Alphanumeric Data: Minimising Privacy Concerns in Smart Environments

Dennis Lupiana, Rose Tinabo, Fredrick Mtenzi, *Ciaran O'Driscoll & Brendan O'Shea
School of Computing
**School of Electronic and Communications Engineering*
Dublin Institute of Technology

Abstract

Users' privacy concerns have been a major obstacle to the successful realisation of the research in Smart environments (SmEs). SmEs are computing environments that proactively and seamlessly support users. In order for a SmE to proactively support users within its surroundings, it must possess knowledge about users' activities. To acquire this knowledge, a lot of information about users needs to be collected. In addition, despite feeling they are out of control due to seamless operations in SmEs, users often feel intimidated by physical monitoring through video cameras. These factors raise a major concern for users' privacy. This article proposes an alternative approach for acquiring knowledge about users' activities without jeopardising their privacy. A description of a solution for identifying users and locations is provided. The solution uses devices' and sensors' IDs to automatically identify users and locations. Since IDs are composed of alphanumeric data types, without actual users' identities, the proposed solution will have a significant impact on minimising users' privacy concerns toward SmEs.

1. Introduction

Smart environments (SmEs) are computing environments that offer decision making capabilities to proactively and seamlessly support users' routine tasks. The fundamental aim of SmEs is to free human cognitive ability from the hassle of manual operations of devices, which keeps on increasing at unimaginably speed as the days pass. Additionally, SmEs leverage devices within users' proximity to offer users with appropriate services and information.

In SmEs users rarely experience interfaces when interacting with devices; almost every operation is performed in a "blackbox" leaving users not in control. In addition, in order for SmEs to proactively and seamlessly support users, a lot of users' information need to be collected. Often, a lot of this information is irrelevant to what is required to drive decisions of the environment. For instance, unless the SmE is for recommending different teaching modes to lecturers, collecting information about student dozing in class will be trivial. It is even more problematic when video cameras are used for collecting users' information. Video cameras not only collect unnecessary information, but also can intimidate users. These factors jeopardise users' privacy and therefore despite the benefits SmEs have to users, their acceptance is still very low.

Most importantly in SmEs, however, is the knowledge about users' activities. Therefore, if users' activities are explicitly represented, then users' identities and other spatial information are sufficient to drive SmEs' decisions in supporting users. This article proposes an alternative approach for acquiring knowledge about users' activities. A detailed description of a solution for identifying users and locations is provided. This solution has a significant impact on improving users' attitude towards SmEs. Instead of physically monitoring users, the proposed solution monitors users through their devices' ID, which are typically alphanumeric data. This solution reduces the amount of personal identifiable information (PII) to be collected and hence minimises users' privacy breach.

The rest of this article is organised as follows; Section 2 provides a background on SmEs highlighting why privacy is a major concern. Section 3 provides a background on privacy in UbiComp, analysing existing definitions and highlighting driving factors. Section 4 reviews some of the existing solutions on ensuring privacy in UbiComp. Section 5 provides an analysis and discussion of the review of the existing solutions to privacy, pointing out unique features of the proposed solution. The design, consequences and improvements of the proposed solution is provided in section 6. The conclusion and future work is provided in section 7.

2. Smart Environments

Smart Environments (SmEs) are highly sensory and device integrated computing environments that responsively and proactively support their users [1]. SmEs are the successors of UbiComp environments. Unlike UbiComp environments, which provide physical infrastructures for device connectivity in support of users, SmEs add intelligence capabilities to physical environments. Therefore, instead of offering physical connectivity to devices, SmEs offer decision making capabilities to proactively and seamlessly configure and operate users' devices, and provide appropriate services and information according to their working conditions.

In Europe, a similar research is referred to as Ambient Intelligence (AmI); devoted on implementing novel systems to responsibly support users in their environments [2, 3]. The principal aim of SmE, or AmI, is to proactively and seamlessly assist users in their daily routines and therefore freeing their cognitive abilities for other brain-demanding tasks. In particular, research in SmEs is addressing generic questions such as; why should human continue to struggle with manual overheads of their routine tasks while technology has become part of their life?

