Available online at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodclaw.htm

ELSEVIER

Computer Law & Security Review

Internet of things: Privacy issues revisited

Rolf H. Weber *

University of Zurich, Switzerland

Keywords:
Data minimization
Internet of things
Quality of data
Privacy challenges
Privacy enhancing technologies
Transparency

ABSTRACT

The Internet of Things presents unique challenges to the protection of individual privacy. This article highlights the growing need for appropriate regulatory as well as technical action in order to bridge the gap between the automated surveillance by IoT devices and the rights of individuals who are often unaware of the potential privacy risk to which they are exposed. As a result, new legal approaches for the protection of privacy need to be developed.

1. Starting point: challenges posed by the Internet of Things

1.1. *Technological background*

The Internet of Things (IoT) as an emerging global Internet-based information architecture that facilitates the exchange of goods and services is gradually gaining importance. The ITU defined the IoT as the development of item identifications, sensor technologies and the ability to interact with the environment.[1] In the meantime, the definition has been widened and it is now encompassing a broad spectrum of device forms that are used in a number of varying settings.

The most commonly known usage of the IoT is based on RFID (radio frequency identification device) technology that aims at preventing the disappearance of goods. However, other forms such as tracking parts through manufacturing processes and measuring variables such as temperature and humidity in a storage facility are common IoT applications as well. In practice, the level of sophistication and the price of RFID can be quite different, starting with the cheap passive device without a power source and limited storage to an active self-powered RFID possessing advanced storage and communication capabilities.[2]

Some of the data that are collected appear to be trivial but for example data relating to a production process could be highly valuable thus requiring appropriate protection. For private purposes, the IoT can be used to increase household efficiency by allowing the devices to communicate and take action such as place an order for goods when the fridge is empty or turn on the washing machine when electricity is cheap. The effects of malfunction created by wrong data (external and internal reasons) might be substantial in particular if a part of the decision-making process in a factory or household is automated. In such a case, the entire production line could be stopped or a customer could end up with double the quantity of goods he required. Furthermore, today all smart phones carry location sensors in them allowing the permanent tracking of their users. All these IoT devices in some form add value to individuals as well as businesses; however, they also cause risks.

* University of Zurich, Rämistrasse 74/38, 8001, Zurich, Switzerland. Tel.: +41 (44) 634'48'84; fax: +41 (44) 634'43'95.
E-mail address: rolf.weber@rwi.uzh.ch.
Chair Professor for International Business Law at the University of Zurich, Visiting Professor at Hong Kong University, Attorney-at-Law in Zurich.

[1] Definition of ITU (2005), available at <https://www.itu.int/osg/spu/publications/internetofthings/InternetofThings_summary.pdf>.

[2] RFID are classed according to their level of sophistication. Class 1 and 2 are passive RFIDs and Class 3 and 4 are active RFIDs which are commonly connected to a network and exchange data whereas Class 1 and 2 only are read by a scanner without actively collecting and submitting data on their own.

http://dx.doi.org/10.1016/j.clsr.2015.07.002

1.2. *Privacy risks*

The IoT devices collect a vast amount of information and, therefore, they also carry a great potential of privacy risks in relation to the use of the data and its access. Particularly the identification of an individual and his behavioral patterns is a growing concern. As IoT devices are increasingly used in all fields of daily life, such as in the health care sector, a great amount of commonly considered private information is stored and collected.

With the growth of these technologies, new safeguards for privacy and data integrity must be created. The IoT has a limitless potential to improve the daily life, for example in health care by allowing the collection of health information (e.g. with new FitBit/Jawbone devices recording basic health information through a wristband or electronic patient chip cards) which can be used to identify disease correlations and support new treatment options as well as remotely monitor the process of the treatment, however, the chances are correlating with the challenges. Similarly, with the help of Big Data analytics the accumulated raw data are highly valuable as specific patterns can be extracted, but the privacy risks naturally inherent are immense as the IoT data could allow the identification of an individual and thus his condition.

The IoT devices usually collect certain data that are often aggregated with other device data and thereafter sent via a router to a communication device (Wi-Fi or cellular) that transfers the data to a cloud server for processing. During this procedure various protocols and compression technologies are employed as the storage space on the devices is extremely limited and cannot cope with the big headers which for example are used for the Internet Protocol IPv6. Currently, providers attempt to filter data as closely as possible to the device that created it since this method avoids unnecessary transmissions and reduces safety risks.

Notwithstanding the fact that discussions about the normative framework governing the IoT are going on for the last five years[3] available legal assessments are still not stable. Furthermore, technologically and practically the interconnection between the devices and infrastructures has not yet reached a level that would allow its application in real life to a broad extent. However, this situation is changing with more and more services being offered based on IoT technology.

1.3. *Need for legal stability*

In view of the large range of IoT applications it is obvious that the new technological opportunities have organizational, social, and cultural implications. At the same time, various legislative instruments place limits on the IoT and its use in daily life; therefore, a single legal description cannot easily be developed. Moreover, data protection laws and privacy laws related to specific types of data must be considered. From a general perspective, the EU Data Protection Directive (DPD) is influencing the processing of data if the data collected are qualified as personal data. Other sector-specific regulations in particular in the USA (e.g. Health Insurance Portability and Accountability Act [HIPAA]) also have an effect on the data collection and the privacy of the data.

