

Working Paper on Telemetry and Diagnostic Data

Adopted at the 71st Meeting on 7th-8th June 2023

Written Procedure after this meeting

Version 0.99 2023-10-11

Introduction

1

Goal of this paper

1. The goal of this paper is to give recommendations to relevant stakeholders on how the privacy and data protection impact of the collection and processing of diagnostic or telemetry information can be limited, and on how the interests of vendors and developers could be balanced against the users' rights to privacy.

Scope and Terminology

2. This paper will look at the privacy and data protection issues related to the collection of telemetry and diagnostic data as implemented in mobile and desktop client operating systems and apps. The same issues also exist in numerous service applications deployed by organizations inside their own IT infrastructure, such as database and videoconferencing servers and the related client software. It will also cover issues with the use of frameworks for the collection of telemetry and diagnostic data used by developers of mobile apps and cloud-based web services.

3. The paper will focus on information collected and transmitted by devices or applications for technical purposes unrelated to the primary purpose of the device or application. Such purposes include monitoring the performance of devices and services, troubleshooting and debugging, quality assurance and service improvement as well as research and development of new products and services. It will not cover information that serves the primary purpose of the device or application in question for example in connected medical devices such as pacemakers or insulin pumps, or in various remote-control scenarios.

The Definition of Personal Data

4. In this paper we will define as personal data any information relating to an identified or identifiable natural person. In the same way we will only consider such data anonymised if they cannot be linked to a natural person by any means reasonably likely to be used.
5. In discussions around the topic of telemetry and diagnostic data it often turns out that vendors and developers have a very different understanding of what constitutes personal data than data protection advocates or authorities, even more so as the definition also varies between jurisdictions. In many cases developers or vendors only consider information as personal data (or Personal Identifiable Information) if it is directly linked to a known individual and contains a set of “personal identifiers” such as a national ID number or an email address.
6. Even when developers and vendors recognize that they are processing data that falls under a broader definition of personal data, they will often raise the objection that they will not process the data in relation to that individual but will only use the data in “de-identified” or “anonymized” form. In these cases, the “de-identification” or “anonymization” usually only consists in the removal of direct identifiers.
7. The IWGDPT would like to underline that data falling under the above definition are personal data even if the connection to an individual is not direct (i.e. through natural identifiers such as first and last name or an official ID number), but via a pseudonym or circumstantial information. Data that could be linked to an identifiable individual is personal data even if the entity processing the data has not identified this person and does not have any intention to do so. This includes pseudonymized data that could potentially be used to single out and have an impact on the specific individual.

Distinguishing Between Telemetry and Diagnostic Data

8. In this paper, we will distinguish between the two types of data, telemetry and diagnostic data, in the following way:
9. By telemetry data, we will mean data that is collected and transmitted by a device or application on a more or less continual basis. Telemetry data usually consists of information on operational behavior or environmental parameters but may also include elements like location information.
10. By diagnostic data, we will mean data that is collected in specific situations or events, e.g., under certain error conditions. Diagnostic data will often also contain information on operational parameters but may also include “user data” in the form of memory dumps or files that were open at the time the error was detected.
11. The issues and recommendations for telemetry data on the one hand and diagnostic data on the other are similar at the generic level. We will use both terms throughout this paper in order to point out that both types of data collection and processing share many data protection issues and that looking into finer details will only be necessary if much more specific recommendations are required.

Background

12. During the past two decades, the degree of “connectedness” in electronic devices has dramatically increased. Many devices now have an “always-on” internet connection either via a local or wireless network to which they are connected or through a network connection of their own, in this case usually via a mobile network. The availability of such connections and of affordable computing power to process the large amount of data collected has prompted developers and vendors of devices and software to integrate diagnostic functionalities into their products that implement a wide range of use cases, ranging from reporting configuration information or submitting error reports to an almost constant monitoring of the operational state of the device or software in question. Similarly, “legacy” operating systems and applications commonly used in desktop or laptop computers or servers often collect and transmit diagnostic or telemetry data (see para 9 and 10 above,) and providers of cloud-based services collect information about the usage of their services by their customers. It is worth noting that in the case of a cloud-based service, information collection can take place both on the client and on the server side.

