Queue Management Algorithm for Satellite Networks Based on Traffic Prediction

YUXIA BIE\textsuperscript{1}, ZHIHAN LI\textsuperscript{1}, ZHI HU\textsuperscript{2}, AND JIAMEI CHEN\textsuperscript{1}\textsuperscript{*}

\textsuperscript{1}College of Electronic Information and Engineering, Shenyang Aerospace University, Shenyang 110136, China
\textsuperscript{2}Software College, Shenyang Normal University, Shenyang 110034, China

Corresponding author: Zhi Hu (hu-zhi521@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61901284, in part by the Liaoning Revitalization Talents Program under Grant XLYC1907195, and in part by the Aeronautical Science Foundation of China under Grant 2020Z017054001.

\textbf{ABSTRACT} In connexion with the effect of the self-similar characteristic of satellite network service traffic on queueing performance, a prediction model with optimised triple exponential smoothing is first established in this paper. This model performs network traffic prediction based on the dynamic triple exponential smoothing model and optimises the smoothing coefficient of the model through the differential evolution algorithm; a cubic function based on traffic prediction is further proposed to improve the adaptive random early detection (ARED) queue management algorithm. Based on the network traffic prediction results and the ARED, this algorithm uses the cubic function to perform nonlinear processing on the packet drop probability function. The simulation results show that the prediction model with optimised triple exponential smoothing has a high prediction accuracy, and the improved ARED algorithm based on the cubic function of traffic prediction in the presence of data bursts in self-similar traffic. It can effectively reduce the packet loss rate and improve the throughput, so as to better control the network congestion caused by selfsimilar traffic in satellite network.

\textbf{INDEX TERMS} Dynamic triple exponential smoothing model, differential evolution algorithm, cubic function, adaptive random early detection.

\section{I. INTRODUCTION}

In recent years, satellite networks have realised seamless global information transmission in real time and are widely applied to numerous fields such as national defence and military affairs, meteorological environment detection, positioning and navigation, and telemedicine. Large amounts of multimedia information such as images, audio, video, and streaming broadcast are generated and forwarded in satellite networks and, at the network aggregation nodes, the data generated by the satellite itself are aggregated with the data coming from other nodes.

The satellite network studied in this paper is an integrated high-capacity information network comprising different types of satellites, constellations, space stations and other spacecrafts with different orbits and performance, and corresponding ground facilities, connected together through intersatellite and satellite-ground links. It’s capable of accurate information acquisition, rapid information processing and efficient information transmission by utilizing high-speed on-satellite processing, switching and routing technologies. Among the spacecrafts included in the network, geosynchronous satellites are used as aggregation nodes of the network, and other spacecrafts as access nodes. The services transmitted by the backbone nodes include services generated by the aggregation nodes itself, services from other spacecraft and from the ground network, with high data rate and high burst rate. In cases of data burst, the queuing performance of the network aggregation nodes would be greatly affected. Applying the active queue management mechanism on the satellite aggregation nodes can effectively avoid congestion by reporting the network congestion information to the source end through packet loss of a small amount of information to reduce the information sending rate of the source end.

Research has shown that the service traffic of the terrestrial Internet generally exhibits self-similarity [1]; the self-similar services of the terrestrial network enter the satellite link, and the traffic after aggregation and propagation in the
satellite network still shows self-similarity, that is, the self-similarity of the traffic is also ubiquitous and is propagated in the satellite network, making the self-similarity of the network traffic an unavoidable inherent characteristic [2]. The self-similarity of satellite network services is embodied in the long-range dependence of the traffic, which puts the aggregation nodes in a burst state for a long period of time and greatly affects queuing performance. The self-similar data flow is predictable, and network traffic prediction can perceive the burst state of queuing in advance. Therefore, the establishment of a high-precision traffic prediction model and research on a queue management algorithm for satellite networks based on traffic prediction can limit the queue length in advance and control network congestion.

To solve the aforementioned problems, in this paper, the approximate self-similar traffic with a Pareto distribution generated from the aggregation of a large number of ON/OFF sources is analysed as the basic model, a queue management algorithm for satellite networks based on traffic prediction is proposed, traffic prediction is conducted using a dynamic triple exponential smoothing model, and the smoothing coefficient of the model is optimised through the differential evolution (DE) algorithm to increase prediction accuracy and reduce complexity. Furthermore, based on the adaptive random early detection (ARED) algorithm, the cubic function is used to perform nonlinear processing on the packet drop probability function, and the results of the traffic prediction are substituted into the queue management algorithm to propose an improved ARED algorithm based on the cubic function of traffic prediction. This algorithm can effectively control queue length, reduce the packet drop rate, increase throughput, and make the channel transmission more stable.

The main algorithm in this paper is as follows.

First, a dynamic triple exponential smoothing (Holt-Winter) model is constructed for the self-similar traffic in the satellite network. The smoothing coefficient $\alpha$ in the dynamic model is not a traditional fixed value, and it can respond in a timely manner when faced with traffic data that fluctuate greatly.

Second, the selection of $\alpha$ directly determines the prediction accuracy. In this paper, based on the dynamic triple exponential smoothing model, the DE algorithm is used to optimise $\alpha$, and a traffic prediction model with optimised triple exponential smoothing is proposed to increase prediction accuracy.

Third, the optimised traffic prediction model is introduced to queue management, and the new cubic function is used to perform nonlinear processing on the packet drop probability function in ARED.

Finally, the overall process for the queue management of satellite networks based on traffic prediction is designed, and the packet drop probability obtained is used to control the average queue length, packet drop rate, and throughput in channel transmission to effectively prevent channel congestion.

The remainder of this paper is organised as follows. The relevant prediction models and queue management models are analysed in Section II. The service model of the satellite network is described in Section III. Based on the dynamic triple exponential smoothing model, the DE algorithm is employed to optimise the traffic prediction model with triple exponential smoothing in Section IV. Section V focuses on improving the packet drop probability function in connexion with the ARED algorithm using the cubic curve, followed by the introduction of the traffic prediction model into queue management. The simulation results and performance analysis are discussed in Section VI.

