# Study for Best Data Obfuscation Techniques using Multi-Criteria Decision-Making Technique

Gayatri C. Deshmukh S.Y M. Tech. (Project Management) Department of Production Engineering and Industrial Management College of Engineering, Pune, India

#### ABSTRACT

In this paper, we describe the formatting guidelines for IJCA Journal Submission Data obfuscation (or also known as masking of data) is the process of hiding real data with random characters or data i.e. the process of protecting sensitive data from thefts and hackers. Obfuscation is applied in order to secure data that is classified as personal identifiable data, personally or commercially sensitive data. Keeping the data valid for the use of test cycles. It must also appear consistent. The objective is to protect the privacy of individuals which is getting vital for operative functioning over the internet. Privacy enforcement is being handled primarily through governments for development or testing purposes and study of various data obfuscation techniques for different applications and their comparison study using statistical parameters. In this paper, we study the comparison of various data obfuscation techniques. The results strongly suggest that replacement methods can be used across the domains starting such as finance, banking, military, health care sector, and identity management domain. Different data obfuscation models such as encryption, shuffling, substitution, and masking out are compared.

#### **General Terms**

Computer Science, Project Management, Decision Making, Information System

#### **Keywords**

Data obfuscation, Multi Criteria Decision Making, Encryption, Substitution, Shuffling

# 1. INTRODUCTION

## 1.1 What Is Data Obfuscation?

Data obfuscation is a method of creating a structurally similar but inauthentic version of an organization's data that can be used for purposes such as software testing and user training [1]. A functional substitute must be provided when the real data is not required is the main purpose of data obfuscation process. Even though most of the IT Organizations have rigorous security controls to guard production data in storage or in business use, sometimes that same data has been used for operations that are less secure. This kind of issue occurs usually when these operations are outsourced and the organizations have less control over the environment. Due to compliance legislation, many organizations prefer not to keep real data in their working environment. In data obfuscation, the data format remains the same, only with changed values. The data may be changed using different kind of techniques such as encryption, character shuffling, and character or word substitution.

# **1.2 Why data obfuscation?**

S. M. Patil, PhD Assistant Professor, Department of Production Engineering and Industrial Management College of Engineering, Pune, India

For any organization protecting their customer information is one of the most vital tasks. There could be legal and financial obligations as well. Because of such a risk, now a day, it's getting more significant that companies must ensure their customer information is only made available to those employees who are authorized to view and interact with it.

While encryption could be used, this obfuscation method is expensive and renders the information in an unusable format. Data obfuscation make sure that information is in the valid format that can be used for development and testing and also providing enough valid information to service a company's customers

## 2. DATA OBFUSCATION TECHNIQUE

There are different types of obfuscation techniques. Some of the widely used obfuscation technique are considered for this study.

## 2.1 Substitution

Substitution is simply replacing one value with another. For example, the mask might substitute a Person's first and last names with names from a random phone book entry. The resulting data still constitutes a name, but has no logical relationship with the original real name unless you have access to the original substitution table. It allows the obfuscation to be performed in such a manner that another authentic looking value can be substituted for the existing value. There are several data field types where this approach provides optimal benefit in disguising the overall data sub set as to if it is a masked data set. For example, if dealing with source data, which contains customer records, real life surname or first name, can be randomly substituted from a supplied or customized look up file. If the first pass of the substitution allows for applying a male first name to all first names, then the second pass would need to allow for applying a female first name to all first names where gender equals "F". Using this approach, we could easily maintain the gender mix within the data structure, apply anonymity to the data records but also maintain a realistic looking database, which could not easily be identified as a database consisting of masked data. This substitution method needs to be applied for many of the fields that are in DB structures across the world, such as telephone numbers, zip codes and postcodes, as well as credit card numbers and other card type numbers such as Social Security numbers and Medicare numbers, where these numbers need to conform to a checksum test of the Luhn algorithm [2]

In most cases the substitution files will need to be extensive, so having large substitution datasets as well the ability to apply customized data substitution sets should be a key element of the evaluation criteria for any data obfuscation solution. [2]

