

Selective Encryption Algorithm for Vectoral Geographic Data under Feature Point Grouping Strategy

Qianwen Zhou¹, Ying Xiong¹, Changqing Zhu¹, Na Ren^{1,2}

¹ Nanjing Normal University, Nanjing, China – (zqw@njnu.edu.cn, 1005104582@qq.com, zcq88@263.net)

² Hunan Engineering Research Centre of Geographic Information Security and Application, Changsha, China - renna1026@163.com

Keywords: Vector geographic data, Selective Encryption, Arnold Transform, Haar Transform, Feature points

Abstract

Selective encryption technology is a data security protection method that can balance encryption security and efficiency. Currently, there is ongoing research on applying this technology to vector geographic data. However, existing algorithms often use random selection methods for choosing objects to be encrypted, which results in lower security. To address this problem, we proposed a selective encryption algorithm for vector geographic data based on feature point grouping strategy. The study includes four steps: (1) Based on the user-input encryption ratio, calculate the feature point grouping thresholds for each element in the data. (2) Extract feature point sets for each element based on the grouping threshold. (3) Sort and group the feature point sets, reducing the smallest encryption unit to group objects within individual elements. (4) To reduce encryption costs and enhance the algorithm's resistance to attacks, perform frequency domain coefficient encryption and spatial domain coordinate value encryption in a stepwise manner on the group objects, ultimately producing the encryption results. The experiment showed that: (1) The decrypted encrypted data maintains consistency with the original data coordinates, achieving lossless decryption. (2) The correlation between encrypted data and the original data is significantly disrupted, resulting in a high level of randomness and thereby ensuring the algorithm's robust security. (3) The key space is substantial and highly sensitive, capable of withstanding brute force attacks. (4) The algorithm significantly improves encryption efficiency. (5) The algorithm displays resilience against deletion attacks and noise attacks. In conclusions, the proposed method enhances encryption efficiency while upholding a high level of security, thus it is an efficient selective encryption algorithm for vector geographic data.

1. Introduction

Vector geographic data serve as a strategic resource indispensable for national economic development (Wang, 2017). With the widespread application of geographic information services in various fields, there is an urgent need for online transmission of data. However, the security features of vector geographic data hinder its direct transmission over networks. Addressing the contradiction between the security characteristics of vector geographic data and the openness of cyberspace has become an urgent issue to be addressed.

As the core technology of information security, data encryption provides a feasible solution for the secure transmission of vector geographic data over networks (Zhu *et al.* 2020). It refers to the use of specific mathematical transformation methods to convert original data into ciphertext data, which can only be transformed back into plaintext data by legitimate users (Shannon, 1949). Encryption technology effectively protecting data security and safeguarding the legitimate rights of data owners (Shen *et al.* 2007). In recent years, some research progress has been made in the encryption technology of vector geographic data. Some scholars treat vector geographic data as ordinary binary files for direct encryption (Wang *et al.* 2021), while others consider the characteristics of vector geographic data and employ chaos-based compression, scrambling, or perturbation encryption methods (Park *et al.* 2020). However, these methods suffer from issues especially format disruption and low efficiency, thus affecting the use of data.

To balance security and efficiency requirements, some scholars have begun to explore the technology of selective encryption for geographic data (Sun, 2021). Selective encryption is a technique that encrypts important parts of the data or parts with rich information content, effectively balancing the demands for

security and efficiency (Anbo *et al.* 2018). Currently, selective encryption algorithms have been well researched and developed in the fields of images and videos (Huan, 2022; Zhang, 2022). However, research on selective encryption techniques for geographic data is still in its infancy, and there are many issues that require further investigation.

Selective encryption algorithms for vector geographic data can be classified into spatial domain-based and frequency domain-based algorithms, depending on the encryption objects. Spatial domain-based selective encryption algorithms are designed to consider the basic spatial structure of vector geographic data (Bang *et al.* 2016; Wang *et al.* 2021). These methods often simplify algorithms or extract feature points to encrypt features. For example, Pham *et al.* proposed a multiscale simplification algorithm to selectively simplify polyline and polygon data and obtain feature vertices for encryption (Pham *et al.* 2019). In his study, all nodes are subjected to random Gaussian distribution processing after encryption to enhance efficiency. While these methods address the efficiency issues of encrypting vector geographic data, they may still lack security, as plaintext data of the encrypted parts can be deduced from the unencrypted parts. Frequency domain-based selective encryption algorithms involve transforming spatial domain information into frequency domain coefficients using specific transformation functions (Ngo *et al.* 2016). Partial frequency domain coefficients are encrypted to achieve encryption of spatial domain information. For instance, Giao *et al.* employed the K-means clustering algorithm and combined it with random keys to cluster and XOR encrypt vector geographic data, followed by selective encryption of polygons in the discrete cosine transform domain (Giao *et al.* 2014). These methods have the advantage of simple judgment criteria, leading to fast processing efficiency. However, to improve efficiency, random selection is often used for selecting encryption objects, resulting in insufficient

algorithm security. In summary, selective encryption algorithms for vector geographic data have made significant progress in algorithm efficiency. However, there are still issues regarding the impact of the randomness in selecting encryption objects on algorithm security.

