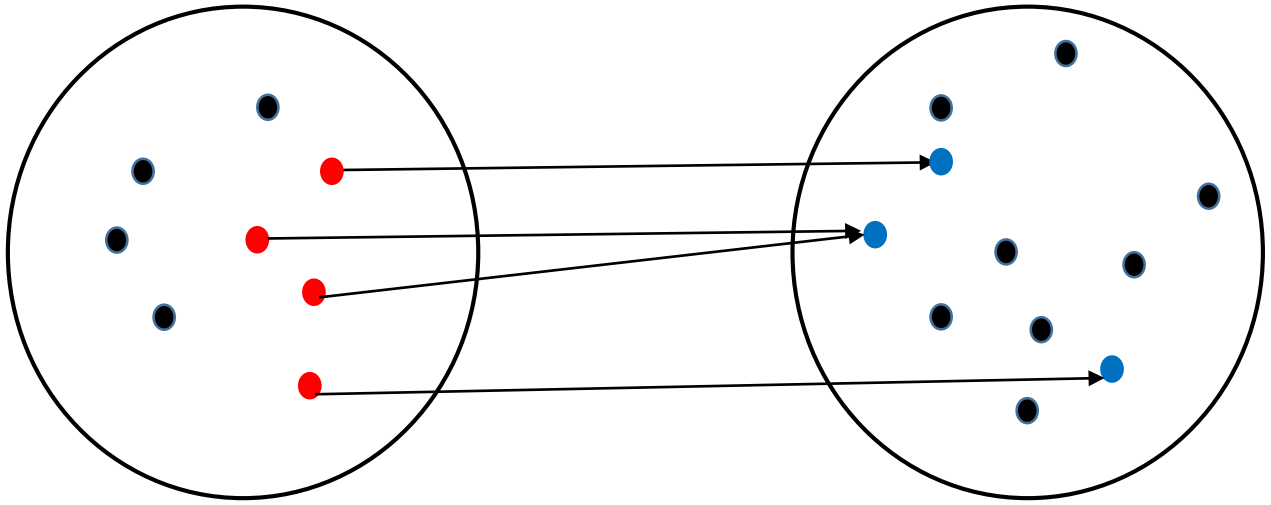
COMP 250

Lecture 26

hashing

Nov. 9, 2016

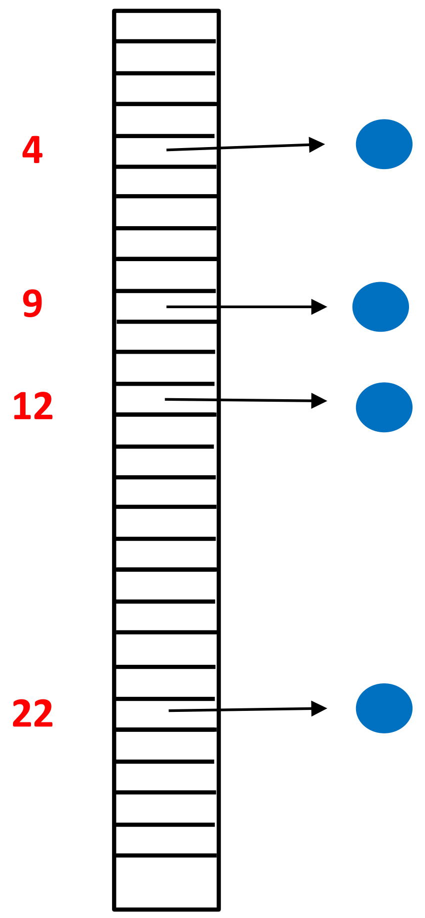
RECALL: Map



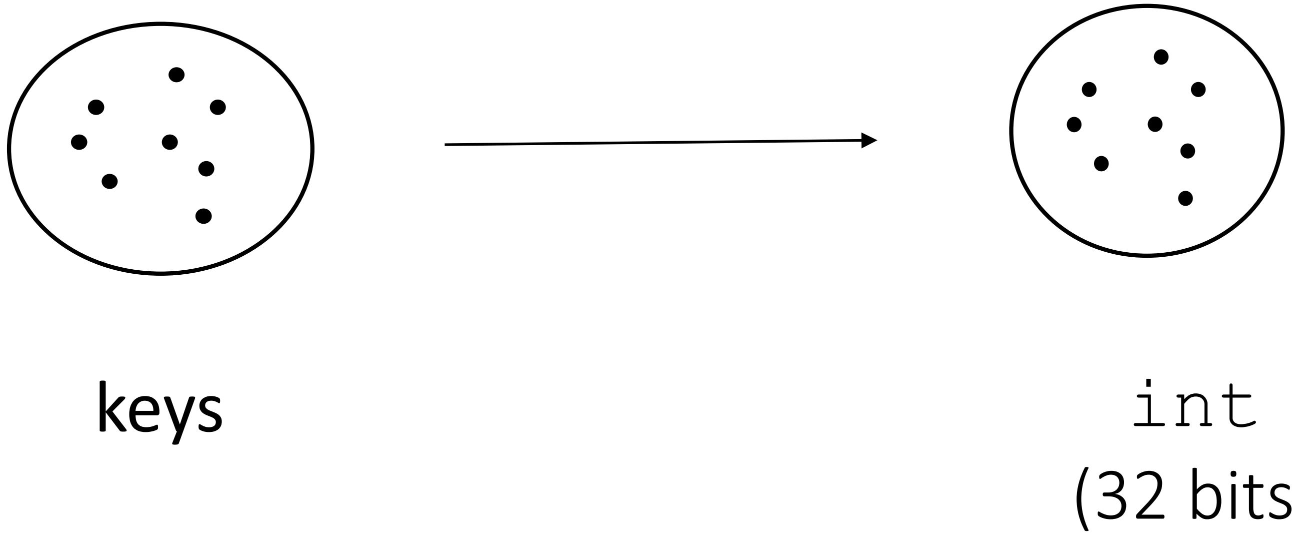
keys (type K)　　　　values (type V)

