Research on Optimal Control of Large File Access upon VPDN

Jing Feng, Kunpeng Jing,Yang Wu
Institute of Meteorology and Oceanography
PLA University of Science and Technology
Nanjing, China
e-mail:jfeng@seu.edu.cn

Xiaoxing Yu
Department of Information and Networks
TELECOM ParisTech (ENST)
Paris, France
e-mail:yu@telecom-paristech.fr

***Abstract*—Facing the optimal problem of meteorological large data files access upon Virtual Private Dial-up Networks (VPDN), transport back of reliable radar data and optimize retrieval of single-server is studied. Focusing on the requirements and radar data application, such as images split joint of multiple weather radars, the strategies of data division and storage are proposed, and an evaluation model named Quality of Experience based on Reliability and Time (QoE-RT) is presented. Consequently, Adaptive File Access Control (AFAC) algorithm is designed and implement, which is consisted of reliable file transmission, rapid file retrieval, file size adjustment and service quality evaluation. The performance is verified by experiments of the file retrieval and transmission in actual environment, which shows that the file access time is decreased by 20%, transport success rate is more than 95%, and the QoE is significantly improved.**

Keywords-optimal cotrol; file access; VPDN; QoE.

I. INTRODUCTION

In the field of meteorological and hydrological information, there are hundred gigabytes (GB) of data collected by satellite, ground observation stations, radar and numerical weather forecasting every day. The size of individual files might be hundreds of megabytes (MB); some raw satellite imagery data is up to a few GB. On the one hand, these data needs to transmit back from the detection/ observation points to the data center, and on the other hand professional users also need to get the national or global meteorological data. For motorized stations and mobile users, Virtual Private Dial-up Networks (VPDN), a mixed network of solid and mobile communication by means of Layer 2 Tunneling Protocol (L2TP) is often used. For such large files like radar mosaic data, satellite images and numerical forecast products, using lossless compression method, even if the compression ratio is 50%, the data size is still around dozens of megabytes, so it is difficult to solve the reliable transmission problems upon VPDN radically. It is necessary to find a new solution, which can conduce to the efficient querying and reliable file transfer.

Although there are many mainstream network storage technologies in large data centers, including Direct Attached Storage (DAS), Storage Area Network (SAN) and Network Attached Storage (NAS), etc.[1], the low-end server clusters are efficient solutions in primary sub-center concerned with space, funds and other conditions. When the network performance and server hardware performance cannot change, the server storage strategy and the setting of maximum file block not only are important factors which affect the Quality of Experience (QoE) of users, but also can be set by user and configurate parameters adaptively with control program. Therefore, how to provide the management efficiency of stand-alone storage is significant for improving server access and cluster configuration.

In general, the optimal control of large file access upon VPDN is rarely mentioned, and how to decide the transmission block size of large file is not explored yet. For example, the user cannot control the partition of file size in FTP, which is not easy to guarantee performance [2]. According to the business requirements, the characteristics of the channel and the network environment, the optimization method is studied for large file transfer by dividing big files into smaller block size.

Analyzing the classic New Technology File System (NTFS), the relationship between access efficiency of its file and directory structure was quantitatively analyzed to obtain the optimal results. File access time and transmission success rate were both the metrics of Quality of Experience based on Reliability and Time (QoE-RT) referring to psychology. Facing the transmission of weather radar data, an Adaptive File Access Control algorithm (AFAC) was developed with the optimal data block segmentation strategy.

The rest of the paper is organized as follows. The background and motivation is described in Section II; the strategies for big file optimal transport and storage are analyzed in Section III considering the influence of NTFS file directory structure and file size; in Section IV, AFAC algorithm is presented and evaluated by QoE-RT. Finally, the conclusion and future work are given in Section V.

II. BACKGROUND AND MOTIVATION

In remote areas with poor communication of terrestrial broadband, deploying meteorological equipment such as automatic weather stations, motorized radar, using wireless network to transmit data back is a convenient and economical way. Although the third-generation (3G) network is able to meet the basic requirements of small quantities of data transmission, it cannot guarantee the weather radar data (dozens or hundreds of megabytes) reliable transmission. The optimal control mechanisms of data transmission and access need to be studied between sub-center and measurement station to improve service abilities in limited condition.