These regulations target at certain types of information, however, in the context of the IoT the definitions used are not sufficient because the IoT raw data are not "personal" on its face as it does not identify an individual. Only through combination and analytical methods can the identity of the individual as subject of data protection regulation be ascertained, which then could potentially submit the data collection to the EU DPD. As the collection by IoT devices is carried out in an automated manner, the risk of being non-compliant with these laws is inherent in their design. Nevertheless, IoT services' providers as well as consumers do not have a clear picture of the available legal provisions; such kind of normative uncertainty is detrimental to the business.

Therefore, in light of the vast technological developments over the last decade new rules are necessary for the IoT. Even if the IoT applications are quite different causing problems in the harmonization processes, the regulation of a global technology requires a worldwide approach in order to be most effective. In light of the difficulties associated with reaching an agreement on basic data protection and privacy issues, this solution is unlikely to be realized in the near future. Rather a more nuanced approach taking into account technological standards as enablers of data protection as well as national data protection regulations is the more likely scenario.

1.4. *First regulatory efforts by the EU*

The first supranational organization having dealt with the business and legal environment of the IoT, namely the European Commission, appointed a large group of experts to examine the relevant aspects of a possible IoT normative framework;[4] however, these activities have come to an end. Nevertheless, not only the expert reports are available but also the results of a public consultation that collected about six hundred responses to a broad questionnaire identifying IoT challenges.[5]

As far as privacy and data protection are concerned, the public consultation showed diverging results regarding the issues raised in the questionnaire. The industry was of the opinion that the current data protection framework would be sufficient, whereas a large majority of interested citizens and consumer organizations claimed that a greater focus on privacy and data protection in the context of the IoT would be needed. New instruments such as data protection impact assessments have been largely welcomed.[6] This reflects a common understanding that enterprises wish to expand their business operations whereas consumers still value their fundamental privacy rights and seek a choice as to what information enterprises can use and collect.

According to the public consultation, special emphasis must be placed on user consent as well as on the right of the

[3] For an early overview see R.H. Weber, Internet of things – New security and privacy challenges, CSLR 26 (2010), 23–30 causing the present title "revisited"; see also R.H. Weber/R. Weber, Internet of Things – Legal Perspectives, Zürich 2010.

[4] European Commission, Internet of Things, Reports, January 16, 2013, available at <https://ec.europa.eu/digital-agenda/en/news/conclusions-internet-things-public-consultation>.

[5] For an overview see also R.H. Weber, Internet of Things – Governance quo vadis? CLSR 29 (2013), 341, 342/43.

[6] Reports (supra note 4), 3.

users to delete data. Furthermore, since the possibility to build extensive personal profiles can be hardly avoided, data anonymization is important in the context of data sharing.[7] In addition, the transparency of data collections and the accountability of data collectors are also central points in the privacy discussion.[8]

Subsequent to the withdrawal of the European Commission from the political arena in the IoT field, mainly countries in East Asia (China, Japan) as well as the Federal Trade Commission (FTC) in the United States are continuing the discussions on IoT matters.[9] Obviously, as far as privacy issues are concerned, the different levels of data protection in the mentioned countries create challenges in coming to a common understanding. In view of the fact that Europe recognizes a relatively high data protection level it is important that privacy issues continue to be discussed in the IoT context.

1.5. *Key elements of the rule-making processes*

The IoT is a global network; in the future not only business will be affected by the IoT, but also civil society using devices and equipment connected to the IoT. Consequently, if regulators consider developing new rules, certain key elements related to the different IoT applications must be taken into account:[10]

(i) *Technology* (such as the RFID technology) must be "global" in the sense that the same technical processes are applied all over the world in order to ensure interoperability and security. This is the case as industry standards have emerged in particular in relation to passive RFID which is used for theft protection in stores.[11] However, only passive ultra high frequency (UHF) RFID is currently regulated by a single global standard. Business and trade would be substantively complicated if differing national laws would place varying obligations on its use. Therefore, regulatory efforts should be based on globality, as an important pillar of IoT regulation.

(ii) *Ubiquity* refers to the extension (scope) of the technological environment; the IoT rules, including data protection, privacy laws and technology standards must be designed to ubiquitously encompass persons, things, plants, and animals. This is necessary as the IoT can take many forms and impacts many spheres of human life. Such a regulatory framework should include privacy rules as well as general guidelines on the use of RFID technology.

(iii) *Verticality* means the potential durability of the technical environment, i.e. it is important for the life of the IoT that technical measures last long enough to not only enable its use in the supply chain until it reaches the final customer, but also for example in the waste management context. A major challenge is caused by the fact that the interaction between IoT generated data and customer generated data (even if unintentionally produced) can have an impact on the technical environment. In the first scenario the data are generated by electronic devices autonomously, based on the way they were designed and programmed. The second situation occurs when the user of an electronic device enters data knowingly, i.e. to improve the function of a device and its utility.

(iv) *Technicity* is an important basis for the development of rules protecting privacy objectives; several differentiations must be taken into account, namely (i) the complexity of the techniques (active and passive, rewritable, processing and sensor provided products), (ii) the complexity of background devices (reader or other linked media) and the maximum reading range which is designed to cover transparency demands.

