



Automated Network Backup and Recovery Using a Safe-Set Baseline and Hash-Based Change Detection

Hamida Algoula*

*Department of Communications Engineering and Flight Information Technology
Technical college of Civil Aviation and Meteorology Technology (Asbiea)*

Tripoli, Libya

hamidamahamed@gmail.com

Received: 15/8/2026 Revised: 17/8/2026 Accepted: 05/9/2026 Publication: 13/10/2026

Abstract:

This paper is relating to PC data back up and the recovery, the techniques and systems for ordaining the data backup and recovery. Also, this paper is implementing the distribution.

With various apps in the PC system and constant developments of Information Technology, fast increase the quantity of the data, which became a critical factor in the working life. The research based on the understanding the importance of the network backup and restoring data to managers, clients, server level collected from the system constructions. The program is able to manage the gathering of network data and restore and backup it automatically.

To protect against data loss the user can be saved by using backup and data recovery program. System backup and recovery - is a software or hardware and software system to create copies of data at regular intervals for subsequent recovery. In addition to protection against loss of data backup systems also allow for the continuity of the work of employees to organize due to the rapid recovery of the operating system (if its image), or restore data on another computer.

Keywords: Data backup; Data recovery; Network backup; Delta extraction; Hash index; Safe-set; Cloud computing.

1.Introduction:

As time passes the high-tech technologies characterized by computers and networking technologies gaining a great importance by the users of PCs. Therefore, to keep in safe the documents become significant for the people. The data kept in the computers are very precious for many working people to lose the data means to lose the job, money, and most importantly time. The main importance of hardware and software of the computer is to accomplish different operations relating with data usage. The fast internet apps developments, cloud storage and cloud computing technologies becoming increasingly popular in present environment. To rely only on one application program, to produce applications with large data capacity systems, thousands of servers working collectively. Therefore, to be sure that applications with big capacities

of data is backed up, the occurred treat is been deleted immediately and the data is recovered automatically in all server computers is becoming a significant issue in the data security field [10][11].

The paper analyzed classifications of the file back-up software and the methodology of data recovery and back-up; studied the way of network data storing and data backup works, according to the way implemented by managers, customers, and the server layers. The model is able to understand the network file management, restore and back-up automatically [1][3].

2. Traditional Backup Approaches

To be protected from the loss of the important data, the data should be stored electronically and have to back up. Traditionally the process of data activation (or back-up) involved in writing a sample, also taking mirror image of the file to different data room devices on a particular time which was generally at night. When the data was backed up to storing devices, the storing devices were moved physically to secured place. If the initial data has been damaged or lost the sample of the file was accessible and could be copied from it. The disadvantages of traditional way of backing up the data are that devices to store the data took extra space and also it was time consuming to move storage devices from one place to another. And also because the need to back up data was increasing the number of devices for storing was also increasing, so the cost to keep data storing was also increasing [3][4][13].

3. Network-Based Backup

The other storage process that was popular in last years is network backup which transmits the data to a secured website. For example, the data management organizations providing services to back-up and achieve the data, and permits an organization to send backup file to a file manager on an agreed time duration. Usually the data in the storage is translated (coded) in order to protect from third party, and relief of information[2][15].

The disadvantage of sending the backed up data the manager to store through internet, internet is line of transmission of the data. Generally, transferring backed up data is hug and the capability of the communication length is exceeding between the data storage manager and the client. When big and several data is to be transmitted, the process of traditional backup data of sampling all the data necessitates to spend too much time and bandwidth which are very obtainable. Moreover, sometimes there is a need of changing only some percentage of the data to back-up for a time[5][6].

4. Delta Extraction Algorithm

One way of using a practical process is to use the delta extraction algorithm. It is controlling the changes that made the company to some backed up files, and modifying the changes immediately. Therefore, this way of modifying backed up data is much easier and faster, also it decreases the need to have a large unnecessary backup storage.

For instance, selecting files are done from the first order through alphabetic numeration in several delta extraction methods. As an example we can take “D” the drive under Windows NT can managed in the next command:

```
D:\autoexecs.bat
D:\boots.ini
D:\Dirs1\
D:\Dir1\Datas1.dat
D:\Dir1\Datas2.dat
D:\Dir1\Subs1\Image1.bmp
D:\Dir1\Subs1\Image2.bmp
D:\Dir1\Texts1.dat
D:\Dir1\Texts2.dat
D:\Dir2\Datas1.dat
D:\Dir2\Datas2.dat
D:\Dir2\Images1.bmp
D:\Dir2\Images2.bmp
D:\pagefiles.sys.
```

These files are for comparison in contrast to delta file mapping, by using alphabetic contrast to decide if the data part is new or old from the previous backed data. If the program realizes that the files had changed then the data is wholly is copied, the previous data which were deleted temporarily from the later back-up are not used for the forthcoming backups. When the data is corrected, it is transferred to the server application from the client.

