Generating real context data to test user dependent systems - application to multi-agent systems

Pedro Oliveira[1][2], Paulo Novais[1], and Paulo Matos[2]

[1] Algoritmi Centre/University of Minho, Department of Informatics, Braga, Portugal,
[2] Institute Polytechnic of Bragança, Dep. of Informatics and Communications,
Bragança, Portugal

Abstract. This paper, deals with the usually need of data to simulate behavior and efficiency of proposed solutions in several fields, and also knowing that personal data always bring privacy and security issues. This work wants to promote a balanced solution between the need of personal information and the user's privacy expectations. We propose a solution to overcome these issues, and don't compromise the balance between security and personal comfort based on generating real context data of users, that allow to test user dependent systems.

Keywords: adaptive-system · AmI · security · privacy · simulation· multi-agent system

1 Introduction

Systems that deal with personal data always bring privacy and security issues, also the balance of these issues, with the need that persons have in interact with spaces in a transparent way and that those spaces smartly adapt to their preferences.

That said, in this project, is proposed a solution to overcome these issues, and don't compromise the balance between security and personal comfort.

In this field some work was done [2][10][4][9], we evaluate and continue to improve that work and develop a more focused solution to solve the presented problem.

Pursuing this effort, we need several information of hundreds of users to test a multi-agent system that simulates the comfort of these users, like in most of the research projects carried out, data are usually needed to simulate behaviour and efficiency of the proposed solutions, since it is necessary to use them to validate and test the research carried out at the most diverse levels.

The development of this project also arises from the need to gather information on the comfort preferences (temperature, humidity, musical playlist, musical genre, etc.) of multiple users. And besides that, also have the information of how each user adapts their preferences to the place where it is [7]. This information, in addition to being necessary on a large scale (hundreds of users), would also be necessary in a very broad time-frame, always longer than one year. As it is known that comfort preferences normally vary according to a seasonal character, namely at the seasons level.

Getting data with this dimension and involving so many users is a difficult task and, in addition to the users collaboration, would require a high cost, regarding the equipment needed to collect this information [8].

Thus an algorithm was created, which simulates not only the variation of preferences of the different users, but also their daily life, taking into account the different places that it frequents (home, work, leisure places). In addition, the relationship between users is also established, introducing the concept of family and co-workers.

2 Materials and Methods

2.1 Security and Privacy

The technological revolution that is felt, particularly in behavioural analysis fields, IoT or big data, brings significant new challenges, including those related to the type of user information that can be collected, and the knowledge that can be obtained derived from the compilation of this information. Although not necessarily existing the user's authority to collect such information.

This IoT revolution has yet clearly identified problems. In particular, the privacy and security of user data. Foreseeing the dissemination of intelligent spaces, of which the user can, and want to take advantage of the interaction between systems, and the consequent sharing of personal data, this is a theme that needs resolution in a short-term [5].

Obviously at this point there will be the requirements for concessions and commitments on the part of the user.

IoT increases the risk of personal privacy, and the confidentiality and integrity of data in organizations. Some IoT applications for consumers, particularly those related to health and wellness, which store sensitive personal information, and that consumers may not want to share.

Using this solution, all the problems of security and privacy are overcomed, without the need of data collection from real users [1].

2.2 The Simulation Algorithm

The developed simulation algorithm has several predefined assumptions that allow the simulation to be as close to reality as possible. These assumptions are defined in the code as input variables and customized according to the simulation needs.

The types of schedule are defined, taking into account the different possibilities, in this case four, then depending on the type of schedule selected, the value for minimum and maximum delay is also defined.

Table 1 defines the four delays to the different schedules, in this table we have the minimum and maximum delay time, that the algorithm uses to generate the random between these limits, and to introduce the generated value in each situation.

Table 1: Parameters and different delays

Parameter (Delay)	**Minimum (minutes)**	**Maximum (minutes)**
Enter Work	-10	10
Enter Home	-30	30
Exit Home	-40	40
Lazer Hours	-90	90

The diagram presented at figure 1 illustrates more clearly the different processes that the algorithm develops.

The process is started by choosing the number of local systems associated with the user, between 1 and 5, depending on this choice the user is associated with the local systems defined. Consequently when selected the local system associated to the home, users are generated between (0-3) that define the concept of family. With regard to the local system associated with the workplace, coworkers are generated between (0-3) associated with the same workplace. Then, for each generated user, the corresponding time type is defined, so when the introduction of user history for each day is started, it is coherently associated with a type of schedule. Then, it is started the process of entering history information for each user, this introduction is carried out consecutively for each day. Thus during the daily input process, the corresponding time of introduction and delay is generated, taking into account the limits previously defined for this delay. For each daily period the introduction is performed, and new delay random intervals are defined relative to the time set as standard, in order to get as close as possible to the daily reality of the users. In addition to the time, and for each period, are also entered the preferences for the correspondent user. In this case, the following premises were defined:

In 5% of the situations the preference value is changed, introducing in this case the factor of randomness with the limits (-5; 5).

