

NAME

tar — manipulate tape archives

SYNOPSIS

```
tar [bundled-flags <args>] [<file>|<pattern> ...]
tar { -c } [options] [files | directories]
tar { -r | -u } -f archive-file [options] [files | directories]
tar { -t | -x } [options] [patterns]
```

DESCRIPTION

tar creates and manipulates streaming archive files. This implementation can extract from tar, pax, cpio, zip, jar, ar, and ISO 9660 cdrom images and can create tar, pax, cpio, ar, and shar archives.

The first synopsis form shows a “bundled” option word. This usage is provided for compatibility with historical implementations. See COMPATIBILITY below for details.

The other synopsis forms show the preferred usage. The first option to **tar** is a mode indicator from the following list:

- c** Create a new archive containing the specified items.
- r** Like **-c**, but new entries are appended to the archive. Note that this only works on uncompressed archives stored in regular files. The **-f** option is required.
- t** List archive contents to stdout.
- u** Like **-r**, but new entries are added only if they have a modification date newer than the corresponding entry in the archive. Note that this only works on uncompressed archives stored in regular files. The **-f** option is required.
- x** Extract to disk from the archive. If a file with the same name appears more than once in the archive, each copy will be extracted, with later copies overwriting (replacing) earlier copies.

In **-c**, **-r**, or **-u** mode, each specified file or directory is added to the archive in the order specified on the command line. By default, the contents of each directory are also archived.

In extract or list mode, the entire command line is read and parsed before the archive is opened. The pathnames or patterns on the command line indicate which items in the archive should be processed. Patterns are shell-style globbing patterns as documented in *tcsh*(1).

OPTIONS

Unless specifically stated otherwise, options are applicable in all operating modes.

@archive

(c and r mode only) The specified archive is opened and the entries in it will be appended to the current archive. As a simple example,

```
tar -c -f - newfile @original.tar
```

writes a new archive to standard output containing a file *newfile* and all of the entries from *original.tar*. In contrast,

```
tar -c -f - newfile original.tar
```

creates a new archive with only two entries. Similarly,

```
tar -czf - --format pax @-
```

reads an archive from standard input (whose format will be determined automatically) and converts it into a gzip-compressed pax-format archive on stdout. In this way, **tar** can be used to convert archives from one format to another.

-b *blocksize*

Specify the block size, in 512-byte records, for tape drive I/O. As a rule, this argument is only needed when reading from or writing to tape drives, and usually not even then as the default block size of 20 records (10240 bytes) is very common.

-C *directory*

In c and r mode, this changes the directory before adding the following files. In x mode, change directories after opening the archive but before extracting entries from the archive.

--check-links

(c and r modes only) Issue a warning message unless all links to each file are archived.

--chroot

(x mode only) **chroot()** to the current directory after processing any **-C** options and before extracting any files.

--exclude *pattern*

Do not process files or directories that match the specified pattern. Note that exclusions take precedence over patterns or filenames specified on the command line.

--format *format*

(c, r, u mode only) Use the specified format for the created archive. Supported formats include “cpio”, “pax”, “shar”, and “ustar”. Other formats may also be supported; see `libarchive-formats(5)` for more information about currently-supported formats. In r and u modes, when extending an existing archive, the format specified here must be compatible with the format of the existing archive on disk.

-f *file*

Read the archive from or write the archive to the specified file. The filename can be **-** for standard input or standard output. If not specified, the default tape device will be used. (On FreeBSD, the default tape device is `/dev/sa0`.)

-H (c and r mode only) Symbolic links named on the command line will be followed; the target of the link will be archived, not the link itself.

-h (c and r mode only) Synonym for **-L**.

-I Synonym for **-T**.

--include *pattern*

Process only files or directories that match the specified pattern. Note that exclusions specified with **--exclude** take precedence over inclusions. If no inclusions are explicitly specified, all entries are processed by default. The **--include** option is especially useful when filtering archives. For example, the command

```
tar -c -f new.tar --include='*foo*' @old.tgz
```

creates a new archive `new.tar` containing only the entries from `old.tgz` containing the string ‘foo’.

