A Spatial Information Network Traffic Prediction Method Based on Hybrid Model

Jingling Li, Yi Zhang, Wei Liang, Tao Cui, Jun Li

***Abstract*—**Compared with terrestrial network, the traffic of spatial information network has both self-similarity and short correlation characteristics. By studying its traffic prediction method, the resource utilization of spatial information network can be improved, and the method can provide an important basis for traffic planning of a spatial information network. In this paper, considering the accuracy and complexity of the algorithm, the spatial information network traffic is decomposed into approximate component with long correlation and detail component with short correlation, and a time series hybrid prediction model based on wavelet decomposition is proposed to predict the spatial network traffic. Firstly, the original traffic data are decomposed to approximate components and detail components by using wavelet decomposition algorithm. According to the autocorrelation and partial correlation smearing and truncation characteristics of each component, the corresponding model (AR/MA/ARMA) of each detail component can be directly established, while the type of approximate component modeling can be established by ARIMA model after smoothing. Finally, the prediction results of the multiple models are fitted to obtain the prediction results of the original data. The method not only considers the self-similarity of a spatial information network, but also takes into account the short correlation caused by network burst information, which is verified by using the measured data of a certain back bone network released by the MAWI working group in 2018. Compared with the typical time series model, the predicted data of hybrid model is closer to the real traffic data and has a smaller relative root means square error, which is more suitable for a spatial information network.

***Keywords*—**Spatial Information Network, Traffic prediction, Wavelet decomposition, Time series model

I. Introduction

THE spatial information network is a network system that acquires, transmits and processes spatial information in real time using various space platforms such as synchronous orbit satellites, medium orbit, low orbit satellites, stratospheric balloons, manned or unmanned aerial vehicles. The space information network supports high-dynamic and broadband real-time transmission of earth observation, and supports ultra-long-range and large-delay reliable transmission of deep space exploration. It can serve major applications such as ocean navigation, emergency rescue, navigation and positioning, air transportation, and space measurement and control. In practical applications, reasonable traffic planning and prediction can optimize network traffic allocation and improve network resource utilization. Especially for spatial information networks with severe resource constraints, traffic prediction methods are one of the key components of network system design. [1], [2]

The spatial information network business mainly includes conventional satellite communication services and some special communication services. Ekici and other scholars have tracked the backbone network traffic such as the US satellite backbone network and Abilene (American Education Network) [3]: "Since the satellite network is similar in function to the terrestrial network, the traffic carried by both networks has a similar network characteristic". At the same time, there are many other factors affecting spatial information network traffic. In addition to the similar factors to the terrestrial network, these factors include the physical impact of the space environment and various sudden factors. However, taking all factors into consideration will inevitably lead to an increase in algorithm complexity and resource cost, and a compromise between algorithm accuracy and computational complexity can be made according to user requirements.

At present, the traffic models used by terrestrial networks mainly include Least Mean Square (LMS) filters, Kalman filters, neural network models [4], autoregressive (AR) or autoregressive moving average models (ARMA) [5], autoregressive integral moving average (ARIMA) [6], [7], and fractal autoregressive summation sliding average (FARIMA) model [8], [9], combined with wavelet transform prediction model [10], etc. Among them, the LMS filter is not suitable for describing unstable network traffic. The Kalman filter requires the process statistics before implementing prediction, and the real-time prediction of the spatial information network traffic cannot be guaranteed. The neural network or the deep learning algorithm is suitable for describing the instability of the traffic. However, a large number of training sequences are required, and the complexity for the resource-constrained space network is too high. FARIMA is used to capture long correlation characteristics, but the calculation amount is large, and is not suitable for current space applications.

This paper comprehensively considers the complexity of network traffic prediction algorithm and model accuracy. Based on the self-similarity and short correlation of spatial information network traffic, a hybrid time-series traffic prediction method based on wavelet transform is proposed. The method can improve the accuracy of time series traffic prediction, effectively control the computational complexity, and is more suitable for traffic prediction in spatial information networks.

Jingling Li is with the National Key Laboratory of Science and Technology on Space Microwave, China Academy of Space Technology (Xi'an), Xi'an 710100, China (phone:086-029-89253208; e-mail: ljlspirit@foxmail.com).

Yi Zhang and Wei Liang are with the National Key Laboratory of Science and Technology on Space Microwave, China Academy of Space Technology (Xi'an), Xi'an 710100, China (e-mail: lbzytogether@163.com, sw.dong@163.com).