13. The collected data can range from configuration options to full memory dumps of the application or device and could, even if it doesn't inherently contain personally identifying information itself, often be related to natural persons via a registered application or device, or at least via the network layer address of the corresponding device.
14. However, collecting diagnostic or telemetry information is often not limited to technical information in the strict sense or to information related to the operational state of a device, application or service. Some developers and manufacturers have taken to collecting other types of information that is not inextricably linked with the essential usage of a device or application, such as the location of the device or other contextual information. The information thus collected is then used for other purposes and sometimes monetized via other channels, providing another revenue stream for the developer or vendor.
15. In general, already the link between a device or application and the collected data about the use of the device or application will turn those data, as well as any inferred data, into personal data. Certain categories of data among those collected, such as location data, would further contribute to defining their nature of personal data. Elements such as the purpose of their use and the result of the processing, insofar as they have an impact on specific individuals (e.g. profiling to infer behavior and other possible characteristics) should also be considered. In many cases, setting up a newly purchased device will be difficult without creating an account with the device vendor, which may be linked to any information collected about the device. It should be pointed out that creating an account in order to set up a device is in many cases not needed from a functional point of view, and it should be avoided wherever possible. Furthermore, linking devices to accounts will often lead to different accounts belonging to the same individuals, which enables organizations having access to those different accounts and devices to build an enriched profile of that individual without a clear necessity.
16. Even if no other factors were present in the collected data that would allow linking them to an individual, the connection of a device or application to an individual user account will clearly turn all information collected into personal data. However, the IWGDPT wishes to underline that telemetry data can be personal data even if they do not contain a direct identifier such as an ID number or an e-mail address (see also below).
17. One key issue related to this kind of information gathering is that it is often done without the knowledge of the users of the device or application, or at least without informing them about the true extent of the data collection and the purposes for which the information is collected. A second issue is that there are usually few to no ways for users to limit the amount of information collected or to object to the gathering of the information. Even when users are informed about the fact that information is collected, there is often insufficient

information about the kind and amount of information that will be transmitted. Moreover, the purposes for which the information will be used are only rarely stated precisely.

18. Where organizations manage and deploy in their IT infrastructure services and applications (i.e. in a business to business setup, as opposed to a direct business to consumers relationship), the collection and transmission of telemetry and diagnostic data raises additional issues relating to the organizations' control over their processing and their information security and obligations under the applicable privacy and data protection legislation. Furthermore, the collection of information about the way applications and services are used could also lead to conflicts with labor law, for example if they could be used to monitor the employees of the organizations using the service.

Use Cases

19. The following high-level use case descriptions can serve to illustrate the issues at hand:

Use Case 1: Telemetry and Diagnostic Data in Operating Systems

20. Current operating systems (mobile, desktop and server) integrate the functionality to collect and store data about operating parameters of the device, operating system and applications, and about the interaction of users with that device, operating system and the applications. The systems integrate frameworks that allow the collection of hundreds of data types (often also called "metrics"), ranging from various system parameters (CPU and network load, memory usage), over a list of running applications and services up to individual keystroke sequences and memory dumps of individual applications or the entire system memory. Mobile operating systems may also include location information in this data.
21. The collected data will usually be stored locally on the system and will be transmitted to a remote endpoint either at scheduled intervals or in case of certain events (such as the crash of an application, a driver, or the system itself). Device owners and users have at least a theoretical possibility to influence the collection and transmission of the data and to view the data that has been collected while it is still stored on the system itself. In practice, however, the control over the data collection is usually very limited. Even where the operating system

provides a set of controls to the users there is often a minimum level of data collection and transmission that cannot be turned off through user or admin controls alone.

