From Privacy Impact Assessment to Social Impact Assessment

Lilian Edwards
Department of Law
University of Strathclyde
Glasgow, UK

Derek McAuley
School of Computer Science
University of Nottingham
Nottingham, UK

Laurence Diver
School of Law
University of Edinburgh
Edinburgh, UK

Abstract—In order to address the continued decline in consumer trust in all things digital, and specifically the Internet of Things (IoT), we propose a radical overhaul of IoT design processes. Privacy by Design has been proposed as a suitable framework, but we argue the current approach has two failings: it presents too abstract a framework to inform design; and it is often applied after many critical design decisions have been made in defining the business opportunity. To rebuild trust we need the philosophy of Privacy by Design to be transformed into a wider Social Impact Assessment and delivered with practical guidance to be applied at product/service concept stage as well as throughout the system’s engineering.

Keywords—Privacy Law; Privacy by Design; Internet of Things; Ubiquitous Computing; Privacy Impact Assessment; Social Impact Assessment

I. INTRODUCTION

Consumer trust in the Internet of Things (IoT) is at a precarious moment. Emerging from hardware designed to meet retail and engineering needs in B2B rather than B2C environments, in order for IoT applications to be cheap, unobtrusive and mass-produced they have historically paid little attention to values like privacy, security and usability. Now in the era of smart homes, roads, transport, and wearables, this approach is coming home to roost. In terms of device security, the past eighteen months have seen increasingly serious proof-of-concept attacks on IoT devices [1], for example recently reported vulnerabilities in keyless car access [2]. In terms of privacy, stories such as the hack-able baby monitor [3], the Samsung smart TV that listens to your conversations [4] and the search engine for private video streams made publically accessible through poor webcam security [5] have had considerable impact on public confidence. These combine with general widespread concerns about collection and use of personal data in existing e-commerce and social networking consumer contexts, which have largely not been helped in Europe by the halting, compromised progress of the new EU General Data Protection Regulation (GDPR) and the Snowden revelations [5]. Within this EU context the recent Walport Report from the UK asserts that ‘public acceptability and trust are central to the implementation of IoT’ [6]. The FTC in the US expressed a similar view in their 2015 report [7].

We argue that it is crucial to create embedded social, legal and technical processes which ensure either (i) that users are given functional notice and meaningful control over what data they share via the IoT and how it is used, or (ii) guarantee that toxic uses of data will be prohibited. EU data protection (DP) laws, while a “gold standard” for privacy, are widely regarded as inadequate, unenforceable and overly bureaucratic. In particular, prior informed consent, the central pillar of EU data protection, is extremely difficult to implement in ubiquitous pervasive environments as currently implemented. The collection of private data in public places by “smart” systems is also a key problem for existing laws, especially in the US, which tie privacy protection to private places. We conclude that law needs to be supplemented, implemented and often exceeded by “code”, by means of privacy by design methods created specifically for the IoT environment.

II. LEGAL BACKGROUND

There is a growing literature on the potential threat the IoT poses to privacy and increasing public awareness of the IoT as a tool for pervasive surveillance [8]. To give a flavour, in 2015 The Guardian opined:

“We may find ourselves interacting with thousands of little objects around us on a daily basis, each collecting seemingly innocuous bits of data 24/7, information these things will report to the cloud, where it will be processed, correlated, and reviewed. Your smart watch will reveal your lack of exercise to your health insurance company, your car will tell your insurer of your frequent speeding, and your dustbin will tell your local council that you are not following local recycling regulations. This is the ‘internet of stool pigeons’, and though it may sound far-fetched, it’s already happening”[10]

