Special Theme on Privacy and the Internet of Things

Alan Chamberlain · Andy Crabtree

School of Computer Science, University of Nottingham, UK

Hamed Haddadi

School of Electronic Engineering and Computer Science, Queen Mary University London, UK

Richard Mortier

Computer Laboratory, University of Cambridge, UK

Editors' note

When we initially thought about running a special theme on the Internet of Things we were motivated, as many are [e.g., 11, 9, 15, 10], by a concern with the threat to privacy that accompanies the widespread rollout of connected devices. Edith Ramirez, chairwoman of the Federal Trade Commission tasked with protecting consumers in the US, seemed to us to sum the situation up generally:

> "Connected devices [are] collecting, transmitting, storing, and often sharing vast amounts of consumer data, some of it highly personal … … … companies are investing billions of dollars in this growing industry; they should also make appropriate investments in privacy and security. The stakes are too high to do otherwise." [8]

In saying this Ramirez reminds us that privacy and security are key to building societal *trust* into the IoT, an ingredient seen as essential to its success [14]. We note that privacy is not the same as security, however.

While current industry solutions largely put emphasis on encryption as a privacy-preserving measure, there is more to privacy than the confidentiality of data at rest or in motion [7]. Furthermore, encryption often won't stop industry accessing personal data [13], and metadata can be as or more revealing than data itself [1]. Thus, in considering just what this special theme might about it seemed to us important to take account of *what more* might be involved in addressing the privacy risks occasioned by the IoT than security provides for?

Our own interest in 'what more?' is driven by a concern to build accountability into the Internet of Things and to enable ordinary people to control the flow of personal data in everyday life. These entwined issues drive the development of the open source Databox platform (www.databoxproject.uk), which seeks to enable accountable, 'GDPR compliant' [6], personal data processing at the edge of the network. The Databox approach thus *takes computing to the data* [12], rather than data to the computing as per the current 'cloud' paradigm, and this has distinct computational as well as social advantages.

Computationally, as the number of connected devices increases exponentially it will be impractical if not 'resource prohibitive' to transport data for processing over networks to remote data centres [2]. Thus, moving computing to the data reduces network latency and bandwidth contention. Socially, moving data processing to the edge of the network restricts data distribution and with it the accompanying threats to privacy. Thus, instead of shipping data to remote centres for processing, processing can be done locally with the added benefit that *only the results* of processing need be distributed: the data need never leave home, literally and figuratively speaking.

> "The edge of the Internet is a unique place … often just one wireless hop away from … devices … It can be an optimal site for aggregating, analyzing and distilling bandwidth-hungry sensor data … In the Internet of Things, it offers a natural vantage point for … access control, privacy, administrative autonomy and responsive analytics." (ibid.)

As we have previously suggested in PUC [3], this edge approach may also be of distinct economic advantage in enabling individuals to exploit their data for personal benefit. Figure 1 represents a simple use case demonstrating the potential efficacy of this approach.

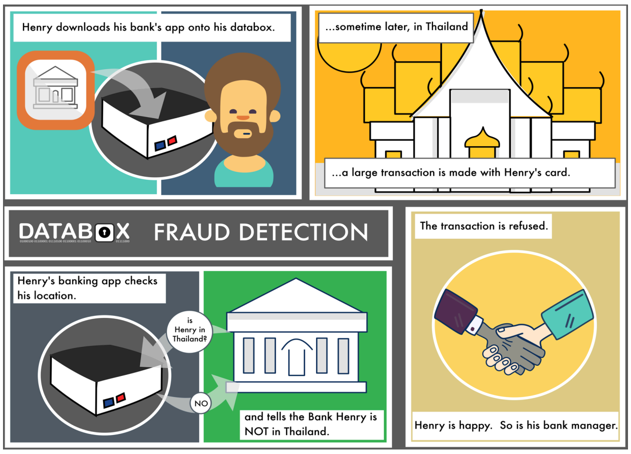
Henry downloads his bank's app onto his databox.
...sometime later, in Thailand
...a large transaction is made with Henry's card.
DATABOX FRAUD DETECTION
Henry's banking app checks his location.
is Henry in Thailand?
NO
and tells the Bank Henry is NOT in Thailand.
The transaction is refused.
Henry is happy. So is his bank manager.

