STUDY ON COPYRIGHT AUTHENTICATION OF GIS VECTOR DATA BASED ON ZERO-WATERMARKING

Anbo Li *, Bingxian Lin, Ying Chen, Guonian Lü

Key laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing, China - mrlab@126.com

Commission VI, SS

KEY WORDS: GIS, Copyright Authentication, Digital Product, Vector Data, Zero-watermarking, Hash

ABSTRACT:

GIS vector data is closely linked with the development of social economy and national defense, which is of high production cost and high precision, so the copyright protection of GIS vector data means a great deal and the copyright authentication of GIS vector data has become an increasingly complex topic. In recent years, researches about copyright protection of GIS vector data mainly focuses on copyright marking techniques delegated by watermarking, but it seems like the traditional watermarking are incapability. In this paper, a new copyright authentication method is presented. The watermark signal is constructed utilizing the original data's macroscopical topology characteristics and microcosmic topology characteristics, and then it is registered to authority organ for notarization and authentication. This method is applied to topography map, and the experiment results show that this method can avoid data distortion bringing by watermark embedding and can resolve conflict between imperceptibility and robustness of digital watermark.

1. INTRODUCTIOIN

Geospatial data is a kind of strategic information resource and is widely used in economic, social and environmental applications. GIS vector data is closely related to the overall situation of socio-economic development and national security attributed to its high production cost and high precision characteristics. The issue on copyright protection makes significant sense to the construction of GIS digital production copyright protection system, the legislation draft on copyright protection, the construction of national information security, as well as the healthy development of geographic information system, electronic navigation, digital city and other geographic information industries.

In recent years, researches about copyright protection of GIS vector data mainly focuses on copyright marking techniques delegated by watermarking. Copyright marking techniques allows user to embed auxiliary information into digital productions and the information embedded are imperceptible during normal use but could be readable by using some detection software. Copyright marking techniques are complementary approaches of encryption techniques, and it could maintain protection function even after the encrypted content is decrypted. Copyright protection techniques are used for copyright authentication, content secrecy and access control based on copyright markings. The single copyright marking technique is used for copyright authentication, and it hardly reaches the multiple copyright protection aims, but it is the necessary and precondition of copyright protection.

The possibility of using copyright marking technologies for ensuring the security of data is intriguing and therefore put forward the researches on copyright marking technologies. In the last ten years, various copyright marking algorithms have been proposed. However, there are a few copyright marking algorithms for GIS vector data and they can be summed up to transform domain methods, frequency domain methods and space domain methods (Zhong Shangping, 2006). Kitamura I (2001) proposed a watermarking method for the vector map, where the watermark could be embedded in a set of polylines by using discrete Fourier transform (DFT); Solachidis V extracted vertex coordinates from vector data, and then embedded the information to DFT coefficient of these vertex coordinates, but this method might cause data distortion; Ohbuchi R (2002) proposed a digital watermarking algorithm for vector digital maps, and a watermark bit is embedded by displacing an average of coordinates of a set of vertices that lies in a rectangular area created on a map by adaptively subdividing the map; Li Yuanyuan (2004) improved Ohbuchi’s method, and in the algorithm presented, the strength of watermark is adaptively modulated according to the classification of density of vertices in each blocks, and a two-value watermark image is embedded repeatedly by displacing the coordinates of vertices; Kutter proposed the second-generation of watermarking technology, the perceptive character are used for watermark embedding to enhance the robustness. Ohbuchi R (2003) treats vertices in the map as a point set, and imposes connectivity among the points by using Delaunay triangulation, and then computes the mesh-spectral coefficients from the mesh created. Modifications of the coefficients according to the message bits, and inverse transforming the coefficients back into the coordinate domain produces the watermarked map. Wang Xun (2004) presented a double embedded robust watermark scheme for vector digital mapping, as an extension to the MQUAD algorithm. The vector map is subdivided into two layers by the character of objects in the map, and a watermark bit is embedded by two different methods in two layers of the same map. Zhang Haitao (2004) probed into the issue of watermark attack from the aspects of

* Anbo Li.

storage structure and classification rule. Zhu Changqing (2006) proposed an anti-compression watermarking algorithm for vector map. Yang Chengsong (2007) proposed watermarking algorithm for GIS vector data based on wavelet transformation and Ming Lianquan (2007) proposed DFT-based algorithm for digital map.