To address the aforementioned issues, this paper proposes a selective encryption algorithm for vector geographic data based on a feature point grouping strategy. The algorithm mainly consists of the following steps. Firstly, the algorithm calculates a feature point grouping threshold based on the user-defined encryption ratio, which represents the proportion of data to be encrypted in the global dataset. Secondly, using the calculated threshold, the algorithm sorts and groups the feature points one by one. Finally, the grouping is based on the threshold, and perform encryption on the grouped objects using frequency domain coefficient encryption and coordinate encryption techniques.

This approach aims to provide a method for selectively encrypting vector geographic data while maintaining security and efficiency. By employing a feature point grouping strategy, the algorithm effectively balances the encryption ratio and ensures the security of the encrypted data.

2. Basic Idea and Preliminaries

2.1 Basic Idea

In designing the selective encryption algorithm for vector geographic data, it is essential to consider both security and efficiency aspects. Therefore, it is necessary to ensure that the selected objects for encryption represent the crucial parts of the vector geographic data. Based on this premise, this paper takes individual features within the data as the fundamental units for encryption, considering these features as the key nodes of the data. Additionally, to address the issue of inability to identify encrypted points during decryption when directly extracting and encrypting features from the global data, this paper employs the encryption ratio provided by the user as the basis for grouping and sorting. Furthermore, to achieve randomized encryption of the overall data at a lower cost, this paper opts to encrypt the frequency domain coefficients. To withstand plaintext attacks, the paper also employs a cat face transformation to randomly encrypt the coordinate values of the data. The flowchart of the algorithm is illustrated in Figure 1.

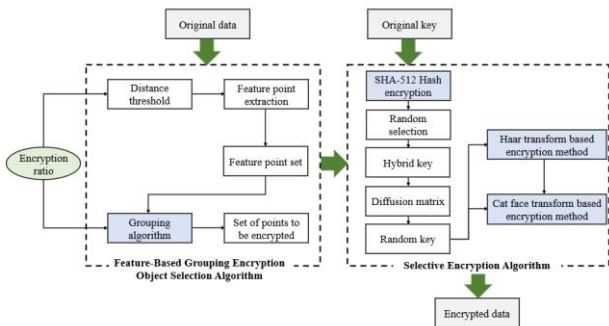


Figure 1. Algorithm flow chart

2.2 Distance Threshold Calculation Method

Vector geographic data is characterized by low redundancy and high precision. However, in order to better represent real-world geographic features, there are often many redundant data entries in the records. These redundancies contribute to a smoother

representation of geographic features. It is believed that these redundancies have little influence on the security of the data. Therefore, the proposed algorithm utilizes simplification algorithms to process the data, thereby selecting the parts that have a greater impact on security and are more significant.

Distance threshold d is the most critical aspect in the feature point extraction algorithm, which is determined based on the scale ratio before and after simplification. The purpose of extracting feature points in this paper is to differentiate the importance of subsequent feature element grouping. For this purpose, let d_{max} denote the maximum distance between all points in feature and the line connecting the two end points. When the distance threshold is $d = d_{max}$, only the two end nodes are selected as feature points; when $d = 0$, all feature points in the element are considered. These two extreme cases result in the same number of grouped feature points, making it indistinguishable. Therefore, to determine the optimal distance threshold scheme, d_{max} is taken as the reference value, and the value of the distance threshold d is adjusted accordingly. The specific calculation formula is as follows:

$$d = d_{max} * (1 - r), r \in (0, 1) \quad (1)$$

where d_{max} = maximum distance of all points from the line joining the two endpoints
 r = threshold adjustment parameter

The data encryption ratio α is the percentage of the number of encrypted points in the original data element. In practice, the encryption ratio is set according to the actual needs, when a higher degree of encryption is required, α is set close to 1. In order to obtain the relationship between the value of the threshold adjustment parameter r and the data encryption ratio α , a line element with 100 points is used here as an example for the experiment. Statistical changes in the value of α for different r are shown in Figure 2.