For privacy purposes the key problem of the IoT is that its devices are explicitly designed to be seamless and as unobtrusive to the user experience as possible; as Weiser puts it, such devices aim to weave themselves “into the fabric of daily life until they are indistinguishable from it”.[11] Traditional privacy regulatory systems including both European data protection law and the US’s Fair Information Principles (FIPs) are historically dependent on notions of “notice and choice”: instances of collection and use of data are presented to users (“data subjects” in EU parlance) who are then given the “choice” to accept or reject these. In EU law, this notion of user consent and autonomy is one of the most important, though not exclusive, grounds legitimating processing of personal data (Data Protection Directive (DPD), Article 7). Even where unobtrusiveness is not a functional specification, IoT devices simply do not usually have means to display privacy notices and/or to “provide fine-tuned consent in line with the preferences expressed by individuals”, as devices are often small, screen-less and lack an effective input
mechanism (e.g. a keyboard or a touch screen) [12]. In private or domestic environments, the problem may be to an extent be solved by the theoretical opportunity consumers have to read the privacy policy of, for example, their Nest thermostat or smart fridge — although often not before purchasing the product, thus bringing to mind the problems around shrink wrap software licensing. Nevertheless such consent, based on terms which are mainly unread and, even if they are, largely not comprehended nor offered with a viable alternative, does not meet the requirement of DP law that consent be “freely given, specific and informed” (Article 2(h)). The problem is compounded in public IoT deployments, for example smart transport systems, because there is usually no opportunity to provide consent, and no choice between alternative providers.

Although this lack of opportunity to provide meaningful consent in IoT environments has different consequences in the EU and the US, it is likely to reduce trust in both. In the EU, it may mean a shift to non-consent grounds for validation of data collection/processing in the IoT, notably, the ground that such processing is in the “legitimate interests” of the data collector and does not significantly impact on the rights of the data subject (Article 7(0), DPD). (Interestingly, this ground is well known to marketing companies but apparently little known to computer science researchers whose focus tends to be wholly on consent, though their concern may be more ethics than law.) “Legitimate interests” is a legal ground lacking transparency to the public and is easily abused in the absence of effective enforcement. In the US privacy protection for collection of location, especially in public places, is minimal and dependent not on consumer rights but on Fourth Amendment protections against search without warrant which are difficult to import to the IoT context [13]. It can be seen that in both legal systems consumer trust is likely to fray.

The IoT is not the only major current problem for privacy law. Many developments, notably the Snowden leaks which exposed mass covert surveillance of electronic communications by security agencies, the rise of data-mining, profiling and targeted data-led discrimination and the inability of consumers to control what happens to their data in the Cloud have all led to a general collapse in faith in the law’s ability to regulate sensibly the transnational flow, monetisation and surveillance of personal data. In Europe the long-delayed GDPR, which reached a (hopefully) final compromise text in January 2016, attempts, largely unconvincingly, to find new solutions to graft on to the existing basic framework of DP law. Although some new remedies such as the “right to be forgotten” and increased fines for breaches may improve enforcement, nothing in the substantive law has really changed to improve privacy in the IoT (though there may still be room for such in the upcoming review of the e-Privacy Directive). Meanwhile the issue of data flows from the EU to the US via the backdoor of “safe harbour” remains unresolved and the proposed “PrivacyShield” compromise will very likely find itself under renewed challenge in European courts [14].

Accordingly our view is that to place user trust in the IoT on a firm footing for both commercial and societal benefit we need to look beyond law to standards which are: (a) higher than legal minima negotiated between states with widely differing privacy and business cultures; (b) implemented and thus enforced by “code” which includes software, hardware and industry codes of practice; (c) created in meaningful dialogue between industry and consumers with a real sense of what data subjects actually want; and (d) achieve global recognition via incorporation into international technology standards (c.f. the use of ISO 27001 in data security). We turn thus to what progress has been made in “privacy by design” (PbD) and in particular to privacy impact assessments (PIAs), both of which will from about 2018 be mandated in the EU by the GDPR, albeit in extremely vague terms.

