In this design, wearable sensors are used to capture raw data (e.g., accelerometers, gyroscopes) that can be used to identify users’ (patients) activities. Data is then sent to the cloud for activity recognition using a mobile phone as an intermediary device. Next, cloud services are used to process the data and send alerts to a nurse to change the exercise plan as necessary. Two software engineers need an IoT application which can be used to monitor patients’ rehabilitation process. This use case is inspired by a real-world application called ‘MyPhysioapp’ (myphysioapp.com) [5]. A doctor has compiled his functional requirements as follows. The doctor has difficulties in seeing his patients frequently due to different reasons (e.g., traveling distance, work schedules, etc.). Further, frequent in-person consultations are not necessary for most circumstances. Each in-person visit incurs costs for both the doctor (government) and the patient. Once the initial consultation is performed, the doctor only needs to track the patient’s progress and does not need to meet the patient unless something exceptional has happened. The doctor is only interested in tracking the patient’s progress. After evaluating the progress every two weeks, the doctor may ask his specialty nurse to change the exercise plan as necessary. Two software engineers have come up with two different designs as follows to fulfill the above functional requirements. The designs are visually illustrated in Fig. 1.

\[ \text{Design 1: In this design, wearable sensors are used to capture raw data (e.g., accelerometers, gyroscopes) that can be used to identify users’ (patients) activities. Data is then sent to the cloud for activity recognition using a mobile phone as an intermediary device. Next, cloud services are used to process} \]

---

**Privacy Challenge at Design Time**

The engineering complexities in the Internet of Things (IoT) have forced engineers to focus most of their efforts on addressing challenges such as interoperability, reliability, and modifiability, resulting in privacy concerns being largely overlooked [1] [2]. IoT applications typically collect and analyze personal data that can be used to derive sensitive information about individuals. Without proper privacy protections in place, IoT applications could lead to serious privacy violations. Over the last few years, we have seen several privacy violations (e.g., baby monitor [3], Google smart speaker eavesdropping[1]). Traditionally, privacy challenges are addressed in an isolated manner by different research communities (e.g., networking, database, software engineering, and human-computer interaction) [1]. More importantly, such independently developed solutions are complicated to adopt and require significant expert knowledge, time, and resources.

In contrast, we propose an end-to-end unified technique that does not require expert knowledge for it to be adopted, therefore reducing the cost associated with designing privacy-aware IoT applications. Our vision is to develop a tool that the software engineering community has not seen before that would not only bring privacy-by-design (PbD) into mainstream engineering but also ensures that IoT applications are compliant with leading laws and regulations (e.g., General Data Protection Regulation (GDPR) [4]) working toward creating a safer and privacy-aware IoT ecosystem. Usability, consistency, trustworthiness, scalability, accuracy, accessibility, and extendibility are major unique characteristics of this tool.

**Contribution:** The primary goal of this paper is to highlight the importance of developing tools to augment the capabilities of software engineers toward designing privacy-aware applications. First, we present an example scenario where we argue the need for a tool. Then, we discuss how such a tool should behave and the functionalities it should provide to the developers. Finally, we provide some insights into how to develop such a tool by linking to existing literature.

**Walking Through an Example**

Let us consider a simplified use case scenario to highlight the challenges in designing privacy-aware IoT applications. A doctor needs an IoT application which can be used to monitor patients’ rehabilitation process. This use case is inspired by a real-world application called ‘MyPhysioapp’ (myphysioapp.com) [5]. A doctor has compiled his functional requirements as follows. The doctor has difficulties in seeing his patients frequently due to different reasons (e.g., traveling distance, work schedules, etc.). Further, frequent in-person consultations are not necessary for most circumstances. Each in-person visit incurs costs for both the doctor (government) and the patient. Once the initial consultation is performed, the doctor only needs to track the patient’s progress and does not need to meet the patient unless something exceptional has happened. The doctor is only interested in tracking the patient’s progress. After evaluating the progress every two weeks, the doctor may ask his specialty nurse to change the exercise plan as necessary. Two software engineers have come up with two different designs as follows to fulfill the above functional requirements. The designs are visually illustrated in Fig. 1.
the raw data and identify the user’s activity patterns. User activity patterns are then compared with the doctor’s recommended rehabilitation plan to produce a progress report. The doctor can review the progress and make recommendations to the nurse regarding any alterations.

