Cloud Sync White Paper

Based on DSM 6.0

Table of Contents

Introduction

Nowadays we're given a wide variety of storage options, ranging from direct attached storage, networked storage and Internet-based public storage services. Each serves a different set of purposes, so we often work with a number of them in conjunction, drawing on the benefits and strengths of all where needed. Cloud Sync is designed to facilitate effortless, realtime data exchange. By offering a single framework for data synchronization between Synology NAS servers and heterogeneous storage interface, Cloud Sync bridges private network storage servers with that of oftentimes proprietary public storage services. With support for protocols such as WebDAV and storage like OpenStack Swift, Cloud Sync also deals with potentially private and on-premise storage types gracefully.

Featuring advanced functionalities like path mapping and encryption, along with fine-grained control over sync directions and traffic controls, Cloud Sync is highly flexible and versatile in its task type offerings. This paper outlines the technical designs of Cloud Sync and offers details of its performance for evaluation.

Product Features

Real-time synchronization. Synology Cloud Sync automizes the synchronization of data between the Synology NAS and cloud storage on a real-time basis, sending local updates instantly up to the remote storage, while pulling down remote changes as frequently as every ten seconds.

Monodirectional or bidirectional settings. Sync direction can be customized by session (a subtask created within a cloud storage connection) to meet different usage scenarios.

Customizable cloud storage polling period. The polling interval can be configured from ten seconds to one day, allowing the user to decide his or her own balance between data refresh rate and system resource consumption.

Multiple subtasks. With each cloud connection, multiple pairs of folders are allowed to be set up in synchroinization. This means local and remote data do not have to share identical directory structure.

One-to-one and one-to-many topology. Cloud Sync allows one local folder to be synchronized to more than one cloud destination, making multiple offsite backups possible.

Option to replicate deletions. With monodirectional syncs, an option is available to prevent file deletions on the destination. This facilitates what is broadly defined as incremental backup, as only additions and modifications are updated to the cloud server, and no deletion is initiated by Cloud Sync automatically.

Data encryption and compression. Synchronized data can be encrypted on the client side with Cloud Sync prior to uploading, to protect them from unauthorized access on the remote server. Compression helps reduce outbound traffic and storage consumption. The next chapter talks more about Cloud Sync's encryption design.

Synchronization

In synchronization, Cloud Sync acts as a client devices to cloud storage servers. Though offering a shared framework and feature set, certain synchronization functions are limited on the server side. Cloud Sync strives to deliver a unified experience as we interact with storage spaces with different designs and capabilities. This chapter talks about some of the things we do to make this possible.

Architecture

Cloud Sync consists of five major components:

- **Unified Sync Framework**: a carefully-designed framework adaptable to various storage interfaces and file systems.
- **Event/List Monitor**: monitors file changes on cloud storage.
- **File System Monitor**: monitors file changes on Synology NAS.
- **Cloud Sync Database**: retains local records of synchronized files and metadata.
- **Web-based User Interface**: provides graphic user interface for commands.