Figure 1. A privacy-preserving use case.

The use case posits a networked world in which access to an individual's personal data is controlled via apps running on the Databox. Thus, as above, Henry installs his bank's fraud detection app on his Databox. When the bank's software encounters suspicious activity on Henry's account, it pings the bank's app on his Databox and the app runs a specific query on Henry's location data 'here and now'. The app does not monitor his location over time. Nor does the query it runs reveal Henry's location to the bank, only that he is *not* in the location where the suspicious activity is occurring. The bank's software can thus prevent fraud and do so by leveraging very personal, even sensitive information in a privacy-preserving way.

The Databox approach allows individuals and developers to benefit from connecting distributed data silos together (e.g., bank records and location data, even from multiple devices). The platform thus fosters privacy-preserving innovation, enabling both commerce *and* consumer to extract value from the utilisation of personal data, and to do so in very familiar ways: through apps and apps stores that make personal data processing explicit and accountable to individuals in the process [4].

Clearly there can be more to privacy than security, computational, social, and economic. We therefore invited original contributions that explore and elaborate what more there might be from multidisciplinary perspectives spanning social, legal, and ethical aspects of the IoT as well as more traditional design perspectives. Nevertheless, by far the largest category of submission we received was concerned with security, which was rather disappointing and seemed to us to underscore the need to move current design thinking on. This is not, of course, to say that security is not important, only that more is required if broad societal trust is to be built into the IoT.

What we are left with when security is removed from the picture is a handful of insightful papers that span ethnography, law, and design. Peter Tolmie and Andy Crabtree's ethnographic study of ***The Practical Politics of Sharing Personal Data*** sensitises the reader to some important issues often overlooked in the IoT and the digital space more generally. These relate to the way that people understand data transactions both locally and in a distributed manner. The paper highlights the complexity of understanding and generalizing notions of 'data sharing', exploring situations where data sharing is purposeful, unwittingly done, or even incidental. In carrying out these studies the authors reveal 'real-world' understandings that elaborate the reasoning and practices ordinarily implicated in dealing with data, and systems that create and use data. One might argue that these studies provide a lens onto what actually happens, rather than what is perceived to happen.

As this paper shows, for example, the nature of the 'personal' is bounded by context, which is to say that what constitutes 'personal' in one situation, is different from notions of 'personal' in another. When one bears that in mind, and takes into account the changing nature of the situation, we are able to see that the practices and terminology associated with such situations are in a state of flux. They are social in nature, negotiated in respect to the situation, and implicated in that situation are a range of features that are reasoned about. It is this reasoning and the 'politics of sharing' it elaborates that Tolmie and Crabtree are able to unpack. In rubbing up against the taken for granted and mundane this paper provokes us to ask what it really means to 'share personal data' and to consider the challenges for the IoT in doing so.

Lachlan Urquhart, Neelima Sailaja and Derek McAuley turn our attention to the European Union's General Data Protection Regulation [6], which comes into effect on the 25th of May 2018, and the challenges involved in ***Realising the Right to Data Portability for the Domestic Internet of Things***. This right, enshrined in Article 20, mandates that individuals should be able to obtain any personal data which has been provided to a data processor "in a structured, commonly used and machine-readable format" *and* have the right to transmit that data to another processor "without hindrance". The paper unpacks emergent issues relating to this regulation, and shows that trying to understand it in a narrow technical sense is not sufficient. Rather, one needs to appreciate that not only are multiple types of data processed and stored, but the data is used by a series of actors in differing ways, and those actors may be legally governed in differing ways.

