Categorization for the Security Awareness Domain and Resource (SADAR) Model in the Organization: A Clustering Statistical Analysis

Muharman Lubis, Rokhman Fauzi and Muhammad Azani Hasibuan
Telkom University, Jalan Telekomunikasi No. 1, Bandung, 40257
Email: muharmanlubis@telkomuniversity.ac.id

Abstract. Employee in the organization or customer from company can significantly affect the overall strength of security of the system as the human factor become the common problem that lead to damage or disaster. Unfortunately, most security solutions offered in the market do not consider user as a critical security element. In this study, it is proposed to evaluate the awareness of the user regarding information security problems from three type of antecedents, which are individual, institutional and environmental. In order to collect data, questionnaires have been sent to the users using systematic sampling based on the domain of pattern concern on the role of responsibility. The cluster analysis method was applied to sample of users in categories related to the various items represented in the type of antecedents. It has objective to place a set of object in such a way that have tendency to be similar to each other in one group and tend to be differ in the opposite group. Then, the development of SADAR emphasized four groups namely define, devote, discover and direct.

1. Introduction
In general, the organizations rely heavily on information to manage and operate their tasks and activities, so the security of information is one of the most important problems faced by researchers and professionals. As the system have been developed, new security issues has been occurred with new type of technology solutions. At first glance, it seems the problem have been solved but in reality it provides vulnerability and weakness in the system that are extremely difficult to dismantle due to the constraints of the time, complexity and compatibility from application. Therefore, the identification of user in term of characteristic, behaviour and perception should be a starting point for the proposal on a possible new concept of global security awareness as majority problem might be root to the awareness level involving the intention, motivation and desire. One solution proposes the concept of trust while the other proposes to the social observation, which refer to the readiness and knowledge complementary. Since users are rarely considered in safety guidelines, those solution face difficulty in the implementation. On the other hand, the user generally considers their members or friend in the environment to understand certain issues of security regardless the accuracy or the correctness of those information. Thus, it is necessary to have framework model to understand the process of guarding and protecting information as well to secure the resource considering the diverse characteristic of employees and customers. The assessment and categorization of the user's awareness should help to come up with new initiative of security solutions, taking into account the user as a key ingredient of the ICT system to develop proper policy and procedure. The objective of raising awareness about information security is to improve everyone
concern that they are exposed to the opportunities and challenges in the current threat landscape, change behaviours that involve human risks and create or strengthen a safe organizational culture.

2. Literature Review

Drawing a dendogram or tree diagram is a usual way of visualizing the progress of cluster analysis by showing the distance level at which a combination of objects and clusters is found [1]. A study was found that users with a university degree were more aware of security problems when using the email system. On the contrary, the users with a single address, with less experience, have less knowledge of the security problems of the electronic mail system were less aware. Although all the groups obtained a higher general utility rate than the less conscious, the first four groups did not score much more, with a difference of less than 5% [2]. In the literature, there is an agreement that employees are the weakest path or channel in the security of organization or institutions. The attempt to protect the information assets should involve the level of knowledge of information security among the IT users of the organization. It can only be achieved through physical and technical security measures and utilize the conscious of security awareness of IT users in the location of work [3]. Importantly, the main objective of information security is to guarantee business continuity and minimize commercial damage by preventing and minimizing the impact of security incidents. When an organization applies information security management, it must retain three main components. First, privacy by protecting sensitive information from unauthorized disclosure or unclear objection. Second, integrity by maintaining the accuracy, correctness and completeness of the information stored. And third, availability by providing continuous service to the vital information to the authorized users when it is necessary [5].

Employees can be affected by any type of information security incident, such as worms, viruses or phishing attacks, either in private or operational contexts. Therefore, the awareness of information security can be formed through this negative events that in the future increase the victim’s supportability by gaining more interest of knowing how to prevent such incidents so the others do not have the same condition [6]. Meanwhile, it is believed that close peers, colleagues, partners and friends have a social impact which lead to provide monitoring and capacity building service to advance development effectiveness or even the opposite result due various factors such as conflict of interest, incoordination, misunderstanding and the income gaps. The idea of social impact is the result of the personal influence and consternation of isolation from the group, as well as the sensitivities to follow rather than to lead [7]. On the other hand, another critical factors to achieve optimum level of management of information security is related to the effective compliance with policies and procedure, which the adequate integration of people, processes and technology implemented smoothly. It can be obtained through several approaches that aligned with training program, key performance indicator and proper incentive. However, the results must be evaluated regularly to see how successful the approach due to different type of culture and work pattern in the organization [8]. Meanwhile, it is also widely recognized that the employee of an organization often become the source of implementation failure and primary reason that information security has been violated, but before raising awareness, the organization must realize what we have, what is worth and what is at risk. According to Kearney [10], people can only avoid the security breaches if they are realized of the risks and learned to perform safe behaviors as integral portion of their regular job. Thus, each organization should promote a culture that encourage them to contribute to the attempts to secure the information assets of the organization and put maximum effort to protect the company against internal and external threats. Furthermore, the opinion survey also confirmed that the role of mediation in the employees' commitment play significant percentage to stimulate best practice of work habits and consistent performance [16]. Ironically, the data subject is willing to disclose personal data to gain popularity and appreciation on their achievements, which they often believe it does not pose any risk at all as long as they are used for good and appropriate purposes [13-15].

