**Imperceptibility Analysis of Watermarked Database**

Hameed Ullah & Aihab Khan

Department of Computer Science

Iqra University

Islamabad, Pakistan

hameedullahpk@gmail.com

Aihab@iqraisb.edu.pk

**Abstract**

**Watermarking is recognized system for hiding copyrights identification inside documents, thus preventing illegal distributions. The key feature of database watermarking is to prove the ownership of data that express who is the owner of data. In the case of electronic data it is important, as data sets are generally tempered and redistributed without appropriate citation or permission of originating data set. The effect on the quality of the data sets, defined as imperceptibility is evaluated through statistical tests. Exceeding the error threshold means that the watermarked data is less of use or less competitive. In this research, we present a non-parametric statistical method for imperceptibility analysis of watermarked datasets. Experiments performed on watermarked data sets obtained by two watermarking techniques. Results showed no significant difference occurred after watermark between original and watermarked database.**

Keywords-Watermarking; Imperceptiblity; Kruskal Wallis Tes;

**I. Introduction**

 Presently widely deployed database systems, database piracy have turned into a serious concern in database applications. When compared with the expenses of generating and managing a database, it is a good deal for a pirate to replicate and redistribute it without its owner’s information or approval. Evidence of ownership is generally attainable by representing that the specific piece of data demonstrates a definite rare property. Digital watermarking gives obvious way out in setting up owner of data. By inserting the watermark as identity of owner, the confirmation of ownership can be determined as if there should arise an occurrence of any dispute, the watermarks can be separated from the database and goes about as a proof[[1](#_ENREF_1)]. After inserting watermark, watermark and data are indivisible. The watermarks must not have a remarkable effect on the usability of the data and they ought to be set in a manner that they cannot be erased by anyone without making the data useless[[2](#_ENREF_2)]. Generally digital watermarking methods applied to copyright protection and integrity of authenticated data[[3](#_ENREF_3)].

In digital media, for example, video, image, the method which inserts an advanced watermark to demonstrate copyright has been developed for a long time. [[4](#_ENREF_4)] proposed an idea to utilize a digital watermark in a database keeping in mind the end goal to secure a database of map information and afterward numerous researchers began to look in this domain. At last Agarwal and Kiernan proposed the 1st implementation approach in 2002[[5](#_ENREF_5)]. In the relational database, they computed one LSB of one numeric attribute of selected tuples, and this the place they mean to insert the watermark. The technique marks just numeric attributes and expects that the marked attributes with small changes in some of attributes are adequate and unrecognized. Majority of numeric attributes required not be marked. The study for watermarking relational database can be divided into two categories: distortion based watermarking and distortion free watermarking [[5](#_ENREF_5)]. Distortion based watermarking schemes directly insert watermark into some data. This will modify data and these modifications serve as watermark. However the modifications must be acceptable otherwise it will make data useless. The main idea of distortion free watermarking schemes is to 1st segment the data into some segments, then utilizing these segments to generate a watermark. Since this type of watermarking do not insert watermark into database, so it does not cause data distortion. [[6](#_ENREF_6)] proposed a reversible watermarking scheme motivated from Agarwal and Kiernan algorithm. Their scheme confirms the rights ownership of database owner and recovers the original database after watermark is recognized and authenticated.

**II. Background**

Database watermarking embedding can be classified in two categories: visible or invisible[[7](#_ENREF_7)]. The visible watermark is detectable and is much the same as noise. Noise deletion process can separate the visible watermark. Generally proposed methods are undetectable to avoid the risk of cracking. Alternatively originality of watermarked data is also essential. The watermarked database will draw attention of attackers or lose its worth if watermark inserting method seriously influences the quality of watermarked data. By this the quality between original data and watermarked must not be genuinely degraded. This property is called Imperceptibility[[8](#_ENREF_8)]. Imperceptibility implies difficult to separate between original data and watermarked data by human visual system[[9](#_ENREF_9)]. By raising the inserted watermarking data capacity upgraded the watermarking scheme robustness which may influence the imperceptibility of watermarked data.

A few quality measures have been proposed in the past to predict the quality of database [[10](#_ENREF_10)] [[6](#_ENREF_6)] [[11](#_ENREF_11)]. In this paper, Kruskal-Wallis test is used to measure the association between original and watermarked datasets. The watermarked data sets are acquired from techniques used by [[10](#_ENREF_10)] & [[6](#_ENREF_6)].

**III. Kruskal-Wallis test**

The Kruskal-Wallis test is a nonparametric statistical test for comparing data sets that are independent or not related [[12](#_ENREF_12)]. When Kruskal-Wallis test leads to significant results then one dataset is not the same as other dataset. Yet test does not recognize where difference occurs.

To compute Kruskal-Wallis test datasets are combined and ranked ordering the values together.

$$KW=\frac{12}{N\left(N+1\right)}\sum\_{k=1}^{k}\frac{R\_{i}^{2}}{n\_{i}}-3\left(N+1\right)$$

*N* is number of values for all datasets, $R\_{i}$$R\_{i}$is the sum of ranks for a particular dataset and $n\_{i}$$n\_{i}$ is the number of values from the corresponding rank sum. When $KW$ $KW$is computed it is compared to a table of critical values to examine datasets for significant differences. When $KW$ is not significant then no difference exists between any of datasets while when $KW$ $KW$ is significant then there exists difference between two datasets.