More than one copyright marking algorithm for GIS vector data have been proposed and some copyright protection systems have been developed. But the current algorithms have some weakness in some aspect. Copyright marking technique can be used for verifying content integrity, preventing forgery and tracking the traitor, but unfortunately it has some shortcomings. Firstly, the copyright marking embedded in the production of GIS vector data is easily to be erased by some operations (coordinate transformation, projection transformation, etc) of GIS software. Moreover, it will cause a certain extent of data distortion, while it needs more watermark information to be embedded for ensuring the robustness of algorithm for copyright protection.

Face to the mass character of GIS vector data and the requirement of high-conformity patterning, it seems like the traditional algorithms of copyright marking are more and more incapability. As a new digital watermark system, zero-watermarking can resolve conflict between imperceptibility and robustness of digital watermark because it does not change the characters of original data, but utilize the characters of original data to construct original watermark information. Also, it is a blind watermark system naturally. Hash function has the ability of converting multi-sizes of messages to fixed size, and it could be used for expressing the visual character of GIS vector data. Therefore, in this paper, these two technologies are combined for the research of zero-watermark generation, embedding, registration and notarization. This solution can effectively verify content integrity and prevent forgery, without distortion of data.

2. PERFORMANCE DEMAND OF HASH FUNCTIOIN FOR ZERO-WATERMARKING OF GIS VECTOR DATA

For effective copyright authentication of GIS vector data, the watermark should be robust to projection transformation, data format transformation, data compression, and other operations. Therefore, this paper focuses on robust zero-watermark. And the hash function should meet the following performance specifications:

(1) Robustness: If data I1 and I2 have similar content, most part of H(I1) should be the same as H(I2).

(2) Collision: If map I1 and I2 have different content, for most part, H(I1) should be different from H(I2).

(3) Security:
a) Unilateralism. Given a output value y, there must be a value x meets H(x)=y, but not vice versa.
b) Anti-collision. Given any value x, there is no other x’ can make H(x)=H(x’).
c) Safety. If the key is unknown, hash value could not be forged or estimated for it is generated under the key. Use different keys, different hash value will obtain.

(4) Complexity: The algorithm of hash function should have low computational complexity.

Besides, some technology characters should be under consideration, such as: locating the tampered area and the length of hash value.

3. ZERO-WATERMARK GENERATION BASED ON DATA TOPOLOGY CHARACTER

Traditional hash algorithms (MD5, SHA-1) are sensitive to information change, and one bit change can lead to highly different hash sequence. GIS vector data might suffer coordinate transformation, projection transformation, data format transformation, data compression, and other operations. Although those operations changed the data information or structure, they do not influence the visual content of the map. Therefore, the hash algorithm should considering the content information of visual domain but not data domain. Considering different render modes lead to different visual domain information, so the hash function is constructed by the topology relationship of data.

A map is a representation of the earth, or part of it and the map includes symbols that represent features as the natural and social economy phenomenon which is generalized and painted by some mathematical rule (Ma Yongli, 1998). For the same geography area, the macroscopical topology relationships of maps produced by different manufactures are same, while the microcosmic topology relationship will be different for using different generalization algorithms.

The greatest threat faced by zero-watermark is data compression. To meet the high precision and topology invariance requirements of GIS vector data, the topology relationship will not be influenced by lossless or lossy compression. Besides, the notarization mechanism of registering to authority organ could support comparability judgement of original registered data and the data suffered data compression attack, and this would be the evidence for copyright authentication.

