ISSN 1996-3343

Asian Journal of Applied Sciences

http://knowledgiascientific.com

Knowledgia SCIENTIFIC
A Place to Publish Outstanding Research

Asian Journal of Applied Sciences 7 (5): 343-359, 2014
ISSN 1996-3343 / DOI: 10.3923/ajaps.2014.343.359

A Conceptual Model for Holistic Classification of Insider

Ikuesan Richard Adeyemi, Shukor Abd Razak and Mazleena Salleh
Information Assurance and Security Research Group, Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

Corresponding Author: Ikuesan Richard Adeyemi, Information Assurance and Security Research Group, Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia Tel: +60143249376

ABSTRACT

The process through which an insider to an organization can be described or classified is lined within the orthodox paradigm of classification in which an organization considers only subject with requisite employee criterion as insider to that organization. This is further clouded with the relative rigidity in operational security policies being implemented in organizations. Establishing investigation process in instances of misuse occurrence and/or ascertaining efficiency of staff member using such paradigm is maligned with endless possibilities of uncertainties. This study therefore, proposes a holistic model for which insider classification can be crystallized using the combination of quantitative research process and analysis of moment structure evaluation process. A full comprehension of this proposition could serve as a hinge through which insider misuse investigation can be thoroughly carried out. In addition, integrating this paradigm into existing operational security policies could serve as a metric upon which an organization can understand insider dynamics, in order to prevent misuses and enhance staff management.

Key words: Insider distinction, insider investigation, subject-object relationship, dynamic insider, organization employee, user interaction

INTRODUCTION

The 2010 Cyber security watch survey with over 500 respondents uncovers a hidden reality that shows that considerable percentage of cybercrimes are committed by neither an insider nor an outsider, thus tagged unknown. Such attacks includes unauthorized access to use of information, system and network, intentional exposure of private/sensitive information. Whilst some attacks could be deniably noninsider origin such as spyware, others are arguably attacks that emanates from within the organization. This level of relatively high unknowns could be attributed to the relative ambiguity and anonymity inherent in the traditional information security policies which views an insider from the paradigm of subject with legitimate access right and clearance privileges only. This paradigm of defining insider based on subject-object relationship in isolation does not reveal the reality of operational process of human behaviour. Furthermore, it neglects the sociological paradigm of human interaction (psychosocial attributes), while it exposes the inherent vulnerabilities in systems (Magklaras and Furnell, 2001). The implication of these ambiguity and obscurity range from the lack of detection and inaccurate identification of insider misuse, to anonymity inclusion which hinders the possibilities of investigation, while either tarnishing reputation of organization or inhibiting operational efficiency amongst others. This study therefore, introduces a conceptual model for insider description which can be adapted into existing policies,

to project clarity in meaning and clarification of an insider. This is however not similar to the usual practice of defining insider based on intent, in which an insider is classified according to benign intention, malicious intention or erroneous action. To the best of our knowledge, this is the first time such paradigm is empirically modelled to classify insider from the longitudinal and vertical axis which presupposes that an insider to an organization, scopes beyond traditional paradigm, especially where insider misuse investigation is called into play. This study thus explores the dimensions of users, classified as insider to an organization. It begins with a brief overview of existing research on insider taxonomy followed by a theoretical bases and a description of the empirical procedures adopted for this study. The result of the study is then presented followed by a detailed discussion of the implication of the result.

RELATED WORKS

Insider taxonomy of geographical delineation identifies subjects based on logical and physical presence to an organization infrastructure as graphically represented in Fig. 1. Graphical representation in Fig. 1 shows four descriptive classifications: Classes A, B, C and D which identifies an insider as someone with physical and/or logical privilege to use a particular system/facility. Other literatures (Magklaras and Furnell, 2004; Magklaras *et al*., 2011; Neumann, 1999; Roy Sarkar, 2010; Salem *et al*., 2008) adopted the description of an insider as a subject with legitimate access right to an object.

These literatures are targeted at delineating benign insider from a malicious insider with various other taxonomies that examines the description of a malicious insider. Thus the literatures neglects the primary criteria for classification of insider. A thorough anatomy of who an insider really is spans beyond the ideal paradigm presented in these literatures. However, Adeyemi *et al*. (2013) presents the state of the art surveys on insider taxonomy. The study identified four distinct abstractions within which an insider can be defined. This includes:

- A current employee by an organization (CE)
- A laid-off employee (LE)
- A contract employee (CS)
- An affiliates to a current employee/contract staff (NCE)

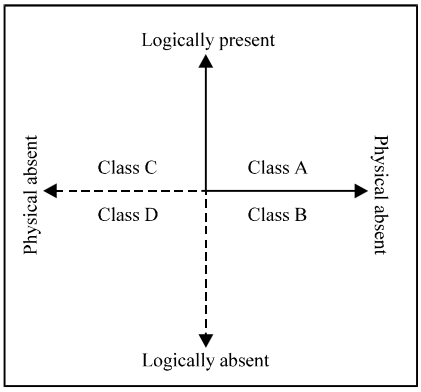

Fig. 1: Geographical taxonomy of insider

Though the study attempted to generically classify insider without prejudice to malicious taxonomy, it however failed to present an empirical validation to substantiate the abstractions. This study builds on the abstraction in Adeyemi *et al.* (2013), by extracting the psychosocial attributes using output matrix, series of self administered questionnaire and a structural modelling evaluation process. The proceeding section details the methodology and process implored in this study.

THEORETICAL DESCRIPTION OF PROPOSED INSIDER TAXONOMY

The four abstractions in (Adeyemi *et al.*, 2013) is plotted on an auto-covariance matrix/description to derive an output matrix depicted in Table 1. This output matrix depicts inter-relational inference which results in a forty dimension description that form various classes of personnel, ranging from current employee, contract staffs, collaborators and other stakeholders. These classes reveal the psychosocial tendencies in which human interaction is described, which considers it erroneous to rule out the influence, affluence and effect each class could generate.