III. RELATED WORK IN PbD AND PIAS

Privacy by Design is an approach to systems design that takes privacy into account throughout the engineering process. It adopts a risk-based approach that identifies the impact that the project might have on the privacy of individuals and examines how both human and computer processes can be designed to mitigate such risks. The approach will be required under the GDPR for all future developments that will process personal data and are “likely to result in a high risk for the rights and freedoms of individuals”. Recital 61 of the pre- prologue draft of the GDPR asserted that: “the principle of data protection by design requires that data protection be embedded within the entire life cycle of the technology, from the very early stage, right through to its ultimate deployment, use and final disposal”. A key element of PbD as mandated by the GDPR is PIAs, which have already begun to be applied to IoT systems [15] and have previously been mandated for government projects in many countries, including the UK and USA, so there exists a broad body of knowledge in industry concerning the approach (although relatively little prior art in the private sector). The UK data protection regulator, the Information Commissioner, defines a PIA as “a process which assists organisations in identifying and minimising the privacy risks of new projects or policies.” [16] “Privacy by Default” also becomes a legislative principle so that, by default, a product or system’s settings should be set such that it processes only personal data that are necessary for a specific purpose. A major impact here will be on the current use of tracking cookies by online marketing companies, whose revenues support many of the major free online platforms.

PbD has its genesis in work done in the 1990s by the then-Information and Privacy Commissioner for Ontario, Ann Cavoukian. Based around seven core principles [17], the concept aims to promote the idea of privacy being “good for business” on the basis that increased user trust is commercially beneficial [18]. Although perhaps useful in policy making, the principles (and the GDPR provisions) have been criticized for providing insufficient concrete guidance on how to implement privacy within the designs of digital systems. Commentators have suggested they are vague and recursive [19], that they “do not address technology producers [or] allow real technology design” [20] and that they provide systems designers with “little clue on how they should go about designing in’ privacy” [21]. One paper even asks, owing to the lack of technical underpinnings in the PbD literature, why “the word ‘design’ was included in the first place” [19].
With EU institutions favouring market-driven privacy technologies, it is arguable that such ambiguity is unacceptable first because the GDPR will have direct effect and secondly because only those enterprises with significant legal and financial resources will be willing to take the risk of speculatively interpreting the legislation, developing PbD technologies, then subsequently operating in areas of the market where personal data are processed. This underlines the need for standardization of PbD technologies in order that smaller entrants are not frozen out of the market, with all the chilling of innovation that this would entail.

It has been argued that traditional legal approaches are too reductive of the social interactions from which legal activity flows to be of practical use. The law does not, and cannot, operate in a logical vacuum; it is necessarily part of a broader social context [22]. Despite such arguments and the sceptical treatment of PbD in some of the literature, research over the past decade has developed practical methods for building in compliance at the design stage of the product lifecycle. This work suggests that some of these problems may be ameliorated by dealing with representations of the law (and of the system to be regulated) in their simplest form, before they are “contaminated” by the vicissitudes of real-world evidence. If a system’s inherent design can be made compliant according to an abstracted set of basic principles, we can avoid the difficulties involved in attempting to perform legal and evidential reasoning “in the field”. This is preferable to bolting on privacy enhancing technologies (PETs) after-the-fact: instead of adding encryption, anonymization or improved security as an afterthought, the values (and specific requirements) represented in the regulation can be embodied in the design of the system from the off, and its behaviour can be adapted to complement, rather than merely tolerate, such technologies.

Over the past decade, logicians have attempted to bridge the gap between internal business rules and the external regulatory framework using logical abstractions as a means of testing compliance [23, 24], going so far as to create formal (computational) languages precisely for this purpose [25]. Other research focuses on ontologies as a means of representing abstractions of both the relevant legal and digital system, in order to enable near-direct comparison of their constituent entities and the relationships between them [26]. Lastly, some very recent work has been done in the area of private medical records sharing, using Business Process Modelling and Notation (BPMN) as a means of formalization against which regulatory norms can be applied [27]. While these approaches are unlikely to replace the bespoke advice of an experienced human lawyer any time soon, arguably they don’t have to, since they are applying reasoning at a point in the development of the system where the input of real-world contingencies, which necessitate the need for human lawyers, need not be considered. To that extent, they might be a useful tool in the arsenal of the non-legally trained systems designer who wants to ensure her new idea is at least minimally compliant from the outset. Work has gone some way to achieving this goal, albeit usually in a fairly manual fashion whereby “wizards” or questionnaires require manual feedback from the user in order to traverse the tree of regulatory norms to arrive at a determination of compliance [26, 30]. Our vision is for a more automated system which could compare formalized abstractions of legal and digital systems within the software development environment, thus apply the concept of “by design” at the most effective stage.