II. Characteristic Analysis

Although the traffic of the spatial information network has similar network characteristics to the ground network traffic, it has different characteristics from the ground. The main business traffic of the spatial information network can be divided into relay, communication and digital transmission businesses. For the remote sensing satellites, the main business traffic is various image data, which has large data volume and low real-time performance, and it is also highly correlated. And the business traffic of communication satellites or relay satellites has strong suddenness and real-time.

In general, spatial information network traffic has self-similarity and impulsiveness [11].

The business traffic description of the network can be mathematically represented as a random process of a time series. Set $X = \{X_n, n = 1,2,3, \ldots\}$ to be a generalized stationary process, where X_n is the time series of traffic, which is stationary. X_n represents the number of service packets arriving at the network in the nth unit time interval. If the total number of packets arriving from 0 to t is expressed as $N|t|$, then X_n can be expressed as:

$$X_n = N\left|nT\right| - N\left|(n-1)T\right| \tag{1}$$

In the case of a known time series X_n, non-overlapping combination is performed to X_n, and then divide the sequence into blocks in order, as shown in (2).

$$\begin{aligned} X_n^{(m)} &= \frac{X_{nm-m+1} + X_{nm-m+2} + \ldots X_{nm}}{m} \\ &= \frac{1}{m}\sum_{i=nm-m+i}^{nm} X_i \end{aligned} \tag{2}$$

The process shown in (2) is a time series aggregation process. If the aggregation process $X^{(m)} = \{X^{(m)}, n = 1,2, \ldots\}$ does not change with the magnitude of m, then the time series X has a self-similarity characteristic. Self-similarity is one of the long-correlation properties, but time series exhibiting long correlations are not all self-similar.

$$\begin{aligned} f(\lambda) = \{&\left|\lambda\right|^{-2H-1} + \sum_{j=1}^{\infty}[(2\pi j+\lambda)^{-2H-1} + (2\pi j-\lambda)^{-2H-1}]\} \\ &\times c(1-\cos\lambda) \end{aligned} \tag{3}$$

As shown in (3), $f(\lambda)$ represents the power spectral density of the stationary time series, H represents the Hurst parameter of the self-similar process, when $H \in (0.5,1)$, it can be determined that a time series has self-similarity. At the same time, its value also reflects the degree of self-similarity of network traffic.

Through the above analysis, the following conclusions can be drawn: the spatial information network traffic has periodicity, suddenness, and long correlation under long-term span and is non-stationary. In the short-term, the spatial information network traffic changes little and is basically stable.

In this paper, the method of time series hybrid model based on wavelet transform can respectively model the long periodicity part and the short stable part of network traffic, which is more consistent with the traffic characteristics of the spatial information network.

III. Hybrid Model and Prediction Process

A. Traffic Prediction Algorithm Design

Fig. 1 shows the block diagram of the time series model establishment and prediction process in the spatial information network. The algorithm flow for traffic prediction is as follows:

1) The original traffic is decomposed by db3 wavelet decomposition algorithm to make it de-correlated, then obtain the approximate component a1 and the detail components d1, d2, and d3. Each of its detail components has exhibited a short correlation, while the approximate component still has the same long correlation as the original signal.
2) Model the approximate component with long correlation properties using ARIMA to obtain an approximate component model A_1. And the detail components with short correlation characteristics are modeled by AR/MA/ARMA algorithms to obtain the detail component models D_1, D_2, D_3.
3) The predicted components a_1' and d_1', d_2', d_3' are obtained by the models A_1, D_1, D_2 and D_3.
4) Fitting the predicted components to obtain predicted data of the original data.

B. Specific Algorithm Description

The prediction process of the approximate component and the detail component in Fig. 1 is similar. The difference is that the approximate component is non-stationary data, and it needs to be smoothed, and the detail components do not need to be processed. The key steps of the algorithm are described in detail below:

1. Discrete Wavelet Decomposition

Firstly, wavelet transform is performed on the spatial self-similar network data. And the traffic is decomposed over different frequency domains. Let $\varphi(t) \in L^2(R)$, whose Fourier transform is $\hat{\varphi}(\omega)$, and when $\hat{\varphi}(\omega)$ satisfies the complete reconstruction condition or the identity resolution condition:

$$\int \frac{\left|\hat{\varphi}(\omega)\right|^2}{|\omega|} d\omega < \infty \tag{4}$$

Then the function $\varphi(t)$ can be called a mother wavelet or a base wavelet. The following formula is obtained by panning and stretching the wavelet $\varphi(t)$.