## 2.2 Masking Out

Character scrambling or masking out of certain fields is also another simplistic yet very effective method of protecting sensitive information. It is really an extension of the previous method of nulling out but there is greater emphasis on keeping the data real and not fully masked altogether. This is commonly applied to credit card data in production systems. For instance, an operator in a Call Centre might bill an item to a customer's credit card. They then quote a billing reference to the card with the last 4 digits of XXXX XXXX XXXX 6789. As an operator they can only see the last 4 digits of the card number, but once the billing system passes the customer's details for charging, the full number is revealed to the payment gateway systems. This system is not very effective for test systems but is very useful for the billing scenario detailed above. It is also commonly known as a dynamic data obfuscation method. [2]

#### 2.3 Shuffling

The shuffling method is a very common form of data obfuscation. It is similar to the substitution method but it derives the substitution set from the same column of data that is being masked. In very simple terms, the data is randomly shuffled within the column. However, if used in isolation, anyone with any knowledge of the original data can then apply a "What If" scenario to the data set and then piece back together a real identity. The shuffling method is also open to being reversed if the shuffling algorithm can be deciphered. Shuffling however has some real strengths in certain areas. If for instance, the end of year figures for financial information in a test data base, one can mask the names of the suppliers and then shuffle the value of the accounts throughout the masked database. It is highly unlikely that anyone, even someone with intimate knowledge of the original data could derive a true data record back to its original values. [2]

#### 2.4 Nulling / Deletion

Sometimes a very simplistic approach to obfuscate is adopted through applying a null value to a particular field. The null value approach is really only useful to prevent visibility of the data element. In almost all cases it lessens the degree of data integrity that is maintained in the masked data set. It is not a realistic value and will then fail any application logic validation that may have been applied in the front-end software in the system under test. It also highlights to anyone who wishes to reverse engineer any of the identity data that data obfuscation has been applied to some degree on the data set. [2]

## 2.5 Encryption

Encryption is often the most complex approach to solving the data obfuscation problem. The encryption algorithm often requires that a "key" be applied to view the data based on user rights. This often sounds like the best solution but in practice the key may then be given out to personnel without the proper rights to view the data and this then defeats the purpose of the obfuscation exercise. Old databases may then be copied with the original credentials of the supplied key and the same uncontrolled problem lives on. Recently, the problem of encrypting data while preserving the properties of the entities got a recognition and newly acquired interest among the vendors and academia. New challenge gave birth to algorithms called FPE (format preserving encryption). They are based on the accepted AES algorithmic mode that makes them being recognized by NIST. [2]

# 3. MULTICRITERIA DECISION MAKING TECHNIQUES

Multiple-criteria decision-making (MCDM) or multiplecriteria decision analysis (MCDA) is a sub-discipline of operations research that explicitly evaluates multiple conflicting criteria in decision making (both in daily life and in settings such as business, government and medicine). [3] In our daily lives, we usually weigh multiple criteria implicitly and we may be comfortable with the consequences of such decisions that are made based on only intuition. [4]

#### Why do we focus on Multi-Criteria Decision Analysis as our chosen decision-making technique?

Conflicting criteria are typical in evaluating options: cost or price is usually one of the main criteria, and some measure of quality is typically another criterion, easily in conflict with the cost. In purchasing a car, cost, comfort, safety, and fuel economy may be some of the main criteria we consider – it is unusual that the cheapest car is the most comfortable and the safest one. In portfolio management, we are interested in getting high returns but at the same time reducing our risks, but the stocks that have the potential of bringing high returns typically also carry high risks of losing money. In a service industry, customer satisfaction and the cost of providing service are fundamental conflicting criteria [3]

There are various kinds of Multi-Criteria Decision Analysis Methods as explained below:

#### 3.1 SAW Method

This is also called the weighted sum method (Fishburn, 1967) and is the simplest, and still the widest used MADM method. Here, each attribute is given a weight, and the sum of all weights must be 1. Each alternative is assessed regarding every attribute. The overall or composite performance score of an alternative is given by Equation 1.

