哈希函数

Hash Functions

刘 铎

liuduo@bjtu.edu.cn

哈希函数

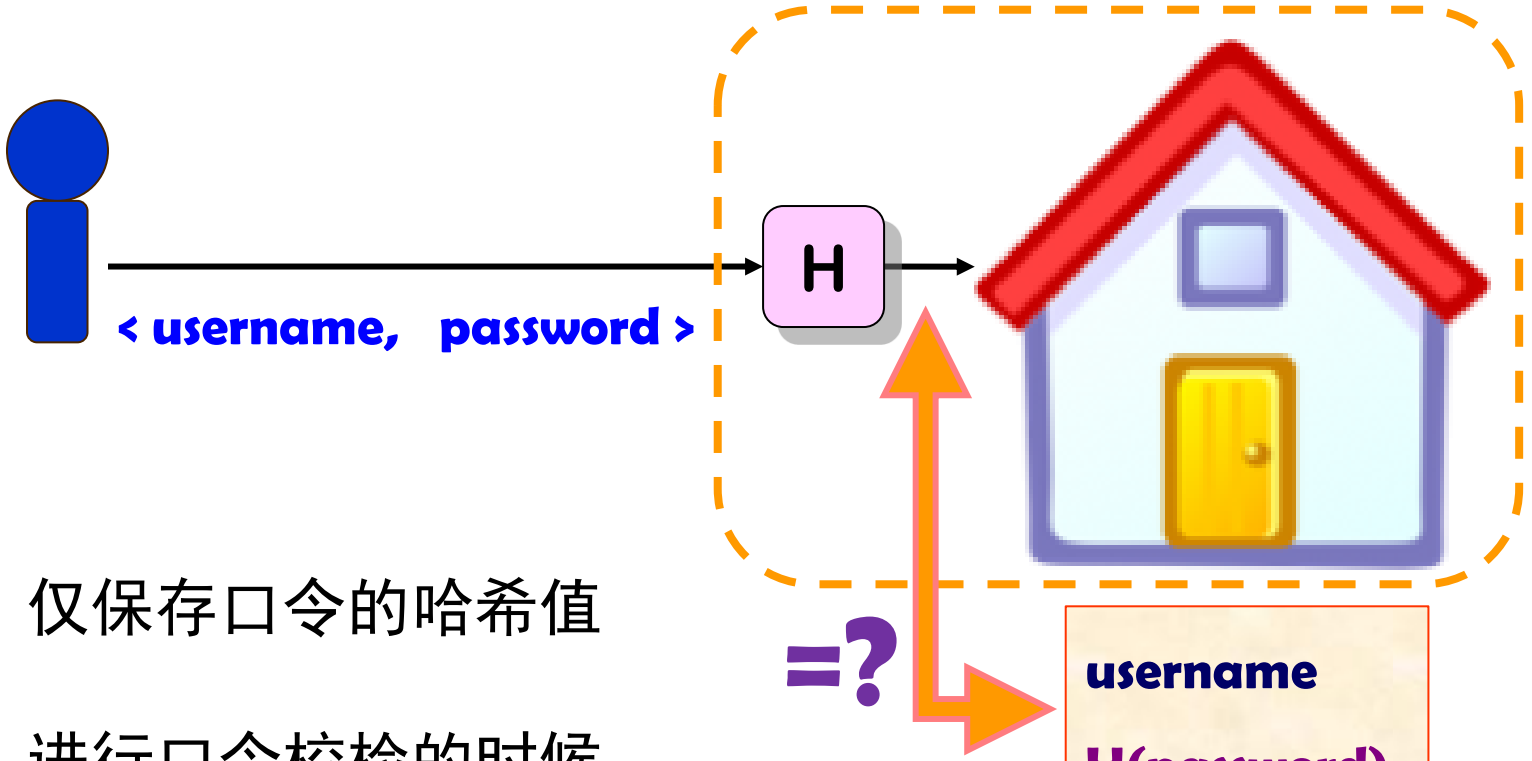
- 设 A 为有限集合，n 为一确定正整数，则 A^* 到 A^n 的函数 $H: A^* \rightarrow A^n$ 可称作一个**哈希函数（hash function）**。

