Proceedings of the Pakistan Academy of Sciences:
A. Physical and Computational Sciences 54 (1): 55–70 (2017)

ISSN: 2518-4245 (print), 2518-4253 (online)

Pakistan Academy of Sciences

Research Article

Analysis of Vulnerability of Keys in a Watermarking System for Attack Susceptibility

Wahid Rehman[1*], Aihab Khan[1], and Basheer Ahmad[2]

[1]Department of Computing and Technology, Iqra University, Islamabad, Pakistan
[2]Department of Management Sciences, Iqra University, Islamabad, Pakistan

Abstract: Recent research brought up numerous techniques for copyright protection and tamper proofing of relational databases along with proof of robustness, etc. However, these techniques are short of presenting a generalized method for susceptible key-based attacks. In this research, we proposed a framework for the analysis of watermarking system against susceptibility to key attacks. We identified two primary concepts of attack models, SKMDs (Single Key Multiple Datasets) and MKsSD (Multiple Keys Single Dataset). These attack models make variants of single and multiple datasets by the usage of single and multiple keys for watermark insertion. The relationship between various pairs of original and watermarked datasets is then statistically analyzed to determine the linearity among datasets. The strength of the attack models is measured by multivariate and discriminant analysis methods like Wilks' lambda, Pillai's trace test, and Box's M test. The empirical analysis shows that MKsSD model in a watermarking system has high significance as compared to SKMDs. We conclude that SKMDs model is more vulnerable to key-based attacks than MKsSD model even by varying watermarking system parameters.

Keywords: Watermark, dataset security, key-based attack, single key multiple datasets (SKMDs), multiple keys single dataset (MKsSD)

1. INTRODUCTION

Since the last decade, watermarking is successfully being used for copyright protection and tamper-proofing of digital assets [1-3]. In general, a secret watermark is embedded into the original data to generate watermarked data for ownership proof, etc. of digital assets. The characteristics of a watermarking system presented in the literature include robustness, capacity, fragility, imperceptibility, blindness and key-based systems. Our main focus is on the attribute of key. The key-based systems which require a secret key for watermark insertion/generation and detection/verification [5, 6] for ownership proof of relational databases.

The recent literature on watermarking [1-3, 5-8] primarily encompass watermark insertion and detection algorithms along with proof of robustness which primarily measures the resistance in the removal of embedded watermark. In general, the robustness of a watermarking system that refers to the difficulty of eliminating an embedded mark without abolishing the quality of the host database is usually addressed in the domain of subset addition, deletion, and modification attacks. Besides relational databases [1-7, 10-12], the key-based watermarking schemes are also presented in the domain of text [13, 14], images [8, 15-17], audio [9, 18] and video [19], etc. However, the existing literature in these data domains does not adequately address the susceptibility to key attacks. In this paper, we address the attack models for key-based systems in relational database domain; however, the methods and schemes presented are also equally applicable to other data domains as well.

The study of watermarking security by Lafaye [2] in the context of relational database presents two attack models; SKMDs (Single Key Multiple Datasets) and MKsSD (Multiple Keys Single Dataset).This scheme analyze the ambiguity on watermarked data positions for the AHK watermarking algorithm [1]. The scheme proposed

Received, April 2016; Accepted, March 2017
*Corresponding author: Wahid Rehman: Email: wdrehman@hec.gov.pk

by Lafaye [2] is chosen as a model for our work to analyze the susceptibility of key attacks for SKMDs and MKsSD. In general for SKMDs, the ambiguity for marks detection drops off and tends to zero for the attacker as the number of conspiring user increases. When an attacker obtains the contents of watermarked dataset, eventually, he 'she becomes aware of the marked bit positions that are used in the original dataset, whereas it is not viable in the MKsSD context to correctly predict all watermarked bit positions in multiple datasets.

The robust watermarking scheme proposed by Gupta et al. 2011 [10] worked on bucket attack model in the domain of numeric dataset. This attack produce bucket of data values with similar MSBs (most significant bits) to determine the watermarked bits from the bits collected in the bucket. Though, their scheme is key-base as it uses a secret key for watermark embedding and detection, but, this scheme does not address the susceptibility of key attacks. Khanduja et al. [3] proposed a watermarking scheme for relational databases known as Bacterial Foraging Algorithm (BFA). The proposed method uses a secret key to create database partitions and then mark is embedded in each partition. Though, their scheme is resilient to subset insertion, deletion, alteration, attribute re-ordering and linear transformation attacks but does not address the key-based attacks. Another robust watermarking scheme presented by Sion et al. [11] also utilizes the secret key to create unique and non-overlapping subsets to embed a watermark. Though their scheme does have an analysis of subset selection, addition, alteration and re-sorting attacks, but it also lacks analysis of key-based attacks.

In the domain of fragile watermarking, Camara et al. [6] present a fragile watermarking scheme to determine the authenticity of a relational database. Their scheme utilizes the secret key to partition a database into a square matrix for embedding a watermark into matrix diagonal. Though, their scheme proves to be fragile but does not provide means to defy the key-based attacks. In order to solve the original content leakage problem, Iwakiri et al. [20] proposed a scheme for fragile watermarking based on incomplete cryptography. Their proposed scheme destroys the quality of original contents to make the trial contents for conveying users over a network. The quality of trial contents is controlled with a watermarked key at the incomplete decoding process, and the user information will be embedded in the incompletely decoded contents simultaneously. Hoang et al. [21] proposed a new framework for remote multimodal biometric confirmation system based on fragile watermarking for transmitting multi-biometrics from client to server for authentication. Also, their scheme is key-based as it uses a secret key for watermarking insertion and detection. A Chaos sequence based fragile watermarking scheme for 3D models in the spatial domain is presented by Wang et al. [22]. Their scheme produces a Chaos sequence and a secret key, which is used to produce the embedded watermark into three LSBs. A hash-based dual fragile watermarking algorithm is proposed by Qian et al. [23]. In the proposed scheme, a host speech signal is transformed into a matrix and a sensitive hash function (MD5) along with a secret key generates a fragile watermark. It is to be noted that though fragile scheme, like the robust watermarking schemes are also key-based, however, both these schemes do not address the susceptibility analysis of key- based attacks.

Singh et al. [25] proposed a technique to prove joint ownership of digital images by inserting invisible digital patterns in the image. This digital pattern is made from biometric features of more than one subject in a strategic matter, so that the identification of individual subject can be done and the multiple ownership of the digital images can be established. The watermarking scheme proposed by Fu et al. [24] for joint ownership verification of relational databases. In this scheme, each owner has the same secret key that is used for the identification of ownership. In general, the main characteristic for ownership proof of joint ownership is the usage of shared secret keys.

