



SECURSTORE
...because recovery matters

DS-Delta Processing

Efficient storage usage and optimizing backup speed

How do we control and monitor SecurStore?

When evaluating the efficiency of a backup solution, very often the critical issues are data transmission/storage efficiency and the speed of backup. Especially for off-site backup (but not only), the data transmission efficiency is the most important, since the less data the backup solution needs to transmit off-site, the faster the backups.

- All good backup solutions implement at least incremental backups, so that files that do not change are not re-transmitted and as such, the storage usage is better and backup speed is increased. Of course, it is desirable that such incremental backups are done forever, without the need of performing full backups from time to time (as in the case of tape backups), since the full backups will always be storage inefficient and will take a longer amount of time than the incremental backups.
- For static data that is duplicated in many places across the LAN (like Operating System files), data can be transmitted even faster by sending a copy of a specific file only once and the other ones sent as simply links to the initial one. This process is called "Common File Elimination" and is used to save both storage space (only on copy of duplicate files stored on-line) as well as to increase backup speed (files do not need to be sent off-site since a copy already resides there)
- For dynamic data (databases, file being worked on), files will change frequently and after they are changed, the backup solution needs to protect the files by sending them again to the off-site location. Of course, it is desirable not to send complete copies of these files since they may be large compared to the amount of data that was changed in them. The process of sending only the changes in a file once an initial backup was completed is called "Delta Block Processing".

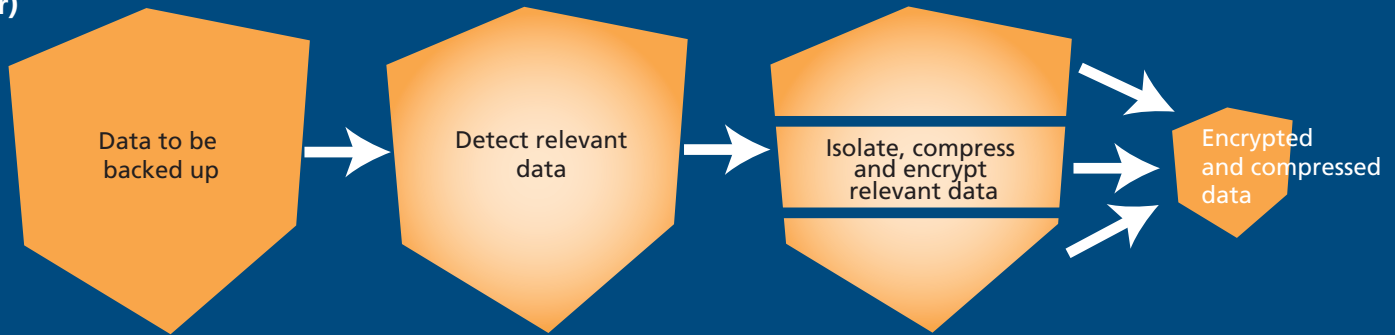
DS-Delta, reducing storage usage and increasing backup speed

- When backing up dynamic data (like databases, documents in process), it is almost always the case that these files are being gradually created. This means that no database will grow to 100GB in one day and then stop changing. It is more likely that a file will start small and then gradually change/grow every day. For a backup solution this means that the size of actual changes in a file are relatively small compared to the total file size.
- For a backup application it is very important to realize that once the initial backup is complete for a large file, for subsequent backups it is much more efficient to backup only changes in a file, since most of the data is already stored on-line. The ability to detect such changes and send then off-site without the need of sending a complete copy of the file is called "Delta Block Processing".

