The Quest for Privacy in the Consumer IoT

Johanna Ullrich, Artemios G. Voyiatzis, Edgar R. Weippl
SBA Research, Vienna, Austria
Email: (firstletterfirstname)(lastname)@sba-research.org

***Abstract*—Privacy remains among the toughest challenges for the consumer-facing Internet of Things (IoT). Privacy-by-Design (PbD) is the most recent attempt to address it. Thereby, privacy goals become part of the technical specification and are resolved directly in the development process. This contemplation opposes existing approaches that retrofit protection measures as an afterthought, often even after the introduction of the "things" in the market. PbD is not solely a technological approach; it is directly addressed by the European General Data Protection Regulation (GDPR) that is presumably going to come into force in 2018.**

In this paper, we highlight the drawbacks of the retrofit approach when applied to IoT, using as a case the IPv6, one of IoT's key networking technologies. We argue that PbD is a resolution of specific significance (if not by now the only one) promising to directly solve the privacy challenges. Nevertheless, we identify a significant omission: neither legislation nor technology mandate the consumer involvement.

I. Introduction

The Internet of Things (IoT) is shaping our future and is fundamentally changing today's Internet as well as our lives. In principle, the IoT aims to invisibly embed objects ("things") surrounding us into the Internet; and as this recent paradigm includes lots of sensors, it is expected to generate large amounts of data [1]. Today's consumer-grade IoT devices comprise among others light bulbs[1], TV sets[2], audio systems[3], fitness trackers[4], and heating systems[5]. IoT is expected to provide more convenience, but also energy efficiency, health and training improvements, and increase time efficiency.

Information security sets its goals according to the triad of **C**onfidentiality, **I**ntegrity and **A**vailability (CIA) [2]. Confidentiality protects sensitive data from access by unintended parties; integrity from unauthorized modification of data, and availability from outage. For sure, the most important thereof from a consumer's point of view is confidentiality. While they are able to stand a fitness tracker's temporary outage (availability) and malformed heart rate measurements to a certain extent (integrity), a consumer is typically anxious of personal data leaks (confidentiality) – commonly also referred to as "data privacy" within the information security community.

From a legal perspective, privacy is a very fundamental right and is essential for exercising other (fundamental) rights, e.g., the right of thought, the right of expression and information, the right of assembly. Allowing the exclusion of others, the right of privacy guarantees a "personal sphere" for individuals and enables free development. Within the European Union, the right of privacy is laid down in Article 7 of the *Charter of Fundamental Rights of the European Union* claiming that *"Everyone has the right to respect for his or her private and family life, home and communications"* [3], and is moreover found in one form or another in national constitutions.

The right of privacy includes an individual's control over access to personal data and thus, protection of data when provided to others is the other side of the coin. The right of privacy typically goes hand-in-hand with the right of personal data protection and urges the application of adequate means of protection[6]. The European *Charter of Fundamental Rights* claims in Article 8 that *"Everyone has the right to the protection of personal data concerning him or her"* [3].

Facing the importance of privacy protection, its realization during typical product development is an afterthought: In a first step, a prototype fullfilling functional requirements is developed ignoring the non-functional requirement of privacy protection. Later, for achieving compliance with standards, regulations, and legislation, privacy is tried to be achieved by retrofitting some extra protections. Privacy issues are similarly tackled in a reactive way – (1) by retrofitting before market launch, (2) by patching, in case of a problem occuring within product life time, or (3) by total ignorance leaving privacy shortcomings as they are. Legal data protection frameworks work in the same reactive manner through punishment – and even that rather seldom supporting the "anything goes" approach by total ignorance. Apparently, this solution is far from perfect.

With respect to IoT, the inflicts on privacy become even worse, considering that:

- IoT devices have longer product life cycles than those in traditional information technology. This means that what is in place at a certain point in time will stay there for years or even decades. In the case of sensor nodes that harvest energy from the surrounding environment, they will eventually stay until their breakage [4].
- IoT equipment is not always reachable. It might be deployed at physically inaccessible places or comes "online" for small periods for reasons of energy efficiency [5] – both making patching difficult. The devices' strict constraints, e.g., with regard to computing power, further limit patching.