Unlike in the early days, where computers were locked inside a particular room, nowadays we carry computers wherever we go. Most importantly, these computers, or devices as often referred, are capable of observing their surroundings, and some can even collect, process and give feedback. In addition, with the improvement of research in sensors and wireless networks, it has now become possible to observe and interact with our physical environments. Therefore, with these capabilities, research in SmEs is devoted to empower our environments with intelligence capabilities to respond to our working conditions and utilise existing devices to support us.

Research challenges in SmEs include acquisition, representation, reasoning and sharing of semantics about our daily routines. The goal is to enable our physical environments to reconstruct whatever we have been doing in order to proactively support us. For instance, if user $\{U_1, U_2,...,U_n,\}$ are in location L_l at time T_t, having a meeting, then the SmE should be able to figure-out whenever the similar scenario occurs to provide appropriate services. Typically, devices in a SmE are self-configured when entering a meeting venue and appropriate files are sent to the meeting audience. However, to acquire such an understanding, a lot of information about users must be collected and therefore jeopardising their privacy.

3. Privacy and UbiComp

Privacy has been largely discussed and still is a major issue in UbiComp [4-5]. Among the features that have been associated with privacy in UbiComp is the pervasiveness and invisibility of computing resources of which sensing and monitoring is at the core [6, 23]. Since the research in SmEs is at the forefront of accomplishing invisible availability of computing resources, inspired by UbiComp technologies, therefore SmEs inherit security and privacy issues of UbiComp.

In order to discuss privacy in the context of UbiComp it is necessary to know what privacy is. However, defining privacy is a challenging task and as a result there is no a single agreed definition of privacy among researchers. Few reasons have been associated as to why defining privacy is difficult.

According to Westin [7], privacy is a relative concept and therefore it is difficult to generalise its definition. This implies that whatever A may consider as confidential, it may not be to B. Many researchers [5, 6, 8, 9] refer to Westin's view of privacy [7] as their starting point for discussing privacy and an indication of the difficulty in arriving at a definition of privacy.

"*no definition [of privacy]… is possible, because [those] issues are fundamentally matters of values, interests and power*" [7].

Similarly, Bellotti and Selen [10] conclude that privacy definition has to be dynamic. It has to consider a situation of an individual, and technological and cultural changes. For instance, in a medical emergence the invasion of personal information could mean saving someone's life, and therefore the benefits outweigh the potential risks of intrusion. In addition, the emergent of social networking technologies such as Facebook, has changed people's attitude to privacy; people are now willing to publish their personal information.

In line with the dynamic view of privacy definition [10], Palen and Dourish [11] define privacy as a "*process of negotiation boundaries in relation to disclosure, identity and time*", and is an organic process that *continually changes* over time. Although Palen and Dourish [11] have not articulated a specific factor to be considered, like Ballotti and Selen [10], they refer to privacy as the *dynamic* process. This view is referred by researchers [9, 12] and is considered to be a suitable basis of a design approach for privacy-sensitive applications.

Privacy has been considered by many researchers and in particular how people regard privacy has been widely researched. O'Driscoll [13] presents a detailed review of privacy in the context of UbiComp. From this review it is clear that there are many contributing factors that people consider in relation to privacy in UbiComp environments. In particular people are concerned with who will use or have access to personal information collected in the environment, how will this be used and how sensitive is the information. Also the situation in which the information will be used is also a major concern for people.

Unfortunately, we cannot convince users otherwise regarding UbiComp environments and in particular SmEs. Therefore a better technological solution is required to intelligently support users without jeopardising their privacy. However, prior describing our solution let us first review what others have done to ensure privacy in UbiComp.

4. Related Work

Langheinrich [6] proposes a set of privacy principles to be adopted in UbiComp environments in order to ensure privacy. These principles are based on the well known Fair Information Practices; notice, choice and consent, proximity and locality, anonymity and pseudonymity, security, and access and recourse [14]. The practices have also been adopted as general rules for the development of privacy enhanced UbiComp systems such as the European Disappearing Computer Privacy Design Guidelines [15]. The approaches range from abstract frameworks to specific protocols and technologies [16].