3. Research Methodology

Every node in the cluster tree consist of a set of homogenous data, which are joined together in the multi stage process by using a degree of similarity. The process continues by same principle until all the nodes
are in the tree, which gives a visible preview of the data throughout the cluster group. Therefore, the total number of clusters is not predefined before the tree starts with the hierarchical method was used as the most common approach for group analysis [11]. The measurement of the Euclidean distance of the difference was also chosen because of the standardized data gradually and algorithmically and there were no extreme values in the sample while the main objective was to have groups of a similar size [12]. It is necessary to standardize the variables when the values are at different levels or the difference varies a lot, which is not the case in this example. After classifying and evaluating the groups, each user group was analyzed with respect to awareness of security issues along with additional variables. The number of clusters is defined graphically as a result of the group analysis. The steps by which an answer algorithm of the resulting graph must be intercepted and discovered, based on the number of clusters and the distance between them. In this case, the algorithm was stopped between 25%, 22%, 17% and 9% in the group process and 8% and 6% in the ward linkage to complete the assembly process due to very large distance between the groups and the results in groups of similar size with gender was used as label case.

### Table 1. Agglomeration Schedule for IC, RT and RI.

| Stage | Cluster Combined | Coefficients | Stage | Cluster First Appears | Next Stage |
|-------|------------------|--------------|-------|-----------------------|------------|
|       | Cluster 1 | Cluster 2 | | Cluster 1 | Cluster 2 | |
| 1     | 1       | 5         | .193  | 0         | 0         | 3         |
| 2     | 9       | 10        | .210  | 0         | 0         | 4         |
| 3     | 1       | 3         | .216  | 1         | 0         | 5         |
| 4     | 8       | 9         | .227  | 0         | 2         | 9         |
| 5     | 1       | 4         | .233  | 3         | 0         | 7         |
| 6     | 2       | 6         | .245  | 0         | 0         | 8         |
| 7     | 1       | 7         | .247  | 5         | 0         | 8         |
| 8     | 1       | 2         | .254  | 7         | 6         | 9         |
| 9     | 1       | 8         | .257  | 8         | 4         | 0         |

### Table 2. Agglomeration Schedule for AT, BV and CT.

| Stage | Cluster Combined | Coefficients | Stage | Cluster First Appears | Next Stage |
|-------|------------------|--------------|-------|-----------------------|------------|
|       | Cluster 1 | Cluster 2 | | Cluster 1 | Cluster 2 | |
| 1     | 3       | 10        | .202  | 0         | 0         | 3         |
| 2     | 2       | 9         | .210  | 0         | 0         | 5         |
| 3     | 3       | 7         | .224  | 1         | 0         | 4         |
| 4     | 5       | 6         | .226  | 3         | 0         | 5         |
| 5     | 2       | 3         | .227  | 2         | 4         | 6         |
| 6     | 2       | 4         | .251  | 5         | 0         | 7         |
| 7     | 2       | 6         | .263  | 6         | 0         | 8         |
| 8     | 1       | 2         | .268  | 0         | 7         | 9         |
| 9     | 1       | 6         | .272  | 6         | 0         | 0         |

### Table 3. Agglomeration Schedule for PC and TP.

| Stage | Cluster Combined | Coefficients | Stage | Cluster First Appears | Next Stage |
|-------|------------------|--------------|-------|-----------------------|------------|
|       | Cluster 1 | Cluster 2 | | Cluster 1 | Cluster 2 | |
| 1     | 1       | 2         | .194  | 0         | 0         | 3         |
| 2     | 5       | 6         | .203  | 0         | 0         | 5         |
| 3     | 1       | 3         | .229  | 1         | 0         | 4         |
| 4     | 1       | 7         | .239  | 3         | 0         | 5         |
| 5     | 1       | 5         | .240  | 4         | 2         | 6         |
| 6     | 1       | 4         | .257  | 5         | 0         | 0         |
Table 4. Agglomeration Schedule for SP and PP.

