



北京邮电大学

Beijing University of Posts and Telecommunications



# 文件内容比对

蒋砚军 北京邮电大学计算机学院

### ■ 基本功能

- ◆ cmp和散列算法：判断两个文件内容是否相同
- ◆ 文件数据完整性验证
- ◆ diff：列出两个文本文件之间的区别
- ◆ 版本管理

## ▶ 两文件逐字节比较: **cmp**

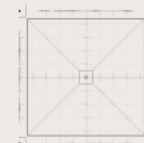


### ■ 用法

◆ `cmp file1 file2`

### ■ 功能

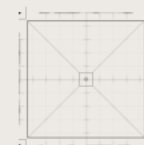
- ◆ 逐字节比较两个文件是否完全相同
- ◆ 两个文件完全相同时，不给出任何提示
- ◆ 两个文件不同时，打印出第一个不同之处
- ◆ 在Windows中有类似的命令**COMP**



## ► md5sum/sha1sum:文件内容比较



- ◆使用MD5算法（散列函数）根据文件内容生成16字节hash值，比较hash值是否相同，就可断定两文件内容是否完全相同
- ◆使用SHA-1算法的命令名为**sha1sum** (20字节hash值)
- ◆常用于数据完整性（Data Integrity）验证和判断位于网络不同机器上的两个文件内容是否相同
- ◆其他散列函数也可以用来完成这一任务：**sha512sum**



## ► md5sum/sha1sum:文件内容比较



```
$ md5sum src.tar proto.txt
```

```
4faffc1f3714a693d7844dcb949ce020  src.tar  
937d369a72c240a01a6a7a9efba919bd  proto.txt  
2b00042f7481c7b056c4b410d28f33cf  log.txt
```

```
$ md5sum src.tar proto.txt log.txt > myfiles.sum
```

仅将myfiles.sum传送到另台计算机，在另台计算机上运行程序，比较同名文件的内容是否一致

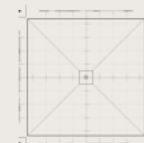
```
$ md5sum -c myfiles.sum
```

```
src.tar: OK  
proto.txt: OK  
log.txt: OK
```

**失误率**

**MD5:  $2^{-128} = 3.4 \times 10^{-38}$**

**SHA-1:  $2^{-160} = 4.7 \times 10^{-50}$**





北京邮电大学

Beijing University of Posts and Telecommunications



谢谢