To emphasise the proposed principles, Langheinrich [5] proposes Privacy Awareness System in order to provide a sense of accountability, rather than to provide security and privacy guarantees, which is difficult to be achieved. In this approach the system notifies users when collecting their data. In addition, the system implements adequate privacy policies for the collected data, and also provides essential means to inform users on how their personal data is being processed.

Jendricke et *al.* [4] propose context-driven identity management as another approach for ensuring privacy in UbiComp environments. This identity management approach enables users to express and enforce a preferable level of privacy depending on the situation which they are involved in. Similar to the everyday life, the identity manager (which operates in each user's device) allows the device to present different subsets of the user's identity depending on the perceived context.

On the other hand, Zugenmaier and Hohl [17] propose the use of encryption or omitting any reference of a user from a content of a message in order to protect the user's identity from anyone who is able to access the content of the message. Zugenmaier and Hohl [17] admit the difficult of masking the source and destination addresses of the message and therefore suggests alternative approaches to ensure privacy in UbiComp environments. These approaches includes: the use of DC-network approach [18, 19]; Mix concept [20] such as onion Routing, SG mixes, Web mixes and Crowds; to replace user's device ID with a temporary address and "classical" anonymising techniques such as using internet cafes [17].

5. Analysis and Discussion

Most of the proposed solutions on privacy concerns operate on a broader view of UbiComp (computing resources everywhere). However, it is a challenge to quantify what is computing everywhere? What distance interval should be considered to conclude if computing resources are available everywhere? Or should we consider our physical objects embedded with computing power within a single room as computing everywhere?

Considering the previous computing paradigms, however, it is plausible to argue that the availability of computing resources in UbiComp should not be limited to a single room or at least to a single orientation. In the UbiComp paradigm, users should be able to work elsewhere supported by myriads of interconnected devices [21]. Therefore if not in the entire city, country, or world then at least computing resources should be available in a workplace. From our point of view, this is the operating definition of UbiComp in the research of SmEs.

Therefore although users are still kept in the dark from nonintrusive operations of SmEs, at least they have a *private zone*. In particular, if users are identified through their devices' IDs, which are typically alphanumeric, will drastically reduce PII to be collected and hence significantly minimise users' privacy concerns. Although there are some occasions where users wish to be completely invisible from their colleagues, these requirements are too personal and often outweighed by the benefits SmEs offer in an organisation.

Additionally, Langheinrich proposals [5, 6] are implicitly observed in SmEs because the whole design process is user-centred. Therefore users are aware of the collection of their information within the SmE. It is even more feasible when using devices' IDs to identify users because the environment will only be capable of monitoring the known users. However, using video cameras as users' identification mechanism does not isolate guest users from the known users and therefore it raises privacy concerns.

Therefore moving away from physical monitoring of users is a plausible solution. However, apart from reducing PII to be collected by monitoring users through their devices, the proposed solution also adopts pseudonymisation technique to protect devices' IDs. Therefore, the system can securely use devices' IDs within a workplace while giving users an option to turn off their devices' Bluetooth or protect their RFID tags from being read whenever they leave the building. Already technology exists to control the accessing of RFID tag information [22].

6. Proposed Solution

Fundamentally, of most importance in SmEs is the knowledge about user's activities in order to effectively support them. More often, this knowledge is acquired through physical monitoring of users – typically using video cameras and audio sensors. This approach has significant impact on security and users' privacy. Therefore, an alternative approach is

required to reconstruct these activities without physically monitoring users. In the School of Computing at the Dublin Institute of Technology, Ireland we are devoted to design of such a solution.

6.1. Solution Design

The unique feature of our solution is the indirect monitoring of users. Instead of video and audio recording, which can be so intimidating to users, we propose to monitor users through their sensor-enabled mobile devices. In this solution, we use radio frequency identification (RFID) and Bluetooth technologies to automatically identify users (Device ID_x = $User_x$) and locations (Sensor ID_x = Location $_x$). We also separately represent human activities and present such knowledge as a knowledge model. Through this approach, users' identities, location and time information will be used to reconstruct human activities and therefore to effectively support them.

reduces users' discomfort towards SmEs and hence it may increase acceptability of SmEs in the society.

