Routing Strategy of Complex Networks Based on Topology Information

Min-sheng TAN, Chang CAI*, Lin-feng ZHU, Lin DING and Shi-ying XIA
School of Computer Science & Technology, University of South China, Hengyang, Hunan, 421001, China
*Corresponding author

Keywords: Complex networks, Routing strategy, Effective path, Local visibility and Detecting communities.

Abstract. A routing strategy using nodes’ effective path and dynamic information weight distribution was proposed, which based on the improved efficient path routing strategy in complex network. The strategy’s weight factor was optimized by the new strategy. Considering the insufficient of local routing strategies in global performance and the community division technology in complex network, a kind of local visibility routing strategy according to the thought of community division was presented. On the condition of community division of the Bgll condensation algorithm, the topology structure information of neighbor nodes and all the community nodes were only stored in each node’s routing list in the new strategy. When node transaction in community, global routing strategy was adopted and when between communities, minimum load strategy will be used. The simulation results indicate that in the different connection density networks, the performance of the new strategy achieves the best when community division’s degree of modularization is the highest. Contrasted to other local routing strategy, the optimized strategy holds incomparable superiority on throughput, transmission delay and the packet loss.

Introduction

Since Watts and Storgazt found the small world characteristics, Barabasi and Albert put forward the scale-free complex network model, the research of the complex network became a new hotspot in fields like physics, control, computer science, etc. Internet, the power networks, aviation transport network and others could be abstracted as complex networks for research, and the transmission capability of these networks are closely related to our life. Therefore, the throughput of complex networks has been paid more and more attention by researchers [1,2,3]. The transmission capability is affected by two main types of factor [4], one is the structural factors, such as the connection way of the network, the processing power of nodes, the bandwidth of the connection, etc. The other is routing strategy factor [5,6], a good routing scheme can greatly improve the network throughput. It is hard to alter the network structure, for instance, if add a connection between two routers, we need to establish a communication line, which would be very expensive and not real. And it's quite feasible to use good routing scheme to improve the transmission capability of the network. So this research focuses on routing strategy in network throughput.

In the past, a number of researches were about random and regular network routing strategy, they drew some valuable conclusions. Recently, the routing strategy of complex network started to gain attentions. Some good routing schemes were proposed by researchers [7,8,9], divided into two kinds. 1) Global routing strategies, most of them were improvement based on traditional shortest path routing strategies. The literature [10,11] proposed a routing strategy based on nodes' valid path has great breakthrough in performance, which is more than ten times than the former ones; 2)Routing strategies based on local node information, a lot of local routing strategies developed from the strategy. The literature [12] proposed a routing strategy according to the neighbor nodes' degree probability distribution, the strategy only needs to know the degree information of neighbor nodes, distributes data by probability of nodes’ degree.
Routing Measure

Under different routing strategies, the network has different throughput. Measured by packets transmitted in every step $R$, what we concern about is the critical value of a network's packets transmission amount $R_c$, when $R < R_c$, the number of produced packets and transmitted packets are basic the same, which means they are in equilibrium. When $R > R_c$, the congestion starts to emerge, the produced packets could not be sent to the destination in time entirely. Generally, we use the order parameter to describe the kind of state in the network proposed by the literature (Pablo et al., 2005).

$$H(R) = \lim_{t \to \infty} \frac{C < \Delta N_p >}{\Delta t}$$

In the formula, $C$ is a constant, $\Delta N_p = N(t+\Delta t)-N(t)$, $\Delta N_p$ is the change of the amount of packets in the network in a span $\Delta t$; $< >$ is the average calculated through choosing enough spans. $H(R)$, is the change rate of the packets amount in the network when $R < R_c$, $H(R)$ is less than or equal to zero, when $R > R_c$, $H(R)$ equal to zero is the throughput of the network $R_c$. The calculation method of average network transmission delay is as the formula (2) shows.

$$< T > = \frac{\sum_{x \in \Omega_n} T(x)}{|\Omega_n|}$$

$\Omega_n$ displays the set of packets that arrived at the destination in the network in enough long time, $|\Omega_n|$ represents the amount of elements in the set $\Omega_n$, $T(x)$ is the time step that the packet $x$ spends in reaching the destination.

