

## Research of File Backup Method Based on Double Cache and Minifilter Driver

Qiu Shaoming<sup>1, a</sup>, Tang Guobin<sup>1, b</sup>, Wang Yunming<sup>2, c</sup>

<sup>1</sup> College of Information Engineering, Dalian University, LiaoNing, China.

<sup>2</sup> School of Automatic, Nanjing University of Science and Technology, Jiangsu, China.

<sup>a</sup>semonqsm@126.com, <sup>b</sup>guobin1310@163.com, <sup>c</sup>wang19871128@126.com

**Keywords:** file backup, double cache, mini-filter driver, data storage.

**Abstract.** Aiming at the shortage of the low efficiency of file backup, large amount of data, highly occupancy rate network resource, Put forward a file backup method based on double cache and mini-filter driver. Method uses Cache Manager creates two caches in kernel model, at the same time combined with real-time monitoring of mini-filter driver to the file of operation information, turning it into a log classification records that procedures easily identify, and processing the file backup in cache, so as to achieve fast and efficient file backup. Experimental results show that this method can effectively reduce the disk reading and writing frequently, reducing the network resource occupancy rate when carry on the file backup, substantially increasing the speed of file backup.

### Introduction

At present, the data has become the core of the information system, once the key data of the information systems lost will bring inestimable loss to the user, such as the United States “9.11” incident, the world trade center and more than 40% of the company's business almost paralyzed or face closure, the fundamental reason is mainly due to the loss of a large number of business data, therefore, data backup technology has been referred to an unprecedented height<sup>[1]</sup>. The traditional file backup method need to transfer data between client and server, and must also check the file data, at the same time a large number of users request to the server backup data, it will occupy a lot of processor resources, also will cause the congestion of network<sup>[2]</sup>, which greatly reduces the efficiency of backup. Literature [3] propose a backup method based on event, real-time monitoring client file operations and generating operation log, then according to the generated operation log to achieve the file backup and recovery. Literature[4] propose a backup scheme based on index, this scheme uses two-dimensional linked list for the backup files indexed, and the same text only store a copy, so as to reach to the purpose of saving storage space; However, the cached data in memory easily disappear because of the computer shut down unexpectedly, and reading the log caused frequently operate disk and make the backup efficiency is reduced, at the same time also could be solve the time difference between the monitor and the background backup problems, index increased the burden of the system, reduces the backup efficiency, these literature designs are not radically shorten the backup and recovery time, and solve the problem of high occupancy network.

Based on this, this paper propose a scheme of file backup based on double buffering mini-filter driver, this scheme of minifilter driver exists in the operation system kernel space, and create two buffers of backup file, combined with operation log generated by mini-filter driver and correlated to the backup file in the buffer zone, greatly improves the backup efficiency, reduces the occupancy network, and saves the storage space.

### The principle of mini-filter driver

Mini-filter driver is a kind of layered filtering model for new filter driver, it is located between the I/O manager and file system, file system filter driver ignored the implementation of internal file system driver, it focus on the files operation<sup>[5-6]</sup>, the I/O manager according to the customer's end of file operation request to constructs IRP (I/O request packet) and sends IRP to the file system driver, file system filter driver can capture and intercept the IRP before the IRP was sent to target device, when

the IRP had been processed by upper drive, filter driver recorded the stack position of next layer filter in IRP, and returned it to the I/O manager, I/O manager sequentially submitted the IRP to the next driver program for processing, until the IRP will be processed by bottom drive, then IRP returned to the I/O manager, I/O manager submitted the result to requesting process by IRP processed, and released the IRP thread for resources applying, so as to realize the purpose to control file system operations.

## Research of double buffer and minifilter driver technology

**File monitor of minifilter model .** Traditional file monitoring is responsible to the client of the target machine's file operations and carry on real-time monitoring through a filter driver, but it lacks of necessary filtering rules, mini-filter driver aimed at the specified location and specific file types carrying on real-time monitor, through adding a layer of filtering intermediate layer between the file system and the I/O manager, intercept the I/O request packet (IRP). When sends an I/O request, the I/O manager to create a IRP sequence of operations, to track the request. So when the target machine operate files and the relevant software, monitoring module according to the filtering rules real-time monitoring the files, and will save the files operation in specific format for the log file.

**Improvement of incremental extraction algorithm.** The calculation of incremental data for existing file backup method is an incremental data extraction for client sends instructions to server, which will greatly increase the interaction between the client and the server, when the network is better, the backup files will not be affected, but once the network condition worse, this will bring the huge influence to the file backup, greatly increasing the network traffic.