**Design 2:** In this design, wearable sensors are not only used to capture raw data but also to identify activities (using the micro-controllers attached to the wearable). Timestamped activities are then sent to the mobile phone. The nurse, based on the doctors’ recommendations, creates the exercise plan and sends it to the patient’s mobile phone. The mobile phone then compares the timestamped activity data and the exercise plan to determine how well the patient is performing the exercises. The mobile phone sends a weekly progress report to the doctor. Based on the report, the doctor advises the nurse and the nurse alters the exercise plan accordingly.

It is important to note that both designs satisfy the doctor’s functional requirements. However, Design 2 is certainly ‘better’ than Design 1 in terms of privacy awareness.

Based on this use case scenario, we extract two research questions as follows:

- How can we define and operationalize ‘better’ IoT application designs (in terms of privacy)?
- How can we automatically convert a weaker design (e.g., Design 1) into a better design (e.g., Design 2)?

We consider privacy as a trade-off function. Applying a certain privacy-preserving measure into a certain IoT application may impact the implementations in terms of costs, complexity, usability, fault tolerance, responsiveness, etc. Therefore, our aim is not to prescribe a certain design over others. Instead, we want the developer to be informed about privacy-by-design choices before they make their final design decisions. In this regard, we propose a usable privacy-aware IoT application design tool that will inform the privacy-aware design choices to the developers. Previous investigations have shown that applying privacy principles into IoT applications is time-consuming and difficult [6]. The term ‘Usable Privacy’ means we emphasize both privacy and usability equally. More specifically, we highlight the importance of putting the user (in this case, the developer) at the center and consider their requirements, constraints, priorities, and needs when designing a tool.

**TARGET ROLES AND AUDIENCE**

We believe such a tool can be beneficial to different types of stakeholders as follows:

**Design Tool for Software Engineers (designers/architects):** Primary stakeholders of this tool would be software engineers. We expect them to use this tool to sketch their potential IoT application designs and get validated before moving to the implementation phase. This tool will provide different types of suggestions which engineers can use to improve their IoT application designs in terms of privacy. More importantly, we do not expect this tool to act as a black box that just spits out application designs. Instead, each recommendation will be justified by the tool so engineers can understand why the tool is making a certain recommendation.

**Compliance Tool:** This tool will also be useful to demonstrate certain compliance needs (e.g., GDPR [4]). It will have the capability to automatically generate a compliance report for each IoT application design, briefly explaining the design decisions and risks associated with them so that the compliance officers can determine whether to approve or not. We envision that, in the future, such a compliance tool could be useful (or maybe required) when submitting an IoT application to IoT app stores.

**Education and Awareness Tool:** We also expect this tool to be used for enhancing privacy awareness among students from school level to university level. Over the last few years, many programming environments and languages have been developed to help young children learn how to code (tynker.com, scratch.mit.edu). Similarly, we believe that everyone should learn about privacy at a young age, so over time, they will become responsible software engineers who care about privacy. A Privacy Mindset [6] is important to be developed by software engineers.

**TOOL-ASSISTED PRIVACY-AWARE IoT APPLICATION DESIGN**

The tool we propose is something that the engineering community has not seen before. It is inspired by many existing tools used by the engineering community (e.g., UML design tools). First, let us illustrate how the proposed tool (and underlying technology) is expected to work in practice using Fig. 2.

