer's home, but also collects 'home layout data' including 'the location of everything from [walls and] lamps to home security cameras and thermostats.'" Roomba says that they may sell your data to other companies (Elvy, 2018). With everyday household appliances that users expect to have one function, the lack of informing the consumer allows deceptive data collection by tech companies.

Data collection consent isn't an obstacle for tech companies to collect user data. Most IoT makers follow a "quasi-contract" where consumers implicitly give consent to privacy policies by choosing to purchase devices, but the question is whether this choice is actually binding (Tschider, 2018). Even if a legitimate option to consent is given to a consumer, Barocas and Nissenbaum (2014) note, "the value of a particular individual's withheld consent diminishes the more effectively a company can draw inferences from the set of people that do consent as it approaches a representative sample". When this occurs, "the company can rely on readily observable data to draw probabilistic inferences about an individual, rather than seeking consent

to obtain these details" (Barocas & Nissenbaum, 2014). Tech companies frequently rely on arbitration clauses to silence consumers who have had problems with data collection (Bannan, 2016). While arbitration clauses are necessary for large companies who can't settle the amount of claims they receive in a normal court setting, it allows IoT makers to continue with bad data practices without informing consumers.

Conclusion

Consent is a significant topic in the conversation about internet privacy. Additional research into the psychology of consumer consent with various contexts would be beneficial. Understanding that unrestricted data collection is a matter of consumer error, as well as tech company manipulation, can aid in further research into consent. As a society, America fails to assign enough importance to consent. Making consent a priority in data privacy may need to be spurred on by advocates of consent on other issues.

Regulation in IoT data protection is lacking. Federal and local legislation has tried to avoid addressing topics involved with internet privacy. In an age of exponential increases over time of machine computing power, data protection becomes increasingly more important. The FTC has been failing at protecting often sensitive and impactful data from both tech companies and hackers. Research into why these deficiencies exist are necessary for understanding a lack in internet privacy policy.

Even with new IoT data regulation, researchers must continue to investigate whether any impact has been made. Tech companies are always a step ahead of regulations with new AI algorithms and using loopholes present in new laws. Until personally identifiable data and other privacy related terms get defined precisely by a central figure, with all tech companies abiding by those definitions, IoT makers will continue to vaguely define their data collection methods.

Recognition that IoT implementation is still in its early stages is important for researchers to identify what small scale problems exist before IoT expands to systems such as smart cities.

References

Bandara, E., Tosh, D., Foytik, P., Shetty, S., Ranasinghe, N., & De Zoysa, K. (2021). Tikiri - Towards a lightweight blockchain for IoT. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2021.02.006

Bannan, C. (2016, August 14). The IoT threat to privacy. *TechCrunch*. https://techcrunch.com/2016/08/14/the-iot-threat-to-privacy/#:~:text=The%20most%20dangerous%20part%20of,how%20it%20is%20being%20used.&text=As%20a%20result%2C%20the%20privacy,left%20without%20any%20real%20remedy.

Barocas, S., Nissenbaum, H. (November 2014). Big Data's End Run Around Procedural Privacy Protections. *Communications of the ACM*. *57*(11), 31-33. https://cacm.acm.org/magazines/2014/11/179832-big-datas-end-run-around-procedural-privacy-protections/fulltext

Day, M., Turner, G., & Drozdiak, N. (2019, April 24). *Amazon's Alexa team can access user's home addresses*. Bloomberg. https://www.bloomberg.com/news/articles/2019-04-24/amazon-s-alexa-reviewers-can-access-customers-home-addresses

Elvy, S.-A. (2018). Commodifying Consumer Data in the Era of the Internet of Things. *Boston College Law Review*, *59*(2), 424–522.

Ferreira, C. M. S., Garrocho, C. T. B., Oliveira, R. A. R., Silva, J. S., Cavalcanti, C. F. M. da C., & Qiu, M. (2021, February 15). IoT Registration and Authentication in Smart City Applications with Blockchain. *Sensors (14248220)*, *21*(4), 1323. https://doi.org/10.3390/s21041323

Hirsch, P. B. (2019, January 14). The goose that laid the golden eggs: personal data and the Internet of Things. *Journal of Business Strategy*, *40*(1), 48–52. https://doi.org/10.1108/JBS-10-2018-0176

Internet Society. (2019, September 19). Policy Brief: IoT Privacy for Policymakers. https://www.internetsociety.org/policybriefs/iot-privacy-for-policymakers/

John, A. S. (2019, May 9). Amazon Echo Dot Kids Violates Privacy Rules, Advocates Claim. *Consumer Reports*. https://www.consumerreports.org/privacy/amazon-echo-dot-kids-violates-privacy-rules-advocates-claim/