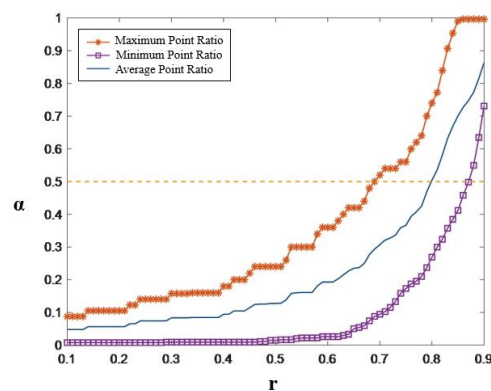


Figure 2. Proportion of feature points

As can be seen from Fig. 2, when the parameter r is close to 0, the data encryption ratio α is close to 0; when the parameter r is close to 1, the data encryption ratio α is close to 1, which overlaps with the previous analysis. In this paper, in order to make the number of grouped feature points have a certain degree of differentiation, we chose 0.5 as the data encryption ratio, the corresponding parameter r takes the value of 0.8, and the converted distance threshold is $d = 0.2 \times d_{max}$.

2.3 Encryption Object Selection Algorithm

In this section, use the encryption ratio α input by the user to calculate the distance threshold d corresponding to the elements to be encrypted, and then group the elements to be encrypted to obtain the encryption group G . Based on the number of feature points within the group to determine the group of elements to be encrypted. Finally, complete the first step of the algorithm for the selection of encrypted objects for the geographic elements.

Step1: Extraction of feature point sets. Assuming the vector geographic data contains m features, forming the feature set $PL = \{O_i | i \in [1, m]\}$, the specific operations are as follows:

(1) For each feature O_i in PL , connect the start and end points to form a line, and calculate the set of distances formed by the perpendicular distances between the points (excluding the start and end points). Sort the set to obtain the maximum distance value $d_{\max i}$.

(2) Based on Equation (1) and considering the user-input encryption ratio α , determine the feature distance threshold d_i . Then, extract the set of feature points that exceed the distance threshold.

Step2: Feature grouping. For each feature O_i in PL , its set of feature points needs to be assessed to determine if it can serve as an encryption object. The specific assessment rules are as follows:

(1) Calculate the average number of feature points num per group. The calculation formula is given by Equation (2). When $N = 5$, the grouping of features is illustrated as shown in Figure 3.

$$num = \left\lfloor \frac{p_i}{N} \right\rfloor \quad (2)$$

where p_i = the total number of points for feature O_i
 N = the number of groups

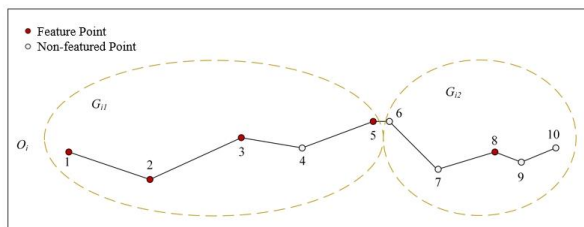


Figure 3. Schematic diagram of element grouping

(2) For each group G_{ij} within O_i , compute the array V_i containing the number of feature points in each group. Sort this array in descending order to obtain the sorted array V'_i of feature point counts.

(3) Based on the user-input encryption ratio α , calculate the threshold $k = \alpha \times N$. If $V_{ij} \geq k$, G_{ij} is considered to require encryption.

2.4 Random Key Generation

Key processing is also one of the steps to improve the algorithm's security. Since Hash functions have one-way transformation property, they can resist known plaintext attacks, chosen plaintext attacks, and ciphertext attacks during the encryption process. Therefore, to further enhance the security of the algorithm, before data encryption, the proposed algorithm uses the SHA-512 Hash algorithm to generate a random key for the original data.

Step1: The original key K is used as input to the Hash function for multiple encryption iterations. Each encryption result is a 128-bit hexadecimal array sequence, denoted as $S = [s_1, s_2, \dots, s_{128}]$. It is then converted into a binary array $S = [s_1, s_2, \dots, s_{512}]$.