-j (c mode only) Compress the resulting archive with `bzip2(1)`. In extract or list modes, this option is ignored. Note that, unlike other **tar** implementations, this implementation recognizes `bzip2` compression automatically when reading archives.

-k (x mode only) Do not overwrite existing files. In particular, if a file appears more than once in an archive, later copies will not overwrite earlier copies.

--keep-newer-files

(x mode only) Do not overwrite existing files that are newer than the versions appearing in the archive being extracted.

-L (c and r mode only) All symbolic links will be followed. Normally, symbolic links are archived as such. With this option, the target of the link will be archived instead.

- l** This is a synonym for the **--check-links** option.
- m** (x mode only) Do not extract modification time. By default, the modification time is set to the time stored in the archive.
- n** (c, r, u modes only) Do not recursively archive the contents of directories.
- newer date**
(c, r, u modes only) Only include files and directories newer than the specified date. This compares ctime entries.
- newer-mtime date**
(c, r, u modes only) Like **--newer**, except it compares mtime entries instead of ctime entries.
- newer-than file**
(c, r, u modes only) Only include files and directories newer than the specified file. This compares ctime entries.
- newer-mtime-than file**
(c, r, u modes only) Like **--newer-than**, except it compares mtime entries instead of ctime entries.
- nodump**
(c and r modes only) Honor the nodump file flag by skipping this file.
- null**
(use with **-I**, **-T**, or **-X**) Filenames or patterns are separated by null characters, not by newlines. This is often used to read filenames output by the **-print0** option to **find(1)**.
- numeric-owner**
(x mode only) Ignore symbolic user and group names when restoring archives to disk, only numeric uid and gid values will be obeyed.
- O** (x, t modes only) In extract (-x) mode, files will be written to standard out rather than being extracted to disk. In list (-t) mode, the file listing will be written to stderr rather than the usual stdout.
- o** (x mode) Use the user and group of the user running the program rather than those specified in the archive. Note that this has no significance unless **-p** is specified, and the program is being run by the root user. In this case, the file modes and flags from the archive will be restored, but ACLs or owner information in the archive will be discarded.
- o** (c, r, u mode) A synonym for **--format ustar**
- one-file-system**
(c, r, and u modes) Do not cross mount points.
- options options**
Select optional behaviors for particular modules. The argument is a text string containing comma-separated keywords and values. These are passed to the modules that handle particular formats to control how those formats will behave. Each option has one of the following forms:
 - key=value* The key will be set to the specified value in every module that supports it. Modules that do not support this key will ignore it.
 - key* The key will be enabled in every module that supports it. This is equivalent to *key=1*.
 - !key* The key will be disabled in every module that supports it.

module:key=value, *module:key*, *module:!key*

As above, but the corresponding key and value will be provided only to modules whose name matches *module*.

The currently supported modules and keys are:

iso9660:joliet

Support Joliet extensions. This is enabled by default, use **!joliet** or **iso9660:!joliet** to disable.

iso9660:rockridge

Support Rock Ridge extensions. This is enabled by default, use **!rockridge** or **iso9660:!rockridge** to disable.

gzip:compression-level

A decimal integer from 0 to 9 specifying the gzip compression level.

xz:compression-level

A decimal integer from 0 to 9 specifying the xz compression level.

mtree:keyword

The mtree writer module allows you to specify which mtree keywords will be included in the output. Supported keywords include: **cksum**, **device**, **flags**, **gid**, **gname**, **indent**, **link**, **md5**, **mode**, **nlink**, **rmd160**, **sha1**, **sha256**, **sha384**, **sha512**, **size**, **time**, **uid**, **uname**. The default is equivalent to: “device, flags, gid, gname, link, mode, nlink, size, time, type, uid, uname”.

mtree:all

Enables all of the above keywords. You can also use **mtree:!all** to disable all keywords.

mtree:use-set

Enable generation of **/set** lines in the output.

mtree:indent

Produce human-readable output by indenting options and splitting lines to fit into 80 columns.

zip:compression=type

Use *type* as compression method. Supported values are store (uncompressed) and deflate (gzip algorithm).

