The classification of algorithms for Privacy Management in Ubiquitous Environments

Valderi R. Q. Leithardt [1] [2], Carlos O. Rolim[1], Anubis G. M. Rossetto[1], Guilherme A. Borges[1], Jorge Sá Silva[2], Claudio F. R. Geyer[1]

[1]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

[2] Universidade de Coimbra, Sistemas Largamente Distribuídos, Departamento de Engenharia Informática (DEI) _ Pólo II - Coimbra - Portugal

{vrqleithardt, carlos.oberdan, agmrossetto, gaborges, geyer}@inf.ufrgs.br, sasilva@dei.uc.pt

Abstract. *Control and data management in ubiquitous environments is not a trivial activity owing to the heterogeneity of the users, applications and devices, required to exchange information. However, various problems have been found in the literature with regard to privacy information and related to the data used in ubiquitous environments. This paper offers a solution by means of statistical classification algorithms that can be used for control and privacy management. On the basis of the algorithms used in the tests, it proved to be possible to control and manage information by providing definitions of the variables and parameters for users, devices, and ubiquitous environments.*

1. Introduction

In the last ten years, advances in mobile communication technologies have led to a change in the computing paradigm. The traditional model is static and relatively predictable with based workstations, and has created a highly dynamic environment with constant changes caused by user mobility. This feature is enhanced by the use of multifunctional mobile devices such as cell phones and smartphones, as well as educational environments such as interactions Teleduc, Moodle, etc. This change can be seen as another stage towards the concept of Ubiquitous Computing (Ubiquitous/ Pervasive Computing) introduced by Mark Weiser (1991).

We are now living in an inter connected society, with e-mails, cell phones, palms, chats, information search engines, news sites, online communities, SMS, VoIP and other tools that until recently were not part of our daily routine either at work or leisure. According to Abech et al. (2012), the popularity of mobile devices to access the internet makes it feasible to obtain educational content regardless of time or place. In this new scenario of technological changes, there are new challenges and new forms of relationships that affect human behavior and hence all social factors involving education. Among these challenges is the question of privacy control which is of great importance since data and a shared location are unavailable without prior knowledge and authorization. This is a considerable problem given the increasing ease of access to computing resources. This kind of information can be best managed by the ubiquitous environment.

Thus it is also necessary to have a control of privacy, since the user may not need or want to locate or share his/her data at all times. The shared information can be best managed by the pervasive or ubiquitous environment since this is a means, for example, of reducing unnecessary data processing, and increasing the level of security and management of services. According to Cristiano Andre da Costa et al. (2008), the popularity of mobile devices to access the internet makes it feasible to obtain educational content regardless of time or place. In this new scenario of technological changes, there are new challenges and new forms of relationships that affect human behavior and hence all social factors involving education.

Thus, a model of ubiquitous privacy control is needed that meets as many requirements as possible related to the physical and virtual environment. In the literature, several studies can be found that address the question of privacy control and these are aimed at the user and the devices, services or communications employed. With regard to these features, this paper seeks to make a proposal for a privacy model that is designed for a ubiquitous academic environment, related to educational computing in an ubiquitous real world. Security issues will not be addressed in this ubiquitous computing, as there are already techniques to prevent attacks or disclosure of encryption information. Nor will this study ad-dress the question of the restrictive controls of users and devices, or its services and forms of communication.

The main concern of this work is the privacy model proposed for ubiquitous environments that provides definitions of parameters and criteria for an individual environment. Appropriate data classification algorithms were used for the control and privacy management environment and these were based on rules.

2. Related Word

In ubiquitous environments there are many problems and issues that need to be discussed, in particular, the control and management of privacy. According to Warren and Brandeis (1890), privacy is intrinsically linked to the perception of each individual about what it represents, such as a threat to their personal property or physical or moral integrity. Thus, it can be inferred that the privacy setting is something very abstract and subjective, and takes account of the diverse needs of each individual. These needs are not homogenous and may depend on cultural areas such as religion, tradition, customs, education or politics, or more subjective concerns such as user privacy or everyday factors such as age, health status, job responsibilities, mood, and leisure activities.

