The Legal Problem of the right to access in ambient intelligence systems

Rens van der Heijden
rensvdheijden@gmail.com
University of Twente
RU student number: 4095057

July 26, 2012

Abstract

This document will discuss the legal problem posed by the right of access to the processing logic of automated individual decisions, as defined in Article 12a jo 15 in Directive 1995/46/EC within the context of cyberspace. Furthermore, it will be argued that this problem requires at least in part technical solutions and two of these. Finally, the interpretation these solutions give to the legal norm will be discussed.

1 Introduction

Since 1995, Directive 1995/46/EC [5] has been a guideline for the legal requirements for any public and private initiative that processes the personal data of citizens of the European Union. However, since the rise of the Internet and the wide acceptance of new technologies like smartphones and tablet computers, the framework that this legislation provides is limited. Many new interpretations of the legislation have been made in order to maintain consistency and legal certainty for the natural persons protected by this directive. For example, in 2008, the German constitutional court has set guidelines for remote search, where law enforcement is allowed to search the computer of a suspect under strict legal safeguards. Similarly, with upcoming technologies like the smart grid[1], vehicular networking[2] and other ambient intelligence technologies, new interpretations of the legal norm are required. However, since ambient intelligence makes decisions on behalf of someone, either a natural person or a legal person, only a new interpretation will not be sufficient to protect the fundemental rights of natural persons. This document will discuss the legal problem in detail in Section 2 and propose technical design solutions to ensure the protection of the rights granted by [5], in the context of this new technology[3] in Section 3. The effects of these solutions, particularly the new interpretation of the current legal norm that they provide, is discussed in Section 4. Finally, this document is concluded in 5.

2 The legal problem

Directive 1995/46/EC [5] defines the conditions that should be met for processing of personal data. This section discusses why data processed by ambient intelligence systems is personal data and presents the challenges posed by Directive 1995/46/EC. For the purpose of this analysis, two example applications will be used; the smart grid and vehicluar networking. However, the intent of this analysis is to be as general as possible.

[1] The smart grid is envisioned to conserve power by making management of generators and power usage more dynamic. This is achieved by smart power meters that start certain power consuming procedures at the right time, and by sending detailed power usage estimates to the utility company. The data is then used to predict future power usage, to predict how many generators are necessary.

[2] In vehicular networking, vehicles on the road will use wireless communication technologies to provide safety, entertainment and information applications. For example, automated breaking, traffic data collection and cooperative adaptive cruise control and eventually automated driving.

[3] Similar legal problems also exist in existing systems, like GSM and Facebook, but design solutions will not help there. Changing the design of an existing IT system is problematic at best; it is also the cause of many failing projects in the IT sector.

2.1 Personal Data?

Article 2 defines personal data as 'any information relating to an identified or identifiable natural person'. An identifiable person refers to a person that can be directly or indirectly identified, either through an identifier or through one or more factors specific to his 'physical, physiological, economic, cultural or social identity'. Notably this definition includes indirect identification, making it difficult to be completely certain when data is sufficiently anonymized.

In the context of the smart grid, the data collected typically consists of detailed information about how much power is used at which time. The granularity is expected to be anywhere between minutes and hours. Particularly for high granularity meters, one can determine economic and social identity from the patterns in power usage, even without information about which devices are used.

For vehicular networking, the personal data involved is again in particular the economic and social identity of the natural person. In this case, the location and route used by the vehicle is the most important data that the network uses. This location is broadcasted by each vehicle periodically, typically with a frequency of 10 Hz [1], presenting an enormous privacy risk if not protected properly. Additional data collected by traffic information systems includes average speed and density of vehicles on a road. However, such additional information is typically aggregated over a large section of a road and is thus not very privacy sensitive.

2.2 Automated Decisions

The second relevant part of [5] is Article 15. This article requires that Member States provide every person the right not to be subject to fully automated decisions that produce legal effects that relate to him and intended to evaluate personal aspects relating to him (par. 1). The only exceptions, subject to the rest of the directive, are that the decision is authorized by a law that also provides safeguards to protect the data subject (par. 2b) or that the decision is taken as part of entering or performing a contract, provided this contract is requested by the data subject or that there are sufficient safeguards to protect the data subject (par. 2a).