If a provided option is not supported by any module, that is a fatal error.

- P** Preserve pathnames. By default, absolute pathnames (those that begin with a / character) have the leading slash removed both when creating archives and extracting from them. Also, **tar** will refuse to extract archive entries whose pathnames contain **..** or whose target directory would be altered by a symlink. This option suppresses these behaviors.
- p** (x mode only) Preserve file permissions. Attempt to restore the full permissions, including owner, file modes, file flags and ACLs, if available, for each item extracted from the archive. By default, newly-created files are owned by the user running **tar**, the file mode is restored for newly-created regular files, and all other types of entries receive default permissions. If **tar** is being run by root, the default is to restore the owner unless the **-o** option is also specified.
- q (--fast-read)**
(x and t mode only) Extract or list only the first archive entry that matches each pattern or filename operand. Exit as soon as each specified pattern or filename has been matched. By default, the archive is always read to the very end, since there can be multiple entries with the same name and, by convention, later entries overwrite earlier entries. This option is provided as a performance optimization.
- S** (x mode only) Extract files as sparse files. For every block on disk, check first if it contains only NULL bytes and seek over it otherwise. This works similar to the **conv=sparse** option of **dd**.

--strip-components *count*

(x mode only) Remove the specified number of leading path elements. Pathnames with fewer elements will be silently skipped. Note that the pathname is edited after checking inclusion/exclusion patterns but before security checks.

-s *pattern*

Modify file or archive member names according to *pattern*. The pattern has the format */old/new/[gps]* where *old* is a basic regular expression, *new* is the replacement string of the matched part, and the optional trailing letters modify how the replacement is handled. If *old* is not matched, the pattern is skipped. Within *new*, *~* is substituted with the match, 1 to 9 with the content of the corresponding captured group. The optional trailing *g* specifies that matching should continue after the matched part and stopped on the first unmatched pattern. The optional trailing *s* specifies that the pattern applies to the value of symbolic links. The optional trailing *p* specifies that after a successful substitution the original path name and the new path name should be printed to standard error.

-T *filename*

In x or t mode, **tar** will read the list of names to be extracted from *filename*. In c mode, **tar** will read names to be archived from *filename*. The special name “-C” on a line by itself will cause the current directory to be changed to the directory specified on the following line. Names are terminated by newlines unless **--null** is specified. Note that **--null** also disables the special handling of lines containing “-C”.

-U (x mode only) Unlink files before creating them. Without this option, **tar** overwrites existing files, which preserves existing hardlinks. With this option, existing hardlinks will be broken, as will any symlink that would affect the location of an extracted file.

--use-compress-program *program*

Pipe the input (in x or t mode) or the output (in c mode) through *program* instead of using the builtin compression support.

-v Produce verbose output. In create and extract modes, **tar** will list each file name as it is read from or written to the archive. In list mode, **tar** will produce output similar to that of **ls(1)**. Additional **-v** options will provide additional detail.

--version

Print version of **tar** and **libarchive**, and exit.

-w Ask for confirmation for every action.

-X *filename*

Read a list of exclusion patterns from the specified file. See **--exclude** for more information about the handling of exclusions.

-Y (c mode only) Compress the resulting archive with **bzip2(1)**. In extract or list modes, this option is ignored. Note that, unlike other **tar** implementations, this implementation recognizes **bzip2** compression automatically when reading archives.

-z (c mode only) Compress the resulting archive with **gzip(1)**. In extract or list modes, this option is ignored. Note that, unlike other **tar** implementations, this implementation recognizes **gzip** compression automatically when reading archives.

-Z (c mode only) Compress the resulting archive with **compress(1)**. In extract or list modes, this option is ignored. Note that, unlike other **tar** implementations, this implementation recognizes **compress** compression automatically when reading archives.