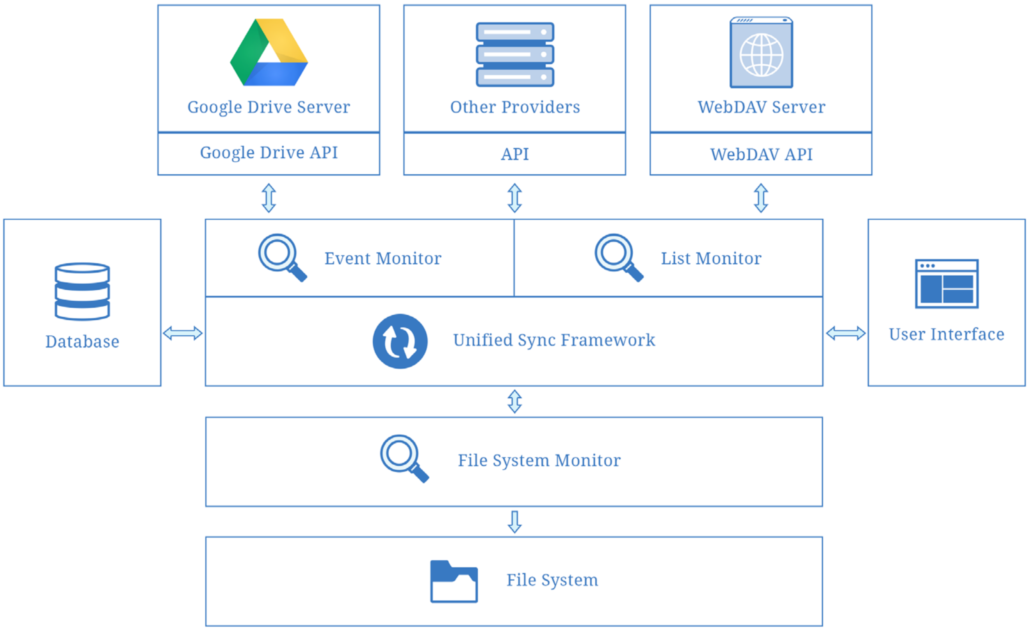


Figure 1: Cloud Sync architecture

Cloud Sync works in real-time, sending locally updated data to the remote and downloading remotely updated files to the local whenever a change takes place. To do so, Cloud Sync implements local file system and cloud change notifications.

File System Monitor (Local change notification)

Cloud Sync leverages DSM's advanced Inotify API to monitor file changes that take place locally on the NAS. This means the changes on the NAS are always instantly updated to the remote cloud storage.

Event/List Monitor (Remote change notification)

Cloud Sync keeps track of remote changes that take place on the cloud storage by polling. Polling mechanisms vary with different storage services and protocols. In general, we can categorize them into event-based providers and list-based providers.

○ Event-Based Providers

Change providers offer APIs that allow third-party vendors to fetch the delta in between each polling event. In such cases, Cloud Sync periodically, according to the configured polling period, demaiders and list-based providers.

Supported providers in this category include (as of April 2016):

- Amazon Cloud Drive
- Baidu Cloud
- Box
- Dropbox (including Dropbox for Business)
- Google Drive (including Dropbox for Business)[1]
- Microsoft OneDrive (including Office 365 and OneDrive for Business)[2]

○ List-Based Providers

For providers or protocols unable to offer delta information, Cloud Sync leverages LIST function to compare the directory structure in the local and remote sync folders. In this case, Cloud Sync generates the difference in between each polling interval on the client side, with computing powers on the NAS server. As the scanning of local directory trees call for system activity, connection with list providers can prevent the NAS from entering hibernation in cases where local file system monitored by Cloud Sync cannot be fully cached.

Supported providers in this category include (as of April 2016):

- Amazon S3
- S3-compatible storage
- hicloud S3
- hubiC
- IBM SoftLayer
- Google Cloud Storage
- Megafon MegaDisk
- OpenStack Swift-compatible storage
- Rackspace
- SFR NAS Backup
- WebDAV
- Yandex Disk

[1] Google Drive features an ID-based file system, which allows an account to own multiple files sharing the same file name.To resolve the resulting file name conflicts on the NAS, Cloud Sync records the IDs of every synced file into the database, and appends a serial number to the end of files that have the same name on the cloud.

[2] Tasks created over OneDrive's old API do not have event-based provider capabilities.

Consistency Check

Cloud Sync keeps a local database to record the synchronization status of each file. In scenarios where comparison of the local and remote is required for consistency check, the attributes recorded in the database come in very handy. The database is also very helpful in verifying change events and reducing API usage and unnecessary downloads/uploads.