Johnson, S. D., Blythe, J. M., Manning, M., & Wong, G. T. W. (2020). The impact of IoT security labelling on consumer product choice and willingness to pay. *PLoS ONE*, *15*(1), 1–21. https://doi.org/10.1371/journal.pone.0227800

Joseph, T. (2018, December 28). Role of Data Analytics in Internet of Things (IoT). *Fingent*. https://www.fingent.com/blog/role-of-data-analytics-in-internet-of-things-iot/

Kim, N. S., Telman, D. A. (2015). Internet Giants as Quasi-Governmental Actors and the Limits of Contractual Consent, *Missouri Law Review*, *80*(3), 723-770. https://scholarship.law.missouri.edu/mlr/vol80/iss3/7/

McKay, K. A., Bassham, L., Turan, M. S., & Mouha, N. (March 2017). Report on Lightweight Cryptography. *NISTIR 8114*. https://doi.org/10.6028/NIST.IR.8114

NetScout. (2018). Netscout threat intelligence report 2H 2018. https://www.netscout.com/sites/default/files/2019-02/SECR_001_EN-1901%20-%20NETSCOUT%20Threat%20Intelligence%20Report%202H%202018.pdf

Ng, Alfred. (2019, July 2). Amazon Alexa keeps your data with no expiration date, and shares it too. *cnet*. https://www.cnet.com/home/smart-home/amazon-alexa-keeps-your-data-with-no-expiration-date-and-shares-it-too/

Pedreschi, D., Ruggieri, S., Franco, T. (2008). Discrimination-aware data mining. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 560-568. https://www.researchgate.net/publication/221654695_Discrimination-aware_data_mining

Peppet, S. R. (2014). Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent. *Texas Law Review*, *93*(1), 85–178.

Regulation 2016/679. *On the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. European Union, Council of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e1797-1-1

Rehman, A. U., Rehman, S. U., Khan, I. U., Moiz, M., & Hasan, S. (November 2016). Security and Privacy Issues in IoT. *IJCNIS*, *8*, 147-157. https://www.researchgate.net/publication/313574376_Security_and_privacy_issues_in_IoT

Rose, K., Eldridge, S., Chapin, L. (2015). The Internet Of Things: An Overview. *The Internet Society (ISOC)*. https://www.internetsociety.org/wp-content/uploads/2017/08/ISOC-IoT-Overview-20151221-en.pdf

Sadler, M. (2017). Securing Our Connected World. *Department for Digital, Culture, Media, and Sport*. https://dcmsblog.uk/2017/10/securing-connected-world/

Schneier, B. (2017). Click Here to Kill Everyone. *Intelligencer*. https://nymag.com/intelligencer/2017/01/the-internet-of-things-dangerous-future-bruce-schneier.html

Stanton, B., Theofanos, M. F., Prettyman, S. S., Furman, S. (2016). Security Fatigue. *IT Pro*, 26-32. https://inside.mines.edu/UserFiles/File/ccit/security/NIST-Security_Fatigue.pdf

The Online Privacy Protection Act of 2003, Cal. Bus. & Prof. Code § 22575-22579 (2004). https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=BPC&division=8.&title=&part=&chapter=22.&article=

Tschider, C. A. (November 2018). Regulating the Internet of Things: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age. *Denver Law Review*, *96*(1), 87–143. https://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=135154524&site=ehost-live&scope=site.

Tung, L. (2019, May 10). Amazon can't yet completely delete Alexa voice transcriptions. *ZDNet*. https://www.zdnet.com/article/amazon-cant-yet-completely-delete-alexa-voice-transcriptions/

Verzun, E. (2020, April 28). How Tech Giants Benefit From Data Collection? *EvgenVerzun.com*. https://evgenverzun.com/how-tech-giants-benefit-from-data-collection/#:~:text=Big%20firms%20not%20only%20use,if%20they%20could%20flourish%20further.

Wachter, S. (2018a). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer Law & Security Review*, *34*(3), 436–449. https://doi.org/10.1016/j.clsr.2018.02.002

Wachter, S. (2018b). The GDPR and the Internet of Things: a three-step transparency model. *Law, Innovation & Technology*, *10*(2), 266–294. https://doi.org/10.1080/17579961.2018.1527479

Zheng, S., Apthorpe, N., Chetty, M., & Feamster, N. (November 2018). User Perceptions of Smart Home IoT Privacy. *Proc. ACM Hum.-Comput. Interact, CSCW, 2*, Article 200, 20 pages. https://doi.org/10.1145/3274469