1.6. *Regulatory agenda*

In view of the difficulty to develop a genuine legal environment for the IoT it is advisable to start any rule-making processes on the basis of the technological designs of IoT applications. Notwithstanding the fact that any regulation on an international level is highly ambitious and hard to achieve within the time frame warranted by the growth of the IoT, the normative framework to be established should achieve a global reach and be applicable to every device on earth from its becoming until its destruction, obviously depending on the concrete IoT application. The present lack of international rules requires the leadership of the industries producing and using these devices to self-regulate in order to avoid a large number of potentially varying (data protection) laws across states. Thus, it seems appropriate that the standard-setting by the industry itself should be encouraged as long as this model meets the demands of the market and offers the parties subjected to the IoT the choice as to the level of privacy protection they wish.

Based on this assessment, the following considerations start with a discussion of technological issues, in particular the prevailing devices and software requirements, the privacy enhancing technologies and privacy by design measures as data protection undertakings, the efforts for improving anonymity, and the technological innovations combating newly emerging risks. Thereafter, the main part of the article assesses the challenges for privacy regulations in the IoT environment by analyzing the types of privacy infringements as well as the transparency and data minimization requirements. In this context, special attention should be directed to the problems of data quality and data control as well as to interoperability and connectivity issues. Therefore, legal efforts need to be supported by appropriate industry standards.

2. Security and technology environment

2.1. *Devices and software requirements*

Cellular devices provide an access point and gateway for other lower technologies such as simple sensors to communicate their

[7] Reports (supra note 4), 4.

[8] Reports (supra note 4), 4.

[9] Center for Policy on Emerging Technologies, The Internet of Things: Roundtable with FTC Commissioner Brill, 2014, available at <http://www.ftc.gov/system/files/documents/public_statements/203011/140226cpetspeech.pdf>.

[10] See also Weber (supra note 3), 26/27.

[11] For a detailed overview of RFID standards see Impinj Inc., RFID Industry Standards, <http://www.impinj.com/resources/about-rfid/rfid-standards/>.

data to a network. Currently microchips are becoming cheaper to produce since IoT sensor prices are dropping below 50 cents per unit. The smaller devices such as RFID will be a main driver of growth in this area over the next decade.[12]

However, the storage space on simple passive RFID is extremely limited, thus an ongoing flow of information cannot be saved on an RFID tag. Instead of storage, a supply of the collected information via distributed servers on the Internet is made available through linking and cross-linking with the help of an Object Naming Service (ONS).[13] The ONS is authoritative (linking metadata and services) in the sense that the entity having (centralized) change control over the information about the Electronic Product Code (EPC) is the same entity that assigned the EPC to the concerned item.[14] The ONS is based on the well-known Domain Name System (DNS); but for this reason, the ONS also inherits all of the well-documented DNS weaknesses, such as the limited redundancy in practical implementations and the creation of single points of failure.[15]

Depending on the IoT device access by a third party would be possible at any point in this chain, starting at the device itself. However, as ongoing aggregated data are of real value the access point (Wi-Fi router or cellular device) is a main entry for hackers and other interested third parties. At this location all data come together before being compressed and sent to a cloud server for processing. The data are generally only encrypted or anonymized at a later stage on the cloud server; this happens after processing and having extracted the valuable information from the data.[16] In order to use various types of data new standards are now being designed by giving the research community open-access to the hardware specifications thus allowing to tailor the web applications to the data collected. In doing so the industry also enables criminal parties to profit from such information while at the same time available open source data can be used to identify criminals.[17] Therefore, a uniform security standard should be developed by the industry using IoT technology in order to ensure the safety of the data at every step from data collection to processing.

2.2. *Security and privacy requirements*

The described technical architecture of the IoT has an impact on the security and privacy of the involved stakeholders. Privacy includes the concealment of personal information as well as the treatment of the data. Not only the state is interested in collecting the respective data for public utility[18] and national security, but also private actors such as marketing enterprises.[19]

Since manifold processes are concerned, a high degree of reliability is needed. In the literature, the following security and privacy requirements are described that aim at achieving these goals:[20] (i) *Resilience to attacks:* The system has to avoid single points of failure and should adjust itself to node failures. (ii) *Data authentication:* As a principle, retrieved address and object information must be authenticated.[21] (iii) *Access control:* Information providers must be able to implement access control on the data provided. (iv) *Client privacy:* Measures need to be taken that only the information provider is able to infer from observing the use of the lookup system related to a specific customer; at least, inference should be very hard to conduct.

Private enterprises using IoT technology will need to include these requirements into their risk management concept governing the business activities in general in order to limit their exposure.[22]

2.3. *Privacy enhancing technologies and privacy by design measures*

A number of technologies have been developed in order to achieve information privacy goals; particularly the following privacy enhancing technologies (PET) are of importance in light of new technologies such as cloud computing and IoT:[23] (i) virtual private networks;[24] (ii) transport layer security;[25] (iii) DNS security extension;[26] (iv) onion routing;[27] and (v) private

[12] Radio Frequency Identification Regional Centre WM, The costs of an RFID implementation, available at <http://www.wmrfid.org/index.php/what-is-rfid/the-costs.html>.

[13] Weber/Weber (supra note 3), 6.

[14] EPC Global, Object Naming Service (ONS) Version 1.0.1, at para 4.2, available at <http://www.gs1.org/sites/default/files/docs/epc/ons_1_0_1-standard-20080529.pdf>.

[15] Weber/Weber (supra note 3), 6/7.