**(Step 1)** Software engineers will draw their application designs using a pre-defined set of notations. Key components will be nodes (device profiles) and data flows. To ease the process, common device profiles will be provided. This process will look like a UML diagram design process. (Step 2) Engineers will then specify the service which they plan to run. (Step 3) They can either assign each service to a node or just leave them unassigned for the algorithms to do that in a later step. (Step 4/5) Optionally, engineers can provide additional information related to data management (e.g., 90 days of data retention) and context (e.g., healthcare domain). Additional information will help the algorithms to better design IoT applications. The rest of the steps are invisible to engineers and triggered by a single click. (Step 6) Algorithms automatically assign each service into nodes appropriately by considering device capabilities, runtime requirements of the services, and other relevant context information. (Step 7) Algorithms incorporate privacy protection features into the design. This step may also reassign the services into different nodes, if necessary. This is one of the key features of this tool. (Step 8) Algorithms examine the privacy awareness at both node and composition levels. Then, all the results will...
be combined to produce the overall privacy index and present it to the engineers. Engineers may consider changing their initial designs to improve the privacy index. (Step 9) The terms and conditions unique for each IoT application design are automatically generated.

So far, we designed the IoT application assuming the target environment is static. This works sufficiently for use cases like above. However, some types of applications would require adaptation at run time to better serve the users. In reality, IoT environments are highly dynamic. Therefore, IoT applications should be able to adapt at run time. Toward this direction, we believe this tool should provide simulation capabilities so the engineers can evaluate how their application might adapt at runtime under different circumstances. Currently, several tools are being developed within the context of security, such as OWASP Threat Dragon (threatdragon.org), Microsoft Threat Modeling Tool, IriusRisk (iriusrisk.com), and CAIRIS (cairis.org). However, none of them explicitly focus on privacy.

**STATE OF THE ART**

Our vision lies at the intersection of IoT, software engineering, and human-computer interaction, where the primary focus would be on usable privacy.

**Software Engineering:** Several existing frameworks have been proposed to help elicit privacy requirements and to design privacy capabilities in systems. The original privacy-by-design framework was proposed by Ann Cavoukian [7]. This framework identifies seven foundational principles that should be followed when developing privacy-sensitive applications. Building on the ideas of engineering privacy by architecture versus privacy-by-policy presented by Spielerman and Cranor [8], Hoepman [9] proposed an approach that identifies eight specific privacy design strategies: minimize, hide, separate, aggregate, inform, control, enforce, and demonstrate. In our previous work [10], we dissected Hoepman’s [9] high-level strategies into a more prescriptive granular set of 30 guidelines. The STRIDE framework [11] was developed by Microsoft to help software engineers consider security threats. It is an example of a framework that has been successfully used to build secure software systems by the industry. In a similar vein, LINDDUN [12] is a privacy threat analysis framework that uses data flow diagrams (DFD) to identify privacy threats. LINDDUN focuses on eliminating a set of pre-identified privacy threats using a systematic review of data flow diagrams. The LINDDUN approach requires engineers to draw DFD diagrams and manually evaluate them to identify privacy issues using the instructions provided. This type of approach is useful but requires a significant amount of time, effort, and some level of privacy knowledge and expertise. Due to human involvement, it is also prone to errors and lacks consistency, and is therefore less trustworthy. All the above-mentioned approaches are not focused on design-time aspects where IoT architectures are highly fluid (change over time), and therefore need to support run-time adaptation. Privacy (and security) patterns [13] have also been developed to facilitate privacy-aware application designs for web and mobile domains. Some notable work is done by the EU FP7 funded PRIPARE 2, where they have developed privacy patterns and categorized them using Hoepman’s [9] eight strategies. Though such patterns are useful, usage is limited due to manual work that needs to be done (also requires some expert knowledge) by engineers in order to apply them in real-world application designs.

**Human-Computer Interaction (HCI):** HCI techniques are used to increase the privacy awareness of both engineers and end-users. Luger et al. [14] aim to make emerging European data protection regulations more accessible to the public by using a series of privacy ideation cards. They have extracted 40 design principles by examining EU General Data Protection Regulations. These high-level principles are similar to the guidelines and strategies mentioned before but in a visual format. Based on previous efforts on nutrition, warning, energy, and standardized banking privacy notifications, Privacy Label [15] has been proposed as a method to communicate and understand privacy policies using a visual medium. Terms of service are often too long to read and are ignored by end-users. Terms of Service; Didn’t Read (ToS; DR) (tosdr.org) aims to address this issue by introducing a rating and classification scheme. It looks for pre-defined factors in a ToS document to generate

