

Privacy Requirements Classification Method for System Design

Nor AsiakinHasbullah, NorulzahrahMohdZainudin, Noor Afiza Mat Razali,
NorshahriahWahab

Abstract—It is important to get users’ privacy requirements through data or information classification during the system design. Currently, the citizen-centric perspective of privacy requirement is not well understood. To fill this gap a study with the objectives of to investigate citizens’ privacy requirements and need through their privacy preferences has been done. From the data analysis, the citizen-centric preferences’ set was developed based on the classification of personal and sensitive information that has been obtained through a survey of 350 respondents. The result is configured into a reference table and sensitivity classification tool respectively. Therefore, we suggested the tool to be used as a classifying method to classify sensitive and personal information for system design.

Index Terms: Classification method, privacy requirements, privacy preferences, reference table, sensitivity classification tool.

I. INTRODUCTION

The development of an effective e-government system needs to take into account citizens’ opinions and preferences on sensitive information. Distinguishing between sensitive and personal data is an important element during system development in order to secure privacy and as a security mechanism in system life. There are lack of classification of citizen's sensitive information in the e-government according to the user's preferences. The privacy requirements for systems development are based on what the designers conceptualize [7] and current e-government developments are in the hands of the systems designer [8]. The important of data sensitivity classifications was based on [9] who cited that serious attempt to determine the privacy sensitivity level of each personal data element during the system design phase. The objectives of this research were to classify citizens’ privacy requirements and need through their privacy preferences. It is important to have the citizen involvement in getting the information classification as they are the future user of the systems.

II. TYPES OF INFORMATION

From the BusinessDictionary.com, the definition of personal data is not listed in their definition, but it defines personal information as recorded information about an identifiable individual while sensitive information as privilege or proprietary information which if compromised through alteration, corruption, loss, misuse or unauthorized

disclosure could cause serious harm to the information owner. Meanwhile, from the Microsoft Corporation’s privacy guidelines there are two types of information, which are [10], [1] (i.) Personally Identifiable Information (PII) (ii.) Sensitive Personally Identifiable Information (Sensitive PII).

Table 1: Definition of personal and sensitive data

Legislation	Personal Data	Sensitive Data
Information Commissioner Office of the UK (2013) [12]	Data which able to identify the owner including their expression of opinion and indication of intention	Information of Racial and ethnic origin; Political Opinion; Religious beliefs or other belief; Member of a trade union; Physical or Mental Health or condition; Sexual life; The commission or alleged commission of any offense; Offense committed or alleged committed, Disposal of such proceeding or the sentence of any court in such proceeding
Malaysia Personal Data Protection Act (2010) [11]	Any information in respect of commercial transactions, which— (a) is being processed wholly or partly by means of equipment operating automatically in response to instructions given for that purpose; (b) is recorded with the intention that it should wholly or partly be processed by means of such equipment; or (c) is recorded as part of a relevant filing system or with the intention that it should form part of a relevant filing system, that relates directly or indirectly to a data subject, who is identified or identifiable from that information or from that and other information in the possession of a data user, including any sensitive personal data and expression of opinion about the data subject; but does not include any information that is processed for the purpose of a credit reporting business carried on by a credit reporting agency under the Credit Reporting Agencies Act 2010	Any personal data consisting of information as to the physical or mental health or condition of a data subject, his political opinions, his religious beliefs or other beliefs of a similar nature, the commission or alleged commission by him of any offence or any other personal data as the Minister may determine by order published in the Gazette

Revised Manuscript Received on September 14, 2019.

Nor Asiakin Hasbullah, Computer Science Department, National Defence University of Malaysia, Kuala Lumpur, Malaysia.

Norulzahrah Mohd Zainudin, Computer Science Department, National Defence University of Malaysia, Kuala Lumpur, Malaysia.

Noor Afiza Mat Razali, Computer Science Department, National Defence University of Malaysia, Kuala Lumpur, Malaysia.