 ISBN: 978-1-61208-478-7

A. VPDN–based Solid and Mobile Mixed Network

Motorized radar usually is deployed on demand with around 100KM detection range. Sometimes multiple radars compose an observation network, which requires each radar data must be transmitted to the data center within the stipulated time to split joint images. For this kind of field mobile environments, VPDN is a cost-effective way of transport, which combines solid lease line and public mobile network with designated access point, but data transmission is instability in wireless section.

Adopting tunnel technique, the user data is encapsulated by special layer-2 protocol and transported in VPDN [3]. For example, China Telecom employs L2TP to establish VPDN, and L2TP use the Point to Point Protocol (PPP) link layer protocol unit to load data. It allows the both connection endpoint of the Layer-2 link protocol and PPP serving at different devices via a packet-switched network, e.g., IP network, which extends the PPP model and enables PPP sessions to cross a frame relay network or the Internet [4]. The case of networking is shown in Figure 1.

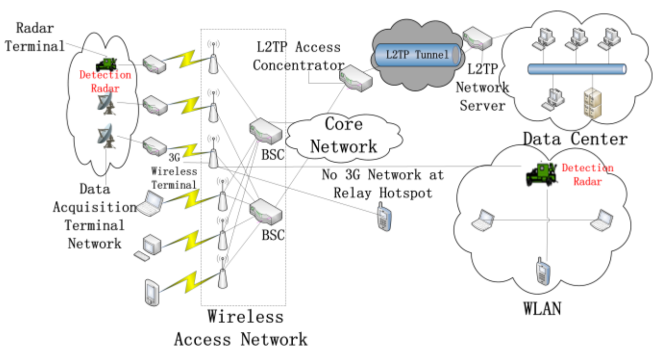

Figure 1. The case of VPDN topology with L2TP

After testing, a phenomenon was found: when the file size exceeds a certain threshold, the transmission failure rate greatly increased upon VPDN. Therefore, autonomous adjustment the size of the data block is the key to improve success transmission rate of big file.

B. Key Optimal Issues for Retrieval in Single-Server

In the context of this paper, the radar data is uploaded to the nearest data center/sub-centers, and is stored on the local servers. As valuable data, it will be archived and shared momentarily with the two service ways: Push and Pall. In practical work, we found that when the number of files on the server is more than ten thousand, query retrieval time has increased significantly. Meanwhile, the transport is also very difficult and often fails when the file size is over 1GB upon FTP, even in the LAN environment.

The efficiency of mobile terminals access data center and download big files via VPDN, will directly affect the user experience. If the reading time is denoted by T_s, it can be expressed in (1).

$$T_s = T_1 + T_2 + T_3 \tag{1}$$

T_1 is locating time of root directory that is related with the file system itself; T_2 is the time of traversing the index tree that is concerned with structure of the tree; T_3 concerns with the time of control and disk access related to hardware performance. Under the same file system and hardware, T_1 and T_3 are fixed basically, so T_2 is the only changeable factor that impacts on access efficiency.

III. Access Control Algorithm and Optimal Strategies of Big File

In this section, the influence factors are analyzed quantitatively from network behavior characteristics and the server's file system.

A. Ralationship betwween File Transport Success Rate and Data Size upon VPDN

Investigating VPDN fundamental network protocol, L2TP protocol consists of two components: control messages and data messages. Control messages establish and maintain the tunnel and connection session, and reliable transport is achieved by using flow control and congestion control. Data messages are encapsulated as the UDP packet in unreliable transmission over public networks.

L2TP packet encapsulation structure is shown in Figure 2.