22. Data collected this way will at least be linked to a particular physical device. The connection to a natural person will be possible via a registration of the device itself or the operating system installed on the device, or when the person using the device is connected via an “online” account such as a Google account on Android, an Apple ID on iOS or a Microsoft account on Windows devices. If the device location data is collected as well, this contributes further to the personal data nature of the information collected by that device, since movement profiles will in most cases be relatable to individuals.
23. Furthermore, the collection and processing of information about what applications are installed on a device and running on the OS and the location information represent elements of specific concerns since this information can be used to build meaningful user profiles, including sensitive data and characteristics, ranging from religious and philosophical beliefs to health status. This applies even more on mobile devices, which are nowadays the information gateways to many people’s everyday life activities, where the location information and other ambient parameters can be also collected and thus enrich the user profile (see also use case 2).
24. Apart from data relating to the person or persons using a device, there is also the possibility that data concerning other natural persons is included in the data collection if files or memory dumps that are part of diagnostic snapshots contain personal information related to these third persons. This kind of data collection bycatch is usually not identified among processed personal data by the developers and therefore creates an often neglected data protection risk, as data of highly personal and sensitive nature may end up in data sets that are not assessed by their custodians, and not treated as such.

Use Case 2: Telemetry and Diagnostic data in Client-side Apps

25. In a similar way as operating systems, apps on fixed or mobile devices often collect telemetry or diagnostic data, in general using the facilities of the underlying operating system by including one of several available frameworks or libraries, or by implementing a collection functionality of their own. In case of use of third party libraries, the app might transmit telemetry or diagnostic data not only to the app developer, but also to the third parties providing the libraries they embedded in the app.
26. As with data collected by the operating system, the data will usually be linked to a particular device both by using a device or OS identifiers (see use case 1) and via an “installation ID”, or

to a natural person through the information used to register the app. The device identification might also proceed indirectly, e.g., by fingerprinting characteristics of the device, in particular if cross-app identification is desired.

27. Depending on the application in question, the amount and sensitivity of the data can be comparable to the information collected by operating systems. One limiting factor in this case is that apps especially in modern mobile operating systems are usually sandboxed and only have access to their own data and to a subset of other environmental information (such as location and other sensor data) which in many cases can be controlled by the users via app system settings in the operating system. However, app restrictions imposed by the operating systems are not able to solve the issue to a satisfactory granularity level, because in many cases an app may have legitimate reasons to e.g., request location access, but might use that access to include location information in telemetry or diagnostic data, as well.
28. The amount of control that users have over the data collection by client-side apps is comparable to the case of operating systems. That said it is worth noticing that there seems to be a wider spectrum in the use of telemetry in apps than in operating systems, with some apps generating only little to no telemetry data and others generating large amounts of telemetry on several levels.

Use Case 3: Telemetry and Diagnostic Data in Server-side Applications

29. Probably the most difficult use case for privacy and data protection regarding telemetry and diagnostic data collection is the collection of such information by server-side components in client-server or cloud-based scenarios. In this case, operators do not usually talk about telemetry data, since the collection takes place at the server side under their full control and does not aim at measuring the operational status of the remote device and its environment, but focuses on monitoring the usage of the applications. This collection still enables the operator of the service to build profiles of users' behavior and to combine data collected at server side with data collected at client side (if applicable). Although many challenges for users are similar (e.g. lack of information and control about the amount and type of data collected), users have even less control over the collection, as they have no possibility at all to physically monitor what data are collected unless an audit is carried out at the operator's data centre.

Privacy / Data Protection Risks

30. We consider the following non-exhaustive set of privacy and data protection risks that arise from the collection and processing of telemetry and diagnostic data.