$$\varphi_{a,b}(t) = |a|^{\frac{1}{2}} \varphi\left(\frac{t-b}{a}\right) \tag{5}$$

This paper adopts the fast-binary discrete wavelet transform, namely the Mallat algorithm [12]. The Mallat wavelet decomposition algorithm can be regarded as the frequency division of the original traffic through the high frequency and low frequency filters, and the multi-scale wavelet decomposition and reconstruction formulas are as follows:

Decomposition formulas:

$$a_k^{j+1} = \sum_{l \in Z} a_l^j \left(\phi_{j+1,k}(x), \phi_{j,l}(x) \right) \tag{6}$$

$$d_k^{j+1} = \sum_{l \in Z} c_l^j \left(\varphi_{j+1,k}(x), \varphi_{j,l}(x) \right) \tag{7}$$

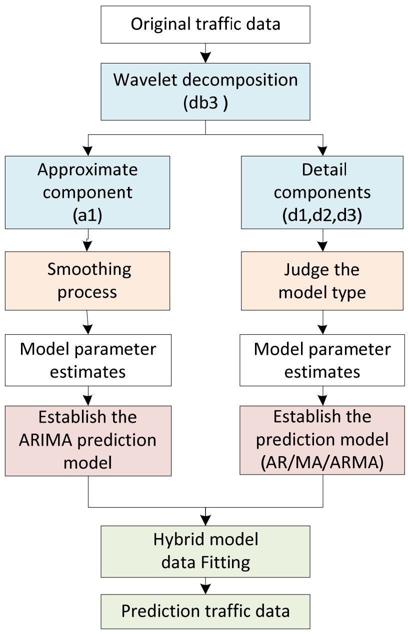


Fig. 1 Time Series Model Establishment and Prediction Process

Reconstruction formula:

$$a_k^j = \sum_{l \in Z} a_l^{j+l} \left(\phi_{j,k}(x), \phi_{j+l,l}(x) \right) + \sum_{l \in Z} d_l^{j+l} \varphi_{j,k}(x), \varphi_{j+l,l}(x) \tag{8}$$

The Mallat wavelet decomposition algorithm can decompose the original signal of the spatial information network into an approximate component (i.e. a low frequency part) and a detail component (i.e. a high frequency part). The approximate component reflects the overall contour characteristics and trends of network traffic, the detail components reflect the effects of dynamic factors such as random disturbances of network traffic.

2. Judge the Prediction Model Type

According to the autocorrelation and partial correlation tailing and truncation characteristics of different detail components, it can be judged which type of model is used for modeling (including AR, MA and ARMA time series models), as shown in Table I.

TABLE I
PARTIAL CORRELATION AND AUTOCORRELATION FUNCTIONS OF THREE MODELS

Model	AR(p)	MA(q)	ARMA(p,q)
Partial correlation	truncation	tailing	tailing
autocorrelation	tailing	truncation	tailing

3. Determine the Model Order

Take the ARIMA model as an example. In the model $\mathrm{ARIMA(p, d, q)}$, the AIC criterion is used to determine the order p and q of the model mainly by judging whether the transformation process of the predictor is closest to stochastic process in the autoregressive model. If the upper bound P_0 of p and the upper bound Q_0 of q are known, for each pair$\mathrm{(k, j)}$, $0 \leq \mathrm{k} \leq P_0, 0 \leq \mathrm{j} \leq Q_0$, the AIC function can be calculated as:

$$AIC(k,j) = \ln\left(\sigma^2(k,j)\right) + \frac{2(\mathrm{k+j})}{N} \tag{9}$$

where σ^2 represents an estimate of the white noise variance. The minimum point $\mathrm{(p, d)}$ of $\mathrm{AIC(k, j)}$ is called the AIC order of point $\mathrm{(p, d)}$. If the minimum value is not unique, first select the minimum value corresponding to smallest $(\mathrm{k} + \mathrm{j})$, and then take the minimum value corresponding to smallest j.

4. Hybrid Prediction Model

In this paper, the wavelet transform is used to reveal the global and local characteristics of the spatial information network traffic, and reduce the self-similarity of signals in the frequency domain. The decomposed detail component signals exhibit a short correlation, reflecting the effects of dynamic factors such as random perturbations or abrupt changes, while the long correlation approximation component signals preserve the contour features and trends of the original stream. In this way, it is equivalent to modeling and predicting the initial traffic through a multi-class hybrid time series model. And the prediction results of different components are fitted to obtain the prediction traffic data, which can improve the flow prediction accuracy.