IP header (Public network)	UDP header	L2TP header	PPP header	IP header (Private network)	Data

Figure 2. L2TP PDU format

From the analysis of L2TP, it is known that the user data will be regarded as a transparent load of UDP encapsulated even if it adopts TCP-based application layer protocols (e.g., FTP) over Internet, so once packet loss, the TCP retransmission message will be treated as unreliable transport. PPP sends data units by the order, and has higher loss rate while the data block is too large or the transport rate is too fast. Once the packet loss rate rises, it will lead to retransmission mechanism of TCP failed to achieve. The default number of TCP retransmissions is three, so if there is no acknowledgment message coming back, the TCP connection will be ended after three times of retransmission,

In order to successfully transport big files, the most direct way is to split big files into small pieces. However, dividing and merging the file will increase overhead, so we must find out the maximum size of the file block and have higher transport success rate. In the actual environment, 2G, 3G and 4G networks are mixed, China telecom 3G mobile devices were used in experiment scenario considering most area covered by 2G and 3G networks. China Telecom uses CDMA2000 standard, compatible with GSM, and the basic rate as shown in Table I.

TABLE I. China Telecom Rate of 2G and 3G

Standard	**2G**		**3G**
Technique	***GSM***	***CDMA2000***	***CDMA2000***
Down rate	236 kbps	153 kbps	3.1 Mbps
Up rate	118 kbps	153 kbps	1.8 Mbps

A set of data transport experiments were taken dividing the typical data file of C-band Doppler weather radar (110MB) into different size. The sending end accesses VPDN router (HUAWEI SRG1200) through WLAN router (HUAWEI EC177, supports CDMA2000 2G and 3G), and the VPDN router is located in our lab as a sub-center, which is linked China telecom via the leased line. For different length of data block (e.g., from 1MB, 2MB increase to 21M.) at different signal strength environments. Using FTP command, we tested the success rate by means of Wireshark which is a capture toolkit. The statistic results are shown in Figure.3.

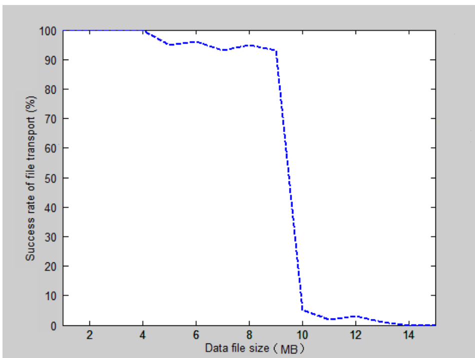

Figure 3. Success rate of different data size upon VPDN

The experimental results indicate that when the data block is less than 4MB, success rate of transport is 100%, and the success rate is more than 90% while the data block is between 4MB and 10MB. Especially, the success rate is less than 3% once the data block exceeds 10MB.

B. Relationship between NTFS File Directory structure and Query Efficiency

As described in Section 2-B, user-defined directory structure plays an important role in efficiency of local query access, and there is a large space-optimized for widely used working group server. Mainstream local file systems include NTFS, File Allocation Table (FAT) and Extended File System (EXT). Their main feature is shown in Table II. NTFS file system has maximum data upper limit, minimum index complexity, and is currently the most widely used local file storage system as well.

TABLE II. COMPARISON OF TYPICAL FILE SYSTEMS

Property	NTFS	FAT	EXT2
Cluster size	0.5KB-4KB	4KB-16KB	4KB
Max block	2TB	4GB	2GB
Max partition	2TB	128GB	4TB
Index structure	B+tree	Chained	Chained
complexity	O(logN)	O(N)	O(N)
OS	Windows	DOS, Win 95	Linux

NTFS adopts B+ tree as index structure, and all data structures including directory are considered as file and indexed in the form of database [5]. When the file number is too large, it is necessary to establish multi-level indexing mechanism in order to reduce operation of read blocks and accelerate the speed of queries. In NTFS partition, partition information is stored in the different property files, where the most important file is Master File Table (MFT). It is core file of NTFS partition, which stores and identifies the basic information of all files. It consists of metadata about files, including the creating time, location, length, file name with a fixed length of 1KB.

For the small files and directories of less than 1KB, their content will be stored in MFT. Otherwise, only the location information will be saved in MFT [6]. In the process of reading big files, the file system will frequently visit MFT, so MFT has crucial impact on the access performance of the operating system.