The algorithm of zero-watermark generation is as follows:

(1) Map subdividing based on macroscopical topology character
Map subdividing is the precondition of adaptive watermark generation. Taking the contour layer in map for example, we extract the peaks of the terrain referred to the contour and create the TIN by these points .Then the map is subdivided by the triangle of the TIN.(Figure 1)

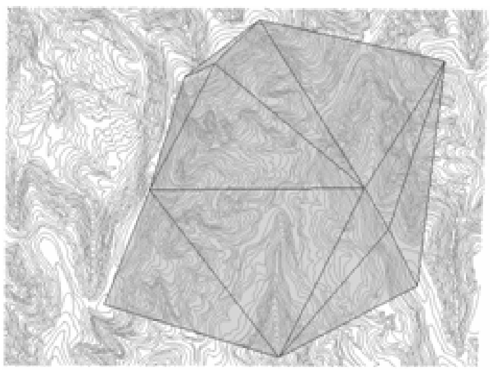


Figure1. Map subdividing based on macroscopical topology character

GIS vector data's macroscopical topology character is determined by natural and social economy phenomena , the topology of the map data which express the same geography region by different mapping companies is in certain similarity . The stability and the geography of the data's macroscopical character determined that the subdividing data will not change through the operation of projection transformation, coordinate transform ,data contract .etc ,thus the arithmetic based on it will be of good robustness.

(2) In the every subdivided triangle, we need to calculate the mathematic morphological factors including geometry factors and topology factors .In detail:

a) When overlapping point layers, figure out its contained points number or the other mathematic morphological factors.

b) When overlapping polyline or polygon layers, the intersected mathematic morphological factors such as polyline number, perimeter ,Euler are able to be the microcosmic character factors (Tang changing, 1990)

c) In addition, when overlapping the terrain layers, the microcosmic character factors can be extracted from what based on the terrain morphological characters and basic terrain character, such as Terrain Complexity, Terrain Curvature and region terrain parameter (Zhou qiming, 2006).

(3) Zero-watermark generation

Hash sequences based on data topology character, have to utilize quantization-disposal and compression-coding to compress, transform to more short and fixed-length bit sequences, and form Zero-watermark of GIS vector data. Detailed method is as follows:

a) Based on quantization encrypt method, change the float data to twain-value data.

b) Based on extent scan arithmetic, sort microcosmic data character string got from each seed block to form hash sequences.

c) Utilizing Huffman compression-coding to compress and obtain final hash value $Hash_{orign}$.

4. ZERO-WATERMARK REGISTRATION AND NOTARIZATION

While other watermarks are embedded into digital product, zero-watermark is registered into IPR depository. The producer of original data takes his data production and zero-watermark generated to authority organ, and if the data and watermark pass the censor, they will be registered and published for notarization efficacy. Attention should be paid to the following problems:

(1) The pirate might make a slight change to a registered data and register the data as his. For avoiding this case, strict scrutiny should be applied to the registering zero-watermark.

(2) For ensuring copyright authentication of data which suffered serious attack, the original data and metadata, blue print of data production and other relational documents should be registered at the same time with zero-watermark.

5. ZERO-WATERMARK EXTRACTION AND AUTHENTICATION

When the copyright of some data is infringed, Zero-watermark of the data can be extracted by third-party watermark extraction software permitted by the state. The watermark extracted is compared to the registered watermark, and the comparability of them can be determined. Therefore the copyright of the data might be affirmed and the evidence could be provided.

When the original data transferred by the network or the GIS vector data undergoing some normal operations, the data will be a little different to the original one. Also the zero-watermark of the latter data will not be same with the original one. Therefore, comparability measuring is necessary for watermark extraction and this enhances the robustness of zero-watermark.

The formula for computing the distance between two hash sequences is:

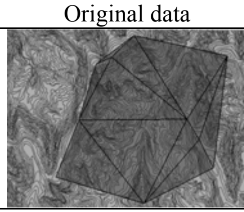
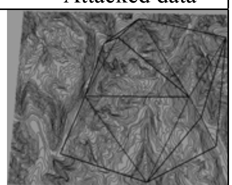
$$dis = norm\left(\frac{\left| hash_{new} - hash_{origin} \right|}{\sqrt{norm(hash_{new}) \bullet norm(hash_{origin})}} \right)$$

Set threshold T, if dis≤T, then match successful. Otherwise, if dis>T, then match failing. The norm here refers to 2-norms. Other comparability measuring method is also receivable.

6. EXPERIMENT RESULT AND PERFORMANCE ANALYSIS

6.1 Robustness validation

Undergoing the operation of projection transformation, coordinate transform ,data clip .etc , the contour layer subdividing character which takes the contained polyline number for example is shown in the Figures(Figure 2),the stability of the data character determined that this arithmetic is of good robustness.