[1]e.g., Philips Hue, see http://www.meethue.com/
[2]e.g., by Samsung, see http://www.samsung.com/us/experience/smart-tv/
[3]e.g., by Raumfeld Streaming Systems, see http://www.raumfeld.com
[4]e.g., Jawbone UP2, see http://jawbone.com
[5]e.g. by heatmiser http://www.heatmisershop.co.uk

[6]Measures of personal data protection appear to comply with information security's term of privacy protection.

- The IoT encompasses a vast amount of devices collecting a lot of data. Replacement of all "things" due to a known privacy vulnerability can be a logistic nightmare, which means that these vulnerabilities are likely to remain there. Beyond that, leaked consumer data are once for all revealed, even in case of retro-respective punishment, and might negatively affect the consumer now or at some point in the future.

In this paper, we address the challenge of adequate privacy in the consumer-facing IoT. In a first step, we shed light on IPv6 – a key technology of the IoT – and discuss how its shortcomings seriously inflict consumer privacy. By this example, we highlight the risks of retrofitting, and argue that privacy has to be taken into account right from the beginning. Privacy-by-Design (PbD) is such an approach, that is also addressed by the new European General Data Protection Regulation. Nevertheless, we argue that consumer involvement remains weak and bears the risk of making compliance a farce.

The remainder of the paper is organized as follows: Section II presents the privacy issues surrounding IPv6, the IoT key networking technology, while Section III introduces the approach of PbD as a potential solution. The discussion of Section IV highlights the weak consumer involvement and identifies its consequences. Section V concludes our paper.

II. A Problem Statement at the Example of IPv6

The IoT relies heavily on already-available technologies with regard to sensors and actuators, networking or computing. Per se, reuse of available technology is desirable as it allows technological advance in a time- and cost-effective manner. Within the discipline of software engineering, reusability is even an explicitly stated goal.

Breaking Implicit Assumptions: Though, reusability can also go wrong as experienced with the introduction of cloud computing. Numerous well-known technologies were reused. However, their novel combination and changed purpose has led to various pitfalls; mainly because implicit assumptions were broken in the new environment.

One example is protection by means of firewalls [6]. In a traditional setting, a firewall is installed at the border of a private network, e.g., of a corporation, and is configured to filter malicious traffic from the Internet. The hosts within the private network are considered benign, as these hosts are controlled by the same entity. Within cloud computing, insiders are unknown to each other as basically everybody can become an insider.

Similarly, the IoT risks privacy vulnerabilities due underlying key technologies. In this section, we focus on the key technology of the *Internet Protocol version 6* (IPv6). IPv6 is by no means a technology that is likely to cause severe privacy vulnerabilities, but by far the only as ubiquitous as required for networking.

A Rough Guide To IPv6: Recognizing the fact that the Internet is running short of addresses, the *Internet Engineering Task Force* (IETF) initiated the standardization of a new protocol. This protocol is IPv6 [7]. As address shortage has been the main issue even in the pre-IoT era, the IoT takes advantage of IPv6 for obvious reasons. The vast amount of devices will require just as much addresses, and this would be infeasible with its predecessor IPv4.