Norshahriah Wahab, Computer Science Department, National Defence University of Malaysia, Kuala Lumpur, Malaysia.

The term “personal data” are commonly referred as “Personally Identifiable Information” (PII) and thus being used interchangeably. In the EU Data Protection Directive, the term “personal data” being employs the same function as PII [16].

The other type of information in the Microsoft Corporation’s privacy guidelines is the Sensitive Personally Identifiable Information (Sensitive PII) which is a subset of PII that considered being so important to the individual. This information type must be specially protected. Included in this category is any data that could [1]:

- i. Be used to discriminate
- ii. Facilitate identity theft
- iii. Permit access to a customer's account

Most of the policies in data protection and privacy of personal data define sensitive data as “A subset of Personal Data, and refers to any Personal Data pertaining to racial or ethnic origins, trade union membership, medical or health conditions, political or religious beliefs, sex life, or criminal history” [14], [15]. The data classification of PII and Sensitive PII were stated precisely in the Microsoft Corporation’s privacy guidelines [1] which are as follows:

Table 2: The classification of PII and sensitive PII [1]

Personally Identifiable Information (PII)	Sensitive Personally Identifiable Information (Sensitive PII)
Personally Identifiable Information includes, but is not limited to: name, address, phone number, fax number, e-mail address, financial profiles, medical profile, social security number, and credit card information.	Credit card numbers, bank account information, race, ethnic origin, religious or philosophical beliefs, political opinions, trade union memberships, sexual lifestyle, physical or mental health, mother’s maiden name, passwords or PINs

From the literature the definition of personal data did not mention the data type, but it was precisely stated for sensitive personal data. There is a room to contribute to personal data types according to the availability of personal data through information sharing to online databases and social media. It is important to identify the data types that could cause intrusion to one’s privacy if known to others. The other issue to look at is the allegation of Microsoft Corporation and other data protection and privacy of personal data policy that Sensitive PII is a subset of PII. There should be an experiment or test to prove that Sensitive data is a subset of personal data.

III. METHODOLOGY

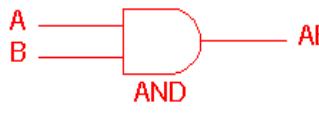
This research conducted a survey to 350 respondents from selected area in Selangor and Kuala Lumpur. The survey was to get citizens' privacy preferences, which they have to answer in accurate time only. The respondents have to classify the information as *Sensitive/Personal/Not Sure*. For the survey, a non-probability sample has been applied to the sampling technique which is convenience sampling in collecting the data. Furthermore, a method to get the sensitivity classification will also be proposed to simplify the classification method through a set of rules. Validation of the proposed classification method by the expert has also been done in this phase. The data types listed to be classified in the questionnaire are based on the UK list of data types

from their 2006 survey, the current sensitive information recognized by the Malaysian law and legislation and available data types that can be achieved through social media and the internet.

IV. SENSITIVITY CLASSIFICATION

The method to classify the information for this research are by adopting a truth table, which is commonly used to represent logic gate function for Boolean expression. The input and output information of any logic gate or circuit can give a visual representation for the switching system’s function and can be plotted into standard table besides the Boolean expression. The output results will depend on the combination of the input. The following table illustrates the two inputs AND gate:

Table 3: The AND gate [2]

Symbol	Truth Table		
	A	B	AB
	0	0	0
	0	1	0
	1	0	0
	1	1	1
Boolean Expression = A.B	Read as A AND B gives AB		

The Boolean expression could be written as A.B or only as AB [3]. This research applied the truth table of AND for the classification of the sensitive and personal information as it provides three distinct categories of information after classification. There are A (Sensitive) and B (Personal) and AB for the selection of different categories between the two variables as depicted in Truth Table in Table 4. AB will actually represent the intersection of the sets. AND returns TRUE if all of the arguments evaluate to TRUE [4]. Through the analysis, a comparison of respondents selected information for Personal, Sensitive and Not Sure was made within two and four variables by using SPSS 16.0. Using descriptive analysis of cross tabulation, the classification methods used are as depict in Table 4 and 5.