ENVIRONMENT

The following environment variables affect the execution of **tar**:

LANG	The locale to use. See <code>environ(7)</code> for more information.
TAPE	The default tape device. The -f option overrides this.
TZ	The timezone to use when displaying dates. See <code>environ(7)</code> for more information.

FILES

`/dev/sa0` The default tape device, if not overridden by the `TAPE` environment variable or the **-f** option.

EXIT STATUS

The **tar** utility exits 0 on success, and >0 if an error occurs.

EXAMPLES

The following creates a new archive called `file.tar.gz` that contains two files `source.c` and `source.h`:

```
tar -czf file.tar.gz source.c source.h
```

To view a detailed table of contents for this archive:

```
tar -tvf file.tar.gz
```

To extract all entries from the archive on the default tape drive:

```
tar -x
```

To examine the contents of an ISO 9660 cdrom image:

```
tar -tf image.iso
```

To move file hierarchies, invoke **tar** as

```
tar -cf - -C srcdir . | tar -xpf - -C destdir
```

or more traditionally

```
cd srcdir ; tar -cf - . | (cd destdir ; tar -xpf -)
```

In create mode, the list of files and directories to be archived can also include directory change instructions of the form **-Cfoo/baz** and archive inclusions of the form **@archive-file**. For example, the command line

```
tar -c -f new.tar foo1 @old.tgz -C/tmp foo2
```

will create a new archive `new.tar`. **tar** will read the file `foo1` from the current directory and add it to the output archive. It will then read each entry from `old.tgz` and add those entries to the output archive. Finally, it will switch to the `/tmp` directory and add `foo2` to the output archive.

An input file in `mtree(5)` format can be used to create an output archive with arbitrary ownership, permissions, or names that differ from existing data on disk:

```
$ cat input.mtree
#mtree
usr/bin uid=0 gid=0 mode=0755 type=dir
usr/bin/ls uid=0 gid=0 mode=0755 type=file content=myls
$ tar -cvf output.tar @input.mtree
```

The **--newer** and **--newer-mtime** switches accept a variety of common date and time specifications, including “12 Mar 2005 7:14:29pm”, “2005-03-12 19:14”, “5 minutes ago”, and “19:14 PST May 1”.

The **--options** argument can be used to control various details of archive generation or reading. For example, you can generate `mtree` output which only contains **type**, **time**, and **uid** keywords:

```
tar -cf file.tar --format=mtree --options='!all,type,time,uid' dir
```

or you can set the compression level used by `gzip` or `xz` compression:

```
tar -czf file.tar --options='compression-level=9'.
```

For more details, see the explanation of the `archive_read_set_options()` and `archive_write_set_options()` API calls that are described in `archive_read(3)` and `archive_write(3)`.

COMPATIBILITY

The bundled-arguments format is supported for compatibility with historic implementations. It consists of an initial word (with no leading - character) in which each character indicates an option. Arguments follow as separate words. The order of the arguments must match the order of the corresponding characters in the bundled command word. For example,

```
tar tbf 32 file.tar
```

specifies three flags **t**, **b**, and **f**. The **b** and **f** flags both require arguments, so there must be two additional items on the command line. The *32* is the argument to the **b** flag, and *file.tar* is the argument to the **f** flag.

The mode options **c**, **r**, **t**, **u**, and **x** and the options **b**, **f**, **l**, **m**, **o**, **v**, and **w** comply with SUSv2.

For maximum portability, scripts that invoke **tar** should use the bundled-argument format above, should limit themselves to the **c**, **t**, and **x** modes, and the **b**, **f**, **m**, **v**, and **w** options.

Additional long options are provided to improve compatibility with other tar implementations.