The approximate component is modeled using the ARIMA model $\mathrm{ARIMA(p, d, q)}$, which can be described as:

$$\Phi(B)\nabla^d X_t = \Theta(B)\varepsilon_t \tag{10}$$

$\Phi(\mathrm{B})$ and $\Theta(\mathrm{B})$ represent polynomials of degree p and number q, whose expression is:

$$\Phi(B) = 1 - \Phi_1(B) - \ldots - \Phi_p(B)^p \tag{11}$$

$$\Theta(B) = 1 + \Phi_1(B) + \ldots + \Phi_q(B)^q \tag{12}$$

In addition, $\mathrm{B}x_t = x_{t-1}$, B is called the delay factor, and the difference factor of d is ∇^d, the relationship between B and ∇^d is$\nabla = 1 - \mathrm{B}$. The binomial expansion of ∇^d is:

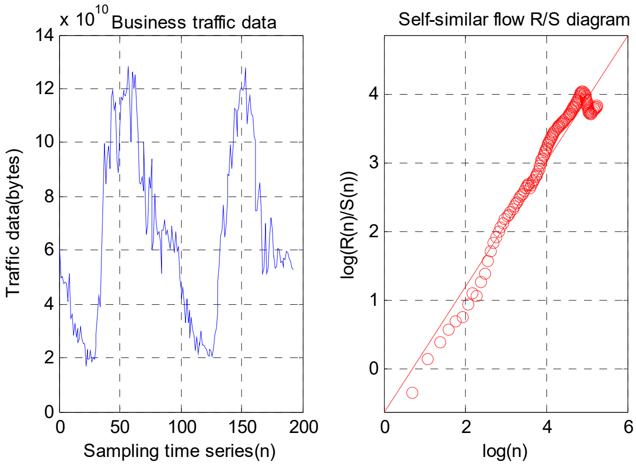
$$\nabla^d = (1-B)^d = \sum_{k=0}^{\infty} \begin{bmatrix} d \\ k \end{bmatrix} (-B)^k \tag{13}$$

where $\begin{bmatrix} d \\ k \end{bmatrix} = \frac{d(d+1)\dots(d+k-1)}{k!}$. For the detail components, after judging the model type by step (3), it is modeled and predicted according to the time series model method.

IV. Simulation and Verification

A. Simulation Model

This paper uses the MAWI working group [13] to verify the traffic flow data collected on the ground backbone network. The simulated traffic data used were collected from May 9, 2018 to May 10, 2018 for a total of 48 hours. Counting once every 15 minutes, for a total of 192 data. It is calculated that the average speed of the ground backbone network reaches 660 Mbps, which can simulate the traffic flow data of the backbone satellite network. The data of the network traffic is shown in Fig. 2(a).



(a) Original network traffic data (b) R/S value of traffic

Fig. 2 Network traffic figure

Observing the network traffic shown in Fig. 2(a), it can be found that the traffic is non-stationary, with obvious burstiness and periodicity. The R/S value of the network traffic is obtained by R/S estimation method, and the parameter H value is determined by the slope of the straight line fitted by the least squares method, and the self-similarity of the network traffic can be verified, as shown in Fig. 2 (b). Its Hurst parameter is $H \approx 0.8093$, and it can be seen from $H \in (0.5,1)$ that the network traffic exhibits obvious self-similarity characteristics.

The db3 algorithm is used for wavelet decomposition, and the result is shown in Fig. 3. The approximate component conforms to the original traffic fluctuation trend, and the detail component reflects the short-term variation of the original traffic. The autocorrelation and partial correlation properties of each detail component after wavelet decomposition are analyzed. As shown in Fig. 4, the detail components after wavelet decomposition show the tailing characteristics of both autocorrelation and partial correlation estimation, which are modeled by ARMA model. The approximate component is modeled using the ARIMA model.