哈希函数

- 哈希函数也称**散列函数**或**杂凑函数**
- 可以将任意长度的输入数据（字符串）打乱、混合、压缩，映射成一个定长的输出字符串
- 于是创建一个叫做“摘要”的数字“指纹”，使得数据量变小，并将数据格式固定下来

哈希函数

用途1——登录系统



仅保存口令的哈希值

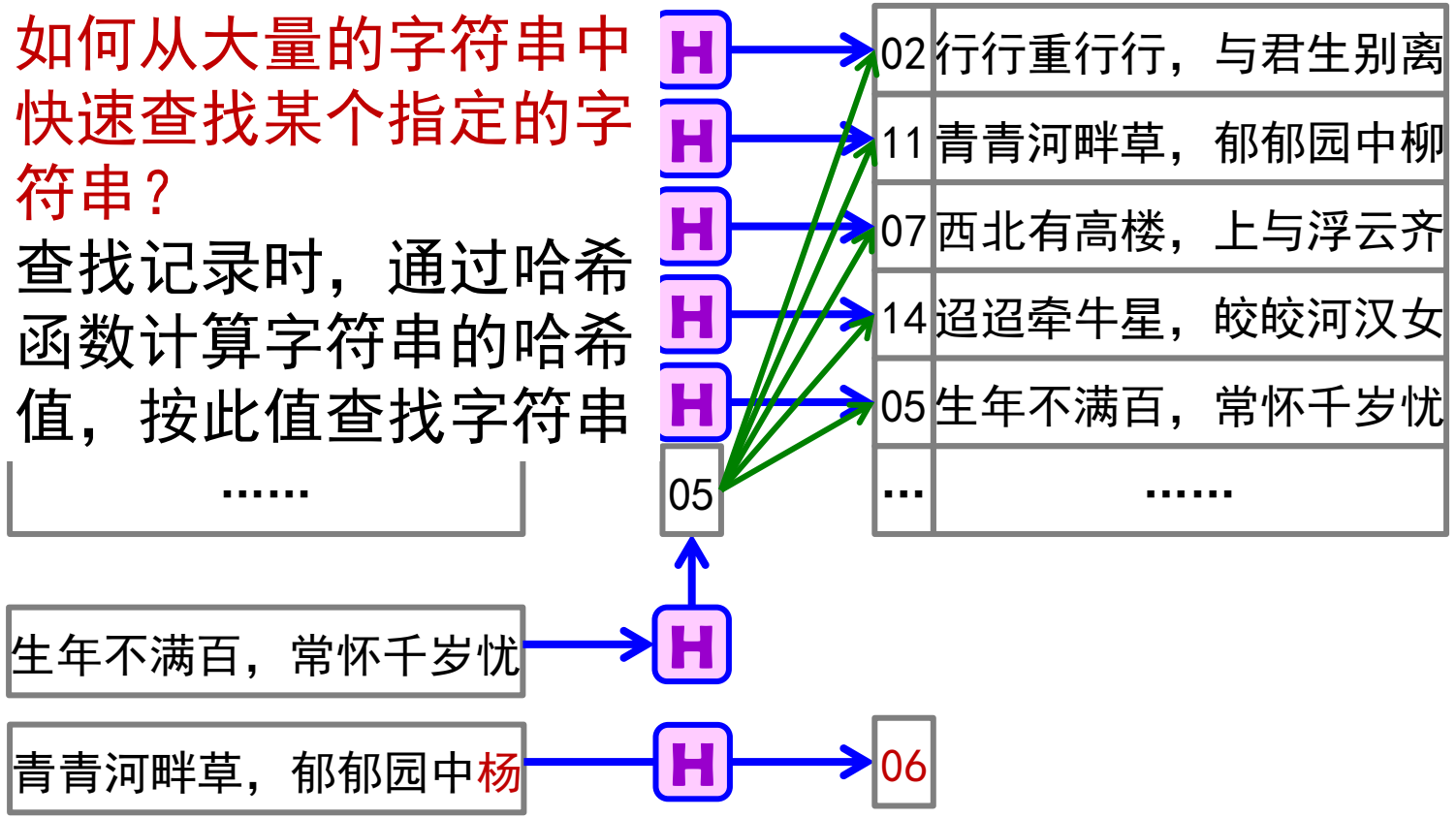
进行口令校检的时候
仅需比对哈希值即可

哈希函数

用途2 —— 加快查找速度

如何从大量的字符串中快速查找某个指定的字符串？

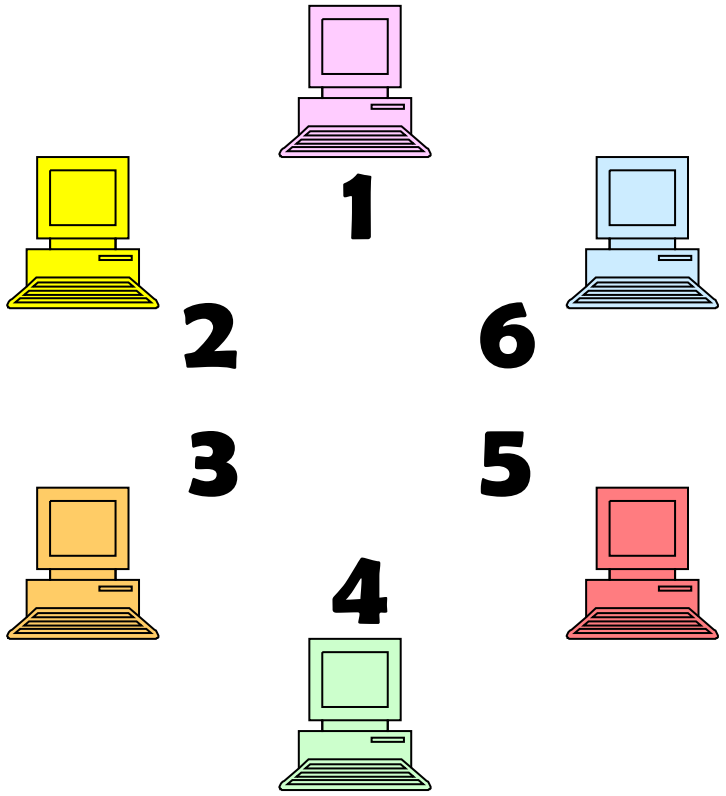
查找记录时，通过哈希函数计算字符串的哈希值，按此值查找字符串



哈希函数

用途3 —— 负载均衡

No.
16126207



$H(No)$

根据任务的编号，通过哈希函数计算得到服务器列表中服务器的序号

将该任务发送给该服务器

哈希函数

- 并非所有这样的函数都是“好”的、适合实际应用的哈希函数
- 一个好的哈希函数一般要满足以下两个要求：
- (a) 冲突尽可能少
  - *H*必定不是单射
  - 必定存在不同的自变量产生相同的哈希值
  - 这种现象称为**冲突(Collision)**或**碰撞**
  - 好的哈希函数应尽可能减少冲突的出现
- (b) 散列值应尽可能均匀地分布在整个值域范围内

哈希函数