In many respects this paper starts to raise serious questions about the socio-technical nature of data. Thus data is something that comes into contact with a set of external forces, that can interpret, access and replicate it, and make it difficult to know how one's data has been used and by whom, even at a very basic level. Once data has been released is it possible to track or to understand the ways in which systems come into contact with and use it? Issues such as this are compounded in IoT systems, where one might expect a multitude of artefacts to be in contact with each other locally and in a distributed manner. Urquhart et als' work brings us back to a very obvious set of issues that are equally hard to negotiate: a world of contracts, consent and control. A world that is perhaps as difficult for designers and researchers to navigate and understand, as it is for people dealing with data in their own homes.

Thomas Pasquier, Jatinder Singh, Julia Powles, David Eyers, Margo Seltzer and Jean Bacon draw us further into the world of regulation in their paper ***Data Provenance to Audit Compliance with Privacy Policy in the Internet of Things***. The concern here is to understand how the IoT can be developed to satisfy the accountability requirements of data protection regulation. With GDPR on the horizon, there is significant need for work like this to enable privacy-by-design. The paper looks particularly at challenges posed by lack of transparency and accountability in IoT information flows, a considerable challenge for data protection as information often flows between different contexts, from domestic to work, private to public, in the IoT ecosytem.

Pasquier *et al.* move design forwards in this crucial area by leveraging the CamFlow infrastructure to audit information flows and demonstrate that systems handling personal data satisfy both regulatory and user requirements. CamFlow implements Linux Security Modules (LSMs), which enable and enforce consent and integrity checking by tagging data as part of its flow. The approach has been shown to work in a distributed architecture and this paper makes the case for use in distributed IoT applications. Whether or not the approach will scale in real time has yet to be demonstrated, though it is undoubtedly the case, as the authors remind us, that data protection requirements *will* apply to significant volumes of data generated within the IoT, and that it is incumbent on developers to put mechanisms in place that provide evidence of where data has flowed in the ecosystem.

Ilaria Torre, Odnan Ref Sanchez, Frosina Koceva and Giovanni Adorni's paper addresses a key human challenge in ***Supporting Users to Take Informed Decisions on Privacy Settings of Personal Devices***. Understanding privacy policies and communicating them to users is a major challenge in an increasingly complicated ecosystem of devices, sensors, apps and permissions. The number of sensors on smart devices, the variety of data types available, and the number of apps and brokers which collect, utilise, and trade this data, is on the rise and there is increasing interest in this aspect of human-data interaction and privacy communication.

Torre *et al.* address the challenge in the context of understanding privacy preferences for wearable devices. This is an overwhelming choice for most individuals, especially as the inference possibilities are unknown in the present and future. The authors evaluate their inference threat notification framework using a relatively large dataset provided by weight loss 'LoseIt' users. The authors present an engaging analysis of LoseIt users personal choices and the privacy risks and dilemmas that accompany quantified self devices. The use of a Bayesian learning method for the inference of private cross-data information is effective and the authors have made the paper reproducible by going into the effort of using open datasets. Importantly, the authors use machine-learning tools to provide inference protection and configuration recommendations for the users to manage their privacy in an intuitive manner.

Developing efficient privacy preserving measures for connected devices is key to Joseph Korpela and Takuya Maekawa's paper ***Privacy Preserving Recognition of Object-based Activities Using Near-Infrared Reflective Markers***. This paper addresses important privacy concerns that arise when image recognition is leveraged by the IoT. Image recognition is widely proposed as a useful mechanism in the IoT ecosytem, but there is increasing societal concern over the current trend towards introducing connected audio-visual systems to collect data before shipping it to the cloud for processing, as they have great potential to invade users' privacy. The system presented here seeks to obviate privacy concerns by not using visible light wavelengths. Thus, even if video data needs to be processed outside the home, it will contain far less information that might lead to a privacy leak.

The lab-based evaluation provided by the paper demonstrates that, while low-cost, the accuracy achieved is comparable with the non-privacy preserving systems. While concerns always remain about just how well the provision for privacy can be measured and evaluated in the lab, and there is need for in-the-wild evaluation to understand how such technology might really be applied and used, there is evidence here of the utility of this approach and its ability to enable recognition of quite fine-grained activities in a privacy-preserving way.