SECURITY

Certain security issues are common to many archiving programs, including **tar**. In particular, carefully-crafted archives can request that **tar** extract files to locations outside of the target directory. This can potentially be used to cause unwitting users to overwrite files they did not intend to overwrite. If the archive is being extracted by the superuser, any file on the system can potentially be overwritten. There are three ways this can happen. Although **tar** has mechanisms to protect against each one, savvy users should be aware of the implications:

- Archive entries can have absolute pathnames. By default, **tar** removes the leading / character from filenames before restoring them to guard against this problem.
- Archive entries can have pathnames that include . . components. By default, **tar** will not extract files containing . . components in their pathname.
- Archive entries can exploit symbolic links to restore files to other directories. An archive can restore a symbolic link to another directory, then use that link to restore a file into that directory. To guard against this, **tar** checks each extracted path for symlinks. If the final path element is a symlink, it will be removed and replaced with the archive entry. If **-U** is specified, any intermediate symlink will also be unconditionally removed. If neither **-U** nor **-P** is specified, **tar** will refuse to extract the entry.

To protect yourself, you should be wary of any archives that come from untrusted sources. You should examine the contents of an archive with

```
tar -tf filename
```

before extraction. You should use the **-k** option to ensure that **tar** will not overwrite any existing files or the **-U** option to remove any pre-existing files. You should generally not extract archives while running with super-user privileges. Note that the **-P** option to **tar** disables the security checks above and allows you to extract an archive while preserving any absolute pathnames, . . components, or symlinks to other directories.

SEE ALSO

`bzip2(1)`, `compress(1)`, `cpio(1)`, `gzip(1)`, `mt(1)`, `pax(1)`, `shar(1)`, `libarchive(3)`, `libarchive-formats(5)`, `tar(5)`

STANDARDS

There is no current POSIX standard for the tar command; it appeared in ISO/IEC 9945-1:1996 (“POSIX.1”) but was dropped from IEEE Std 1003.1-2001 (“POSIX.1”). The options used by this implementation were developed by surveying a number of existing tar implementations as well as the old POSIX specification for tar and the current POSIX specification for pax.

The ustar and pax interchange file formats are defined by IEEE Std 1003.1-2001 (“POSIX.1”) for the pax command.

HISTORY

A **tar** command appeared in Seventh Edition Unix, which was released in January, 1979. There have been numerous other implementations, many of which extended the file format. John Gilmore’s **pdtar** public-domain implementation (circa November, 1987) was quite influential, and formed the basis of GNU tar. GNU tar was included as the standard system tar in FreeBSD beginning with FreeBSD 1.0.

This is a complete re-implementation based on the `libarchive(3)` library.

BUGS

This program follows ISO/IEC 9945-1:1996 (“POSIX.1”) for the definition of the **-l** option. Note that GNU tar prior to version 1.15 treated **-l** as a synonym for the **--one-file-system** option.

The **-C** *dir* option may differ from historic implementations.

All archive output is written in correctly-sized blocks, even if the output is being compressed. Whether or not the last output block is padded to a full block size varies depending on the format and the output device. For tar and cpio formats, the last block of output is padded to a full block size if the output is being written to standard output or to a character or block device such as a tape drive. If the output is being written to a regular file, the last block will not be padded. Many compressors, including `gzip(1)` and `bzip2(1)`, complain about the null padding when decompressing an archive created by **tar**, although they still extract it correctly.

The compression and decompression is implemented internally, so there may be insignificant differences between the compressed output generated by

```
tar -czf - file
```

and that generated by

```
tar -cf - file | gzip
```

The default should be to read and write archives to the standard I/O paths, but tradition (and POSIX) dictates otherwise.

The **r** and **u** modes require that the archive be uncompressed and located in a regular file on disk. Other archives can be modified using **c** mode with the `@archive-file` extension.

To archive a file called `@foo` or `-foo` you must specify it as `./@foo` or `./-foo`, respectively.

In create mode, a leading `./` is always removed. A leading `/` is stripped unless the **-P** option is specified.

There needs to be better support for file selection on both create and extract.

There is not yet any support for multi-volume archives or for archiving sparse files.

Converting between dissimilar archive formats (such as tar and cpio) using the `@-` convention can cause hard link information to be lost. (This is a consequence of the incompatible ways that different archive formats store hardlink information.)

There are alternative long options for many of the short options that are deliberately not documented.