Investigating the Datafication of Your Car

Sarah Elizabeth Gillespie
Cybersecurity and Privacy Institute
Northeastern University
Boston, MA, USA
gillespie.s@northeastern.edu

Christo Wilson
Cybersecurity and Privacy Institute
Northeastern University
Boston, MA, USA
cbw@ccs.neu.edu

I. Introduction

Years of research examining "smart" Internet of Things (IoT) devices have found that developers of these devices have poor privacy practices [1], [2]. This includes collection of personal information without notice or consent [3]–[5] and transmission of data to online advertisers and data brokers [6].

Automobiles are the next frontier for the "datafication" of consumer devices. To the best of our knowledge, all recent model year vehicles available in major market are *connected cars*: they include always-on internet connections, collect and transmit data about the vehicle and the driver, and incorporate companion smartphone apps.

A recent report from the Mozilla Foundation highlights privacy concerns around connected cars [7]. This report drew on automakers' privacy policies to identify the different types of data that automakers claim to collect and disclosures around data selling and sharing to third-parties. The practices revealed in the report are very concerning, in part because vehicle ownership is a de-facto requirement for modern life and car owners have little—if any—ability to opt-out of vehicle's data collection. Recent reporting revealed that automakers are sharing driving data with insurance companies and that this is causing real-world harms to vehicle owners [8].

In this study we propose to use data subject access requests issued under the CCPA and GDPR to investigate the privacy implications of connected cars. The main shortcoming of the Mozilla study is that the report relies on disclosures from privacy policies, which may under- and over-disclose data collection and sharing practices. For example, studies of privacy policies from websites and mobile apps have found that they sometimes contain vague language that permits all data to be collected and shared [9]—thus revealing nothing about actual collection and sharing practices—or they fail to disclose all practices.

Our goal is to obtain at least one data report from every major auto manufacturer to examine the type and granularity of data being collected, as well as compare the provided data to the manufacturer's privacy disclosures. Additionally, we will examine the types of data that are not included in the reports to uncover potential violations of data subject access rules.

II. Related Work

Related studies have begun to investigate connected cars, including the readability of their privacy policies [10], [11], the legality of data collection from cars [12]–[15], and the security of vehicle data [16], [17]. Other studies focus on using vehicle data to train self-driving cars or traffic safety applications [18].

III. Methodology

We plan to perform a crowdsourced audit of data collection by connected auto manufacturers. We will recruit multiple vehicle owners to request the data for their own vehicle—across brands, manufacturers, and recent model years—from the automakers and third-parties that are known to receive vehicle data (e.g., Experian [8]) and donate this data to us. Since some manufacturers only honor data requests from locations with comprehensive data privacy legislation (e.g., California and Europe) [19], we will focus on recruiting participants in these locations. We do not anticipate needing a representative population for this study, and we recognize that the data in question is very sensitive, so we are planning to directly recruit participants from friends, family, and colleagues.

It would be challenging to collect a data report from an instance of every single car make, model, and options package produced between the year 2018 and 2024, so our goal is more modest: to get at least one report for each brand of car discussed in the Mozilla Foundation report [7]. Further, it may be unnecessary to gather data reports from each make and model of car from a given manufacturer because, to save on development costs, connected cars often share an underlying telematics platform. For example, Lexus is a sub-brand of Toyota, and we expect that cars from both brands will share much of the same data collection infrastructure.

Once we have data reports from participants, we plan to compare their content across vehicles and manufacturers. We will also compare them to (1) automakers' stated privacy policies and (2) data made available to third parties by automakers to identify discrepancies. Given that the number of automakers is small, we do not anticipate the need for automation to review the privacy policies [20]. If we are able to recruit multiple participants who own vehicles from a given automaker, this will enable us to assess the consistency of data reports within that automaker. Finally, if we are able to recruit participants in the U. S. and Europe, this may facilitate trans-continental comparisons of data collection by connected cars.

References

[1] J. Ren, D. J. Dubois, D. Choffnes, A. M. Mandalari, R. Kolcun, and H. Haddadi, "Information exposure from consumer iot devices: A multi-

dimensional, network-informed measurement approach," in *Proceedings of the Internet Measurement Conference*, 2019, pp. 267–279.

[2] P. Emami-Naeini, J. Dheenadhayalan, Y. Agarwal, and L. F. Cranor, "Which privacy and security attributes most impact consumers' risk perception and willingness to purchase iot devices?" in *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2021, pp. 519–536.

[3] E. Pan, J. Ren, M. Lindorfer, C. Wilson, and D. R. Choffnes, "Panoptispy: Characterizing audio and video exfiltration from android applications," *Proc. Priv. Enhancing Technol.*, vol. 2018, no. 4, pp. 33–50, 2018.

[4] A. Subahi and G. Theodorakopoulos, "Ensuring compliance of iot devices with their privacy policy agreement," in *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*. IEEE, 2018, pp. 100–107.

[5] G. Chalhoub, M. J. Kraemer, N. Nthala, and I. Flechais, ""it did not give me an option to decline": A longitudinal analysis of the user experience of security and privacy in smart home products," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–16.

[6] C. Leung, J. Ren, D. Choffnes, and C. Wilson, "Should you use the app for that? comparing the privacy implications of app- and web-based online services," in *Proceedings of the 2016 Internet Measurement Conference*, 2016, pp. 365–372.