 $Pi = \sum_{j=1}^{M} WjMij$  (Equation 1)

Previously, it was argued that SAW should be used only when the decision attributes can be expressed in identical units of measure (e.g., only dollars, only pounds, only seconds, etc.). However, if all the elements of the decision table are normalized, then SAW can be used for any type and any number of attributes. In that case, Equation 3.1 will take the following form:

 $Pi = \sum_{j=1}^{M} Wj(Mij) normal$  (Equation 2)

Where (mij) normal represents the normalized value of mij, and Pi is the overall or composite score of the alternative Ai. The alternative with the highest value of Pi is considered as the best alternative. The attributes can be beneficial or nonbeneficial. When objective values of the attribute are available, normalized values are calculated by (mij) K/ (mij) L, where (mij) K is the measure of the attribute for the K-th alternative, and (mij) L is the measure of the attribute out of all alternatives considered. This ratio is valid for beneficial attributes only. A beneficial attribute (e.g., profit) means its higher measures are more desirable for the given decisionmaking problem. By contrast, non-beneficial

Attribute (e.g., cost) is that for which the lower measures are desirable, and the normalized values are calculated by (mij) L/ (mij) K. If the restriction that the sum of all weights is to be equal to 1 is relaxed, then Equation 3 can be used and this method is called simple multiple attribute rating Technique (SMART).

$$Pi = \frac{\sum_{j=1}^{M} Wj(Mij)normal}{\sum_{j=1}^{M} Wj}$$
(Equation 3)

#### 3.2 WPM Method

This method is similar to SAW. The main difference is that, instead of addition in the model, there is multiplication (Miller

and Starr, 1969). The overall or composite performance score of an alternative is given by Equation 4.

 $Pi = \prod_{j=1}^{M} (Mij) normal^{Wj}$  (Equation 4)

The normalized values are calculated as explained under the SAW method. Each normalized value of an alternative with respect to an attribute, i.e., (mij) normal, is raised to the power of the relative weight of the corresponding attribute. The alternative with the highest Pi value is considered the best alternative. [5]

#### 3.3 TOPSIS Method:

The TOPSIS method was developed by Hwang and Yoon (1981). This method is based on the concept that the chosen alternative should have the shortest Euclidean distance from the ideal solution, and the farthest from the negative ideal solution. The ideal solution is a hypothetical solution for which all attribute values correspond to the maximum attribute values in the database comprising the satisfying solutions; the negative ideal solution is the hypothetical solution for which all attribute values in the database. TOPSIS thus gives a solution that is not only closest to the hypothetically best, that is also the farthest from the hypothetically worst. The main procedure of the TOPSIS method for the selection of the best alternative from among those available is described below:

Step 1: The first step is to determine the objective, and to identify the pertinent evaluation attributes.

Step 2: This step represents a matrix based on all the information available on attributes. This matrix is nothing but the decision table shown in Table 3.1. Each row of this matrix is allocated to one alternative, and each column to one attribute. Therefore, an element mij of the decision table 'D' gives the value of the j-th attribute in original real values, that is, non-normalized form and units, for the i-th alternative. In the case of a subjective attribute (i.e., objective value is not available), a ranked value judgement on a scale is adopted. Once a subjective attribute is represented on a scale, then the normalized values of the attribute assigned for different alternatives are calculated in the same manner as that for objective attributes.

Step 3: Obtain the normalized decision matrix, Rij. This can be represented as

 $Rij = \frac{Mij}{[\sum_{j=1}^{M} Mij^2]^{1/2}}$  (Equation 8)

Step 4: Decide on the relative importance (i.e., weights) of different attributes with respect to the objective. A set of weights wj (for j=1, 2, ..., M) such that  $\sum wj =1$  may be decided upon.