Each (key, value) pairs is an “entry”.
For each key, there is at most one value.

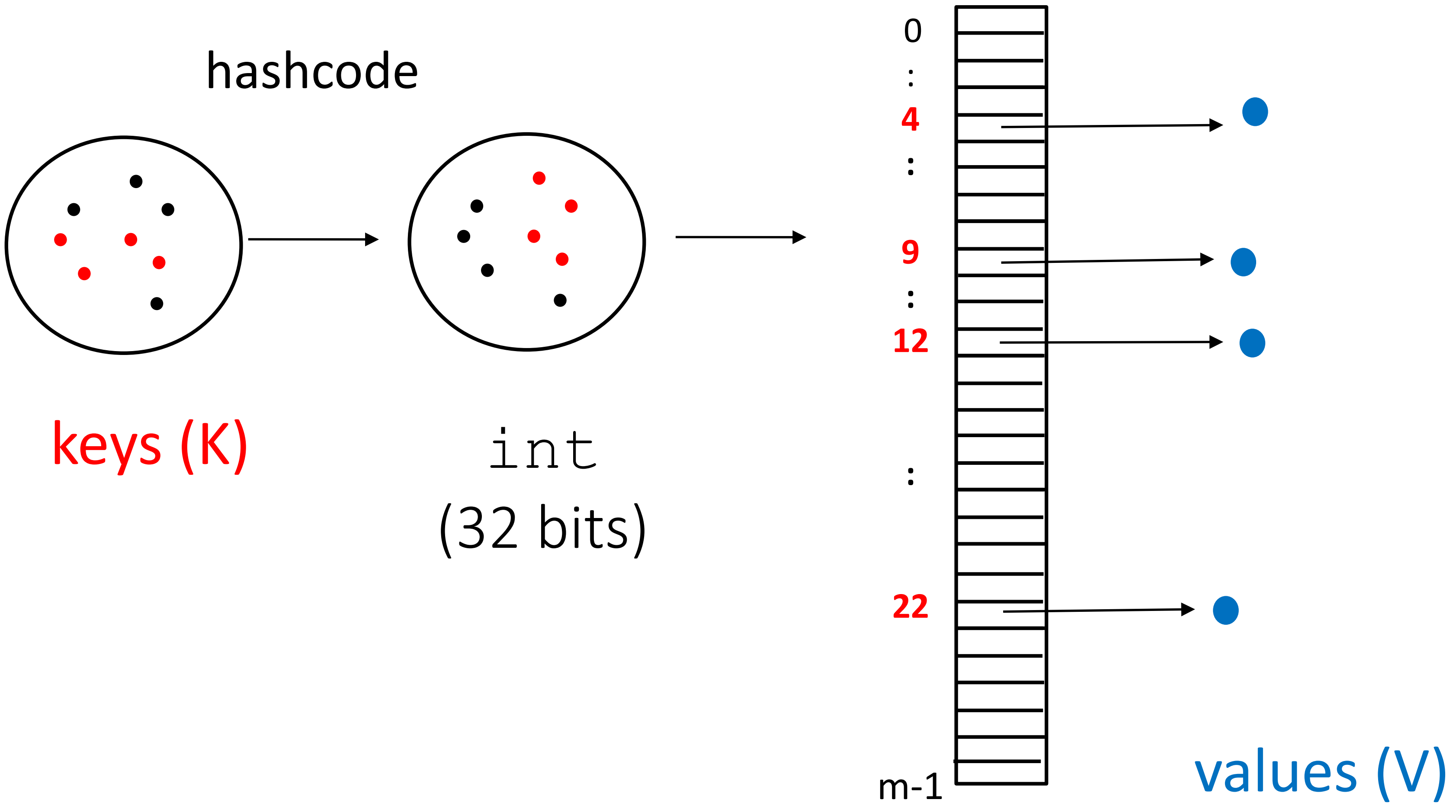
RECALL Special Case: keys are unique positive integers in small range