Step2: Take $S' = [s_{502}, s_{503}, \dots, s_{512}]$ and use it as the initial value, combined with Equation (3) to calculate the sequence of indices p_i selected from the original vector geographic data.

$$N' = \text{mod}(\text{hex2dec}(S'), N) \quad (3)$$

$$p_i = P(N, 1)$$

where $\text{hex2dec}()$ = the conversion of hexadecimal to decimal
 N = the number of elements in the original data

Step3: Encrypt the coordinates corresponding to S and p_i for a second time using the encryption formula as shown in Equation (4).

$$PAD = \text{SHA512}(\text{SHA512}(p_i(x)) \oplus \text{hex2dec}(\text{SHA}(S))) \quad (4)$$

where \oplus = the XOR operator

$p_i(x)$ = the horizontal coordinate value corresponding to the selected original data sequence
 PAD = the generated random pad

Step4: The diffusion matrix D , as defined in Equation (5), is used to perform diffusion on the PAD values. The random key PAD' is obtained as described in Equation (6).

$$D = \begin{bmatrix} 1 & 1 & 1 & k \\ 1 & 1 & 1 & k+1 \\ 1 & 1 & 1 & k+2 \\ 1 & 1 & 1 & k+3 \end{bmatrix} \quad (5)$$

$$PAD' = (PAD \times D) \text{ mod } 255 \quad (6)$$

where k = any real number

2.5 Data Encryption

The encryption algorithm proposed for vector geographical data mainly consists of two steps: firstly, utilizing the Haar wavelet function to encrypt spatial domain coordinates into encrypted frequency domain parameters, thereby achieving a random perturbation effect on the overall shape; secondly, employing the Cat Face Transform to computationally encrypt spatial coordinates, enhancing the algorithm's resistance to plaintext attacks. Assuming the current sequence to be encrypted is Seq , the encryption process is as follows:

Step 1: Coordinate randomization. Extract the first node of each encryption group as the reference node and form the encryption sequence $S = \{S_i = (x_i, y_i) | i = 1, 2, \dots, num\}$. Based on the random sequence generated by the chaotic system, a random sequence g of length num is extracted. This sequence is then multiplied element-wise with the encryption sequence S to perform coordinate randomization, resulting in S' . Calculate the mean and standard deviation (\bar{X}, σ) of this sequence for subsequent decryption key transmission.

Step2: Frequency component encryption. Firstly, Split S' into two sequences x' and y' , and process them as input signals for the Haar wavelet transform to obtain the approximation coefficient Ac and detail coefficient Dc . Then, encrypt the generated function coefficients using the random sequence g_i according to the encryption formula as shown in Equation (7).

$$\begin{aligned} Ac'_i &= Ac_i * g_i \\ Dc'_i &= Dc_i * g_i \end{aligned} \quad (7)$$

where Ac_i = the approximation coefficient
 Dc_i = the detail coefficient
 g_i = the random sequence

Use Ac'_i and Dc'_i as inputs for the inverse transform, apply the Haar wavelet inverse transform to the coefficient sequences to obtain the encrypted point sequence.

Finally, perform a shift operation on the coordinates of the encrypted point sequence, with the mean \bar{X} as the central reference point. Modify the coordinates of the encrypted points using the standard deviation offset to obtain the final encrypted result coordinates.

Step3: Intra-group encryption. For the encryption group $G_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}$, where n represents the total number of points, p_{i1} is taken as the reference point and be encrypted in step2 to obtain p'_{i1} . Other points within the group are encrypted using the following formula:

$$p'_{ij} = (-1)^j * \sin(i + j) * p'_{i1} + p_{ij} \quad (8)$$

where i = the encryption group index
 j = the index of the point within this encryption group
 p'_{i1} = encrypted reference point
 p_{ij} = the point to be encrypted

Step4: Coordinate encryption. For each feature point p_i , extract the integer part of its coordinates as $[x_i, y_i]$ and determine the number of transformation iterations n based on the random numbers generated by the key. The basic format of the cat face transform matrix is given as $\begin{pmatrix} 1 & b \\ a & ab+1 \end{pmatrix}$, In this

section, let $a=1, b=1$. The formula for coordinate encryption is shown in Equation (9).

$$\begin{cases} x'_i = x_i + y_i \\ y'_i = x_i + 2 * y_i \end{cases} \quad (9)$$

where (x_i, y_i) = the original coordinates of the feature point
 (x'_i, y'_i) = the encrypted coordinates

Obtain the encrypted integer part of the coordinates $[x'_i, y'_i]$, and combine it with the decimal part of the original coordinates to form the encrypted data. To ensure that the coordinates of the encrypted data are within a certain range, further standardization of $[x'_i, y'_i]$ is required. In this section, proportional standardization is chosen to process the coordinates, limiting the range of the encrypted coordinates to within 100. The final result xi'' is obtained according to Equation (10).

$$x_i'' = \frac{x'_i}{x_{max}'} * 100 + x_{max}'' \quad (10)$$

where x_{max}' = the maximum encrypted horizontal coordinate value

2.6 Data Decryption

Data decryption is the reverse process of data encryption, where the most important step is to locate the first point of the encrypted object point group. The specific implementation steps are as follows.