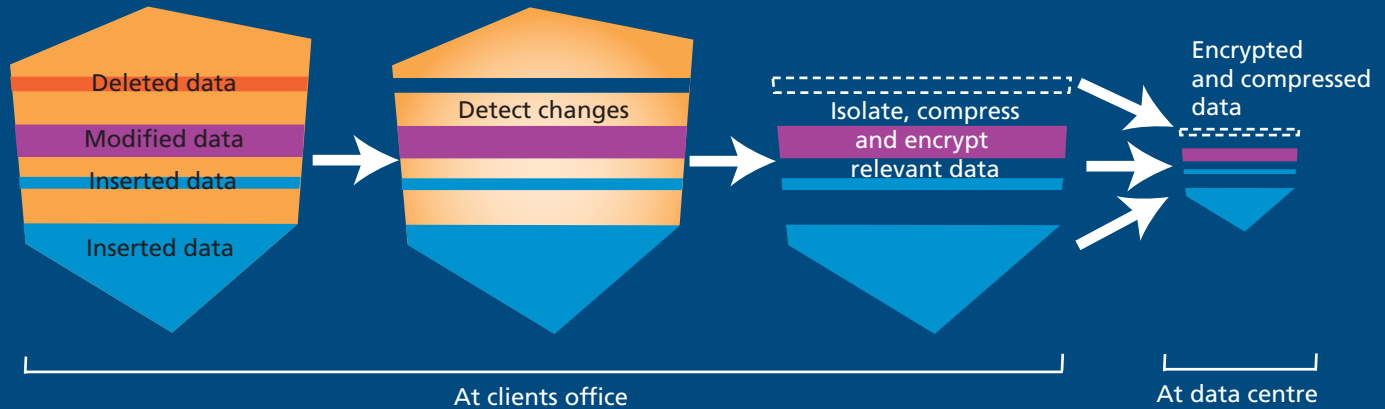
Who does this?

- In order for the Delta Block Processing to make sense, the change in a file should be small compared with the total size of the file. This actually translates into the requirement that files should be medium to large, although the definition of "large" is not very clear and depends on file type. The SecurStore DS-Client will perform delta processing for files larger than 32KB (a registry setting can be used to change this default, however it is not recommended)
- The DS-Delta algorithm is generic: it will work on any binary file, regardless of file type and content. This enables the SecurStore DS-Client to backup any type of database and file and perform delta processing on it.
- The DS-Delta algorithm is effective even when backing up the first generation of a file by removing redundant data blocks.

Generation 1 (Master)



Generation 2



DS-Delta functionality

- The main functionality behind the DS-Delta algorithm is the ability of detecting changes inside a file from one generation to another. Even more, the SecurStore DS-Delta can perform "Delta inside Delta", which means it can detect if a file has relevant changes inside its data, without the need to reuse data from a previous generation: it will re-use data from the existing generation.
- The reuse of data from a previous generation (or from the same generation in case of "delta inside delta") is done by dividing the file into multiple basic building blocks (hence the name "Delta Block Processing") and by comparing these blocks with the ones in the previously backed up data. The DS-Client will transmit only new/changed blocks, thus saving space and increasing backup speed.
- The DS-Delta algorithm is able to detect not only data addition to a file, but also data insertion inside the file, data deletion from the file and data modified inside the file.

Conclusions

- The DS-Delta algorithm greatly reduces the amount of data that needs to be re-sent to the off-site location. By reducing the amount of data that needs to be sent, the DS-Client is able to perform faster backups with reduced network traffic and reduced storage requirements
- The SecurStore DS-Delta is able to remove redundant data from a file even without a previous generation (Delta inside a Delta).
- The DS-Delta Algorithm is one of several data reduction functionalities inside the DS-Client:
- Incremental forever technology allows only changed files to be backed up (incremental backup at file level)
- Common file elimination is performed for backed up files, ensuring that duplicated files are not re-transmitted to the DS-System
- DS-Delta algorithm backs up only changes occurred from one file generation to the next (incremental backup at block level)
- Compression (LZOP or ZLIB) will be applied to data before encryption and transmission, maximizing storage/network efficiency

Below is a real example of 250GB SQL database and how delta processing and compression work together in minimizing the data to be transmitted thus increasing the speed of backups as well as keeping cost down without compromising clients data in anyway. The client can always fully restore any generation at a maximum speed and reliability.

Date (1)	Generation (2)	Original size GB (Protected) (3)	Comp. & delta blocked Gb (4)	Compression ratio (5)
23-Jan-09	Master (Gen 1)	250	69,0	72%
24-Jan-09	Generation 2	250	3,8	
25-Jan-09	Generation 3	250	4,6	
26-Jan-09	Generation 4	250	4,9	
27-Jan-09	Generation 5	250	4,6	
		1.250 GB (6)	86,9 GB (7)	93% (8)

1. Date of backup 2. Master generation is only performed in the beginning when initial backup is performed. After that only Delta generations are performed. Delta generation extracts changes from within the particular file. Delta generations happens every time a file changes between backup sets 3. Native size of the SQL database at the client side. During restores client can pick any generation to restore 4. Amount of data stored at Data Centre after compression and Delta-processing 5. Compression ratio between protected and stored data 6. Amount of data protected 7. Amount of data stored and what is invoiced for 8. Total compression ratio