Step 5: Obtain the weighted normalized matrix Vij. This is done by the multiplication of each element of the column of the matrix Rij with its associated weight wj. Hence, the elements of the weighted normalized matrix Vij are expressed as:

$$Vij = Wj * Rij$$
 ... (Equation 9)

Step 6: Obtain the ideal (best) and negative ideal (worst) solutions in this step. The ideal (best) and negative ideal (worst) solutions can be expressed as:

$$V + = \left\{ \left( \sum_{j=1}^{Max} \frac{Vij}{j}, j \in J \right), \left( \sum_{j=1}^{Min} \frac{Vij}{j}, j \in J' \right), i = 1, 2, \dots N \right\}$$

$$V += \{V1+, V2+, V3 + \dots Vm+\}$$

$$V - = \{ \left( \sum_{j}^{Min} j, j \in J \right), \left( \sum_{j}^{Min} j, j \in J' \right), i = 1, 2, ..., N \}$$
$$V - = \{ V1 -, V2 -, V3 - \cdots Vm - \}$$

Where J = (j = 1, 2... M) / j is associated with beneficial attributes, and J' = (j = 1, 2... M) / j is associated with nonbeneficial attributes Vi+ indicates the ideal (best) value of the considered attribute among the values of the attribute for different alternatives. In the case of beneficial attributes (i.e., those of which higher values are desirable for the given application), Vi + indicates the higher value of the attribute. In the case of non-beneficial attributes (i.e., those of which lower values are desired for the given application), Vj+ indicates the lower value of the attribute. Vi - indicates the negative ideal (worst) value of the considered attribute among the values of the attribute for different alternatives. In the case of beneficial attributes (i.e., those of which higher values are desirable for the given application), Vj- indicates the lower value of the attribute. In the case of non-beneficial attributes (i.e., those of which lower values are desired for the given application), Vj indicates the higher value of the attribute

Step 7: Obtain the separation measures. The separation of each alternative from the ideal one is given by the Euclidean distance in the following equations.

$$S_i^+ = \left\{ \sum_{j=1}^{M} (V_{ij} - V_j +)^2 \right\}^{0.5}$$
,  $i = 1, 2, ... N$ 

$$\left\{ \sum_{j=1}^{M} (Vij - Vj -)^2 \right\}^{0.5} , i = 1, 2, ... N$$
... (Equation 10)

Step 8: The relative closeness of an alternative to the ideal solution, Pi, can be expressed in this step as follows.

$$Pi = \frac{S_i^-}{S_i^+ + S_i^-} \qquad \dots \text{ (Equation 11)}$$

Step 9: A set of alternatives is generated in the descending order in this step, according to the value of Pi indicating the most preferred and least preferred feasible solutions. [5]

## 3.4 VIKOR Method

 $S_i =$ 

The foundation for compromise solution was established by Yu (1973) and Zeleny (1982) and later advocated by Opricovic and Tzeng (2002, 2003, 2004, and 2007) and Tzeng *et al.* (2002a, 2002b, 2005). The compromise solution is a feasible solution that is the closest to the ideal solution, and a compromise means an agreement established by mutual concession. The compromise solution method, also known as the VIKOR (VIšekriterijumsko KOmpromisno Rangiranje) method, was introduced as one applicable technique to implement within MADM. The multiple attribute merit for compromise ranking was developed from the Lp-metric used in the compromise programming method (Zeleny, 1982).

$$L_{p,I} = \{\sum_{j=1}^{M} (wj [(mij)max - (mij)] / [(mij)max - (mij)min])p \} \} / p \qquad \dots \text{ (Equation 12)}$$

Within the VIKOR method L1, i and Lp, i are used to formulate the ranking measure. The main procedure of the VIKOR method is described below:

Step 1: The first step is to determine the objective, and to identify the pertinent evaluation attributes. Also determine the best, i.e., (mij) max, and the worst, i.e., (mij) min, values of all attributes.

Step 2: Calculate the values of Ei and Fi:

$$Ei = \sum_{j=1}^{\infty} wj[(mij)max - (mij)]/[(mij)max - (mij)min]]$$

 $Fi = Maxm of \{wj [(mij)max - (mij )] / [(mij)max - (mij)min] , j = 1, 2, ..., M\}$ Step 3: Calculate the values of Pi: Pi = v ((Ei - Ei - min) / (Ei - max - Ei - min)) + (1 - v) ((Fi - Fi - min) / (Fi

$$-max - Fi - min))$$

Where Ei-max is the maximum value of Ei, and Ei-min the minimum value of Ei; Fi-max is the maximum value of Fi, and Fi-min is the minimum value of Fi... v is introduced as weight of the strategy of 'the majority of attributes'. Usually, the value of v is taken as 0.5. However, v can take any value from 0 to 1.