For 15% of the situations, the preference value is changed by introducing the factor of randomness with the limits (-3; 3). In all other situations, the preference value is entered without any change.

In this way, the information for this project required to the simulation is obtained, with as many users as necessary, as well for the necessary time period, in this case a 720 day time window was defined, with 1000 users in the system, which originates more than three millions historical records in the database.

Next, at table 3, are showed the parameters defined for the four schedules, we try to put these schedules as real as possible, and with that we have the most regular shifts of work, like starting work at 08h or 09h of the morning,

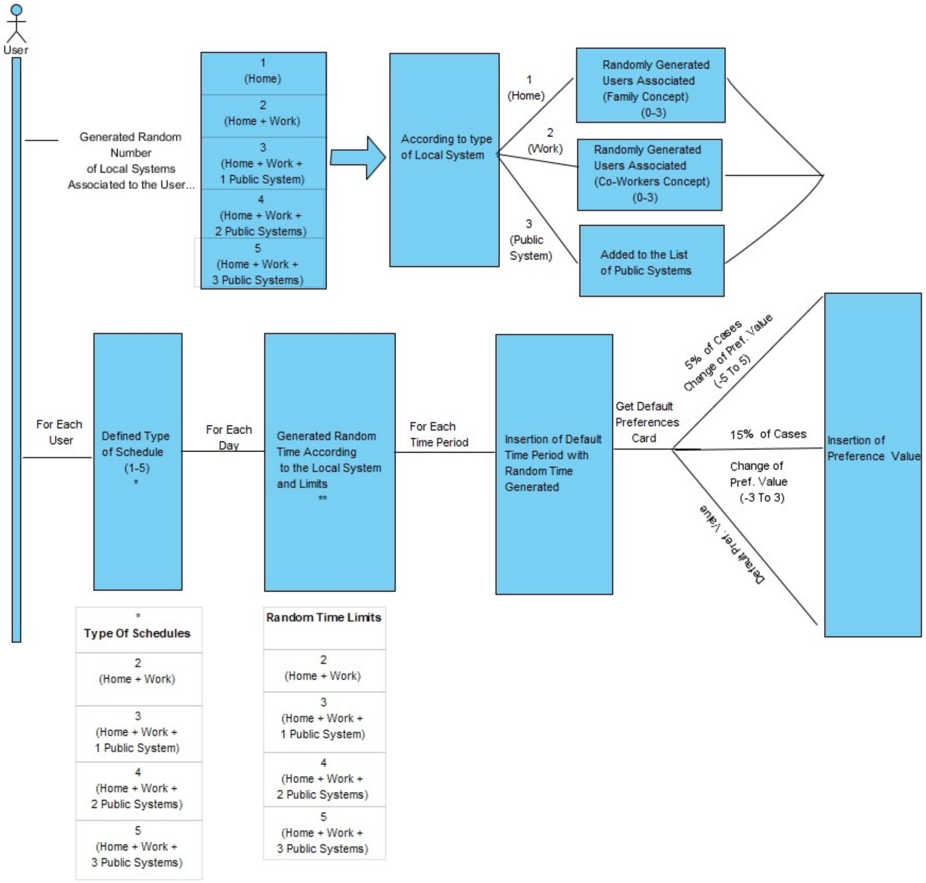

User
Generated Random Number of Local Systems Associated to the User...
1 (Home)
2 (Home + Work)
3 (Home + Work + 1 Public System)
4 (Home + Work + 2 Public Systems)
5 (Home + Work + 3 Public Systems)
According to type of Local System
1 (Home)
2 (Work)
3 (Public System)
Randomly Generated Users Associated (Family Concept) (0-3)
Randomly Generated Users Associated (Co-Workers Concept) (0-3)
Added to the List of Public Systems
For Each User
Defined Type of Schedule (1-5) *
For Each Day
Generated Random Time According to the Local System and Limits **
For Each Time Period
Insertion of Default Time Period with Random Time Generated
Get Default Preferences Card
5% of Cases Change of Pref. Value (-5 To 5)
15% of Cases
Change of Pref. Value (-3 To 3)
Default Pref. Value
Insertion of Preference Value
* Type Of Schedules
2 (Home + Work)
3 (Home + Work + 1 Public System)
4 (Home + Work + 2 Public Systems)
5 (Home + Work + 3 Public Systems)
Random Time Limits
2 (Home + Work)
3 (Home + Work + 1 Public System)
4 (Home + Work + 2 Public Systems)
5 (Home + Work + 3 Public Systems)

Fig. 1: Diagram - Users simulation algorithm

start at 16h in the afternoon, or the work by night that starts at 00h. We set the most normal period of work (eight hours), and with that we can define the correspondent hours of exit home to go work, the enter home hour after work, and also the exit home to do some leisure hours, that is related to the period that the user is not working.