The abstractions CE, LE, CS and NCE refers to subject in an organization, while the abstraction (also referred to as superset) CELE, CECS, CENCE, CSLE, CSNCE and LENCE refers to interaction between each identified subject. Subject-object paradigm is based on access privileges permitted to a subject on an object. Example of such is the relationship defined by the BellLa-Padular confidentiality model (Bell, 2011) in which a subject is bounded by the clearance criterion to an object. Moreover, human interaction paradigm entails the possible communication between subjects of same or differing classes, over an object of group of objects. This process may involve the use of experience garnered from previous observation, deliberate communication with other subject or even collaboration with other subjects. Example of such instance could be an IT expert who doubles as a bank cashier in addition to being a staff to technical consultancy firm which deals with banks. Another example of such a superset could be a university staff that doubles as member of two faculties within the university or instances where married couples are stationed in different department within same organization. Against this backdrop, it could be asserted that an insider transcends the boundary of subject-object paradigm. Table 2 gives a classified description which this study sought to empirically evaluate to explicate the dimension of human interaction paradigm in insider definition.

These five classifications expanded in Table 2 can be further represented using mathematically notations as shown in Eq. 1-4.

Single Variant Subject Class (SVSC): This is a subject-object paradigm. Subject in this class possess only the requisite access right and access knowledge for specific assigned responsibility in an organization.

Unique Single Variant Subject Class (USVSC): Subject in this class possess knowledge formed through the interaction between two subjects in an organization, each belonging to different departments, but same access right and requisite access knowledge within the organization. Equation 1 gives a mathematical composition of SVSC extrapolated for the outcome matrix in Table 1:

$$DVSC_{(A,B)}\,class = \sum_{n=1}^{4} A_n B_n \tag{1}$$

Table 1: Auto-covariance (outcome) matrix of insider

	Subjects-object paradigm			
Human interaction paradigm	CE	LE	CS	NCE
CE	A1B1	A2B1	A3B1	A4B1
LE	A1B2	A2B2	A3B2	A4B2
CS	A1B3	A2B3	A3B3	A4B3
NCE	A1B4	A2B4	A3B4	A4B4
CE-LE	A1B5	A2B5	A3B5	A4B5
CE-CS	A1B6	A2B6	A3B6	A4B6
CE-NCE	A1B7	A2B7	A3B7	A4B7
LE-CS	A1B8	A2B8	A3B8	A4B8
LE-NCE	A1B9	A2B9	A3B9	A4B9
CS-NCE	A1B10	A2B10	A3B10	A4B10*

*Letters "A" and "B" are arbitrary representation with no connotative meaning to the outcome matrix

Table 2: Classification of insider

Single variant subject	Unique single varian subject	Double variant subject	Unique double variant subject	Triple variant subject
CE	CE→CE	CE↔LE	CE↔CE→LE	CE↔LE→CS
LE	LE→LE	CE↔CS	CE↔CE-→CS	CE↔LE→NCE
CS	CS→CS	CE↔NCE	CE↔CE→NCE	CE↔CS→NCE
NCE	NCE→NCE	LE↔CS	LE↔CE→LE	
		LE↔NCE	LE↔LE→CS	
		CS↔NCE	LE↔LE→NCE	
			CS↔CE→CS	
			CS↔LE→CS	
			CS↔CS→NCE	
			NCE↔CE→NCE	
			NCE↔LE→NCE	
			NCE↔CS→NCE	

→: Possible communication process or acquired knowledge by a subject, ↔: Mutual communication process

This class forms the diagonal of the 4x4-subset matrix, of the auto-covariance matrix defined in Table 1. This class of subject shares the same probability of existence with SVSC.

Double Variant Subject Class (DVSC): These refer to subjects that possess knowledge formed through the interaction between member of different access knowledge, but not necessarily different access right, vice versa. Such knowledge tends to transcend the requisite access knowledge of a singular subject. Equation 2 illustrates the mathematical representation of this class:

$$DVSC_{(A,B)}\,class = \sum_{n=2}^{4} A_1B_n - \sum_{n=3}^{4} A_2B_n - A_3B_4 \quad (2)$$

Unique Double Variant Subject Class (UDVSC): These subjects possess higher knowledge of the organization, because of multi-interaction among subjects of different departments

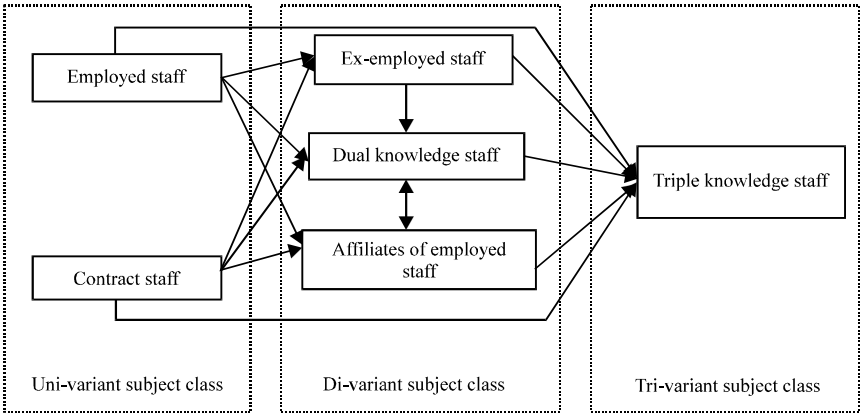

Fig. 2: Theoretical model for insider taxonomy

within an organization such that at least, two of the subjects have same access right but not necessarily within same access knowledge:

$$UDVSC_{(A,B)}\ class = \sum_{n=5}^{7} A_1B_n + \left[\sum_{n=8}^{9} A_2B_n + A_2B_5\right] + \sum_{n=3}^{5} A_3B_{2n} + \left[\sum_{n=9}^{10} A_4B_n + A_4B_7\right] \quad (3)$$

Triple Variant Subject Class (TVSC): These are subjects capacitated with the knowledge accumulated from at least three subjects within an organization, each with a distinct access knowledge and/or access right. Equation 4 gives the mathematical representation of this class of insider:

$$TVS(A,B)\ class = \sum_{n=8}^{10} A_1B_n \quad (4)$$

Equation 1-4, can be further grouped using common factor analysis. SVSC and USVSC can be called a single class, DVSC and UDVSC can be referred to as a double class, while TVSC can be called a triple class. Thus Eq. 1 and 2 represents a uni-variant subject class, Eq. 3 and 4 represents a di-variant subject class and Eq. 5 represents a group of tri-variant subject class. The interaction between subjects in each group forms the cardinal of the theoretical model as shown in Fig. 2.