II. RELATED WORK

Researchers have studied network traffic prediction models extensively. The traditional Markov model is a typical short-range dependence model that fails to satisfy the demands of long-range dependence network prediction [3]. With the development of machine learning, prediction models such as back propagation (BP) neural network and support vector machine (SVM) have emerged. Although the BP neural network can depict the long-range dependence and short-range dependence of the network traffic very well, it suffers from difficult selection of parameters, slow rate of convergence, and susceptibility to premature convergence, making it unsuitable for real-time prediction of network traffic [4]. The SVM model can be a solution only for machine learning with small samples and is difficult to implement for large-scale training samples [5]. The most widely applied prediction models are the time series models, which are divided into the moving average model, stationary time series model, and exponential smoothing model [6], as briefly described below. The moving average model can effectively eliminate random fluctuations in prediction, but it can only use the mean value of the latest set of actual data for prediction, which does not consider the later trends, making it suitable for instant prediction in which the data neither increase rapidly nor decrease rapidly. The stationary time series model has a simple structure but requires that the predicted data be stable, as this model cannot capture the pattern of bursty data. The exponential smoothing model is compatible with the advantages of the above two types of time series models and is adapted to long-range dependent, self-similar traffic data; this model gives the past data a gradually diminishing degree of influence as they move away, calculates the corresponding exponential smoothing value, and then acts in concert with the optimised model to greatly improve the prediction accuracy.

The existing studies are mainly based on neural network flow models, multi-fractal wavelet models, multiple fractal finite shock response neural network prediction models, etc. This line of research is in the budding phase and there is no guarantee for the prediction accuracy and there hasn’t been detailed discussion about the condition variables in the model. The most commonly used traffic prediction algorithms are either neural network or combined prediction algorithms. In this paper, a time series model combined with a
three-time exponential smoothing algorithm and a differential evolution algorithm is used which could predict self-similar traffic accurately and be linked to the subsequent queue management algorithm seamlessly.

The research of network traffic characteristics is based on the theory of time series analysis, which is aimed at the modeling of traffic prediction. As a more comprehensive algorithm in time series, cubic exponential smoothing method is almost suitable for analyzing all application problems of time series. Time series is an objective record of the historical behaviors of the system under study, which is to analyze these historical records, find out the statistical dependence between data, grasp the structural characteristics and operating rules of the system, and predict the future behavior of the system. The research shows that the network traffic model has self-similarity statistically, which can accurately describe the network traffic, properly establish functions and dependencies on the existing data, analyze the inherent patterns of the data, control and predict the future data. The improved cubic exponential smoothing method is established to smooth the burst data stream three times, and the coefficient and flow can be predicted accurately after the cubic smoothing, which has important reference value for eliminating network congestion and realizing the management and planning of complex networks. Cubic exponential smoothing algorithm has strong adaptability to the self-similarity and non-regularity of satellite network traffic. It can reduce the impact of burst traffic on the overall prediction results, and has high prediction accuracy for satellite network traffic prediction research.

Queue management algorithms can be passive or active [7]. The typical representative of passive queue management is the tail drop algorithm, which is simple in principle but poses problems of deadlock, full queue, and global synchronisation. The random-drop and the drop-front algorithms are derived subsequently to effectively avoid the deadlock problem, but the full queue and global synchronisation problems remain unaddressed. Later, proposed the use of active queue management algorithms [8], which are represented by the random early detection (RED) algorithm, the random exponential marking (REM) algorithm based on optimisation theory, and the ARED algorithm. The RED algorithm proposed by Floyd et al. is the earliest version of the active queue management algorithm; this algorithm greatly improves the operating quality of the network but is highly sensitive to parameter settings [9]. Steven et al. combined the optimisation theory with the REM algorithm, which, however, has drawbacks such as slow convergence and slow reaction to bursts of traffic [10]. Nichols and Jacobson proposed a new active queue management algorithm named Controlled Delay and shortened as CoDel [11], which provides a partial solution to the router’s over-caching problem. The Controlled Delay active queue management algorithm CoDel, uses the delay of the packets in the queue as the evidence for determining congestion, and strictly controls the delay. When each packet arrives, the queue length size is checked, and if it is bigger than the maximum queue length \( \min_{th} \), the packet is discarded, otherwise it enters the queue, and a time stamp is added to the header of the packet; if the current queue delay is longer than the queue expected delay, the queue enters the discard state, and the data packet is continuously discarded, and when the queue delay drops below the queue expected delay, the queue enters the non-discard state. The CoDel algorithm needs to check the queue size, record the time when the packet enters the queue, and get the delay of the packet in the queue by calculating the difference between the current time and the time when the packet enters the queue. The performance can then be analyzed comprehensively by ARED algorithm based on the average queue length, the queue threshold, transmission rate, link utilization, packet loss rate. ARED algorithm itself contains a discreet adaptive mechanism that can improve the stability of the algorithm.

The ARED algorithm is capable of realising the automatic adjustment of parameters, thereby significantly increasing the system channel utilisation and greatly decreasing the packet drop rate of the system; however, the packet drop probability function in ARED is a linear function, and the overly rapid linear growth rate easily leads to queue oscillation [12]. For this reason, the S-shaped ascending half-Cauchy distribution function is used in some studies to perform nonlinear adjustment to the packet drop probability function of the ARED algorithm; consequently, the packet drop rate decreases, and the queue management system becomes stable, whereas changes in the queue length is not considered in advance.

In connexion with the self-similar characteristic of the satellite network traffic, an algorithm is proposed in this paper to introduce traffic prediction in queue management for satellite networks.

### III. TRAFFIC MODEL OF SATELLITE NETWORK

The satellite network communication system consists of space and ground segments, as shown in Figure 1. The space segment mainly refers to the orbiting satellites and the ground stations that control the orbiting satellites. These ground stations mainly implement tracking, remote sensing, and remote control functions to provide necessary satellite management and control and hence ensure the normal in-orbit operation of the satellites. The ground segment mainly refers to the user terminals that conduct communication through the satellites, including fixed terminals, mobile terminals, and movable terminals.

