Adaptive Routing Forwarding Strategy Based on Neural Network Algorithm

Chao Ma, Mingchuan Zhang, Fenghua Zhang, Yaming Chen, Xinlu Wang, Qianpeng Li

Abstract—With the profound changes in global digital media, the focus of Internet users has gradually shifted to how to quickly obtain information without paying attention to where the information is stored. However, the current TCP/IP network protocol architecture cannot adapt to the rapid development of today's content applications. In order to adapt to the changes in the Internet, information-centric networking (ICN) has received extensive attention. Besides, the optimization of the user service request scheduling problem is the core issue affecting the performance of the ICN, and it is one of the hot research topics in the ICN network. To solve this problem, this paper proposes an adaptive routing forwarding strategy based on neural network algorithm. Through the modeling of the classic architecture named data networking (NDN) network delay model of ICN network, a neural network algorithm is used to delay prediction, and a forwarding strategy mechanism based on predict delay is designed to innovate in the NDN. The interface information Stat is added to the forwarding information base (FIB) of the network component to implement the dynamic selection of the forwarding routing. In addition, routing dynamic self-adaptation adjustment mechanism and fault rerouting function are designed in consideration of the situation of route congestion and interruption. Simulation results show that this strategy effectively reduces network delay and improves network performance.

Index Terms—ICN, Dynamic priority, Neural network, Time Delay Prediction, Forwarding strategy

I. INTRODUCTION

The initial purpose of the design of the Internet was to solve the problem of resource sharing. With the rapid increase in the amount of information on the Internet, people's sharing of the resources has gradually become the acquisition and distribution of the content. Users are more concerned with how to obtain the content itself than the location of the content. The “assumptions” of traditional Internet design can no longer meet the needs of new applications. Therefore, the new Information-centric Networking (ICN) [1-2] has already emerged. Different from the traditional Internet architecture for data transmission based on IP addresses, the ICN architecture focuses on the content or information that the users care about, and it is dedicated to designing the content as a thin waist of the future Internet architecture. In addition, the in-network caching[3] is an important feature of the ICN. Its core idea is to cache content through internal routers of the network. After that, the user requests routing in the network. If the cache is hit along the way, it can respond directly without having to get from the content source (server) each time. By the nearest service method, it can achieve the purpose of improving the content transmission efficiency, reducing network traffic, and alleviating server access pressure. However, the cache technology also results in a new problem, namely, how to coordinate the response of the routing node and the server to the user service request, to avoid duplication, inefficient response and so on. It can be said that the routing forwarding of service requests on the node is a core issue affecting the performance of the ICN network and it is also one of the hot research topics of the current ICN network.

In this paper, I take the network delay as a reference index, proposing an adaptive routing and forwarding strategy based on neural network algorithm. Firstly, on the named data networking (NDN) which is the classic architecture of ICN network build a network hybrid delay model. And using the time delay data generated by this model as the input layer data of neural network to train the wavelet neural network, then doing experiments on the performance of time delay prediction. At the same time, the NNDP routing and forwarding strategy is designed by using the predicted delay as a guide. In addition, taking into account the situation of route congestion and interruption, the dynamic self-adaptive routing adjustment mechanism and fault rerouting are designed, which greatly improves the satisfaction rate of service requests. Simulation results show that the proposed strategy can reduce the average network latency and improve the network performance.

II. THE OVERVIEW OF NDN

In the NDN network, communication is dominated by the recipient and is communicated by exchanging interest packets and data packets. The identification of interest packages and data packages is achieved through special naming rules. The user sends an interest packet with a data name to the network and the router forwards the interest packet to the packet by identifying the name. Once the interest packet reaches a node that wants a packet, the node
will return a packet containing the name and data so that the information for the two nodes can be bound. As shown in Figure 2.1[4], the router forwards packets based on the routing of the node to which the interest packet is transferred.

![Interest packet and data packet in NDN](image)

**Fig. 2.1 Interest packet and data packet in NDN**

In order to complete the forwarding of interest packets and data packets, the NDN network includes three components[5]: a forwarding information base (FIB), a content store (CS), and a pending interest table (PIT). In Figure 2.1, these three components, also called the neural network based delay prediction (NNDP) module, can decide whether, when and where to forward each interest packet.