Cloud Sync allows each user to configure whether to enable advanced consistency check in the sync process for each session, in which case an additional attribute - file hash - will be compared for consistency verification. The support for advanced consistency check requires the cloud storage to offer hash information. Hash availability is displayed in the charts below:

Action	**Advanced consistency check enabled**				**Advanced consistency check disabled**			
	type	**size**	**mtime**	**hash**	**type**	**size**	**mtime**	**hash**
Comapre cloud and local file in the event of a relink	v	v		v	v	v		
Compare the cloud and local file attribute before download	v	v	v	v	v	v	v	
Determining whether to rename a file due to conflict (compare downloaded file attribute with a local file of the same name)	v	v		v	v	v		

Platform	File hash
Dropbox	X
Google Drive	O (md5)
Baidu	O (md5)
Box	O (sha1)
hubiC	O (md5)
Amazon S3 (including hicloud S3 and SFR NAS Backup)	O (md5)[3]
WebDAV	X
HiDrive	X
Yandex Disk	X
Amazon Cloud Drive	O (md5)
MegaFon MegaDisk	X
OpenStack Swift (including IBM SoftLayer and Rackspace)	O (md5)[4]
Google Cloud Storage	O (md5)[5]
Microsoft OneDrive	O (sha1)
Microsoft OneDrive for Business	X

[3] Not available for Amazon S3 files uploaded via multi-part upload.

[4] Not available for OpenStack Swift files uploaded via DLO (Dynmamic Large Object).

[5] Cloud Sync uses Google Cloud Storage's base64 encoded md5 hash. The crc32 hash of composite object is not adopted.

Encryption Algorithm

Storing data at a remote and public location often results in heightened risk of hackers or unauthorized parties getting hold of the data. Cloud Sync offers the option to password-protect data backups to assuage such concerns.

In addition to HTTPS transmission encryption, the method adopted to prevent data from hacker interception during transfer, Cloud Sync also provides data encryption to ensure data on the cloud is safe from being accessed by all other entities, including the public cloud service itself.

Cloud Sync's data encryption is designed to allow decryption even when the source NAS is out of service. The following will detail the encryption flow and decryption options.

○ Encryption

With data encryption enabled, each file from the sync task will is encrypted by AES using a new randomly generated 256-bit key. The AES session key is then encrypted using 2048-bit RSA key and the user-defined primary key (password) to be kept together with the uploaded file.