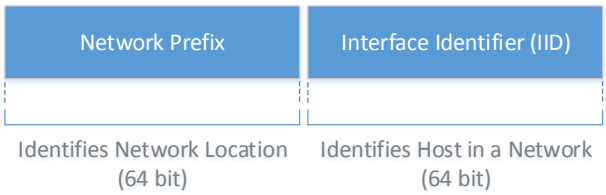

Fig. 1. IPv6 Address Structure

"Old" IPv4 address just have 32 bits while "new" IPv6 allows 2^{128} addresses – more than enough to provide every single square millimeter on earth with more addresses than the whole IPv4-based Internet has in total. The new addresses have a distinct format comprising two parts of equal size, as depicted in Figure 1:

- *Network Prefix:* The first 64 bits form the network prefix, and are dependent on the location of a host (for example a laptop) within the Internet. It is comparable to a post code: It is dependent on the village that you live at, and changes when moving to another one.
- *Interface Identifier:* The remainder 64 bits form the interface identifier (IID) that uniquely identifies a host within a certain network prefix. This is comparable to a name, that identifies you uniquely in your village. There is a minor difference, that you are allowed to freely choose your name as long as it is different from all the others in the same village.

IPv6 addresses are configured by every host individually. It connects the network prefix that is announced by the router with its self-chosen interface identifier [8]. From that time on, it is reachable from all over the Internet by means of this address. Different ways of choosing the interface identifier are available:

- *Manual Configuration:* The operator decides for a number of their choice. While this works well for small number of hosts, it will not become the method of choice for consumer IoT devices. On the one hand, the configuration of the sheer amount of IoT devices is time consuming. On the other hand, it does not appear reasonable for consumers to configure addresses themselves.
- *Modified EUI-Format* [9]: Every networking device has a 48-bit unique identifier. This so-called Media Access Control (MAC) address remains stable for the whole lifetime of a device. The Modified EUI-Format forms an interface identifier therefrom by inserting a fixed pattern to reach the length of 64 bits.
- *Privacy Extension* [8]: The interface identifier is changed on a daily routine, at the best in a random manner.

Privacy Shortcomings of IPv6 Addressing: As already mentioned above, manual configuration appears unlikely for

IoT devices leaving the choice between the *Modified EUI-Format* and the *Privacy Extension*. However, both have distinct privacy drawbacks.

The *Modified EUI-Format* includes a unique identifier into the address, and the address is visible to others on the Internet. This means that all addresses of a single devices, even those from different locations in the Internet, are equivalent in the latter 64 bits of their address, and all Internet activity can be traced back. At the example of a stationary temperature sensor, this implies that an adversary is able to determine the activity of the sensor. At the example of a portable fitness tracker, this would mean that an adversary could even track the physical movement by means of the respective network prefixes [10].

The *Privacy Extension* was originally developed to overcome this drawback by means of randomly-changing the interface identifier. However, the proposed algorithm is far from perfect and an adversary can forecast future interface identifiers [11]. Thus, even devices using the privacy extension can be traced – admittedly with higher computational effort for the adversary.

***Implications on IoT Privacy**:* We lack an IPv6 addressing standard that protects privacy, and at the moment it appears as there will not be any solution soon, neither for the known Internet nor the IoT. However, this situation is far more dreadful for the latter.

Ordinary PCs, smart phones, etc. are patchable and updates deliver the correction of a privacy vulnerability whenever necessary. Likewise, cloud-based infrastructures are even easier to update as they are operated by a single entity each. However, this does not necessarily hold as a high number of consumer IoT devices will not have this possibility, and thus remain as they were deployed for the remainder time of operation. On the one hand, they are of constrained resources. On the other hand, they will not be permanently online for reasons of power saving. Neither, it will be possible to replace all of them due to their sheer amount, and sometimes they might be deployed at rather inaccessible places.

IoT devices will have significantly longer life cycles than traditional devices in information technology. The constrained resources are not limited to resources like memory that allow to process only lightweight algorithms, but also to the availability of certain interfaces: IoT devices range from smart cars to intelligent light bulbs, i.e., there is a lot of variety, and the availability of certain interfaces is dependent on the application scenario. A camera might be used to scan fingerprints or the iris in order to increase security, but devices like a thermostat are unlikely to have such a camera included.

Finally, it is a cost factor: In the automotive industry, an average recall affects 10,000 cars, and US$ 200 are spent per car for rectification of defects [12]. Considering the obstacles in this high-revenue market, one can assume the arising difficulties in the utter competitive consumer IoT mass market that aims to save every single penny.