Lack of Awareness on Telemetry and Diagnostic Data as Personal Data

31. An important risk in connection with the processing of telemetry or diagnostic data is a widespread lack of awareness on the developers' and vendors' side that telemetry or diagnostic data is in fact often personal data. This implies that developers and vendors do not see the need and the obligation set up a data protection by design and by default approach, nor to be compliant and accountable in respecting principles such as lawfulness, data minimization, purpose specification etc.
32. On the customers' side there is often a lack of awareness what information is collected, or that telemetry and diagnostic data is collected in the first place, which is often due to a lack of information about the data collection (see below).

Lack of Transparency

33. As long as customers and users are not informed about what information is collected under which circumstances, when it is transmitted and to whom, for which purposes it is processed by the recipients and which risks collection, transmission and subsequent processing entail, or if they are unaware of the information collection altogether, they have no possibility to object to the collection when possible and in any way limit their exposure to any potential adverse consequences of the use of that information.
34. Even in cases where a vendor or developer informs users about the fact that information is collected and explicitly assumes the role of a data controller, there is often insufficient information about the precise nature (e.g. data types) and amount of information (e.g. frequency or granularity) that is collected for each purpose of the collection.

Insufficient Data Minimization

35. The information collected via telemetry and diagnostic data collection may encompass information relating to persons using a device or application or may contain personal data that is processed on the device or application, which can also be sensitive or be used to infer sensitive characteristics of individuals. If the information is not strictly necessary and proportionate for the purpose for which it is initially collected, but is stored just in case it might come in handy at a later stage, this creates a risk that the data will either get into the wrong hands in the course of an information security incident, or the data might be used for a different purpose later (see also below).

Insufficient Storage Limitation

36. In addition to collecting more data than necessary, operators frequently store the data for longer than required, based on the reasoning that it might help diagnose some as-yet unknown issues in the future. This is especially relevant for telemetry data, since this data mostly reflects the operational state of a device or application at a given point in time without specifically looking at any concrete issue. Diagnostic data, which is only collected under specific circumstances, will often also be kept beyond the time needed to resolve the specific issue that triggered the collection, e.g., for “reference purposes”.

Insufficient Purpose Limitation or Linking with Other Personal Data

37. If the purposes for which telemetry data or diagnostic information are processed are not clearly defined, then this will likely lead to a situation where more information is collected and stored than is necessary to achieve the purposes for which it is eventually used. Even if the data collected is strictly necessary for the primary purpose relating to the device or system’s functional requirements, but there are no processes in place that ensure that the data is not used for other purposes, this may lead to situations where persons are adversely affected by such secondary uses. Personal data from diagnostic or telemetry data might also be linked with other information a developer or vendor may possess, leading to additional risks, e.g., in connection with profiling or limiting of people’s consumer rights. Furthermore, device information could be used as a basis for fingerprinting to (re-) identify users even when anonymised data are collected from that device.

Lack of Fairness

38. Even if users are informed of the collection of telemetry or diagnostic data, there is often no possibility for them to object to this collection or to control the amount of information collected or transmitted. The question of fairness comes up even more pointedly in cases where the users' choice is limited to either submitting to the collection of their data or not using the device or service in question at all. Especially where the users' freedom to choose whether or not to use the device or application is limited (e.g. in employment or educational settings), this will create unfair pressure on the affected persons.
39. Another aspect of lack of fairness, linked with a lack of transparency, is a constellation in which a vendor provides controls and tools with which users can limit data collection and view collected data, but the controls do not give a sufficient amount of control (e.g. don't allow to turn off data collection completely) or the tools only show a subset of the collected data.

Insufficient Processes to Guarantee Individuals' Rights

40. Individuals must be able to exercise their rights to their personal data under the applicable legislation. As developers or vendors often lack sufficient knowledge of the requirements resulting from different legal frameworks, the corresponding processes are often not in place. This makes it difficult or impossible for individuals to exercise their rights.

Lack of Measures to Ensure Confidentiality / Integrity / Availability of Data

41. As with all personal data, a failure to ensure the confidentiality, integrity and availability of the information can lead to risks for the persons concerned. While this may seem of somewhat less importance in the context of telemetry or diagnostic data, violations of confidentiality or integrity of such data can adversely affect the persons concerned at least in the same way as in other scenarios in which personal data is processed.