6.2. Design Consequences

In most cases, an RFID tag is imprinted with PII such as the name of the user and therefore anyone with appropriate RFID reader within such environment can access such information. Similar problem can be experienced in Bluetooth-enabled devices. However, since devices are resource rich compared to passive RFID tags, it can be solved through imposing security mechanisms such as encryption or pseudonymisation algorithms.

Since our solution is using passive RFID tags which have limited computing resources, the only possible solution is not to imprint users' names in the tags.

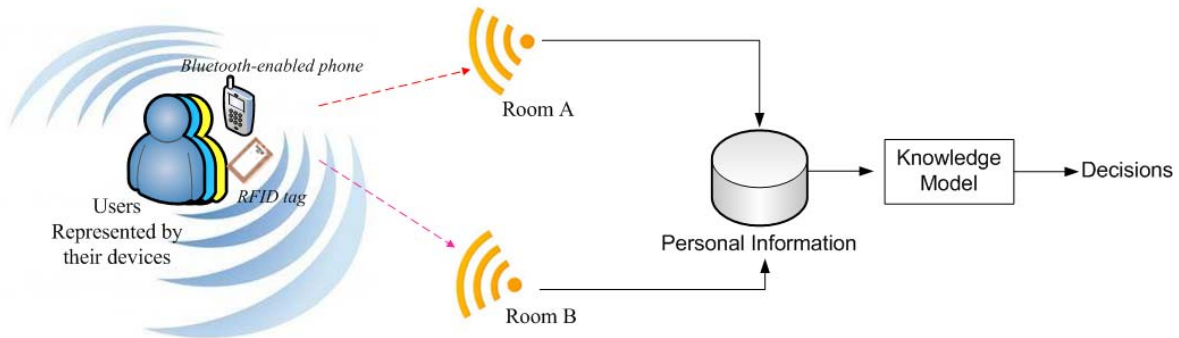

Figure 1. Conceptual Design of the Proposed Solution

As shown in figure 1, when a device is detected by a particular sensor embedded in a physical environment, the captured IDs are mapped to personal and location information to determine the device's owner and his current location. The user and location information will then be used to reconstruct user's activities to proactively support him. Therefore, depending on who is detected, where and when the SmE, through the knowledge model and other mechanisms, will be able to figure-out what the user is supposed to be doing and offer appropriate services. However, since our interest is on security of personal information we have not shown location information in the diagram.

Unlike video and audio data, which is rich in personal identifiable information (PII), devices' IDs are typically alphanumeric data and therefore limit the amount of PII to be collected. In addition, the separation between personal information and the knowledge about users' activities reduces the severity of security breach which subsequently reduces the risks of users' privacy breach. Furthermore, because users are not physically monitored, this solution also Therefore the tags will only have their unique IDs which will be associated to users. Through this approach, the tag user can only be identified through mapping device's ID and the personal information. This approach reduces the risk of users being directly identified by intruders.

However, the challenge remains as to how to protect devices' IDs from intruders. It is more challenging because the mapping between a device ID and user is one-to-one; always the same ID will be mapped to the same person. With this approach an intruder can easily access and link device's ID to a particular user and therefore access his personal information and even impersonate him.

6.3. Solution Improvements

To improve this solution, therefore, we propose to impose a security layer on both device's ID and personal information, see figure 2. However, since our solution is using passive RFID tags which are resource limited, the security layer on device's ID will be implemented on Bluetooth-enabled devices and in particular mobile phones.

Like many other mobile devices, computation power is limited and therefore the security measures required should be light. This is particularly essential in SmEs because response time is an essential factor to be considered. Additionally, in SmEs it is important to identify users in order to support them and therefore the security measures required must be retraceable. In most cases encryption algorithms require high computation power and are time consuming. In addition, managing encryption keys is so challenging. Similarly, anonymisation techniques do not establish a link between the anonymised data and its original source.