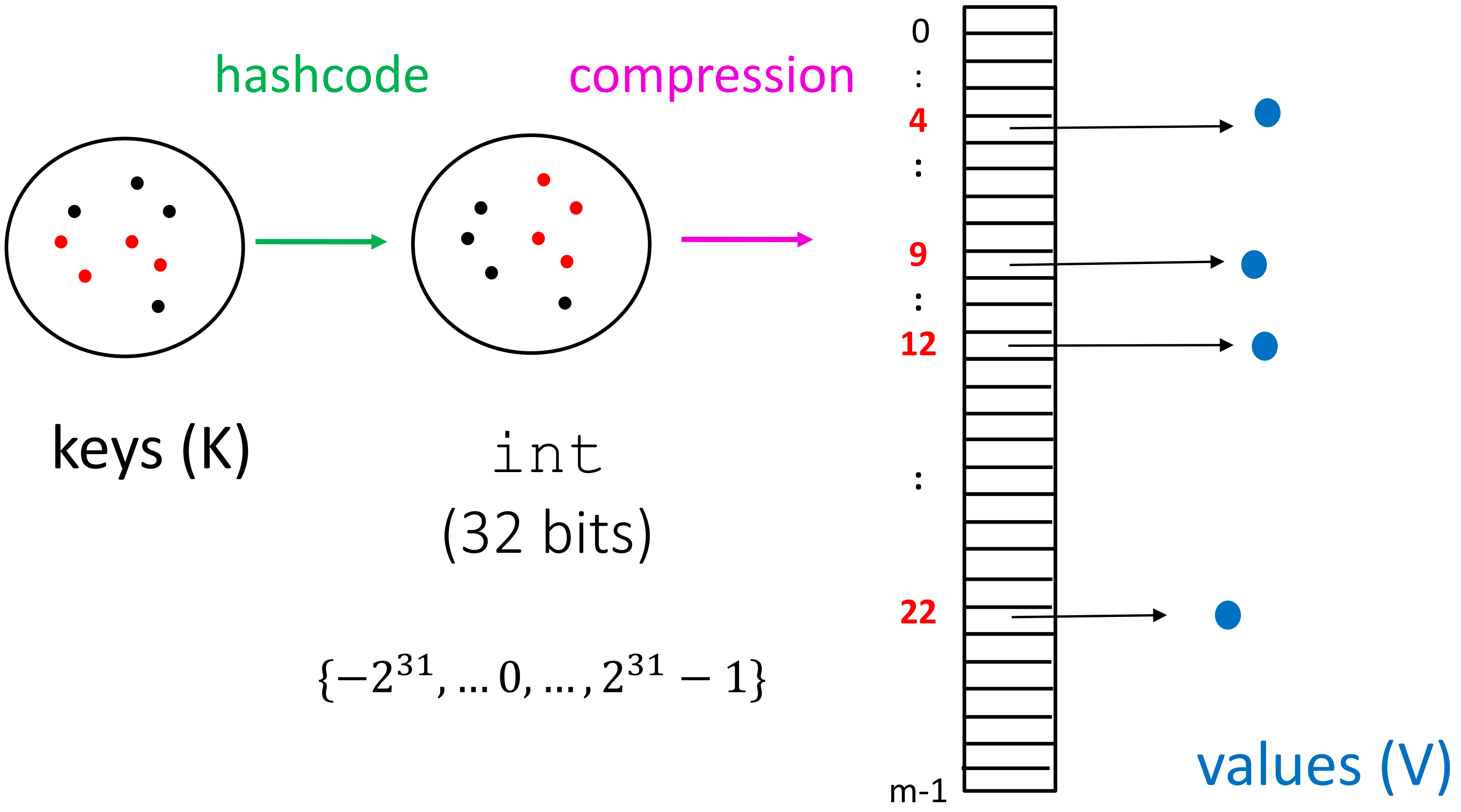


hashcode



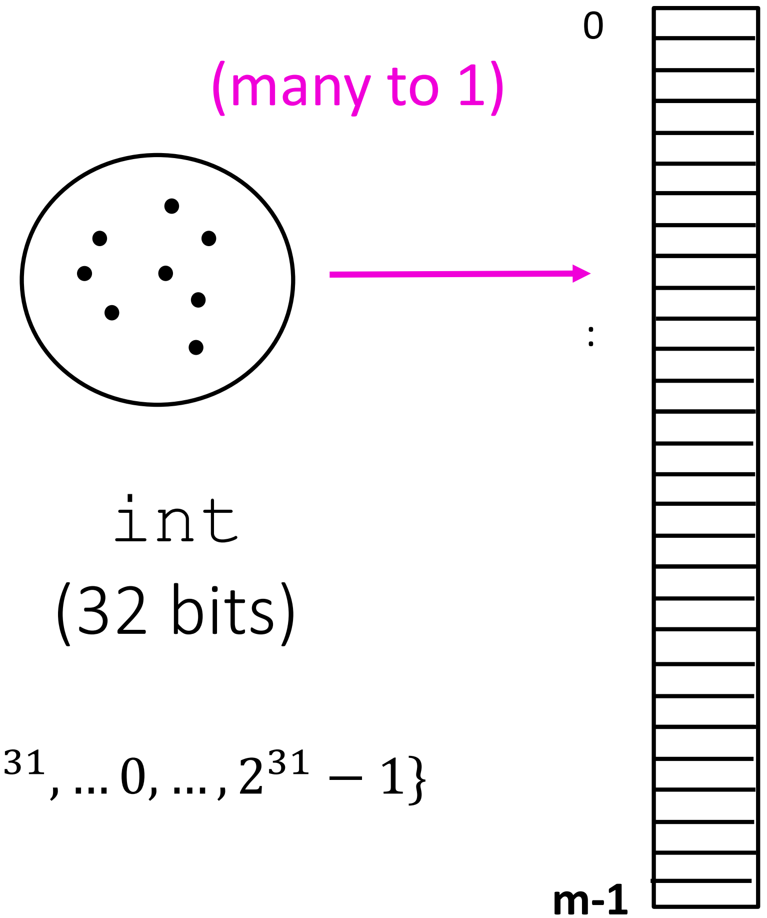
Today's plan: Map Composition




hashcode
compression
keys (K)
int
(32 bits)