Thus, it has been observed from existing literature that the watermarking schemes utilize a secret key for watermark embedding and verification, however, these schemes do not address the susceptibility of key-based attacks. As the security of a watermarking system primarily depends on a secret key, therefore, it is vital to analyze the key-based attacks. In this research, we proposed a framework for the analysis of the watermarking system in the domain of relational databases against susceptibility of key-based attacks. We identified two primary concepts of attack models, SKMDs (Single Key Multiple Datasets) and MKsSD (multiple keys single dataset) adopted from lafaye work [2] as discussed earlier. The usage of single

key for watermarking multiple instances of the database becomes vulnerable for all watermarked databases, as if an attacker somehow discovers the secret key during transmission, etc. However, if single or few users perform watermark embedding or detection operations, the usage of multiple keys may become performance overhead as the usual and periodical key change may be sufficient in such situations. In using multiple keys for watermarking multiple databases, if a secret key is compromised, only the security of specific database would be compromised, while the other watermarked databases would still remain secure. In contrary, by using a Single key for watermarking multiple databases, if a secret key is compromised, then the security of all the watermarked databases is compromised. Another issue that may arise with the usage of single key is that the watermarking process by a particular user cannot be verified which may result in a dispute. However, the usage of multiple Keys for watermarking multiple databases may give rise to key management and to ensure the security of multiple keys. Besides several disadvantages, the advantage of using single key based watermarking system is ease to carry out the watermark embedding and detection process as and when required. The proposed SKMDs and MKsSD attack models make variants of single and multiple datasets by the usage of single and multiple keys for watermark insertion. The relationship between various pairs of original and watermarked datasets is then analyzed to determine the susceptibility of key-based attacks. The strength of the attack models is measured by multivariate and discriminant analysis methods like Wilks' lambda, Pillai's trace test, and Box's M test. The proposed scheme can be employed for data warehouse or cloud database which usually contains a large database repository across an entire enterprise. In a single database environment, a relational database can be considered as a collection of multiple related tables.

The rest of the paper is organized as follows: Section 2 elaborates the proposed framework along with the model. Experimental results along with its analysis are shown in Section 3 and we conclude our findings in Section 4.

2. PROPOSED FRAMEWORK

Fig. 1 shows the proposed framework for key-based attack analysis of a watermarking system. The proposed framework encompasses three segments; watermark encoding and decoding, statistical analysis, and interpretation of results. In the first segments, an original dataset D_0 is transformed into a watermarked dataset $D\omega$ by AHK watermarking algorithm [1] and recognition of inserted watermarks is achieved by watermark decoder. After insertion process is completed then attacker channel can be considered as a way where the watermarked data is stored or communicated, for example, this channel may be a public network where the watermarked data is being communicated to some destination. The embedded watermarks can be achieved by watermarking decoding algorithm [1]. The detection process of watermark is a blind system and it may not require original data nor the watermark. So, the watermarking decoding process uses the watermark dataset $D\omega$using the same secret keyS_k. In statistical analysis stage, various statistical tests like Wilks's Lambda test, Pillai's Trace test, Box's M test, etc., are used to determine the significance of SKMDs and MKsSD against key-based malicious attacks. The last segment interprets the analysis results on the basis of conclusions that are derived. The SKMDs and MKsSD watermarking models are presented as follows:

2.1. Single Key Multiple Datasets (SKMDs)

The SKMDs watermarking model is shown in Fig. 2. In SKMDs model, the original multiple datasets $D_0, D_1, D_2, \ldots D_{n-1}$ are watermarked by using single secret key $S_k$to generatewatermarked variants $V_0, V_1, V_2, \ldots V_{n-1}$where $V_0 = D_\omega^0,\ V_1 = D_\omega^1, V_2 = D_\omega^2, \ldots\ldots\ldots, V_{n-1} = D_\omega^{n-1}$.In general, the watermarked datasets by SKMDs model are generated by using the following relations: $D_\omega^0 = \omega(D_0, S_k), D_\omega^1 = \omega(D_1, S_k)\ldots., D_\omega^{n-1} = \omega(D_{n-1}, S_k)$.

In general, for SKMDs, the ambiguity for marks detection drops off and tends to zero for the attacker as the number of conspiring user increases. When an attacker obtains the contents of watermarked dataset, eventually, he 'she becomes aware of the marked bit positions that are used in the original dataset. In using a Single key for watermarking multiple datasets, if a secret key is compromised, then the security of all the watermarked datasets also becomes vulnerable.

2.2. Multiple Keys Single Dataset (MKsSD)

The MKsSD watermarking model is shown in Fig. 3. In contrary to SKMDs model, the MKsSD model watermarks a single dataset$D_0$by using n

Table 1. Original forest dataset sample (OFDS).

$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	$X9$	$X10$
2606	45	7	270	5	633	226	225	138	6256
2507	22	9	120	14	732	215	221	143	5534
2962	88	16	190	23	6095	242	212	95	3811
2864	118	18	201	74	4567	248	221	93	4849
2827	160	28	134	65	3948	235	233	108	5474
2840	153	26	134	42	4613	241	231	102	4833

Table 2. Watermarked forest dataset sample (WFDS).

$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	$X9$	$X10$
2606	45	7	270	5	633	**234**	225	138	6256
2507	22	9	120	14	732	215	221	143	5534
2962	88	**24**	190	23	6095	242	212	95	3811
2864	118	18	**193**	74	4567	248	221	93	4849
2827	160	28	134	**73**	3948	235	233	108	5474
2840	153	26	134	42	4613	241	231	102	4833

Table 3. Wilks' lambda test results in case of SKMDs with $\xi = 4$.

Fraction of watermark (%)	***P*-value**				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
29	0.046	0.076	0.133	0.188	0.277
30	0.000	0.041	0.017	0.024	0.103
37	0.000	0.012	0.005	0.013	0.044
40	0.000	0.003	0.001	0.003	0.005
29	0.046	0.076	0.133	0.188	0.277

Table 4. Wilks' lambda test results in case of MKsSD with $\xi = 4$.

Fraction of watermark (%)	***P*-value**				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
29	0.342	0.437	0.411	0.256	0.306
30	0.461	0.386	0.385	0.399	0.276
37	0.040	0.043	0.023	0.010	0.024
40	0.010	0.029	0.031	0.004	0.014

Table 5. Wilks' lambda test results in case of SKMDs with $\xi = 6$.

Fraction of watermark (%)	***P*-value**				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
2	0.202	0.283	0.210	0.225	0.487
3	0.003	0.006	0.004	0.002	0.033
5	0.000	0.000	0.000	0.000	0.000

different secret keys $S_{ki}$for i = 0, 1, 2 ….n-1, i.e. $sk_0, sk_1 \ldots sk_{n-1}$. In this model, a secret keyS_{ki} is kept variant and the dataset $D_0$is kept constant to generate watermarked variants $V_0, V_1, V_2, \ldots V_{n-1}$ where $V_0 = D^0_{\omega 0}$, $V_1 = D^0_{\omega 1}, V_2 = D^0_{\omega 2}, \ldots D^0_{\omega,n-1}$. In general, the watermarked datasets by MKsSD model are generated by using the following relations: Also, $D^0_\omega = \omega(D_0, S_{k0})$, $D^1_\omega = \omega(D_0, S_{k1})$, $D^{n-1}_\omega = \omega\left(D_0, S_{k,n-1}\right)$.

In the MKsSD context, it is not viable to correctly guess all watermarked bit positions in single or multiple datasets. In using multiple keys for watermarking single or multiple datasets, if a secret key is compromised, only the security of specific dataset would be compromised, while the other watermarked datasets would still remain secure.