Step 4: Arrange the alternatives in the ascending order, according to the values of Pi. Similarly, arrange the alternatives according to the values of Ei and Fi separately. Thus, three ranking lists can be obtained. The compromise ranking list for a given v is obtained by ranking with Pi measures. The best alternative, ranked by Pi, is the one with the minimum value of Pi.

Step 5: For given attribute weights, propose a compromise solution, alternative Ak, which is the best ranked by the measure P, if the following two conditions are satisfied (Tzeng et al., 2005):

Condition 1: 'Acceptable advantage' P (Ak) - P (Al)  $\geq$  (1/ (N-1) Al is the second-best alternative in the ranking by P.

Condition 2: 'Acceptable stability in decision making' alternative Ak must also be the best ranked by E and/or F. This compromise solution is stable within a decision-making process, which could be: 'voting by majority rule' (when v > 0.5 is needed) or 'by consensus' (when  $v \approx 0.5$ ) or 'with veto' (when v > 0.5).

If one of the conditions is not satisfied, then a set of compromise solutions is proposed, which consists of:

Alternatives Ak and Al if only condition 2 is not satisfied Alternatives Ak, Al..., Ap if condition 1 is not satisfied; Ap is determined by the relation  $P(Ap) - P(Al) \approx (1/(N-1))$ .

VIKOR is a helpful tool in MADM, particularly in a situation where the decision maker is not able, or does not know how to express preference at the beginning of system design. The obtained compromise solution could be accepted by the decision makers because it provides a maximum 'group utility' (represented by Ei-min) of the 'majority', and a minimum of individual regret (represented by Fi-min) of the 'opponent' (Opricovic and Tzeng, 2002, 2003, 2004, 2007). The Compromise solutions could be the basis for negotiations, involving the decision makers' preference by attribute weights. [5]

## 4. CONSTRAINTS

Whenever any decision is to be made is, it is always being made on certain group of criteria. These criteria should be well defined and must be used as vital input to decision making. The criteria which affect the decision may have either positive or negative impact on the decision. Those attributes which affect decision positively are known as beneficial attributes, such as profit generated. On the other hand, criteria which affects decision negatively is known as non-beneficial attribute in the decision making while selecting best data obfuscation technique, following criteria are found to be important:

## 4.1 Reversible (REV)

If some technique which is being used for obfuscation can be reversed to get original data then it can be hazardous in case of data security. There are further aspects such as what is probability of achieving original data, what is probability that after reversing the data got is same as original data, etc. Accumulating all these aspects according to expert's judgement method, the weightage got is 0.3. As this attribute negatively affects decision, this is non-beneficial attribute, that is, lower the impact higher the reference being given to this attribute.

## 4.2 Referential Integrity Maintenance (RI)

Many times, some attribute in one table is mapped to some another attribute or element in a different table, this is called as reference. In order to keep database schema in proper format, all these references must be maintained. This is known as referential integrity maintenance. Obfuscation technologies must maintain referential integrity when data the process of obfuscation is being done. This will make sure that the data is present in valid format. Therefore, any obfuscation technique which maintains this referential integrity must be given preference. This is beneficial attribute which is given weightage of 0.5 as by the experts.

#### **4.3** Ease of Development (ED)

There are efforts required to implement the any obfuscation technique. As this is one part of complete project, organizations can't invest huge amount of efforts in implementing obfuscation technique. So, lesser efforts required preference is given to that technique. So higher the ease of development, higher the preference being given to that technique. That's why weightage given to this attribute is 0.2 by exerts and this is beneficial attribute.

# 5. EXPERIMENTS & RESULTS

Questionnaire method for data collection was used to get the data points for experimentation.

# 5.1 Questionnaire Method

The questionnaire is one of the best method for collecting data in research projects. A questionnaire is a research instrument consisting of a series of questions (or other types of prompts) for the purpose of gathering information from respondents. [6] The questionnaire mainly focused on above 3 attributes i.e. Ease of development of algorithm, Maintenance of referential integrity & Reversibility.