---

2 www.privacypatterns.eu
a rating. PrivacyGrade.org also addresses the same challenge by developing a smart assistant chatbot that can answer questions regarding the privacy policies of different websites. PrivacyGrade.org proposes a method to grade smartphone apps from a privacy perspective by evaluating permission requirements. Privacy Bird (privacybird.org) is a tool that allows end-users to find out what web sites will do with their data by reading privacy policies written in the Platform for Privacy Preferences (P3P) standards. Mozilla has proposed privacy icons that aim to replace (or complement) text-based privacy policies with a visual and consistent set of icons. Naeini et al. have explored privacy expectations and preferences in the IoT domain. Their results show that end-users’ privacy expectations vary based on contextual factors. Further, end-users prefer to be informed about data management practices. End-users become comfortable in sharing data when honestly communicated. All the approaches mentioned above consider privacy as an afterthought and look at the problem from a third-party point of view. The similarity is that they all propose mechanisms to look at privacy as an afterthought (i.e. looking at a finished product or a service). Further, none of these approaches focus on IoT applications, which are complex in nature and have challenging characteristics such as distributed architecture and unpredictability at runtime.

**Benefits and Impact of the Tool**

**Assisting Good Actors:** It is important to note that the tool we envision is not aiming to stop bad actors with malicious intent. It is designed to assist good actors in designing better privacy-aware applications. More importantly, we aim to assist software engineers (SMEs), freelance developers, and IoT hackers/makers who cannot afford to hire privacy engineering expertise or learn privacy engineering capabilities due to practical cost and time constraints. Further, we do not aim to develop a perfect tool (in the short term) that produces perfect results and eliminates human intervention. We intend to highlight and demonstrate the importance of providing useful and usable tools for good actors to do their job better.

**On the Job Privacy Education:** We believe that such a tool will help novice (in terms of privacy) software engineers to learn privacy challenges and alternative design by design strategies on-the-go without investing significant time and cost on learning them separately.

**Facilitating Constructive and Critical Discussion through Transparency:** Such a tool will allow software engineers (and other stakeholders) to think and act more transparently. It will demonstrate the importance of providing useful and usable tools for good actors to do their job better. It will help novices (in terms of privacy) software engineers to design IoT applications by manipulating program elements graphically. We expect such a design process would expedite the step 6 and step 7. For example, without starting from the scratch, algorithms can pick a similar solution to the problem at hand (i.e., IoT design drawn by the developer) from a library and revise it appropriately.

**Automated planning and scheduling:** These techniques can be used to implement step 6 and step 7 as well. We can convert the problem of integrating privacy protection measures (into an IoT application design) to an AI constraint solver problem. Then, we can use existing AI planning algorithms and frameworks such as Optaplanner (optaplanner.org) to generate new plans (i.e., new designs).

**Explainable AI:** For example, step 9 can benefit from explainable AI techniques. Developers are not only interested in the end outcome of the tool (i.e., designs that shows how to apply privacy-protecting measures), but also the reasoning and justifications of the recommendations provided by the tool. It is important to tell the developer why it is better to apply ‘minimisation’ at the first node than applying it at a later node. Explainability is really important to convince engineers with a significant amount of experience where they do not like to follow a tool without proper reasoning or justification.

**Economic Impact:** Regulatory entities increasingly introduce new laws and regulations that businesses need to follow. Such adherence requires privacy expertise which often needs to be acquired through consultancy services. Such costs are significant for SMEs due to their limited budget. Our technology will help SMEs to reduce (or eliminate in the long term) privacy consultancy. Privacy-by-design will also help SMEs to reduce costs in the long term (e.g., re-engineering cost, privacy violations and legal costs, and loss of customers). It is important to note that GDPR fines can go up to 20 million Euros or 4 percent of annual global turnover, whichever of both is highest. Especially, such stiff financial penalties could significantly impact SMEs.