For ambient intelligence, this article is one of the more challenging parts of [5]: the fundament of ambient intelligence is the generation of automated decisions. Certainly, the use of such systems by the data subject can be seen as consent; however, some ambient intelligence systems may exist in a public area or on the Internet, which means that they will also analyze natural persons that have not given consent. In addition, notice that consent is not sufficient to avoid the requirements posed by this article.

Data collected by an ambient intelligence will be aggergated into one or more profiles of natural persons, or at least correlations between them, in order to create the intelligence[4]. This makes it not only difficult to exclude natural persons from analysis, but also undesirable from the accuracy point of view. Because exclusion of natural persons from the collection process is not an option, the next step is to exclude natural persons only from decisions made by ambient intelligence. This is challenging, as ambient intelligence is usually intended to be ambient; out of sight and autonomous, performing tasks for its users[5]. However, since the article is formulated as an opt-out mechanism, it should be technically possible to exclude specific natural persons.

Finally, note that automated individual decisions are not specifically defined. Thus, it could be easy to circumvent the law by claiming a human always presses an OK button before a decision is made; this article assumes such loopholes are fixed in the implementations of this directive.

2.3 Access and Objection Rights

Articles 12 and 14 of [5] provide the data subjects with two important rights; the right to access the information collected about them (12) and the right to object to the collection of data (14). For the purpose of this discussion, the focus will be on Article 12, which constitutes the main legal problem of this document.

More specifically, Article 12 allows the data subject to obtain: confirmation about whether data is collected about him, the purpose, the catagories of data, the recipients of the data, the data itself and knowledge of the logic involved in processing if such processing is automated (12a). Here, the article may clash with intellectual property rights[6], though it may be argued that such rights are inferior to fundemental human rights. Thus it is up to the data controller to document its activities in such a way that he can comply with this article.

[4] Of course, it is debatable whether recognition and application of patterns constitutes intelligence, but for the purposes of this document this question is not really interesting.

[5] The users are not necessarily the natural persons; for example, in the case of the smart grid, the utilities are (also) users.

[6] Recall the recent news where facebook denied these requests on grounds of intellectual property and trade secrets.

Note that for a single human being, it is practically impossible to understand all the relations in a normal-sized database, let alone a large database with information generated by an ambient intelligence system. In order to comply with the directive, the data controller must provide access to all the data relating to the data subject, the type and purpose of this processing and the logic that performs any automated decisions with respect to the data subject. It is practically impossible for a human to collect this information in a reasonable timespan unless the data is properly organised. Thus, a technical solution is required to either organise or collect the data in such a way that responses to information requests can be handles 'without excessive delay or expense'.

In addition, Article 12b allows for rectification, erasure or blocking of data which does not comply with the directive, in particular inaccurate or incomplete data. This should not be a problem for ambient intelligence, as higher accuracy is in the interest of such a system. However, such a requirement may also be challenging when the risk of abuse is high, or when the system is highly distributed.

Finally, Article 12c provides a requirement of notification to third parties that have received the data after a change occurs through Article 12b, except when this involves disproportional effort. It is unclear whether this notification also entails a request for the same change to each third party; however the data subject should be able to request the appropriate information to request such a change if it is not implicit in Article 12c. One scenario where incorrrect or inaccurate data may remain is when a contract between the data controller and a third party expires before the data subject requests correction by Article 12c.

In Germany, Article 34 of the Datenschutzgesetz provides a strict definition of what data must be provided to the data subject concerning automated processing [3]. This data includes probability values for six months preceding the request, the types of data used to compute these probabilities and their meaning for this specific data subject in an understandable way. In some ways, the information provided on request is thus much more limited than what is required in the data protection directive; on the other hand, it ensures that requests cannot be waved away citing intellectual property issues or unreasonable cost.

3 Technical solutions

This section will discuss how to achieve what is often called 'Privacy by Design'[7] in order to protect the fundemental rights of natural persons. Two general approaches will be discussed; one aims at avoiding the Directive by not using personal data, while the other aims at storing sufficient details to provide the information required by Article 12 without releasing the software itself.