$\{-2^{31}, \dots 0, \dots, 2^{31} - 1\}$
0
:
4
:
9
:
12
:
22
m-1
values (V)

compression: $i \rightarrow |i| \ mod \ m,$

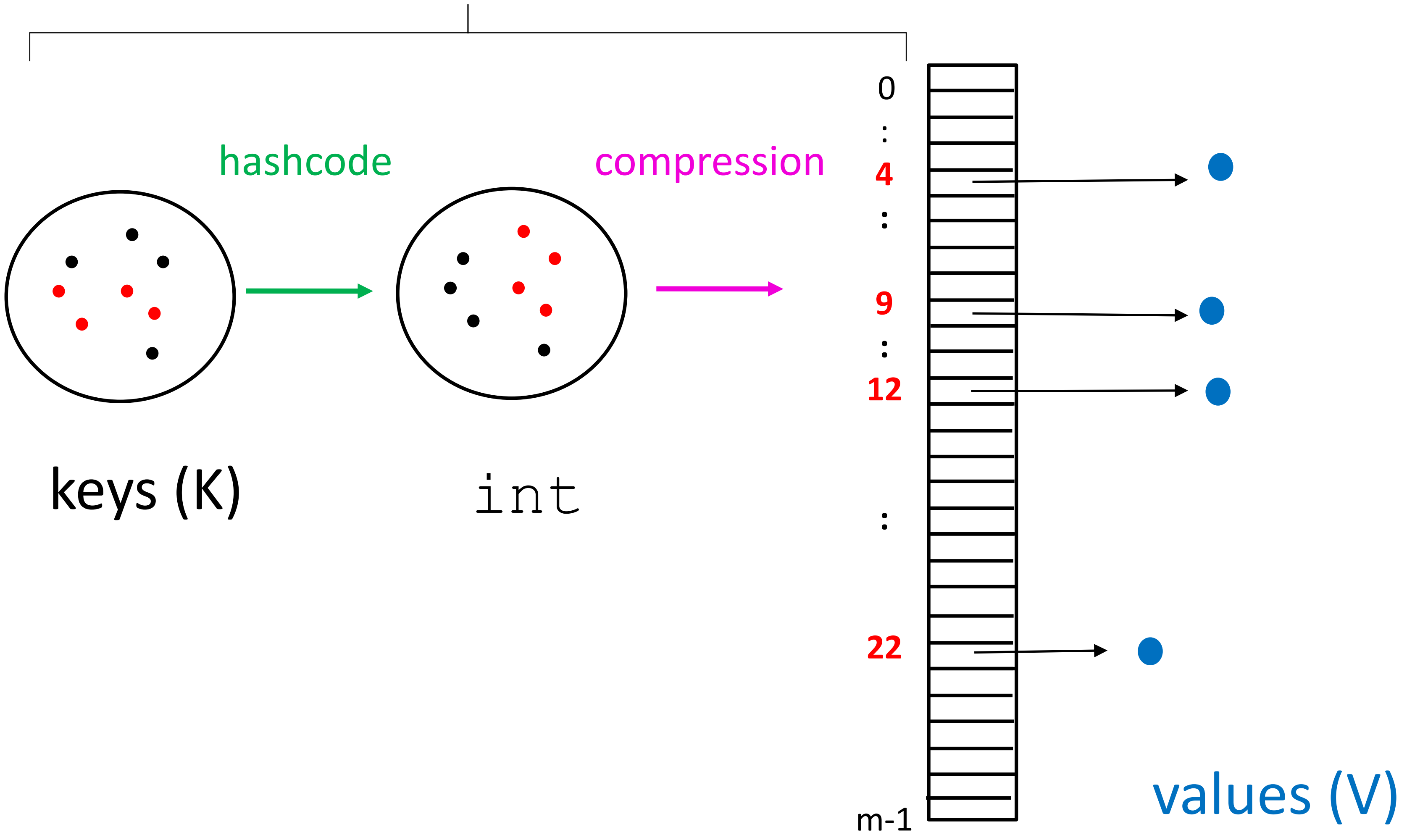
where m is the length of the array.