In view of the existing file backup problems, advanced an improved incremental data extraction algorithm, using incremental data indexing storage, first stored the incremental data and instructional, and when the network condition in good states, sent united them to the server and constructed, such as much as possible to reduce the communication between the client and the server.

The algorithm of calculate the incremental data of file as follows:

(1) Suppose that the backup file was  $f_{old}$  at  $t_1$ , divided  $f_{old}$  into the file block in  $m$  bytes, final block might be less than  $m$ .

(2) Calculate the weak rolling checksum and MD5 check of block of  $f_{old}$ , and storage in  $hash < key, value >$  to database, weak rolling checksum algorithm:

$$\begin{cases} a(k,l) = (\sum_{i=k}^l X_i) \bmod M \\ b(k,l) = (\sum_{i=k}^l (l-i+1)X_i) \bmod M \\ s(k,l) = a(k,l) + 2^{16}b(k,l) \end{cases} \quad (1)$$

$s(k,l)$  is weak rolling checksum of block  $X_k, X_l, M = 2^{16}$ .

Through (1) can computed  $j+1, j+k+1$  rolling checksum:

$$\begin{cases} a(j+1, j+k+1) = (a(j, j+k) - X_j + X_{j+k+1}) \bmod M \\ b(j+1, j+k+1) = (b(j, j+k) - (k+1)X_j + a(j+1, j+k+1)) \bmod M \\ s(j+1, j+k+1) = a(j+1, j+k+1) + 2^{16}b(j+1, j+k+1) \end{cases} \quad (2)$$

(3) Suppose that backup file is  $f_{next}$  at  $t_2$ , and repeat the step (1).

(4) Use each block of weak rolling checksums of  $f_{old}$  search blocks of  $f_{next}$  for any blocks at any offset that match the checksum of some block of  $f_{old}$ , compute the weak rolling checksum for a block of length starting at each byte of  $f_{next}$  in turn, first we match the of  $hash < key, value >$ , if match, compare MD5, if match, record the first bytes of block, updated to the next offset and search proceeds until the end.

(5) Record and store the block of incremental and checksums, so as to use the double cache to execute the file reconstruction.

**The file backup based on double buffering and minifilter driver.** When the client sends backup instruction, the backup program began to read the monitor log and request a backup file, the buffer manager will let files "intelligent prereading" into buffer.

In order to scheduling manage the creation buffer, in the backup module constructs a buffer control management module (BCM), it records corresponding information when create a buffer, so you can easily scheduling buffer. When receiving the backup file transfer instruction, we can schedule the two buffers through BCM. Call buffer 1 perform file reconstruction between the blocks of incremental and the original file blocks, call buffer 2 encrypted and compressed and rename the reconstruction file. The capacity of file was smaller and code more concise after deal with by double buffer, At the same time the database of file information storage module will be save the sensitive information (such as the file name, path, size, storage path, modify date) from source file and synchronous store them to Oracle database tables. This not only simplifies communication protocol, shorten the backup time, but also can be conveniently and effectively manage the backup files, also can find and locate the files fast and efficiently and accurately in a file recovery. Compared to traditional file recovery method which traversed the whole disk folder and search for the files though file name, it not only save location time but also reduce CPU usage, thus greatly saving the file to find the time and effectively improve the efficiency of file recovery. Figure 1 illustrates the realization process of file backup based on double buffer and mini-filter driver.

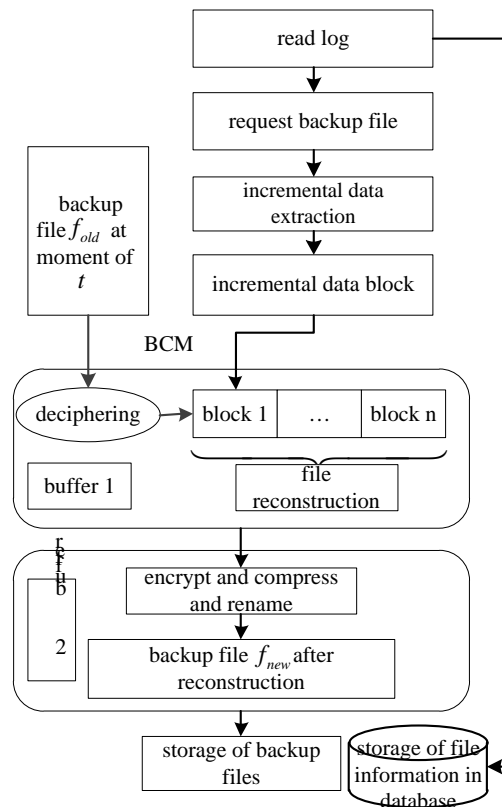


Fig 1. Reconstruction of backup file

## Experiment and analysis

The experimental environment: Server: service machine, Intel Xeon4 processor, frequency of 2.13GHz, memory: 2G, the operating system: Windows 2000 Server SP4 client: 2 client; Core2 Duo dual core processor, frequency of 1.86GHz, memory: 2G, operating system: Windows XP SP3; network environment: full duplex 1000M, network flow measurement tools: NetMeter, version 7.0.4.