With respect to IPv6, this means that IoT devices using today's state-of-the-art of addressing are leaking data [11], and cannot be considered as privacy-aware anymore. Particularly bitter is the fact that devices using the privacy extension were presumably even sold as being especially privacy-affine; but nevertheless, they will now permanently leak data.

III. The Privacy-by-Design Approach

The example of IPv6 drastically highlights the shortcomings of plain technology reuse, how such behavior is prone to become an enduring situations in the IoT and endangers consumer privacy. The crux is that once the devices are deployed, they remain there without any chance for later modification. It therefore seems essential that privacy protection is included before market launch, at the best already at the very first prototype, i.e., privacy should be included into the device by design. Thus, the dictum of the time is *Privacy-by-Design* (PbD) – in legislation as well as in technology.

The *General Data Protection Regulation* (GDPR), the future legal framework on data protection within the European Union, addresses PbD directly. It is estimated that GDPR will become effective in 2018. In general, it is expected to have a positive impact on data protection, e.g., with respect to its territorial scope, harmonization among member states and the right to be forgotten.

Focusing on PbD, Article 23 of GDPR defines PbD as *"... the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, [...], which are designed to implement data protection principles, ..."* [13]. In addition, the security-by-default becomes obligatory: *"The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed; [...] such measures shall ensure that by default personal data are not made accessible without the individuals intervention to an indefinite number of individuals"* [13].

The law is technology-neutral for a good reason. Technology and especially information and communication technologies are evolving in a much faster pace than legislation. Thus, law has to articulate itself in rather general terms to provide an umbrella for current and future situations. Nevertheless, its vagueness leaves open many issues that are of importance for practical application. Certification might indeed come at some point in the future, but is not available at the moment – presumably as experience with PbD is still rather low.

The *European Union Agency for Network and Information Security* (ENISA) elaborated however on PbD, and accumulated today's knowledge on the methodology on PbD [14]. At a very high level, PbD's line of action aims to reach a goal that is dictated by law with means of technology as depicted in Figure 2; obviously, requiring some form of cooperation. However, legal demands cannot be directly realized through technological solutions; they are rather used as a first step to derive more detailed goals. Then, appropriate privacy technologies that fit the purpose best are chosen.

The overall approach, as highlighted in [14], is depicted in Figure 3.

Fig. 3. Privacy-by-Design Methodology

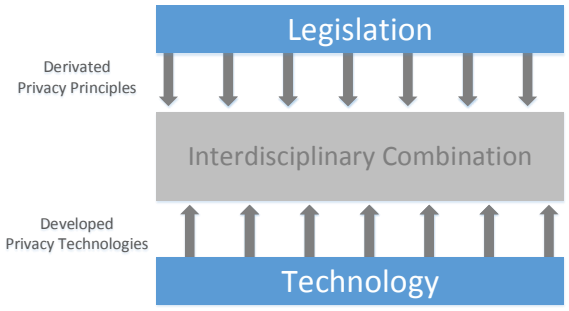

Fig. 2. Privacy-by-Design: An Interdisciplinary Combination of Legislation and Technology

- *Legal Framework:* Law sets the greater goal of privacy that has to be met by IoT devices; however, the legal framework does not necessarily have to call for PbD like the GDPR.
- *Derivation of Privacy Principles:* The greater goals are broken down into smaller principles. Inferred from European legislation, ENISA sees among others the following principles: purpose binding of data processing, an explicit consent of the concerned person, data minimization or accountability in order to gain clear responsibilities.
- *Privacy Risk Analysis:* Nevertheless, the above privacy principles are still not in the form of a technical specification. Thus, a *Privacy Risk Analysis* (PIA) is conducted to define the latter. Typically, a PIA includes the identification of stakeholders and risks, and the development of recommendations.
- *Privacy-Aware Design & Implementation:* Finally, the step towards concrete realization is come. Therefore, numerous design patterns and strategies are available. They can be considered as best-practise, and are heavily used in software engineering, not only to reach privacy expectations.
- *Evaluation:* Applying PbD does not suffice, it also has to be proven that the final system is compliant to the goals of the beginning. This is equivalent to the idea of testing in software engineering in order to guarantee that software meets the specification.