The original datasets D_{0i} are selected from the original database D_0 repository. These original datasets D_{0i} are the input for watermark insertion algorithm which is adopted from AHK [1].This algorithm marks only numeric attributes and assumes that the marked attributes are such that minor changes in some of their values are acceptable and does not disturb the usability and integrity of the dataset. All of the numeric attributes of a dataset need not to be marked and the dataset owner is responsible for deciding which attributes are appropriate for marking. Thus, the tuples, attributes in a tuple, bit positions in an attribute, and specific watermark bits are all determined by AHK algorithm [1] and using secret key S_k generated pseudo randomly. The secret key S_k information is only known to the owner of the dataset. This transformation of bit positions makes the watermark. The watermark can be detected with high probability once the secret key is shared. After watermark insertion in each variant $V_0, V_1, V_2, \ldots V_{n-1}$ of datasets are shown as watermarked datasets D_ω in Fig. 2 and 3.

3. EXPERIMENTAL RESULTS

This section elaborates the statistical techniques used in analysis for key-based attacks of a watermarking system. Two types of datasets, original and watermarked datasets, are compared regarding means, variances, and covariances/correlations. The techniques used for key-based attacks analysis includes Wilks' lambda Test (comparison mean), Pillai's Trace Test (comparison variance), and Box's M Test (comparison covariance/correlation). These techniques are used in experiments with Forest cover datasets (10 attributes and 1, 16,000 tuples) available at UCI (University of California, Irvine) machine learning repository [27]. The experiments are performed on an Intel (R) Core (TM) i7, CPU 1.3 GHz with 6GB RAM and 500GB hard drive. The software packages that are used to tabulate results are SPSS version 21, MATLAB, MS Excel as back- end tool and JAVA as front- end tool.

Table 1 and 2 show original and watermarked dataset respectively which is generated by using AHK watermarking algorithm [1], as discussed earlier.

3.1. Hypothesis Testing

A statistical hypothesis testing ensures that the result obtained from a population sample does not occur by chance and then demonstrate the result for the entire population if alternative hypothesis is true. Thus, hypothesis testing is carried out to test the significance of the results.

H_0:The results are statistically significant which shows that the original and watermarked datasets are same.

H_a:The results are statistically significant which shows that the original and watermarked datasets are different.

The null hypothesis H_0 and alternate hypothesis H_a is selected for testing are as follows:

3.1.1. Wilks' Lambda Test

Wilks' lambda is a statistical test which is used in MANOVA (multivariate analysis of variance). In our experiments, Wilks' lambda compares the difference between the mean of designated groups of datasets i.e. original and watermarked dataset. The mean difference between original and watermarked dataset is judged by the P-value for Wilks' lambda statistic. If aP- value is less than 0.05, the result is statistically significant and it shows mean scores across original and watermarked dataset are different. The Wilk's Lambda is evaluated by using the following relation as discussed by Field [26].

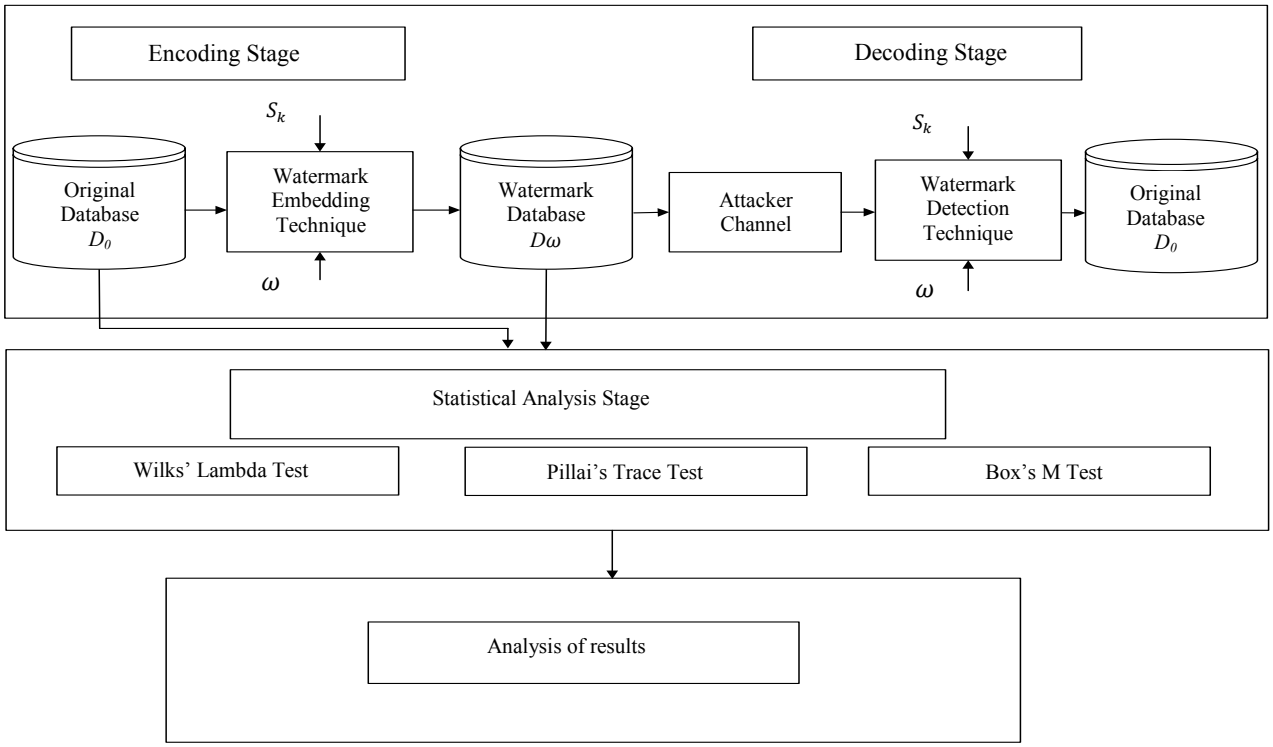

Fig. 1. Proposed framework for key-based attacks analysis of original and watermarking system.

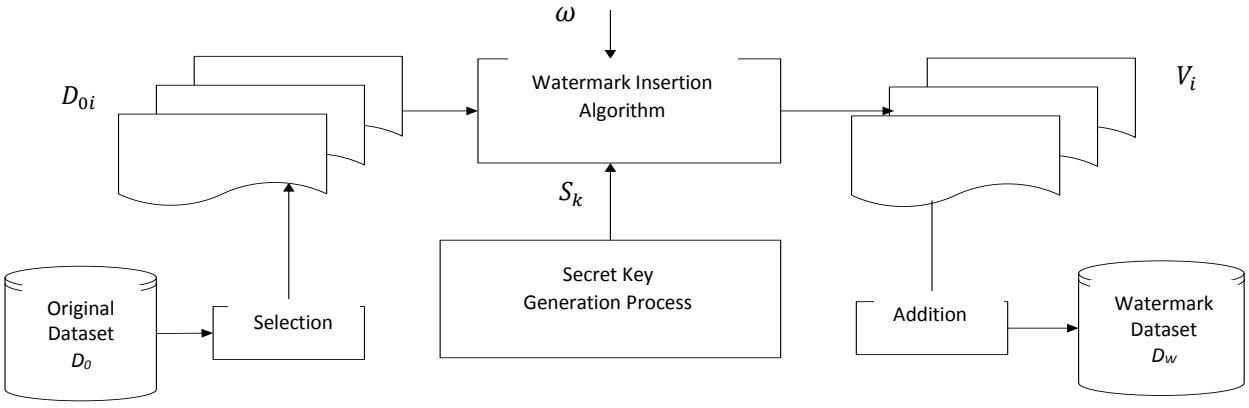

Fig. 2. Single key multiple datasets watermarking model (SKMDs).