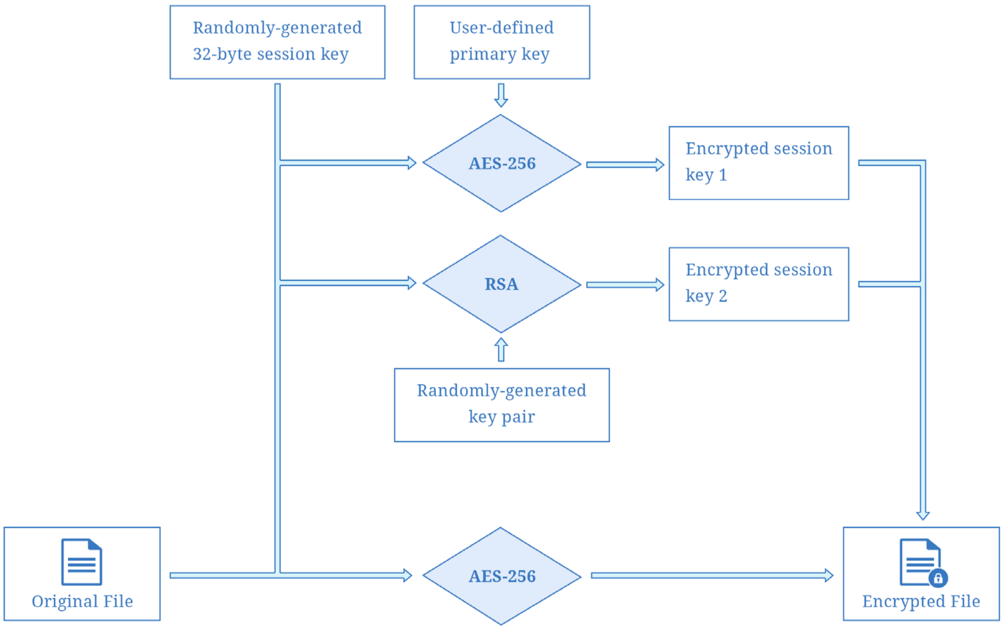


Figure 2: Encryption flow diagram

○ Decryption

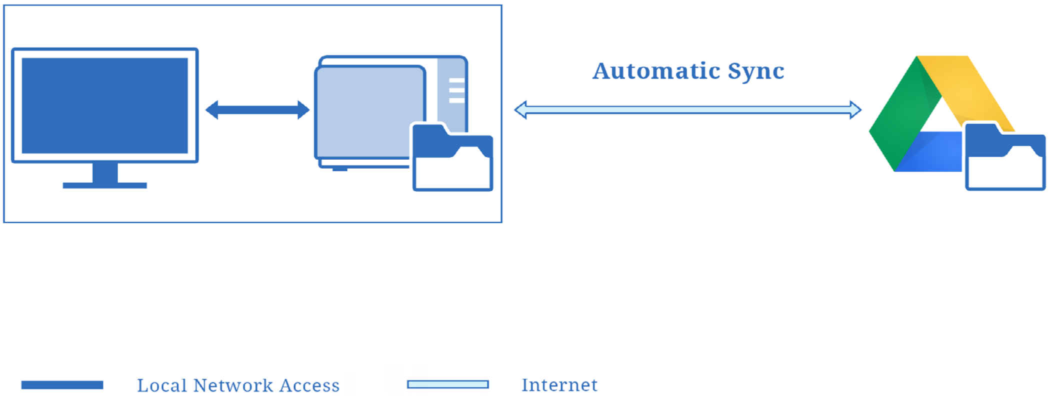
An encrypted file is automatically decrypted during Cloud Sync download. That is to say, when two NAS servers are linked to the same public cloud, files can be exchanged in between without sacrificing the confidentiality achieved by data encryption. The files encrypted and uploaded by either server will be readable by the other NAS that has the same password.

In cases where the NAS is stolen or unavailable, and the data on the cloud is needed, a decryption tool is provided to decrypt data manually on Windows or Ubuntu. With the decrption tool, users can decrypt a file or a folder using either the primary key or the RSA private key generated during task creation.

Usage Scenarios

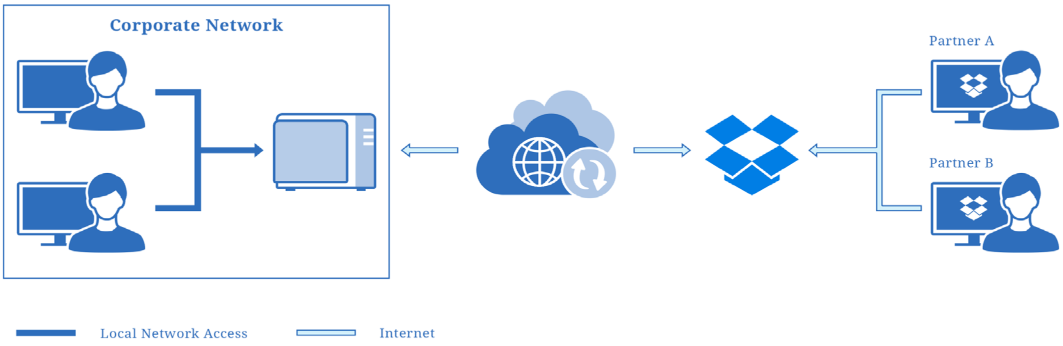
Public Cloud Gateway

Featuring real-time bidirectional synchronization, Cloud Sync makes Synology NAS an ideal public cloud gateway. Caching public cloud data on the NAS server, Cloud Sync allows fast local access to such data, reducing the latency often experienced when accessing data on the public cloud. This also offloads the transfer between the local and the cloud from the PC workstation to the NAS.