On the one hand, this process appears straight forward as it is similar to software development processes. However, the devil is in the detail. By now, there is not much experience on PbD. Thus, the number of patterns and strategies are yet limited, and may cause demanding challenges in the design process, at least at the first attempts. One critical step within this development is definitely the translation from one field of expertise into another. Legislation defines societal goals, that have to be accordingly met and solved by technology; but in principal, technology appears suitable to solve the issue.

IV. Towards Consumer Participation

With the introduction of the *General Data Protection Regulation* (GDPR), privacy-by-design is directly addressed by the legal framework of the European Union in order to guarantee sufficient privacy protection. Similarly, computer science has privacy-by-design on its radar in order to make privacy goals a part of the specification and overcome current shortcomings, particularly the pronounced patching mentality. Following this approach, basic principles for privacy are derived from the legal framework, and these should be met in the development phase. The solutions are likely to be based on the experience gained from the area of *Privacy Enhancing Technologies* (PETS). Such a preactive approach is of utter importance for a privacy-preserving *Internet of Things* (IoT) because IoT devices cannot be easily replaced or patched in case of a vulnerability. Our example of IPv6 demonstrated that otherwise already deployed (and data leaking) devices will leak even more personal data over their long period of operation.

Summarizing the whole issue, legislation introduces laws calling for protection; science provides potential solutions, and businesses have to develop privacy protecting devices potentially following the privacy-by-design approach. What appears strange in this discussion is that consumers remain uninvolved to a large extent; but they buy and use IoT devices, spend most time with these devices, and it is also their personal data that are processed! Manifold examples are available, including: smart TVs with voice control that transmit recorded user commands to their vendor [7] and baby monitors that post photos and/or videos of the babies directly on the Internet for the entire world to watch [8]. In these cases, the consumers were totally unaware of these privacy leaks, they were neither informed about this functionality nor asked for their permission. But also legislation lacks consumer participation: The reform of EU data protection rules, especially the negotiations of the council, remained shielded from the public. The latter had to rely on leaked data [15] in order to examine the current draft. This battle seems to be over, but we doubt that privacy-by-design works without consumer participation as important information is just missing for proper development, as highlighted in the following paragraphs.

[7] http://techcrunch.com/2015/02/10/smarttv-privacy/
[8] http://bit.ly/1PLdJRV

Which data should be collected? Consumers willingly provide their personal data for some purposes, but refrain from doing so for others. These notions are dependent on the respective IoT device's functionality, and thus have to be derived anew in each design process. Beyond, the notion of privacy might vary among different cultures, even within the European Union and such differences have to be included as well.

Which means of privacy protection is appropriate? Beyond suiting technological and legal aspects, the IoT device has to fit the consumer needs as well. This is of importance in case privacy protection requires additional consumer intervention in comparison to a non-protecting device, e.g., entering a password when accessing. If the privacy gain in return for the effort remains low, it is likely that protection is turned off. In such a case, also the obligation towards privacy-by-default becomes a farce. The ratio on gain and return however is dependent on the processed data and its notion of privacy, both again heavily dependent on societal assumptions. A measure's usability plays a decisive role, too.

Consumer participation within PbD processes would be the only way to take their wishes for privacy seriously by transforming the latter into device specifications of the development process. Otherwise, businesses will continue developing devices that are ostensibly compliant, but create discomfort at the consumer. The current and also the future legal framework of the European Union do not make provisions for such participation. This might be a consequence of the underlying model of data protection regulation. The person whose data are processed plays a rather passive role. This model however does not appear appropriate anymore. Consumers are far more active, provide their data voluntarily and benefit. Participation would reflect this change. Along these line, the discussion about the person as a (partial) controller in the sense of data protection legislation takes place.