**Social Impact:** Such a tool can impact society in two different ways. First, it will create a benchmark (i.e., privacy assessment report) to facilitate privacy by design decisions among engineers. Second, the same benchmark (in the Terms and Conditions) will help end-users (i.e., consumers of the IoT applications) to understand and choose IoT products based on transparent and trustworthy T&C.

**Research Directions**

**Design Notations and User Interactions**

We envision this tool to follow the visual programming paradigm [18]. The proposed tool is expected to be used by engineers to design IoT applications by manipulating program elements graphically. We expect such a design process would be natural for engineers as they are typically familiar with design approaches such as Unified Modelling Language (UML) and Data Flow Diagrams (DFD). Such familiarity will help engineers to familiarize themselves with the tool quickly. Ideally, the visual programming language will be inspired by the data flow diagram notations. A data flow diagram is a graphical representation of the ‘flow’ of data through a system (in this case, an IoT system), modeling its process aspects. However, it is important to note that DFDs are flexible enough to be represented in different levels of complexities. Therefore, it would be a fine balance between maintaining simplicity while allowing engineers to customize their systems in detail. In addition to the DFD based information, the tool should have ways to gather other related contextual information. The tool should show errors/warnings when vital pieces of information are not provided by the engi-

---

2. https://wiki.mozilla.org/Privacy_ICons
neers (e.g., data retention period). Knowledge-bases can be used to provide assistance and recommendations for the engineers (typical data retention period based on the domain, data types, applicable laws, and so on) [19].

**Privacy Patterns and Knowledge Modelling**
Incorporating privacy-preserving techniques into IoT applications is a complex and time-consuming process [6]. Traditionally, in software engineering, such complexities are handled by introducing design patterns. Design patterns are general repeatable solutions to commonly occurring problems. Design patterns can also speed up the design and development process by providing tested and proven solutions. We believe that this tool should be knowledge-driven. This means that algorithms should not require constant upgrades while the knowledge-bases will grow overtime, enabling new features and capabilities. To achieve this, we propose to create a privacy pattern library (by both developing new privacy patterns and organizing existing privacy patterns). Ontology-based knowledge models can be developed to model information about each privacy pattern in a common structure. Such a common structure and semantic interoperability allow algorithms to manipulate patterns in a semantically meaningful way. Pattern candidates need to be extensively analyzed to find out their characteristics (e.g., usability, complexity, abstractness, relationship to other patterns, and composability) and to categorize them from different perspectives (e.g., functionality and level of granularity). Such an analysis would be vital in the next phase.

**Context-Aware Planning and Adaptation**
The design and development of IoT applications require both software and hardware to work together across multiple different types of nodes (e.g., micro-controllers, system-on-chips, mobile phones, miniaturized single-board computers, cloud platforms) with different capabilities under different conditions (e.g., CPU, memory, energy, data communication, knowledge availability, energy limitations, latency tolerance limitations, domain requirements). Therefore, the privacy-preserving techniques that can be applied on a given node vary depending on the context. The question that needs to be answered is ‘How do we optimally allocate responsibilities to each node based on the context when designing a privacy-aware IoT application?’ First, we need a knowledge base that can be used to reason about different IoT application design choices. Second, we need algorithms that can decide which privacy patterns are to be used in different nodes individually and as a whole. Toward this, techniques developed by the web service composition community could be useful. This challenge could be addressed by formulating it as a service composition (with constraints) problem [20].

The IoT ecosystem is highly dynamic. Therefore, IoT applications should be able to adapt to the context changes at run time. However, it is difficult to predict how such adaptation would work at runtime. We believe that the proposed tool should be able to provide engineers with some insights on how their applications would adapt. Let us consider a home care example scenario. Assume three IoT systems are deployed in a home as follows: 1) care receiver wears a smart band that tracks health; 2) a smart bed that can adjust itself, 3) a smart carpet that tracks movements. Assume that none of these IoT systems is originally designed to detect falls and notify caregivers. For example, a novel IoT application may be originally developed to detect falls by using smart band data. The challenge is, how should such an IoT application behave if the smart band fails (e.g., hardware failure)? Can the IoT application adapt and reconfigure itself based on context? (e.g., the IoT application won’t be able to use data from the smart band and smart carpet to detect falls, instead of the failed wrist band). As the IoT application changes at runtime, privacy-preserving measures may also need to change accordingly to support the adaptation. The tool should be able to simulate scenarios to evaluate the quality of the application design as well as its adaptability.