Table 4: Truth table for Sensitive (S) and Personal (P) for two variables

Variable1	Variable2	Result
S	S	S
S	P	SP
P	S	SP
P	P	P

Table 5: Truth table for Sensitive (S) and Personal (P) for four variables

Variable1	Variable2	Variable3	Variable4	Result
S	S	S	S	S
S	S	S	P	SP
S	S	P	S	SP
S	S	P	P	SP
S	P	S	S	SP
S	P	S	P	SP
S	P	P	S	SP



S	P	P	P	SP
P	S	S	S	SP
P	S	S	P	SP
P	S	P	S	SP
P	S	P	P	SP
P	P	S	S	SP
P	P	S	P	SP
P	P	P	S	SP
P	P	P	P	P

The rules of the truth table of logic function AND in this study are, the entire variable Sensitive then the results will be sensitive, the entire variable Personal then the results will be Personal, otherwise will be Sensitive and Personal. The reasons for using AND operator was because the classification of sensitive and personal is not absolute, for example, if sensitive is greater than personal ($S > P$), but still there are respondents than classify it as personal. The only thing is, it does not more than the respondents that classified the information as sensitive. Using the truth table in this method the following set of rules was suggested:

```

If
  (V1=S> P and V2=S > P) then
S
Else if
  (V1=P> S and V2=P> S) then
P
Else
  S ∩ P
    
```

Fig. 1:Set of rules for locality

To get the S from the set of rules is to get both of the variables (rural and urban) classified the information as Sensitive. While, to get the P will need both of the variables to classify the information as Personal. Otherwise, it will be categorized as Sensitive AND Personal. An analysis between four variables was made using SPSS 16.0 through cross-tabulation to reveal another detail view of the response. A comparison of sensitivity classification was made within races.

```

If
  (V1=S>P and V2=S>P
  and V3=S>P and V4=S>P) then S
Else if
  (V1=P>S and V2=P>S
  and V3=P>S and V4=P>S) then
P
Else
  S ∩ P
    
```

Fig. 2:Set of rules for races

To get the S from the formula is to get all the four variables classified the information as Sensitive. While for P, all the variables need to classify the information as Personal. Otherwise, it will be categorized it as Sensitive AND Personal.

V. FINDINGS DISCUSSION & RESULTS

The results for privacy preferences for all categories are tabulated in sets for sensitivity classification.

A. Between Locality

$S = \{\text{Biometric Information, Physical or Mental Health, Sexual Life Information, Financial Detail, Credit Card Number, Political Opinion}\}$

$P = \{\text{First Name, Full Name, Gender, Racial or Ethnic Origin, Home Address, Country/State/City of Residence, Marital Status, Employment History, Workplace, Academic Qualification, Job Position, Vehicle Registration Number, Email Address}\}$

$S \cap P = \{\text{Date of Birth, Place of Birth, Age, Identity Card Number, Religion or Belief, Genetic Information, Job Grade and Salary, Telephone Number, Income Detail, Mother's Maiden Name, Offense or Criminal record, Personal Photo, Property/asset, Map of House, Your Voice, Your Video}\}$

B. Between Genders

$S = \{\text{Biometric Information, Physical or Mental Health, Credit Card Number, Sexual Life Information, Financial Detail, Income Detail, Offense or Criminal Record}\}$

$P = \{\text{First Name, Full Name, Place of Birth, Age, Gender, Racial or Ethnic Origin, Home Address, Country/State/City of Residence, Marital Status, Employment History, Workplace, Academic Qualification, Job Position, Telephone Number, Vehicle Registration Number, Email address, Property/asset, Map of House}\}$

$S \cap P = \{\text{Identity Card Number, Date of Birth, Genetic Information, Religion or Belief, Job Grade and Salary, Political Opinion, Mother's Maiden Name, Personal Photo, Your Voice, Your Video}\}$