$\{-2^{31}, \dots 0, \dots, 2^{31} - 1\}$

“hash values”

hash function : keys → {0, …, m-1}



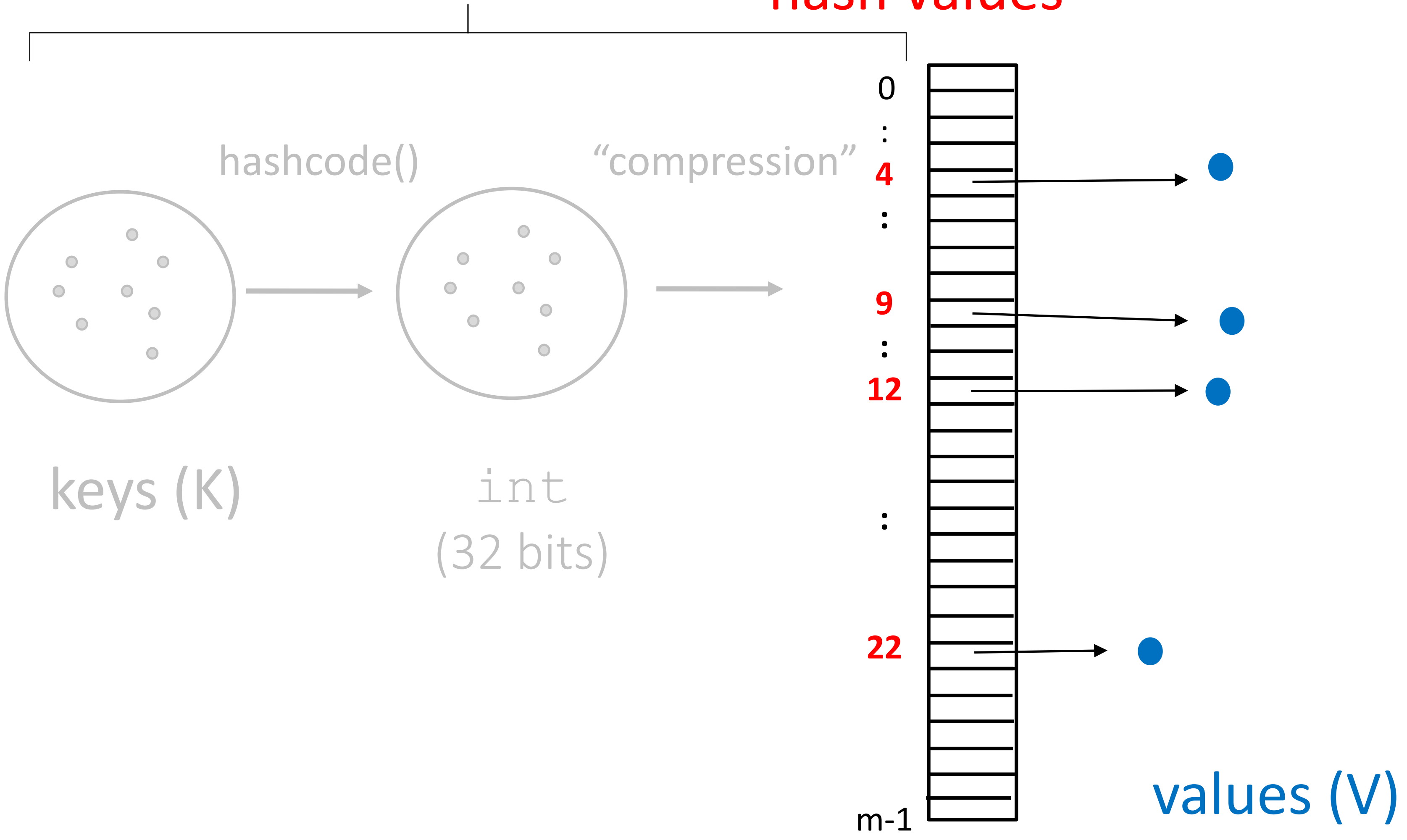
“hash function” ≡ compression ° hashCode

hash code	hash value (hash code % 7)
41	6
16	2
25	4
21	0
36	1
35	0
53	4

Heads up! “Values” is used in two ways.