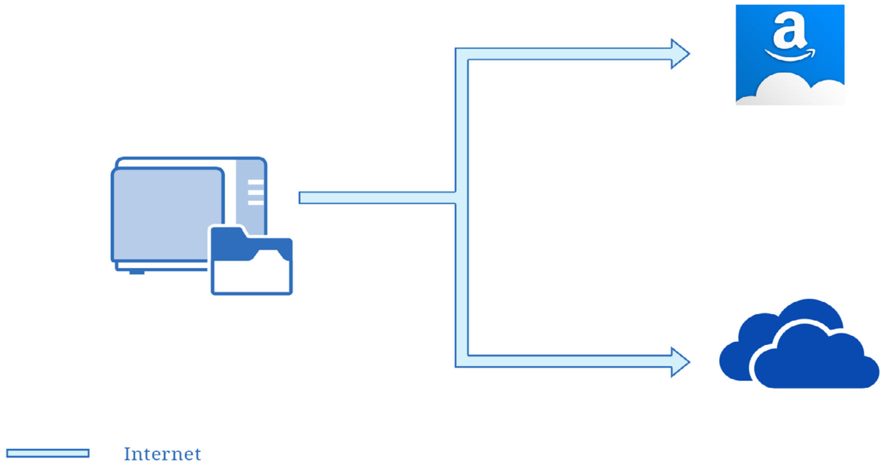
Collaboration across Multiple Storage Spaces

Cloud Sync facilitates collaboration with external parties over public cloud in scenarios where the partner is not given an account to the NAS storage server. In such cases both the local employee and the external correspondent can work in the comfort and speed of their workstations and let the public cloud and Cloud Sync take care of the exchange automatically.



Offsite Backup to more than one public cloud

Instead of setting up and maintaining a data center in a remote location, users can easily leverage public storage space as a cost-effective offsite backup destination by enabling upload-only monodirectional sync.



NAS as centralized data store of multiple cloud storage

While working on public storage seems inevitable these days, it is always reassuring to have Google Drive and Dropbox backed up in the security of your private NAS storage. This can be easily configured with Cloud Sync's download-only monodirectional sync. Cloud Sync also backs up your online Google Doc into Microsoft Office or JPEG formats.

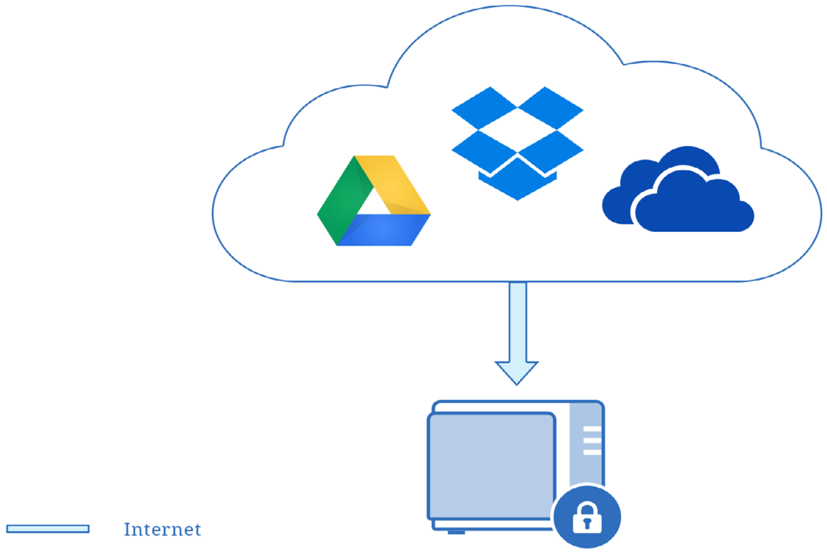


Figure 6: Version linked list

Performance Benchmark

Testing Bed

In the performance testing, we set up five different NAS to install Cloud Sync. The specifications of these five models are listed below.

- RS3614xs+: Ext4 on RAID 5 with twelve 1TB hard disks
- DS3615xs: Ext4 on RAID 5 with twelve 1TB hard disks
- DS716+: Ext4 on RAID 1 with two 1TB hard disks
- DS416: Ext4 on RAID 5 with four 1TB hard disks
- DS216j: Ext4 on RAID 1 with two 1TB hard disks