Step1: Data decryption. Extract the coordinates from the obtained ciphertext data to obtain $[x_i'', y_i'']$, and obtain the maximum values $[x_{max}'', y_{max}'']$. Based on $x_{max}'' = x_{max}' - 100$, obtain the maximum value of the x-coordinate before standardization, with similar handling for the y-coordinate. Using Equation (11), reverse-standardize all coordinates to obtain the encrypted coordinates $[x'_i, y'_i]$, and complete the decryption of the data after reverse-standardization.

$$xi' = \frac{(xi'' - x_{max}'') * x_{max}'}{100} \quad (11)$$

where x_i'' = the encrypted and standardized x-coordinate
 x_{max}' = the maximum x-coordinate after encrypting

$$\begin{cases} x_i = 2 * x'_i - y'_i \\ y_i = y'_i - x'_i \end{cases} \quad (12)$$

where (x'_i, y'_i) = the reverse-standardized coordinate
 (x_i, y_i) = the decrypted coordinate

Step2: Finding the encrypted object. For each feature of the vector geographic data, use $\bar{X} + 3\sigma$ to determine each point. When the point exceeds this standard positive or negative value, it is considered as the first point of the encrypted point group,

and the subsequent $n - 1$ points are considered as other points in the group.

Step3: Group decryption. For the encrypted group $G_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}$, where n is the total number of points in the group and p_{i1} serves as the reference point. The decryption formula for other feature points within the group is as shown in Equation (13).

$$p_{ij} = (-1)^j * \sin(i + j) * p'_{i1} - p'_{ij} \quad (13)$$

where i = the encrypted group number
 j = the point number within the encrypted group
 p'_{i1} = encrypted reference point
 p'_{ij} = the encrypted point

Step4: Frequency domain coefficient decryption. Reverse the encrypted point sequence coordinates by performing a reverse shift operation, using the mean value \bar{X} as the centre point, and modify the coordinate values using the standard deviation to obtain the pre-shift coordinates. Then, perform the same operation as step 2 of the encryption operation to obtain the decrypted point sequence.

3. Experiments and Results

The paper selected four Shapefile datasets as experimental data, including river data, road data and administrative boundary data of Jiangsu Province, and building data of Shanghai, China. The detailed information of the data is shown in Table 1, and the visualization of the data is presented in Figure 4.

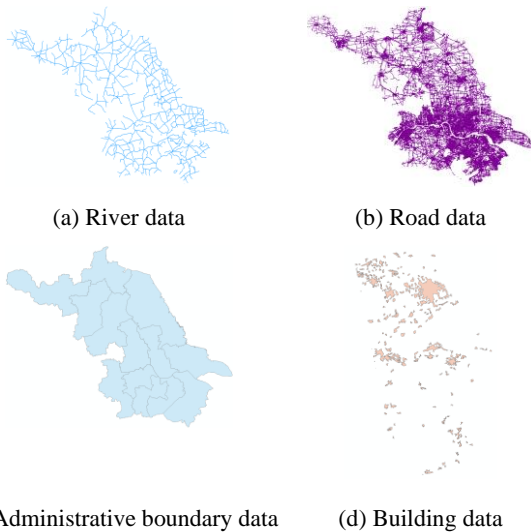


Figure 4. Visualization of experimental Data

Data	Type	Element number	Point number	Data size / (KB)
River data	Line	752	12773	286
Road data	Line	148784	1368822	27196
Administrative boundary data	Surface	13	15547	244
Building data	Surface	204	19729	264

Table 1. Details of experimental data

3.1 Visualization of Results

Based on the selective encryption algorithm proposed, encryption and decryption experiments were conducted on the four types of experimental data mentioned above. The initial key chosen for encryption in this paper is key=[administrator]. To evaluate the effectiveness of encryption, visual comparisons were made between the data before and after encryption and decryption as shown in Table 2.

Index	Original Data	Encrypted Result	Decrypted Result
1			
2			
3			
4			

Table 2. Visualization of the results after encryption and decryption

From Table 2, it can be observed that the visual differences between the encrypted vector geographic data and the original data are significant. The encrypted data cannot provide valid information from the original data and cannot be processed or analysed normally. However, the decrypted data shows no visual differences compared to the original data.