Together, the papers included in this special theme elaborate privacy-related topics that extend beyond security: understanding how people actually share data and the challenges this raises for the IoT; satisfying legal requirements concerning data portability and the accountability of data processing; enabling users to make informed privacy decisions concerning, and implementing measures to reduce, the risks to privacy created by connected devices.

We have no doubt that this small collection of disparate papers merely scratches the surface of the IoT privacy challenge, though perhaps their broader value, to borrow from sociologist Harold Garfinkel [5], is to treat them as "aids to a sluggish imagination" designed to drive reflection on an "obstinately familiar world" in which privacy is currently subsumed to a large extent under the auspices of security. There *is* more to the matter, and we thank our authors for demonstrating the point and driving it home with clarity.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council [grants EP/M001636/1, EP/M02315X/1, EP/N028260/1].

References

1. Apthorpe, N., Reissman, D. and Feamster, N. (2016) "A smart home is no castle: privacy vulnerabilities of encrypted IoT traffic", *Workshop on Data and Algorithmic Transparency* (DAT '16), November 19. New York, http://datworkshop.org/papers/dat16-final37.pdf [accessed 03-07-2017].
2. Chiang, M. and Shi, W. (2016) Workshop Report on Grand Challenges in Edge Computing, National Science Foundation. http://iot.eng.wayne.edu/edge/finalreport.php [accessed 03-07-2017].
3. Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C., Mortier, R. and Haddadi, H. (2016) "Enabling the new economic actor: data protection, the digital economy, and the Databox", *Personal and Ubiquitous Computing*, vol. 20 (6), pp. 947-957.
4. Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C. and Mortier, R. (in print) "Accountable Internet of Things? Outline of the IoT Databox model", *Proc. of WoWMoM*, pp. 1-6, Macau, IEEE.
5. Garfinkel, H. (1967) *Studies in Ethnomethodology*, Englewood Cliffs, Prentice-Hall.
6. General Data Protection Regulation, *Official Journal of the European Unio*n, vol. 59, pp. 1-88, May 2016.
7. Kamp. P-H. (2013) "More encryption is not the solution", ACM Queue, vol. 11 (7), pp. 1-4.
8. Ramirez, E. (2015) "Privacy and the IoT: navigating policy issues", *International Consumer Electronics Show*, January 6, Las Vegas. www.ftc.gov/public-statements/2015/01/privacy-iot-navigating-policy-

issues-opening-remarks-ftc-chairwoman-edith [accessed 03-07-2017].

9. Roman, R., Zhou, J. and Lopez, J. (2013) "On the features and challenges of security and privacy in distributed internet of things", *Computer Networks*, vol. 57, pp. 2266-2279.
10. Sicari, S., Rizzardi, A., Grieco, L. and Coen-Porisini, A. (2015) "Security, privacy and trust in Internet of Things: the road ahead", *Computer Networks*, vol. 76, pp. 146-164.
11. Weber, R.H. (2010) "Internet of Things – new security and privacy challenges", *Computer Law and Security Review*, vol. 26, pp. 23-30.
12. What is Databox? www.youtube.com/watch?v=hJQiIrkJniU&list=PL9UQcjjmGF3cUoLmvvfrepPAM62sxx5_3&index=1 [accessed 03-07-2017].
13. Winstein, K. (2015) "Introducing the right to eavesdrop on your things", *The Agenda Magazine*. www.politico.com/agenda/story/2015/06/internet-of-things-privacy-concerns-000107 [accessed 03-07-2017].
14. World Economic Forum (2014) *Rethinking Personal Data: A New Lens for Strengthening Trust*. www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf [accessed 03-07-2017].
15. Ziegeldorf, J., Morchon, O., and Wehrle, K. (2014) "Privacy in the Internet of Things: threats and challenges", *Security and Communication Networks*, vol. 7 (12), pp. 2728-2742.