The services at the satellite communication aggregation nodes come from self-generated service and service generated at the satellite or ground segments. Each node provides various types of services, and the various services of multiple nodes aggregate at the nodes of the satellite network, thus leading to a burst of traffic on the nodes. In order to truly describe the characteristics of satellite network traffic, it is necessary to develop an accurate and reliable operating model of network traffic. Previous researchers have proposed many traffic prediction mechanisms to generate self-similar network models, such as Fractal Brown Motion (FBM), Fractional Gaussian Noise (FGN), Multi-fractal Models (MFM),...
Multi-fractal Wavelet Models (MWM), Fractional Auto-Regressive Integrated Moving-Average (FARIMA) model and ON/OFF superposition model with heavy-tailed distribution. Among the above mentioned models, the FBN, FGN and FARIMA models have specific modeling formulas, which make it easy to analyze the model performance. However, the disadvantage is that they suggest no physical meaning, thus unable to indicate the causes of traffic generation. The ON/OFF model could explain the causes of Network Traffic Self-Similarity by aggregating a large number of independent ON/OFF sources to generate auto-correlated aggregated traffic. When the number of ON/OFF source reaches infinity, the aggregated traffic converges to fractal Brownian motion, which can explain the single fractal behaviour at long time scales. Nonetheless, this model cannot explain the multi-fractal characteristics of multi-fractal models at short time scales. These models mainly include MFM and MWM which usually assume that they conform to a random distribution and iterate from top to bottom to obtain a multi-fractal traffic flow. However, this approach doesn’t carry much physical significance.

The ON/OFF model is used in this paper. The lengths of the duration when a single service is in the ON state and the OFF state are typically independent and identically distributed random variables. The ON cycle random variable \( A \) satisfies \( E(A) = 1/\nu \), and the OFF cycle random variable \( B \) satisfies \( E(B) = 1/\mu \); then, the average velocity of the ON/OFF source is \( \nu\mu/(\nu + \mu) \). The duration of the ON/OFF state has a finite mean and infinite variance. Mathematical derivation shows that the time intervals of ON and OFF follow a heavy-tailed distribution and have strictly alternating ON and OFF cycles, and the random process generated by the superposition of a large number of ON/OFF sources has self-similarity.

The strictly alternating ON and OFF cycles are used to describe the alternation of the data sources of the satellite network nodes between sending data and nonsending data states [13]. Network traffic with self-similar characteristics is generated through the superposition of multiple independent ON/OFF service sources, and the heavy-tailed distribution of the ON cycle and OFF cycle of the ON/OFF service sources can be expressed using the Pareto distribution.

The Pareto distribution is the most commonly used heavy-tailed distribution, and its distribution function is

\[
F(x) = 1 - \left( \frac{k}{x} \right)^\lambda, \quad 0 < \lambda, \ x > k
\]

The parameter \( k \) determines the minimum value that the random variable can take, while the parameter \( \lambda \) determines the mean and variance of the random variable. If \( \lambda \leq 2 \), the distribution has infinite variance; if \( \lambda \leq 1 \), the variance has an infinite variance and mean. The superimposed services generate self-similar traffic as long as the ON or OFF cycle follows a heavy-tailed distribution. Therefore, in this paper, the traffic of the satellite network aggregation nodes can be considered the joint action of a series of ON/OFF sources superimposed with the Pareto distribution, and the network traffic generated by multiple ON/OFF service sources is sent to the receiving end through the aggregation nodes, as shown in Figure 2.

**IV. TRAFFIC PREDICTION MODEL WITH OPTIMISED TRIPLE EXPONENTIAL SMOOTHING**

**A. PREDICTION MODEL WITH DYNAMIC TRIPLE EXPONENTIAL SMOOTHING**

The exponential smoothing model for time series was first proposed by Holt in 1958. The principle is that the exponential smoothing value of any period is the weighted average of the actual observed value of the current period and the exponential smoothing value of the previous period. The smoothing technique is used to weaken the effect of short-term random fluctuations on the series and to smoothen the series, thereby obtaining the time series smoothing value as the prediction parameter for the short term [14]. This algorithm is...
The exponential smoothing model has different times of smoothing. In particular, the prediction model with triple exponential smoothing (Holt-Winter) mainly performs correction in connexion with the nonlinear trend in the time series and is able to adapt to the nonlinear, self-similar, and long-range dependent changing trends of the satellite network traffic [5]. The Holt-Winter algorithm includes three smoothing equations and one prediction equation. It calculates the weighted average of the observed values from each period according to the time sequence, and the result is used as the traffic prediction value, showing that the influence of historical data on future values decreases with time [16].

Suppose the time series is \( y_1, y_2, \ldots, y_t, \ldots \), the smoothing coefficient is \( \alpha, 0 < \alpha < 1 \), and the triple exponential smoothing equation is

\[
S_t^{(1)} = \alpha y_t + (1 - \alpha)S_{t-1}^{(1)} \\
S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha)S_{t-1}^{(2)} \\
S_t^{(3)} = \alpha S_t^{(2)} + (1 - \alpha)S_{t-1}^{(3)}
\]

where \( S_t^{(1)}, S_t^{(2)}, \) and \( S_t^{(3)} \) are the single, double, and triple exponential smoothing values, respectively. \( S_t^{(1)} \) is expanded successively as follows:

\[
S_t^{(1)} = \alpha y_t + (1 - \alpha)[\alpha y_{t-1} + (1 - \alpha)S_{t-1}^{(1)}] = \cdots = \alpha \sum_{n=0}^{\infty} (1 - \alpha)^n y_{t-n}
\]

Equation (3) indicates that \( S_t^{(1)} \) is the weighted average of all historical data, using the weighting coefficients of \( \alpha \), \( \alpha (1 - \alpha) \), and \( \alpha (1 - \alpha)^2 \cdots \), respectively. It can be seen that the predicted value of period \( t \) as well as the actual observed values of period \( t - 1 \) and all other previous periods increase by \( (1 - \alpha)^n \), that is, the observed values closer to the prediction period are assigned larger weights, and the observed values farther away from the predicted value are given smaller weights, with the weights decreasing exponentially from near to far. The principle for expanding \( S_t^{(2)} \) and \( S_t^{(3)} \) is the same as above [17].