| Stage | Cluster 1 Combined | Cluster 2 | Coefficients | Stage Cluster First Appears | Cluster 1 | Cluster 2 | Next Stage |
|-------|-------------------|-----------|--------------|----------------------------|-----------|-----------|------------|
| 1     | 1 2               |           | .162         | 0 0                        | 0 0       | 0 2       |
| 2     | 1 3               |           | .183         | 1 0                        | 1 0       | 3         |
| 3     | 1 5               |           | .211         | 2 0                        | 2 0       | 5         |
| 4     | 4 9               |           | .219         | 0 0                        | 0 0       | 5         |
| 5     | 1 4               |           | .234         | 3 4                        | 3 4       | 8         |
| 6     | 5 8               |           | .254         | 0 0                        | 0 0       | 7         |
| 7     | 5 7               |           | .260         | 5 0                        | 5 0       | 8         |
| 8     | 1 6               |           | .271         | 5 7                        | 5 7       | 0         |

Bigger distance in dendogram presents difference between groups and it is presented graphically as higher jump. This procedure results in fourteen clusters that are representing fourteen groups of users. Basically, there are four groups in this study, which are individual antecedents (self-attitude/AT, self-behavior/BV, self-cognitive/CT), institutional antecedents (policy compliance/PC, training program/TP), environmental antecedents (peer performance/PP, social pressure/SP) and mediator antecedents (intention to comply/IC, perceived threats/RT, religious indicator/RI). Meanwhile, there are 4 ATs (1: interest, 2: common, 3: pirated, 4: safety), 3 BVs (1: access, 2: experience, 3: check), 3 CTs (1: advice, 2: read, 3: advanced), 3 PCs (1: rule, 2: quick, 3: standard), 4 TPs (1: occurrence, 2: counseling, 3: regulation, 4: consequences), 4 PPs (1: obedience, 2: capacity, 3: accommodation, 4: performance), 5 SPs (1: plan, 2: affect, 3: share, 4: workload, 5: penalties), 4 ICs (1: impact, 2: limitation, 3: trust, 4: responsibility), 3 RTs (1: negligence, 2: attack, 3: disrupt) and 3 RIs (1: facilities, 2: program, 3: organization). Through agglomeration schedule for IC, RT and RI showed that cases 1 and 5 have the smallest distance (coefficient 0.183), so it is merged with case 3 in the stage 3, case 4 in the stage 5, case 7 in the stage 7, case 2 in the stage 8 and lastly case 8 in the stage 9. Meanwhile, cases 9 and 10 have the second smallest distance (coefficient 0.210), which is merged with case 8 in the stage 4 and case 1 in the stage 9. The other three agglomeration schedule follow the same patterns from small to large coefficient until all cases merge together. It showed the total error made at each stage of clustering when two separate cases in the first phase, then clusters of cases are carried out together to present a new cluster. Meanwhile, a large value of the error show that two discrete things have been combined, which there is a prominent typology at fusion level.
Figure 1. Hierarchical cluster using average linkage (between groups).

In general, the adjacency between two clusters are measured as between-groups linkage (average linkage) for average value of every possible distances between the cases of clusters A and B while Ward’s method for a cluster the total of squares distances of every cases from the centroid. The other method can be used is nearest neighbor (single linkage) with minimum of all possible distance, centroid clustering (other linkage) with distance between the centroids, furthest neighbor (complete linkage) with maximum, within-groups linkage (other linkage) with average and median clustering (other linkage). In short, it can be interpreted that persons with higher religious indicator such as encourage campus to provides good facilities for praying (RI1), happy when campus promotes the religious program and feel that religious organization give positive impact to their productivity will have trusted campus management to protect personal data (IC3), have high motivations to anticipate the threats by aware on the danger of negligence (RT1) but feel that it does not matter to violate the IS rule as long as no impact at all (IC1). Meanwhile, persons who fell it is common to see somebody’s personal data (AT2) and glad to read Internet article on privacy protection (CT2), will have preference to use pirated software (AT3) and think that technology advancement has double edge sword (CT3). On the other hand, persons who feel campus game counseling on regular basis for IS issues (TP2) and think it is necessary to learn relevant IT regulation (TP3) will have understood easily the rule of campus (PC1) and access related privacy protection quickly (PC2). In general, dendogram is counting the number of lines that cross with the observations in that group are represented by the branches of the dendogram that spread out below the line. For example, at a height of 6 (x-axis), there are three lines that defines a three-cluster solution. By following the line down through all its branches, there are the names of the variables that are included in these three clusters, which then at height of 5, two clusters broken down into small groups. In short, the development of SADAR can be classified into 4 groups of D, which are cluster 1 (Define): PC1, RT1, CT2, IS1, IS2, PP3, BV3, PC2, TP2, IS3, SP1, RI3, TP2, PP1, PP2, IC3, SP5, then cluster 2 (Devote): IC1, PC3, PP4, RT2, TP3, PP4, RT2, TP3, CT3, RT3, CT1, RI2, RI1, then cluster 3 (Discover): IC4, TP1, AT2, AT1, SP2, BV2, SP4 and cluster 4 (Direct): AT3, IC2, SP3, BV1, AT4.
4. Conclusion