Leveraging the outcome matrix, the theoretical model shown in Fig. 2 stations this study as an interactive integrated composition of the various classes of subjects considered to form the cardinal in which an insider is formed. This research thus set the paradigm of insider formation as an extension to the traditional definition of insider.

EMPIRICAL PROCEDURE OF THE PROPOSED INSIDER DESCRIPTION MODEL

This section details the quantitative factorization of the conceptual model. Self administered questionnaire instrument was developed for this process. This study adopts the process defined in Eq. 5 (Bartlett *et al.*, 2001; Israel, 1992):

Table 3: Sample size calculation

Organization classification	Population variance (%)	Precision (%)	Confidence level (%)	Estimate of population (%)	Estimated response rate (%)	Sample size
Public	50	3	95	200	70	168
Private	50	3	95	100	60	81
Government	50	3	95	100	80	81

$$n = \frac{\left\{\left(p(1-p)\right)/\left(\frac{A^2}{Z^2}\right)+\frac{P(1-p)}{N}\right.}{R} \quad (5)$$

where, n = Sample size, N = Estimate of population, P = Estimated variance in population, A = Desired precision, Z = Confidence level, R = Estimated response rate.

Table 3 elaborates on the adopted selected values for each parameter, which constitutes the sample size.

As shown in Table 3, the respondents covers three tiers of organization: Private, governement and public. The private institution represents section in observation which are owned and controlled by private individuals, such as insurance, banks and so on. The government institution represents section in investigation caseload that is controlled by the government of a nation such as the government operational centres and government administrative centres. The public institution on the other hand represents sections in investigation which are controlled by the Government through proxies such as the public libraries and public universities. Two public universities were selected to represent the public institution, a government operational centre is selected to represent the government institution and a commercial bank was selected to represents private institutions, all in two different states in Malaysia. A total of 153, 89 and 82 respondents were collected for public, government and private institutions respectively.

The questionnaire design is structured into five-phase 34-item, self rated semantic differential pattern to reflect the description of insider identified in existing literatures (Al-Morjan, 2010; Anderson, 1980; Hunker and Probst, 2011; Magklaras and Furnell, 2004; Neumann, 1999; Roy Sarkar, 2010). These phases include description of employed staff, contract staff, sacked staff, affiliates and double knowledge staffs. The pattern matrix from the analysis of the result further reveals the delineation of phases herein defined. The distribution and collection of the questionnaire spanned 7 weeks (3rd February, 2013 to 25th March, 2013). The respondents comprise top management, middle management, administrative staff, contract staff and technical staffs. We do not claim that this sample covers the entirety of the population of Malaysia but serves as representative of the Malaysia workforce.

However, a 10 days pilot survey comprising 15 respondents was conducted to initially ascertain the skewness of the questionnaire. As expected the outcome of the analysis is positively skewed. Various tests are conducted on the data, such as:

Kaiser Meyer Olkin (KMO): KMO is a sampling adequacy test, while Bartlett's measure is a Sphericity test. Kiser Meyer Olkin varies from 0-1. A value of 0 indicates low correlation among variable, while a value close to 1 indicates the high correlation. High correlation indicates that the factor analysis is reliable. Recommended range of 0.5-0.7, 0.7-0.8, 0.8-0.9 and above 0.9 to be mediocre, good, great value and superb value, respectively.

Bartlett's measure: Bartlett's measure is a test of null hypothesis, which states that the correlation matrix is an identity matrix. This illustrates that the variables are not related, thus unsuitable for factor analysis. Values of range <0.05 indicates that the data is suitable for factor analysis. Otherwise, the data is not suitable for structure detection.

Missing Completely at Random (MCAR) Test: The MCAR test is also referred to as Little's test. It is a test for biases in dataset with missing variables. A statistically significant result of MCAR indicates that the missing data in a dataset is biased, while an insignificant statistical result indicates a completely missing at random.

RESULT OF THE EMPIRICAL PROCESS

The percentage of male and female respondents are 38.3 and 60.5%, respectively which is synonymous with the findings in (Ashari, 2012) where professional and management occupational level of male to female is 39.4-60.6%. A concise description of the demography of the respondents is shown in Table 4, 35.39% of the overall respondents are top management staffs.

Little's test of MCAR: This test was carried out using the process described in Fig. 3. As shown in Fig. 3, Expectation Maximization (EM) algorithm is used. The results, as shown in Table 5, which indicates percentage of the missing values, shows that the percentage of the highest missing value (1.5%) is less than the negligible threshold of 2%.

As shown in Table 5, the result of the EM test shows it is not statistically significant (0.796), thus, indicating a randomized MCAR. Hence, the Little's MCAR test implies that data imputations can be carried out on the dataset to complete the missing values.

Data screening test: In order to identify the possible number of factors within which these data can be classed, we adopted the exploratory factor analysis on SPSS statistical tool. Exploratory Factor Analysis (EFA) is carried out on the data in accordance with the procedure described in Fig. 4.

As shown in Fig. 4, KMO and Bartlett's test are carried out on the dataset. It is used to evaluate the adequacy and appropriateness of the set of observed variable (dataset) for EFA. Table 6 shows the result of the test.