[7] M. R. Jen Caltrider and Z. MacDonald. (2023) It's official: Cars are the worst product category we have ever reviewed for privacy. [Online]. Available: https://foundation.mozilla.org/en/privacynotincluded/articles/its-official-cars-are-the-worst-product-category-we-have-ever-reviewed-for-privacy/

[8] K. Hill. (2024) Automakers are sharing consumers' driving behavior with insurance companies. [Online]. Available: https://www.nytimes.com/2024/03/11/technology/carmakers-driver-tracking-insurance.html?smid=nytcore-android-share

[9] J. R. Reidenberg, J. Bhatia, T. D. Breaux, and T. B. Norton, "Ambiguity in privacy policies and the impact of regulation," *The Journal of Legal Studies*, vol. 45, no. S2, pp. S163–S190, 2016.

[10] C. Bodei, G. Costantino, M. De Vincenzi, I. Matteucci, and A. Monreale, "Vehicle data collection: A privacy policy analysis and comparison." in *ICISSP*, 2023, pp. 626–633.

[11] N. Liu, A. Nikitas, and S. Parkinson, "Exploring expert perceptions about the cyber security and privacy of connected and autonomous vehicles: A thematic analysis approach," *Transportation research part F: traffic psychology and behaviour*, vol. 75, pp. 66–86, 2020.

[12] H. B. Peacher, "Regulating data privacy of connected vehicles: How automotive giants can protect themselves and their golden goose," *Alb. LJ Sci. & Tech.*, vol. 30, p. 74, 2020.

[13] D. Gill, "The data act proposal and the problem of access to in-vehicle data and resources," *Available at SSRN 4115443*, 2022.

[14] M. C. Gaeta *et al.*, "Data protection and self-driving cars: The consent to the processing of personal data in compliance with gdpr," *Communications Law*, vol. 24, no. 1, pp. 15–23, 2019.

[15] A. Domeikiene, "Data collected and generated by cars and its ownership in relation with the gdpr user rights," Master's thesis, 2017.

[16] B. Chah, A. Lombard, A. Bkakria, R. Yaich, A. Abbas-Turki, and S. Galland, "Privacy threat analysis for connected and autonomous vehicles," *Procedia Computer Science*, vol. 210, pp. 36–44, 2022.

[17] R. Dave, E. Boone, K. Roy *et al.*, "Efficient data privacy and security in autonomous cars," *Journal of Computer Sciences and Applications*, vol. 7, no. 1, pp. 31–36, 2019.

[18] R. Kandiboina, S. Knickerbocker, S. Bhagat, N. Hawkins, and A. Sharma, "Exploring the efficacy of large-scale connected vehicle data in real-time traffic applications," *Transportation Research Record*, 2023.

[19] *Which States Have Consumer Data Privacy Laws?*, Bloomberg Law, 2023, accessed 12 March 2024. [Online]. Available: https://pro.bloomberglaw.com/insights/privacy/state-privacy-legislation-tracker/

[20] B. Andow, S. Y. Mahmud, J. Whitaker, W. Enck, B. Reaves, K. Singh, and S. Egelman, "Actions speak louder than words: Entity-Sensitive privacy policy and data flow analysis with PoliCheck," in *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, Aug. 2020, pp. 985–1002. [Online]. Available: https://www.usenix.org/conference/usenixsecurity20/presentation/andow

[21] *Your Privacy Rights*, Toyota Motor Sales, U.S.A., Inc., 2024, updated as of January 1, 2024. [Online]. Available: https://www.toyota.com/support/privacy-notice/

[22] *Privacy Policy*, 2022, last Updated December 21, 2022. [Online]. Available: https://www.subaru.com/support/privacy-policies.html

[23] *Genesis Motor America Privacy Policy*, Hyundai Motor America, 2023, effective date: August 21, 2023. [Online]. Available: https://www.genesis.com/us/en/my-privacy-rights

[24] *Customer Privacy Notice*, Tesla, 2023, updated May 2023. [Online]. Available: https://www.tesla.com/legal/privacy

IV. Appendix

A. Funding

Sources of Funding: none.

V. Preliminary Results

We have already acquired three data reports from major automakers. The Hyundai Motor America/Genesis Motor report pertaining to a GV60 included, among other things, marketing insights on the owners gender, income, age, ethnicity, and religion; and a list of notifications regarding unsuccessful remote door lock and electric charge commands with the specific street address of where those events occurred.

The Toyota America report [21] pertaining to a Subaru [22] Solterra included similar information, as well as a disclaimer asserting that, while Toyota had collected more data from the vehicle, they would not produce it as it might compromise the privacy of non-vehicle owners who had driven or ridden in the car. The limited information in these two reports surprised us because we expected to find records of all mobile app requests—not just unsuccessful requests—and data about the vehicle's operations—as permitted in the vehicle company's privacy policies [23]—such as geolocation, driving speed, use of vehicle features, and possibly images from exterior cameras.

In contrast, the Tesla report pertaining to a Model 3 contained specific data about how and when the car was used. This data included details about how and where the Tesla was charged; millisecond-by-millisecond, highly granular driving data (e.g., break pedal application, accelerator pedal position, autopilot lane changing, etc.); and all mobile app requests. There was a specific place in the data file where critical safety event video footage would have been provided if this specific Tesla had any critical safety events. Notably, the report did not include geographic data of the vehicle's location nor obvious marketing data [24].