**Experimental method.** Perform backup and recovery with two clients to the server, two processes have independently, with one using traditional backup method for backup of the file, another user the described in this paper; two backup are aim at the same files of different sizes. (1) Recorded the backup time respectively. (2) Recorded the variation of network stream when backup by NetMeter.

Experiment (1): the comparison with traditional file backup and double cache filter driver file backup of time is shown in Figure 2.

Experiment (2): The comparison with network flow is shown Figure 3.

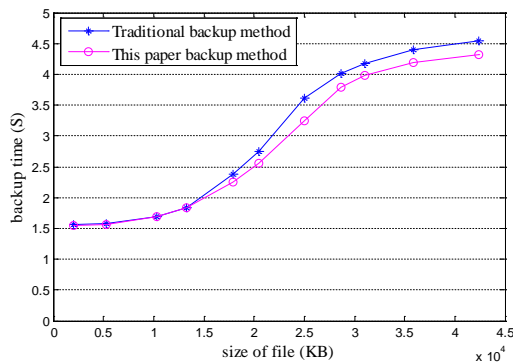


Fig2. Contrast to backup time

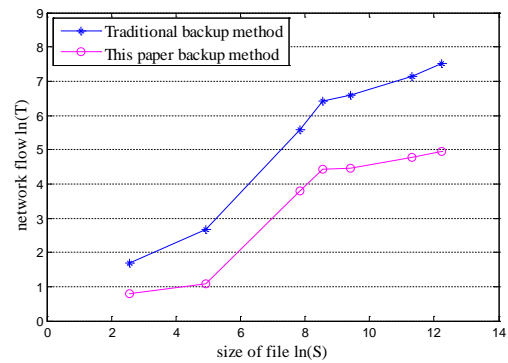


Fig 3. Contrast to network flow

**The result analysis.** (1) looked from figure 2, different size of the file, the file backup method used in this paper compared with the traditional file backup which effectively reduce the backup time required; with the increase in the size of the file, backup method is adopted in this paper in a certain range reduce the file backup time in a some degree.

(2) looked from figure 3, when the backup file size changes, the network flow generated by traditional backup method and backup of this paper, the relationship between the backup file size and network flow, this paper backup method compared with traditional backup method can significantly reduce the network flow.

## Conclusion

This paper uses double buffering mini-filter driver technology to realize file backup. Through the Windows buffer manager and file system create a two buffer for the backup file, so as to realize on compressed backup file and rename processing, effectively reduces the storage space of the backup file, save the maximum file backup time, and it is convenient to reasonable and effective manage the backup file; at the same time implement the file information database table storage to the backup file, not only can effectively manage the backup file, but also can find and locate the position of file fast and accurately for the recovery files, which improve file backup and recovery efficiency.

## References

- [1] Xiang Fei, Liu Chuanyi, Wang Chunlu, Fang Binxing. The new data disaster recovery strategy based on cloud computing environment. Journal of China Institute of communications, 34(6), pp. 92-93,2013.
- [2] Yang Xinhao. The network congestion control in the study of several algorithms based on control theory. Nanjing: School of automation, Nanjing University of Science and Technology, 2010.
- [3] Li Yiran, Li Tao, Hu Xiaoqin, Ma Xiaoxu. Research and realize of file backup base on event. Computer engineering and design. 31(18), pp. 3957-3959, 2010.
- [4] Lin Guoqing, Wang Jing, Chen Ruwei. The file backup scheme based on index. electronic design engineering. 19 (19), pp. 165-167, 2011.
- [5] Zhao Zhongmeng, Yao Hangtian, A Data Backup Method Based on File System Filter Driver. WuHan: proceedings of Software Engineering (WCSE), 2010 Second World Congress on, December pp. 19-20, 2010.
- [6] Wang Lina, Xiong Qi, Wang Dejun, Pan Tengfei. Research on the application of capturing continuous data technology and disaster recovery based on file filter driver [J], Journal of Chinese Computer System. 2011, 32 (3): 477-482.