

TAR(5) BSD 文件格式手册 TAR(5)

名称 tar - 磁带归档文件的格式

描述 tar归档格式将任意数量的文件、目录和其他文件系统对象（符号链接、设备节点等）收集为单一的字节流。该格式最初设计用于固定块大小操作的磁带驱动器，但现已广泛用作通用打包机制。

通用格式 tar归档由一系列512字节的记录组成。每个文件系统对象需要一个头部记录（存储基本元数据：路径名、所有者、权限等）和零个或多个包含文件数据的记录。归档结束由两个全零字节的记录标识。

为兼容使用固定块大小的磁带驱动器，读写tar文件的程序每次I/O操作总是处理固定数量的记录。这些“块”始终是记录大小的整数倍。早期实现支持的最大块大小为10240字节（20条记录）。尽管现代高速磁带驱动器常用1MiB（2048条记录）或更大的块大小，这仍是大多数实现的默认值。（注：此处术语“块”和“记录”并非完全标准；本文档遵循John Gilmore在pdtar文档中建立的约定。）

旧式归档格式

原始tar归档格式经过多次扩展，以包含不同实现者认为必要的附加信息。本节描述AT&T; UNIX第7版中tar命令实现的变体，这似乎是早期广泛使用的tar程序版本。

旧式tar归档的头部记录结构如下：
`struct header_old_tar { char name[100]; char mode[8]; char uid[8]; char gid[8]; char size[12]; char mtime[12]; char checksum[8]; char linkflag[1]; char linkname[100]; char pad[255]; };`

头部记录中所有未使用的字节均用空值填充。

name 路径名，以空终止字符串存储。早期tar实现仅存储常规文件（包括这些文件的硬链接）。早期常见约定使用尾部“/”字符表示目录名，从而允许归档和恢复目录权限及所有者信息。

mode 文件模式，以ASCII八进制数存储。

uid,gid 所有者的用户ID和组ID，以ASCII八进制数存储。

size 文件大小，以ASCII八进制数存储。

对于常规文件（regular files）而言，该字段表示紧随文件头之后的数据量。值得注意的是，早期tar实现版本在提取硬链接（hardlinks）时会忽略此字段。现代写入程序应始终为硬链接条目存储零长度值。

BSD 2016年12月27日 1

TAR(5) BSD 文件格式手册 TAR(5)

`mtime` 文件的修改时间，以 ASCII 八进制数表示。该数值表示自纪元开始（1970年1月1日 00:00:00 UTC）以来的秒数。注意应避免使用负值，因其处理方式存在不一致性。

`checksum` 头部校验和，以 ASCII 八进制数存储。计算校验和时，先将校验和字段设为全空格，再对头部所有字节进行无符号算术求和。此字段应以6位八进制数字后接空字符和空格的形式存储。需注意，早期许多 `tar` 实现在校验和字段中使用有符号算术，可能导致系统间归档文件传输时的互操作性问题。现代健壮的读取器会采用两种方式计算校验和，只要任一结果匹配即接受该头部。

`linkflag`, `linkname` 为保留硬链接（hardlinks）并节省磁带空间，具有多重链接的文件仅在首次遇到时写入归档。后续遇到时，`linkflag` 会被设为 ASCII 字符 '1'，且 `linkname` 字段记录该文件首次出现的名称（注意常规文件的 `linkflag` 字段值为空）。

早期 `tar` 实现对这些字段的终止方式存在差异。AT&T; UNIX 第7版中的 `tar` 命令遵循以下约定（早期 BSD 手册页亦有记载）：路径名必须以空字符终止；`mode`、`uid` 和 `gid` 字段必须以空格加空字节结尾；`size` 和 `mtime` 字段必须以空格结尾；校验和以空字符加空格终止。早期实现会在数字字段前导填充空格，这种惯例一直持续到 IEEE Std 1003.1-1988 ("POSIX.1") 标准发布。为实现最佳可移植性，现代实现应在数字字段前导填充零。

前POSIX归档格式 IEEE Std 1003.1-1988 ("POSIX.1") 的早期草案是 John Gilmore 的 `pdtar` 程序及1980年代末至1990年代初众多系统实现的基础。这些归档通常遵循下文所述的 POSIX `ustar` 格式，但存在以下差异： - 魔数（magic value）由五个字符 "ustar" 后接空格组成，版本字段包含一个空格字符加空字符 - 数字字段通常采用前导空格填充（而非最终标准推荐的前导零填充）

- 前缀字段（prefix field）通常不被使用，这使得路径名被限制为旧式归档文件的100个字符长度。

POSIX `ustar`归档格式 IEEE标准1003.1-1988 ("POSIX.1") 定义了一种标准tar文件格式，要求兼容的tar(1)实现能够读写该格式。这种格式通常被称为"ustar"格式，得名于其文件头中使用的魔数（magic value）。（该名称是"Unix Standard TAR"的首字母缩写）。它在历史格式基础上通过新增字段进行了扩展：

```
``c struct header_posix_ustar { char name[100]; char mode[8]; char uid[8]; char  
gid[8]; char size[12]; char mtime[12]; char checksum[8]; char typeflag[1]; char  
linkname[100]; char magic[6]; char version[2]; char uname[32]; ... }; ``
```

BSD December 27, 2016 2