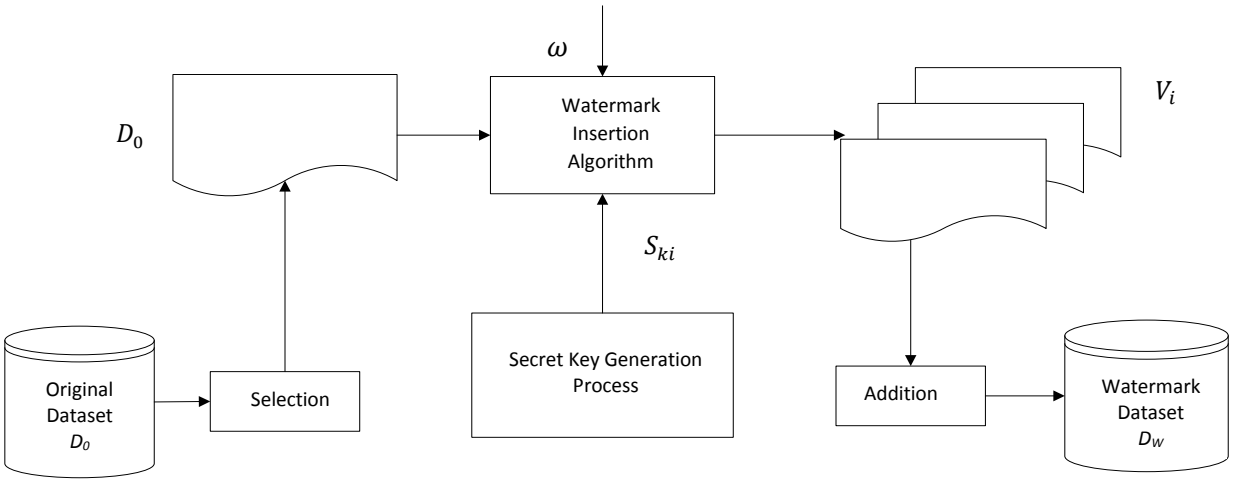

Fig. 3. Multiple keys single dataset watermarking model (MKsSD).

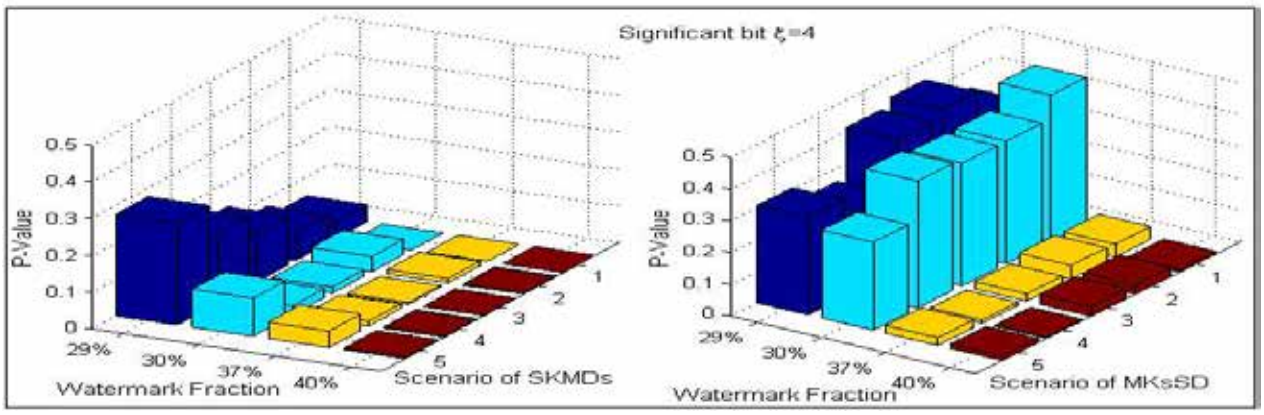

Fig. 4. Variation in P-values (Wilks's Lambda Test*)*.

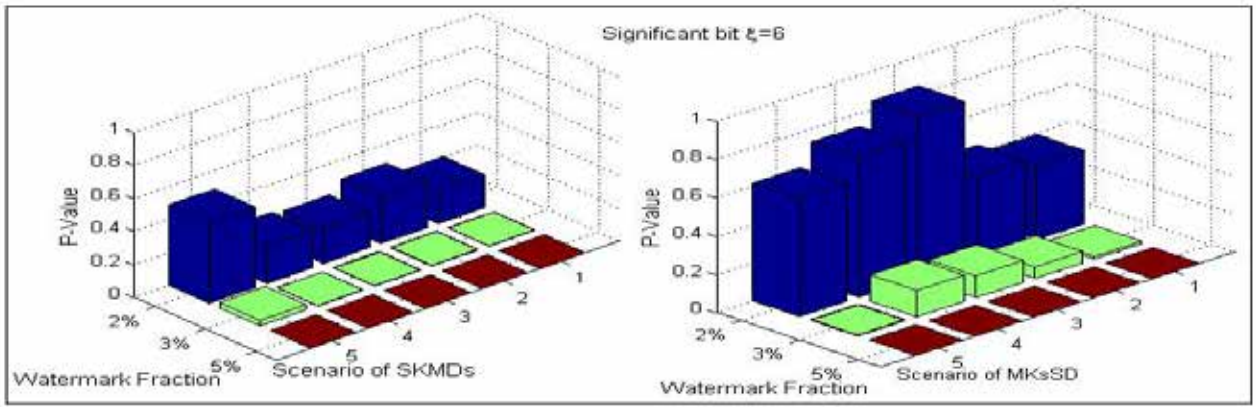

Fig. 5. Variation in P-values (Wilks's Lambda Test).

$$\Lambda_{P,h,e} = \frac{|E|}{|E+H|} = \prod_{i=1}^{P} \frac{1}{1+\lambda_i} \dots\dots\dots\dots (1)$$

Wilks's lambda is the ratio of error sum of squares and total sum of squares. So, in the equation, the symbol $\prod$ is similar to sign of summation $\sum$ but here it means multiply instead of addition, λ_i is the eigenvalue.

3.1.1.1. Results analysis

The results of Wilks' lambda test are shown in Table 3–8 and Fig. 4 - 6. In case of SKMDs, the result of the Wilk's lambda test using the LSB (least significant bit) ξ=4 indicates that the group mean differences md do not appear to be significantly different up to 29% of inserted watermarks i.e. P>0.05 in all five different datasets. When the fraction of watermarks $\omega$is increased to 30%, 37% and 40% then the Wilks' lambda indicates that the group mean differences md appear to differ because the significant values are less than 0.05 (P<0.05). When 29% and 30% of watermarks insertion are induced in a case of MKsSD, the P-value is greater than 0.05, showing that the mean differences md between original and watermarked datasets are not significantly different. When the watermark insertion is increased to 37% and 40%, the P-value is decreased which is less than 0.05, showing that the mean differences mdbetween the two groups i.e. original and watermarked groups are significantly different, which proves that the mean scores are significant statistically in case of MKsSD at 37% and 40% watermarks are inserted.

In another set of experiments, the LSB(least significant bit) ξ is increased to 6th bit and 8th bit then up to 2% of watermark insertion, the P-value is greater than 0.05 i.e.$P > 0.05$, indicating that the mean differencesmd are significantly same between original and watermarked datasets in both the cases SKMDs and MKsSD. If watermark insertion is increased to 3% and 5%, then$P$-values are less than 0.05 i.e.$P < 0.05$, indicating significant results of mean differences md among original and watermarked datasets which show that the mean differences md are statistically significant.