In the server computer the data is processed in the same way as it was back up. The data is consequently compound with the old data, by using alphabetical comparison to recognize changes if there is any. If there are some new documents (files) they are been saved wholly[2][5].

5.Limitations of Delta Extraction

Apart of using the “delta extraction algorithm”, the back-up process will take longer time once hug amount of the files are being involved, and if complex and wide back-

up data is been involved in the process. Sometimes some data contains up to ten thousand separate folders, which takes long time to come over issues such as slow work, not enough time and crashes of the system which happens very frequently. The user's app in performing delta algorithm greatly relying in sorting processes. These sorting processes are long, because it is allowing to change, to match with the initial files that are backed up previously and stored. Generally, user's side of processing the data to change and transmit the records are consuming hug amount of time[13].

6.The comprehensive information about methods of data backing up in internet

Creating a copy of the data is a fairly simple process, but the real needs of users are often very varied and complex. For example, many users want to be able to make backups of any point or store very large amounts of data. For companies relevant is the problem of managing large amounts of data storage and rapid recovery. For each class of problems, solutions, and there are a variety of backup and data recovery[2][13].

The new method to back up the information by the network connection involves next stages as will be explained further. The method includes receiving the data information for the first backing up data and thereafter will receive periodically data to back up or update the information which has been changed. Throughout this back up procedure, the system analyses new information and decides whether to create new file, or to move it to update the old file, or avoid data which is not changed. operating system of the computer delivers the data information on a stable and not randomly commands. For instance, at a user's app, the selection of folders are processing in depth first order as they shown in the folder scheme, that storing won't be able to be stored in alphabetic way. Using the same folders that were given in the previous example with the "D" driver in Windows NT which can work as next:

```
D:\autoexecs.bat
D:\pagefiles.sys
D:\Dir1\
D:\Dir1\Datas1.dat
D:\Dir1\Datas2.dat
D:\Dir1\Texts2.dat
D:\Dir1\Sub1\Images1.bmp
D:\Dir1\Sub1\Images2.bmp
D:\Dir1\Texts1.dat
D:\Dir2\Datas2.dat
```

D:\Dir2\Images2.bmp

D:\Dir2\Images1.bmp

D:\Dir2\Datas1.dat

D:\boots.ini

The system tracks data folder locations using a pointer scheme numeric values (e.g. 001,002,003) that represent addresses rather than file contents and pointer values shift when folders are added or removed so newer folders receive higher values. A baseline copy called the safe-set, maintained by the storage manager, is updated over time; during backup the current folders are compared to the safe set by the common sort order and their pointer values, and any current folders with larger pointer values are identified as new and appended to the safe-set. This comparison works because both the live set and the safe-set use the same sorting scheme, so even small pointer-value changes reveal added or changed folders [2][7][8].

7. Hash Index and Change Detection

The established system is an empirical hashing index to analyze whether the current folder is an updated of the initial folder in the safe-set. The algorithm that recognizes the change in a data folder through creating “hash total” of the records is called hashing. The hash total is the collection of numbers from different points from the data. Therefore, the current total is being compared to the previous hash total of the same folder. If there is any changes in the data folder, the result will show dissimilar hash total. Because, added numbers are been changed or moved to other locations so the hash total is changed. Therefore, when the hash total is dissimilar, the data must be changed and folder has to be updated in the safe-set[7][8].

For the total of hash for every folder in the safe-set there is a special place called “hash index”. It been created with safe-set in order to provide a baseline of hash totals for every single folder. The model is analyzing whether the current folder is updated relating to the initial folder through comparing total of hash of the current folder with total hash of the initial folder in the safe-set. If both hashes are differing from each other the information in the data has been changed from the last backup. Then the current folder is being moved and saved with the new updates in safe-set like an updated folder. However, when the total of the hash is the same it means that the folder has not changed, then data is just been skipped.

8. Storage Costs and Access-Frequency Tiering

In order to achieve efficiency and economy in the safe-set the storage of the data is used. There are many different costs for the storage media. For example, to have an

access to high speed is used next devices: RAID disc, near line, optical, and tape. Usually, more the storage media is faster more it is expensive. Therefore, it is better to store data or update the data folders with lesser expenses and smaller storage media[2].