![Forwarding model in NDN](image)

**Fig. 2.2 Forwarding model in NDN**

The Figure 2.2 describes the forwarding strategy of interest packets and data packets in an NDN network[6]. When the interest packet reaches a router, the router first detects whether its CS contains a packet that matches the name of the interest packet. If there is such a data packet, the data packet in the CS is directly sent back along the input interface of the interest packet; if there isn’t, it is detected in the PIT of the router whether the name of the interest packet is included. If the name of the interest packet already exists in the PIT, it indicates that the interest packet that has been previously owned by other consumers with the same name is received by the router and has already been forwarded from the routing node. Therefore, it is no need forwarding again, and then the router adds an input interface of the interest packet to the existing PIT entry. If there is the name of the interest package in the PIT, it will be added to the PIT. At the same time, the interest packet is further forwarded according to the FIB. When the router receives the returned data packet, it will use the packet’s name to find the PIT entry. If a named matching PIT entry is found, the router forwards the data packet to the interface of the interest packet recorded by the PIT and the cached data also deletes the PIT entry. Otherwise, the data packet will be discarded. Each interest packet is set to have a lifetime. When the lifetime expires, the PIT entry is automatically deleted.

### III. TIME DELAY PREDICTION BASED ON NEURAL NETWORK

#### A. Time Delay Modeling of the NDN Hybrid Networks

The change of the delay in the network is uncertain and it is affected by some factors, such as the number of users in the network and the link status. However, the delay every day is always regular. Therefore, in order to simulate the delay data in the real NDN networks better, this paper aims to propose the modeling of the NDN delay,

\[
D_b = k_1 \cdot D_1 + k_2 \cdot D_2 + k_3 \cdot D_3
\]  

In above four formulas, \(D_b\) is the initial value of the delay; \(t\) is the time; \(f_1\) is the function of the mutation; \(f_2\) is the function of the jump; \(b_1\) and \(b_2\) are the width of the function; \(T_1\) and \(T_2\) are the periods of the function; \(rand\) is a random function. According to \(t\), the parameters \(k_1-k_{10}\) change ensures the randomness of the delay variation.

The NDN network caching mechanism is considered by \(D_1\) and the delay will show a downward trend for a period of time. However, the number of network users will still be affected in the network and will not affect the fluctuation of the overall trend. \(D_2\) considers that the sudden increase in the number of users in the network leads to a sharp increase in delay, and in a short time the delay is normal again; \(D_3\) considers the existence of the cache mechanism in the NDN network, the data packet may suddenly change the cache location, resulting in a time-varying transition; \(D_b\), a hybrid model, mixes \(D_1, D_2\) and \(D_3\) randomly to train NDN network model.

#### B. The Basic Principle of Wavelet Neural Network

Artificial neural network is a mathematical model of distributed information processing by imitating the working characteristics of biological neural networks[7].

Neural networks are currently an algorithm model with very powerful predictive functions. A complete neural network consists of an input layer, a hidden layer, and an output layer, and each layer consists of multiple neurons. The following is the neural network algorithm to be used in this paper.

Wavelet analysis is a new field in applied mathematics and engineering disciplines. Compared with the Fourier transform, wavelet analysis can extract information from signals more effectively. The application of wavelet analysis is closely related to its theoretical research[7]. For signals that are stable over time, the ideal tool for signal processing is still Fourier analysis. However, the signals in practical...
applications are mostly non-stationary, and wavelet analysis can decompose unstable signals and then superimpose them by a series of wavelet functions. These wavelet functions are generated by a mother wavelet function through a series of translations and scaling. By using this irregular wavelet function, discrete signals with local characteristics can be arbitrarily approached to more truly reflect the change of the original signal. For example, the delay of the network system, it has a strong randomness, so it is certainly not a stable signal, and the tool that is very suitable for unstable signal analysis is wavelet analysis.

BP neural network is one of the most widely used neural network models. BP neural network has strong self-learning ability and can infinitely approach any unknown model without revealing the mathematical equation describing the mapping relationship.[8] Its self-learning rules use the steepest descent method. The weights and thresholds of each layer of the BP neural network are constantly adjusted by the magnitude of backpropagation errors, so that the squared error sum of the network is the minimum.