From the above experimental results it has been observed that the ξ and fraction of watermark insertion ω increases, the mean differencesmd also increase but the P-values decrease consistently. Thus, the observations can be shown as the following relations.

$$P \propto \frac{1}{\xi} \dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

$$md \propto \omega \dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

Where, md is the value of the mean difference, ω is the fraction of a watermark, ξ is the least significant bits and P is the significant value. The variation in P-value with respect to the fraction of watermarks inserted in a case of SKMDs and MKsSD are shown in Fig.4, 5, 6. In general, it has been observed that high P-value (P<0.05) indicates that the mean differences between original and watermarked datasets are statistically same and also suitable the data usability of a watermarked datasets and does not show any valuable evidence to the unauthorized users. A low significant P-value of Wilks' Lambda test (usually less than 0.05) specifies that there is a significant difference among the two group's i.e. original and watermarked datasets and the watermarked data values are most visible in those attributes which have smaller values.

3.1.2. Pillai's Trace Test

Pillai's Trace test is used to compare the variance between group's i.e. original dataset and watermarked dataset are statistically same from each other or not. The variance differences between original and watermarked dataset is judged by the$P$- value for Pillai's trace statistic. If aP- value is less than 0.05, the result is statistically significant and it shows variance between original and watermarked datasets are different. The Pillai's Trace is evaluated by using the following relation [26].

$$\mathrm{V} = trace[H(H+E)^{-1}] = \sum_{i=1}^{S} \frac{\lambda_i}{1+\lambda_i} \dots\dots\dots (4)$$

Pillai's trace test is the ratio of model sum of squares and total sum of squares. So, in the equation, the symbol λ_i is the given values and S represent the number of variants.

3.1.2.1. Results analysis

The results of the Pillai's trace are shown in Table 9–14and Fig. 7 – 9, indicates that the groups variances between original and watermarked datasets using the LSB (least significant bit)ξ=4

Table 6. Wilks' lambda test results in case of MKsSD with $\xi = 6$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
2	0.387	0.432	0.843	0.731	0.616
3	0.013	0.060	0.114	0.138	0.005
5	0.000	0.000	0.000	0.000	0.000

Table 7. Wilks' lambda test results in case of SKMDs with $\xi = 8$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
2	0.000	0.010	0.007	0.002	0.161
3	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000

Table 8. Wilks' lambda test results in case of MKsSD with $\xi = 8$.

Fraction of watermark (%)	P-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
2	0.138	0.061	0.069	0.106	0.131
3	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000

Table 9. Pillai's trace test results in case of SKMDs with $\xi = 4$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
29	0.046	0.076	0.133	0.188	0.277
30	0.000	0.041	0.017	0.024	0.103
37	0.000	0.012	0.005	0.013	0.044
40	0.000	0.003	0.001	0.003	0.005

Table 10. Pillai's trace test results in case of MKsSD with $\xi = 4$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
29	0.342	0.437	0.411	0.256	0.306
30	0.461	0.386	0.385	0.399	0.276
37	0.040	0.043	0.023	0.010	0.024
40	0.010	0.029	0.031	0.004	0.014

Table 11. Pillai's trace test results in case of SKMDs with $= 6$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
2	0.202	0.283	0.210	0.225	0.487
3	0.003	0.006	0.004	0.002	0.033
5	0.000	0.000	0.000	0.000	0.000

Table 12. Pillai's trace test results in case of MKsSD with $\xi = 6$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
2	0.387	0.432	0.843	0.731	0.616
3	0.013	0.060	0.114	0.138	0.005
5	0.000	0.000	0.000	0.000	0.000

Table 13. Pillai's trace test results in case of SKMDs with $\xi = 8$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
2	0.000	0.010	0.007	0.002	0.161
3	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000

Table 14. Pillai's trace test results in case of MKsSD with $\xi = 8$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
2	0.138	0.061	0.069	0.106	0.131
3	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000

Table 15. Box's M test of equality of covariance matrices results of SKMDs with ξ=4.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
12	0.082	0.499	0.602	0.590	1.000
15	0.000	0.014	0.016	0.000	0.042
17	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000

are not significantly different because *P*-value is greater than 0.05 (P<0.05) in all five different data sets up to 29% watermark insertion. When the fraction of watermarks insertion is increased to 30%, 37% and 40%, in this case, the test shows significant results i.e. P<0.05, indicating that the group variances are significantly different. But in a case of MKsSD, the results of the test indicate that there is no significant difference up to 30% of watermark insertion between the two groups of variance i.e. original and watermark group. When the fraction of watermarks insertion is increased to 37% and 40% then Pillai's Trace test shows the significant results that is *P*-value or significance value of the test is less than 0.05, indicating that the variances for the two groups are significantly different.

In other set of experiment, the LSB (least significant bit) ξ is increased to 6th bit then up to 2% of watermark insertion in both cases i.e. SKMDs and MKsSD, the *P*-value is greater than 0.05 ($P < 0.05$), showing that the variances are significantly same between original and watermarked datasets. When the significant bit ξ i.e. 8th bit is selected and tested the experiments at 2%, 3% and 5% of watermark insertion in case of SKMDs, the *P*-value is less than 0.05 ($P < 0.05$) which proves that the variances are statistically significant. But in a case of MKsSD, the results are statistically same up to 2% of watermark insertion. When 3% and 5% of watermark insertion is selected, the *P*-value decreases and becoming significant.

The P-values for each of the datasets in both cases i.e. SKMDs and MKsSD are shown in Fig.7–9. In these Figures, the high *P*-values indicate that the variances between original and watermarked datasets are significantly same and the data usability of a watermarked datasets is also acceptable. The low *P*-values showing significant results. These significant results indicate that no relationship exists between original and watermarked datasets.

3.1.3. Box's M Test of Equality of Covariance Matrices

Box's M test compares the variance covariance matrices of original and watermarked datasets. The test statistics of Box's M follows F-distribution. The difference between two variance covariance matrices is judged by the *P* -value for Box's M statistic. Usually a P-value less than 0.05 is considered to be significant.

The Box's M test is evaluated by using the following relation [27].

$$M = (N - q)\log_e|S| - \sum_{i=1}^{q}(n_i - 1)\log_e|S_i| \dots\dots (5)$$

In the equation q represents total groups that we actually compare across variables, N represent number of subjects in each sample n_i is the number of subject values. S calculate the estimated pooled within-group covariance andS_i presents the cell covariance matrix. The M value is then transformed into the approximation based on the F-distribution to calculate the significance value.\

3.1.3.1. Results Analysis

The results of Box's M test of the assumption of equality of covariance matrices using P<0.05 as a criterion are shown in Table 15 – 20 and Fig. 10 - 12. So that Box's M (in a case of SKMDs) is not significant at all five different datasets using the LSB (least significant bit) ξ=4. When the percentage of watermarks ω insertion are selected up to 12% i.e. P>0.05 indicating that there are no significant differences between the covariance matrices of original and watermarked datasets. When the fraction of watermarks ω insertion increases to 15%, 17% and 20% then Box's M test shows the significant results that is *P*-value or significance value of the test is less than 0.05, indicating that the covariance matrices for the two groups are significantly different. But in a case of MKsSD, the Box's M test is also not significant (P>.05) up to 15% watermark insertion and shows the significant results when the watermarks insertion is increased to 17% and 20%. So, P<0.05 it suggests that the covariance matrices for the two groups are significantly different. Hence, the result is statistically significant at 17% and 20% of watermarks insertion.