The result presented in Table 6, shows that the dataset is adequate and appropriate for EFA. Both KMO value (0.916) and Bartlett's measure value (<0.001), indicates the appropriateness of factor analysis otherwise known as structural detection on the dataset.

Table 4: Synopsis of demography of respondents

		Marital status (%)		Educational level (%)		Job experience (%)	
	Sample size						
Gender	(%)	Married	Single	High	Low	< 4 years	>4years
Female	60.5	70.07	25.85	74.15	25.85	36.05	63.95
Male	38.3	70.97	24.73	83.87	15.05	35.48	64.52

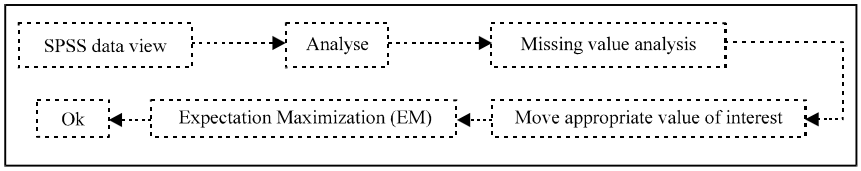

Fig. 3: Steps to little's MCAR test

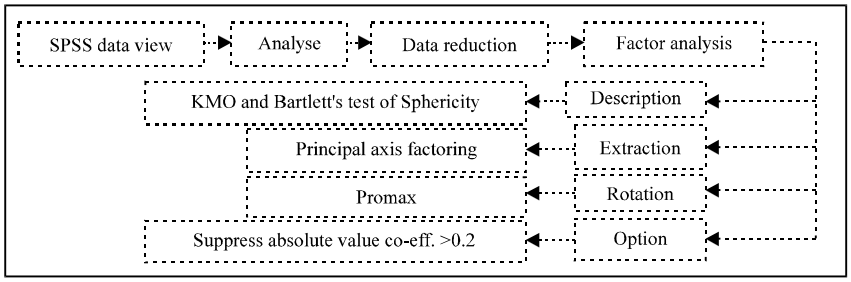

Fig. 4: Exploratory factor analysis procedure in SPSS

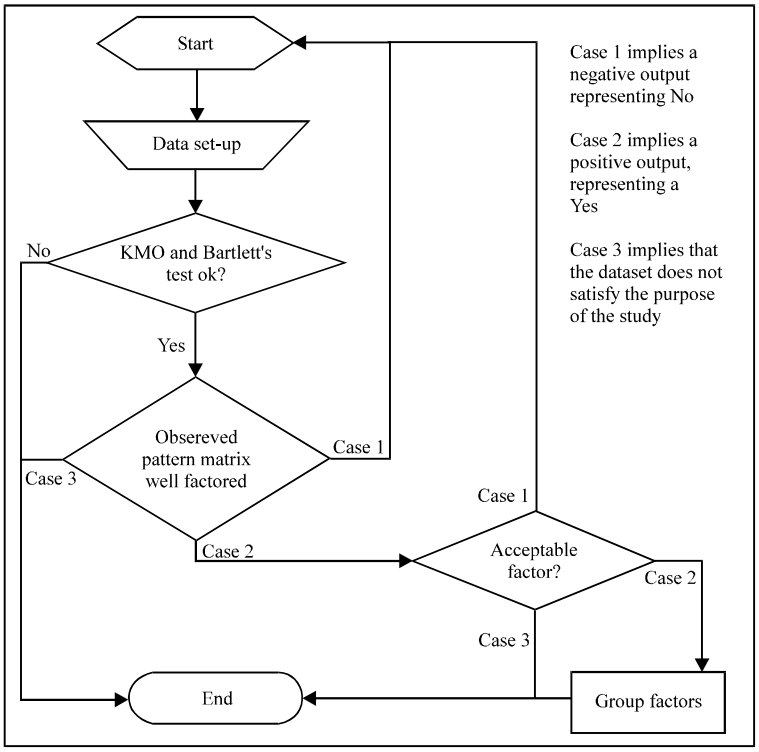

Fig. 5: Flow chart of factor analysis process

Factor analysis: Factor analysis is done to determine the number of underlying factors in a surveyed data, as well as to eliminate outlier variables. In order to derive statistically significant structure from the dataset, EM imputation is first carried out on the dataset to fill up all missing data. Figure 5 shows the flow chart of the factor analysis process.

Table 5: Little's MCAR test

Parameters	Values
Chi-Square	374.453
df	398
Significance	0.796

Table 6: KMO and Bartlett's test

Test	Values
KMO test	
Measure of sampling adequacy	0.913
Bartlett's test of sphericity	
Chi-square	6.063E3
df	561
Significance	0.000

The flow chart, as shown in Fig. 5, starts with the preparation of the data for analysis as indicated by the symbol. A manual process of identification of observable variables proceeds the preparation phase, as indicated by the symbols. This is then followed a decision on output of KMO and Bartlett's test. Further decision on the output of the KMO and Bartlett's test is carried-out through observation and processing of the pattern matrix. Case 1 indicates that the desired output criteria are not met, case 2 indicate that the desired output are satisfied while case 3 indicates that the desired output cannot be satisfied by the dataset. Upon a desired factor output, the factors are then grouped to reflect the result. The output of the pattern matrix using a principal axis factoring and Promax with Kaiser normalization is shown in Table 7.

Six factors were initially observed. However, a synopsis of the survey instrument using the principle of interpretability of factors and theoretical expectation of number of constructs (Costello and Osborne, 2005) as described in the outcome of the auto-covariance matrix presented in Table 2, this study adopted a five factor classification system. Cross loading problem was observed in the factors. Cross loading occurs when a single observable variable is loaded by different factors. In order to achieve better factorization, this study eliminated observable variables with variance cross-factor loading of more than 0.2 value.

However, after the deletion (approximate of 14% of the total observable) the study decided to terminate the elimination process. Moreover, at the point, the pattern matrix reflects the distinction among the factors, as shown in Table 7. For easy description, the factors are relabelled to reflect the underlying characteristics of the factors as detailed in Table 7. The relabelled factors are described as construct in the proceeding sections.