The sample items from questionnaire is given below

#### Table 1. Questionnaire for Data Collection

Reversibility (REV)		
Chances of Reverting algorithm		
Security aspects maintained by the algorithm		
Referential Integrity (RI)		
Referential Integrity Constrain		
Changes of rework in case of loss of referential integrity		
Reusability of algorithm		
Efforts Required (ER)		
Consistency		
Effect of size of data		
Space complexity of algorithm		
Efforts required for obfuscation		

The scale used for questionnaire was of 5 units (0-5) which implied 0 being worst & 5 being the best scores. Ten experts were consulted and the scores given by these experts were averaged out. The experts consulted were personnel from various level such Solution architect, project manager, technical consultant, developers, testers, etc. The summary obtained after expert's onion is given in Table 2.

Sr. No.	Obfuscation Technique	Average of Experts Votes		
		ER	RI	REV
1	Substitution	3.90	4.00	0.70
2	Masking Out	2.15	2.67	2.90
3	Nulling	2.90	0.47	2.70
4	Encryption	3.25	3.73	3.10
5	Shuffling	3.10	2.80	2.90

 Table 2. Scores for Different Alternatives

The study was carried out using various decision-making methods as mentioned in section 3. The experimental results were obtained as follows

#### 5.1.1 SAW Method

In the SAW Method, as mentioned weighted sum is calculated. After applying the equation 1 & 2 on the alternatives as mentioned in Table 2, we got following results

Table 3. Normalized Table (SAW Method)

Obfuscation Technique	ED	RI	REV
Substitution	1.000	1.000	1.000
Masking Out	0.551	0.667	0.241
Nulling	0.744	0.117	0.259
Encryption	0.833	0.933	0.226
Shuffling	0.795	0.700	0.241

#### Table 4. Results by SAW Method

Obfuscation Technique	Pi
Substitution	1.000
Masking Out	0.516
Nulling	0.285
Encryption	0.701
Shuffling	0.581

Thus, Preferences obtained by SAW Method are as follows (Table 5):

#### Table 5. Preferences by SAW Method

Obfuscation Technique	SAW
Substitution	1
Encryption	0.701
Shuffling	0.581
Masking Out	0.516
Nulling	0.285

#### 5.1.2 WPM Method:

In the WPM Method, as mentioned weighted Product is calculated. After applying equations as mentioned in section 3.2 on the alternatives as mentioned in Table 2, we got following results

 Table 6. Normalized Table (WPM Method)

Obfuscation Technique	ED	RI	REV
Substitution	1.000	1.000	1.000
Masking Out	0.551	0.667	0.241
Nulling	0.744	0.117	0.259
Encryption	0.833	0.933	0.226
Shuffling	0.795	0.700	0.241

Table 7. Results by WPM Method

Obfuscation Technique	Pi
Substitution	1.000
Masking Out	0.473
Nulling	0.215
Encryption	0.596
Shuffling	0.522

Thus, Preferences obtained by WPM Method are as follows:

Table 8. Preferences by WPM Method

<b>Obfuscation Technique</b>	WPM
Substitution	1
Encryption	0.596
Shuffling	0.522
Masking Out	0.473
Nulling	0.215

#### 5.1.3 TOPSIS Method:

After Applying TOPSIS method (as given in Section 3.3) for Multi-Criteria decision making we got the

Table 9. Normalized Table (TOPSIS Method)

Obfuscation Technique	ED	RI	REV
Substitution	0.112	0.300	0.036
Masking Out	0.062	0.200	0.149
Nulling	0.083	0.035	0.138
Encryption	0.093	0.280	0.159
Shuffling	0.089	0.210	0.149

#### Table 10. Results TOPSIS Method

Obfuscation Technique	Pi
Substitution	1
Masking Out	0.510
Nulling	0.094
Encryption	0.662
Shuffling	0.548

Thus, Preferences obtained by TOPSIS Method are as follows (Table 11):