3.2 Statistical Quantitative Analysis

To evaluate the security and lossless decryption capability of the proposed algorithm in this paper, it is necessary to conduct statistical quantitative analysis on the encrypted and decrypted data, including coordinate difference analysis, adjacent feature point correlation analysis, and entropy analysis.

3.2.1 Coordinate Difference Analysis

The coordinate difference of vector geographical ciphertext data is an important indicator to express the resistance of the algorithm to statistical analysis attacks. Existing vector geographical data encryption algorithms mostly use the Mean Square Error (MSE) of point sets before and after encryption to represent this. The calculation formula of MSE is shown in Equation (14).

$$MSE = \frac{1}{n} \sum_{i=1}^n [(x_i - x'_i)^2 + (y_i - y'_i)^2] \quad (14)$$

where n = the total number of data points
 (x_i, y_i) = the coordinate before encryption

(x'_i, y'_i) = the coordinates after encryption

Data	The MSE after encryption	The MSE after decryption
River data	1.66e+04	0
Road data	1.69e+04	0
Administrative boundary data	9.98e+12	0
Building data	2.17e+04	0

Table 3. The MSE after Encryption and Decryption

A larger MSE value indicates a greater difference in coordinate points between the data. From Table 3, it can be observed that the MSE between the encrypted data and the original data is significantly high, indicating a considerable deviation and thus a disruption in the data utility, which implies a good security level of the encryption algorithm. On the other hand, the MSE between the decrypted data and the original data is 0, indicating that the decrypted data perfectly matches the original data, demonstrating the lossless decryption capability of the proposed algorithm.

3.2.2 Analysis of Adjacent Feature Point Correlation

Neighbourhood correlation is one of the important characteristics of vector geographic features. The smaller the correlation between adjacent coordinates of encrypted vector geographic data, the stronger the resistance of the encryption algorithm to statistical analysis attacks. The algorithm proposed used Pearson correlation coefficient as the basis to measure the correlation between adjacent coordinates of data before and after encryption and decryption. The calculation formula is as follows:

$$\begin{aligned}
 E(x) &= \frac{1}{n} \sum_i^n x_i \\
 D(x) &= \frac{1}{n} \sum_{i=1}^n (x_i - E(x))^2 \\
 \text{cov}(x, y) &= \frac{1}{n} \sum_{i=1}^n ((x_i - E(x)) * (y_i - E(y))) \\
 r_{xy} &= \frac{\text{cov}(x, y)}{\sqrt{D(x)}\sqrt{D(y)}}
 \end{aligned}
 \tag{15}$$

where $E(x)$ = the mean of x
 $D(x)$ = the standard deviation of x
 $\text{cov}(x, y)$ = the covariance between x and y
 r_{xy} = the correlation coefficient between adjacent coordinate

The correlation coefficients before and after encryption can be computed as shown in Table 4.

Data	Original Data		Encrypted Result		Decrypted Result	
	X	Y	X	Y	X	Y
River data	0.999	0.999	0.089	0.132	0.999	0.999
	973	971	328	946	973	971
Road data	0.999	0.999	0.233	0.661	0.999	0.999
	998	998	127	721	998	998

Administrative boundary data	0.999	0.999	-	-	0.999	0.999
	990	991	0.146	0.315	990	991
			314	103		
Building data	0.999	0.999	-	-	0.999	0.999
	874	959	0.080	0.022	874	959
			110	691		

Table 4. Correlation of Data Adjacent Coordinates

From Table 4, it can be observed that the correlation coefficients of the encrypted data are relatively small and close to 0, indicating that the correlation between adjacent points in the data has been disrupted. Overall, the results demonstrate that the selective encryption algorithm proposed effectively disrupts the correlation between data points and possesses strong resistance to statistical analysis attacks.

3.2.3 Information Entropy Analysis

Information entropy is an important evaluation metric for reflecting the randomness of data encryption algorithms. The higher the information entropy, the stronger the resistance to attacks of the encrypted data. Therefore, this paper evaluated the security of the proposed method by calculating the entropy of encrypted vector geographic data. For each feature L_i , the formula for calculating information entropy is as shown in Equation (16) (Cover and Thomas, 2006). The overall information entropy H_L of the encrypted data L is the sum of the entropies of all encrypted elements.

$$\begin{aligned}
 H_{L_i} &= H(K) + H(PAD') \\
 &= |K| * \log_2 |K| + |PAD'| * \log_2 |PAD'|
 \end{aligned}
 \tag{16}$$

where K = the original encryption algorithm key
 PAD' = the random pad

The comparison of information entropy between the algorithm and the reference algorithm (Bang *et al.* 2016) is shown in Table 5.