MODEL DESIGN AND VALIDATION

This study adopts the sequential procedure in SPSS-AMOS tool defined in Fig. 6. The construct defined in Table 8 serves as the input for this phase of the study. The flow-chart for this phase is presented in Fig. 7.

As shown in Fig. 6, the designed model comprises a measurement model and a structural model. The measurement model examines the dataset for the fabricated constructs using multiple fit indices for Goodness of Fit (GOF). These indices include Root Mean Square Error of Approximation (RMSEA), Comparative Fit Indices (CFI), ratio of Chi-square and Degree of Freedom (CMIN/DF), Composite Reliability (CR) and Average Variance Explained (AVE). The choice of these indices is based on the recommendation in (Hair *et al.*, 2010; Costello and Osborne, 2005). Measurements are based on the thumb rule for each of the GOF indices, as shown in Table 9.

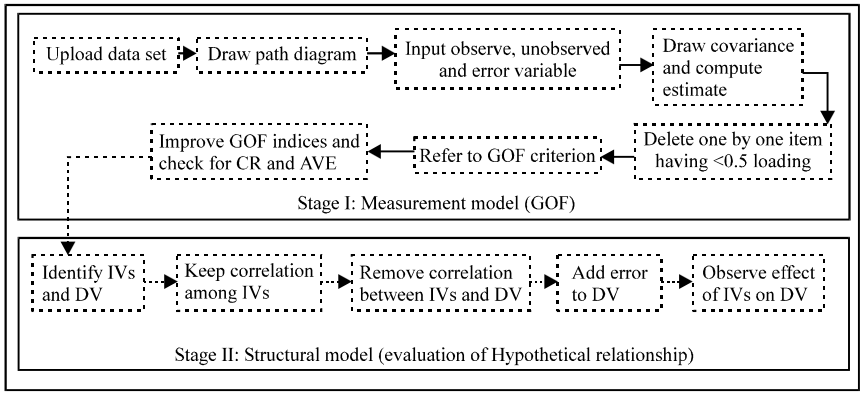

Fig. 6: Model testing and hypothesizing process

Table 7: Pattern matrix[a]

	Factor				
Observable variable	1	2	3	4	5
Employed staff 1	0.722				
Employed staff 2	0.812				
Employed staff 3	0.578				
Employed staff 4	0.724				
Contract staff 1	0.542			0.277	
Contract staff 2	0.747				
Contract staff 3	0.378	0.288			
Sacked staff 1				0.591	
Sacked staff 2	0.273			0.572	
Sacked staff 3				0.694	
Sacked staff 4	0.303			0.550	
Sacked staff 5				0.637	
Affiliates 2			0.542		
Affiliates 3			0.605		
Affiliates 4			0.508		
Affiliates 5			0.884		
Affiliates 6			0.890		
Multiple access		0.646			
Multiple access 2	0.217	0.574		-0.258	
Multiple access 3	0.241	0.688			
Multiple access 4		0.327		0.246	
Dual knowledge	-0.408	0.207		0.396	
Dual knowledge 2		0.703			
Dual knowledge 4		0.679			
Multiple knowledge		0.773			
Multiple knowledge 2		0.742			
Triple knowledge				0.207	0.633
Triple knowledge 2					0.945
Triple knowledge 3					0.769
Extraction method: Principal axis factoring					
Rotation method: Promax with Kaiser normalization					

[a]Rotation converged in 7 iterations

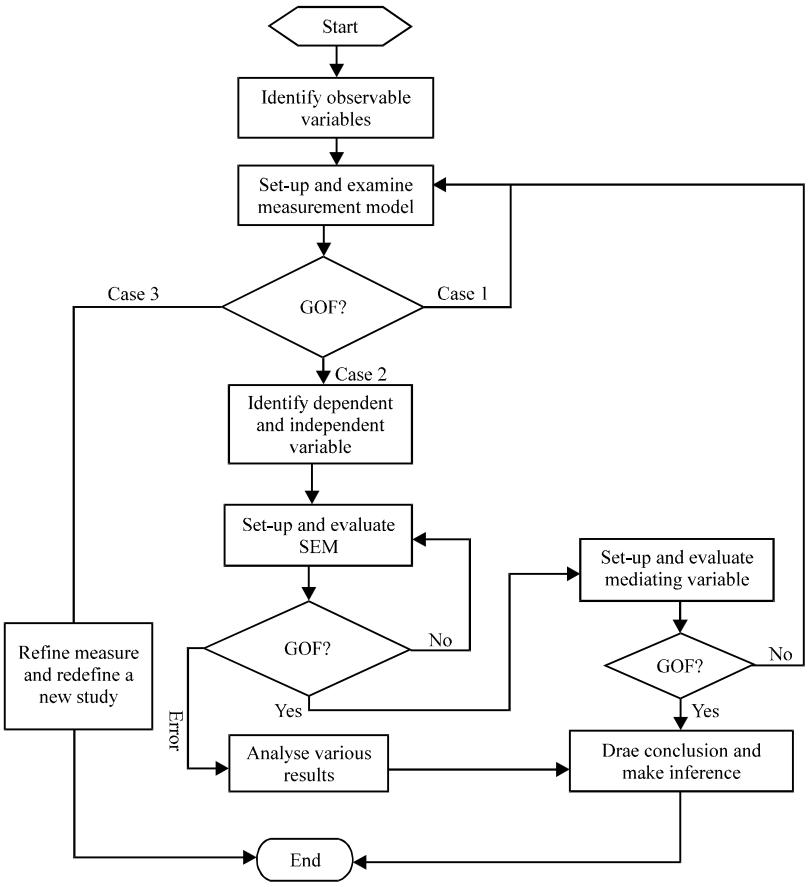

Fig. 7: Flow chart for proposed model

Following the flow chart in Fig. 7, the model was observed to achieve a substantial GOF criterion. For the first iteration process as shown in Table 10, the covariance of all the constructs (at this phase, the constructs are generally classified without distinction on latent or independent variable).

