Accurate Identification Technique of Access Path Information for Hierarchical Geographic Traffic Network under Big Data Construction

Professor Dr. Soroush Niknamian

BSC in Applied Mathematics (IAUCTB), BSC in Mechatronics Engineering (SUT)

and PhD in Cell and Molecular Biology and Military Medicine, Liberty University

(LU), USA

Email: so.ninamian@gmail.com

Abstract: The traditional identification method of access path information requires a lot of data packets, and uses a fixed probability, and the identification accuracy is low. To this end, a new method for accurate identification of access path information of hierarchical geographic traffic network is proposed. Under the construction of big data, the access path information is marked with a certain probability, and part of the access path information of the corresponding router is recorded. According to the collected data packet, the accessed node reconstructs the path by using the access path information identified in the packet, to determine the identity range selection and identification format. According to the big data, the hierarchical geographic traffic network router is constructed. According to p_1 , it is judged whether the current data packet is identified, and the identification result is judged by p_2 . The particle cluster algorithm is used to optimize the identification result of the parameter and the penalty factor to complete the accurate identification of access path information of hierarchical geographic traffic network. The experimental results show that compared with the

traditional method, the proposed technique has high accuracy and strong convergence.

Keywords: big data; hierarchical geographic transportation network; access path; information; identification

1. Introduction

With the rapid development of China's economy, high-speed transportation has played an important role in the development of the national economy as an important means of transportation. The national expressway has developed from a number of routes connecting important cities to a high-speed road network covering the country's criss-crossing and extending in all directions. It provides more route choices for road transporters, reduces transportation time, effectively disperses traffic pressure and relieves traffic pressure to a certain extent. With the rapid increase of vehicles in China, traffic problems have become an important factor restricting China's economic development. How to provide effective traffic management and advanced traffic information services has become an important issue that the transportation industry needs to solve urgently. The key is the real-time path information identification problem under large-scale hierarchical geographic traffic network. In the era of big data, people enjoy rich resources, but they also increase the burden on users to get resources [1, 2]. With the wide application of online shopping, chat software and e-mail services in people's lives, the role of hierarchical geographic transportation network in different fields is increasing [3]. While providing network connectivity to other users at different time and place, it also faces serious network security issues. Under the construction of

big data, access path information identification of hierarchical geographic transportation network is an effective defense technique, which can solve the network security problem of hierarchical geographic transportation network [4, 5].

There are many researches on the identification of access path information. Among them, the PPM method identifies the IP address fragment in the corresponding range. However, when the access path is higher than 16 hops, the reconstructed path used for identification needs many data packets [6]; Hash function encoding method can identify the complete IP address, greatly reducing the number of data packets required for access path reconstruction, however, it uses fixed probability [7] and Huffman coding identification method [8], which can track the completion of the attack source through identification, but the detailed identification range is not pointed out. In the case that the identification range is insufficient, the identification package information needs to be saved to the hierarchical geographic transportation network router, which not only consumes storage space, but also has low precision. Literature [9] proposed a solution to accurately identify the driving path of expressway vehicles by using mobile positioning technique, and analyzed the feasibility of the scheme from two aspects: technical difficulty and economic cost, and obtained the conclusion that the practical solution was feasible, economical and practical, but this method also had the problem of low identification accuracy.

Aiming at the shortcomings of the above methods, this paper proposes a new big data construction technique to accurately identify the access path information of hierarchical geographic transportation network. The identification of the access path information is

completed by the dynamic probability packet marking, which provides a basis for path detection.

2. Material and methods

2.1 Basic Probability Packet Identification

The basic process of packet identifying is as follows: under the construction of big data, the data packets in the hierarchical geographic transportation network are saved to the packet by the router, so that all the data packets received by the target host contain the complete access path through the network [10,11].

Probability packet identification means that the router completes the identification of the access path information with a certain probability, and records part of the access path information of the corresponding router. With the collected data packet, the accessed node can reconstruct the access path through the access path information in the packet, to guarantee the security of hierarchical geographic transportation networks [12, 13].