The module is monitoring the access for every single data frequently. The grouping of the data folders is done according to the frequency access. In the result the smallest frequent accessed set of the data folder will be relocated in the less expensive storage area, whereas the biggest frequency accessed set of the data folders will be relocated to the fastest storage area. In this way it will increase the economic performance by minimizing the storage cost whereas lowering recovery and back-up intervals utilizing by the diverse storage media in the best efficient way. The controlling of the access frequently is a continuous process and the model will merge/move folders in the storage areas according to their separate access of frequency changes[9][16][17].

9.The data recovery and backup requirement analyses

The backup and recovery software is very important for managers; the important files are always backed up in different location, if the original file is lost or damaged there will not be need in put new efforts to create new document.

The file back-up is the requirement of file recovery, and the file recovery is the crucial goal of file back-up, and both procedures work together to make sure that data is save in the storage[3][17].

In the present time, there are many issues and difficulties exist in the computer system. Firstly, most computer users don't have professional skills and knowledge in computer system, so they cannot understand the system and working conditions clearly. Also, are not able to backup and process data effectively. Secondly, even though some specific units can recover the system, due to privacy requests the software program is generally can't send to other users. Moreover, due to the variety of computer applications systems created in the present, the locations of them are dispersed delivery state, and when problems occur the administrators are not able to resolve occurred problems[1][15][17].

The data security is a significant for the users of computers, because users are facing many risks in the network area data info, and to guaranty the safeness of the files the recovery and backup is the best way. Traditionally the backup and recovery system was done locally, but this method lost the meaning because of various problems with hardware and software disasters (like illegal cracks, viruses and etc.). For example, when the user faces failure of the local system, if the user is using the traditional way

he will lose the data or won't be able to have access to the data. Therefore, to use distant network recovery and back up data technology will be the most suitable for the user. In addition a perfect way of backup system is to include possibilities of the failure hardware and software system.

10.Recovery and Backup and other related Technologies

Backup is by the object used for restoring the data in the case of damage or failure of the program, so that the user won't lose the information which he needs. The object of the backup system is the information in the PCs; it is preventing information from natural disaster, hardware failures, software mistakes, loss or damages from user made illegal or unwanted operations[3].

The back-up data is established on recovery, the reason backing up the data usually is to reduce the risk of data room system. There are many different recovery/backup purposes to develop different types of backup and recovery methods, the generally used technique for backup and recovery system explained below[4][14].

11.The distribution of backups and restore

When the server is distributed locations are relatively decentralized and not able to or to operate with unnecessary expensive SAN system. There are various data services that user can use in every location, increased stored devices straight forward link with the server. The server to arrange the increased storage corresponds with "storage management" system, to delete relating storage method. This method is very easy and simple, quick arrangements of implementations, but the negative part is that the implementing the data in large scope is not very practical way to do. The distribution of back-up process arrangement is provided below in Fig. 1 [3] [15].

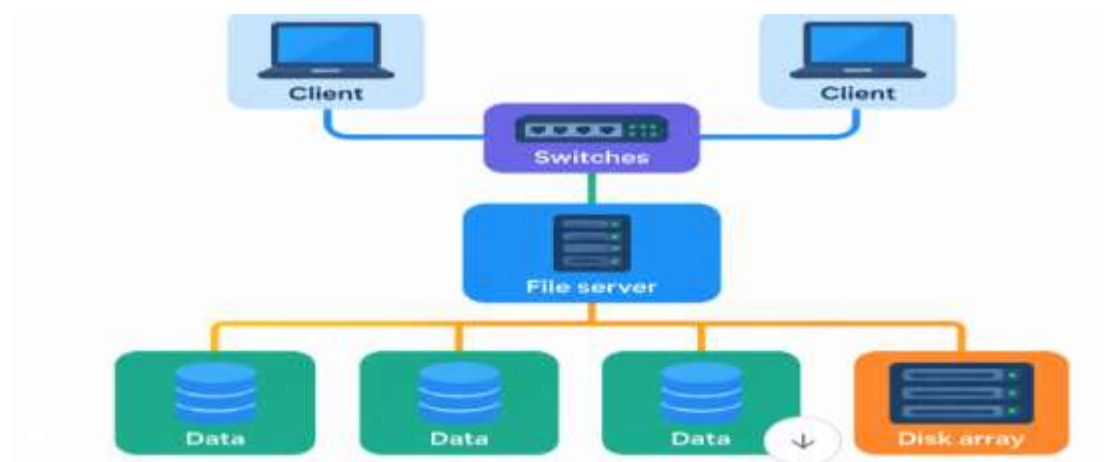


Figure 1. Distributed backup system structure.