[16] Q. Xiao/T. Gibbons/H. Lebrun, RFID Technology, Security Vulnerabilities, and Countermeasures, 366, available at <http://cdn.intechopen.com/pdfs-wm/6177.pdf> (KAP).

[17] Wynzard Group, Using Wynyard Advanced Crime Analytics On Open Source Data, available at <https://www.wynyardgroup.com/us/news-events-blog/using-wynyard-advanced-crime-analytics-on-open-source-data/>.

[18] For example, US Energy Information Administration, How many smart meters are installed in the U.S. and who has them?, available at <http://www.eia.gov/tools/faqs/faq.cfm?id=108&t=1>.

[19] S. Lueng, 5 Ways the Internet of Things Will Make Marketing Smarter, Salesforce Blog 20. March 2014, available at <http://blogs.salesforce.com/company/2014/03/internet-of-things-marketing-impact.html>.

[20] See B. Fabian/O. Gunther, Distributed ONS and its Impact on Privacy, 1223, 1225, available at <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber¼04288878>.

[21] R.H. Weber/A. Willi, IT-Sicherheit und Recht, Zürich 2006, 284.

[22] See also E. Grummt/M. Müller, Fine-Grained Access Control for EPC Information Services, in: Floerkemeier/Langheinrich/Fleisch/Mattern/Sarma, The Internet of Things, Springer, Berlin (2008), at 35–49.

[23] For further details see Weber (supra note 3), 24/25.

[24] VPN allows the transfer of data through a secured line between an endpoint and the VPN server. The user essentially connects to the internet through that server which is protected by sophisticated firewalls and other protective measures enabling the user to benefit from this infrastructure.

[25] TLS is a cryptographic protocol that allows for a secure communication over a network.

[26] Domain Name System Security Extensions (DNSSEC) provide DNS clients with authentication of DNS data and ensure data integrity.

[27] Onion routing aims at masquerading ones identity through the use of a vast amount of servers. However, it appears that the FBI has developed capabilities to identify individuals using the TOR system.

information retrieval.[28] These technologies aim at ensuring the security of communications as well as the preservation of the identity of a user in instances when such information is not required by another party. They can be seen as methods which all in their own way play an important part in increasing the privacy of users and the data transmitted.

In addition, Privacy by Design (PbD) also forms a cornerstone of future privacy protection. PbD requires the adherence to seven basic principles including a proactive approach to protection measures, privacy as default setting, privacy embedded into the design of the technology, full functionality, end-to-end security spanning the life-cycle of the device, visibility and transparency allowing stakeholders to verify the privacy claims made as well as respect for user privacy.[29] Under the newly proposed EU Data Protection Regulation this approach has been taken up in Article 23 requiring the controller and processor to implement appropriate and proportional organizational measures in order to protect the rights of the data subject. It was also highlighted that this requirement spans the entire lifecycle of the data and must be considered at the time of determining the means of processing as well as the processing itself.[30]

The IoT heavily relies on PET for its further expansion as the technology, in its current state, creates a wide array of privacy risks for individuals who are under constant surveillance either directly or indirectly by these devices. PET must bridge the gap between the simple technology employed in IoT devices and the privacy needs of the users subject to a device's monitoring capabilities. Furthermore, the mentioned PbD features are to be implemented at the design stage of the IoT device: (i) The data communicated by these devices need to be secured. (ii) The transmission must be anonymized so as to obscure the source as well as the type of data. (iii) The storage of the information collected by the devices on a central server and the data analytics (i.e. Big Data) or other behavioral pattern deriving technologies must be limited in order to prevent identification of individuals in cases in which this is not essential to the provisioning of the service.

Any such information as to the data usage must be unambiguously communicated in order to allow for informed consent. The manner in which the data can be used varies between jurisdictions; however, within the EU as far as personal data are concerned, only a few exceptions are available such as the mentioned informed consent or the requirement to fulfill a contractual obligation entered into by the data subject. In the U.S. limitations only apply to certain types of data such as medical information (HIPAA).

Without clear limitations on privacy infringements through appropriate PET measures the IoT will be inhibited in its expansion as various laws such as the EU data protection framework and other sector-specific laws (i.e. U.S. HIPAA) restrict the collection of personal data, unless strict requirements as to the data subject's consent or other legal justification are present as well as appropriate security measures are taken.

2.4. *From confidentiality to anonymity*

Confidentiality is usually provided in some ways in existing privacy technologies as the first objective of privacy is to protect personal context data from being accessed by unauthorized persons. If personal data become public, confidentiality and hence privacy is lost. Privacy as confidentiality represents solutions for anonymizing the collected data (including communications) and minimizing the collection of data.

Anonymity[31] of data relies on cryptographic solutions in order to achieve properties like (i) unlinkability (two information items or two actions of the same user cannot be related), (ii) undetectability (an attacker is not able to distinguish whether an information item exists), (iii) unobservability (it is not possible to detect whether a system is being visited by a given user), and (iv) communications content confidentiality.

A new model in respect of disclosing information to a third party which protects anonymity of the persons from which the data were collected while still being of value to the third party is called k-anonymity; it aims at reducing the risk of re-identification by linking data sets. In order to achieve this objective, previously the disclosing party had to ensure that there were no outside data available leading to the identification of individuals when combined. Contrary to this approach being highly resource intensive the new k-anonymity addresses the problem of directly matching externally available data and claims that an individual cannot be identified within a set of k users. Thus, a release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release.[32] In practice the data have to be structured as otherwise this method is not able to ascertain the variables which it needs to protect from resulting in the identification of an individual.