#### Table 11. Preferences by TOPSIS Method

Obfuscation Technique	TOPSIS
Substitution	1
Encryption	0.662
Shuffling	0.548
Masking Out	0.510
Nulling	0.094

# 5.1.4 VIKOR's Method:

After Applying VIKOR's method (as given in Section 3.4) for Multi-Criteria decision making we got the following tables for values of Ei& Fi:

#### Table 12. Calculation of Ei & Fi Values

<b>Obfuscation Technique</b>	Ei	Fi
Substitution	0	0
Masking Out	0.664	0.275
Nulling	0.864	0.500
Encryption	0.412	0.300
Shuffling	0.536	0.275

#### Table 13. Results VIKOR Method

<b>Obfuscation Technique</b>	Qj	Si	Ri
Substitution	0	0	0
Masking Out	0.659	0.664	0.275
Nulling	1.000	0.864	0.500
Encryption	0.538	0.412	0.300
Shuffling	0.585	0.536	0.275

Thus, Preferences obtained by VIKOR Method are as follows (Table 14):

Table 14	. Preferences	by	VIKOR Method
----------	---------------	----	--------------

Obfuscation Technique	VIKOR
Substitution	0
Encryption	0.538
Shuffling	0.585
Masking Out	0.659
Nulling	1

#### 6. RESULTS

The results obtained from different decision making techniques are as follows:

Simple Additive Weighing Method (SAW Method) indicated that most preferred technique for data obfuscation is substitution, while it rated Encryption as second preferred option.

#### Table 15: Scores Obtained by SAW Method

Obfuscation Technique	Pi
Substitution	1.000
Masking Out	0.516
Nulling	0.285
Encryption	0.701
Shuffling	0.581



Figure 1: Results by SAW Method

Weighted Product Model Method (WPM Method) showed same results as of Simple Additive Weighing Method. It, also, rated substitution as most preferred technique for data obfuscation.

<b>Table 16: </b> \$	Scores Ob	tained by	WPM	Method
----------------------	-----------	-----------	-----	--------

Obfuscation Technique	Pi
Substitution	1.000
Masking Out	0.473
Nulling	0.215
Encryption	0.596
Shuffling	0.522



Figure 2: Results by WPM Method

By TOPSIS method Substitution found to be best suited technique and nulling to be worst suited technique for data obfuscation.

**Table 17: Scores Obtained by TOPSIS Method** 

<b>Obfuscation Technique</b>	Pi
Substitution	1
Masking Out	0.510
Nulling	0.094
Encryption	0.662
Shuffling	0.548



By VIKOR Method, Substitution was rated to be least rated which indicates most preferred alternative, it also provides Acceptable stability in decision making as Si & Ri are also preferred among the other alternatives.

<b>Obfuscation Technique</b>	Qj	Si	Ri
Substitution	0	0	0
Masking Out	0.659	0.664	0.275
Nulling	1.000	0.864	0.500
Encryption	0.538	0.412	0.300
Shuffling	0.585	0.536	0.275

Table 14. Scores Obtained by VIKOR Method



Figure 4: Results by VIKOR Method

By applying various kinds of decision making methods on the available alternatives based on the given weightages of the attributes we got following results:

 Table 15. Consolidated Results achieved by SAW, WPM,

 TOPSIS and VIKOR Method

Obfuscation Technique	SAW	WPM	TOPSIS	VIKOR
Substitution	1.000	1.000	1.000	0.000
Encryption	0.701	0.596	0.662	0.538
Shuffling	0.581	0.522	0.548	0.585
Masking Out	0.516	0.473	0.510	0.659
Nulling	0.285	0.215	0.094	1.000



Figure 5: Consolidated Results achieved by SAW, WPM, TOPSIS and VIKOR Method

## 7. CONCLUSION AND FUTURE SCOPE

Thus, after comparing the output given by decision making methods such as SAW, WPM, TOPSIS, VIKOR, etc. We got following preferences:

#### Table 16. Ranking Obtained using decision making

Obfuscation Technique	Rank
Substitution	1
Encryption	2
Shuffling	3
Masking Out	4
Nulling	5



Figure 5: Consolidated Results achieved by SAW, WPM, TOPSIS and VIKOR Method

We got the Preference being given to 'Substitution' technique. This gives inference that for data obfuscation substitution technique is best suited in order to make optimized use of resources, maintain integrity & security aspects.