- 设 $A=\{0, 1, 2, \ldots, 9\}$，则每一个非负整数都可以看作 A^* 中的一个元素，对于给定的正整数 m，可定义函数 f 为：
  $f(x) = x \bmod m$
- 则 f 是 A^* 到 A^n 的哈希函数（不一定是满射），其中 $n = \lceil \log_{10} m \rceil$
- 例
  - 学生的学号范围取值为20170000至20172999，可取其模1000后的余数作为其哈希值（即学号的末三位）

哈希函数

- 对于密码学中使用的安全哈希函数，有如下要求：
  - 快速性：已知 m，计算 $H(m)$ 是容易的。
  - 单向性：已知 $c=H(m)$，求 m 在计算上是不可行的。
  - 弱抗碰撞性：对给定的消息 m_1，找到另一个与之不同的消息 m_2，使得 $H(m_1)=H(m_2)$ 在计算上是不可行的。
  - 强抗碰撞性：找到两个不同的消息 m_1 和 m_2，使得 $H(m_1)=H(m_2)$ 在计算上是不可行的。
  - 敏感性：$c=H(m)$，c 的每一比特都与 m 的每一比特相关，并有高度敏感性，即每改变 m 的一比特，都将对 c 产生明显影响。

哈希函数

- 例
  - 学生的学号范围取值为20170000至20172999，可取其模1000后的余数作为其哈希值（即学号的末三位）
  - ↑该哈希函数在密码学上是不适用的

End