However, this approach although a valuable contribution to privacy protection is susceptible to background knowledge and homogeneity attacks.[33] Thus, a further refined variant has been proposed, namely the L-diversity; a block of data is L-diverse if it contains at least L well represented values for the sensitive attribute S.[34] However, the first versions of this method were only effective where the release of data was static and not ongoing as this would open the door for potential

[28] Private information retrieval (PIR) allows a user to retrieve information through a protocol from a server without revealing what item is retrieved.

[29] Privacy by Design, 7 Foundational Principles <https://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/>.

[30] Draft EU Data Protection Regulation (as of 7.10.2013) and R.H. Weber, Privacy management practices in the proposed EU regulation, International Data Privacy Law, 2014, Vol. 4, No. 4.

[31] For a general overview see R.H. Weber/U. Heinrich, Anonymization, London (2012), passim.

[32] L. Sweeney, k-anonymity: a model for protection privacy, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10 (2002), 557–570.

[33] A. Machanavajjhala/J. Gehrke/D. Kifer, l-Diversity: Privacy beyond k-anonymity, Proceedings of the 22nd International Conference on Data Engineering. Los Alamitos: IEEE Computer Society, 24–35 (2006).

[34] P. Wang/J. Wang, L-diversity Algorithm for Incremental Data Release, Applied Mathematics & Information Services 7 (2013), 2055–2060.

identification. Due to the increased research in this field, the L-diversity method is applicable to incremental data disclosure.[35]

Other anonymity metrics have also been developed. However, what degree of anonymity is sufficient for a particular scenario is dependent on the debatable legal and social consequences of a data breach. In this context, differential privacy aims at providing means to maximize the accuracy of queries from statistical databases while minimizing the chances of identifying its records.[36]

Anonymity in communication has the objective of protecting traffic data from concealing who talks to whom. Even if the content of a communication is kept confidential, sensitive information may be leaked by traffic data that include locations and identities of the communicating parties, time, frequency and the volume of communication. Providing anonymous communication is challenging since many communication protocols use unique identifiers.[37]

2.5. *Privacy laws confronted with technological innovations*

Various technological innovations have enabled the growth of the IoT and are now part of the privacy challenges that have arisen over the past decade. These technologies include (i) location-based services, (ii) sensor networks, (iii) delay tolerant networks, and (iv) privacy-friendly smart grids.

These technologies are a source of potential privacy risks but could also be part of the solution if they are designed and applied in a way that conforms with privacy standards. In particular, the data collection devices can be designed to include basic privacy protection features from the start. New so-called G2 RFID technology allows the user to hide part or all of the tags memory and the ability to read or alter the data varies depending on the proximity to the tag allowing for more control and reduction of data tampering or theft.

Nevertheless, with more and more devices being used in daily life new risk are created. These emerging risks encompass (i) automatically generated data that are not necessary for service provisioning but its collection could potentially have severe privacy implication, (ii) data scattered across large distributed systems lead to a loss in control, and (iii) de-anonymization through linking data collected lead to an ever growing amount of devices.

The quality of the data collected is increasing with any additional information such as locations and environments (i.e. temperature/lights/sounds) leading to a higher value of the data for interested parties. With every further data set the validity and accuracy of another data set can be verified. For example if your iPhone location appears to be in Puerto Rico but the temperature measured by your FitBit on your arm wrist states –10 degrees one of the values must be wrong. Thus, potential device errors can be identified with every further information set that becomes available. Additionally, the amount of data created is rising exponentially enabling a clearer picture of the individual who is observed by these devices directly or indirectly based on i.e. the electricity use which can be attributed to individual devices.[38]

The aggregated data from various IoT devices add up to a total surveillance. For example the house sensors will know when to start the coffee machine and to pull up the blinds, thus the time when somebody gets up is known to the data collectors. The amount of coffee as well as the used products from the fridge will determine the number of people residing in the house on that day including their eating habits. The car will then communicate the driven route through its GPS system and the onboard entertainment will know the driver's favorite music. The automated seatbelt warning system will know how many people are in the car. Mobile phones with their tracking and recording capabilities can ascertain a person's whole day through the data collected by IoT devices. These collected data are in most cases not encrypted as its face value as singular data are very low. However, the risk lies in the combination of the automatically collected data from various sources into one database. Without privacy technologies such as automatic anonymization by replacing the unique identifier already at the earliest possible point the potential privacy infringements are steadily growing with the expansion of technology.

Automated processes must be implemented in IoT devices to ensure privacy by encrypting and anonymizing data such as location data that can be attributed to an individual person. Furthermore, the risk of reversing the anonymization through advanced technologies must be limited by increasing the security measures surrounding the data collection as well as implementing clear regulation as to the use, distribution and sale of such data.

Deletion rights and automatic data deletion is also an important aspect of ensuring privacy as the amount of data collected is growing exponentially and saved in various scattered databases thus an individual should retain the right to have the data deleted after a certain period or upon request.