Data	Number of points	Entropy of proposed algorithm/(dB)	Entropy of reference algorithm/(dB)
River data	12773	98892	4608
Road data	15547	123010	4608
Administrative boundary data	19729	160170	4608
Building data	1368822	16136000	4608

Table 5. Comparison of information entropy

Since the encryption algorithm in reference [17] uses a uniform key to encrypt data, its information entropy remains constant for different datasets. Table 5 reveals that the information entropy value of this algorithm is greater than that of the algorithm in reference, and the information entropy value increases with the number of elements. These results indicate that this algorithm possesses higher security.

3.3 Key Analysis

3.3.1 Key Space Analysis

The analysis of the key space for selective encryption primarily involves assessing whether the random password generated by the chaotic system mapping can resist brute force attacks. In this algorithm, the initial key used is $key=[administrator]$, and the initial parameters of the chaotic mapping (x_0, μ, n) are calculated from the initial key and plaintext information. Therefore, the calculation of the key space only needs to consider the key K . The SHA-512 Hash algorithm chosen in this algorithm achieves a key space of 2^{512} , which means that even with a computer capable of performing 10 trillion calculations per second, the time required for exhaustive attacks would be:

$$\frac{2^{512}}{10^{13} \times 3600 \times 24 \times 365} = 4.25 \times 10^{133} \text{ years}$$

Therefore, the encryption algorithm's key space security is significant, demonstrating strong resistance against exhaustive attacks and the ability to withstand brute force cracking.

3.3.2 Key Sensitivity Analysis

In the decryption experiments, it was observed that using the correct initial key could decrypt and retrieve the original data. To verify the sensitivity of the key, the last bit of the key was modified, and the decryption results of the correct key and the slightly modified correct key were compared. The results are shown in Table 6.


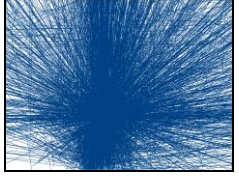

Index	Decryption key	Decryption result
1	70616E64616D616461 (Correct Key)	
2	70616E64616D616465	
3	70616E64616D616462	

Table 6. Decryption Results for Different Decryption Keys

From Table 6, it can be observed that in Experiment 1, where the correct key was used to decrypt the encrypted data, the decrypted result visually matched the original plaintext image perfectly. In Experiments 2 and 3, where decryption was attempted using incorrect keys with minor modifications to the last hexadecimal digit. By visually comparing the decryption results, it is evident that even though the keys only differed by a change in the last hexadecimal digit, the decryption results were incorrect when using the wrong keys. This demonstrates the high sensitivity of the algorithm to the key.

3.4 Encryption and Decryption Efficiency

To verify the efficiency of the encryption and decryption process in the proposed algorithm, a comparison was made with the selective encryption algorithms proposed in references (Pham *et al.* 2019) and (Sun, 2021). The experiments were conducted with different sizes of vector geographic data using the same encryption key. The results are presented in Table 7.

Data size/(KB)	Coordinate point count	Proposed algorithm encryption time/ (s)	Encryption time for Pham's algorithm/(s)	Encryption time for Sun's algorithm / (s)
286	12773	0.48	4.73	1.65
27196	1368822	144.52	264.85	163.38
45347	2743561	248.35	598.21	317.64
404219	20533665	1658.37	7427.14	-*

Table 7. Comparison of encryption efficiency

*: "-" indicates that the algorithm could not produce a result.





From Table 7, it can be observed that as the size of the data and the number of coordinate points increase, the efficiency of the algorithm gradually decreases. The algorithm in Pham's algorithm using DES encryption method results in longer overall encryption time. The algorithm in Sun's algorithm requires higher memory usage due to the use of global permutation. The results indicate that the proposed algorithm has a significant advantage in handling large datasets, with an average efficiency improvement of 34.51%.

3.5 Analysis of Resistance to Attacks

Resistance to attacks is one of the crucial indicators of the interference resistance capability of encryption algorithms. Considering the practical application scenarios of network transmission, this study assessed the robustness of the proposed algorithm through deletion attacks and noise attacks.

3.5.1 Deletion Attack Experiment

Deletion attack is a commonly used test for assessing attack capabilities. In this experiment, deletion attack refers to the removal of some elements from the vector geographic data. The experiment involved deleting 10%, 30%, 50%, 70%, and 90% of the encrypted test data. The results of the deletion attack experiment are shown in Table 8.