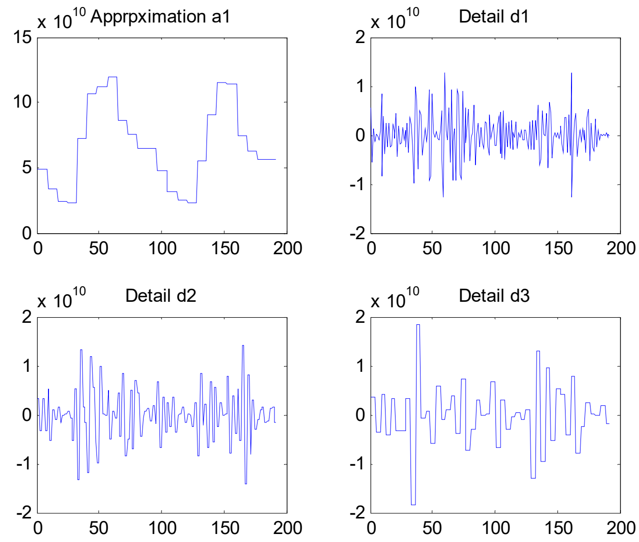


Fig. 3 Approximate component and detail components obtained by wavelet decomposition

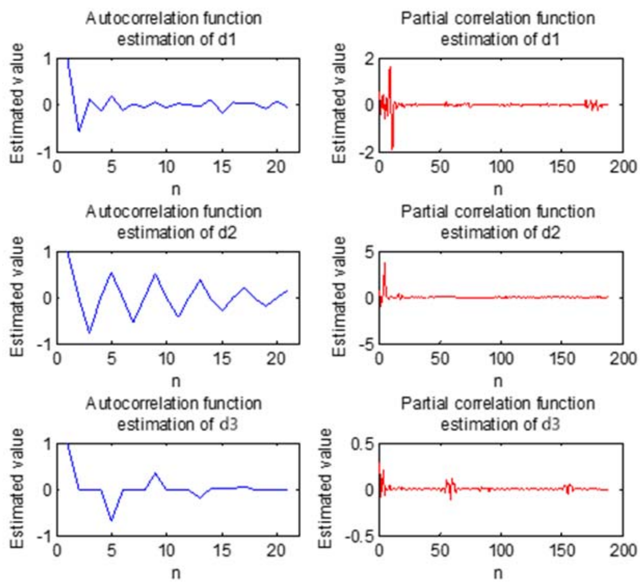


Fig. 4 Autocorrelation and partial correlation estimation of components

The prediction step size of the time series models is set to 1, the training sequence length is set to 184, the prediction sequence length is 8, and the prediction models are established for the approximate component and the detail component, respectively. The prediction results of each component are compared with the decomposition components as shown in Fig. 5.

As shown in Fig. 6, the prediction data obtained by the ARIMA model and the hybrid model are compared with the original data. It can be seen from Fig. 5 and Fig. 6 that comparing with the ARIMA model, the predicted value of the hybrid model is more in line with the change trend of the original flow.

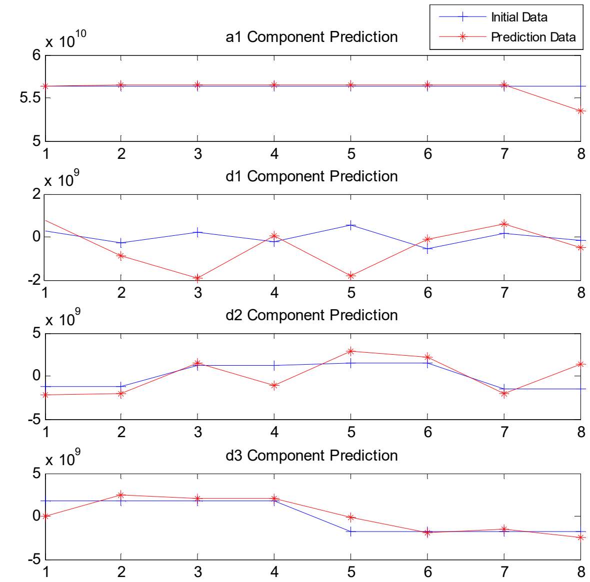


Fig. 5 The prediction data compared with the original data of each component

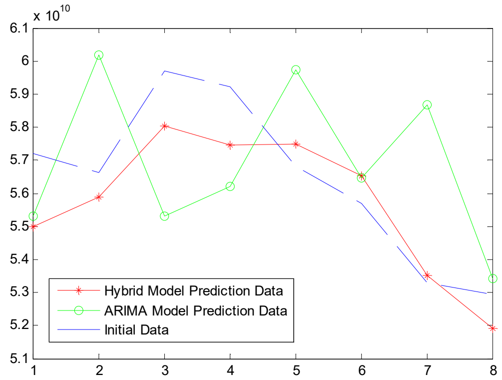


Fig. 6 The prediction data compared with the original data

B. Model Performance Evaluation

The accuracy of the prediction results is measured by the relative root mean square error (RRMSE) parameter. The smaller the value of RRMSE, the higher the prediction accuracy of the model, as shown in the following formula:

$$RRMSE=\sqrt{\frac{\sum_{t=1}^{M} abs\left(X\left(t\right)-X'\left(t\right)\right)}{M\times X\left(t\right)}} \tag{14}$$

where, X(t) is the actual flow value, X́(t) is the predicted flow value, and M is the predicted data quantity.