PbD is evolving. In the beginning it had a technology-centric focus, albeit a deficient one with respect to practical guidance on implementation. More recently it has moved towards becoming “a conceptual model for building an entire privacy program” [30], widening the focus to include organizational measures, as reflected in Article 23 of the GDPR. Scholars have noted the challenges of engineering privacy without such a concomitant consideration of these “softer” organizational aspects [19, 21, 28]. An integral element in this is the PIA [16, 29]. Use of PIAs appears to be confined mostly to governments and a few very large corporations, although in the latter case there is scant evidence on the quality of the assessments being carried out [31]. It is becoming clear, however, that as more types and sizes of private sector enterprises adapt their products to the IoT economy, and thus become data controllers, there will be an increasing need for effective and easy-to-follow compliance methodologies and tools.

IV. FROM PbD TO SIA

We have argued that PbD has yet to be truly integrated into the design of systems and become a core component of systems engineering. However, we also need to proceed even earlier into the design phase and consider the business drivers and constraints that the engineering team is presented with. Many systems, especially in B2B relationships, are not designed with privacy as a prime consideration. This is for a number of reasons illustrated below. The problem is particularly acute in relation to minimisation of personal data acquisition and sharing. IoT systems can be built to either collect or retain only anonymous data, or to only share data within a household or between one individual’s devices with no second or third party access to any data. However, many systems are not so designed.

For example, in order to achieve rapid deployment of a system for detecting bus occupancy, one project chose to use video cameras and image processing software to count the number of passengers on a given vehicle. A simpler, more reliable and anonymous approach would have been a switch under each seat – we do it for car seat belts already! However, that would have involved more time and capital expenditure, albeit that the operational costs in the long term (beyond the timescale of the specific project) would have been much less, and in the future the capital costs would be ameliorated by bus manufacturers fitting such sensors in the factory.

The IoT baby monitor noted above is a classic example where there is a need for video data, but no functional or commercial reason to have the video available to anything other than a set of specifically nominated devices – end to end encryption with offline key exchange could secure such a system against even a concerted government sponsored attack. That these systems have not been so designed comes down to technical expediency or, perhaps more bluntly, lack of competence: a suitable user-comprehensible offline key
exchange would require design thinking and the cryptography would have to be implemented correctly, while NAT punching and other cunning technical means to overcome network middle-boxes would also be required.

Under-explored problems also exist in commerce as to reuse of data and supply chain dynamics. One business may capture data from one IoT device and the same or another business may reuse the data for other purposes. Unspecified uses of data are one of the major sources of lack of trust in IoT, e.g., at the point of acquisition of Nest by Google, it was not clear if household activity data from Nest users might be integrated with other Google datasets. More recently there has been concern about smart TVs that may collect viewing habits and introduce intrusive screen adverts. EU DP law already stipulates that data must be collected for known and particular purposes and that any reuse (if not separately legitimised) must be compatible with that original purpose. However, this rule has been substantially undermined by the international growth of Big Data and data mining as a revenue stream.

Most worryingly in the IoT, not only data, but the data-collecting device itself, possibly built in an entirely industrial context, may be reused in a more complex product designed for consumer use further up a supply chain. When we consider the supply chain of B2B businesses, often SMEs, whose work is bundled together to create a finished, customer-facing IoT product, it is obvious that privacy and PbD will not have been a main or perhaps even contemplated concern of the SMEs on the lower stages of the supply chain ladder. In this context the contribution that an impact assessment can make is particularly important. Little existing work on PbD and PIAs has dealt with the problems of SMEs down the line, though the PRIPARE project acknowledged the lack of uptake of and awareness of PbD by SMEs, and in the PIAF report, the Spanish DP authority voiced concerns at placing PIA burdens on SMEs, at least without the aid of special software tools [37]. Our SIA will aim to provide a methodology that can plug the gap currently experienced by innovation-led, but regulation-unaware, SMEs.