TABLE I

Kruskall-Wallis Test for [[10](#_ENREF_10)]Algorithm

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| $$γ$$ | **10000** | **1000** | **100** | **10** | **1** |
| **Attribute** |  |  |  |  |  |
| **Timbre average1** | 1 | 1 | 1 | 1 | .998 |
| **Timbre average2** | 1 | 1 | 1 | 1 | .998 |
| **Timbre average3** | 1 | 1 | 1 | 1 | .997 |
| **Timbre average4** | 1 | 1 | 1 | 1 | .997 |

TABLE II

Kruskal-Wallis Test for [[6](#_ENREF_6)]Algorithm

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| $$γ$$ | **96** | **48** | **24** | **12** | **6** |
| **Attribute** |  |  |  |  |  |
| **Timbre average1** | .991 | .989 | .966 | .675 | .541 |
| **Timbre average2** | .993 | .973 | .991 | .615 | .428 |
| **Timbre average3** | .993 | .970 | .928 | .664 | .596 |
| **Timbre average4** | .994 | .957 | .979 | .695 | .631 |

We next estimate experimentally the effect of watermarking on the medians of all watermarked attributes. Experiments carried out for both algorithms using gap parameter *γ* =1, 10*,* 100*,* 1000*,* 10000 and γ = 6*,* 12*,* 24, 48, 96. We noticed that the change in the median was very little for all attributes in all cases. Kruskal-Wallis is close to 1 for most of the attributes. Significance value less than 0.05 for kruskal wallis test exibits that a significant change exists between the medians of original and watermarked database. High median difference exibits a decline in the imperceptibility among the original and watermarked datasets.

As indicated by the Kruskal-Wallis test attributes, following resutls obtained on the basis of experiments presented in table 1 and 2. In table 1, when watermark with gap parameter *γ* = 1000, 1000, 100 and 10 is embedded, the $KW$ is 1 for all attributes. The median difference is very small and sig value is greater than, i.e., 0.05, showing that the medians of original and watermarked datasets are statistically same. When watermark inserted with *γ* = 1, $KW$ is close to 1 for all attributes and sig value is still greater than .05 which shows no significant change occurred after watermark. Table 2, when watermark with *γ* = 96, 48 and 24 is embedded ; $KW$ is close to 1 for all attributes. the median difference is very small and sig value is greater than 0.05, showing that the difference between the medians of both datasets is not significant. When watermark embeded with *γ* = 12 and 6, $KW$ decreased gradually for all attributes but sig value is still greater than .05 shows that both data sets are statistically same. Results specify that both datasets have same medians as described by the test. It is clear from the tables that the difference between the original and watermarked database is imperceptible, showing more resemblance in the data sets. Original and watermarked data sets look statistically identical and one cannot discriminate at a quick look between both datasets.

**IV. Conclusion**

In this article Kruskal-Wallis test has been used to determine the imperceptibility of original and watermarked dataset. This test can be useful to any distortion based watermarking method. Both watermarking schemes were found efficient because imperceptibility of watermarked data sets is not decreased with the increase in watermark. We finish up with original and watermarked data sets are indistinguishable as assessed by Kruskal-Wallis test so the watermarked datasets have huge data integrity and hence data usability is also high. Both data sets look reasonably identical and it cannot be useful for attacker. In addition, both watermarking schemes are imperceptible as determined by statistical test.

**References**

[1] V. Khanduja, O. P. Verma, and S. Chakraverty, "Watermarking relational databases using bacterial foraging algorithm," *Multimedia Tools and Applications,* vol. 74, pp. 813-839, 2015.

[2] R. Agrawal and J. Kiernan, "Watermarking relational databases," in *Proceedings of the 28th international conference on Very Large Data Bases*, 2002, pp. 155-166.

[3] M.-H. Tsai, H.-Y. Tseng, and C.-Y. Lai, "A Database Watermarking Technique for Temper Detection," in *JCIS*, 2006.

[4] S. Khanna and F. Zane, "Watermarking maps: hiding information in structured data," in *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, 2000, pp. 596-605.

[5] M.-R. Xie, C.-C. Wu, J.-J. Shen, and M.-S. Hwang, "A Survey of Data Distortion Watermarking Relational Databases," *International Journal of Network Security,* vol. 18, pp. 1022-1033, 2016.

[6] M. E. Farfoura, S.-J. Horng, J.-L. Lai, R.-S. Run, R.-J. Chen, and M. K. Khan, "A blind reversible method for watermarking relational databases based on a time-stamping protocol," *Expert Systems with Applications,* vol. 39, pp. 3185-3196, 2012.

[7] D. N. E. Renault and M. Dhanuskodi, "Trends in Computer Science, Engineering and Information Technology."

[8] J.-B. Feng, I.-C. Lin, C.-S. Tsai, and Y.-P. Chu, "Reversible watermarking: current status and key issues," *IJ Network Security,* vol. 2, pp. 161-170, 2006.

[9] S. Roy and A. K. Pal, "A robust blind hybrid image watermarking scheme in RDWT-DCT domain using Arnold scrambling," *Multimedia Tools and Applications,* pp. 1-40, 2016.

[10] R. Agrawal, P. J. Haas, and J. Kiernan, "Watermarking relational data: framework, algorithms and analysis," *The VLDB journal,* vol. 12, pp. 157-169, 2003.

[11] A. Khan, S. A. Hussain, M. S. H. Khayal, and Quratulain, "Correlation coefficient analysis of original and watermarked dataset for data usability," *Journal of the Chinese Institute of Engineers,* vol. 36, pp. 411-421, 2013.

[12] G. Corder and D. Foreman, "Nonparametric statistics for non-statisticians: a step-by-step approach. 2009," ed: Wiley.