ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

Security framework for data publishing using Distance Based Mining

Sumathi.B, MathiVanan.M

Student of M.E, Dept. of Computer and Communication Engineering, MNSK College of Engineering, Pudukkottai, Tamil Nadu, India [1]

Assistant Professor, MNSK College of Engineering, Pudukkottai, Tamil Nadu, India [2]

ABSTRACT: This work presents a right-protection mechanism that can provide detectable evidence for the legal ownership of a shared dataset, without compromising its usability under a wide range of machine learning, mining, and search operations. We provide mechanisms for establishing the ownership of a dataset consisting of multiple objects. The algorithms also preserve important properties of the dataset, which are important for mining operations, and so guarantee both right protection and utility preservation. The watermarking methodology preserves important distance relationships, such as: the Nearest Neighbours (NN) of each object and the Minimum Spanning Tree (MST) of the original dataset. This leads to preservation of any mining operation that depends on the ordering of distances between objects, such as NN-search and classification, as well as many visualization techniques.

KEYWORDS: Nearest neighbours (NN), minimum spanning tree (MST), restricted isometric property (RIP)

I. INTRODUCTION

Data owners need to maintain the principal rights over the datasets that they share, which in many cases have been obtained after expensive and laborious procedures. To right protect we use watermarking. Watermarking allows the user to hide innocuous pieces of information inside the data. The value of watermarking becomes increasingly important because of the proliferation of digital content, and because of the ease of data sharing particularly through data clouds.

It is essential to discover the maximum watermark intensity for right protection. This provides assurances of better detectability and hence better security for the right protection scheme. Our goal is to discover how to right-protect a dataset, but at the same time guarantee preservation of the outcome of important distance-based mining operations. We provide two variants: one that preserves Nearest-Neighbours (NN) and another that preserves the Minimum Spanning Tree (MST). Therefore, the output of any algorithm based on these two properties will be preserved after right protection.

II. RELATED WORK

Traditional multimedia watermarking techniques considered only a single object and did not analyze distortions in the object relationships when dealing with watermarking multiple objects. Watermarking essentially adds noise to a given dataset, and so it may distort the original object distances. Spread spectrum approach is normally implemented with the help of Minimum spanning tree and nearest neighbor algorithms, which embeds the watermark across multiple frequencies of each object and across multiple objects of the dataset. As such, it renders the removal of the watermark particularly difficult without substantially compromising the data utility. The robustness of the watermark embedding depends on the choice of coefficients. We embed the watermark in the coefficients that exhibit, on average over the dataset, the largest Fourier magnitudes. This makes the removal of the watermark difficult; in order for it to be masked out, it would mean that the dominant frequencies of the dataset have to be distorted.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

III. PROPOSED ALGORITHM

A. *Design Considerations:*

This paper suggests the fast variants drastically reduce the number of calculations of quadratic coefficients as well as the solutions of quadratic inequalities. It is paramount to optimize for maximal security, detectability of the watermark, and at the same time minimize visual distortion of objects. Therefore, we seek to find the maximum embedding power $p*$, so that the desired properties are maintained. In certain cases, it may be possible to embed the watermark very strongly and still maintain the original distance relations.

B. *Description of the Proposed Algorithm:*

The previous algorithms exhaustively search the design space for discovering the maximum watermark embedding power. Here, we show how to use the restricted isometric property to properly prune the search space without compromising NN and MST preservation post-watermarking.

Fast NN-Preservation

This proposed algorithm prove a sufficient condition for preservation of the Nearest Neighbor of an object x. We show that if the ratio of the Euclidean distance between x and some other object y in the original dataset over the distance is greater than or equal to a threshold depending solely on $p$max, then y does not violate the NN of x after the watermark embedding, regardless of the details of the dataset or the watermark embedding.

Fast MST-Preservation

The pruning attained by the Fast preservation algorithms ranges from two to five orders of magnitude. This results in enormous savings the execution time is sped up by up to 2 orders of magnitude. In addition, we store the original distances of each object x from objects in its neighborhood, so that when computing squared watermarked distances, the constant term of the quadratic need not be recomputed. The above interpretation highlights the reasoning for a potentially drastic reduction in time complexity attained by the *Fast MST-Preservation* algorithm.

IV. PSEUDO CODE

Step 1 Assume a sequence x ∈ Cn with corresponding set of Fourier descriptors X, a watermark W ∈ Rn and power p ∈ [0, 1] which specifies the intensity of the watermark. A multiplicative watermark embedding (W, p) generates a watermarked sequence by replacing the magnitudes of each Fourier descriptor of x with the watermarked magnitude while not altering the phases, specifically.