Wavelet neural network combines wavelet analysis and BP neural network. Based on the structure of BP neural network, the wavelet basis function is used as the neural network of the transfer function of hidden layer nodes in BP network [9]. The structure of the wavelet neural network is shown in Figure 3.1. In the figure, $x_1, x_2, \ldots, x_k$ is the input signal of the wavelet neural network and $Y_1, Y_2, \ldots, Y_m$ is the prediction output of the wavelet neural network. $a_{ij}, \omega_{jk}$ respectively are the connection weights from the input layer to the hidden layer and the hidden layer to the output layer of the wavelet neural network.

![Wavelet neural network structure diagram](image)

Assume that the input signal of wavelet neural network is $x_i$ ($i=1, 2, \ldots, K$), the number of input layer neurons is $k$, the number of hidden layer neurons is $l$, and the number of output layer neurons is $m$. Then the output of the hidden layer is:

$$h(j) = h_j \left( \sum_{i=1}^{k} w_{ij}x_i - b_j \right) / a_j \quad j = 1, 2, \ldots, l$$

In the formula, $h(j)$ is the output value of the $j$ node of the hidden layer of the wavelet neural network, and $h(j)$ is the wavelet basis function: $a_j$ and $b_j$ respectively are the translation factor and the expansion factor of the basis function.

The basis function used in this section is Morlet's mother wavelet basis function. The Morlet mother wavelet basis function is a type of wave function with fast attenuation and finite length. Its mathematical model is:

$$y = \cos(1.75x)e^{-x^2/2}$$

The output of the wavelet neural network output is:

$$y(k) = \sum_{j=1}^{l} \omega_{jk}h(j)$$

Similar to the BP neural network, the weight self-modifying algorithm of the wavelet neural network is modified by the steepest gradient descent method. The difference is that the wavelet neural network not only corrects the connection weights in the neural network online, but also corrects the basis functions. The parameters so that the predicted output value approaches the actual value faster. The wavelet neural network parameters and weights are corrected as follows:

1. Calculate prediction error

$$e(k) = y_i(k) - y(k)$$

In the formula, $y_i(k)$ is the actual output of the time delay model, and the $y(k)$ is the predicted output value.

2. Online correction of weights and coefficients of wavelet basis functions based on prediction errors:

$$\omega_{ij}^{k+1} = \omega_{ij}^{k} + \Delta \omega_{ij}^{k+1}$$

$$\omega_{jk}^{k+1} = \omega_{jk}^{k} + \Delta \omega_{jk}^{k+1}$$

In the formula $\Delta \omega_{ij}^{k+1}, \Delta \omega_{jk}^{k+1}$ is calculated from the steepest descent method of wavelet neural network error:

$$\Delta \omega_{ij}^{k+1} = -\eta \frac{\partial e}{\partial \omega_{ij}^{k}}$$

$$\Delta \omega_{jk}^{k+1} = -\eta \frac{\partial e}{\partial \omega_{jk}^{k}}$$

$\eta$ is the learning rate of the wavelet neural network.

The training steps of the wavelet neural network algorithm are as follows:

1. Initialization. Initialize the wavelet basis function parameters $a_j, b_j$, and the connection weights $\omega_{ij}, \omega_{jk}$, and set the learning rate $\eta$ of the wavelet neural network.

2. Sample classification. The network delay data generated by the delay model of the ICN hybrid network is divided into training samples and test samples. The delay training samples are used to train the wavelet neural network, and the delay test samples are used to test the prediction accuracy of the wavelet neural network.

3. Delay forecast. The delay training sample is used as the input value of the wavelet neural network and the error of the prediction delay and the real delay is calculated.

4. Correction of weights and parameters. According to the value of the error, the gradient method is used to modify the
weights of the wavelet neural network and the parameters of the wavelet basis function, so that the delay of the predicted value is closer to the actual delay value.

5. Determine whether the predicted delay error satisfies the given accuracy. If it is not satisfied, return to step 3 to continue training.