The complete identification process is divided into two parts: identification and path reconstruction [14]. In the process of identification, the router in the hierarchical geographic transportation network identifies the packets passing through it with a certain probability p for the access path information. In this section, the access path information and the 32-bit hash value of hierarchical geographic traffic network are divided into 6 slices, each slice is 10 bits, and one slice is arbitrarily selected and added to the ID field of the IP packet, and the slice offset is recorded through the 4-bit space [15, 16]. The topology of hierarchical geographic traffic network typically has about 20

hops, less than 30 hops, and the identified access path information can be placed in a 15-bit ID field.

In the process of re-routing the route access path, the accessed node reconstructs the path according to the access path information stored in the package to provide security for the access security of the hierarchical geographic transportation network [17]. Usually, seven fragments with the same distance domain but different offsets are selected and combined. If the distance domain is 0, the identified access path information is an IP address. The router IP of hierarchical geographic transportation network in upstream is obtained by using the part adjacent to upstream to complete the reconstruction of the access path.

The identification efficiency of access path information of the hierarchical traffic network under the big data construction is generally measured by the convergence time, which is essentially determining how many packets are needed to complete the recovery of the intrusion path.

It is difficult to obtain an appropriate identification probability. Supposing that $B = \{S, R_1, R_2, \dots, R_d, V\}$ is used to describe the access path from S to V . The probability that a packet is identified by the router and not identified by other routers is as follows:

$$\beta_i = p(1-p)^{d-i} \quad (1)$$

Where the probability that a packet is not identified by the router in B can be described as:

$$\beta_0 = (1-p)^d \quad (2)$$

Under the big data construction, after a packet is identified by the router, it will be re-identified by other routers [18-20], that is, the historical identifier value will be overwritten, and $\beta_d \geq \beta_{d-1} \geq \dots \geq \beta_2 \geq \beta_1$, that is to say, the shorter the interval between the router and the intruder is, the greater the possibility that the intruder receives the router identification information is; otherwise, the greater the interval between the router and the intruder is, the less likely the intruder receives the router identification information does[21, 22]. The probability that an intruder receives at least one tagged packet can be obtained by:

$$\sum_{i=1}^d \beta_i = \sum_{i=1}^d \left(p \prod_{j=0}^{d-i} (1-p)^j \right) = 1 - (1-p)^d \quad (3)$$

Assuming that the number of packets required to reconstruct the access path is described by N , its expected value can be found by:

$$E(N) \leq \frac{k(\ln kd)}{p(1-p)^{d-1}} \quad (4)$$

Where k is used to describe the number of segments for all edge ids. If the visitor does not receive a packet with a route tag, the visitor will not be able to discover that the router is in the access path [23, 24]. Therefore, in order to reconstruct the access path, the visitor needs to receive at least one tagged packet from all routers in the access path.

In order to achieve the above conditions, the following conditions must be met:

$$N\beta_1 = Np(1-p)^{d-1} \geq 1 \quad (5)$$

Since β_1 is the smallest and N is controlled by the accessed node, the number of packets identified by the S and R_1 edges is the least. The edge value is called the weakest chain, which improves the convergence speed and has a bad influence on the

identification result. If the following conditions are met:

$$\beta_d = \beta_{d-1} = \dots = \beta_2 = \beta_1 = \frac{1}{d} \quad (6)$$

According to the Coupon Collector question, the number of packets N required to reconstruct the access path is minimized. The best structure of the research problem in this section is that all the packets identified by the router along the access path arrive at the visitor, so the probability that the visitor receives a identified packet should be the initial identification probability of the packet.

2.2 Identification range selection and identification format determination

The header is mainly composed of a header and an extension header. The basic header consists of a version, a communication type, and a label. The detailed information is described in Table 1.

Table 1. Header format

Edition	Type of communication		Flow label
	Load length (15 bit)	Next header	Hop Limit
	Source address (128bit)		
	Destination address (128bit)		

The access path information of hierarchical geographic traffic network can be identified in the identification range without being transferred to the router under big data construction, so the identification of access path information can be replaced by the $de(\text{delimiter})$ field instead of the sf field, and the ls field does not change; furthermore, the distance field is added to identify the primary link number, i.e. the distance field is added to identify the first link number. The number of hops separated by the visited nodes is recorded [25, 26], which is regarded as the basis of judging

whether the path reconstruction is completed in the process of access path reconstruction. The detailed format is described in Figure 1.