A third option, often advocated by privacy activists and programmers, is releasing the source code and design of the software in order to provide complete insight in what the software does[8]. While this has many benefits, like the fact that it allows for a thorough security analysis and there can be no doubt about the interpretation of what the software does, there are also problems. First, software is protected by copyright, which may not allow its publication. Second, publishing the source code means that third parties may be able to copy the techniques used in this software, or use the knowledge to manipulate the system to their wishes. Finally, there is no guarantee that the source code that is published corresponds to the source code that is released. Similar problems play a role with design documents.

The two approaches that will be discussed are each illustrated by an example; pseudonymity to guarantee that the data cannot be linked to any one natural person, and accountability, to ensure that the data required by [5] can always be provided in sufficient detail. These solutions themselves will work, if applied to the right situation; the discussion emphasizes potential challenges and possible remaining legal problems, as well as their interpretation of the law.

3.1 Pseudonymity

Pseudonymity is considered one of the core requirements of vehicular networks, providing a delicate balance between privacy, cost and functionality [2]. However, it may also be used in any other application where the processing logic is part of a system that is not under the direct control of its seller, such as an energy meter in the smart grid [4].

In vehicular networks each vehicle transmits beacon messages to its surrounding vehicles, to allow them to communicate and provide services like break notification and adaptive cruise control. To protect the network against abuse, each message is signed with a digital signature. If the signature is directly associated with an identity, as is the case for typical digital signatures, all privacy for the driver is clearly lost; anyone can set up a few access points in a city and track every vehicle that

[7] This was also discussed in the Privacy Seminar last year. The example used was from `http://privacybydesign.ca`.

[8] In addition, this is not really a technical solution.

passes. The solution for this problem is providing a number of identities (pseudonyms) the vehicle can use to sign its messsages. By making these identities random strings, it becomes difficult to track vehicles over long distances for outside attackers.

While one might argue that the data is not personal data, and therefore Article 12a is not a problem, Article 15 1 still applies; each vehicle that receives a message from a vehicle will use that message as input for an automated decision. This automated decision significantly affects the driver of this vehicle, because each vehicle decides its position on the road based on the location in the message. However, until a device for vehicular networking becomes mandatory by law, one can consider the purchase of such a device as as explicit consent to be subject to automated decisions as defined in Article 15 1.

In addition, it is not decided that pseudonyms are not personal data, specifically because it has been shown to be possible to attack many pseudonym schemes to revoke the privacy of the users. Finally, note the authority that issues the identities can still track each vehicle, because it knows which vehicle is assigned which identity[9]; thus, that authority processes personal data, because arbitrary network data can then be mapped to each vehicle[10]. This authority is most likely not involved in the programming tasks for vehicular networks; it is expected that these activities will be delegated to sellers of hardware or vehicle manufacturers. This then illustrates a clear conflict between Article 12 (a) and the intellectual property rights of the manufacturers, if pseudonyms are really considered personal data.

Thus, using pseudonyms as a method to hide the identity is not a perfect solution. However, it saves a lot of additional effort for developers if pseudonyms are implemented in such a way that they really hide the vehicle from identification by any particular organisation. If recoverable at all, the data is allowed to be collected by Article 13 1 for e.g. criminal investigations in the case of an accident.

3.2 Accounting

A completely different approach, more in line with the approach that Facebook and most other commercial companies have[11], is to store the profiles and types of information collected, so that they can be easily sent to a user that requests this information.

In a typical relational database, adding an additional table that associates each user with the profiles he is or has been part of is not technically challenging and if done properly also not cost-prohibitive. This method also allows for the addition of probability values and other interesting information. The data can then be stored, including its history, in a data warehouse. This is a type of database that is oriented more to long-term storage, aggregate queries[12] and preservation of history, as opposed to typical databases, which frequently change and only provide the basic information stored in the database. However, if such a database is modified by an ambient intelligence system, which may build, associate and de-associate profiles on the fly, the amount of changes is critical for whether such a system is feasible. Storing the history of such a system could prove too expensive, especially for small businesses that cannot afford data warehousing solutions.

This method could be used in a smart grid system, where a large data warehouse is already available to perform analysis to express when generators should turn on and to predict the power usage for the near future. Because the data warehouse is already available, it should be fairly simple to associate households with their different consumption profiles over time. Such a relation to one or more profiles can be expressed with a probability vector, each probability expressing the similarity with each existing profile. These vectors can be stored for 6 months (or perhaps longer, if they are used to create long term profiles) to comply with the German requirements. It is reasonable to expect the purposes of these probabilities to be documented in an understandable way to benefit the organisation as well as compatibility with the law.