The relationship between files and folders in NTFS is established through the index. If files belong to a same folder, their indexes are record by the property table of the folder's father in MFT. When the number of files within a folder is larger than a certain limit, the file index of the folder forms a B+ tree [7], and the root of the B+ tree is still stored in MFT record of the father directory. The depth of the tree has a crucial impact on the search efficiency, the greater the number of tree's level, the longer the traversal time. The file system has different directory tree structure following the number of files, and then affects the efficiency of retrieval.

For instance, if a big directory tree with three-level B+ tree is completely full according to limit of order, every B+ tree in first level can store 30 entries of index; the one in second level can store 900 entries; the one in third level can store 27930 entries [8]. When the index entries are more than 27,930, it will generate the fourth level's B+ tree. Flat tree structure can achieve the best efficiency of retrieval, so a reasonable directory structure will enable index tree to be balanced and enhance the reading efficiency.

Having retrieval experiments with 10,000 files and 100,000 files respectively, on behalf of the three-level and four-level B+ tree structure, it covers the general application of large directory. By tracking the experimental data and contrasting the results, the best number region of directory and file was estimated. The sample files (about 800KB in size) were put in the N directories averagely, recording its search time. Where, the value of N is set as follows: 5, 10, 20, 50, 100, 200, 500, 1000.The trends of retrieval efficiency are shown in Figure 4.

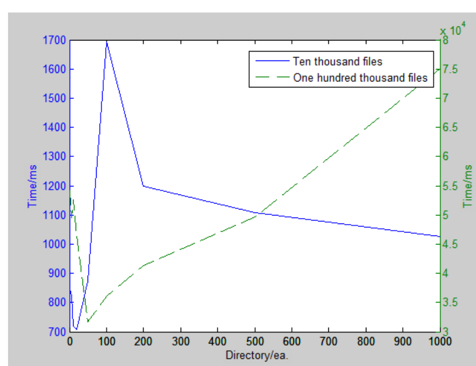

Figure 4. Comparison of file retrieval efficiency

Experiment results say that there is minimum retrieval time for 10,000 files with 20 directories, and for 100,000 files with 50 directories. At this point the B+ tree is balance, and the searching time is least for retrieval the leaf nodes.

C. Adaptive File Access Control Algorithm

Considering the reliable transmission of big files upon VPDN and retrieve optimization in single-server, the algorithm for AFAC is designed, which consists of four parts such as application configuration, network recognition, quality control and data transmission, and its components are shown in Figure.5.

1) User applications: provide system parameter setting interface, making it can be dynamically extended;

2) Network recognition: identify network status, or allow user settings, choose modes from CDMA, EVDO, etc.*;*

3) Quality control: complete data encapsulation and decapsulation, integrity verification and so on;

4) Data transmission: send and receive the data block using TCP mode, adjust file directory and file block size.

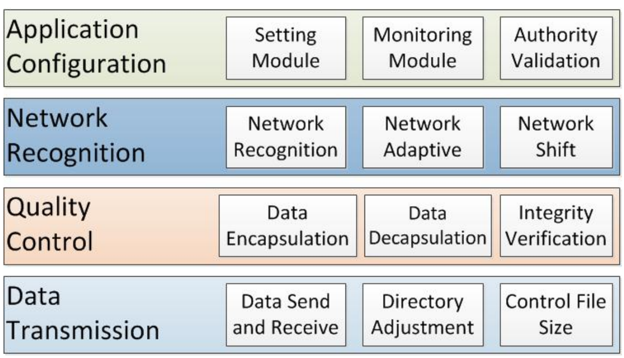

Figure 5. AFAC components

According to the experimental results in Section 3.2, set the initial number of directory is 20, when number of files is more than 10,000, each appending 1000 files with an additional directory. Based on wireless network technique, set the maximum size of the data threshold (the default value is 10240KB), sending data block by using dividing strategy when the threshold of file size is exceeded and merging them at receiving end. For smaller data files, after jointing them to a bigger file within the threshold, send it and decapsulate respectively at the sending and receiving end. Using AFAC for the transport of radar files, make sure the integrity of the layer-scan documents in order to effectively utilize the successful transmission of data blocks.