According to Cristiano et al. (2008) data that characterize a context may range from the physical world to the virtual world, and sometimes the two are merged. People often do not think of physical environments (e.g. an office, shop floor, stadium, and classroom) and virtual environments (e.g. a desktop computer, or the features of a mobile phone) as separate areas [Rodrigues et al. 2006]. Thus, the problem becomes even greater owing to the interaction between devices, users, applications, and communications between these environments. The work of involves the transfer of the control of music files based on the location of WIFI points. This author offers a different interpretation to what is given in this study, he seeks to define the types and sizes of information that must be transferred with regard to section and location. However, in our view, the question does not concern the environment itself but the point

of access to it. Thus, it is necessary to couple several other systems to supply information about the ubiquitous environment, and hence how it should proceed.

Henricksen et al. (2005) describe the hierarchy of control based on facts and user preferences. They also describe a context model of the application that controls the facts and individual occurrences, by seeking information on various ubiquitous sources; thus, it is not the privacy control in the ubiquitous environment. The work described by Lachello and Hong (2007), conducts a survey of several privacy issues addressed in the context of human computer interaction (HCI); the work also provides an overview of several points that should be tackled such as trends in the field and research being carried out. The main contribution of this work is that it addresses several key issues including the protection of the pervasive environment. In Bardram, Kjaer and Pedersen (2003) an authentication based solution is provided that is based on several examples of communication such as Radio Frequency Identification (RFID), and offers a single mechanism to manage different authentication protocols in ubiquitous environments. Despite carrying out authentication iterations with the pervasive system, there is not a change of perspective in the environment, nor any attempt to address issues related to individual privacy.

The work of Santarosa and Comfort and Basso (2010) seeks to support digital social inclusion in the technological dimension including principles of accessibility this exposes many points of weakness in the virtual learning environments, in particular the privacy control system in the devices used. In Tao and Peiran (2010) there is a research inquiry into the protection of data for individual transactions between users and "things" (Internet of things, IOT), based on the use of cards, tags and other devices used every day. It also outlines some specific situations in which IOT is used with categories and applications where concepts are defined in terms of a specific situation: for example, medical treatment is defined as private identification, but only based on the user´s location and restricted to the pervasive goal. In the work carried out by Gotthard and Zorzo (2007), there is an examination of the technologies that assist in the process of teaching and learning which are being discussed in various fields of knowledge where the issue of privacy is handled by a user agent. On the basis of these research studies, a comparison is made in Table 1, where the existing solutions are displayed. The left column has abbreviations and the number of references that are cited. The first line of this comparative table outlines the approaches required for privacy management in pervasive environments and these were set up on the basis of the literature. The following definitions are used:

-> Addresses (A): the work deals with the Question addressed;

-> It does not address (NA): The work does not deal with the Question addressed;

-> Describes not (ND): information not found to address the Question;

-> Developing (D): The item is still being developed; it is often pointed out in tests, validations, obtained results or future work.

Several studies describe the particular solution that applies to a pervasive or ubiquitous environment, but do not state clearly how to go about the management and control of ubiquitous environments. In the next section, there will be an examination of the application scenario where the privacy management model in ubiquitous

environments is applicable. Based on the comparison with and related work, we define a privacy model for ubiquitous environments that will be outlined in Section 3.

Table 1. Comparative State-of-the-art Privacy Management in Ubiquitous Environments.