Figure 1. Identification Format

In Figure 1, the high de is replaced by 0 and placed before the effective access path information flow. All routers have 4 adjacent contacts, corresponding to 4 links, and only 4 bit encoded link numbers. When a visitor sends a data packet, the number of hops is less than 16 hops, the distance domain is 5 bits, all link numbers are 3 bits, the link numbers are stored in order, and the number of bits of ls is 3×16 bits.

2.3 Determination of dynamic probability

When a fixed probability is selected, the access path identification information of the routers that carrying the short interval between the visitors will be identified repeatedly the routers that have passed through, and the visited node has unacceptable router identification information with a short interval from the visitor [27, 28]. Supposing that the access path is described by $H = (U, S_1, S_2, \dots, S_d, V)$, U represents the visitor, V represents the visited node, S_i is used to describe the router of the access path. When the probability of access path identification is p , assuming that the total number of packets sent by the visitor is described by n , the visited node must receive at least one access path packet identified by the router nearest to the visitor, that is, $np(1-p) \geq 1$.

Under all big data structures, the number of convergent packets needed for reconstructing access paths is the smallest [29], when the packet of hierarchical geographic traffic network is identified by the current router and the probability that it will not be identified by the downstream router is $1/d$.

In any access path, assuming that i is the distance traveled by the packet q from the access node. The router identifies the access path packets with the identification probability $p_i = 1/i$. At this stage, the probability of packets being identified by any path is $1/d$, that is:

$$\begin{cases} p_d = 1/d \\ p_{d-1}(1-p_d) = 1/d \\ p_{d-2}(1-p_{d-1})(1-p_d) = 1/d \\ \vdots \\ P_2(1-p_3)\cdots(1-p_{d-2})(1-p_{d-1})(1-p_d) = 1/d \\ P_1(1-p_2)(1-p_3)\cdots(1-p_{d-2})(1-p_{d-1}) = 1/d \end{cases} \quad (7)$$

There are:

$$\begin{cases} p_d = 1/d \\ p_{d-1} = 1/(d-1) \\ \vdots \\ p_2 = 1/2 \\ p_1 = 1 \end{cases} \quad (8)$$

By analyzing the above equation, we can see that when the change rate is $1/j (j \in [1, d])$, the number of packets needed by the access side to reconstruct the access path is the lowest, the mean value is d , and the corresponding expected value is $d \ln d$.

A large number of studies have shown that the majority of the data packets transmitted

through the router in the hierarchical geographic traffic network are less than 25 hops.

According to the rate of change $1/j$ [30], all router identification probabilities are:

$$p_1 = \begin{cases} 1/(32-s,ttl), w,ttl < 32 \\ 1/(64-s,ttl), 32 \leq w,ttl < 64 \\ 1/(128-s,ttl), 64 \leq w,ttl < 128 \\ 1/(255-s,ttl), 128 \leq w,ttl < 255 \end{cases} \quad (9)$$

In the process of identifying the access path information, the ratio of the required data packets between the domains and in the domains is about 1:2, so in the case of $p_2 \leq 1/3$, the access path information is identified.

2.4 Access path information preprocessing of hierarchical geographic traffic network

Since the user demand information of the hierarchical geographic traffic network is data with semi-structure characteristics, it is unbalanced. Before the data identification, the frequency of the occurrence of the feature words is used to represent the vector features in the access path information of the hierarchical geographic traffic network. The space vector model describes the user information of the access path network. At the same time, the TF×IDF function is used to calculate all the feature weights in the access path information, and the access path information preprocessing of the hierarchical geographic traffic network is realized.

The frequency of occurrence of feature words is used to represent the vector feature D in the access path information of the hierarchical geographic traffic network. The expression is:

$$D = (f_1, f_2, f_i, \dots, f_M), \sum_i f_i = 1 \quad (10)$$

Assuming that l_i denotes the feature term in the access path information of hierarchical geographic traffic network; $\omega_i(D)$ denotes the weight of l_i in the information vector feature D , and uses the space vector model to represent the access path information of hierarchical geographic traffic network, the expression is as follows:

$$V(D) = (\langle l_1, \omega_1(D) \rangle, L, \langle l_i, \omega_i(D) \rangle, L, \langle l_m, \omega_m(D) \rangle) \quad (11)$$