**Operationalization, Measuring, and Rating**
Finally, the challenge is how engineers know, given an IoT application design, whether it is a good design or a bad design (from a privacy perspective). We tend to understand different types of measuring and rating/indexing techniques well (e.g., Body Mass Index, energy ratings, food reference intake, and so on). However, no such mechanism is available to measure the privacy awareness of IoT applications. We believe such mechanisms (e.g., privacy index) would be increasingly important for both engineers and end-users. Engineers can use such an index to evaluate their applications to improve their designs iteratively. End-users can use such an index to understand how each IoT application manages their data. The operationalization of privacy is a challenging task. There are many factors to take into account when generating a privacy index for a particular application design. Some of the major factors are: 1) privacy patterns used (individually and compositions), 2) order of privacy patterns applied, 3) sensitivity of the data involved, 4) potential risks, and so on. After operationalizing, these factors need to combine in a meaningful way. Crowdsourcing techniques [21] may be used to combine expert knowledge and end-user expectations to generate a privacy index. The data flow diagrams drawn by developers can be converted into a semantic network that comprises nodes and edges (e.g., semantic web model). We can then develop algorithms to traverse through the graph and determine what kind of privacy-preserving measures are being incorporated into the design and where (e.g., which part of the architecture). A privacy index can be generated based on this information. For example, if a ‘minimization’ privacy-protecting measure is applied earlier (i.e., within the first node) within the architecture (Design 2 in Fig. 1), it will get a higher privacy index. Similarly, Design 1 (in Fig. 1) will get a comparatively lower index as it applies ‘minimization’ within the third node. Likewise, we can define different types of rules that can be automatically checked to generate a privacy index, which will then be combined into a final privacy index.

Another challenge that goes hand in hand with rating is Terms and Conditions (T&C). It is a well known fact that end users hardly ever read T&C [15] related to any product or service, besides IoT applications. Typically, T&C are written as piles of text and therefore are difficult for a human to understand the most important information within the end-users’ short interest and attention span. From engineers’ point of view, putting together a T&C document is also a time-consuming task that requires a lot of effort and specialist expertise (e.g., legal professionals) and also less trustworthy (due to human involvement). We propose to capture and model privacy expert knowledge using knowledge-based AI techniques so that the algorithms can eliminate the necessity for privacy experts and related human errors. Such knowledge could be used to automatically generate the T&C based on the design of the IoT application.

**References**


**Biographies**

Charith Perera (charith.perera@ieee.org) is a senior lecturer at Cardiff University, UK. He received his B.Sc. (Hons.) in computer science from Staffordshire University, UK, an MBA in business administration from the University of Wales, Cardiff, UK, and a Ph.D. in computer science at The Australian National University, Canberra, Australia. Previously, he worked at the Information Engineering Laboratory, ICT Centre, CSIRO. His research interests are the Internet of Things, sensing as a service, privacy, middleware platforms, and sensing infrastructure. He is a member of the IEEE and ACM.

Mahmoud Barhamgi (mahmoud.barhamgi@univ-lyon1.fr) is an associate professor of computer science at Claude Bernard University Lyon 1, France. His research focuses on security and privacy preservation in service-oriented architectures, web, and cloud environments. He received a Ph.D. in information and communication technology from Claude Bernard University Lyon 1.

Massimo Vecchio (mvecchio@fbk.eu) is an associate professor at the eCampus University and senior researcher within the OpenIoT Research Unit of the Fondazione Bruno Kessler (FBK). His current research interests include computational intelligence and soft computing techniques, the Internet of Things paradigm, and effective engineering design and solutions for constrained and embedded devices. Regarding his most recent editorial activity, he is an associate editor of the Applied Soft Computing journal and the managing editor of the IEEE IoT newsletters.