Process	Original data	Attacked data
subdividing		
character (contained polyline number)	FID / POLPNUMBER 0 / 68 1 / 56 2 / 88 3 / 105 4 / 26 5 / 105 6 / 63 7 / 56 8 / 7 9 / 84 10 / 106	FID / FeatureNum 0 / 68 1 / 88 2 / 105 3 / 56 4 / 26 5 / 105 6 / 62 7 / 56 8 / 7 9 / 84 10 / 106

Figur2. Robustness validation and zero-watermarking generation

6.2 Safety analysis

The hash function meets the three basic safety requirements:

(1) Unilateralism. Given any GIS vector data, there must be a hash value, but the data cannot be reckoned from the hash value.

(2) Anti-collision. Given any GIS vector data, it is hard to find another different data which makes the hash value of these two data are the same or similar under the condition of ensuring data quality.

(3) Privacy of key. When the key is unknown, hash value could not be forged or estimated for it is generated under the key. Use different keys, different hash value will obtain.

6.3 Unique Existence analysis

The collision of zero-watermark refers to similar hash values are generated from different data content. In this algorithm, zero-watermark is constructed by the macroscopical topology characteristics and microcosmic topology characteristics of original data. The unique geography character of different data, the random of the processing of data generalization and the limited quantity of GIS vector data product determine that the possibility of hash collision is minimal. So the unique existence of hash value can be ensured.

6.4 Algorithm complexity and sequence length

This algorithm is based on traditional quantization modulation, and it has low complexity and can be easily programme implemented. On the other hand, no matter how mass the data is, the sequence length which is controlled by the security level is of a fixed value. This algorithm can meet the restriction requirements of algorithm complexity and sequence length.

7. CONCLUSIONS

The experiment results show that: On one hand, in this method, the watermark is registered to third-party website and not embedded into the original data, so there will not influence data precision and therefore keep the quality of original data well. Zero-watermark extraction of local map would be useful for local data authentication. On the other hand, the hash algorithm is robust, safe and exclusive.

ACKNOWLEDGEMENT

This research is financially supported by The HI-TECH Research and Development Program of China (863 Program) – Research on Key Technique for Copyright Protection of GIS Vector Data (NO. 2006AA12Z222).

REFERENCES

H. Zhang, Z. Li and L. Sun, 2004,Application of watermark System for Geographic information, Science of Surveying and Mapping, 29(7),pp.146-148.

Kitamura I, Kanai S, Kishinami T. 2000,Watermarking vector digital map using wavelet transformation [A]. In: Proceedings of Annual Conference on the Geographical Information Systems Association (GISA), pp: 417～421.

Q. Zhang, H. Xiang and X. Meng. 2005,Watermarking Vector Graphics Based on Complex Wavelet Transform. Journal of Image and Graphics. 10(4), pp.494-498.

Ryutarou Ohbuchi, Hiro Ueda, Shu Endoh. 2002, Robust watermarking of vector digital maps . In: The IEEE International Conference on Multimedia and Expo 2002 [C] , pp: 1～4.

Ryutarou Ohbuchi, Hiroo Ueda, Shuh Endoh. 2003,Watermarking 2D vector maps in the mesh-spectral domain [A]. In: The fifth International Conference on Shape Modeling and App lications,SMI2003 , pp: 216～228.

S. Zhong, Q. Gao, 2006, The Feasibility Analysis of Normalized-correlation-based Vector Maps Watermarking Detection Algorithm and the Improved Watermarking Algorithm, Journal of Image and Graphics, 11(3), pp. 401-409.

Solachidis V , Tefas A , Nikolaidis N , Tsekeridou S , Kikolaiddis A ,Pitas I. 2001. A benchmarking protocol for watermarking methods. In :Proceedings of 2001 IEEE International Conference on Image Processing (ICIP'01) , pp , 1023～1026.

X. Wang, X. Zhu and H. Bao, 2006,Complementary watermarking algorithm for digital grid map, Journal of Zhejiang University(Engineering Science), 40(6), pp:1056-1060.
Y.L.Ma,1998,Tutorial of cartology,Nanjing University Press,pp:4-5