In our solution, therefore, we are proposing to use pseudonymisation technique. In this approach, a pseudonym generation algorithm will be developed and deployed in a centralised server along with other system's modules. In addition, an application will be developed and installed in users' mobile phones to randomly use the generated pseudonyms. The algorithm will be generating unique pseudonyms of devices' IDs after a certain period of time. A copy of the generated pseudonyms will then be transferred to the corresponding mobile phones to be used by the application. All these operations are underground leaving the user to concentrate with other things.

The idea is to have multiple authentication of the user through a Bluetooth-enabled mobile phone and an RFID tag when sensitive services and information are to be provided. In this solution, the RFID tag will be used to provide users with services and information. In the case of a sensitive service or information is requested, the user will be authenticated via his Bluetooth-enabled mobile phone. In occasions where two identical users – legitimate and intruder – are observed, the environment can differentiate them by sending a confirmation request to be acknowledged by the legitimate user.

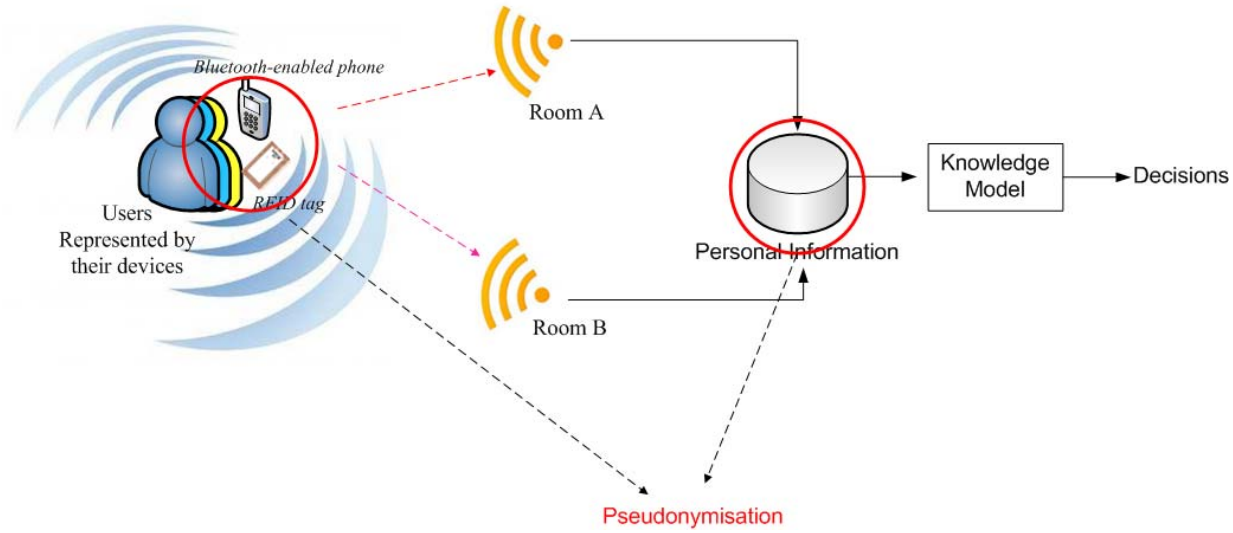

Figure 2. Conceptual Design of the Solution Improvements

7. Conclusion and Future Work

In this article the problem of users' privacy concerns in Smart environments is discussed. The collection of users' information has being identified as the major issue. This article proposes a solution to reduce the amount of personal identifiable information to be collected. The article proposes to indirectly monitor users through their devices. The proposed solution has a significant impact on users' attitude toward Smart environments. Despite reducing the amount of information to be collected, the proposed solution provides an alternative approach to physical monitoring, which is often intimidating to users. In addition, the proposed approach provides a solution to secure the devices' IDs and hence enables users to be securely supported in Smart environments.

The next phase of this work is to implement and validate the proposed solution by building the knowledge model and developing other mechanisms which are essential for facilitating decision making

in Smart environments. In addition, the focus will be on developing the pseudonym generation algorithm and the client application to use them. Already the work on building knowledge model has started. It will also be interesting to learn how the knowledge engineering techniques can be applied to enhance users' authentication process in Smart environments.