C. Between Race

$S = \{\text{Sexual Life Information, Offense or Criminal record}\}$

$P = \{\text{First Name, Full Name, Date of Birth, Place of Birth, Racial or Ethnic Origin, Home Address, Country/state/city of Residence, Marital Status, Employment History, Academic Qualification, Job Position, Vehicle Registration Number, Email Address}\}$

$S \cap P = \{\text{Identity card number, Age, Gender, Genetic Information, Biometric Information, Religion or Belief, Physical or Mental Health, Workplace, Job Grade and Salary, Telephone Number, Financial Detail, Income Detail, Credit Card Number, Political Opinion, Mother's Maiden Name, Personal Photo, Property/asset, Map of House, Your Voice, Your Video}\}$

VI. SYNTHESIS OF THE FINDINGS

To get the synthesis of the findings the following classification method has been applied.

$$\begin{aligned}
 S &= (S_{\text{locality}}) \cup (S_{\text{genders}}) \cup (S_{\text{races}}) \\
 S \cap P &= (S \cap P)_{\text{locality}} \cup (S \cap P)_{\text{genders}} \cup (S \cap P)_{\text{races}} \\
 P &= (P_{\text{locality}}) \cup (P_{\text{genders}}) \cup (P_{\text{races}})
 \end{aligned}$$

Fig. 3:Set of rules to get all of the sensitivity classification

The following sets are the results after applying the set of rules in Fig. 3:



$S = \{\text{Biometric Information, Physical or Mental Health, Credit Card Number, Political Opinion, Sexual Life Information, Financial Detail, Income Detail, Offense or Criminal record}\}$

$S \cap P = \{\text{Identity Card Number, Age, Gender, Genetic Information, Religion or Belief, Workplace, Job Grade and Salary, Telephone Number, Mother's Maiden Name, Personal Photo, Property/asset, Map of House, Your Voice, Your Video}\}$

$P = \{\text{First Name, Full Name, Racial or Ethnic Origin, Home Address, Country/state/city of Residence, Marital Status, Employment History, Workplace, Job Position, Academic Qualification, Vehicle Registration Number, Email Address}\}$

Table 6: Results for citizen-centric privacy preferences

Level of Procedures	Type of Information	Sensitivity Classification
High	Biometric Information, Physical or Mental Health, Credit Card Number, Political Opinion, Sexual Life Information, Financial Detail, Income Detail, Offense or Criminal record	Sensitive
Medium	Identity Card Number, Age, Gender, Genetic Information, Religion or Belief, Workplace, Job Grade and Salary, Telephone Number, Mother's Maiden Name, Personal Photo, Property/asset, Map of House, Your Voice, Your Video	Sensitive AND Personal
Basic	First Name, Full Name, Racial or Ethnic Origin, Home Address, Country/state/city of Residence, Marital Status, Employment History, Workplace, Job Position, Academic Qualification, Vehicle Registration Number, Email Address	Personal

There are three levels of procedures for integrity, security and auditing suggested by [17] which are basic, medium and high. By adopting this level of procedures the above table is derived from the sensitivity classification. From Malaysian information classification [18], there are four classifications which are restricted, confidential, secret and top secret and could only be declared by (i) someone who is the public

officer and (ii) has been appointed by a minister or chief minister of the state [5]. Top secret will be merged with secret information as in the schedule of the Official Secret Act 1972 [5], documents such as Cabinet documents, State Executive Council documents, any documents concerning national security, defence and international relations are classified as official [6]. This paper suggested that the information that has been classified as sensitive should have a higher degree of security mechanism of secret while the sensitive AND personal categories of information should have a medium degree of security mechanism of confidential and the personal information should have a basic security mechanism of restricted.