The smoothing coefficient \( \alpha \) of the traditional triple exponential smoothing model takes a fixed value. This model is applied mainly to stable data models. In medium- and long-term data prediction, significant changes in data defy their timely adjustment, thereby leading to large prediction errors. For this reason, the traditional triple exponential smoothing method was improved to construct a prediction model with dynamic triple exponential smoothing [18], where dynamic smoothing coefficients are proposed based on the traditional algorithm, and the parameters are updated through iteration. This algorithm can reduce the prediction error in the medium- and long-term prediction of complex data and has better stability.

The smoothing coefficient is denoted as \( \alpha_{m,n} \), where \( m \) is the predicted traffic data and \( n \) is the time point of the

data prediction. The dynamic triple exponential smoothing equation is

\[
S_t^{(1)} = \alpha_{m,n} x_t + (1 - \alpha_{m,n})S_{t-1}^{(1)} \\
S_t^{(2)} = \alpha_{m,n} S_t^{(1)} + (1 - \alpha_{m,n})S_t^{(2)} \\
S_t^{(3)} = \alpha_{m,n} S_t^{(2)} + (1 - \alpha_{m,n})S_{t-1}^{(3)}
\]

where

\[
S_1^{(1)} = S_2^{(2)} = S_3^{(3)} = X_{N-3}, X_{N-2}, X_{N-1}
\]

The first three values of the overall data are selected for \( X_{N-3}, X_{N-2}, \) and \( X_{N-1} \) in Equation (5). Then, every time the predicted value is calculated, the initial values \( S_1^{(1)}, S_2^{(2)}, \) and \( S_3^{(3)} \) are updated.

Suppose the horizontal parameter is \( a_t \), the trend parameter is \( b_t \), the seasonality parameter is \( c_t \), and the predicted value of period \( T \) in the future is \( y_{t+T} \). Then, the prediction equation is

\[
y_{t+T} = a_t + b_t T + c_t T^2, \quad T = 1, 2, 3, \ldots
\]

The smoothing coefficient and parameters of each time point in the dynamic triple exponential smoothing model change dynamically. Hence, iterative calculation is required for each period as follows:

\[
a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)}
\]

\[
b_t = \frac{\alpha_{m,n}}{2(1 - \alpha_{m,n})} \left[ (6 - 5\alpha_{m,n})S_t^{(1)} - (10 - 8\alpha_{m,n})S_t^{(2)} + (4 - 3\alpha_{m,n})S_t^{(3)} \right]
\]

\[
c_t = \frac{\alpha_{m,n}^2}{2(1 - \alpha_{m,n})^2} (S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)})
\]

**B. OPTIMISATION OF THE SMOOTHING COEFFICIENT**

In the dynamic triple exponential smoothing model, determining the appropriate smoothing coefficient \( \alpha_{m,n} \) is the key to the entire model. The DE algorithm is an efficient global optimisation algorithm, which is used mainly to solve global optimisation problems with continuous variables. In this paper, the DE algorithm is used to optimise the smoothing coefficient \( \alpha_{m,n} \) to increase the prediction accuracy of the dynamic triple exponential smoothing model.

Through continuous iterative calculations, the DE algorithm keeps good individuals, eliminates inferior individuals, and guides the search process to approach the global optimal solution [19]. The memory ability and selectivity characteristics of the DE algorithm enables it to dynamically track the current search situation to adjust its search strategy and optimise the smoothing coefficient \( \alpha_{m,n} \) of the dynamic triple exponential smoothing model at every moment. The DE algorithm optimises the smoothing coefficient \( \alpha_{m,n} \), which overcomes the problems of low efficiency and inability to self-adapt in the parameter selection of the traditional Holt-Winter model, thus increasing the prediction accuracy and efficiency of the triple exponential smoothing model.
The fitness function is established with the minimum mean squared error (MSE) of samples as the optimisation criterion. The MSE is expressed as follows, with $S_i$, $x_i$, and $I$ representing the predicted value, the true value, and the total number of traffic data, respectively:

$$\text{MSE} = \frac{\sum_{i=1}^{n} S_i - x_i}{I}$$  \hspace{1cm} (10)$$

The problem of obtaining the optimal solution for $\alpha$ can be transformed into the calculation of the minimum MSE between the predicted value and the actual value at each iteration and the corresponding smoothing coefficient and using it as the optimal value for this iteration. The smaller the MSE is, the higher the accuracy is. As shown in Figure 3, the individuals in the population are repeatedly subject to iterative adjustments of crossover, mutation, and selection until the maximum number of iterations is reached or the termination condition of the fitness function is satisfied, and the optimal smoothing coefficient $\alpha_{m,n}$ and the corresponding predicted value at this time are used as the output.

Let $\Delta$ be a vector of individuals in the population, the evaluation function $f(\Delta)$ is used as the fitness function, and the DE algorithm is used to solve for the optimal value $\Delta^*$. After $M$ individuals have undergone $G$ times of evolution, the predicted value of the output is substituted into Equation (10) to solve for the optimal value when the function $f(\Delta)$ takes the minimum value. This problem is the global optimal solution problem of a multivariate function. The process of optimising the smoothing coefficient with the DE algorithm is shown in Table 1.