Step 2 Using the modified magnitudes and the original phases φj, we can revert from the frequency domain back to the space domain and obtain the watermarked sequence using the inverse discrete Fourier transform.

Step 3 The robustness of the watermark embedding depends on the choice of coefficients. We embed the watermark in the coefficients that exhibit, on average over the dataset, the largest Fourier magnitudes.

Step 4 When embedding the watermark, we exclude the first Fourier descriptor (the DC component) $X1$ from consideration,
and leave it intact. The DC component captures the center of mass of object x and is therefore highly susceptible to translational attacks.

Step 5 we embed the watermark in the magnitudesof the Fourier descriptors and leave the phases unchanged; we leave the DC component intact; and we watermark the Fourier descriptors with the largest average magnitudes.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

V. SIMULATION RESULTS

The evaluation of the right-protection scheme, verify that it retains the important distances of the original dataset graph. We then compare the fast algorithms proposed to their exhaustive counterparts. We also assess the resilience of our scheme against a wide range of adversarial data transformations. We test our methods on six datasets from various application areas video-tracking data, handwritten data, and image contour data from anthropology and natural sciences. We treat images of shapes as two-dimensional sequences by extracting the shape perimeter and sequencing adjacent peripheral points.

Running Time (sec) of the NN-Preservation Algorithms

Dataset	NNP	Fast NNP	Speedup
Skulls	0.047	0.005	9.4
Leaves	224.047	0.562	398.6
Fish	10.781	0.047	229.4
Video1	0.047	0.009	5.2
Video2	0.094	0.007	13.4
Hand	1.438	0.012	119.8

Running Time (sec) of the MST-Preservation Algorithms

Dataset	MST-P	Fast MST-P	Speedup
Skulls	0.188	0.016	11.8
Leaves	4937.92	168.156	29.4
Fish	98.937	3.594	27.5
Video1	0.094	0.012	7.8
Video2	0.204	0.021	9.7
Hand	9.563	0.407	23.5

Compared to exhaustive solutions, the fast algorithms (using the restricted isometric property) drastically reduce runtime. As a concrete example: running the *exhaustive* algorithm for NN preservation on 40K objects requires one and a half day. The *fast* version executes in 19 minutes.

VI. CONCLUSION AND FUTURE WORK

The present work presents a *multiplicative* watermarking framework; therefore the algorithm analysis is quite different. More importantly, here we examine fundamental properties of distance distortion due to multiplicative watermarking. The current work represents an extended version in which the watermark embedding has been done so faster than the existing algorithms. If any unauthorized person or hacker tries to change the watermark content it will be immediately notified to the data owner knowledge through short message service.

REFERENCES

1. M. Vlachos, C. Lucchese, D. Rajan, and P. S. Yu, “Ownership protection of shape datasets with geodesic distance preservation,” in
   a. *Proc. 11th Int. Conf. EDBT*, Nantes, France, 2008, pp. 276–286.
2.]J.-P. M. G. Linnartz and M. van Dijk, “Analysis of the sensitivity attack against electronic watermarks in images,” in *Proc. 2nd Int.*
3. *Workshop IH*, Portland, OR, USA, 1998, pp. 258–272.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

4. F. Hartung, J. Su, and B. Girod., "Spread spectrum watermarking: Malicious attacks and counterattacks," in *Proc. SPIE Security Watermarking Multimedia Contents*, vol. 3657, San Jose, CA, USA, 1999.
5. G. Economou, V. Pothos, and A. Ifantis, "Geodesic distance and MST-based image segmentation," in *Proc. 12th EUSIPCO*, Vienna,
6. Austria, 2004, pp. 941–944.
7. R. Agrawal and J. Kiernan, "Watermarking relational databases," in *Proc. 28th Int. Conf. VLDB*, Hong Kong, China, 2002, pp. 155–166.
8. P. Moulin, M. E. Mihcak, and G.-I. Lin, "An informationtheoretic model for image watermarking and data hiding," in *Proc. IEEE Int. Conf. Image Process.*, Vancouver, BC, Canada, 2000, pp. 667–670.
9. R. Sion, M. Atallah, and S. Prabhakar, "Rights protection for relational data," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1509–1525, Dec. 2004.
10. C. Lucchese, M. Vlachos, D. Rajan, and P. S. Yu, "Rights protection of trajectory datasets with nearest-neighbor preservation," *VLDBJ.*, vol. 19, no. 4, pp. 531–556, 2010.

BIOGRAPHY

Sumathi.B is a Lecturer in Computer Science and Engineering Department and also Student of M.E Computer and Communication Engineering at MNSK College of Engineering, Tamilnadu, India. Her research interest includes network security and related algorithms.