The average relative routing table size $<M> = \langle$routing table size$>/N$, $N$ represents the number of nodes in the network. The packet loss rate $Loss(R)=$packets discarded the amount packets transmitted. We define the average packet loss rate. Formula (3) used to quantify the loss rate under the circumstance that the transmission rate is $R$.

$$<\text{loss}> = \left\| \left\{ \left[ R \mid R \leq R_c \right] \right\} \right\|$$

An Improved Routing Strategy Based on Nodes Effective Path

Strategy description. Considering the more connections a node has, the more probability the node could generate congestion, the nodes' effective path concept based on nodes' degree was put forward by domestic scholars Bing-Hong Wang etc (Wang et al., 2006). The distance $d(i \rightarrow j): \beta$ of the effective path $p[i \rightarrow j]=i, x_i, x_{i+1}, \ldots, x_n, j$ between any node $i$ and $j$ could be calculated as the formula (4).

$$d(i \rightarrow j): \beta = \min_{i \in \Omega_n} \sum_{x \in \Omega_n} k(x)^\beta$$

$k(x_i)$ represents the degree of the node $x_i$, $\beta$ is an adjustable parameter. This strategy considers the nodes' degree when calculating the shortest path, thus could avoid letting the shortest paths go through the nodes with largest degree too much, lessen the possibility those nodes congest. When selecting a route, the node still chooses the closest node from the neighbor to transmit. The paper proved that when $\beta = 1$, the network has largest throughput by running simulations, the performance is more than ten times as traditional shortest path strategy. The routing strategy proposed by this paper considers the neighbour nodes' real-time queue length (nodes' dynamic information) in route selecting on the basis of the above strategy, when the node $i$ transmit packets to the destination node $n$, the method calculating the weight $w_j$ of the neighbour node $j$ is shown as formula (5).
The Local Visibility Routing Strategy Based on the Thought of Community Dividing

Considering the performance represented by the general local routing strategy based on neighbor nodes' information in the complex network is very bad, while reducing node storage topology information can effectively improve the local visibility strategy of routing performance nodes, at the same time, the community dividing technology is relatively mature could strongly support the network regional division, this paper puts forward a local visibility routing strategy based on the community dividing thought.

The community dividing based on Bgll cohering algorithm. The routing strategy based on nodes' local visibility need to divide the network firstly when routing packets, a large number of research achievements made by scientists in complex network community dividing support the network dividing in technology. Currently there are many mature communities dividing algorithm, in the new routing strategy put forward by this paper, we choose Bgll algorithm proposed by foreign scholar Blondel and Guillaume to divide the network. Bgll algorithm is a typical cohering algorithm, (Barabasi et al.,1999) which could detect hierarchical community structure. The detailed description of algorithm is given in the literature (Blondel et al.,2008). algorithm description.

The first iteration: network’s each node is allocated a community number, when any node i enters the community of its every neighbor node j, the increment ΔQ of the corresponding community modularity.

\[ ΔQ = \sum \left( \frac{k_i}{2m} - \left( \frac{\sum k_i}{2m} \right) \right) \left( \frac{\sum k_i}{2m} - \frac{\sum k_i}{2m} \right) \]  

(6)

In the formula, \( k_i \) is the sum weight of all edges corresponding to the point i; \( \sum \) is the sum weight of all edges inside the community; \( \frac{k_i}{2m} \) is the sum weight of all edges corresponding to points inside the community; \( \frac{\sum k_i}{2m} \) is the sum weight of all edges connecting to the point i and the community. If there is a positive ΔQ, add the point i into the community where the neighbor node with maximum corresponding value locates in. If all ΔQ are negative, then the node i stay in the initial community. The merging is repeated until the merging would not happen in the network.

The second iteration: firstly, the algorithm constructs a new network, the node of the network is the community detected in the first phase, and the weight of edges between nodes is the sum weight of all edges between two communities. Afterwards, the algorithm divides the new network again by using

\[ w_j = h \cdot d(p(j \rightarrow n) \cdot 1) + (1-h) \cdot l_j \]

represents the distance of the effective path between the node j and n when \( \beta = 1 \). h is an adjustable weight factor. \( l_j \) represents the queue length of the node j at the moment.