VII. CONCLUSION

For practical contribution, this paper proposed a sensitivity classification as a classification method to be used as a tool in classifying personal information into three categories of information which are sensitive, personal and sensitive AND personal. This method used the logic table of the AND operation that came out with the set of rules to be used in the classification. The output of the classification was the citizen-centric privacy preference that contributes as citizen choice in the citizen-centric privacy requirements engineering framework with three levels of procedures for integrity, security and auditing suggested by [17] which are basic, medium and high. It is suggested to convert the classification method with the set of rules to classify the information types to an algorithm by the system developer to develop automated classification system. Therefore, could simplify the analysis to be done and assist system designers to classify information types during the system design phase. Hence, for the empirical contributions, from the survey and the classifying method applied, it is evidenced that sensitive information is intersect of personal information and not a subset of personal information. Through the finding, it is evidence that there are sensitive information which is not personal information and vice versa [13].

APPENDIX

Respondents' Classification

Data Type	Personal		Sensitive		Not Sure	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Identification Information						
First Name	275	78.6	54	15.4	21	6.0
Full Name	222	63.4	108	30.9	20	5.7
Identity Card Number	157	44.9	165	47.1	28	8.0
Date of Birth	173	49.4	153	43.7	24	6.9
Personal Photo	161	46.0	155	44.3	34	9.7
Mother's Maiden Name	133	38.0	174	49.7	43	12.3



PRIVACY REQUIREMENTS CLASSIFICATION METHOD FOR SYSTEM DESIGN

Biometric Information	99	28.3	220	62.9	31	8.9
Social Status						
Place of Birth	196	56.0	114	32.6	40	11.4
Age	184	52.6	128	36.6	38	10.9
Gender	203	58.0	96	27.4	51	14.6
Marital Status	244	69.7	66	18.9	40	11.4
Racial or Ethnic Origin	190	54.3	118	33.7	42	12.0
Religion or Belief	176	50.3	125	35.7	49	14.0
Political Opinion	117	33.4	177	50.6	56	16.0
Offense or Criminal record	113	32.3	169	48.3	68	19.4
Contact Information						
Home Address	192	54.9	125	35.7	33	9.4
Telephone Number	172	49.1	147	42.0	31	8.9
Email Address	201	57.4	96	27.4	53	15.1
Country, State or City of Residents	204	58.3	103	29.4	43	12.3
Map of House	168	48.0	143	40.9	39	11.1
Health Information						
Physical or Mental Health	138	39.4	175	50.0	37	10.6
Genetic Information	136	38.9	181	51.7	33	9.4
Financial information						
Financial Detail	111	31.7	210	60.0	29	8.3
Income Detail	114	32.6	194	55.4	42	12.0
Credit Card No	104	29.7	211	60.3	35	10.0
Education Information						
Academic Qualification	247	70.6	69	19.7	34	9.7
Occupational Information						
Workplace	229	65.4	84	24.0	37	10.6
Job Grade & Salary	161	46.0	154	44.0	35	10.0
Job Position	222	63.4	92	26.3	36	10.3
Employment History	212	60.6	99	28.3	39	11.1
Assets Information						
Vehicle Registration No	186	53.1	126	36.0	38	10.9
Property/assets	179	51.1	131	37.4	40	11.4
Social Information						
Sexual Life Information	120	34.3	184	52.6	46	13.1
Your Voice	161	46.0	148	42.3	41	11.7
Your Video	149	42.6	159	45.4	42	12.0