Supposing that $tf_i(D)$ represents the frequency at which l_i appears in the information vector feature D . The expression is as follows:

$$\omega_i(D) = \varphi(tf_i(D)) \quad (12)$$

According to the above calculation, the TF×IDF function is used to calculate all the feature weights in the access path information of the hierarchical geographic traffic network [31], and the access path information preprocessing of the hierarchical geographic traffic network is realized. The expression is as follows:

$$\varphi = (\sqrt{tf_i(D)} \log(N / m_i + 0.01) / \sqrt{\sum_{i=1}^m tf_i(D)^2 \log^2(N / m_i + 0.01)}) \quad (13)$$

Where N is the total number of information contained in the access path information of hierarchical geographic traffic network; m represents the dimension of the vector feature D in the access path information of hierarchical geographic traffic network; $\omega_i(D)$ can reflect the ability of the information vector's feature term item l_i to distinguish information attributes; if $\omega_i(D)$ appears in the access path information, the wider the range is, the lower the ability of $\omega_i(D)$ to distinguish the access path information set does; in addition, the higher the frequency at which $\omega_i(D)$ appears within a certain interval of the network access path information is, the higher the ability

of $\omega_i(D)$ to distinguish the attributes of the access path information of hierarchical geographic traffic network does.

Through the above steps, the basis for the accurate identification of the access path information of the hierarchical geographic traffic network is provided.

2.5 Identification method of access path information

For the big data to construct the hierarchical geographic traffic network router, it can judge whether to identify the current data packet according to p_1 , and judge the domain identifier or the route identifier through p_2 . The detailed identification process is as follows:

(1) In the case that $d = 0$, $x_2 < p_2$, and the router is a border router, the first domain identifier and the processing of the domain number of the router are completed to obtain $Hash_asn$. $Hash_asn$ is added to the $hash_asn$ domain. The autonomous domain number of the router is set to as_edge , and the domain value is added to 1. After completing the access path information identifier, the data packet is forwarded.

(2) If $d = 1$, $x_2 < p_2$, $f_1 = 0$, and the router is a border router, the second domain identifier is completed; if the router is a border router and the f_2 value is 1, then the router is located. The processing of the autonomous domain number is carried out, to obtain $hash_asn$. The autonomous domain number and the identification domain of the router are added to as_edge , the domain value of d_1 is increased by 1, and the access path information identifier is completed.

The pre-processed access path information training samples of hierarchical geographic traffic network are input into the support vector machine classifier for training [32-35],

and the linear classification problem of information is transformed into a quadratic regression problem, and the support of the access path information is derived. The vector linear discriminant function is used to optimize the discriminant parameters of the support vector machine classifier by using the co-evolution particle swarm optimization algorithm to realize the access path information identification of the hierarchical geographic traffic network.

Assuming that (x_i, y_i) represents the sample of network access path information, and an optimal classification plane (w, b) of information sample is obtained by the support vector machine classifier, which satisfies the following conditions:

$$w \cdot x + b = 0 \quad (14)$$

Where w and b represent the normal vector and the offset vector of the optimal hyperplane of the support vector machine classifier, respectively.

At the same time, the following conditions should be met in the sample of the access path information training set of hierarchical geographic traffic network:

$$y_i(w \cdot x + b) - 1 \geq 0 \quad (15)$$

The linear classification problem of access path information of hierarchical geographic traffic network is transformed into a quadratic regression problem. The expression is as follows:

$$\begin{aligned} & \min\left(\frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^n \xi_i\right)\right) \\ & s.t. \begin{cases} y_i(w \cdot x + b) \geq 1 + \xi_i \\ \xi_i \geq 0, i = 1, 2, \dots, n \end{cases} \end{aligned} \quad (16)$$

Where ξ_i is the slack variable; C is the penalty factor of the support vector machine

classifier.

According to the above calculation, the support vector linear discriminant function of the access path information of hierarchical geographic traffic network can be derived.