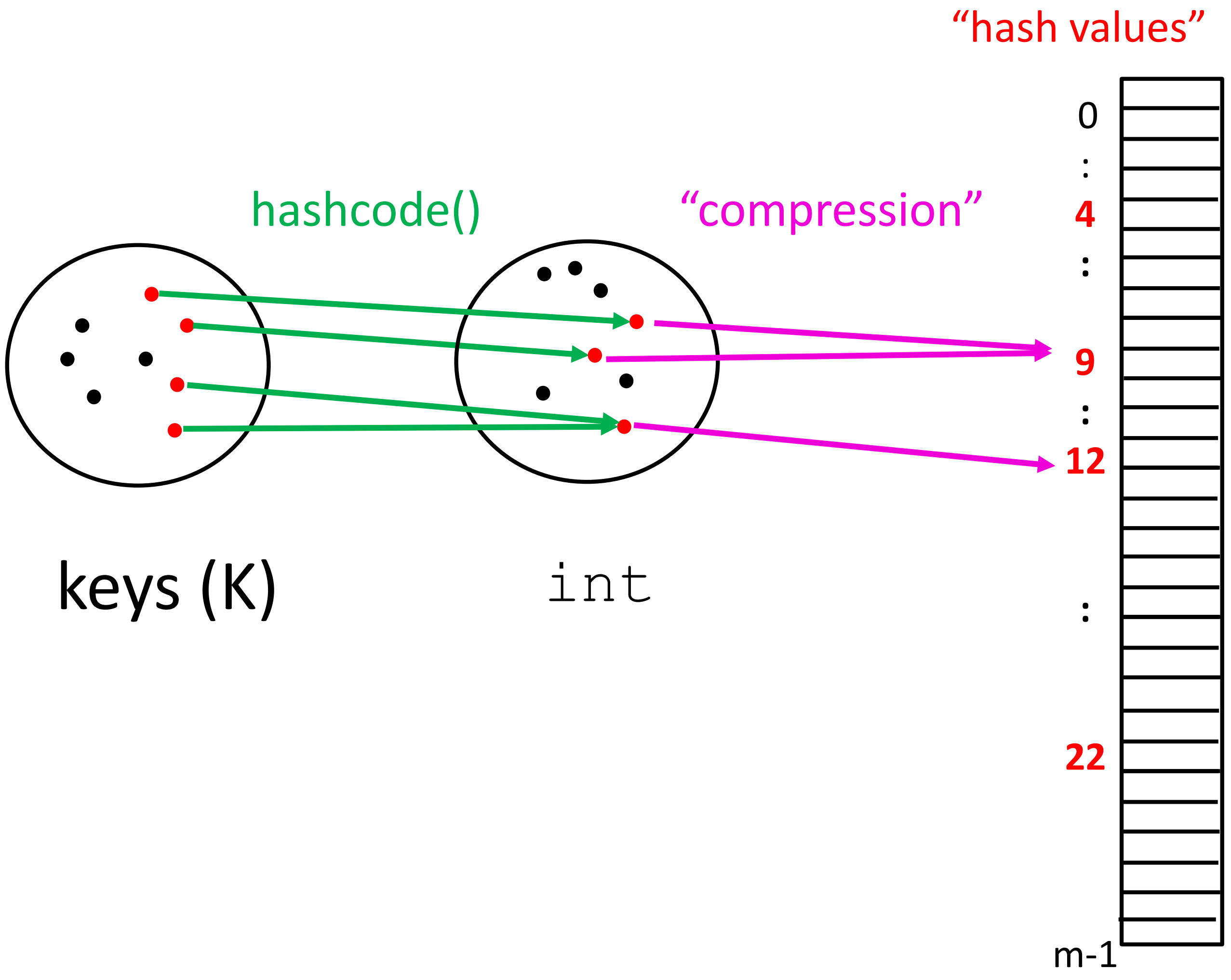
hash function : keys → {0, …, m-1}

“hash values”



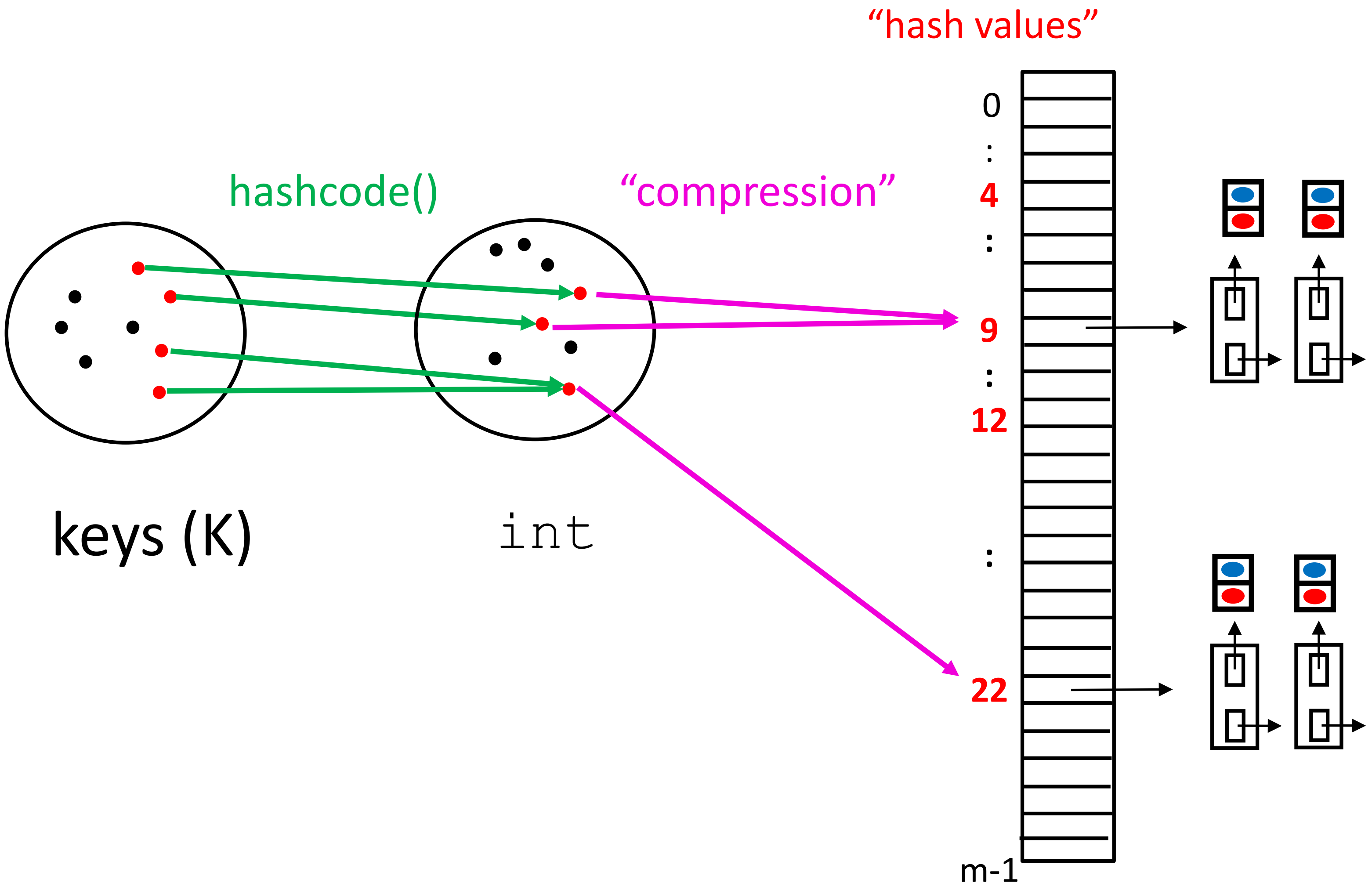
Collision:

when two or more keys k map to the same hash value.



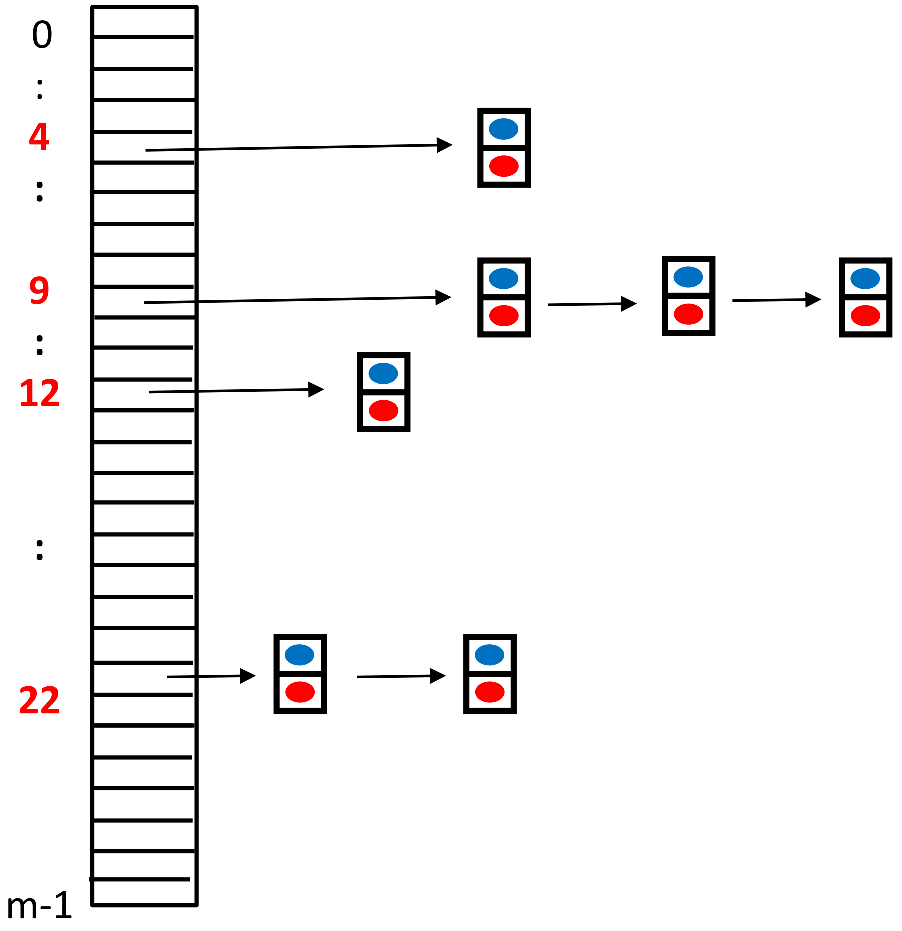
Hash Table (or Hash Map):

each array slot holds a singly linked list of entries



Each array slot + linked list is called a **bucket**.

Note simpler linked list notation here.