Approach solution	User	Device	Applications	Services	Communication	Environment	Privacy
Leithardt et al 2010	A	NA	A	D	A	D	NA
Rodrigues 2006	A	NA	A	A	A	NA	NA
Görlach, Heinemann e Terpstra 2005	A	A	A	NA	D	NA	NA
Kalempa 2009	D	NA	A	A	A	ND	D
Henricken et al 2005	D	D	A	A	D	NA	NA
Iachello e Hong 2007	NA	ND	D	D	A	NA	NA
Bardram, Kjaer e Pedersen 2003	A	D	A	A	A	NA	NA
Santarosa, Conforto e Basso 2010	A	D	D	NA	NA	D	NA
Tao e Peiran 2010	D	A	D	A	D	NA	D
Gotardo e Zorzo 2007	A	D	A	A	D	NA	D
Proposed model	A	A	A	A	A	A	A

3. Proposed Model

The use of a privacy management mechanism for ubiquitous environments that meets the requirements of a given scenario has been identified as a problem. Together with this, there is a need to collect personal information to operate these systems within ethical and legal constraints, because the privacy of people is involved. A proposed model for privacy management for ubiquitous environments that complies with these requirements is illustrated in Figure 1.

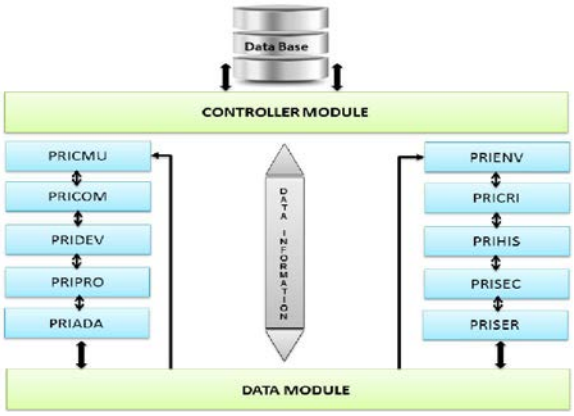

Figure 1. Model for a Privacy Manager adapted from [Leithardt et al., 2013].

The proposed privacy manager model consists of several components that are necessary to control ubiquitous environments. They can be individually described as follows:

Data Base: Rules for Information Storage and settings for users, devices, and the ubiquitous communication environment.

Controller module: the purpose of this is to receive access requests and perform the database control of the tables directly, with necessary information in accordance with the requests and access settings by maintaining control of the ubiquitous environment.

Data Module: In this module, calculations will be made of all the variables and parameters received from the other modules.

PRICRI: this module contains the rules and definitions of criteria and environmental settings such as access, use, sharing, location and other variables that can be manipulated or replaced in accordance with a) the environment settings and b) established criteria and standards;

PRICMU: Module management and privacy control of user information, which will handle definitions of related features for individual user preferences such as: temperature, light, authorized shares (such as information that someone wishes to share with other users and with his/her own environment), location data and other user preferences.

PRIDEV: management module and privacy control devices. This module will handle the data on the devices if these devices are in the environment itself and will then interact with it.

PRICOM: management module and communications privacy control. This concerns which forms of communication will be employed within the ubiquitous environment and how they will be used.

PRIADA: management and adjustment control module. This module will handle the information related to the adaptation of software and hardware in ubiquitous environments.

PRISER: environmental services management module. This module is responsible for the availability of services that can be used individually in each environment such as, shared information from other environments, devices, communications, location of users, environmental availability and its components that interact with users.

PRIHIS: this module will store and handle information on the historical user, environment, devices and other variables that may include other factors depending on the context.

PRIPRO: this module will carry out transactions of controls related to the user profile management.

PRISEC: this module will carry out the controls and management with regard to the safety of both the user and environment.

PRIENV: this module will register the attributes related to the environment.

All the modules operate independently and have their own characteristics and features that may vary according to the rules that have been previously established registered and enforced. Once these rules have been set out, each module sets its parameters based on the settings of the previous module. Thus, it is possible to have multiple environments with different rules and definitions for the same ubiquitous environments and the same user can use one or more different environments each with a

defined criterion. This can change depending on the device used in the communication as well as other factors that will be calculated on the data module. A preliminary strategy to address this need is to use algorithms for processing and classification of data used to manage the criteria set out in privacy management model. Therefore, some algorithms used in the literature were used and compared with the initial results of the tests described in the next section.