The RRMES parameter of the predicted data obtained by the ARIMA model is calculated to be 2.2198%, and the RRMES parameter of the predicted data obtained by the hybrid model is 1.4132%. It can be seen that the hybrid time series model proposed in this paper has better prediction accuracy than the ARIMA time series model.

V. CONCLUSION

In this paper, a hybrid time series model is proposed to model and predict spatial information network traffic. Through simulation analysis and performance evaluation, it is proved that compared with the ARIMA model, the proposed hybrid model has better prediction accuracy than the ARIMA model in the case of increasing the computational complexity of the algorithm. This model can provide an important basis for traffic planning, routing design and network behavior monitoring of spatial information networks.

ACKNOWLEDGMENT

This work is supported by the Supported Foundation by National Key Lab. Of Science and Technology on Space microwave (6142411182103).

REFERENCES

[1] Hadjadi Aoul Y, Ahmed T. Challenges, opportunities, and solutions for converged satellite and terrestrial networks (J). IEEE Wireless Communications Magazine, 18(1):46-52, 2011.
[2] Wang H. Efficient and robust traffic engineering in a dynamic environment (D). New Haven, USA: Yale University, 2008
[3] Ekici E, Akyildiz I F, Micheal D B. Datagram routing algorithm for LEO satellite networks(C)//Proceedings of 2000 IEEE International Conference on Computer Communications. Tel-Aviv, Israel: IEEE, 2000. 500-508.
[4] Matyas W, Liang Y. Predicted sum: a robust measurement-based admission control with online traffic prediction (J). IEEE Communications Letters, 11(2):204-206, 2007.
[5] Deng Zi-li, Optimal Estimation Theory with Applications: Modeling, Filtering, and Information Fusion Estimation (M), Harbin Institute of Technology Press, 2005.
[6] George B, Gwilym J, Gregory R, Time Series Analysis: Forecasting and control (M), Posts & Etlecom Press, 2005.
[7] Du Wei. Self-Similar Traffic Prediction Scheme for Satellite Internet Services (D). School of Electronics and Information Technology, 2014.
[8] Corradi M, Garroppo R.G, Giordano Pagano M. Analysis of F-ARIMA processes in the modeling of broadband traffic (A). IEEE International Conference on Communications (C). 3(1): 964-968, 2000.
[9] Chen Xiao-tian, Liu Jing-Xian, Network traffic prediction based on wavelet transformation and Fariima (J), Journal on Communications, 32(4): 153-165, 2011.
[10] Bai Xiang-Yu, YE Xin-Ming, Jiang Hai, Network Traffic Predicting Based on Wavelet Transform and Autoregressive Model (J), Computer Science, 34(7): 47-49, 2007.
[11] Wei De-bin, PAN Cheng-shen, Han Rui, Self-similarity Analysis of Satellite Network Traffic (J), 40(5):67-69, 2013.
[12] Yu I, Kim C.A Novel Shot-Term Load Forecasting Technique Using Wavelet Transform Analysis (J). Electric Machines and Power systems, 28:537-549, 2000.
[13] Wide Mawi Working Group: http://mawi.wide.ad.jp/mawi/ditl/ditl201, 2018.

Jing Lingli (M'06–SM'09) received a B.S. in Communication Engineering from Xidian university in 2006 and an M.S. in Communication and Information Systems from Xidian university in 2009. She became a Member (M) of WASET in 2019. She has been with China Academy of Space Technology (Xi'an) for more than 10 years. Her technical interest areas include mixed Optical/elctrical satellite switching technology and distributed cluster satellite network technology. She has designed or led the key technology of satellite swithing system for clients. She is now focused on the field of cluster satellite topology control and routing technology.

Yi Zhang received a B.S. in Communication Engineering from Xidian university in 2008 and an M.S. in Communication and Information Systems from China Academy of Space Technology (Xi'an) in 2007. She has been with

China Academy of Space Technology (Xi'an) for more than 12 years. Her technical interest areas include mixed Optical/elctrical satellite switching technology and satellite data processing technology.