Recommendations

Recommendations for Lawmakers

42. The IWGDPT encourages legislators to check whether the issues described in this paper are adequately addressed by the relevant legislation applicable in each country.
43. Clarify in law, if necessary and proportionate, under what conditions and for what purposes data may be processed and what data may be processed without consent. Require appropriate safeguards for the different types of processing, including, e.g., anonymization.
44. Encourage the possible adoption of telemetry and diagnostic processing standards.

Recommendations for Developers / Vendors / Manufacturers

45. Acknowledge the fact that information collected via telemetry or diagnostics functions will constitute personal data in many cases, even if the purpose for which the data is collected does not necessarily require relating it directly to individual persons.
46. Clearly define and document the purposes for which you are collecting data the types of data necessary for each purpose and be able to justify why certain data needs to be processed to achieve a specific purpose and whether the designed processing is proportionate.
47. Based on necessity and proportionality considerations,
 - a) limit the volume of telemetry and diagnostic data that is collected and transmitted for telemetry and diagnostic purposes and restrict the categories of information that will be collected,
 - b) limit the time for which telemetry and diagnostic data is stored,
 - c) avoid collecting potentially sensitive personal data.
48. By default, collect at most the information that according to the applicable laws may be collected without consent. More specifically, collect by default only data strictly necessary to correctly operate the service or device.
49. Provide a possibility to set up a device without having to create an account using personal data that also gives access to relevant firmware updates. This does not preclude (additional) offers requiring personalized accounts for features or services that actually require it.

50. Provide clear information to customers about what information will be collected under what circumstances, for what purposes, and whether it will be transmitted or transferred (and to whom). This information should be provided to the users as early as possible during the setup process, and for any information collection that goes beyond what may be collected without consent, provide a possibility to continue the setup process without giving consent. The information should be given in plain language, avoiding when possible technical jargon. When the amount information on telemetry / diagnostic data is significant, it should be provided in layers, allowing users an easy and flexible access to different information sections.
51. Put processes in place to allow people to exercise their data subject rights under the applicable legislation.
52. Provide easily accessible controls that enable users
 - a) to easily view any telemetry or diagnostic information collected (“data viewer”),
 - b) to disable or at least limit the transmission of telemetry or diagnostic data, and
 - c) to delete or reset any explicit identifiers that are used in the collection process.

If separate apps or applications are used to provide these controls, they should be made a part of the “core package” of an install and there should be no need to separately download and install additional software.

53. Put safeguards in place to ensure the confidentiality, integrity and availability of personal data that may be contained in diagnostic or telemetry data.
54. Minimize the sharing of personal data in telemetry and diagnostic information.
55. Use pseudonymization techniques whenever possible and work with aggregated and anonymized information as much as possible. Make use of existing privacy enhancing technologies to minimise the transfer of personal data. If working with aggregated or anonymised data is not feasible for the set purpose, mitigate the risk of identification by using techniques such as differential privacy. Keep any data that could still be related to individual customers or users only as long as necessary. For example, if certain data is only collected to generate statistics about the number of installs or the usage of certain features, then any identifying components in the data should be deleted immediately after collection or might not need to be collected at all if there is no need to refer to the same device or user.
56. Implement the principles of data protection by design and by default also in the context of diagnostic and telemetry data collection. Apart from data minimization, transparency, storage limitation and information security measures, this also means that updates should

not override or reset any settings that users have already made regarding telemetry and diagnostic data collection and processing.