12. Built in the storage area network recovery and back-up

The solution for the network back-up is built in the storage area, two different independent app modules “LAN” and “server”. In Fig.2 is shown the independent LAN back-up and recovery data application. LAN is the backup application server that is through visual fiber room network connects directly to room devices; room spaces by LUN are transparent allocated to the servers. Server data storage application does not take network bandwidth to up, transmission of the file stream doesn’t disturb app. SAN is the real backup data for app server, to protect the storage space the application is been used immediately with backup mechanism. Figure 2 provides the below the instruction of the construction of network back-up and recovery system built on the room area[3] [9].

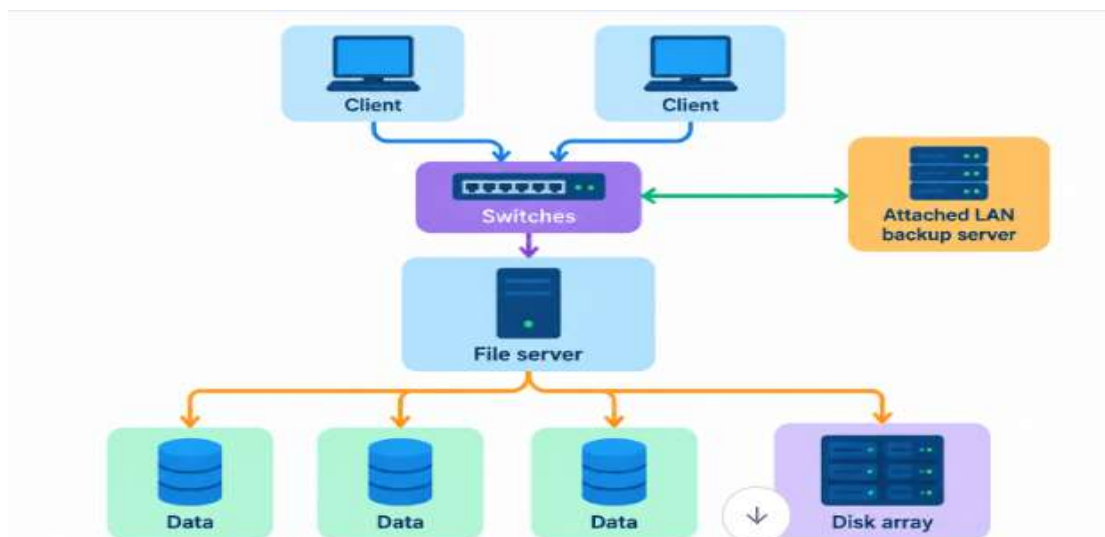


Figure 2. Structure of network backup and recovery based on the storage area

13. Network Backup System Architecture

Contemporary networks typically connect many clients to multiple servers; clients generally hold only user data while servers provide shared storage, backup storage and applications[3][9]. Common backup arrangements include: each user backing up to a personal tape drive (a low-cost, high-capacity, portable medium), storing important files on a remote server whose folders are systematically backed up, and Windows NT4/NT Domain clients having their data backed up to a tape drive attached to the NT server[1]. These approaches combine user responsibility and centralized protection depending on deployment.

Restores are often manual: a user requests recovery and an administrator locates the tape containing the latest backup. This method is reliable but places a burden on both

users and administrators[4]. To address this, many designs adopt a three-layer architecture client, management and service layers implemented in software and hardware [18], [19], [20]. The client layer controls PC data; the management layer provides backup services, adapter support and a certification center; and the service layer supplies storage server resources. This structure enables centralized, automated backup, archiving, storage and recovery, and employs PKI-based digital certificates and trust chains to authenticate devices and secure operations [1][12].

Conclusion

Under the networking environment the development of the app of cloud computing technologies for the file distribution room became a significant development way for back-up software. This research constructed the cloud software which backup and restore the data, is designed by insufficient hardware to support the centralized back-up data, and also delivers an easy recovery mechanisms. The system fully considers the entire backup and recovery procedure, the user is safe from unwanted issues by backing up the data.