Deletion ratio/ (%)	Remaining element count	Decrypted data
10%	677	
30%	526	
50%	377	
70%	225	

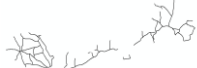
90%	77	
-----	----	---

Table 8. Experimental Results of Deletion Attack

From Table 8 it can be observed that since the encryption algorithm operated at the element level rather than the entire document, deleting elements from the data did not affect the decryption process of other elements. It demonstrates that the proposed encryption algorithm exhibits strong robustness against deletion attacks.

3.5.2 Noise Attack Experiment

In actual data transmission processes, communication channels often suffer from noise interference. In this experiment, different signal-to-noise ratio (SNR) intensities of Gaussian noise are embedded into the encrypted data, and the results after decrypting with the correct key are shown in Table 9.




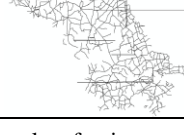
Noise interference intensity/ (dB)	Decrypted data
10	
1	
0.1	
0.01	

Table 9. Experimental results of noise attack

From Table 9, with a higher ratio of useful data information to noise interference, the quality of the decrypted data gradually decreased. However, it is evident from the table that under an interference intensity of 0.01dB, the decrypted data still maintains the general shape and positional information of the original data. These results indicate that the encryption and decryption algorithm proposed in this paper possesses good resilience against noise attacks.

4. Conclusion

The proposed algorithm in this paper is a selective encryption algorithm for vector geographic data based on a feature point grouping strategy. It utilizes the sensitivity of vector geographic features input by users to extract feature points and encrypt them in groups. The selected feature point groups undergo encryption of frequency domain coefficients and coordinate points. Experimental results demonstrate that the proposed method maintains high security while improving encryption efficiency. It exhibits significant advantages in handling large volumes of data and resisting attacks even under encryption.

References

- Anbo, L., Ying, C., Yao, M.M., 2018: Quantitative Measurement of Geometrical Information for Sensitive Features in Secret-related Vector Digital Maps. *Journal of Geoinformation Science* 20(01), 7-16.
- Bang, N., Lee, S., Moon, K., 2016: Encryption algorithm using polyline simplification for GIS vector map. *Journal of Korea Multimedia Society* 19(8), 1453-1459.
- Cover, T., Thomas, J., 2006: Elements of Information Theory. *Tsinghua University Press*.
- Giao, P., Kwon, G., Lee, S., 2014: Selective encryption algorithm based on DCT for GIS vector map. *Journal of Korea Multimedia Society* 17(7), 769-777.
- Huan, L.X., 2022: Selective Chaotic Image Encryption Based on Salient Object Detection. *Modern Information Technology* 6, 84-88+91.
- Meng, B., 2018: Research on Digital Image Encryption Algorithm Based on Chaotic Technology. *Tianjin University*.
- Michalis, P., Dowman, I., 2008: A Generic Model for Along-Track Stereo Sensors Using Rigorous Orbit Mechanics. *Photogrammetric Engineering & Remote Sensing* 74(3), 303-309.
- Ngoc, G., Moon, K., Lee, S., 2016: GIS Map Encryption Algorithm for Drone Security Based on Geographical Features. *2016 International Conference on Computational Science and Computational Intelligence*.
- Park, J.H., Lee, D.H., 2010: A hidden vector encryption scheme with constant-size tokens and pairing computations. *IEICE transactions on fundamentals of electronics, communications and computer sciences* 93(9), 1620-1631.
- Shannon, C., 1949: Communication theory of secrecy systems. *The Bell system technical journal* 28(4), 656-715.
- Shen, C.X., Zhang, H.G., Feng, D.G., 2007: Overview of Information Security Overview. *Science in China Series E: Information Sciences* 129-150.
- Sun, X.H., 2021: Research on Selective Encryption Algorithm for Vector Geographic Data Based on Geospatial Information *Nanjing Normal University*.
- Wang, J.Y., 2007: Development Trends of Cartography and Geographic Information Engineering. *Acta Geodaetica et Cartographica Sinica* 000(005), 1-6.
- Wang, X.L., Yan, H.W., Zhang, L.M., 2021: Vector Map Encryption Algorithm Based on Double Random Position Permutation Strategy. *ISPRS International Journal of Geoinformation* 10, 311.
- Zhang, Z.R., 2022: H.264 video selective encryption communication system based on chaotic encryption algorithm. *Guangdong University of Technology*.
- Zhu, C.Q., Ren, N., Zhou, Z.C., 2020: Research Status and Prospect of Security Technology for Geographic Big Data. *Modern Surveying and Mapping* 43(06), 9-13.