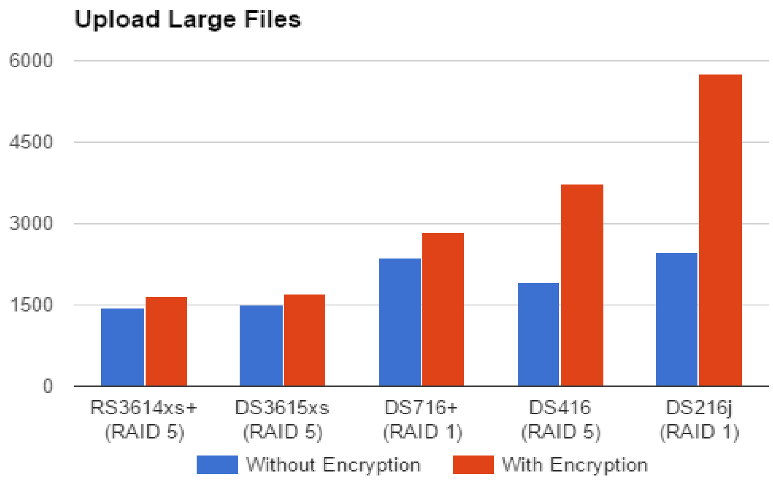
Each of the above NAS is installed with DSM v7321 and Cloud Sync v0716, and the Cloud Sync is connected with a WebDAV server with the following specifications:

- WebDAV Server: Windows IIS
- Hard disk: Intel 535 120G
- Memory: 16 GB

Large Files

The charts below shows the evaluation results for the large file case:

- Number of files: 20
- Size of each file: 5 GB



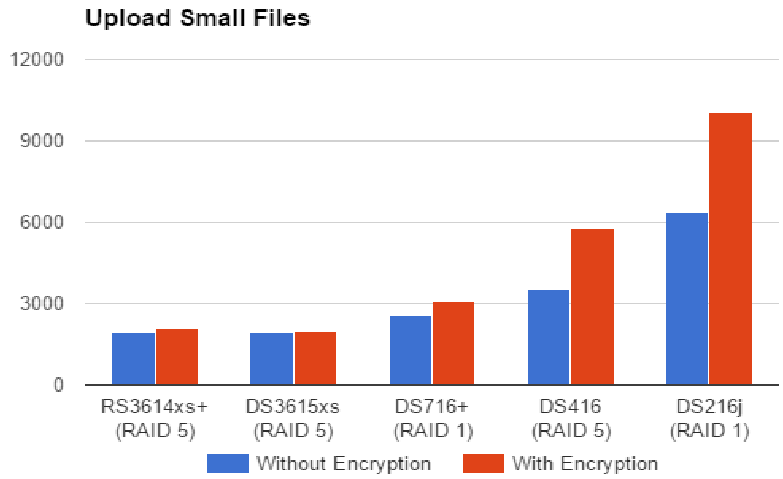
For the blue bars, it shows the results for the cases without encryption. From these results the Disk I/O is simply the bottleneck. Therefore, models with more powerful Disk I/O (that is, the ones to the left) yield better results. It is noteworthy to mention that, the DS416 model yields a better results than DS716+ due to the different RAID configurations. The DS716+ uses RAID 1 with 2 disks, and the DS416 uses RAID5 with 4 disks.

For red bars, it shows the results with encryption. Similarly, the Disk I/O is the bottleneck for models with strong computation power. However, for the models with weaker CPU (e.g., DS216j and possibly DS416), the CPU becomes the bottleneck.

◦ Small Files

The charts below shows the evaluation results for the small file case:

- Number of files: 100,000
- Size of each file: 1 MB



For models with powerful computation power (i.e., RS3614xs+ and DS3615xs), the Disk I/O is still the bottleneck for the testing tasks, producing similar results no matter the encryption is enabled or not.

Contrarily, for models with weaker computation power, it takes some time to accomplish the database operations and the encryption operation. As one can see, the rightmost low-end models are with longer operation time, and it takes even more time for the cases with encryption.

Conclusion

With Cloud Sync, one can seamlessly synchronize files between a Synology NAS and multiple cloud storage. Despite the traditional file synchronization, client-side encryption offers a great solution to overcome the safety concerns for public cloud storage, attending to both availbility and confidentiality simultaneously. Furthermore, with fine-grained configurable options, Cloud Sync meets the needs for most usage scenarios.