Recommendations for Institutional / Commercial Customers

57. Use the available information on telemetry and diagnostic data and their privacy data protection features, as identified above, including the ability to be in control of their collection and transmission as criteria when selecting vendors / products.
58. In contracts with prospective vendors do not accept the “take it or leave it” approach. If not yet provided for in the standard contract clauses used by the vendor, negotiate terms and safeguards that ensure that you remain in control of the use of the product or services, including of their diagnostics and telemetry features and relevant data protection and security aspects.
59. Comply with all data protection obligations connected with the transmission of telemetry and diagnostic data relating to your employees and other users under your supervision to developers, vendors and other recipients.
60. Use product configuration, rights and service management, and network controls to limit and disable the flow of telemetry and diagnostic data, the processing of which unduly impinges on the privacy of employees and other users under your supervision.

Recommendations for Consumers / End Users

61. Use the available information on telemetry diagnostic data and their privacy data protection features, as identified above, including the ability to be in control of their collection and transmission as criteria when selecting products.

Recommendations for DPAs, Privacy and Data Protection Advisors and Advocates

62. Raise awareness among developers about data protection principles regarding telemetry and diagnostic data and, when necessary, enforce compliance in line with your competence.

63. Demand clear legal requirements from legislators to ensure that the data protection obligations are clearly defined and executable.
64. Develop an accountability framework that developers or vendors can use to implement good practices for collecting and processing telemetry and diagnostic data. This should include suggestions for processes to select purposes, define what and how much data may be collected, for the handling of the collected data, and for the information that should be made available to users about the collection of the data.

Links / References

- [1] European Data Protection Board; EDPB Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications; https://edpb.europa.eu/our-work-tools/documents/public-consultations/2020/guidelines-12020-processing-personal-data_en
- [2] European Data Protection Board; EDPB Guidelines 4/2019 on Article 25 Data Protection by Design and by Default; https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en
- [3] Norwegian Data Protection Authority; Software development with Data Protection by Design and by Default; <https://www.datatilsynet.no/en/about-privacy/virksomhetenes-plikter/innebygd-personvern/data-protection-by-design-and-by-default/>
- [4] German Federal Office for Information Security; SiSyPHuS Win10: Study on System Integrity, Logging, Hardening and Security relevant Functionality in Windows 10; https://www.bsi.bund.de/EN/Service-Navi/Publikationen/Studien/SiSyPHuS_Win10/SiSyPHuS.html
- [5] Threatpost; Apple, Google Both Track Mobile Telemetry Data, Despite Users Opting Out; <https://threatpost.com/google-apple-track-mobile-opting-out/165147/>
- [6] Douglas J. Leith; Mobile Handset Privacy: Measuring The Data iOS and Android Send to Apple And Google; https://www.scss.tcd.ie/doug.leith/apple_google.pdf
- [7] Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder; Telemetriefunktionen und Datenschutz beim Einsatz von Windows 10 Enterprise; https://datenschutzkonferenz-online.de/media/dskb/TOP_30_Beschluss_Windows_10_mit_Anlagen.pdf

[8] The Register; Android's Messages, Dialer apps quietly sent text, call info to Google
https://www.theregister.com/2022/03/21/google_messages_gdpr/

[9] Privacy Company; New DPIA for the Dutch government and universities on Microsoft Teams, OneDrive and SharePoint Online; <https://www.privacycompany.eu/blogpost-en/new-dpia-for-the-dutch-government-and-universities-on-microsoft-teams-onedrive-and-sharepoint-online>

[10] Rob O'Leary; VS Code - What's the deal with the telemetry?;
<https://www.roboleary.net/tools/2022/04/20/vscode-telemetry.html>

[11] Microsoft Windows 10 Home and Pro investigation by the Autoriteit Persoonsgegevens (Dutch DPA), August 2017;
https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/public_version_dutch_dpa_informal_translation_summary_of_investigation_report.pdf

[12] European Data Protection Board - 2022 Coordinated Enforcement Action - Use of cloud-based services by the public sector - Adopted on 17 January 2023. See in particular section 3.7;
https://edpb.europa.eu/system/files/2023-01/edpb_20230118_cef_cloud-basedservices_publicsector_en.pdf