For example, the volume-scan file of X-band Doppler weather radar is around 9MB including 6 layer-scanned documents, and the volume-scan file of C-band Doppler weather radar is around 110MB including 14 layer-scanned documents. Thus, for X-band radar data, when the network state is good enough, it is transmitted without dividing, but for the C-band radar data, it is almost impossible to successfully complete the data transport without dividing.

Normally, a group of radar data is generated every six minutes, which consists of base data, images and products, occupying total hard disk space approximately 18.11MB (X-band radar). If the radar is on all day, it performs about 260 scans and generates approximately 4.65GB data. In order to optimize storage, a layer-scanned data can be stored as a file in favor of adjusting and optimizing the transport data block.

Definition 1: The data size of guaranteed delivery success rate p is denoted M_p, and the data block size of guaranteed delivery success rate 100% is denoted M_c. In the same network conditions, there is common sense: $M_p > M_c$.

In order to describe the part pseudo code of AFAC accurately, several functions and processes are given as follows:

divide_it(F, N):divide the file F into the size of N data block.
send_it(F):send the file F, and return “0” means fault.
traversal_it(D, N): look for file that size smaller than N in the directory D, and return “0” means no such file otherwise the file names.
merge_it(L,F): merge multiple files L into F.

Assuming the size of file F is x, the part pseudo code of AFAC for data sending module are listed as follows.

```
Data sending (F, x, Mp, Mc)
While (x!=0){
1  if  x< Mc  then {S=0;goto 5;} // merge the smaller files
2     else if  x< Mp  then {
           S= send_it(F);
           if S=0 goto 4; //fault, divide F into smaller block
              else x=0; }
   end}
3          else // divide F following  Mp
             { F1=divide_it(F, Mp);
                S= send_it(F1);
                if S=0 goto 4;
                else { F=F-F1;  x=x-Mp;}
   end}
4          F2=divide_it(F, Mc); // divide F following  Mc
            S= send_it(F2);
            if S=0 goto 1;// fault, resend
              else { F=F-F2; x=x- Mc; }
   end}
5   L=traversal_it(D, Mc);
    If L=0 then {
       while( S=0) {S= send_it(F);}// fault, resend
       x=0; end }
     else { merge_it(L,F)；goto 1}
   end}
```

Figure 6. The part pseudo code of AFAC

IV. PERFORMANCE EVALUATION BASED QOE-RT

Considering the application background of real-time access to weather radar file, the experience to network Quality of Service (QoS) was investigated, hoping to obtain a comprehensive evaluation for above research results.

A. QoE Model of File Transport upon VPDN

There are three main QoE evaluation methods such as statistics [9], artificial intelligence [10] and psychology [11]. The first two of them are suitable for multi-index evaluation system, and need certain expert knowledge. The psychology-based on evaluation method refers to the laws of psychology, and does not need complex training and computing. It can intuitively give the accurate QoE function model related with some QoS parameters. Although it cannot solve the problem affected by multiple factors, it has better effective in some applications.

The complete raw data of weather radar must be transported to information center, and the secondary products can be made. Therefore, the base data has to reach the data center with 6 minutes and no error for the real-time application, e.g., now-casting. For images split joint, the radar data detected by several radars must arrive in the data center within 15 minutes synchronously. In this application mode, the timeliness and correctness are two indicators that are cared by user, and can be used to establish the corresponding evaluation model.

So, the QoE model based on Reliability and Time (QoE-RT) was proposed referencing the psychological principle based on the relative waiting time and reliability factors. Classic Weber - Fechner law is applicative to large-scale range of irritation in middle intensity, which reflecting the relationship between QoE (i.e., the amount of feeling) and physical (e.g., waiting time) is not a simple linear growth: the longer the waiting time, the higher degree of tolerance. This relationship is shown in (2).

$$dp = k\frac{ds}{s} \tag{2}$$

Where dp presents the change of feeling, s indicates the amount of physical stimulus, $\frac{ds}{s}$ expresses the relative change of the physical stimulus, k is a proportion coefficient.

However, in above application scenario, having a smaller waiting time scale and greater irritation, the user experience of time is opposite to Weber - Fechner law: the longer the waiting period, the lower the degree of tolerance.