From the observation, the CMIN/DF criteria of ≤3.00 was achieved as well as the Composite Reliability (CR). However, other criterions were not achieved. To improve the fitness of the model, unobservable variables possessing highest regression coefficient weight, which falls within the same constructs are correlated. Thus, the 2nd through the 4th iteration process of the correlates unobservable variables e8 and e9, e5 and e6 and e22 and e23, respectively. After these iterations, the convergence validity measure (Average Variance Explained: AVE) of the model was observed to fall within the threshold of value >0.5.

Moreover, observable variable SS2 was observed to have regression weight lesser than 0.5. Thus, SS2 was removed from the model. Further iterations were carried out as detailed in Table 10, until an acceptable statistical validity (adequate model fit and construct validity) was achieved. The path diagram of the measurement model is presented in Fig. 8. After the 10th iteration process, the measurement model was observed to attain statistical reliability and validity.

This implies therefore that these sets of constructs can be used to study relationship between employees of an organization and affiliates as well as ex-employees which constitutes the five

Table 8: Model construct description

Factor	Underlying characteristics	Construct description	Label
1	Describes perceived employed staff in an organization	Employed staff	ES
2	Employed staff with knowledge of multiple department in an organization	Dual Knowledge Staff	DK
3	Affiliates of an employed or ex-employed staff	Affiliate of employed staff	AFF
4	Describes subject perceived to be an ex-employee of the organization	Ex-Employed Staff	SS
5	Describes subjects with knowledge of more than two department in an organization	Triple Knowledge staff	TK

Table 9: Criterion for GOF (Source Hair *et al*., 2010)

Indices	N≤250, m≤30
CMIN/DF	≤3.00
CFI	≥0.92
RMSEA	≤0.07
AVE	≥0.5
CR	≥0.7

N: No. of sample size, m: No. of observable variable

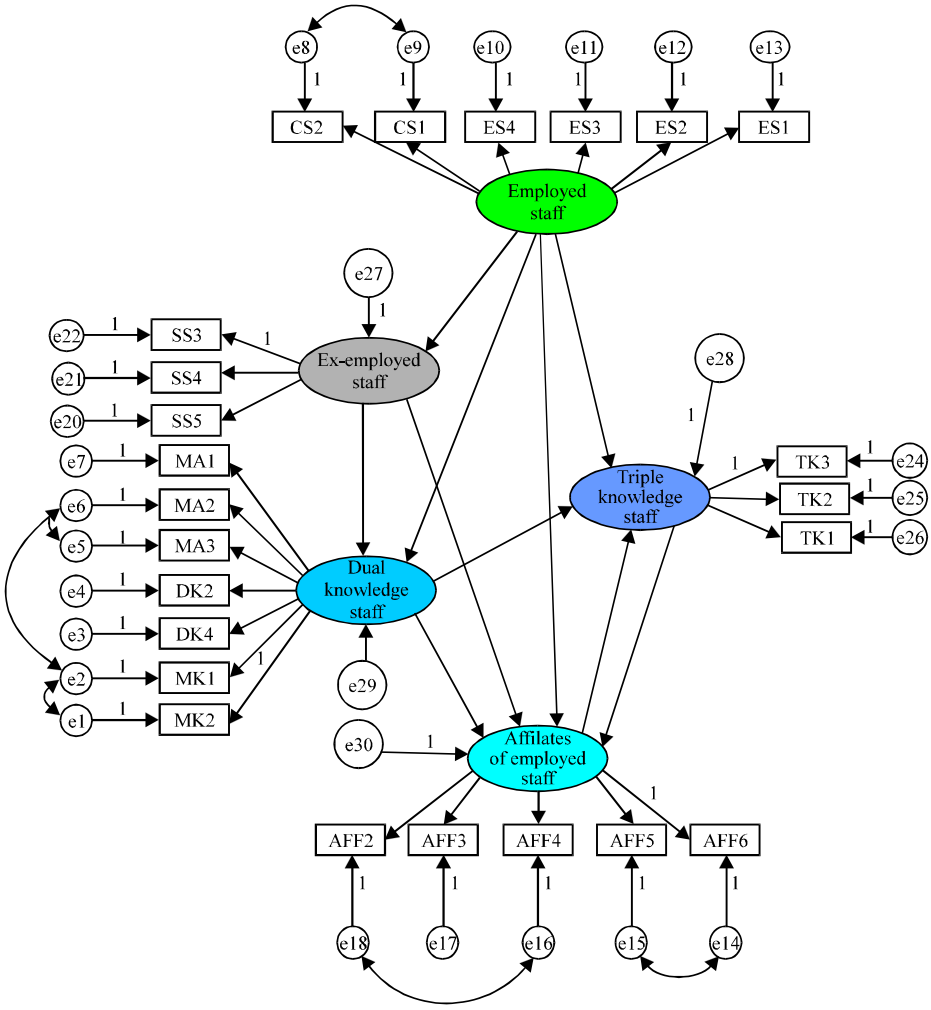

Fig. 8: Path diagram of structural equation modelling for insider taxonomy

constructs carved out in this study. In order to carry out the structural theory test, stage II of Fig. 6 was implemented. It begins with the identification of independent and dependent variables. With reference to the equations in appendix A, an employed staff through interaction with other employed staff within and without same department, can gain knowledge which could be adequate for access knowledge and/or right. Similarly, interaction between employed staff and affiliates of an employed staff as well as ex-employed staff could yield substantial information for access knowledge and/or right. Hence, this study identifies these constructs thus:

- Employed Staff (ES): Independent construct
- Dual Knowledge Staff (DK): Dependent construct
- Ex-employed staff (SS): Dependent construct
- Affiliate of employed/ex-employed staff (AFF): Dependent construct
- Triple Knowledge Staff (TK): Dependent construct