REFERENCES

- 1 Microsoft Corporation, Privacy guidelines for developing software products and services. 2008, Available: <https://download.microsoft.com/download/9/3/5/935520EC-D9E2-413E-BEA7-0B865A79B18C/Privacy%20in%20Software%20Development.pptx>.
- 2 W. Sangosanya, D. Belton, and R. Bigwood, Basic gates and functions. 2017, Available: <http://www.ee.surrey.ac.uk/Projects/CAL/digital-logic/gatesfunc/#andgate>.
- 3 Electronic Tutorials, Boolean algebra truth tables. 2017, Available: http://www.electronicstutorials.ws/boolean/bool_7.html.
- 4 S. Cheusheva, Using logical functions in Excel: AND, OR, XOR and NOT. 2017, Available: <https://www.ablebits.com/office-addins-blog/2014/12/17/excel-and-or-xor-not-functions/>.
- 5 Official Secrets Act 1972, Laws of Malaysia Act 88. 2017, Available: <http://www.agc.gov.my/agcportal/uploads/files/Publications/LOM/EN/Act%2088.pdf>.
- 6 N. Ismail, and E. L. Y. Cieh, Beyond Data Protection: Strategic Case Studies and Practical Guidance. London: Springer Science and Business Media, 2013.
- 7 J. Holgersson, E. Söderström, F. Karlsson, and K. Hedström, "Towards a roadmap for user involvement in e-government service development," International Conference on Electronic Government, 2010, pp. 251-262.
- 8 Y. Taher, W. J. V. D. Heuvel, S. Koussouris, and C. Georgousopoulos, "Empowering citizens in public service design and delivery: A reference model and methodology," European Conference on a Service-Based Internet, 2011, pp. 129-136.
- 9 G. Skinner, H. Song, and E. Chang, "An information privacy taxonomy for collaborative environments," Information Management and Computer Security, 14(4), 2006, pp. 382-394.
- 10 TRUSTe's, 3rd party data collection principles from TRUSTe. 2012, Available: <http://www.truste.org>.
- 11 Laws of Malaysia, Act 709: Personal Data Protection Act 2010. 2010, Available: http://www.pdp.gov.my/images/LAWS_OF_MALAYSI_A_PDPA.pdf.
- 12 Information Commissioner's Office, Guide to the General Data Protection Regulation (GDPR). 2013, Available: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/>.
- 13 P. M. Schwartz, and D. J. Solove, Reconciling personal information in the United States and European Union. 2013, Available: http://scholarship.law.gwu.edu/faculty_publications/956.
- 14 University of York, General data protection regulation. 2016, Available: <https://www.york.ac.uk/records-management/dp/policy/introduction-policy>.
- 15 A. Schulman Inc., Policy on data protection and privacy of personal data. 2016, Available: https://www.aschulman.com/UserFiles/ASI_Data_Privacy.pdf.
- 16 EU Data Protection Directive, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, art. 2(a). O.J. (C 93), ("Data Protection Directive"), 1995.
- 17 R. Turn, "Classification of personal information for privacy protection purposes," ACM National Computer Conference and Exposition, 1976, pp. 301-307.

- 18 National Archives, Malaysian information classification. 2010, Available: <http://www.arkib.gov.my>

AUTHORS PROFILE



Nor Asiakin Hasbullah is a Senior Lecturer in Department of Computer Science in the National Defence University of Malaysia (UPNM) with a Ph.D. in Information Technology and Quantitative Sciences from MARA University of Technology Malaysia. Her research interests are in the field of privacy, data protection, information security and ethics in ICT. She is a member of Malaysia Board of Technologist and Informatics Intelligence Special Interest Group, UPNM. She has published and presented most of her research findings to various international conferences and articles in international journal.



Norulzahrah Mohd Zainudin is a Lecturer in the Department of Computer Science at Faculty of Defence Science and Technology, National Defence University Malaysia (UPNM). She received her MSc at Universiti Putra Malaysia and joined Military Academy of Malaysia in 2002. Her main research interests are in the areas of Forensic Computing, Online Social Networks and Computer Intelligence. She has published a number of papers in international journals and conferences. Currently she is a member of Informatics Intelligence Special Interest Group, UPNM.



Noor Afiza Mat Razali is a senior lecturer in the National Defence University of Malaysia and visiting lecturer at Management and Science University. She worked in mass media and information technology sectors for 9 years prior to becoming researcher and lecturer with expertise in Information and Cyber Security field, Disaster Management and Malaysia-Japan relation. Her research interest includes Network, Cyber and Information Security, Education for Cyber Security, Kansei Engineering, Environmental and Disaster Management



Norshahriah Wahab, currently as Senior Lecturer at Computer Science Department in Faculty of Defence Science and Technology, UPNM. Her expertise area covers Visual Informatics and Human Computer Interaction.