In other set of experiment, the LSB(least significant bit) ξ is increased to 6th bit and 8th bit, the P-value decreases in both cases SKMDs and MKsSD that is 0.000 which is less than 0.05 at 1%-3% watermarking insertion, indicating that there are significant differences between the covariance matrices of original and watermarked datasets.

The experimental results show that when increases the watermarks insertion between

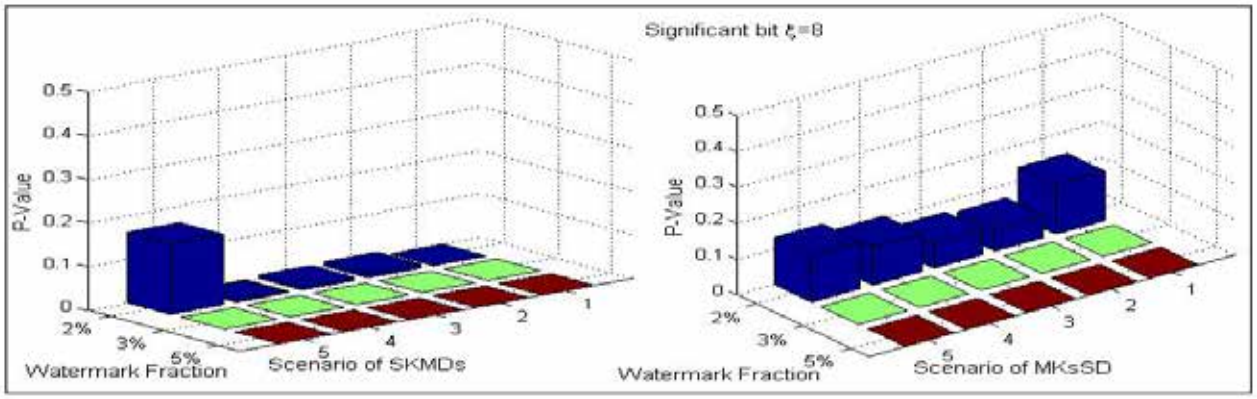

Fig. 6. Variation in P-Values (Wilks's Lambda Test).

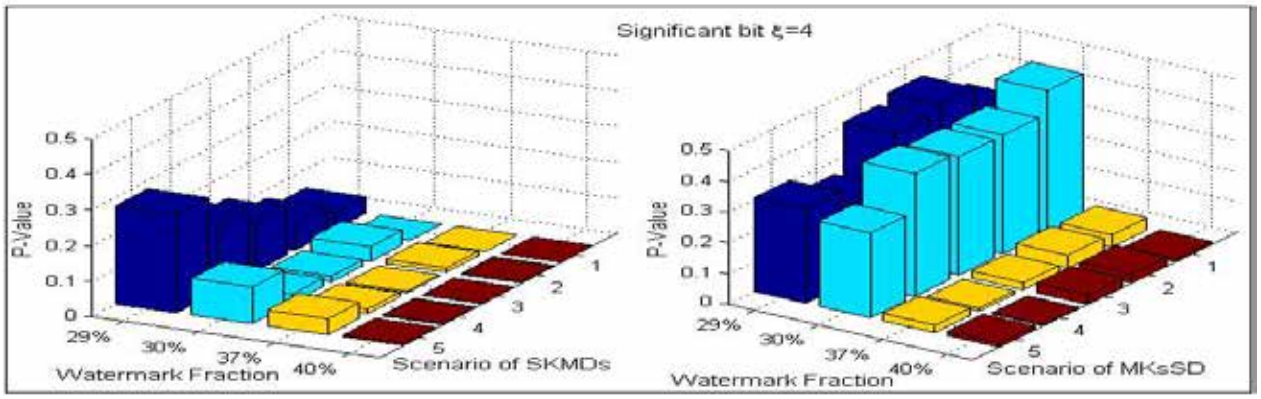

Fig. 7. Variation in P − value (Pillai's Trace Test).

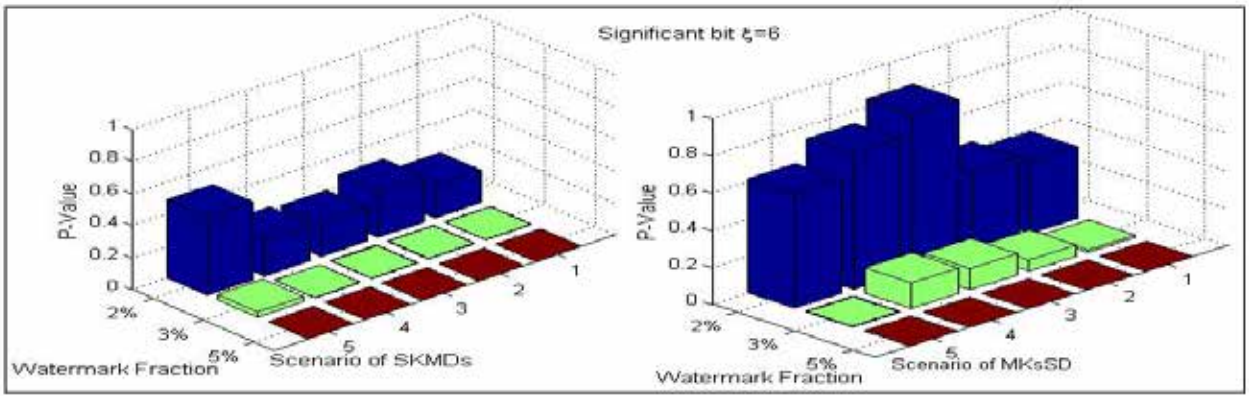

Fig. 8. Variation in P − value (Pillai's Trace Test).

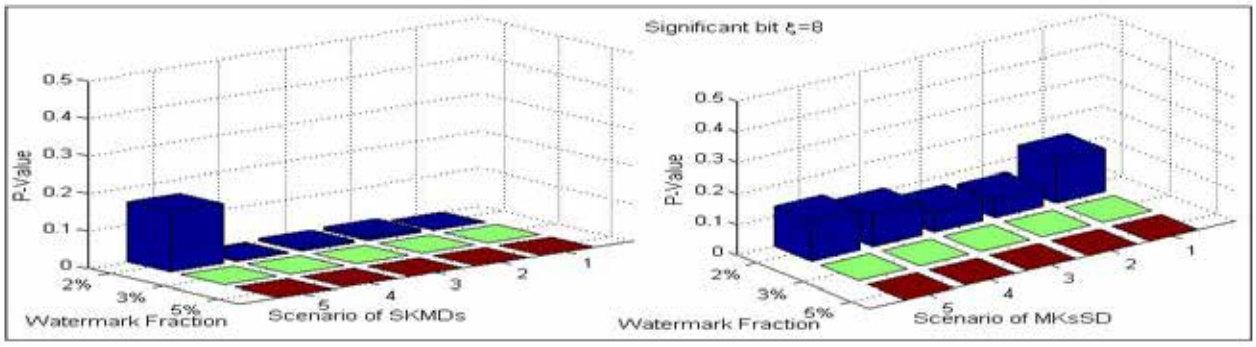

Fig. 9. Variation in P − value (Pillai′s Trace Test).