Further study can be carried out, to compare effectiveness of above mentioned obfuscation techniques, in various environment such as On-the-fly data masking, data masking on cloud, etc. Depending upon type of environment, size of data, frequency of data updates, etc, a study can be made to analyze the performance of obfuscation technique under various situations and live environment.

#### 8. REFERENCES

- Ravikumar G K, Manjunath T N, Ravindra S Hegadi, Umesh I M, 2011, A Survey on Recent Trends, Process and Development in Data Masking for Testing
- [2] A Net 2000 Ltd. White Paper, 2016, Data Masking: What You Need to Know, What You Really Need To Know Before You Begin
- [3] Madurika, HKGM, & Hemakumara, GPTS., 2015, Gis Based Analysis for Suitability Location Finding in The Residential Development Areas of Greater Matara Region. International Journal of Scientific & Technology Research, 4(8)
- [4] Rew, L., 1988, Intuition in Decision-making Journal of Nursing Scholarship. 20 (3): 150–154. doi:10.1111/j.1547-5069. 1988.tb00056
- [5] R. Venkata Rao, 2007, Decision Making in the Manufacturing Environment Using Graph Theory and Fuzzy Multiple Attribute Decision Making Methods

- [6] Gault, RH, 1907, "A history of the questionnaire method of research in psychology". Research in Psychology.
- [7] Evangelos Triantaphyllou, 2011,"Multi-Criteria Decision Making Methods: A Comparative Study".
- [8] Securosis, L.L.C., 2012, Understanding and Selecting Data Masking Solutions: Creating Secure and Useful Data
- [9] An Oracle White Paper, 2013, Data Masking Best Practice
- [10] L. Karlitasari, D. Suhartini and Benny,2017, "Comparison of simple additive weighting (SAW) and composite performance index (CPI) methods in employee remuneration determination"
- [11] Ricardo Jorge Santos, Jorge Bernardino, Marco Vieira,2011, A Data Masking Technique for Data Warehouses
- [12] Rupa Parameswaran, Douglas M Blough, 2007, Privacy Preserving Collaborative Filtering using Data Obfuscation
- [13] David E. Bakken, Rupa Parameswaran And Douglas M. Blough, Andy A. Franz And Ty J. Palmer, 2004, Data Obfuscation: Anonymity and Desensitization of Usable Data Sets
- [14] Wei Xu, Fangfang Zhang and Sencun Zhu, 2012, The

Power of Obfuscation Techniques in Malicious JavaScript Code: A Measurement Study

- [15] A Hidayat, H Mustafidah, and A Suyadi. 2015, Penerapan Metode Simple Additive Weighting (Saw) Untuk Sistem Pendukung Keputusan Penilaian Kinerja Dosen Di Universitas Muhammadiyah Purwokerto. Prosiding SENATEK 2015. Technique Faculty. Universitas Muhammadiyah Purwokerto. Purwokerto. ISBN 978-602-14355-0 -2, pp. 367-374.
- [16] A Memariania, A Aminib and A Alinezhadc, 2009, Sensitivity Analysis of Simple Additive Weighting Method (SAW): The Results of Change in the Weight of One Attribute on the Final Ranking of Alternatives. Journal of Optimization in Industrial Engineering. Article 2. Volume 2. Issue 4. Page 13-18.
- [17] G S Pandian, N Jawahar, and SP Nachiappan, 2013, Composite Performance Index for Sustainability. IOSR Journal of Environmental Science, Toxicology and Food Technology (IOSR-JESTFT) e-ISSN: 2319-2402, p-ISSN: 2319-2399.Volume 3, Issue 1. PP 91-102.
- [18] Turban, Efrain, Aronson, and Jay, 2001, Decision Support System and Intelligent System, Prentice Hall, New Jersey.
- [19] Valensia, 2012, Sistem Pendukung Keputusan penilaian Kinerja Karyawan Dengan Menggunakan Metode Simple Additive Weighting (SAW).