4. Prototype Testing and Preliminary Results

The implemented prototype, illustrated in Figure 2 shows the scenario used where the privacy settings server (4), (which acts as an authority for the mobile devices (2) and environments), receives the inferred symbolic locations through physical locations (1) and the mobile devices that the users upload. Each mobile device belongs to one user at a time while in use, and the same device is used by more than one user at different times. On receiving the symbolic location for a medium processed by the user's mobile location, the server passes by the model described in the previous section, and checks the changes that need to be adapted/ suited to the new context of the user with regard to environmental privacy and directs them via communication (3).

Figure 2. Architecture implemented

The server and the model are implemented with the aid of the Java EE programming language, which currently supports communications WebService Rest, Google Cloud Messaging (GCM) and Serial Communication. A PostgreSQL relational database was also employed. The mobile clients have been trained for the Android platform, and their native APIs were used for access to information about location, the prototype currently:

-> Makes environmental changes using different locations such as GPS and NFC Tags;

-> Identifies what types of access each user profile possesses when entering an environment;

-> Enables or disables the smartphone functionality in accordance with the privacy required by the environment.

One problem was how to classify the degree of access that a user has to the environment. This problem was solved by identifying the variables that provide the level of Access required by t the user to the environment, which are as follows: Profile and user frequency in the environment, an environment, weekday, shift and working Day.

The first three profiles (Unknown, Transient and User) are automatically identified and allowed to proceed through the system (evolutionary profiles), while the last three (Responsible, Student and Manager) are assigned manually by the environmental manager. This configuration is necessary because a large number of users will probably not access all the environments known by the system. In this way, the system itself can distinguish between ordinary and new users while it is running, thus dispensing with the settings of the system manager. By contrast, ordinary users will not be able to access all the resources, especially in private settings, or automatically be allowed to proceed for security reasons. For example, the customer in a cafeteria might not be able to access the box, as such, since there are profiles that have to be assigned to the user manually.

The rules for the progression of evolutionary profiles are configurable in the system. In this study the frequency (F) of the user (u) is used to determine when it should evolve into a profile (R) before it can advance through the environment or be resolved as a less permissive profile. In this case, for each environment, the implemented rules define, (a) the location of the lower frequency ranges (I) and upper (S) at which a change of profile should be made. If the environment has no evolutionary profiles, the equation below expresses the progression rule that is implemented.

$$P_{u,a} = \{P_{u,a} + 1\ , if\ F_{u,a} > S_a\ ;\ P_{u,a} - 1\ , if\ F_{u,a} < I_a\ ;\ P_{u,a}\ , otherwise\}\ \forall\ P \in [1, n]$$

In the case of non-evolutionary profiles, the frequency is also used to increase or reduce the level of access to users. However, the user profile remains the same and only the type of access is changed. In both cases, it was assumed that the frequency can take on three distinct levels: frequent, normal and infrequent.

Three types of environment were also taken into account: restricted, private and public; it is assumed that the binary value is true for weekdays, and false, for weekends; There were two shifts, day and night; Finally, the variable working day indicates whether a day is useful or not with regard to the location based on the day of the week or holidays since there are no days which have working hours for certain environments. The combination of all the variables that are possible in the case scenario studied resulted in a rating with 383 possibilities. In the second instance, after being identified, the variables were assigned to the combinations of the following types of access: Locked, Guest, Basic, Advanced and Administrative. These data were used for training and testing in seven different classification algorithms (Table 2), to determine the one with the highest degree of accuracy.