C. Wavelet neural network prediction based on delay model of NDN hybrid network

(1) Selection of training and speed samples

This section selects the sample data, in order to be able to truly restore the delay characteristics of the actual network system, according to the $D_h$ hybrid model, with reference to the size of the general network delay, resulting in 200 different time delay samples. Among them, 50% are training samples and 50% are test samples.

(2) Establish a delay prediction model

The wavelet neural network selected in the simulation has 5 input nodes, 5 output nodes, and 10 implied neuron nodes. The input is:

$$X = (\tau_1, \tau_{k-1}, \tau_{k-2}, \tau_{k-3}, \tau_{k-4})$$

The output is:

$$y(k) = (\bar{\tau}_{k+1}, \bar{\tau}_k, \bar{\tau}_{k-1}, \bar{\tau}_{k-2}, \bar{\tau}_{k-3})$$

In the formula, $\tau_k$ is the real time delay value at time k, and $\bar{\tau}_k$ is the time delay value predicted at time k, where $\bar{\tau}_{k+1}$ is the predicted value at the next moment.

The value of error e:

$$e = (\bar{\tau}_k - \tau_k)^2 + (\bar{\tau}_{k-1} - \tau_{k-1})^2 + (\bar{\tau}_{k-2} - \tau_{k-2})^2 + (\bar{\tau}_{k-3} - \tau_{k-3})^2$$

(11)

Using training data, the weights and parameters are constantly updated based on the size of the error. The final system predictive model is:

$$y(k) = \sum_{j=1}^{5} \omega_j h_j + \sum_{i=1}^{10} \omega_i h_i$$

(12)

Apply the trained wavelet neural network delay prediction model as shown in Equation (12) to verify the prediction effect of the model in the Matlab simulation environment. Figure 3.2 is a comparison of the data generated by the delay model of the ICN hybrid network and the prediction data. By comparison, it can be found that when the delay data is mutated, the prediction error reaches 0.01s, but it can be ignored. In general, the trained wavelet neural network has good predictability for most delay times and has high prediction accuracy.

IV. ROUTING FORWARDING STRATEGY ALGORITHM BASED ON TIME DELAY PREDICTION

A. Store Forwarding Model

Through the delay predicted by the trained neural network in Section 3, the forwarding route is selected. This article adds an interface information Stat to the FIB table innovatively. It uses simple numbers to indicate the status. When forwarding, it can determine which route to take by simply comparing the interface content of each route. At the same time, the delay table is stored at the node, which includes five historical delay data ($D_1$-$D_5$) and the delay data predicted from the neural network. The content of the Stat interface added in the FIB table is changed according to the size of the predicted delay data.

B. Forwarding Algorithm Based on Time Delay Prediction

The Figure 4.1 shows the forwarding process in the entire NDN network. Due to the addition of the Stat interface, it is assumed that there are several routes with named-pointed data packets on the premise that there are forwarding routes. If there is only one, the interest packet is forwarded in this direction, and if there are more than one, the data in Stat is judged to execute the forwarding strategy.

C. NNDP

(1) The data is forwarded in the direction of the smallest Stat interface.

(2) If there is the same Stat interface, all the smallest interfaces are forwarded.

(3) After the transfer is completed, the delay in $D_1$ is deleted, and the time delay data is moved to the left in turn, and the new actual time delay is stored in $D_5$, and the sixth
delay is predicted. After the Stat is updated, if the forwarding interface is changed, the historical delay data in the delay table is changed to the delay data of the new route.

D. Stat Updating Method

(1) There is only one state 0 for the interface of the same interest packet and 0 is the smallest Stat.

(2) When an interest packet interface is added, the delay of the interface is predicted and compared with the delay of the existing interface. The Stat of the interface is assigned according to the comparison result.

(3) If all Stats are the same, all ports will be forwarded. After forwarding, the Stats of each interface are first stored with their respective returned delays, and the current minimum delay is stored in the delay table. Then the delays are sorted. The smallest is 0, followed by an increase, and the number is not repeated. For the same time delay, the more time-delay value is treated as the smaller Stat value according to the position in the FIB table.