In this way, the information for this project required to the simulation is obtained, with as many users as necessary, as well for the necessary time period, in this case a 720 day time window was defined, with 1000 users in the system, which originates more than three millions history records in the database.

This process takes few hours, namely for the values defined before, for instance in an average computer it takes three to four hours the process to generate the full dataset of information and the insertion at the database.

The following listing at listing 1.1 demonstrates a little part of the algorithm, in the case where we select the type of system selected.

Table 2: Randomness factor

% of Cases	**Randomness Factor**	
●	Minimum	Maximum
75%	●	●
15%	-3	3
5%	-5	5

Table 3: Parameters/Schedules

Parameter (Hour)/ Schedule	**Schedule A**	**Schedule B**	**Schedule C**	**Schedule D**
Enter Work	08h00	09h00	16h00	00h00
Enter Home	17h10	18h10	00h10	08h10
Exit Home	07h50	08h50	15h50	23h50
Enter Lazer	20h00	21h00	10h00	15h00

Listing 1.1: Part of algorithm implementation

```
switch (randomNumberOfLocalSystemsByUser) {
case 1:
TypeOfSystem = 1;
System.out.println("Home");
idLocalSystem++;
addMoreUsersToSystem(idMasterUser, TypeOfSystem);
DescLocalSystem = ("Home:␣" + idMasterUser + "␣+" +
    idsUsersAtSystem);
genenerateUserSimul.AddLocalSystemWS(idLocalSystem,
    DescLocalSystem);
genenerateUserSimul.insertHistoryRecord(idMasterUser,
    idsUsersAtSystem, idLocalSystem, TypeOfSystem);
break;
case 2:
```

3 Results

3.1 Statistical Analysis

For the simulation results validation, we used the distribution of frequency analysis, using histograms, that is the graphical representation in columns/bars of the dataset previously tabulated and divided into classes [6].

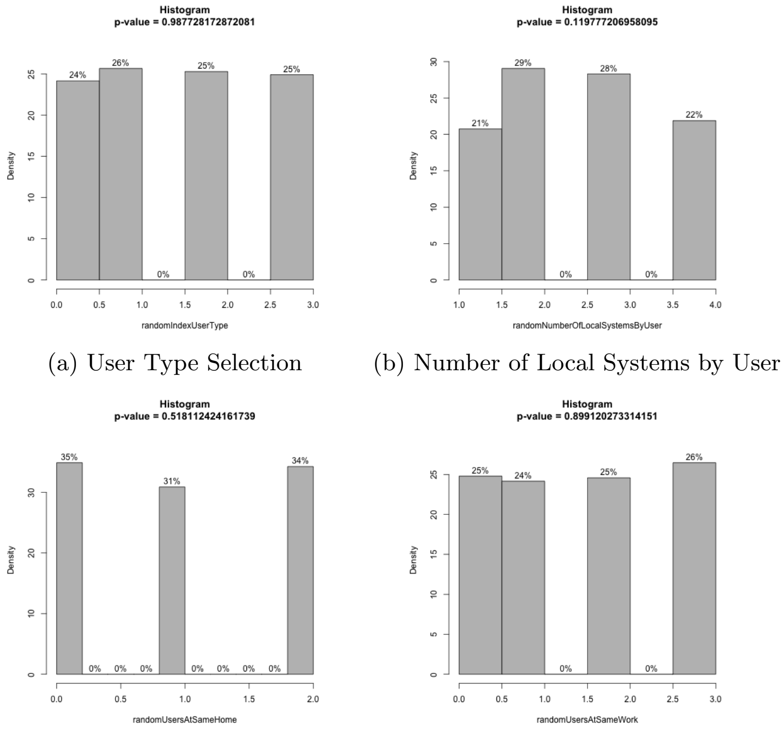
In this case for the different classes, the density of each is shown, as well each percentage, which it represents within each distribution.

As can be verified by the histogram analysis in figure 2 and 3, it can be concluded that the different classes are evenly distributed.

As can be seen by analysing the histograms, we can conclude that the different classes are distributed equally.

Describing figure 2, we can see in (a) the type of user (0-3), at (b) the number of local systems for each user (0-4), in (c) the number of users at the same home (0-2) (family concept) and at (d) the number of users at the same work (0-3) (co-workers concept).

In all histograms for each distribution, we calculate the hypothetical frequency called the P-value, also known as the "observed significance level" [3], and also validate that all the percentages density are very close, what confirm a evenly distribution.