The expression is as follows:

$$F(x) = \text{sgn}(\sum_{i=1}^n a_i y_i x_i + b) \quad (17)$$

Where, a_i represents the Lagrange multiplier. In the access path information of hierarchical geographic traffic network, a sample corresponding to a_i that is not 0 becomes a support vector. For the nonlinear classification problem of access path information of hierarchical geographic traffic network, by introducing the kernel function $k(x_i, x_i)$, the above equation (15) can be transformed into the following form:

$$F(x) = \text{sgn}(\sum_{i=1}^n a_i y_i k(x_i, x_i) + b) \quad (18)$$

Using the radial basis kernel function, the calculation equation for kernel function $k(x_i, x_i)$ is:

$$k(x_i, x_i) = \exp(-\frac{\|x_i - y_i\|^2}{2\sigma^2}), i = 1, 2, L, n \quad (19)$$

Where σ is the kernel function parameter.

In summary, in the process of constructing the support vector machine classifier for access path information of hierarchical geographic traffic network, particle swarm optimization is needed to optimize parameter 2 and penalty factor C . The optimized identification results of the access path information of hierarchical geographic transportation network are as follows:

$$v_{id}(\kappa+1) = wv_{id}(\kappa) + h_1r_1(pBest_i(\kappa) - x_{id}(\kappa)) + h_2r_2(gBest_i(\kappa) - x_{id}(\kappa)) \quad (20)$$

$$x_{id}(\kappa+1) = x_{id}(\kappa) + v_{id}(\kappa+1) \quad (21)$$

Where, h_1 and h_2 represent the acceleration factors of the feature particles of the access path information in hierarchical geographic traffic network; r_1 and r_2 represent random numbers. In summary, the fast identification of the access path information of the hierarchical geographic transportation network is completed.

3. Results

3.1 Experimental design

The experiment simulates the identification router that can complete the dynamic identification through NS2, and designs an experimental environment based on dynamic identification. The designed experimental environment is described in Figure 2. The router capable of realizing dynamic identification is installed in the actual hierarchical geographic traffic network. The computer 1, the computer 2 and the computer 3 are the delivery test machines. Because the simulation environment is constrained by the space and the device, the network is replaced by the router, and the current experimental environment will be identified. Probability disposition is set to 20%.

Figure 2. Topology of hierarchical geographic traffic network

3.2 Technical identification process and results

Experiment 1: In the case where the visitor is the computer 1 and the accessed one is the computer 3, the visitor transmits an access request to the accessed node, and the access path information identified in the IP header is identified by the interviewee, and the identification flow is described in Table 2.

Table 2. Identification process record results in Experiment 1

Identification router	Data packet							
	1	2	3	4	5	6	7	8
a1			a1				a1	
a2		<u>a2</u>	#	a2	<u>a2</u>		#	a2
a5		#	#	#	#	a5	#	#
Computer 3		a2	a1	a2	a2	a5	a1	a2

In Table 2, the computer 3 represents the collection result of the identification access path information by the accessed end, # represents that the data packet is no longer identified, and a2 represents the upper routing access path information identified by the corresponding data packet when the identification probability of the router a2 is 1. After the identification is completed, the router converts the identification probability to 20%, and a2 represents the routing information identified by the packet with a probability of 20%, and the identifier is 1. The identification information collected at the computer 3 is described in Table 3.

Table 3. Collection results of identification information in Experiment 1

Data packet	Distance domain	edge1	edge2	edge3	edge4
2	3	196	170	0	260
3	1	196	170	100	8

4	2	196	170	0	260
5	3	196	170	0	260
6	3	220	252	2	6
7	2	196	170	100	8
8	1	196	170	0	260

Through the identified access path information, the subordinate access path can be reconstructed: 192.168.92.3→192.168.0.100→220.231.1.2→computer 3, and that the path is accurate is verified.

Experiment 2: In the case where the visitor is the computer 2 and the visited one is the computer 3, the visitor transmits an access request to the accessed node. The access path information of hierarchical geographic transportation network that has been identified by the packet header is collected at the accessed node, and the identification process is described in Table 4.

Table 4. Record results of the identification process in Experiment 2

Identification router	Data packet							
	1	2	3	4	5	6	7	8
a3	<u>a3</u>					<u>a3</u>		
a4	#	a4			<u>a4</u>	#	<u>a4</u>	
a5	#	#	c5		#	#	#	a5
Computer 3	a3	a4	a5		a4	a3	a4	a5

The identification result of the access path information collected at the computer 3 is described in Table 5.