Once selecting a route, the node chooses the neighbor node with minimum \( w_j \). This routing strategy consider the real-time queue information of neighbor nodes while considering nodes' effective path, which is a dynamic routing strategy. It could avoid transmitting packets to a node when the node is in the congestion state, at the same time, it could improve the throughput of the network greatly.

Routing algorithm. The algorithm (Jeremie et al.,2006) preprocesses the global routing table before the strategy executed--calculating the effective path length using Floyd algorithm. The strategy firstly seeks the node ts where the packet t locates and does some simple route selecting, then get all the neighbor nodes’ comprehensive weight \( w_j \) and the neighbor node s with the minimum \( w_j \) by calculating the comprehensive weight \( w_j \) using global routing table. Input: the packet number t, destination node number n, the packet list list (F*R,3) (F represents the simulation steps, R is the global data transmission rate in the network), global routing table node (N,N) (N represents the amount of nodes),the adjacency matrix of BA network nodes (N,N), the real-time queue length list of nodes Packet (N,1), weight factor h. Output : the node numbers of the packet's next hop node. The time complexity of the algorithm: O(n3).
the first iteration algorithm. The algorithm repeats the second phase until no more advanced community cohered. \textit{Bgl algorithm} itself can generate a hierarchical community structure.

**Strategy description.** The routing strategy put forward by this paper is a sort of local visibility routing strategy based on the community dividing thought, the premise of its implementation is as follows.

- Any node could know the corresponding information of all neighbor nodes in real-time.
- For any packet needs to be transmitted, the destination node item contains the community number where the destination node locates. The strategy is described as below.

When selecting a route, the node i firstly searches its neighbor node table, if the destination node j is the neighbor of the node i, then transmits directly, otherwise executes following steps.

**Step 1:** Look up the community number where the destination node locates, if the community number is the same as the community number of the node i (packets are transmitted in a community), then we choose a global routing strategy to transmit, otherwise, the algorithm executes the next step.

**Step 2:** Iterate all the neighbor nodes, gets all nodes that the community number equals to the destination node's community number, transfers the nodes with the least load, or else executes next step.

**Step 3:** If there exist nodes with different community number from the destination node, at the same time, these nodes do not belong to the community where the transmitting node locates, chooses the node with least load from them to transmit, otherwise, the algorithm executes the next step.

**Step 4:** Choose the node with least load from neighbor nodes to transmit.

**Routing algorithm:** The local routing table of every node would be preprocessed before the strategy executed based on the result of the community dividing, the global routing table inside every community will be got by using Floyd algorithm (Hughes et al., 1995). When executing the strategy, the algorithm (Conan et al., 2008) firstly searches the node ts where the packet t is locating, dose simple route selecting according to the destination node n, if the destination node is not among the neighbor nodes of ts, then executes the strategy as following situations.

- Look up the community number cn of the destination node n, if cts=cn, packets are transmitted inside the community, choosing a global routing strategy to transmit (shortest path routing strategy).
- Traverse all the neighbor nodes, gets the node set \( \Omega = \{ k / k \in N, \& c_k = c_n \} \), \( c_k \) is the neighbor node set of nodes i, \( i \) represents the community number corresponding to the node k, if \( \Omega \) is not empty, chooses the node m with least load (the real-time queue length) to transmit.
- Traverse all the neighbor nodes and gets the node set \( \Omega = \{ k / k \in N, \& c_k \neq c_n \& c_k \neq c_i \} \), if \( \Omega \) is not empty, then chooses the node \( m' \) with least load to transmit.
- Choose the node m with least load to transmit.

Input: the packet number t, the destination node number n, the packet list list (F*R,3), the community structure mapping table community (N), the local routing table local (N,N), the adjacency matrix of BA network Nodes (N,N), the nodes' real-time queue length list Packet (N). Output: the next hop node number s of the packet. The time complexity of the algorithm (Hughes et al.,1996) is O(n3).