The model and technique to back up the data through network is been developed over time. The technique contains making a baseline copy of the data folders which are to be achieved (used). If the data is constantly backed up, then the model controls the presence of new added folders through comparing the order sort of the current data folders with the order sort of the baseline copy. If there are any new folders then they are being saved and kept in the baseline copy until the next back up (or changes). The model is controlling for any changes in the current folder through comparing the hash total number of the current data folder with the hash total number of the data folders in the baseline copy. If there are changed folders then they are being merged to their matching data folders in the baseline copy.

References.

- [1] Xu, W., & Luo, J. (2015, April). The research on electronic data backup and recovery system based on network. In *2015 International Conference on Intelligent Systems Research and Mechatronics Engineering* (pp. 1892-1895). Atlantis Press.1892–1895, doi: 10.2991/isrme-15.2015.387.
- [2] Murray, B., Lewitzky, A., & Boldt, T. (2006). *U.S. Patent No. 7,134,041*. Washington, DC: U.S. Patent and Trademark Office.
- [3] Preston, W. C. (2007). *Backup and recovery* (Vol. 40). O'Reilly..
- [4] Chervenak, A., Vellanki, V., & Kurmas, Z. (1998, March). Protecting file systems: A survey of backup techniques. In *Joint NASA and IEEE Mass Storage Conference* (Vol. 99, pp. 17-32). Citeseer.
- [5] Mackerras, P., & Tridgell, A. (1996). The rsync algorithm. *Depart. CS, ANU, Canberra, Tech. Rep. TR-CS-96-05*.

- [6] Muthitacharoen, A., Chen, B., & Mazieres, D. (2001, October). A low-bandwidth network file system. In *Proceedings of the eighteenth ACM symposium on Operating systems principles* (pp. 174-187).
- [7] Quinlan, S., & Dorward, S. (2002). Venti: A new approach to archival data storage. In *Conference on file and storage technologies (FAST 02)*.
- [8] Zhu, B., Li, K., & Patterson, R. H. (2008, February). Avoiding the disk bottleneck in the data domain deduplication file system. In *Fast* (Vol. 8, pp. 1-14).
- [9] Patterson, D. A., Gibson, G., & Katz, R. H. (1988, June). A case for redundant arrays of inexpensive disks (RAID). In *Proceedings of the 1988 ACM SIGMOD international conference on Management of data* (pp. 109-116).
- [10] Mell, P., & Grance, T. (2011). The NIST definition of cloud computing..
- [11] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50-58.
- [12] Cooper, D., Santesson, S., Farrell, S., Boeyen, S., Housley, R., & Polk, W. (2008). RFC 5280: Internet X. 509 public key infrastructure certificate and certificate revocation list (CRL) profile.
- [13] Hutchinson, N. C., Manley, S., Federwisch, M., Harris, G., Hitz, D., Kleiman, S., & O'Malley, S. (1999, February). Logical vs. physical file system backup. In *OSDI* (Vol. 99, pp. 239-249).
- [14] Santry, D. S., Feeley, M. J., Hutchinson, N. C., Veitch, A. C., Carton, R. W., & Ofir, J. (1999, December). Deciding when to forget in the Elephant file system. In *Proceedings of the seventeenth ACM symposium on Operating systems principles* (pp. 110-123).
- [15] Cox, L. P., Murray, C. D., & Noble, B. D. (2002). Pastiche: Making backup cheap and easy. *ACM SIGOPS Operating Systems Review*, 36(SI), 285-298.
- [16] Wilkes, J., Golding, R., Staelin, C., & Sullivan, T. (1996). The HP AutoRAID hierarchical storage system. *ACM Transactions on Computer Systems (TOCS)*, 14(1), 108-136.
- [17] Kaczmarek, M., Jiang, T., & Pease, D. A. (2003). Beyond backup toward storage management. *IBM Systems Journal*, 42(2), 322-337.
- [18] Ben Dalla, L., Medeni, T. M., Agila, A. A., & Medeni, İ. M. (2024). Architectural Synergy: Investigating the Role of Artificial Neural Networks in Enabling Deep Learning. *The International Journal of Engineering & Information Technology (IJEIT)*, 12(1), 96-103.
- [19] Dalla, L., Medeni, T., Medeni, I., & Ulubay, M. (2025). Enhancing healthcare efficiency at Almasara Hospital: distributed data analysis and patient risk management. *Economy: strategy and practice*, 19(4), 54-72.
- [20] Dalla, L. O. F. B., & Ahmad, T. M. A. (2023). Heart Disease Prediction Via Using Machine Learning Techniques with Distributed System and Weka Visualization. *Journal of Southwest Jiaotong University*, 58(4), 322-333.