Table 5. Identification information collection results in Experiment 2

Data packet	Distance domain	edge1	edge2	edge3	edge4
1	2	20	0	3	15

2	2	160	220	2	15
3	3	230	250	2	6
5	2	160	220	2	15
6	3	20	0	3	15
7	2	160	220	2	15
8	1	220	250	2	6

Through the identified access path information, the subordinate access path can be reconstructed: 10.0.2.15→160.211.1.11→220.231.1.2→computer 3, and that the path is accurate is verified.

3.3 Identification Technique Analysis

3.3.1 Analysis of technical convergence of the proposed technique

The convergence analysis is a key performance indicator for evaluating the identification technique of access path information. The convergence of this section is the number of data packets required to reconstruct the access path based on the completion of the access path information identification. In the probabilistic packet identification process, after the boundary information is divided into k slices when the data packet is identified, the expected range of data packets required to reconstruct the access path can be described as follows:

$$\frac{k}{p(1-p)^{d-1}} \leq E(x) < \frac{k \ln kd + \varepsilon}{p(1-p)^{d-1}} \approx \frac{k \ln kd}{p(1-p)^{d-1}} \quad (22)$$

In the basic packet identification process, the data packet is divided into 8 pieces, that is, at least 8 valid data packets are needed to reconstruct the access path information, so the expected value of the data packet required for reconstructing the access path can

be described as:

$$E(x) = \frac{8 \ln 8d}{p(1-p)^{d-1}} \quad (23)$$

According to the Coupon collector theory, it can be seen that the number of packets required to reconstruct access path information is the smallest when all packets are identified with a probability of $1/d$. In the process of identifying a data packet by dynamic probability identification, the number of packets required by the accessed node to reconstruct the access path is the least, and the expected value is described by $d \times \ln d$. In the process in which the packet side information is divided into k slices during the identification process, the expected value of the data packet required to reconstruct the access path can be described as:

$$kd \leq E(x) \leq kd \times \ln kd \quad (24)$$

If the probability packet marking scheme is identified by dynamic probability identification, the expected packet value of the reconstructed access path can be obtained by:

$$E(x) = 8d \times \ln 8d \quad (25)$$

After the dynamic probability identification technique is adopted, the number of data packet fragments is reduced to four, so the number of data packets required for reconstructing the access path is significantly increased compared with the basic probability identification method, and the expected values are as follows:

$$E(x) = 4d \ln 4d \quad (26)$$

It shows that the proposed technique has high convergence.

3.3.2 Comparison with other methods

Using the simulation experiment, the technical convergence performance of the proposed technique, the technique of literature [7] and literature [8] are compared. An access path is randomly selected in the topology map of the hierarchical geographic traffic network, because the number of hops that the data packet passes in the hierarchical geographic traffic network is substantially less than 25, the path lengths are sequentially selected to be 1, 2, ..., 30. Since the access path information identifier is a probability event, the results obtained in each experiment are largely different, so multiple experiments are required, and the convergence comparison results of the three techniques are described in Figure 3.

Figure 3. Comparison of the convergence of the three techniques

It can be seen from the analysis of Figure 3 that compared with the technique of literature [7] and the literature [8], the convergence of the proposed technique is obviously better, indicating that the technique in this paper not only has accurate identification, but also has good performance.

In order to further verify the accuracy of the method in this paper, the proposed technique is compared with the technique of literature [8] and literature [9].

Figure 4. Comparison of information identification accuracy under different methods

It can be seen from Fig. 4 that the identification accuracy of the access path information of the hierarchical geographic traffic network by using the technique of literature [8] and literature [9] is significantly lower than the information identification accuracy under the method of this paper. The average information identification accuracy under this method is 92. %, and that of the literature [8] and the literature [9] is about 30% and 45% respectively, indicating that the proposed method is obviously different from the other two methods in the identification accuracy of the access information of the hierarchical geographic traffic network

4. Discussion

The urban transportation system is an important carrier for people to carry out daily traffic activities, and is closely related to people's production and life. In recent years, with the gradual enhancement of China's comprehensive national strength and the accelerating industrialization process, the current economic development has shown a steady and positive trend, and people's living standards and quality of life have also been greatly improved. In order to pursue the convenience and comfort of travel, more

and more people began to buy cars. In recent years, the number of motor vehicles in China has gradually increased. While motor vehicles bring people a convenient and fast travel experience, the large increase in the number of vehicles has also brought severe challenges to the road network traffic in many large cities. The continuous increase of road traffic has intensified the contradiction between supply and demand of urban road network, and it is easy to induce traffic congestion on road network. The accurate identification method of access path information for traffic network can effectively alleviate traffic pressure. In addition, the identification of access path information of traffic network is of great significance for the normal operation of the network and traffic safety.