(4) If the predicted delay exceeds $T_1$ ($m_1$ times of the timeout value), all Stats are immediately set to 1. When the predicted delay exceeds $T_2$ ($m_2$ times of the timeout value), the route is forwarded to the route with Stat 1, and if the returned delay is less than $T_2$, the two values with Stat 0 and 1 are interchanged. If the predicted data still exceeds $T_2$, all Stats are set to 1. Among them, $m_1$, $m_2$ are determined according to the specific conditions of the network.

(5) When the route is changed, all the five historical delays are stored as the time delay of the replacement route so as to avoid the influence of the delay of the previous route and reduce the storage pressure of the local time delay table.

E. The Dynamic Adaptive Adjustment of the Route

Because the NDN network itself has high dynamic nodes[9], it will bring topological uncertainty and even route interruption. In order to ensure normal communication even when congestion or failure occurs in the best-optimum route, this chapter designs a dynamic adaptive routing mechanism and rerouted faults.

The routing and link status judgment are based on timer-based timeout retransmission counts[10]. Each routing node sets a timeout retransmission counter for the corresponding content. When a timeout retransmission occurs many times in a short time, the link is considered to be faulty. The routing node can quickly detect the fault and take corresponding measures according to the specific severity to ensure the quality of follow-up communication. The adjustment process is divided into the following three aspects:

(1) When timeout retransmission occurs, if the frequency is not too high, it is considered that the optimal route is congested or faulty and the sub-optimal route communication is used first. Communication is performed by selecting the second smallest route in the Stat interface.

(2) If timeout retransmission occurs frequently in a short time, the previous route is considered to be completely interrupted, and the retransmitted interest packet needs to be broadcast again to the entire network, and the route is reestablished. Then the FIB data is updated in real time through the methods in Sections C and D.

(3) When the faulty node resumes normal operation, the original optimal route may recover, or, due to network conditions, a better route appears, and the mechanism can use the predicted delay and actual feedback delay to make adjustments. The data in the FIB is updated in real time, and the traffic is converged to the current optimal route as soon as possible.

V. SIMULATION VERIFICATION

A. The Introduction of Simulation Experiment

In this chapter, the NDNSim simulator is used to test the performance of the proposed algorithm, and a relatively simple binary tree is selected as the NDN network topology, as shown in Figure 5.1. The network description is as follows:

![Fig. 5.1 Binary tree network architecture in NDN](image-url)

The Node 1 in the network is a user and sends different interest packets (simulation is to set a random request for 1-10 named interest packets, representing 10 data packets). Nodes 2-15 are intermediate routes of the network, and have CS (It can store data packets) and Stat (show which direction to send data packets).

Store 5 most recently returned delay values in the top node 2, and predict the next delay, and apply the algorithm in Section 3.3 according to the predicted delay to perform link 2-3 and link 2-4.

Since it is a binary tree structure network, Stat is set to only 1, 2, and 3 values. 1 represents sending data in 2 directions; 2 represents sending data to left subtree nodes; 3 represents sending data to right subtree nodes. (5) In an NDN network, each node should have CS (cached packets), PIT, and FIB. In order to simplify the network, we make the network initialization and each end node, namely node 9-16, has these 10 types of data packets at the same time, so each node has 2 directions to choose the route, and it will always have 2 directions to reduce PIT.

In order to verify the effectiveness of the adaptive route forwarding strategy based on neural network algorithm, two groups of experiments are conducted in this paper. The following are the introduction and purpose of the two experiments:

The first group of experiments is to test the performance of the transmission mechanism under the guidance of the NNDP strategy and the average delay of the conventional NDN transmission mechanism. In this group, the CS of each node in the network topology buffers up to 5 data packets,
and each time a new data packet arrives, if the CS table does not exist, the new data packet is made to arrive at the top and the data is down updated in turn; if there is a new data packet, the CS table is not updated. The performance of the network is evaluated by sending out the first interest packet and the returned average time delay of the last packet. The main purpose of this group of experiments is to verify that under the guidance of the NNDP, the transmission mechanism can choose the route forwarding with smaller delay to improve the network experience.