The clustering analysis is a flexible exploratory method that allows to repeat the classification in a larger sample size, either for general or specific users, as users of an ICT system in a given organization. The process of classification of the user's awareness based on similar attributes should be done to develop the model and criteria to be considered before execute such training program or security policy within organization. This study has limitation with mostly based on the method used, which is is cluster analysis. Since this is an exploratory method, researchers do not know the expected results, and if the resulting groups will be sufficiently differentiated for further statistical analysis. Specific solutions can be directed to different groups of users with respect to the results of cluster analysis. For example, it can be developed for a particular company after its employees are classified and analyzed, according to the company's security requirements, which vary in organization due to its culture, budget, structure, mission and much more. The initial result of information security awareness based on this method is related to the model development, which is still in the phase 1, or in this case called SADAR1 consist of 4D’s namely define, devote, discover and direct.

Acknowledgment

The publication is part of Research of Platform for Development of Maturity Model to Assess the Level of Information Security Awareness in the Institution. The research has been funded and supported by the Directorate of Research and Community Service (PPM), Telkom University. We also thank Enterprise System Development Research Group, School of Industrial Engineering and all of the participants.

References

[1] Solic K, Tovjanin and Ilakovac V. 2012. Assessment methodology for the categorization of ICT system users security awareness. IEEE MIPRO.

[2] Galba T, Solic K and Nenadic K. 2018. Evidential reasoning approach to behavioural analysis of ICT users’ security awareness. Tehnicki vjesnik 25(2). pp. 309-315.
[3] Lebek B, Uffen J, Breitner MH, Neumann M and Hohler B. 2013. Employees’ information security awareness behaviour: a literature review. *Hawaii Int. Conf. on System Sciences*.

[4] Bashorun A, Worwui A and Parker D. 2013. Information security: to determine its level of awareness in an organization. *IEEE Int. Conf. on Application of Inf. and Comm. Tech.*

[5] Sari PK, Candiwan and Trianasari N. 2014. Information security awareness measurement with confirmatory factor analysis. *IEEE Int. Symposium on Tech. Management and Emerging Tech.*

[6] Haeussinger FJ and Kranz JJ. 2013. Information security awareness: its antecedents and mediating effects on security compliant behaviour. *Int. Conf. on Information Systems*.

[7] Ahlan A, Lubis M and Lubis AR. 2015. Information security awareness at the knowledge-based institution: its antecedents and measures. *Procedia Computer Science*. 72, 361–373.

[8] Eminagaoglu M, Ucar E and Eren S. 2009. The positive outcomes of information security awareness training in companies – a case study. *Information Security Technical Report* 14, pp. 223-229.

[9] Metalidou E, Marigani C, Trivelas P, Eberhagen N, Skourlas C and Giannakopoulos G. 2014. The human factor of information security: Unintentional damage perspective. *Procedia – Social and Behavioral Sciences* 147, 424-428.

[10] Kearney P. 2010. *Security: the human factor*. Cambridgeshire: IT Governance Publishing.

[11] Jobson JD. 1992. *Applied multivariate data analysis*. Springer, pp. 518-616.

[12] Mooi E and Sarstedt M. 2011. *A concise guide to market research*. Cluster Analysis, Chapter 9, Springer, pp. 237-284.

[13] Lubis M, Kusumasari TF and Hakim L. 2018. The Indonesia Public Information Disclosure Act (UU-KIP): Its Challenges and Responses. *Int. J. of Electrical and Com. Engineering*, 8(1), Feb, pp. 94-103

[14] Rosmaini E, Kusumasari TF, Lubis M and Lubis AR. 2018. Study to the Current Protection of Personal Data in the Educational Sector in Indonesia. *Journal of Physics: Con. Series*, 978, 012037.

[15] Lubis M and Kartiwi M. 2014. Privacy concern and perceived benefits towards privacy and election in Indonesia. *Proc. ICT4M*.

[16] Trivelas P. 2011. Work motivation and job performance of frontline employees: the mediating role of organizational commitment. *Proc. IEEM*, 1878-1882.