Beyond, the following question remains: *How can adequate consumer participation be organized?* It appears impossible that every single consumer is included into the development process, especially for devices of the IoT mass market. We believe that state data protection authorities or consumer protection boards are suitable candidates, as they already have experience in adequate consumer representation. However, they will need more specialized staff to perform accordingly.

V. Conclusion

Privacy is still one of the major challenges of the consumer-facing Internet of Things (IoT). Consumers accept temporary outage (availability) or occasional errors in measured values (integrity), but do not condone personal data leaks (confidentiality). Our example of IPv6, an IoT key technology for networking, highlights the drawbacks of today's practice of privacy protection as an afterthought. Protection measures are added just before market launch; once they are put into the field, potential privacy vulnerabilities remain and leak more personal data in their time of operation. We argue that Privacy-by-Design allows to overcome this by its systematic approach; privacy goals are directly included into the design process, and solved accordingly.

Nevertheless, we highlight that neither legislation nor technologies mandate consumer involvement; but consumers are eventually buying, and using the respective devices; it is their data processed. We argue that consumer involvement is a necessity for personal data protection. The notion of privacy varies among cultures, target audience, and purpose of the IoT devices! Raising the question for sustainable anchoring of consumer involvement, we suggest the involvement of state data protection authorities or consumer protection boards.

Acknowledgment

This research was supported by the Bridge Early Stage grant P842485 and the COMET K1 program of the Austrian Research Promotion Agency (FFG).

References

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645 – 1660, 2013.

[2] C. Perrin, "The CIA Triad," 2008. [Online]. Available: http://www.techrepublic.com/blog/it-security/the-cia-triad/

[3] European Parliament, "Charter of Fundamental Rights of the European Union," 2010.

[4] M. Penella and M. Gasulla, "A review of commercial energy harvesters for autonomous sensors," in *Instrumentation and Measurement Technology Conference Proceedings*, 2007, pp. 1–5.

[5] J.-M. Liang, J.-J. Chen, H.-H. Cheng, and Y.-C. Tseng, "An energy-efficient sleep scheduling with QoS consideration in 3GPP LTE-Advanced networks for Internet of Things," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, vol. 3, no. 1, pp. 13–22, 2013.

[6] J. Cropper, J. Ullrich, P. Frühwirt, and E. Weippl, "The role and security of firewalls in IaaS cloud computing," in *Availability, Reliability and Security (ARES), 2015 10th International Conference on*, 2015, pp. 70–79.

[7] S. Deering and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," RFC 2460, December 1998.

[8] S. Thomson, T. Narten, and T. Jinmei, "IPv6 Stateless Address Autoconfiguration," RFC 4862, September 2007.

[9] R. Hinden and S. Deering, "IP Version 6 Addressing Architecture," RFC 4291, February 2006.

[10] T. Narten, R. Draves, and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6," RFC 4941, September 2007.

[11] J. Ullrich and E. Weippl, "Privacy is not an option: Attacking the IPv6 privacy extension," in *Proceedings of the International Symposium on Research in Attacks, Intrusions, and Defenses (RAID)*, 2015, pp. 448–468.

[12] I. Wright, "Higher Factory Utilization Leads to More Auto Recalls," 2016. [Online]. Available: http://www.engineering.com/AdvancedManufacturing/ArticleID/11475/Higher-Factory-Utilization-Leads-to-More-Auto-Recalls.aspx

[13] Presidency of the Council of the European Union, "General Data Protection Regulation - Analysis of the final compromise text with a view to agreement," Dec 2015.

[14] European Union Agency for Network and Information Security, "Privacyand Data Protection by Design - from policy to Engineering," 2014.

[15] "LobbyPlag," 2016. [Online]. Available: http://lobbyplag.eu/map