8. References

[1] D. Lupiana, C. O'Driscoll, and F. Mtenzi, "Defining Smart Space in the Context of Ubiquitous Computing," Ubiquitous Computing and Communication Journal (UbiCC), vol. 4, pp. 516-524, 2009.

[2] J. Augusto, "Ambient Intelligence: The confluence of ubiquitous/pervasive computing and artificial intelligence," Intelligent Computing Everywhere, pp. 213-234, 2007.

[3] K. Ducatel, "Scenarios for ambient intelligence in 2010," Office for Official Publications of the European Communities, 2001.

[4] U. Jendricke, M. Kreutzer, and A. Zugenmaier, "Pervasive privacy with identity management," 2002.

[5] M. Langheinrich, "A privacy awareness system for ubiquitous computing environments," UbiComp 2002: Ubiquitous Computing, pp. 315-320, 2002.

[6] M. Langheinrich, "Privacy by Design – Principles of Privacy-Aware Ubiquitous Systems," presented at Proceedings of the Third International Conference on Ubiquitous Computing (UbiComp 2001), Atlanta, USA, 2001.

[7] A. F. Westin, Privacy and Freedom: New York NY: Atheneum, 1967.

[8] J. I. Hong and J. A. Landay, "An architecture for privacy-sensitive ubiquitous computing," in Proceedings of the 2nd international conference on Mobile systems, applications, and services. Boston, MA, USA: ACM, 2004.

[9] J. I. Hong, J. D. Ng, S. Lederer, and J. A. Landay, "Privacy risk models for designing privacy-sensitive ubiquitous computing systems," in Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques. Cambridge, MA, USA: ACM, 2004.

[10] V. Bellotti and A. Sellen, "Design for privacy in ubiquitous computing environments," presented at Proc. Third European Conference on Computer-Supported Cooperative Work ECSCW'93, Milano, Italy, 1993.

[11] L. Palen and P. Dourish, "Unpacking Privacy" for a Networked World.," presented at Computer Human Interaction, CHI 03, Florida, 2003.

[12] G. Iachello and G. D. Abowd, "Privacy and proportionality: adapting legal evaluation techniques to inform design in ubiquitous computing," in Proceedings of the SIGCHI conference on Human factors in computing systems. Portland, Oregon, USA: ACM, 2005.

[13] C. O'Driscoll, "Privacy in context: Privacy issues in Ubiquitous Computing applications," presented at Third International Conference on Digital Information Management (ICDIM) 2008, 2008.

[14] OECD, "Recommendation of the council concerning guidelines governing the protection of privacy and transborder flows of personal data.," 1980.

[15] S. Lahlou and F. Jegou, "European Disappearing Computer Privacy Design Guidelines V1. 1," Ambient Agoras IST-DC, 2004.

[16] S. Dritsas, D. Gritzalis, and C. Lambrinoudakis, "Protecting privacy and anonymity in pervasive computing: trends and perspectives," Telematics and informatics, vol. 23, pp. 196-210, 2006.

[17] A. Zugenmaier and A. Hohl, "Anonymity for users of ubiquitous computing," 2003.

[18] D. Chaum, "The dining cryptographers problem: Unconditional sender and recipient untraceability," Journal of Cryptology, vol. 1, pp. 65-75, 1988.

[19] M. Waidner, "Unconditional sender and recipient untraceability in spite of active attacks," 1990.

[20] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," Pervasive Computing, IEEE, vol. 2, pp. 46-55, 2005.

[21] M. Weiser, "The computer for the 21st century," Scientific American Vol. 265, No. 3, Sept. 1991, pp94-104, (Reprinted in Communications of ACM July 1993), vol. 3, pp. 3-11, 1991.

[22] N. Marquardt, A. S. Taylor, N. Villar, and S. Greenberg, "Visible and controllable RFID tags," presented at Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, 2010.

[23] D. Lupiana, F. Mtenzi, C. O'Driscoll, and B. O'Shea "Strictly alphanumeric data: Improving privacy in smart environments", Internet Technology and Secured Transactions (ICITST) 2010 International Conference, pp. 1- 3, IEEE, 2010