These experiments were used to select the algorithm that could be used by the server to automatically classify the user access level in unfamiliar surroundings and that had not been configured in the system, or in other words, where all the rules of a well-defined environment can be found. The comparative experiment between the classification algorithms was carried out by Weka tool by Hall, Mark, et al. (2009). The Table with the rules (a combination of all the attributes and their types that result in access) was divided into training and testing sets, through a cross-validation technique with ten subsets (10-fold cross-validation) where 90% of the data is used for training the classification algorithms, and the remaining 10% is used to check the results of these rules (unknown to the classifier). In addition, with this method, the test set is varied

among all the possible data training subsets. The final degree of accuracy shown in Table 2 is obtained from the average of the tests.

Table 2: Comparison of Classification Algorithms

Algorithm Classification	Precision	Correct instances	Incorrect instances
DecisionTable	0.887	343	40
Bayes Network	0.814	322	61
J48	0.887	341	42
Best-First Decision Tree (BT-Tree)	0.871	336	47
RandomTree	0.861	332	51
Nearest Neighbor With Generalization (NNge)	0.848	326	57
MultilayerPerceptron	0.888	341	42

An ontology was designed to formally represent the UbiPri model, [Leithardt et al., 2013]. According to [Bouiadjra et al., 2011], ontologies can be used to represent the context, provide inferences and share the knowledge generated by the application. Similarly, X states that axioms, bodies and vocabulary can be shared with the scientific community. FOEval assessment and evaluation scenarios were used to validate the ontology. FOEval allows users to select a set of metrics that can assist in the evaluation of ontologies. For this study, metrics were chosen with a level of detail and computational efficiency recommended by. Three calculations are made to assess wealth. First the wealth ratio (RR) measures the range of relationships, and assumes that the higher the number of non-hierarchical relationships, the richer the ontology. Another calculation performed is the wealth attribute (RA) which is the average number of attributes that are defined for each class. This may indicate the amount of information in the instance data, since the more attributes are set, the greater the amount of knowledge the ontology conveys. Finally, the rich ontology is calculated (RO) from the RR and RA.

According to [Bouiadjra et al., 2011], RO is set to the sum of RR and RA. RR is defined as the ratio between the numbers of non-hierarchical relationships defined in the ontology, divided by the number of all the (R) attribute wealth relations. RA, in turn, is defined as the number of attributes defined for all classes divided by the number of ontology classes. To make the calculations of the ontology [Gruber, T. R. et al., 1993], data were used (as summarized in Table 3). The results obtained for the ontology were UbiPri [Leithardt et al., 2013] points for RR 1.27; 0.31 points for RA; and 1.58 points for RO. These results demonstrate that UbiPri ontology is richer in relationships than in attributes. Furthermore, there is, on average, 15 classes by attribute and a relationship of 3 classes. The overall level of detailed calculation is defined by the average number of subclasses divided by the number of ontology classes. When we make to obtain the total number of subclasses, including 45 classes. We come to near zero, 0.02 ((48/45) / 45 = 0.02)). UbiPri ontology is divided in terms of classes and subclasses, and is very close to midway between the vertical and horizontal types of the taxonomy.

5. Conclusion and Suggestions for Future Work

Several studies have been conducted que have led to the survey. With the results, we intend that they be used for future research. The scientific investigations showed that it is possible to set criteria, parameters and variables que comply with the particular rules

of each environment. Thus, it can be concluded that it is also the basis of information que is handled automatically. In this study, a comparison preliminary was made of classification algorithms for data privacy and treatment.

These were focused on the environment. However, a good deal of further research and implementations are required in the future, due to the large amount of information que needs to be controlled and managed in ubiquitous environments. The results were generated from real time and in real world scenarios, and were based on information about students at an academic institution. That involves listening to lectures and taking part in other events in classrooms and the auditorium. We carried out a simulation to test the same settings with thousands of students, but due to several factors such as time and physical resources, it was not possible to simulate all profiles and managements. For this reason this fact can be listed among many others the areas for the continuation of research into data privacy in the future which shouldn't be of value in the area of computer studies. Figure 3 shows the location, identification and classification of the environment used in the tests. It was concluded from this que the behavior of algorithms chosen for the sample of 500 users achieved the expected results.