### TABLE 1. The process of optimising the smoothing coefficient with the DE algorithm.

| Step | Description |
|------|-------------|
| 1.   | $t \leftarrow 1$ (Initialisation) / Initialise the vector of individuals |
| 2.   | for $i=1$ to $M$ do / Initialise M individuals, i represents the i-th individual, t represents the t-th generation |
| 3.   | for $j=1$ to D do / represents the value of the j-th dimension of individual i |
| 4.   | $x_{ij}^* = x_{ij} + \text{rand}(0,1)(x_{ij}^{\text{max}} - x_{ij}^{\text{min}})$ |
| 5.   | end |
| 6.   | end |
| 7.   | while $|f(\Delta)| \geq \varepsilon$ or $t \leq T$ do |
| 8.   | for $i=1$ to $M$ do |
| 9.   | $\text{Mutation and Crossover}$ |
| 10.  | for $j=1$ to D do |
| 11.  | $v_{ij} = \text{Mutation}(x_{ij}^*)$ / Mutate individual v |
| 12.  | $u_{ij} = \text{Crossover}(x_{ij}^*, v_{ij})$ / Crossover to generate experimental individual u |
| 13.  | end |
| 14.  | (Greedy Selection) |
| 15.  | if $f(x_{ij}) < f(u_{ij})$ then |
| 16.  | $x_{ij} \leftarrow u_{ij}$ |
| 17.  | else |
| 18.  | if $f(x_{ij} + u_{ij}) < f(\Delta)$ then |
| 19.  | $\Delta \leftarrow x_{ij}$ |
| 20.  | end |
| 21.  | $x_{ij} \leftarrow x_{ij}$ |
| 22.  | end |
| 23.  | end |
| 24.  | $t \leftarrow t + 1$ / Complete a generation of updates |
| 25.  | end |
| 26.  | return the best vector $\Delta^*$ / Return to the optimal value |

### C. TRAFFIC PREDICTION PROCESS WITH OPTIMISED TRIPLE EXPONENTIAL SMOOTHING

The smoothing coefficient and parameters of each time point in the dynamic triple exponential smoothing model change dynamically, requiring iterative calculation in each period. The smoothing value of the data is calculated first, and then the optimal value $\Delta^*$ in Algorithm 1 is designated the dynamic smoothing coefficient $\alpha_{m,n}$, which is substituted into Equations (7), (8), and (9) to obtain the predicted value using Equation (6). The process of traffic prediction performed with the optimised triple exponential smoothing model is shown in Table 2, where m is the predicted traffic data and n is the time point of the data set prediction.

### V. IMPROVED ARED ALGORITHM BASED ON THE CUBIC FUNCTION OF TRAFFIC PREDICTION

In queue management algorithms, when there are too many devices accessed by users at the buffer end, buffer expansion will cause congestions. Many ways including protocol modification such as delay-based congestion control would solve this problem, but it may produce an opposite effect caused by network jitter, dynamically changing access devices based on buffer size or a combination of both. The Active Queue Management (AQM) mechanisms which have many other potential benefits can solve the problem of buffer bloat. Apart from the first-in-first-out (FIFO) and tail-drop algorithms, cache management policies are considered to be aggressive and the corresponding method used by routers to manage queues is called aggressive queue management (AQM). It collects information about congestions by marking the data packet from routers, then transmits the status of routers and switches to the end system to obtain AOM mechanism. The classical algorithm of the AQM algorithm is RED and some other improved algorithms on the basis of the RED algorithm. The Random Early Detection (RED) mechanism detects the onset of congestions in advance and controls data packet...
TABLE 2. Prediction model with dynamic triple exponential smoothing method.

| Algorithm 2: Prediction model with dynamic triple exponential smoothing method |
|-----------------|-----------------|
| **Input:** Satellite network traffic data y_i |
| **Output:** Predicted value of satellite network traffic data, y_{i+T} |

1. y_i ← Satellite network data // Initialisation
2. Call the optimal vector of individuals, \( \alpha \), in Algorithm 1
3. \( \alpha = \Delta \)
4. \( S^{(0)} = \sigma \sum_{i=1}^{m} X_{i} \) // The first three values of the overall data (\( X_{i} \)) are selected as the initial values
5. for \( i = 3 \) to \( m \) do
   \( S^{(i)} = \alpha_{1}S^{(i-1)} + (1 - \alpha_{2})S^{(i-2)} \)
6. end
7. \( S^{(i)} = \alpha_{2}S^{(i-1)} + (1 - \alpha_{3})S^{(i-2)} \)
8. \( S^{(i)} = \alpha_{3}S^{(i-1)} + (1 - \alpha_{4})S^{(i-2)} \)
9. end
10. \( a_{i} = 3S^{(i)} - 3S^{(i-1)} + S^{(i-2)} \)
11. \( b_{i} = \frac{\sigma}{2(\alpha_{2} - \alpha_{3})} \left[ (6 - 5\alpha_{2})S^{(i-1)} - (6 - 5\alpha_{3})S^{(i-2)} - S^{(i-3)} - (4 - 5\alpha_{4})S^{(i-4)} \right] \)
12. \( c_{i} = \frac{\sigma}{2(\alpha_{2} - \alpha_{3})} \left[ (6 - 5\alpha_{2})S^{(i-1)} - (6 - 5\alpha_{3})S^{(i-2)} + S^{(i-3)} + (4 - 5\alpha_{4})S^{(i-4)} \right] \)
13. \( y_{i+T} = a_{i} + b_{i}T + c_{i}T^{2} \) // Predicted value of period \( T \) in the future

A. ARED PACKET DROP PROBABILITY BASED ON THE CUBIC FUNCTION

In the ARED algorithm, when the average queue length \( Q_{\text{avg}} \) is between the lower limit \( min_{th} \) and the upper limit \( max_{th} \) of the threshold, the packet drop probability \( P_{b} \) is linearly related to \( Q_{\text{avg}} \) [20]. However, when \( Q_{\text{avg}} \) is in the vicinity of \( min_{th} \), there are fewer data packets in the queue, the network is comparatively idle, and \( P_{b} \) can be slowly increased or decreased at this moment to increase the throughput of the network; when \( Q_{\text{avg}} \) is in the vicinity of the maximum threshold \( max_{th} \), there are more data packets in the queue, and \( P_{b} \) should be quickly adjusted at this moment to avoid network congestion. In this paper, the cubic function curve is used to conduct nonlinear smoothing processing on the packet drop probability function of ARED to more efficiently avoid network congestion and increase channel utilisation.