The existing identification method of access path information has a large calculation amount or a low identification efficiency, and cannot be well applied to an actual road network. In the context of big data, this paper establishes a hierarchical geographic traffic network router, judges whether to identify the current data packets according to p_1 , and judges the domain identification or routing identifier through p_2 , so that completing the accurate identification of the access path information of traffic network .

5. Conclusions

In this paper, a new big data construction technique to accurately identify the access path information of hierarchical geographic traffic network is proposed. Under the construction of big data, the access path information of the hierarchical geographic traffic network is identified by the dynamic probability identification method. The experimental results show that the proposed technique has accurate results and strong

convergence.

Acknowledgements

This research was supported by the Natural Science Foundation of Fujian Province (No. 2017J01405) of China; the Educational Department of Fujian Province (No. JAT170366) of China and the Educational Department of Fujian Province (No. JAS170258) of China.

References

- [1] Xu, X.S.; Yan, S.J. The optimization of iron and steel enterprises energy management under the background of big data. *Industrial Technique & Economy* 2017, 36(1): 32-40.
- [2] Hao, S.Y. Path selection of personal information protection in big data era. *Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition)* 2016; 18(5): 14-20.
- [3] Wang, X.C.; Bao, F.; Wang, S.T. et al. Transfer learning classification algorithms based on minimax probability machine. *CAAI Transactions on Intelligent Systems* 2016, 11(1): 84-92.
- [4] Han, X.G.; Yao, X.X.; Qu, W. et al. Role-based access control of user and object in clouding computing environment. *Science Technique and Engineering* 2014, 14(29): 229-233.
- [5] Zhai, J.Q.; Tang, Y.X.; Ye, F. et al. Implementation of DDoSattack path labeling

using divide-and-conquer strategy. Journal of Harbin University of Science and Technique 2014, 19(5): 76-82.

[6] Prawin, J.; Rao, A. R. M.; Lakshmi, K. Nonlinear parametric identification strategy combining reverse path and hybrid dynamic quantum particle swarm optimization. Nonlinear Dynamics 2015, 84(2): 1-19.

[7] Han, Z.M.; Wang, X.Z.; Huang, C.Q. et al. The situation semantic identification and applications of the learning path in the learning cloud space. Modern Educational Technique 2017, 27(11): 120-126.

[8] Wang, S.; Kang, J.S.; Zhong, Z.M. et al. Overviews of Torque Ripple Suppression Method with Per-manent Magnet Synchronous Motor Used in Electric Vehicles. Journal of Power Supply 2016, 14(5):24-32.

[9] Zheng, L.; Su, W.; Gao, S. Concurrent multipath transfer based on connection identifier. Computer technique and Development 2017, 27(11): 24-27.

[10] Zhang, Y.; Shum, K. W.; Wong, W. S. et al. Binary sequences for multiple access collision channel: identification and synchronization. IEEE Transactions on Communications 2014, 62(2): 667—675.

[11] Hu, J.H.; Qin, Z.C.; Shi, L. et al. Research on Spatial Information Cloud Service Platform and Application. Journal of China Academy of Electronics and Information Technique 2016, 11(1):51-58.

[12] Hu, X.Y. Cloud storage data dynamic access control method research and simulation. Computer Simulation 2017, 34(3): 365-368.

[13] Shen, X.L.; Shen, J. Optimization method of tracing distributed denial of service

attacks based on autonomous system and dynamic probabilistic packet-marking.

Journal of Computer Applications 2015, 35(6): 1705-1709.

[14] Zhao, Y.L. Research on GUI automated testing in the accounting information system test. Automation & Instrumentation 2016, 23(4):35-36.

[15] Zhou, X; Li, M; Chen Z; et al. An efficient probabilistic packet marking node localization algorithm based on layers-mixed in WSNs. Journal of Electronics & Information Technique 2014, 36(2): 384-389.