Figure 3. User identification and environment

In future work we plan to simulate larger environments, with a greater number of users, devices and criteria definitions than what was used to obtain the results of this work. For this reason, two projects have been submitted to funding agencies of Brazil requesting financial assistance for further studies. This will enable us to identify and develop more robust algorithms for the handling of private data on a large scale. Other important factors also need to be considered for future work, including the following:

- Definition of data processing techniques on a large scale;

- Devising computational metrics for a database;

- Preparation of other environmental ratings, for ex-ample, those restricted to individuals;

- Implementation of distributed algorithms for processing information;

- Definition of bug tracking techniques, as well as others that will be determined in the course of the research.

Acknowledgment

This work forms a part of the UbiArch project - Ubiquitous Architecture for Context Management and Application Development of the Federal University of Rio Grande do

Sul. We would like to thank the Conselho Nacional de Desenvolvimento Cientifíco e Técnologico (CNPq) [National Council of Scientific and Technological Development].

References

Bardram, J. E., Kjær, R. E., & Pedersen, M. (2003) "Context-aware user authentication–supporting proximity-based login in pervasive computing". In UbiComp 2003: Ubiquitous Computing, pp. 107-123. Springer Berlin Heidelberg.

Abech, M., da Costa, C. A., Barbosa, J.,Rigo, S., and Cambruzzi, W. (2012) "Um Modelo de Adaptação de Objetos de Aprendizagem com foco em Dispositivos Móveis". In Anais do Simpósio Brasileiro de Informática na Educação.

Bouiadjra A B. (2011) FOEval: Full Ontology Evaluation In proceeding of: 7th InternationalConference on Natural Language Processing and Knowledge Engineering, NLPKE,Tokushima, Japan.

Cristiano Andre da Costa, Adenauer Correa Yamin, and Claudio Fernando Resin Geyer. 2008. Toward a General Software Infrastructure for Ubiquitous Computing. IEEE Pervasive Computing 7, 1 (January 2008), 64-73.

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. KnowledgeAcquisition, 5, (2):199-220, 1993.

Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter 11.1 (2009): 10-18.

Henricksen, K., Wishart, R., McFadden, T., &Indulska, J. (2005) "Extending context models for privacy in pervasive computing environments". In Pervasive Computing and Communications Workshops, 2005. PerCom 2005 Workshops. pp. 20-24. IEEE.

Iachello, G., and Hong, J. (2007) "End-user privacy in human-computer interaction". Foundations and Trends in Human-Computer Interaction, v. 1, n. 1, pp. 1-137.

Leithardt, Valderi R. Q., Guilherme A. Borges, Anubis G. M. Rossetto, Carlos O. Rolim, Claudio Geyer, Luiz H. A. Correia, David Nunes and Jorge Sa Silva: A Privacy Taxonomy for the Management of Ubiquitous Environments. Journal of Communication and Computer Volume 10, Number 12, December 2013.

Rodrigues, Vagner J.d.S. (2006) "Privacy Management to Context-Aware Applications on Mobile Networks"., http://www.inf.ufg.br/~va gner/publications/Tese-2006-Vagner.pdf .Accessed in January 2016.

Santarosa, L. M. C., Conforto, D., & Basso, L. d. O. (2010) "Eduquito: Ergonomia Cognitiva para a Diversidade Humana". In: Educação, Formação & Tecnologia, v. 3, p. 4-13. ISSN 1646-933X.

Tao, H., &Peiran, W. (2010) "Preference-Based Privacy Protection Mechanism for the Internet of Things". In Information Science and Engineering (ISISE), 2010 International Symposium on, pp. 531-534. IEEE.

Warren, S. D., and Brandeis, L. D (1890) "The right to privacy". Harvard law review, 193-220.

Weiser, M. "The computer for the twenty-first century." Scientific American, vol. 265, no. 3, pp. 94–104, September 1991.