1) PACKET DROP PROBABILITY BASED ON THE CUBIC FUNCTION

The cubic curve is expressed as

\[
s(t) = a_{0} + a_{1}t + a_{2}t^{2} + a_{3}t^{3} \tag{11}
\]

The constraint conditions for the start time 0 and the end time \( T \) are set as \( s(0) = 0 \), \( s(T) = 1 \), \( s'(0) = 0 \), \( s'(T) = 0 \). The following are separately obtained:

\[
a_{0} = 0, \quad a_{1} = 0, \quad a_{2} = \frac{3}{T^{2}}, \quad a_{3} = -\frac{2}{T^{3}}.
\]

Therefore,

\[
s(t) = \frac{3}{T^{2}}t^{2} - \frac{2}{T^{3}}t^{3} \tag{12}
\]

which is translated to the right by \( a \) units to obtain

\[
s(t-a) = \frac{3}{T^{2}}(t-a)^{2} - \frac{2}{T^{3}}(t-a)^{3}, \quad t \geq a \tag{13}
\]

The cubic function curve after translation is shown in Figure 4.

In Equation (13), let \( t \) be the average queue length, \( a \) be the lower limit of the threshold \( min_{th} \), and \( T + a \) be the upper limit of the threshold \( max_{th} \). This equation is used to perform nonlinear processing on the part \( min_{th} \leq Q_{\text{avg}} \leq max_{th} \) in the ARED packet drop probability function to obtain the ARED packet drop probability function improved by the
hence, the value of $\maxp$ must be increased \[22\]. Let $\alpha$ be the increase factor, $\alpha = \min(0.01, \maxp/4)$, and $\beta$ be the decrease factor, which usually takes a value of 0.9. Different drop probability curves are obtained with different $\beta$ values of $\alpha$.

Based on the idea of the ARED algorithm, the target queue length, target, is set with a value range of \( [\text{min}h + 0.4(\maxh - \text{min}h), \text{min}h + 0.6(\maxh - \text{min}h)] \), and the relationship between target and $Q_{avg}$ is used to adaptively adjust $\maxp$ \[21\]. If $Q_{avg}$ is in the vicinity of $\text{min}h$ and $Q_{avg} < \text{target}$, the congestion adjustment is too active, and hence, the value of $\maxp$ must be decreased; if $Q_{avg}$ is in the vicinity of $\maxh$ and $Q_{avg} > \text{target}$, the congestion adjustment is too conservative, and the value of $\maxp$ must be increased \[22\]. Let $\maxp^+$ and $\maxp^-$ represent the maximum drop probabilities obtained after a radical increase and a conservative decrease, respectively, which are expressed as follows:

$$
\begin{align*}
\maxp^+ &= \maxp + \alpha \\
\maxp^- &= \maxp \times \beta
\end{align*}
$$

(15)

where $\alpha$ is the increase factor, $\alpha = \min(0.01, \maxp/4)$, and $\beta$ is the decrease factor, which usually takes a value of 0.9.

Different drop probability curves are obtained with different values of $\maxp$.

Figure 5 shows the variation curve of the packet drop probability $P_b$ with $Q_{avg}$ when $\maxp = 0.5$, $P_b = 0$ when $Q_{avg} < \text{min}h$, and $P_b = 1$ when $Q_{avg} > \maxh$. When $\text{min}h \leq Q_{avg} \leq \maxh$, $P_b$ varies nonlinearly with $Q_{avg}$. If there are fewer data packets in the queue, $Q_{avg}$ is small, and thus, $P_b$ increases relatively slowly; if the data packets in the queue increase continuously, $Q_{avg}$ is large, and thus, $P_b$ increases fast. Based on the ARED algorithm, the algorithm in this paper uses a nonlinear packet drop probability function, which can effectively solve the problem of excessively fast increase in the packet drop rate, does not easily cause queue oscillation, makes the traffic transmission more stable, increases channel utilisation, and resolves the parameter sensitivity of RED, controlling network congestion very well.

B. PROCESS OF THE IMPROVED ARED ALGORITHM BASED ON THE CUBIC FUNCTION OF TRAFFIC PREDICTION

In this paper, the DE algorithm is used first to optimise the smoothing coefficient $\alpha$ in the dynamic triple exponential smoothing model to obtain the predicted results of the traffic. Then, the queue management algorithm is executed and, based on ARED, the cubic function is employed to perform nonlinear processing on the packet drop probability function. Finally, the predicted traffic results are sent to the queue management algorithm, and the improved ARED algorithm based on the cubic function of traffic prediction is proposed. Figure 6 is the flow chart of the algorithm.

Table 3 is the process of the improved ARED algorithm based on the cubic function of traffic prediction. Here, $q$ is the predicted value in Algorithm 2, $Q_{avg}$ is the average queue length, the weight value is $w_q = 0.05$, and $interval$ is the time interval for the change in drop probability.

VI. PERFORMANCE SIMULATION AND ANALYSIS

A. SIMULATION OF THE TRAFFIC PREDICTION MODEL WITH OPTIMISED TRIPLE EXPONENTIAL SMOOTHING

The degree of self-similarity of the traffic is represented by the Hurst parameter value $H$. $0 < H \leq 0.5$ means the network flow has short-range dependence; $0.5 < H < 1$ indicates that the network traffic has self-similarity and long-range dependence, and a larger $H$ value indicates higher dependence and burstiness in the data stream \[23\]. In this paper, the original data stream of the satellite network were simulated in Matlab. The R/S analysis method was used to estimate the Hurst parameter $H = 0.835$ of the original data, showing that the original data stream had high data self-similarity and long-range dependence \[25\]. The prediction model with optimised triple exponential smoothing was used to predict the satellite network traffic and was compared with the autoregressive
TABLE 3. The process of the improved ARED algorithm based on the cubic function of traffic prediction.