V. CREATING THE SIA

The above analysis demonstrates that the opacity around IoT devices, both in what personal data they collect, for what purposes and with what lawful ground of processing (almost certainly not meaningful consent) is highly problematic for the privacy of end-users.

Thus we propose developing a comprehensive methodology for assessing the risks associated with building IoT systems for consumer or public use. This will build upon existing legal, social and technical work on PIAs. Specifically, unlike with most PIA work, we will consider the public interest as well as the interests and rights of enterprises and users. Thus our novel proposal is that although work on PIAs has already considered the IoT [15], we plan to move beyond assessing privacy and data protection risks to a Social Impact Assessment (SIA) protocol, which takes account of the wider aspects of Responsible Innovation [36]. Some work has begun on Ethical Impact Assessments e.g. SATORI, PULSE [32] and the European Data Protection Supervisor (EDPS) has also endorsed the idea of an ethical framework for data processing with particular reference to machine learning [33]. A draft ISO standard WD29134 for PIAs drawing on ethical principles is also in development. However, these early stages of creating a methodology have yet to be fully applied to IoT scenarios and to iterative software product development; nor has stakeholder status and implementation responsibility (user, consumer, producer, reuser, “public society”) been fully explored.

In particular, we have commenced our work bottom up by looking at SMEs building IoT systems, rather than the large public sector systems more typically considered. As noted above, privacy in the IoT is particularly problematic as between commercial interests (B2B), while the main focus thus far has been on traditional enterprise-user relationships. Furthermore, although impact assessments are well understood in the spheres of health and public welfare, experience of them in the commercial B2B SME context is minimal. As preliminary work to this, Edwards is collecting data on current knowledge and implementation of PbD principles in B2B SME IoT companies as part of her Researcher in Residence placement at the Digital Catapult1 from January to June 2016. The SIA protocol will be developed iteratively, starting with retrospective application to, and critique of, existing IoT deployments (to be identified as part of the Catapult work with partners of that organisation) and moving on to developing how it should be applied de novo throughout the development process of new pilot applications.

The next stage of analysis will be to trial the following elements as modules of the SIA:

**Personal data minimization and sharing by design** – this should be built in as part of innovation and business development for new product and service opportunities. The work should determine the underlying minimal technical data requirements and clearly separate these from the business drivers or budgetary constraints. One of the biggest challenges here will be to anticipate reuses of data (or generation of new data) in subsequent systems down the supply chain.

**Security by design** – good security engineering is essential, but we also need to consider whether the mechanisms proposed shift the liability and risk unreasonably from the business to the individual customer, and in particular has the security system actually been made useable for the target customers.

**Transparency by design** – algorithms should be tested for disparate impact, as while appearing neutral, many algorithms reflect built in biases in the input data or coding resulting in discrimination which otherwise would be illegal – evidence from Sweeney (race) [35] is well known but may also apply to less examined areas such as gender and class. Appropriate statistical tests should be performed as part of data capture and software testing to address this [33]. Likewise ordinary users often accept algorithmic output as inextricably right (“Computer says no”) and this has already created historic problems in areas such as credit scoring and more recently in machine learning applications such as Google search and

---

1 http://digital.catapult.org.uk/
EU GDPR is mandating data portability to encourage “Apple” installation and all your devices are “Google”. The will be wasteful and/or tediou

Resilience by design – has the designs considered the likely failure modes, and in particular, whether this would have a disproportionate impact on vulnerable communities such as children, the elderly, the disabled and the digitally illiterate.

Interoperability by design - as compared to our ecosystem of personal portable devices, IoT systems will be composed of portable devices and those embedded in the environment. It will be wasteful and/or tedious to find your new apartment is an “Apple” installation and all your devices are “Google”. The EU GDPR is mandating data portability to encourage competition and prevent foreclosure; such logic would seem appropriate for IoT systems too.

ACKNOWLEDGMENTS

This work is funded with grants from Research Councils UK under the Horizon (EP/G065802/1, EP/M02315X/1) and CREATE (EP/K039695/1) projects.

REFERENCES

[16] Information Commissioners Office, Conducting privacy impact assessments code of practice, Version: 1.0 (Feb 2014), p. 5