Probably over 200 billion sensors will be collecting data in 2020.[39] In light of these estimates a need for further regulatory action is needed, in particular since the current frameworks for data protection are inadequate to address the issues raised by the IoT. For example, the EU Data Protection Directive still has a very old definition of personal data that equates "identifiable" with "identified" and thus results in an expansive view of the definition of personal data.[40] From a legislative perspective, this expansive interpretation will enable a broad application

[35] Ibid, 2059.

[36] C. Dwork, A Firm Foundation for Private Data Analysis, Communications of the ACM, Vol. 54 No. 1, 86–95, available at <http://cacm.acm.org/magazines/2011/1/103226-a-firm-foundation-for-private-data-analysis/fulltext>.

[37] ICANN, Beginner's Guide to Internet Protocol (IP) Addresses, March 2011, available at <https://www.icann.org/en/system/files/files/ip-addresses-beginners-guide-04mar11-en.pdf>.

[38] S. Wicker/R. Thomas, A Privacy-Aware Architecture For Demand Response Systems, Cornell University, Proceedings of the 44th Hawaiian Conference on System Science (HICSS-44), Kauai, Hawaii, January 2011, available at <wisl.ece.cornell.edu/wicker/SWicker_RThomas_HICSS.pdf>.

[39] T. Bajarin, The Next Big Thing for Tech: The Internet of Everything, Time 13 January 2014, available at <time.com/539/the-next-big-thing-for-tech-the-internet-of-everything/>.

[40] P. Schwartz/D. Solove, "The PII Problem: Privacy and a New Concept of Personally Identifiable Information", New York University Law Review, vol. 86, ed. 6, 2011, 1814–1894, 1819.

of data protection laws to many forms of technology.[41] However, the impression prevails that the legislator is not aware of the nature of the IoT and the further technologies currently created. A rather nuanced approach seems necessary at least in respect of automatically created data as essentially the reality is such that everyone is identifiable without some sort of protection such as the above discussed anonymity approaches. Furthermore, IoT devices record data no matter what individual is affected thus leading to the risk of an unaware party being subjected to IoT surveillance.

3. Challenges for privacy regulations in the IoT

Based on the acknowledgment that the technological developments in relation to the IoT lead to substantial privacy risks and that the legal stability, founded on an appropriate regulatory agenda, must be increased, the interim conclusion has been drawn that an improvement of the data security environment is unavoidable. As mentioned, privacy enhancing technologies and privacy by design measures can be applied in order to increase the level of confidentiality and anonymity. Nevertheless, the tensions between technological innovations and traditional privacy laws will not diminish; therefore, the following last chapter attempts to give guidance in designing a legal privacy framework that tackles the queries of forward-looking data protection principles and also takes the important elements of data quality and context quality into account.

3.1. *Determination of the relevant types of privacy infringements*

Looking from the perspective of a normative framework, in the IoT context privacy can be infringed at various stages:[42] the first stage is access by third parties to the collected data, the second is the use and distribution of data by the data collector and the third is the risk of data being combined with other data. Especially the third possibility is often not known by the party who is using the IoT devices supplying the data. In combination with other data sets new information about a person or situation can be generated which are of high commercial value for various data hungry enterprises and marketing firms.

Context information is any information that describes the quality of information. This includes information surrounding the collection of the primary data such as the status and attributes of the data collecting device. A large amount of context information may have to be collected and then aggregated or fusioned, allowing the inference of originally unforeseen context information, for example the power usage of a household, and linking this to travel patterns from mobile phone location data. This is not compliant with the data minimization principle of privacy law requiring the limitation of data collection to the furthest extent possible; in fact this may even allow deriving the user's new information that was unknown or hidden so far, thereby increasing the risks of privacy violation. Low quality data may also allow to inferring false context information with potentially serious repercussions to the user's privacy.

As the data are assumed to be created by the automated devices themselves a perceived higher level of trust is placed on them than on manually entered human data. This is of major concern as for example insurance companies are moving to monitor (i.e. the driving behavior of their customers) via such devices to ascertain the specific risks in accordance with personal characteristics and conduct. Today the customer must consent to and physically install the IoT monitoring device himself. But, in the future such data could be aggregated through the car system as well as through mobile devices carried on a person. The devices are also able to record GPS data and thus the speed a person is traveling. Adverse judgments affecting a person based on such data must be prevented through appropriate device safety measures as well as legislative limitations on data usage.

However, first awareness should be created in society as to the many privacy implications these devices can have on an individual. In particular, the real life consequences and their effect in light of human rights must be addressed. Questions such as to what extent a society wants to be controlled by devices must be adequately resolved. As explained, privacy enhancing technologies and privacy by design can lead the way in ensuring that the boundaries of the IoT are maintained.

3.2. *Quality of data and quality of context*

The quality of data and the quality of context are not yet sufficiently discussed issues even if these phenomena play an increasingly important role in the privacy debates related to the IoT. Quality of the data is to be insured by taking the environment in which it is collected into account. A context attribute may be unknown when there is no information about it, leading to incomplete context information. It may also be ambiguous as there is a risk of having contradictory information from different context sources. An attribute is imprecise when the reported information is correct but not provided with a sufficient degree or precision. Apart from precision, important factors in this regard are probability of correctness, resolution, trustworthiness and up-to-dateness of the data collected.