| Algorithm 3: ARED algorithm improved by the cubic function |
|---------------------------------|
| **Input:** Predicted value of the initial queue length, \( q \) |
| **Output:** Packet drop probability \( P_d \) |
| 1. \( q \leftarrow \text{Ypred}/\text{The initial value is the predicted value of Algorithm 2} \) |
| 2. \( \text{count} \leftarrow -1 \) |
| 3. \( \text{target} \leftarrow \lfloor \text{min}_a + 0.4(\text{max}_a - \text{min}_a), \text{min}_a + 0.6(\text{max}_a - \text{min}_a) \rfloor \) |
| 4. \( a \leftarrow \min(0.01, \text{maxp}/4) \) |
| 5. \( \beta \leftarrow 0.9 \) |
| 6. for each packet arrived |
| 7. calculate new \( Q_{avg} \): |
| \( Q_{avg} \leftarrow (1 - \omega_3)Q_{avg} + \omega_3 \times q \) |
| 8. every interval seconds |
| 9. if \( Q_{avg} < \text{min}_a \) then |
| 10. if \( \text{maxp} < -0.5 \) then |
| 11. \( p \leftarrow \text{max} p + \alpha \gamma \) //Increase the packet drop probability |
| else if \( Q_{avg} > \text{max}_a \) then |
| 13. \( \text{maxp} \leftarrow \text{maxp} \times \beta \gamma \) //Decrease the packet drop probability |
| 14. end |
| 15. if \( Q_{avg} < \text{min}_a \) then |
| 16. \( P_d = 0 \); |
| 17. else if \( \text{min}_a < Q_{avg} < \text{max}_a \) then |
| 18. \( \text{count} = \text{count} + 1 \); |
| 19. \( P_d = \max p \text{maxp} \) |
| 20. else if \( Q_{avg} = \text{max}_a \) then |
| 21. \( P_d = 1 \); |
| 22. end |

The fitting results of the traffic predicted by the different models are shown in Figure 7. Clearly, the prediction model with optimised triple exponential smoothing of this paper significantly outperforms the other two prediction models in terms of fitting the original data and thus are able to more accurately predict the magnitude and trend of the traffic.

The MSE errors of the three traffic prediction models are compared in Figure 8. The prediction error of the ARMA model is the highest, followed by that of the prediction model with traditional triple exponential smoothing, and the error of the prediction model with optimised triple exponential smoothing is significantly lower than that of the other models, showing that the prediction model with optimised triple exponential smoothing has a high prediction accuracy.

Table 4 lists the MSEs of the three algorithms. The MSE error of the algorithm proposed in this paper is smaller than that of the other two algorithms. Hence, the proposed algorithm has the highest accuracy.

B. SIMULATION WITH THE IMPROVED ARED ALGORITHM BASED ON THE CUBIC FUNCTION OF TRAFFIC PREDICTION

Simulation is performed using the improved ARED algorithm based on the cubic function of traffic prediction and compared with that using the ARED algorithm and the ARED algorithm based on traffic prediction. The simulation parameters are shown in Table 5.

1) AVERAGE QUEUE LENGTH

Figure 9 shows the curves of the average queue length for the three algorithms. Between 9 s and 14 s, the average queue length of the algorithm proposed in this paper is significantly longer than that of the other two algorithms. This result is found because data bursts occur in the self-similar traffic.
on the network within this time range, and the proposed algorithm is able to maintain a long queue length in the case of bursts in self-similar traffic to stabilise the transmission performance of the network.

The mean and standard deviation of the average queue length for each of the three algorithms are provided in Table 6. In the presence of data burst in the self-similar traffic, to stabilise the network transmission performance, the mean and standard deviation of the average queue length for the algorithm proposed in this paper are larger than those of the other two algorithms.

2) PACKET DROP RATE

The packet drop rate refers to the ratio of the amount of data dropped to the total amount of data sent in the data transmission process. The curves of the packet drop rates for the three algorithms are shown in Figure 10. Between 9 s and 14 s, the ARED algorithm and the ARED algorithm based on traffic prediction both have a high packet drop rate, while the packet drop rate of the algorithm proposed in this paper is significantly lower than that of the other two algorithms. This result is found because data bursts occur in the self-similar traffic on the network within this time range, and the proposed algorithm is able to maintain a low packet drop rate in the presence of bursts in self-similar traffic.

The mean and standard deviation of the packet drop rate for each of the three algorithms are listed in Table 7. The mean
and standard deviation of the packet drop rate for the ARED algorithm are the largest, followed by those for the ARED algorithm based on traffic prediction, and the mean and standard deviation of the packet drop rate for the algorithm proposed in this paper are significantly smaller than those of the other two algorithms, indicating that the improved ARED algorithm based on the cubic function of traffic prediction in this paper has a lower packet drop rate and a better congestion control performance.

3) THROUGHPUT

Throughput refers to the number of data packets transmitted by the link within a unit time. It is an important indicator reflecting the network transmission capacity. Figure 11 shows the curves of the throughputs for the three algorithms. Between 9 s and 14 s, the ARED algorithm and the ARED algorithm based on traffic prediction have low throughputs, whereas the algorithm proposed in this paper has a high throughput. The reason for this difference is that data bursts occur in the self-similar traffic on the network within that time range, and the proposed algorithm is able to maintain a high throughput in the presence of bursts in self-similar traffic.

The mean and standard deviation of the throughput for each of the three algorithms are given in Table 8. The mean of the throughput for the algorithm proposed in this paper is larger than that of the other two algorithms. Because the proposed algorithm is able to maintain a higher throughput in the presence of data bursts in self-similar traffic, the standard deviation of its throughput is larger than that of the other two algorithms, indicating that the improved ARED algorithm based on the cubic function of traffic prediction proposed in this paper can better control the network congestion caused by self-similar traffic.

VII. CONCLUSION

In this paper, a prediction model with optimised triple exponential smoothing is established, and the results of network traffic prediction are introduced into the queue management algorithm to propose an improved ARED algorithm based on the cubic function of traffic prediction. The proposed prediction model has a higher prediction accuracy. In the presence of data bursts in self-similar traffic, the improved ARED algorithm based on the cubic function of traffic prediction outperforms the ARED algorithm and the ARED algorithm based on traffic prediction in terms of packet drop rate and throughput performance, thereby effectively controlling the network congestion caused by self-similar traffic.

ACKNOWLEDGMENT

The authors would like to thank Shenyang Aerospace University for its support.