[16] Mshiu, E. E.; Gläßer, C.; Borg, G. Identification of hydrothermal paleofluid pathways, the pathfinders in the exploration of mineral deposits: A case study from the Sukumaland Greenstone Belt, Lake Victoria Gold Field, Tanzania. Advances in Space Research 2015, 55(4): 1117-1133.

[17] Feng, B; Guo, F; Tan, S. Attack source traceback scheme based on probabilistic packet marking for IPv6; network. Computer Engineering and Applications 2016, 52(6): 102-106.

[18] Hu, L.; Boos, K. S.; Ye, M. et al. Analysis of the endogenous human serum peptides by on-line extraction with restricted-access material and HPLC-MS/MS identification. Talanta 2014, 127(127): 191-195.

[19] Sun, P.; Wu, Q. RDDoSattack source-tracing technique based on self-adaptive packet-marking. Communications Technique 2015, 48(4): 478-483.

[20] Li, Q.X. Database access path optimization control embedded weight vector code. Bulletin of Science and Technique 2015, 31(10): 88-90.

[21] Zhang, J.; Xia, J.Z.; He, J.B. et al. A IP traceback scheme for direct broadcast-based

DRDoSattack. Journal of Chinese Computer Systems 2014, 35(10): 2313-2318.

[22] Wang, Y.; Xie, W.; Yu, X. et al. An automatic physical access control system based on hand vein biometric identification. IEEE Transactions on Consumer Electronics 2015, 61(3): 320-327.

[23] Jiang, C.; Chao, C.; Huang, Z. et al. ITIS, a bioinformatics tool for accurate identification of transposon insertion sites using next-generation sequencing data. BMC Bioinformatics 2015, 16(1):72-80.

[24] Hu, X.; Feng, Z.; Chen, S. et al. Accurate Identification of Ontology Alignments at Different Granularity Levels. IEEE Access 2017, 5(99):105-120.

[25] Wang, S.; Liu, K. Accurate identification of saturated pixels for high dynamic range measurement. Optical Engineering 2015, 54(4):43-106.

[26] Zhou, Y.; Zhao, Z.; Louët, Y. et al. Large-Scale Spatial Distribution Identification of Base Stations in Cellular Networks. IEEE Access 2016, 3(23):2987-2999.

[27] Roxburgh, A.; Pilgrim, J. L.; Hall, W. D. et al. Accurate identification of opioid overdose deaths using coronial data. Forensic Science International 2018, 287(23):40-46.

[28] Wang, C.; He, L.; Li, D. et al. Accurate Identification of Unknown and Known Metabolic Mixture Components by Combining 3D NMR with FT-ICR MS/MS. Journal of Proteome Research 2017, 16(10):3774.

[29] Goenka, A.; Jeena, P. M.; Mlisana, K. et al. Rapid Accurate Identification of Tuberculous Meningitis Among South African Children Using a Novel Clinical Decision Tool. Pediatric Infectious Disease Journal 2017, 37(3):229-350.

- [30] Dong, X.; Zhang, L.; Milholland, B. et al. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nature Methods* 2017, 14(5):491-493.
- [31] Martinucci, M.; Roscetto, E.; Iula, V. D. et al. Accurate identification of members of the *Burkholderia cepacia* complex in cystic fibrosis sputum. *Letters in Applied Microbiology* 2016, 62(3):221-229.
- [32] Bhavani, R.R. and Jiji, G.W. Image Registration for Varicose Ulcer Classification Using Knn Classifier. *International Journal of Computers and Applications* 2018, 40(2):88-97.
- [33] Shao, M.; Ma, J.; Wang, S. Deep Bound: accurate identification of transcript boundaries via deep convolutional neural fields. *Bioinformatics* 2017, 33(14):267-273.
- [34] Abusitta, A., Bellaiche, M. and Dagenais, M. An Svm-Based Framework for Detecting Dos Attacks in Virtualized Clouds Under Changing Environment. *Journal of Cloud Computing*, 2018, 7(1).
- [35] Banerjee, A. and Bej, S. On Extension of Regular Graphs. *Journal of Discrete Mathematical Sciences and Cryptography* 2018, 21(1):13-21.