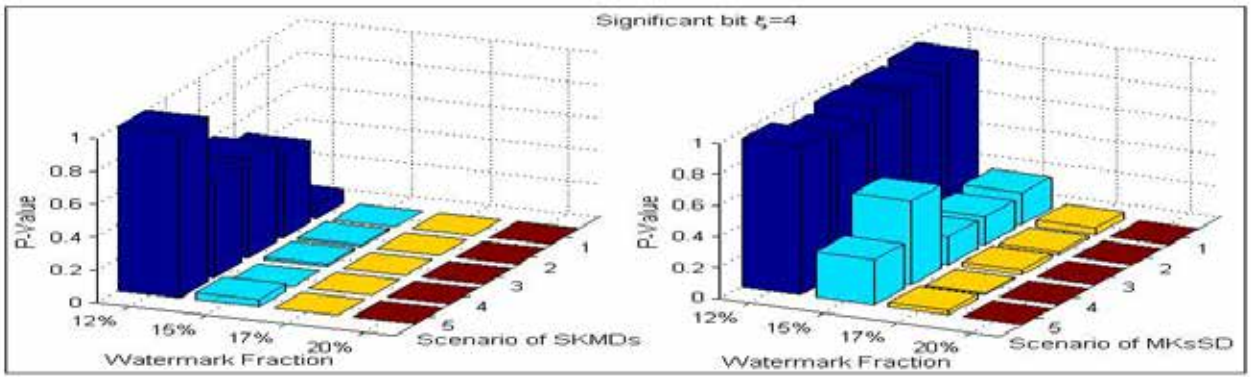

Fig. 10. Variation in P-values (Box's M Test).

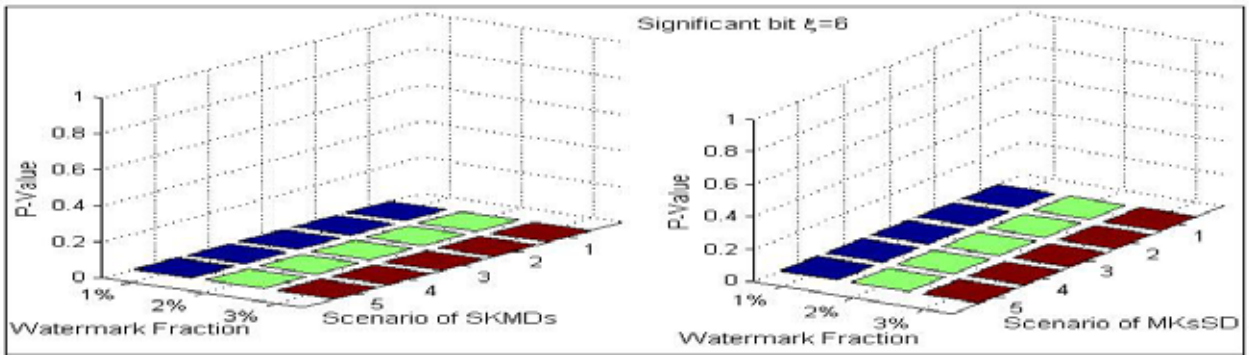

Fig. 11. Variation in P-values (Box′M Test).

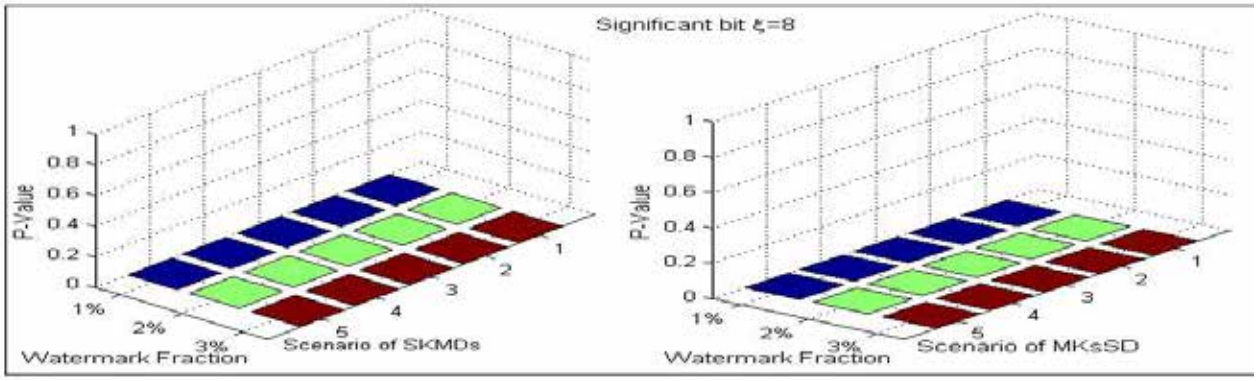

Fig. 12. Variation in P-values (Box′M Test).

Table 16. Box's M test of equality of covariance matrices results of MKsSD with ξ=4.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
12	0.953	0.929	0.935	0.875	0.950
15	0.205	0.180	0.180	0.536	0.289
17	0.043	0.020	0.030	0.010	0.029
20	0.000	0.000	0.000	0.000	0.000

Table 17. Box's M test of equality of covariance matrices results of SKMDs with$\xi = 6$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
1	0.000	0.000	0.000	0.000	0.002
2	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000

Table 18. Box's M test of equality of covariance matrices results of MKsSD with $\xi = 6$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
1	0.000	0.000	0.000	0.000	0.003
2	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000

Table 19. Box's M test of equality of covariance matrices results of SKMDs with $\xi = 8$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
1	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000

Table 20. Box's M test of equality of covariance matrices results of MKsSD with $\xi = 8$.

Fraction of watermark (%)	*P*-value				
	Dataset1	**Dataset2**	**Dataset3**	**Dataset4**	**Dataset5**
1	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000

original and watermarked datasets using the least significant bit (ξ=4) and decreases the P-value as the fraction of watermarks is increased. So, we conclude that high P-value (P>0.05) indicates that the covariance matrices for the two groups are statistically same and also acceptable the data usability of a watermarked datasets and does not show any valuable evidence to the unauthorized users. A low significant value of P for the Box's M test (usually less than 0.05) specifies that there is a significant difference among original and watermark datasets. Note that the larger the least significant bit of ξ, the larger the visibility of watermarked data and the results may provide useful information to the malicious users. The following relation has been observed from the results:

$$P \propto \frac{1}{\omega} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

$$P \propto \frac{1}{\xi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

Where P is the significant value and ω are the fraction of watermarks and ξ is the least significant bits. The significant value decreases as the fraction of watermarks and ξ is increases.

Fig. 10 -12 show variation in the P-value for original and watermarked datasets in a case of SKMDs and MKsSD. In these Figures, when five different datasets up to 12% of watermarks are selected, the P-value is close to 1 in both cases i.e. SKMDs and MKsSD which shows that the covariance matrices are statistically same. When the datasets with 15%, 17% and 20% of watermarks are selected, the P-value decreases, which indicates that the covariance matrices for the two groups are significantly different. When the least significant bit (ξ) is increased to a 6^{th} bit and 8^{th} bit, the$P$-value is decreased as the fraction of the watermark is 1%, 2% and 3%, indicating that covariance matrices for the two groups, i.e., original and watermark groups are significantly different. It means that the watermark data error is visible in those attributes which have smaller values.