REFERENCES

[1] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, “On the self-similar nature of Ethernet traffic,” ACM SIGCOMM Comput. Commun. Rev., vol. 49, no. 5, pp. 56–62, 2019.
[2] H. Elagha and M. Alshafee, “On the self-similar nature of ATM network traffic,” Issues Informing Sci. Inf. Technol., vol. 4, pp. 641–649, Feb. 2007.
[3] A. Domanski, J. Domanski, K. Fiuls, J. Szygula, and T. Czachorski, “Self-similar Markovian sources,” Appl. Sci., vol. 10, no. 11, p. 3727, 2020.
[4] X. Zhang, J. Feng, Z. Yu, Z. Hong, and X. Yun, “Prediction of the slope solute loss based on BP neural network,” Comput., Mater. Continua, vol. 69, no. 3, pp. 3871–3888, 2021.
[5] L. Chen and S. Zhou, “Sarse algorithm for robust LSSVM in primal space,” Neurocomputing, vol. 275, pp. 2880–2891, Jan. 2018.
[6] L. M. L. De Campos, “Time series prediction with direct and recurrent neural networks,” Turkish J. Forecasting, vol. 1, no. 1, pp. 7–15, 2017.
[7] F. Li, J. Sun, M. Zukerman, Z. Liu, Q. Xu, S. Chan, G. Chen, and K.-T. Ko, “A comparative simulation study of TCP/AQM systems for evaluating the potential of neuron-based AQM schemes,” J. Netw. Comput. Appl., vol. 41, pp. 274–299, May 2014.
[8] F. Jiang, C. Feng, C. Zhu, and Y. Sun, “Performance analysis of active queue management algorithm based on reinforcement learning,” J. Européen des Systèmes Automatisés, vol. 53, no. 5, pp. 637–644, Nov. 2020.
[9] S. Floyd. (2000). Recommendation on Using the Gentle Variant of RED, March 2000. [Online]. Available: http://www.icir.org/floyd/red/gentle.html
[10] L. Tan, G. Peng, and S. Chan, “Adaptive REM: Random exponential marking with improved robustness,” Electron. Lett., vol. 43, no. 2, pp. 131–133, 2007.
[11] K. Nichols and V. Jacobson, “Controlling queue delay ACM queue magazine: Networks,” vol. 10, no. 5, pp. 68–81, 2012.
[12] A. S. Belamfedel, E. H. Tissir, and N. Chaibi, “New design of anti-windup and dynamic output feedback control for TCP/AQM system with asymmetrical input constraints,” Int. J. Syst. Sci., vol. 52, no. 9, pp. 1822–1834, 2021.
[13] H.-P. Schwefel, I. Antonios, and L. Lipsky, “Understanding the relationship between network traffic correlation and queueing behavior: A review based on the N-burst ON/OFF model,” Perform. Eval., vol. 115, pp. 68–91, Oct. 2017.
[14] G. Zhu, L. Li, Y. Zheng, X. Zhang, and H. Zou, “Forecasting influenza based on autoregressive moving average and holt-winters exponential smoothing models,” J. Adv. Comput. Intell. Intell. Inform., vol. 25, no. 1, pp. 138–144, 2021.
[15] H. K. Sharma, K. Kumari, and M. Jamir, “Short-term forecasting of air passengers based on hybrid rough set and double exponential smoothing models,” Intell. Automat. Soft Comput., vol. 25, no. 1, pp. 1–14, 2019.
[16] D. A. Adeyinka and M. Nazeem, “Time series prediction of under-five mortality rates for Nigeria: Comparative analysis of artificial neural networks,” J. Netw. Comput. Appl., vol. 41, no. 3, pp. 3871–3888, 2021.
[17] C. Deng, X. Zhang, Y. Huang, and Y. Bao, “Equipping seasonal exponential smoothing models with particle swarm optimization algorithm for electricity consumption forecasting,” Energies, vol. 14, no. 13, p. 4036, 2021.
[18] W. Šulandari and P. C. Rodrigues, “Exponential smoothing on modeling and forecasting multiple seasonal time series: An overview,” Fluctuation Noise Lett., vol. 20, no. 4, Aug. 2021, Art. no. 2130003.
YUXIA BIE was born in China, in 1981. She received the B.S. and M.S. degrees in communication engineering from Shenyang Ligong University, Shenyang, China, and the Ph.D. degree in control science and engineering from the Nanjing University of Science & Technology, Nanjing, China, in 2014. From 2008 to 2016, she worked with the Key Laboratory of Communication and Network, Dalian University. Since 2017, she has been a Teacher with the College of Electronic and Information Engineering, Shenyang Aerospace University. She presides the National Natural Science Foundation Project of China, a Doctoral Startup Foundation of Liaoning Province, and a General Project of Liaoning Provincial Committee of Education. She is the author of more than 20 articles and holds six patents. Her research interests include satellite networks, data multiplexing, machine learning, and UAV communication. Since 2012, she has been a member of the China Computer Association.

ZHIHAN LI was born in Fuxin, Liaoning, China, in 1996. She received the B.E. degree in communication engineering from the City Institute, Dalian University. Since 2017, she has been a Teacher with the College of Electronic and Information Engineering, Shenyang Aerospace University. Her research interests include satellite networks traffic prediction algorithms and queue management algorithms.

ZHI HU was born in Shenyang, Liaoning, China, in 1978. He received the B.S. degree in computer science and technology from Northeastern University, Shenyang, China, the M.S. degree in computer software and theory from Shenyang Ligong University, Shenyang, and the Ph.D. degree in technology of computer application from Northeastern University, in 2016. From 2006 to 2016, he was a Researcher with Neusoft Research, Shenyang. Since 2017, he has been a Teacher with the Software College, Shenyang Normal University. He is the author of six articles. His research interests include ad hoc networks, network routing, and machine learning. He was a recipient of the Liaoning Prize for Science and Technology in 2012.

JIAMEI CHEN was born in Chifeng, Inner Mongolia, China, in 1985. She received the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 2015. She studied at Purdue University as a Visiting Scholar, from 2011 to 2012. She is currently working as an Assistant Professor with the College of Electrical and Information Engineering, Shenyang Aerospace University. Her research interests include heterogeneous wireless networks, resource allocation, and green communication.