Definition 2: Let $l = \frac{d}{D}$ represents the relative waiting value, where d denotes the actual data transport time, and D is time-out limit of data transport. QoE-RT is defined in (3).

$$s = (1-\alpha)(1-k) + \alpha l, 0 \le \alpha \le 1, 0 \le k \le 1 \tag{3}$$

Where, s is evaluation scores of QoE-RT, the lower the score, the better the QoE; α is the adjustment coefficient for adjusting the weights that concerns with the contribution of transport success rate and transport time to evaluation score; k is the reliability factor that expresses transport success rate changing with the alteration of the file length.

Fitting the transport success rate curve shown in Figure 3, the function about reliability factor k was obtained in (4), $k = f(m)$, where, m is the file size (unit: MB).

$$f(m) = \begin{cases} 1, m < 4 \\ \beta m + \gamma, 4 \le m < 10 \\ \beta m + \gamma, 10 \le m < 14 \\ 0, m \ge 14 \end{cases} \tag{4}$$

$$\beta = \begin{cases} -0.01, 4 \le m < 10 \\ -0.125, 10 \le m < 14 \end{cases} \qquad \gamma = \begin{cases} 1.04, 4 \le m < 10 \\ 1.75, 10 \le m < 14 \end{cases}$$

According to the experiment data, the maximum waiting time for successfully transport file was the time to transmit 14MB file, which is computed by the expectations as $E(D) = 80$.

According to the result of Figure 3, the file size is related to the success rate of file transport, therefore it impacts on different QoE levels. Analyzing the experimental data by setting four intervals the influence factors α were concluded in (5).

$$\alpha = \begin{cases} 1, m < 4 \\ 0.8, 4 \le m < 10 \\ 0.2, 10 \le m < 14 \\ 0, m \ge 14 \end{cases} \tag{5}$$

In order to enable the user to intuitively get the results of the evaluation, S scores of QoE-RT were graded as five ranks and shown in Table III.

TABLE III. QOE SCORE MAPPING

S	QoE
(0,0.2]	excellent
(0.2,0.4]	good
(0.4,0.6]	medium
(0.6,0.8]	average
(0.8,1)	inferior

B. Performance Evaluation

In order to evaluate the performance of AFAC algorithm, two methods were compared for single file transport and multiple files transport with QoE-RT score as follows:

1) Using conventional FTP, files are stored in a directory without any control;

2) Using AFAC algorithm, set the initial directory number as 20.

We chose the 10,000 files with different size: 2MB, 6MB, 12MB, 16MB respectively and took 10 tests, and calculated the average data obtained. The result is shown in Figure 7.

For multiple files transport, the experiment scenario is same as the single file transport, and the result is shown in Figure 8.

Contrasting the Table III, it is found that for the same size file, the score of QoE-RT is similar no matter transmitting single file or multiple files. Whereas compared with FTP, AFAC algorithm has a higher level of QoE, it is shown in Figure 9.