DK, SS, TK and AFF are dependent on ES. Therefore, ES is an exogenous construct (predictors), while DK, SS, TK and AFF are endogenous construct (outcome). The hypothesis to validate does as follows:

- H1 (ES→DK): Overall interaction between employee, either in same or different department could be substantial enough to the degree of the existence of dual knowledge staff, either through direct or indirect relationship
- H2 (ES→SS): Overall interaction between ES and SS could be positive such that there exist a common knowledge among them through direct or indirect relationship
- H3 (ES→AFF): Overall interaction between ES and AFF either directly or directly could be positive such that there exist the possibilities of AFF gaining access knowledge and or right
- H4 (ES→TK): Overall interaction among ESs in same or different department as well as among AFFs, could be substantial such that there exist the possibilities of a TK

IV/DV relationship was introduced as shown in Fig. 8 and correlations between the DVs were deleted. In order to evaluate the model and test the hypothesis, we ensured that the direct and indirect relations among construct of interest was added, by connecting single arrow from the predictor to the various outcomes. Furthermore, we also observed the relationship such that the two other endogenous constructs (SS, DK) serves as predictor to the other endogenous constructs (AFF, TK). This was observed to yield a significant regression coefficient in the model. As shown in Fig. 8, both direct and indirect effects are observed on the construct. H1 hypothesize the direct and indirect relationship between ES and DK, H2 hypothesize the direct relationship between ES and SS, H3 hypothesize the direct and indirect relationship between ES and AFF, while H4 hypothesize the direct and indirect relationship between ES and TK.

ES has three structural path of indirect relationship to AFF and two structural path of indirect relationship with TK as expected from the classification matrix in Table 2, thus establishing triple knowledge possibilities. The indirect relationship between ES and DK further supports the theory of the outcome matrix in Table 1. Table 10 shows the overall correlation estimate of the relationship between the constructs in the model. Table 11 shows the comparison between the structural model and the measurement model.

Table 10: Goodness of fit indices for measurement model

Stages	CMIN/DF	CFI	RMSEA	AVE	CR	Description/Outcome
Construct covariance	2.697	0.890	0.072	SS<0.5 others >0.5	All >0.7	Measurement model fit for some indices
e8↔e9	2.635	0.895	0.071	SS<0.5	All >0.7	AVE for SS = 0.482
e5↔e6	2.540	0.901	0.069	SS<0.5	All >0.7	CR for all greater than 0.87
e22↔e23	2.330	0.915	0.064	SS<0.5	All >0.7	To improve AVE, remove factor with least regression weight in the model
Delete SS2	2.378	0.915	0.065	SS<0.5	All >0.7	AVE for SS = 0.496,
e1↔e2	2.303	0.923	0.063	SS<0.5	All >0.7	CR for all greater than 0.754
e18↔e16	2.259	0.923	0.062	SS<0.5	All >0.7	AVE for SS = 0.496,
E14↔e15	2.150	0.930	0.060	SS<0.5	All >0.7	CR for all greater than 0.754
E2↔e6	2.077	0.934	0.058	SS< 0.5	All >0.7	AVE for all greater than 0.5
Delete SS6	2.065	0.939	0.057	All >0.5	All >0.77	CR for all greater than 0.754, No validity and reliability concerns. Model fit is good

↔: Indicates correlation

Table 11: Standardized total effects

	Employed staff	Ex-employed staff	Dual knowledge staff	Triple knowledge staff	Affilates of employed staff
Ex-employed staff	0.528	0.000	0.000	0.000	0.000
Dual knowledge staff	0.704	0.298	0.000	0.000	0.000
Triple knowledge staff	0.249	0.432	0.543	-0.076	0.431
Affilates of employed staff	0.262	0.715	0.460	-0.162	-0.076

Table 12: Comparison indices between structural model for insider taxonomy and the CFA measurement model

GOF Indices	Structural model	CFA model
CMIN	487.434	487.434
DF	236.000	236.000
CMIN/DF	2.065	2.065
P	0.000	0.000
RMSEA	0.057	0.057
CFI	0.939	0.939

There is significant statistical correlation among all the constructs as shown in Table 10, indicating the acceptance of the alternate hypotheses and rejection of the null hypotheses. However, statistical correlation peaks at ES→DK and troughs at ES→TK, describing the possible level of interaction between the constructs.

As shown in Table 12, there is no statistical variance between the two models. This shows that the model provides a good overall model fit which is constituent for both models. It also implies there are no interpretational confounding errors. Interpretational confounding reveals structural misspecification, as well as measurement error. Since there is no variance between the models, it can be said that the model perfectly fit for insider taxonomy.

RESULT DISCUSSION AND LIMITATION

Our result shows that there is no clear cut distinction between an employer and a contract staff in an organization. This result therefore supports the recommendation in Roy Sarkar (2010) which identifies contract staff as a potential employee and should be address as such.

From the result shown in Table 10, various observations can be inferred about the relationship between the dependent variable and the independent variables. There is a statistically significant relationship between the independent construct (ES) and the dependent constructs (DK, SS, AFF, TK). Hypothesis H1, H2, H3 and H4 therefore holds true for each of the relationships, thus establishing the basis for which DK, SS, AFF and TK can be classified in line with ES as the composition of insider to an organization. This study therefore rejects the null hypothesis of the models, favouring the acceptance of the alternate hypothesis, which describes the relationship between the constructs. Moreover, the coefficient of correlation between the independent variable and the dependent variable (0.528, 0.704, 0.249 and 0.262) depicts the operational reality of humans/subjects that contribute to the day to day activities of an organization. This explains the need for the evolved paradigm of insider taxonomy, intention dissection not with standing. From the result therefore, it follows that an insider classification system especially where insider misuse is involved, requires a thorough consideration of the possible connection between different subjects as well as between subjects and their possible affiliates. It may be possible that an affiliate perpetuate a particular act with the access right of a subject, with or without the knowledge of the subject in question. Such clarification would require a proper dissection of all the affiliation related to each subjects. Furthermore, the interoperability between subjects of different classes and clearance level may generate useful artefacts, which could have been otherwise overlooked. Appropriating this result into organizational security framework could be a possible way of identifying possible breaches and curbing insider misuse possibilities. This research thus fills the gap of identifying the operational composition of organization's day to day activities. However, this research is limited in its incapacity to delineate the perspective of insider from each categories of organization. Furthermore, it failed to provide insight into insider perception from societal differences perspective. This is anchored on the premise that (if) human interaction forms the cardinal for which holistic taxonomy of insider can be viewed, then it surmise to state that interaction differs from one society to another.