**Experiments and the Analysis of the Results**

To validate the two routing strategies is effective, we use \textit{Matlab} to do simulation experiments. This paper uses the BA model when constructing the network model, lets \( m = m_0 = 5 \), \( N = 1000 \) and \( m = m_0 = 2 \), \( N = 1000 \) respectively, figure 1 shows the results of simulations.
The Strategy Improving the Effective Path Routing Strategy

To validate the effectiveness and the superiority of the strategy proposed by this paper, we run simulations on the traditional shortest path routing algorithm, the unimproved strategy based on the node's effective path and the improved routing strategy proposed by this paper in the same network model respectively, and compare their performance. Figure 2 describes the parameter $H(R)$ of the shortest path routing strategy corresponding to different data transmitting rate when $T = 10000$ in the scale-free network. It's easy to see that when $R <= 5$, $H(R) \approx 0$; When $R = 6$, the value of $H(R)$ has a mutation, afterwards, the gradient of $H(R)$ increases with the increase of $R$, so we could draw the conclusion that the measure parameter of the network throughput $R_c$ equals to 5 when using the shortest path routing strategy in this case. In the same way, figure 3 shows that the threshold transmission rate of the effective path routing $R_c = 95$ is far greater than the shortest path routing, we can also see that the curve's gradient is less than the one in the figure 2, that is to say, when the congestion emerges in the network, the congestion extent is lower than the one of the shortest path routing.

The purpose of proposing the new strategy is to improve the network throughput (the threshold data transmission rate $R_c$) in quality by improving the effective path routing strategy, therefore, we run simulations on the improved effective routing strategy in the same network topology model in this section. From figure 4, we can see that the weight factor values corresponding to the maximum threshold transmission rate are found in the middle area of all value range, when $h = 0.3, 0.4, 0.5$, $R_c$’s maximum is 160, larger than the $R_c$ corresponding to the unimproved routing strategy under the same circumstance.
To further comparing on the shortest path routing strategy, the effective path routing strategy and the improved effective path routing strategy which is optimized, the comparison of the average transmission delay of the three strategies in the free network state is given in the table. Ave1(T), Ave2(T), Ave3(T) corresponds to the shortest path routing, the effective path routing and the improved effective path routing strategy which is optimized respectively (h=0.3). We can see from the table 1 that, when R=5, the average transmission delay of the shortest path routing is slightly less than the delay of the other two. It shows that when there is no congestion in the network, the traditional shortest path routing strategy is still the most effective and fastest one, but when R>5, the congestion emerges when using the shortest path routing, the average transmission delay increases rapidly, the performance is far worse than the other two strategies. In this case, the improved one is better than the unimproved one, it shows that the improved strategy lessens the average transmission delay in the network effectively while improving the network throughput.

Table 1. The comparison of the average transmission delay <T> using the three routing strategies.

<table>
<thead>
<tr>
<th>R</th>
<th>5</th>
<th>10</th>
<th>30</th>
<th>50</th>
<th>70</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave1(T)</td>
<td>3.62924</td>
<td>296.916</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Ave2(T)</td>
<td>4.10757</td>
<td>4.15094</td>
<td>4.37971</td>
<td>4.67825</td>
<td>5.11994</td>
<td>5.95121</td>
</tr>
<tr>
<td>Ave3(T)</td>
<td>4.10494</td>
<td>4.14236</td>
<td>4.33864</td>
<td>4.59607</td>
<td>4.92005</td>
<td>5.37298</td>
</tr>
</tbody>
</table>

The figure 7 counts the nodes' load (nodes' queue length) of three strategies before and in the congestion state, the simulation time is 10000 steps. The figure a, b corresponds to the situation of R=5 (before the congestion) and R=10 (in the congestion) respectively using the shortest path routing strategy, and c,d corresponds to the nodes' load before and in the state of the congestion using the effective path routing and the improved effective path routing respectively. The nodes' load before and in the state of congestion is extremely uneven by using the shortest path routing. Most nodes' load do not increase, while the load of a few nodes increases for thousands of times (from 3 to 7000); Using the effective path routing, the nodes' load before and in the state of congestion is still uneven, but the gap between the maximum value of the load and the previous one decreases, from 150 to 700, only increases for only more than four times; Differing from the previous two strategies, most nodes' load increase to the range of from 70 to 80 from the range of from 10 to 20. The figures show that the
nodes’ load distribution of the strategy is more even than the unimproved one because of taking more consideration of nodes’ dynamic information when executing the improved routing strategy.

Figure 7. The nodes’ load before and after the congestion emerges using different routing strategies.

The result shows that the new strategy could improve the network throughput while decrease the average transmission delay and optimize the load balance. We found that when the weight factor is 0.3, the performance performed by the new strategy is the best by optimizing the weight factor of the strategy.