(a) User Type Selection

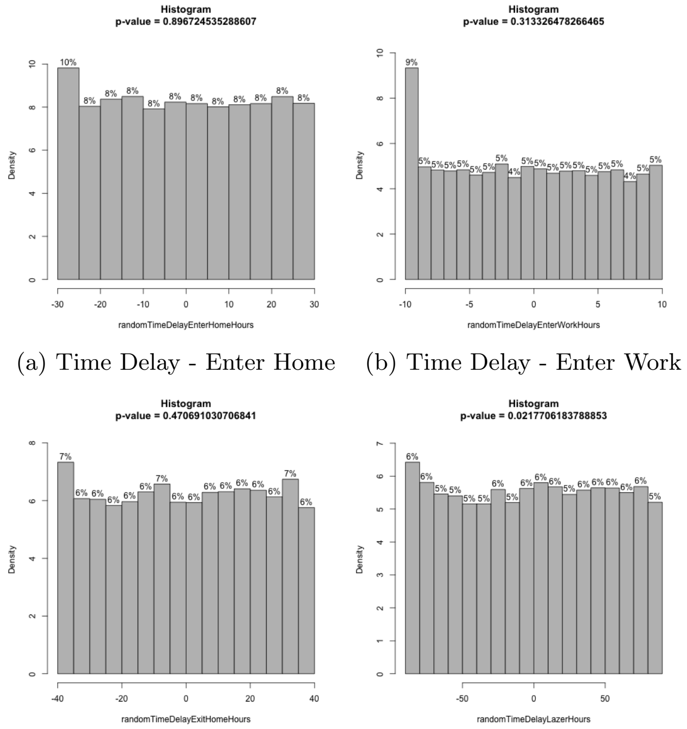
(b) Number of Local Systems by User

(c) Number of Users at same Home

(d) Number of Users at same Work

Fig. 2: Histogram - User Parameters

Describing figure 3, we can see in (a) the type of user (0-3), at (b) the number of local systems for each user (0-4), in (c) the number of users at the same home (0-2) (family concept) and at (d) the number of users at the same work (0-3) (co-workers concept).



(a) Time Delay - Enter Home (b) Time Delay - Enter Work

(c) Time Delay - Exit Home (d) Time Delay - Lazer Hours

Fig. 3: Histogram - Time Delays

4 Conclusions

Concluding, we can say that all the work proposed is developed, and all the objectives are achieved. We can use a fully operational solution to simulate how many users we need, in any time period and with the comfort preferences needed for the problem. This work is very important to all type of investigation, that need massive information, and where is needed the simulation of this kind of data, with defined parameters and limits to be as close to reality as possible.

In this case the information generated is used, to test and improve a multi-agent system, that depends from user information, namely behaviour and comfort preferences but also we can adapt it to simulate other type of information.

For future work, the simulation algorithm can be generalized, as much to accommodate as many cases as possible, and with that, the user only need to define and configure the initial parameters for each case, and after that retrieves how many data as he needs for the specific case defined.

Acknowledgements

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2019.

References

1. Babar, S., Mahalle, P., Stango, A., Prasad, N., Prasad, R.: Proposed security model and threat taxonomy for the internet of things (iot). In: Recent Trends in Network Security and Applications, pp. 420–429. Springer (2010)
2. Cárdenas, M., Sanz, J.G., Pavón, J.: Testing ambient assisted living solutions with simulations. In: IFIP International Conference on Testing Software and Systems. pp. 56–61. Springer (2018)
3. Chaubey, Y.P.: Resampling-based multiple testing: Examples and methods for p-value adjustment (1993)
4. Hoes, P., Hensen, J., Loomans, M., De Vries, B., Bourgeois, D.: User behavior in whole building simulation. Energy and buildings **41**(3), 295–302 (2009)
5. Khan, R., Khan, S.U., Zaheer, R., Khan, S.: Future internet: the internet of things architecture, possible applications and key challenges. In: Frontiers of Information Technology (FIT), 2012 10th International Conference on. pp. 257–260. IEEE (2012)
6. Konishi, S., Kitagawa, G.: Information criteria and statistical modeling. Springer Science & Business Media (2008)
7. Schilit, B., Adams, N., Want, R.: Context-aware computing applications. In: Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on. pp. 85–90. IEEE (1994)
8. Wang, K.I.K., Abdulla, W.H., Salcic, Z.: Ambient intelligence platform using multi-agent system and mobile ubiquitous hardware. Pervasive and Mobile Computing **5**(5), 558–573 (2009)
9. Yan, D., O'Brien, W., Hong, T., Feng, X., Gunay, H.B., Tahmasebi, F., Mahdavi, A.: Occupant behavior modeling for building performance simulation: Current state and future challenges. Energy and Buildings **107**, 264–278 (2015)
10. Zimmerman, G.: Modeling and simulation of individual user behavior for building performance predictions. In: Proceedings of the 2007 Summer Computer Simulation Conference. pp. 913–920. Society for Computer Simulation International (2007)