4. CONCLUSIONS

In this paper, we have analyzed two attack models SKMDs and MKsSD for susceptibility of key-based attacks in a watermarking system. These attack models make variants of single and multiple datasets by the usage of single and multiple keys for watermark insertion. The empirical analysis of these attack models is measured by multivariate and discriminant analysis methods like Wilks' lambda, Pillai's trace test and Box's M test. We observe that MKsSD model has high significance as compared to SKMDs which shows that MKsSD are more secure in a watermarking system, whereas, SKMDs model are more susceptible to malicious key-based attacks. Also, by varying ξ LSB's and fraction of watermarks ω, the MKsSD still shows significance as compare to SKMDs and does not provide any valuable evidence to the unauthorized users. In future, we intend to analyze key-based attacks by using other statistical techniques like clustering such as K-mean, two steps, density based and Hierarchical etc. Also, besides relational databases, the proposed framework can be analyzed on other data domains, such as text, images, audio and video, etc.

5. REFERENCES

1. Agrawal, R., P.J. Haas, & J. Kiernan. Watermarking relational data: framework, algorithms and analysis. *The VLDB Journal* 12(2): 157-169 (2003).
2. Lafaye, J. An analysis of database watermarking security. In: *Third IEEE International Symposium on Information Assurance and Security (IAS 2007)*, Manchester, United Kingdom, 29 August, 2007, p. 462-467 (2007).
3. Khanduja, V, P. Verma, & S. Chakraverty. Watermarking relational databases using bacterial foraging algorithm. *Multimedia Tools and Applications* 74(3): 813-839 (2015).
4. Khan, A., & Husain. A fragile zero watermarking scheme to detect and characterize malicious modifications in database relations. *The Scientific World Journal https://www.hindawi.com/journal/ tswj/2013/796726/* (2013).
5. Melkundi, S., & C. Chandankhede. A robust technique for relational database watermarking and verification. In: *Communication, Information & Computing Technology (ICCICT),* International Conference IEEE, Mumbai, India, 15 January - 17 January, 2015, p. 1-7 (2015).
6. Camara, L., J. Li, R.Li, F. Kagorora, & D. Hanyurwimfura. Block-based scheme for database integrity verification. *International Journal of Security and Its Applications* 8(6): 25-40 (2014).
7. Sruthi, N., A.V. Sheetal, & V. Elamaran. Spatial and spectral digital watermarking with robustness evaluation. In: *Computation of Power, Energy, Information and Communication (ICCPEIC)*,

International Conference IEEE, Chennai, India, 16 April – 17 April, 2014, p. 500-505 (2014).
8. Rohith, S., K.H. Bhat, & B. K. Sujatha. A secure and robust digital image watermarking scheme using repetition codes for copyright protection. In: *Advances in Electronics, Computers and Communications (ICAECC)*. International Conference IEEE, Bangalore, India, 10 October – 11 October, 2014, p. 1-8 (2014).
9. Dhar, P. K, & I. Echizen. Robust FFT Based Watermarking Scheme for Copyright Protection of Digital Audio Data. In: *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Seventh International ConferenceIEEE, Shanghai Songjiang, China, 26 July – 28 July, 2011, p. 181-184 (2011).
10. Gupta, G., J. Pieprzyk, & L. Hamey. Bucket attack on numeric set watermarking model and safeguards. *Information Security Technical Report* 16.2: 59-66 (2011).
11. Sion, R., M. Atallah, & S. Prabhakar. On watermarking numeric sets. In: *Digital Watermarking*, Springer, Berlin, p. 130-146 (2003).
12. Pournaghshband, V. A new watermarking approach for relational data. In: *Proceedings of the 46th Annual Southeast Regional Conference (ACM)*, United States, 28 March, 2008, p. 127-131 (2008).
13. Zhang, L., W. Gao, N. Jiang, L. Zhang, & Y. Zhang. Relational databases watermarking for textual and numerical data. In: *Mechatronic Science, Electric Engineering and Computer (MEC)*, International Conference IEEE, Jilin, China, 19 August - 22 August, 2011, p. 1633-1636 (2011).
14. Khadim, U., A. Khan, B. Ahmad & A. Khan. Information Hiding in Text to Improve Performance for Word Document. *International Journal of Technology and Research* 3(3): 50 (2015).
15. Maheshwari, J. Prasad, M. Kumar, G. Mathur, R.P. Yadav, & R.K. Kakerda. Robust digital image watermarking using DCT based pyramid transform via image compression. In: *Communications and Signal Processing (ICCSP)*, International Conference IEEE, Melmaruvathur, India, 02 April – 04 April, 2015, p. 1059-1063 (2015).
16. Munesh C., S. Pandey,& R. Chaudhary. Digital watermarking technique for protecting digital images. In: *Computer Science and Information Technology (ICCSIT)*, Third IEEE International Conference on, Beijing, China, 09 July - 11 July, 2010, vol. 7, p. 226-233 (2010).
17. Kavipriya, R., & S. Maheswari. Statistical quantity based reversible watermarking for copyright protection of digital images. In: *Green Computing Communication and Electrical Engineering (ICGCCEE)*. International Conference IEEE, Coimbatore, India, 06-08 March, 2014, p. 1-6 (2014).
18. Hu, H.Tsu, & L.Y.Hsu. Robust, transparent and high-capacity audio watermarking in DCT domain. *Signal Processing* 109: 226-235 (2015).
19. Patra, J. C., A. Karthik, & C. Bornand. A novel CRT-based watermarking technique for authentication of multimedia contents. *Digital Signal Processing* 20.2: 442-453 (2010).
20. Iwakiri, M. & T. M. Thanh. Fragile watermarking based on incomplete cryptography for copyright protection. *Applied Informatics* 2(1): 1-20 (2015).
21. Hoang, T., D. Tran, & D. Sharma. Remote multimodal biometric authentication using bit priority-based fragile watermarking. In: *Pattern Recognition. 19th International Conference, IEEE*, Tampa, FL, USA, 08 Dec - 11 Dec, 2008, p. 1-4 (2008).
22. Wang, J.-T., W.-H. Yang, P.C. Wang, &.-T. Chang. A novel chaos sequence based 3D Fragile Watermarking Scheme. In: *Computer, Consumer and Control (IS3C),* International Symposium, IEEE, Taichung, Taiwan, 10 Jun - 12 Jun, 2014, p. 745-748 (2014).
23. Qian, Q., H.X. Wang, Y. Hu, L.N. Zhou, & J.F. Li. A dual fragile watermarking scheme for speech authentication. *Multimedia Tools and Applications* 75: 1-20 (2015).
24. Fu, Y., T. Ye, Z. Qu, X. Niu, & Y. Yang. A Novel Relational Database Watermarking Algorithm for Joint Ownership. In: *Intelligent Information Hiding and Multimedia Signal Processing.* International Conference, IEEE, Harbin, China, 15August -17 August, 2008, p. 985-988 (2008).
25. Singh, A., M. K. Dutta, C. M. Travieso, & K. M. Soni. Digital right management control for joint ownership of digital images using biometric features. In: *Signal Processing and Integrated Networks (SPIN*), International Conference. IEEE, Noida, India, 20 February - 21 February, 2014, p. 164-167 (2014).
26. Field, A. *Discovering statistics Using SPSS*. Sage Publications, Los Angeles (2009).
27. https://archive.ics.uci.edu/ml/datasets/Covertype.