The Local Visibility Routing Strategy Based on the Community Dividing Thought

The result of the community division using Bgll cohering algorithm corresponding to two kinds of complex network is given in the table 2. Because the algorithm is a dividing algorithm based on the hierarchy, the division level increases with the increase of the modularity, when the increment of modularity equals to zero, the division is halted. Level represents the level of the community division, two networks are divided in four levels, the different is the community amount of the network with small M is more than the network with larger M, that is because the network with less connection density is with poor communication, it’s hard for the smaller communities to cohere to form a bigger community.

Table 2. The community division result of different networks M stands for the amount of communities, Q stands for the modularity.

<table>
<thead>
<tr>
<th>Level</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Q</td>
<td>M</td>
</tr>
<tr>
<td>m=2</td>
<td>31</td>
<td>0.51806</td>
</tr>
<tr>
<td>m=5</td>
<td>11</td>
<td>0.28928</td>
</tr>
</tbody>
</table>

Table 3. <M> corresponds to different community level in different networks.

<table>
<thead>
<tr>
<th>L &lt;M&gt;</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>m=2</td>
<td>0.0040</td>
<td>0.0047</td>
<td>0.0194</td>
<td>0.0194</td>
<td>0.0537</td>
</tr>
<tr>
<td>m=5</td>
<td>0.0099</td>
<td>0.0819</td>
<td>0.1102</td>
<td>0.1258</td>
<td>0.1274</td>
</tr>
</tbody>
</table>

The figure 8,9,10 and the table 3 describe the comparison result of the performance performed by the routing strategy based on the community division thought and other typical local routing strategies. Table 3 gives the average routing table size <M> which nodes need to store corresponding to different community division level. The division level L=0 means the average routing table size that nodes need to store corresponding to the general local routing strategy based on neighbor nodes' information. The routing table size corresponding to the global routing strategy is 1, the routing table size could be reduced to one order of magnitude at least using the new strategy (from 1 to 0.1), so the strategy could decrease the routing table size effectively. We can also find that, <M> increases with the increase of the division level corresponding to different connection densities; <M> of the network with larger connection density is greater than the one with smaller density when the level is identical.
The comparison results of the threshold data transmission rate, the average data transmission delay and the loss rate using the shortest path routing, the effective path routing based on nodes and the improved effective routing strategy in the community are described in the figure 8, 9, 10. Three global shortest path strategy, the unimproved effective path routing strategy and the new strategy, numerous propounded according to the community dividing technology. Secondly by comparing the traditional a view to improve the existing effective path routing strategies and a local visibility routing strategy is identical on the whole when the division level of the strategies increases (except the level 0 and 5). In general, the overall performance of the first level is the worst, when the level is up to level 4, the overall performance is the best, especially the loss rate, the loss rate is close to 0 when the division level is 4 using all strategies, and the other three local routing strategies are much higher. Clearly, when the community division level is the highest, namely the modularity is the highest, the performance of the new strategy on the threshold data transmission rate, the average transmission delay, the loss rate is better than other local routing strategies, it’s the same as we anticipate.

**Conclusion and Further Research**

In this thesis firstly a new global routing strategy based on the nodes’ effective path is presented with a view to improve the existing effective path routing strategies and a local visibility routing strategy is propounded according to the community dividing technology. Secondly by comparing the traditional shortest path strategy, the unimproved effective path routing strategy and the new strategy, numerous
simulation experiments in Matlab are made to verify the effectiveness and the superiority of the new strategy. The results indicate that the new effective path strategy improves the throughput of the network while decrease the average transmission delay and optimize the load balance. Besides, the weight factor of the new strategy is optimized and it is proved that the performance achieves the best when the weight factor is 0.3. And the performances all reach the best state when the community division level is the highest in network topologies of different connection density using the local routing strategy based on the community dividing thought. Compared with other local routing strategies, it may draw a conclusion that the new strategy holds its incomparable advantages on throughput, transmission delay and packet loss.

Acknowledgement
This work was supported by Project 61403183 of the National Science Foundation of China, Project 2017JJ4048 of the Hunan Provincial Natural Science Foundation of China, Project 20183350502 of the Hunan Provincial Finance Office Science Research of China.

References