

## rsync

# Keeping in sync

The `rsync` [1] program allows you to synchronize local or remote files and directories using the “rsync remote-update protocol” that first checks for differences between the source and target before data transfer commences, thus minimizing the transfer volume. **BY HEIKE JURZIK**



The `rsync` program helps you to copy files from local or remote systems, ensuring that file properties (e.g. permissions or owners), and complete directory structures are kept. Launching the program is similar to launching `rcp` or `scp`: `rsync file(s) target`. To save the `dir1` directory recursively to the `backup` directory of your local system, type:

```
huhn@asteroid:~$ rsync -r dir1 backup
```

The `-r` (`--recursive`) option ensures that every scrap of information in any sub-directories will be transferred, although this does not include **symbolic links**.

The *skipping non-regular file dir2/link* message indicates that the `link` file in the `dir2` subdirectory is a symlink and will not be copied. To keep any symbolic links to other files on copying, additionally specify the `-l` (`--links`) parameter as in: `rsync -rl dir1 backup`.

To “resolve” symbolic links, instead use the `-L` (`--copy-links`) – the former links are then stored as normal files in the `backup` directory. You can set the `-p` (`--perms`) flag to preserve **file permissions**, and `-t` (`--times`) to keep the **timestamps**.

Only `root` can use an option to keep the file owner (`-o` or `--owner`), as normal

users are just not allowed to give away their files in this way. As the manpage indicates, you can combine all of these flags to create a single option `-rlptg -a` (`--archive`) copies recursively and retains file properties. Again, this requires `root` privileges.

## Patterns

The `--exclude = searchpattern` parameter excludes any files and directories that match the **searchpattern** from the copy operation. If you want to exclude files

```
asteroid:~$ ls -la /home/huhn/linux-magazine/28/
total 16
drwxr-xr-x  2 huhn  users    4096 Jan 14 17:46 ./
drwxr-xr-x  5 huhn  users    4096 Jan 14 17:17 ../
-rw-r--r--  1 huhn  users    5227 Jan 14 17:25 cmdline.txt
asteroid:~$ rsync -a -v /home/huhn/linux-magazine/ /home/huhn/backup/
building file list ... done
created directory /home/huhn/backup/
./
27/
28/
29/
27/edit.html
28/cmdline.txt
./
27/
28/
wrote 37180 bytes read 48 bytes 74456.00 bytes/sec
total size is 36937 speedup is 0.99
asteroid:~$
```

Figure 1: `rsync` makes the first copy of all the files...

```
asteroid:~$ mv /home/huhn/sysadmin.txt /home/huhn/linux-magazine/28/
asteroid:~$ rsync -a -v /home/huhn/linux-magazine/ /home/huhn/backup/
building file list ... done
28/sysadmin.txt
28/
wrote 5453 bytes read 32 bytes 10970.00 bytes/sec
total size is 42164 speedup is 7.69
asteroid:~$ ls -la /home/huhn/backup/28/
total 24
drwxr-xr-x  2 huhn  users    4096 Jan 14 17:50 ./
drwxr-xr-x  5 huhn  users    4096 Jan 14 17:17 ../
-rw-r--r--  1 huhn  users    5227 Jan 14 17:25 cmdline.txt
-rw-r--r--  1 huhn  users    5227 Jan 14 17:50 sysadmin.txt
asteroid:~$
```

Figure 2: ...and next time the same command only copies the changes.

## GLOSSARY

**Symbolic links:** A file reference that an application program will handle exactly if it were referencing the file. If the file the symlink points to is deleted, the link will point to empty space. Symlinks are created using the “ln -s sourcefile targetfile” syntax.

**File permissions:** Besides an owner and a group, files also have specific permissions: The “r” indicates “read” permission, with “w” meaning “write”, and “x” “executable”. The first triplet refers to the file’s owner, the second triplet to the group and the last triplet to any other users on the machine. The combination of letters that appears indicates the permissions, read, write, and execute for the file.

**Timestamps:** The Linux filesystem stores a variety of file information, for example the time elapsed since the last access (atime), status (ctime), or content change (mtime).

**Search pattern:** Normally comprising two elements: the characters to search for, and an count for the number of times these characters are allowed to occur. When used in combination with rsync, this provides you with a variety of complex character sequences, which are covered by several sections of the manpages. Typical wildcards can be used to perform most tasks.

**Wildcards:** Metacharacters that allow substrings to remain unspecified. A question

mark “?” in a string means that a single character at the position marked by the question mark is undefined (e.g. h?llo => hallo, hGlo, h7llo,...); in contrast to this, an asterisk “\*” can represent any number of different characters (e.g. m\*ss => mass, morass, or even mss).

**Environment variable:** The shell provides the user with memory space for storing specific information and to allowing access to application programs. These environment variables comprise a name and the assigned value. The standard Linux shell, bash, uses the export command to set variables, whereas csh uses the setenv command.

ending in .backup from the backup operation, the syntax is as follows:

```
huhn@asteroid:~$ rsync -r
--exclude=*.*backup source target
```

You can supply as many --exclude options as you need to exclude specific files. The --exclude-from = file parameter works in a similar way – but uses a file to supply the search patterns (one pattern per line) that describe the files to be excluded. Of course, this also works in reverse: the parameters --include = searchpattern and --include-from = file allow you to specify search patterns for files to be transferred.

## Encryption Please

For reasons of security, rsync is best combined with ssh. The “Secure Shell” (with the ssh login command and scp for secure file transfer) ensures that your data does not cross the wire in the clear, but is encrypted first. The scp tool soon shows its limitations when transferring files. If an upload or download crashes,

the program does not “remember” the files it has copied, and this can be annoying, especially if you are copying directories recursively (scp -r) or transferring high volumes of data. This scenario also fails to provide for version control, that is, new files simply overwrite existing files with the same name.

The parameter that causes rsync to use ssh is -e ssh or --rsh = ssh (defines a substitute command for rsh as indicated by the equals sign). If you want the ssh command to use a parameter of its own, such as -l (to select protocol version 1), you will need to place that section in quotes:

```
huhn@asteroid:~$ rsync -e "ssh -l" file target
```

otherwise rsync will attempt to parse the option as its own parameter. You can specify encrypted rsync as default, by setting an environment variable:

```
huhn@asteroid:~$ export RSYNC_RSH=ssh
```

After setting this variable, you can omit the -e ssh option. Incidentally, rsync can be combined with the --partial option to transfer larger files:

```
huhn@asteroid:~$ rsync -r
--partial file.avi pluto:films/
Password:
```

You can then type ls -l to list the partial file on the other machine:

```
huhn@pluto:~$ ls -l .file*
-rw----- 1 huhn users 3899392
Jan 6 20:31 .file.avi.yeahkG
```

If the copy operation crashes, the partial file that has already been transferred is kept.

When you re-launch the transfer operation, rsync will first check to see how much of the file has already made the jump and re-start at that point. ■

## INFO

[1] <http://rsync.samba.org/>



More information at:  
[www.linux-magazine.com/Backissues](http://www.linux-magazine.com/Backissues)