The second group of experiments is to test the influence of the dynamic self-adaptive adjustment strategy introduced in routing on the NNDP strategy. In this group of experiments, the nodes cache all 1-10 types of packets, increasing the chance of route interruption and congestion as much as possible. The main purpose of this group of experiments is to verify that in the NNDP routing forwarding strategy, if all nodes have the qualification of caching packets to cause route interruption and congestion, request satisfaction rate of the users will be raised by the dynamic self-adaptive adjustment strategy.

The two network indicators of the two groups are as follows:

Definition1. Request satisfaction rate $R_{satisfy}$: the percentage of data consumers who successfully receives a data packet during the time period T and the number of interest packets generated by the data consumer.

Definition 2. Average response delay $T_{delay}$: The average value of the data consumer from the time when the interest packet is sent out to the corresponding data packet within the time period T.

B. Simulation Results

(1) Performance Comparison between Route Forwarding Strategy based on NNDP and Regular NDN Route Forwarding Strategy

![Simulation comparison chart of average time delay](image1)

According to the provisions of section A, NDNSim software is used to simulate the NDN network. Each time 100 interest packets are sent, the contents of 1-10 are random, but each node's CS can only store up to 5 data packets for a total of 50 times. Figure 5.2 shows the average delay for each simulation after 50 times of 100 data simulations. From the figure, we can see that when the packet number is 25, the average delay has abruptly changed. According to the detection, it is due to the congestion between nodes 2-3, resulting in a sudden increase in delay. However, the prediction delay is implemented under the guidance of the NNDP. The Stat interface is updated in real time and the forwarding route is redesigned based on the updated interface. After the delay is changed, the network quickly returns to normal. The experiments show that the total average delay after applying the algorithm is much smaller than the average delay without applying the algorithm, which directly improves the network performance.

(2) Performance Analysis of Forwarding Strategy Based on Neighbor Awareness Introduced with Assisted Forwarding Mechanism

In the design of this chapter, in order to ensure the normal communication of the network in the event of route interruption and congestion, this paper introduces a route dynamic adaptive adjustment strategy. In order to verify the effectiveness of the introduction strategy, each node caches all types of data packets in this group of experiments. The experimental results are shown in Figure 5.3. In the figure, the abscissa indicates the number of packets, and the ordinate indicates the request satisfaction rate. From the figure, it can be seen that the satisfaction rate of the NDN after the introduction of the route dynamic self-adaptive adjustment strategy has been stable at about 90%, while the satisfaction rate of the non-introduced NDN has been between 75% and 80% fluctuations. The experimental results show that the satisfaction rate of the NNDP that introduces the routing dynamic adaptive strategy is higher than the non-introduced satisfaction rate. It is because the route dynamic self-adapting adjustment strategy will know the network communication situation in real time according to the overloaded counter, so as to select the forwarding route according to the situation, thereby ensuring normal communication and improving the network performance.

![Simulation comparison chart of request satisfaction rate](image2)

Through the visual and clear reflection of average delay and request satisfaction rate in Figure 5.2 and Figure 5.3, it can be found that after applying an adaptive routing forwarding strategy based on neural network algorithm, both the average network time delay and the request satisfaction rate are higher than the conventional NDN transmission. The mechanism is much better. Therefore, it can be concluded
that the adaptive routing forwarding strategy based on neural network algorithm in this chapter can improve the performance of the network to a certain extent and reduce the occurrence of network congestion.

VI. CONCLUSION

In the ICN network, due to its caching mechanism, a large number of copies of information provided by the server are cached in the content memory of the routing node. If a copy of information cached by a routing node can satisfy a user's service request, the corresponding routing node can directly respond without forwarding to the server. This has caused the ICN network to respond to the user service request with such situations as disorder, distribution, crossover and duplication, which will seriously affect the performance of the ICN network.

In this paper, aiming at the above problems, I propose an adaptive routing and forwarding strategy based on neural network algorithm. The strategy innovatively adds interface information Stat on the forwarding information based (FIB) of NDN network component, and predicts the time delay by combining the neural network with strong prediction ability. Guiding the data forwarding path dynamically by analyzing and determining the prediction delay. In addition, I design the routing dynamic adaptive adjustment mechanism and fault reselection function because of the routing congestion and interruption. The simulation results show that the strategy can effectively improve the network transmission performance, and reduce the network congestion to some extent.

REFERENCES