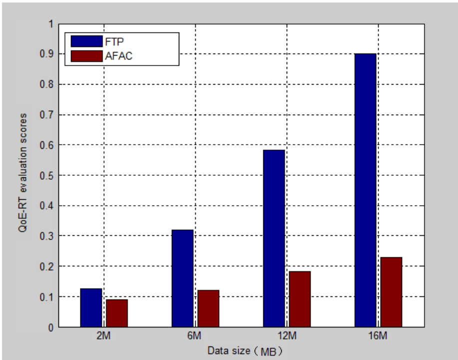

Figure 7. The QoE-RT score of single file access

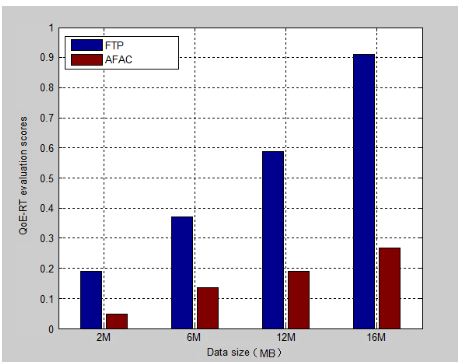

Figure 8. The QoE-RT score of multiple file access

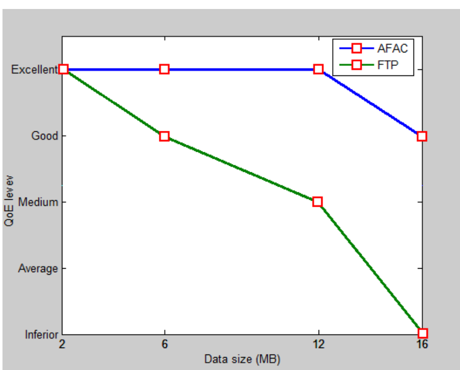

Figure 9. Comparison of both QoE scores FTP and AFAC

V. CONCLUSION AND FUTURE WORK

Facing the efficient real-time transportation and access of meteorological big file in VPDN and NTFS conditions, the optimal control strategy was researched. The AFAC algorithm was put forward based on the initiative dividing big file into several suitable data block and controlling the number of directories and files according to the threshold values shown by experiment results. Referencing psychology QoE evaluation methods, QoE-RT model was established, and made the comparison of performance between AFAC and conventional FTP. The experiment results indicated that AFAC has a better user experience for the condition of weak real-time in weather radar data transport.

In future work, it needs to consider the system overhead of file encapsulation and decapsulation for different network systems, and integrate the algorithm with autonomous network management.

ACKNOWLEDGMENT

The authors thank to the financial support by Nature Fund of Science of China with grant No. 61371119.

REFERENCES

[1] L. Wang, “Mass information resource storage and sharing technology research,” Information System Engineering, 2011, pp. 129-131, doi:10.3969/j.issn.1001-2362.2011.11.068 (in Chinese).

[2] Z. Luo, F. Liu, and Y. Xie, “User-Perceived FTP Service QoS Parameters and Measurement” IEEE International Conference on Network Infrastructure and Digital Content, 2009, pp. 69-73, doi: 10.1109/ICNIDC.2009.5360792.

[3] Y. Xu, “Design and Implementation of 3G-based Wireless VPDN Business Network,” Environmental Monitoring and Forewarning, 2010, pp. 27-30, doi:10.3969/j.issn.1674-6732.2010.05.009 (in Chinese).

[4] V. Rawat, R.Tio, S. Nanji, and R. Verma, Layer Two Tunneling Protocol. RFC 3070, February 2001. Standards Trackpp.3931L2TPv3,3070,doi:http://dx.doi.org/10.17487/RFC3070.

[5] J. Liang and Y. Zhang, “The main data structure of NTFS file System,” Computer Engineering and Design, 2003,pp.116-118,doi:10.3321/j.issn:1002-8331.2003.08.039 (in Chinese).

[6] L. Wang and J. Ju, “Analysis of NTFS file system structure,”Computer Engineering and Design,2006,pp.418-419,doi:10.3969/j.issn.1000-7024.2006.03.019(in Chinese).

[7] Z. Lu and X. Chen, “The optimization of B+ tree index file structure,” Computer Engineering and Design, 2000,pp.40-44,doi:10.3969/j.issn.1000-7024.2000.03.008(in Chinese).

[8] W. Wu, K. Liu, D. Jiang, Q. Su, and Z. Chen, “Dynamic analysis of NTFS B+tree big directory structure,” Computer Engineering and Design, 2013, pp. 1376-1382, doi:10.3969/j.issn.1000-7024.2013.04.046(in Chinese).

[9] R. Huang and Y. Guan, “Data Statistics Analysis,” Beijing:Higher Education Press, 2010(in Chinese).

[10] A. Altuzarra, JM. Moreno-Jiménez, and M. Salvador, “A Bayesian priorization procedure for AHP-group decision making,” European Journal of Operational Research, 2007, pp. 367–382,doi:10.1016/j.ejor.2006.07.025.

[11] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo, “The Logarithmic Nature of QoE and the Role of the Weber-Fechner Law in QoE Assessment,” 2010 IEEE International Conference on Communications, ISSN: 1550-3607, ISBN: 978-1-4244-6402-9 doi:10.1109/ICC.2010.5501894.