Thus, an accounting system that maintains a history of relations and their associated probablitities should be able to fullfill a request without disclosure of additional logic, or copyrighted software. Any patents or other intellectual property do not play a role, because only the output of the process is used. From Articles 12 and 15, it is not clear whether decisions also need to be explained, or whether providing the probablities and their explaination is sufficient to comply with the request.

[9] The reason for this is basically accountability. There have been proposals that remove this requirement, but these are typically infeasible or probabilistic solutions. In addition, there are solutions that rely on an organisational separation, which does not seem strong enough for privacy.

[10] Such an attack has possible practical issues, because the data may be at different authorities. Depending on the proposal, governments, hardware manufacturers, vehicle manufacturers or separate organisations perform this task.

[11] For the purpose of this document, this refers to the few cases where companies actually produce a correct response in the first place.

[12] Aggregate queries are requests to a database that may perform complex operations before presenting the data. For example, they can be used to periodically generate profiles from a data warehouse containing several decades of customer data.

Since organisations will likely choose the most limited interpretation of the law, an accounting system should be sufficient to collect the required data.

4 Effects

The choice of technical solution is greatly impacted by the interpretation of the law by its designer. The previous section has already given a dicussion of possibilities and the related interpretations of law; each given solution relies on a specific assumption and this section will review the effects of these choices.

On the one hand, pseudonymous systems rely on the fact that pseudonymous data is not personal data. As long as the security assertions made for such systems hold, specifically that two pseudonyms can not be linked with more than negligible probability[13], then such a system is secure and thus is not required to respond to the requests based on Article 12 of [5]. However, the recoverability of pseudonyms in vehicular networking illustrates that using pseudonyms is not a magic solution to privacy issues. In general, it may be argued that pseudonyms are not sufficient to remove the label of 'personal data' from the data.

On the other hand, accounting systems provide exactly the information that is required by German law, but designing such a system properly may be cost-prohibitive and the results may not give a complete picture of the logic that a data subject requests. This method relies heavily on its interpretation of the required information about the logic in a request. If indeed the probabilities (and their explanation) used in computations are necessary and sufficient as a response to requests based on Article 12, then accounting is a good option. However, for systems that rely on different inputs for their computation, such as profiles, configuration parameters or general statistics from another data source, then accounting may provide insufficient details.

Finally, releasing source code may provide the easiest solution for organisations, if such source code does not compose a large portion of their income. However, German law prevents such a solution because it specifically requires probability values. While it is reasonable to require these values, this further illustrates that probabilities do not provide the complete picture.

5 Conclusion

This document has reviewed the legal problem of the right of access to the program logic that processes personal data in ambient intelligence. Two proposed technical solutions, pseudonyms and accounting, have been discussed and their suitability and interpretation of the law has been analyzed. While both solutions are adequate, which one should be used strongly depends on the application it is applied to. Because pseudonymity may be very difficult to provide in some situations, and because of the very specific German legislation, accounting may be the best solution for most organisations.

References

[1] ETSI and ITS WG1. Etsi ts 102 637-2 v1.2.1. Technical report, March 2011. Reference: RTS/TIS-0010018.

[2] Matthias Gerlach. Assessing and improving privacy in VANETs. In *4th Workshop on Embedded Security in Cars*, 2006.

[3] Mireille Hildebrandt. Law in cyberspace: Lecture 9 legal proection by design, November 2011. Translations for German law, slides 66-68.

[4] Klaus Kursawe, George Danezis, and Markulf Kohlweiss. Privacy-friendly aggregation for the smart-grid. In *Privacy Enhancing Technologies - 11th International Symposium*, pages 175–191, july 2011.

[5] European Parliament and of the Council of 24 October 1995. Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L 281, November 1995.

[13] Negligible probability may sound vague, but it is a well-defined term in cryptography. For a definition of it, refer to Definition 1.25 of `http://www.win.tue.nl/~berry/2WC13/LectureNotes.pdf`. Defintion 1.26 also defines statistical indistinguishability using the term Negligible, which is exactly the property that two pseudonyms of different users should have when compared to two pseudonyms of the same user.