CONCLUSION AND FUTURE WORK

This study presents the result of a conceptual model for insider taxonomy from the evolving paradigm of classical insider description. Using questionnaire instrument, from three categories of organization, this study models an outcome matrix of insider taxonomy. Statistical analysis tools and structural equation modelling tool was adopted for analysis and modelling process. The sample provided the minimal requirement for which generalized findings can be extracted. The result reveals the "real operational" description of insider constituents, as against the subject-object description. From the result, is it observed that there is a statistical significance between the designed variables which defines "who an insider is". This result can be applied for investigation process of insider crime and security related alerts. Furthermore, this result can be applied in staff training as well as implemented in organizational policies to manage the effectiveness of staffs, evaluate staff propensity to malicious intention or provide interactive policies for effectiveness. This can be done by reviewing the various dependencies and level of interaction between each identified subject. This study is part of an on going research on insider taxonomy in relation to misuse investigation. As part of the continuing work, this study intends to further examine the variability in description of insider from the three distinct organizations in order to understand the level of interaction between the identified variables. Further studies on insider taxonomy based on societal differences will greatly improve this paradigm of insider definition.

APPENDIX A

Social interaction, relationship and interdependence theory suits this study. This theory "presents a logical analysis of the structure of interpersonal relationship", thus offers a conceptual framework for interpersonal situational analysis, as shown in equations given below:

$$B = f(P, E)$$

where, B signifies behavior, f(P, E) represents function of the property of the person (P) and the environment (E) in context:

$$I = f(S, A, B)$$

Interaction (I) is a function of social situation (S) between persons (A) and (B). "The option and outcome of interaction can be represented using a tool from the classic game theory, the outcome matrix". An outcome matrix describes interdependence pattern among people thus useful in describing social situation, in that it describes the intricacies and degree of interaction. This follows suit with Locard's exchange principle of exchange theory of transference.

REFERENCES

Adeyemi, I.R., S.A. Razak, Z. Anazida and N.A.N. Azhan, 2013. A Digital forensic investigation model for insider misuse. Adv. Comput. Sci., Eng. Inf. Technol., 225: 293-305.

Al-Morjan, A.A., 2010. An investigation into a digital forensic model to distinguish between insider and outsider. Ph.D. Thesis, Software Technology Research Laboratory, Faculty of Technology, De Montfort University, UK.

Anderson, J.P., 1980. Computer security threat monitoring and surveillance. Technical Report, James P. Anderson Company, Fort Washington, PA., USA., February 26, 1980.

Ashari, H., 2012. Adaptive leadership: An opportunity for women to have access to higher responsibilities. http://www.parlimen.gov.my/images/webuser/artikel/ro/halisah/Dec%202012%20Adaptive%20Leadership%20Halisah%20Ashari_2.pdf

Bartlett, J.E., J.W. Kotrlik and C.C. Higgins, 2001. Organizational research: Determining appropriate sample size in survey research. Inform. Technol. Learn. Perform. J., 19: 43-50.

Bell, D.E., 2011. Bell-La Padula Model. In: Encyclopedia of Cryptography and Security, Van Tilborg, H.C.A. and S. Jajodia (Eds.). 2nd Edn., Springer, New York, USA., ISBN-13: 9781441959058, pp: 74-79.

Costello, A.B. and J.W. Osborne, 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assess. Res. Eval., 10: 1-9.

Hair, J.F., W.C. Black, B.J. Babin and R.E. Anderson, 2010. Multivariate Data Analysis. 7th Edn., Prentice-Hall, Upper Saddle River, NJ., USA., ISBN-13: 978-0138132637, Pages: 816.

Hunker, J. and C.W. Probst, 2011. Insiders and insider threats: An overview of definitions and mitigation techniques. J. Wireless Mobile Networks Ubiquitous Comput. Dependable Appl., 2: 4-27.

Israel, G.D., 1992. Determining sample size. Fact Sheet PEOD-6, University of Florida, Gainesville. http://www.soc.uoc.gr/socmedia/papageo/metaptyxiakoi/sample_size/samplesize1.pdf

Magklaras, G. and S. Furnell, 2004. The insider misuse threat survey: Investigating IT misuse from legitimate users. Proceedings of the 5th Australian Information Warfare and Security Conference, November 25-26, 2004, Perth Western Australia, pp: 42-51.

Magklaras, G., S. Furnell and M. Papadaki, 2011. LUARM: An audit engine for insider misuse detection. Int. J. Digital Crime Forensics, 3: 37-49.

Magklaras, G.B. and S.M. Furnell, 2001. Insider threat prediction tool: Evaluating the probability of IT misuse. Comput. Secur., 21: 62-73.

Neumann, P.G., 1999. Risks of insiders. Commun. ACM, 42: 160-160.

Roy Sarkar, K., 2010. Assessing insider threats to information security using technical, behavioural and organisational measures. Inform. Secur. Tech. Rep., 15: 112-133.

Salem, M.B., S. Hershkop and S.J. Stolfo, 2008. A Survey of Insider Attack Detection Research. In: Insider Attack and Cyber Security: Beyond the Hacker, Stolfo, S.J., S.M. Bellovin, A.D. Keromytis, S. Sinclair, S.W. Smith and S. Hershkop (Eds.)., Vol. 39, Springer, New York, ISBN: 9781441945891, pp: 69-90.