Quality of context (QoC) raises new issues of confidentiality that are not yet addressed by current research. QoC refers to information and not to the process neither the hardware components that possibly provide the information. A context attribute may be unknown when there is no information available about it, leading to incomplete context information and wrong interpretations. It may also be ambiguous as there is a risk of having contradictory information from different context sources. An attribute is imprecise when the reported information is correct but not provided with a sufficient degree or precision. As context data are by nature

[41] L. J. Sotto, Privacy and Data Security Law Deskbook, Aspen Publishers, New York, 2010, Section 18.02[A].

[42] D. Miorandi/S. Sicari/F. De Pellegrini/I. Chlamtac, Internet of things: Vision, applications and research challenges, Ad Hoc Networks 10 (2012), 1505, available at <http://www.sciencedirect.com/science/article/pii/S1570870512000674#>.

dynamic and very heterogeneous, they also tend to be erroneous and not exactly reflecting the real state of the modeled entity. Therefore, one solution that has been used for a decade is to attach metadata to context information representing its quality.

The following examples illustrate the confidentiality issues raised by QoC:[43]

(i) John has a mobile phone equipped with positioning technology being interested to share his location that is encoded using the Google address component type format. John defines a policy to provide only the region where he is, which consists in sharing the attribute of the administrative area level 2 type according to the Google format. By using this mechanism, John believes that nobody will be able to track him. However, when he is at the border of three regions R1, R2 and R3, within a short period of time (10 min for instance), the location data history will include the three different values R1, R2 and R3. A third-party system may then deduce that John is at the crossing border of the three regions. As a consequence, the actual level of detail is then much more precise than the one expected by John.[44]

(ii) A recent study has proven that in an anonymized dataset "where location of an individual is recorded hourly and with a spatial resolution equal to that given by carrier antenna, four spatio-temporal points are enough to uniquely identify 95% of the individuals". The spatial resolution of antenna goes from 0.15 to 15 km^2. The fact is highlighted that "a point on the MIT campus at 3AM is more likely to make a trace unique that a point in downtown Boston on Friday evening".[45]

These examples show that even low quality context data are useful for data mining algorithms and reliable QoC information will improve the efficiency of such algorithms. QoC should be qualified as sensitive information because its value and the potential effects it can have: The first step in preparing a security attack on networked systems is the "reconnaissance phase" which objective is to collect information about the target system in order to detect possible known vulnerabilities. For example, TCP/IP stack fingerprinting consists in collecting configuration values (e.g. the initial TTL and window size fields) from a remote device during standard network communications. The combination of parameters values may then be used to infer what the remote machine's operating system is because different operating systems, and different versions of the same operating system, set different default values for these parameters. QoC might ease system fingerprinting if different systems set varying default QoC values. As a result, QoC is sensitive information that must be protected too. QoC change is also sensitive information: Context-aware computing in the IoT does not allow people to have the power to switch off the system or to easily disconnect from it. Some researchers have proposed to use the concept of white lie to provide people with this capability.[46]

QoC will also make white lying much more complex to perform, as the following example shows:

Mary is a teenager who provides her location to her parents with a high degree of detail. On Friday evening Mary tells her parents that she is going to visit her grandmother. Actually, she is lying and wants to see her friends who live near her grandmother's house. Thus, she uses the obfuscation mechanism that changes the granularity level of her location information. However, when her parents notice that the granularity level has changed, they can deduce that their daughter lied to them by using algorithms for detecting changes.[47] Therefore, obfuscation mechanisms must consider that white lies cannot be used if QoC information is reliable.

3.3. *Identification of transparency and data minimization requirements*

Transparency tools intend to improve the users' understanding and control of their data profile. Four characteristics that such tools should possess appear to be important:

- provide information about the intended collection, storage and/or data processing;
- provide an overview of what personal data have been disclosed to what data controller under which policies;
- provide online access to the personal data and how they have been processed;
- provide counter profiling capabilities helping the user to anticipate how their data match relevant group profiles, which may affect future opportunities or risks.

Transparency was also acknowledged as an important element of privacy in the Mauritius Declaration on the Internet of Things, approved by the Data Protection and Privacy Commissioners of more than 100 countries on 14 October 2014: "*Transparency is key: those who offer Internet of Things devices should be clear about what data they collect, for what purposes and how long this data is retained*".[48]

Privacy as transparency is an important issue because most PET are useless if people cannot use them efficiently or if they are not implemented in an automated fashion. Privacy as transparency is even more critical for the next IoT-based distributed systems than it is in the existing web based ubiquitous applications. The users (i.e. context data owners)

[43] S. Machara Marquez/S. Chabridon/C. Taconet (2013) models@run.time for privacy and quality of context level agreements in the Internet of Things. Technical report, UMR SAMOVAR, Télécom SudParis.

[44] Google Inc., The Google geocoding API, 2013, available at <https://developers.google.com/maps/documentation/geocoding/>.

[45] Y. de Montjoye/C. A. Hidalgo/M. Verleysen/V. D. Blondel, Unique in the crowd: the privacy bounds of human mobility, 2013 Nat Sci Rep 3:1376.

[46] S. A. Bagüés/A. Zeidler/C. F. Valdivielso/I. R. Matias, Disappearing for a while-using white lies in pervasive computing. In: Proceedings of the ACM workshop on privacy in electronic society, ACM 2007, 80–83.

[47] M. Basseville/I. V. Nikiforov, Detection of abrupt changes: theory and application, vol. 104. Prentice-Hall, Englewood Cliffs 2007.

[48] Mauritius Declaration on the Internet of Things, 36th International Conference of Data Protection and Privacy Commissioners, October 2014, 2, available at <http://www.privacyconference2014.org/media/16596/Mauritius-Declaration.pdf>.

will not only have to control the personal data that can be propagated from the terminals with which they directly interact (smartphone, laptop), but they will also have to handle the control of the data automatically produced by the connected devices they own. These data can surround them or are located in their living environments (home, office, etc.); they could be scattered across a large distributed system while facing issues such as heterogeneousness and scalability. Despite the importance of this issue, limited research has been undertaken in this regard. Based on the international nature of the data collection and use a purely regulatory answer does not seem to be feasible. Rather a combined approach including technical standards set by the industry as well as general principles and independent certification seems to be the most viable option.

Data minimization aims to limit the collection and processing of personal data, its implementation can be done by encrypted aggregation techniques. Other approaches include perturbation and obfuscation: (i) Perturbation means that data get systematically altered using a perturbation function (e.g. adding random numbers). (ii) Obfuscation means that a certain percentage of data is replaced by random values (e.g. replacing with the mean).

Data minimization was developed as a major principle of data protection laws, mainly in Europe. This principle means that data not being relevant anymore for the purpose at the time of collection are not to be stored anymore. In the context of the IoT, data minimization must be balanced against the demands of civil society and businesses for more and more functionality. Once data are altered the question is to what extent it can still be used and what possibilities there exist for putting it back into its original state.

3.4. *Acknowledgment of interoperability and connectivity*

In the year 2014 the Open Interconnect Consortium was created; the new industry consortium focused on improving technological interoperability and defining the connectivity requirements for the billions of devices that will make up the IoT. The Open Interconnect Consortium (OIC) is focused on defining a common communications framework based on industry standard technologies to wirelessly connect and intelligently manage the flow of information among personal computing and emerging IoT devices, regardless of form factor, operating system or service provider.

According to a press release of 8 July 2014, the OIC has assembled leaders from a broad range of industry vertical segments (from smart home and office solutions to automotive and more). They participate in the program and ensure that OIC specifications and open source implementations support companies in the design of products that intelligently, reliably and securely manage and exchange information under changing conditions, power and bandwidth, even in case of lack of an Internet connection.[49] This group, founded in 2014, is also attempting to create a normative framework for the Internet of Things despite the growing use of IoT devices.

What are the implications of IoT devices on the daily life of the individual? Take as example the announced Apple Watch. This is a device will know (1) who you are, (2) where you are via GPS, (3) what you are doing via accelerometer and gyroscope, (4) your health, and (5) even be able to monitor your mood. While there are security measures built into these devices, the ramifications could be significant if there is a failure. In light of growing concerns, Apple has changed its security infrastructure and now encrypts its cloud communication. This allows for safer data transfers but limits the ability to access the device if the user loses his password.[50]

4. Outlook

The main challenge for privacy in the context of IoT remains the management of the vast amount of data collected. Necessary technologies that ensure secure communication and storage of this data are currently being designed and implemented. However, the actual use and possible privacy implications of the IoT data remain largely unaddressed. In order to maintain a minimum level of privacy, industry standards must be created which limit the use and collection of data relating to sensible information such as health, religion or sexual orientation, etc. (being considered to be at the heart of personal privacy) and which are observed in many countries. Information relating to these matters collected by IoT devices must not be stored or processed in any form without the express consent of the individual concerned and appropriate measures should be taken to ensure that the data collected are not those of an unrelated individual.

The solution to the mentioned IoT challenges can only be tackled by a coordinated approach including a clear regulatory framework as well as technical measures which enhance the privacy of the collected data. First steps have been addressed from a technical side by the industry in setting international standards such as for the use of UHF RFID. However, on a regulatory level appropriate laws are still missing. Most of the laws are only concerned with basic data protection issues and do not address the particular and complex requirements of the IoT.

In this regard, action seems particularly necessary in light of research being able to predict the location of a person with 80% accuracy up to 80 weeks into the future.[51] The more data an enterprise can acquire the more it knows about its customers and the better it can market its products. This is the reason why Google recently paid a very high price to buy Nest,

[49] Broadcom, Industry Leaders to Establish Open Interconnect Consortium to Advance Interoperability for Internet of Things, Press Release 9 July 2014, available at <http://www.broadcom.com/press/release.php?id=s858114>.

[50] Apple Inc., iCloud security and privacy overview, available at <http://support.apple.com/en-us/HT202303>.

[51] P. Tucker, Has Big Data Made Anonymity Impossible?, MIT Technology Review, 7 May 2013, available at <technologyreview.com/news/514351/has-big-data-made-anonymity-impossible/>.

a company that produces thermostats and other home monitoring devices.[52] This example demonstrates the movement of private enterprises into the private sphere of their customers by way of IoT devices; furthermore, it underlines the necessity to implement, as discussed, further privacy measures in order to regain control of one's own data. Apart from the technological measures, mainly the general data protection principles are to be made fruitful for application in the IoT context, for example the improved transparency and data minimization principle.

[52] M. Wohlsen, What Google Really Gets Out of Buying Nest for $3.2 Billion, Wired, 14 January 2014, available at <http://www.wired.com/2014/01/googles-3-billion-nest-buy-finally-make-internet-things-real-us/>.