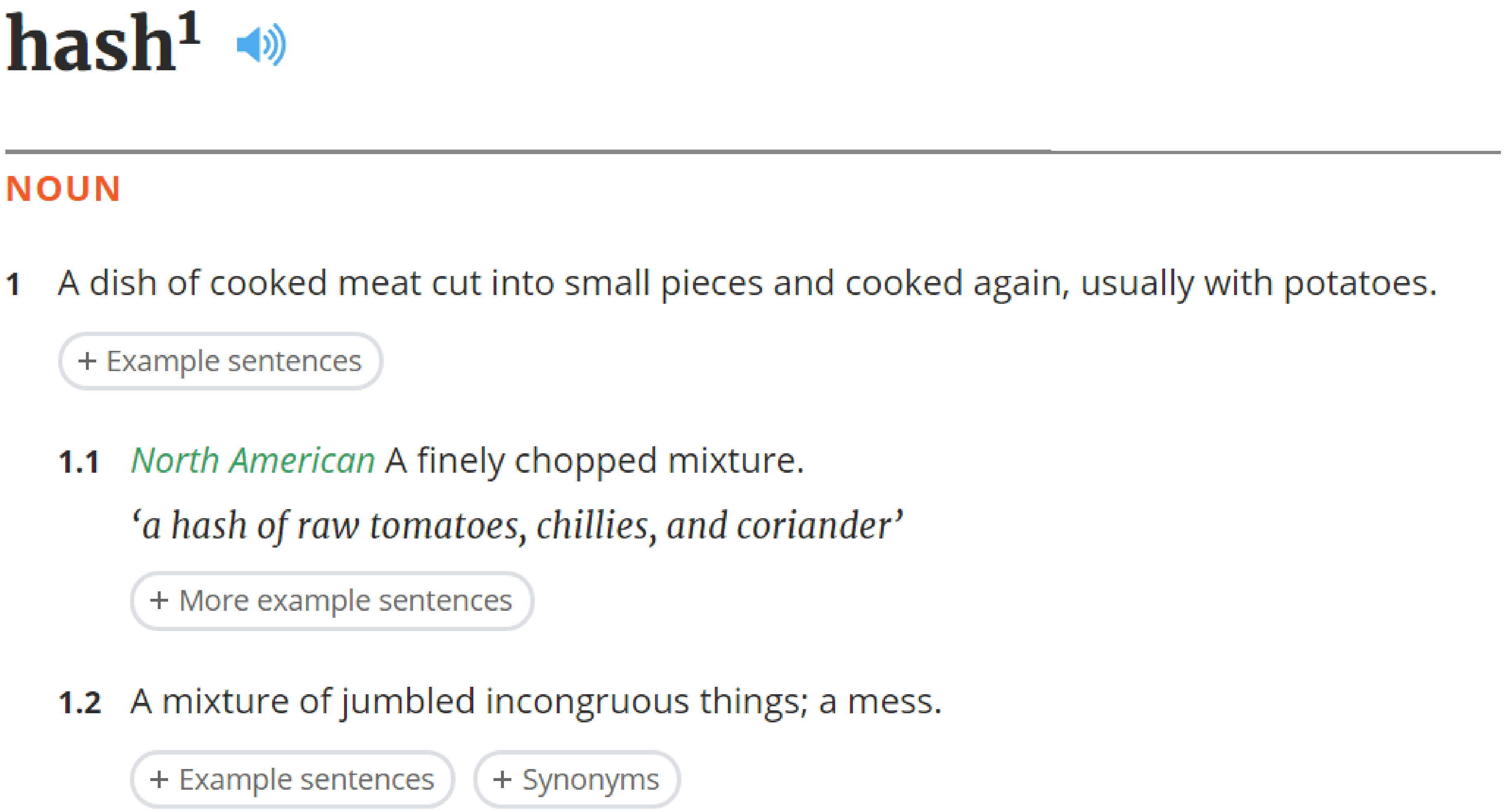


Load factor of hash table

$$\equiv \frac{number\ of\ (key, value)\ pairs\ in\ map}{number\ of\ buckets,\ m}$$

One typically keeps the load factor below 1.

https://en.oxforddictionaries.com/definition/hash#hash_Noun_200



hash[1]

NOUN

1 A dish of cooked meat cut into small pieces and cooked again, usually with potatoes.

+ Example sentences

1.1 *North American* A finely chopped mixture.

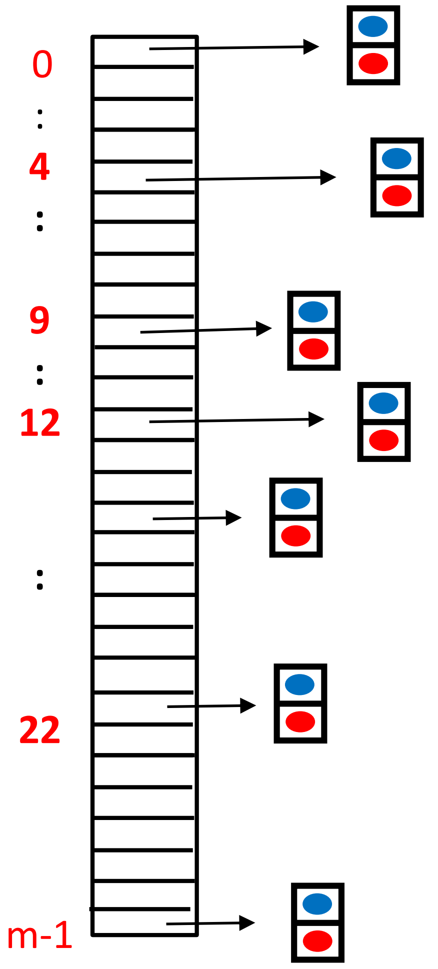
‘a hash of raw tomatoes, chillies, and coriander’

+ More example sentences

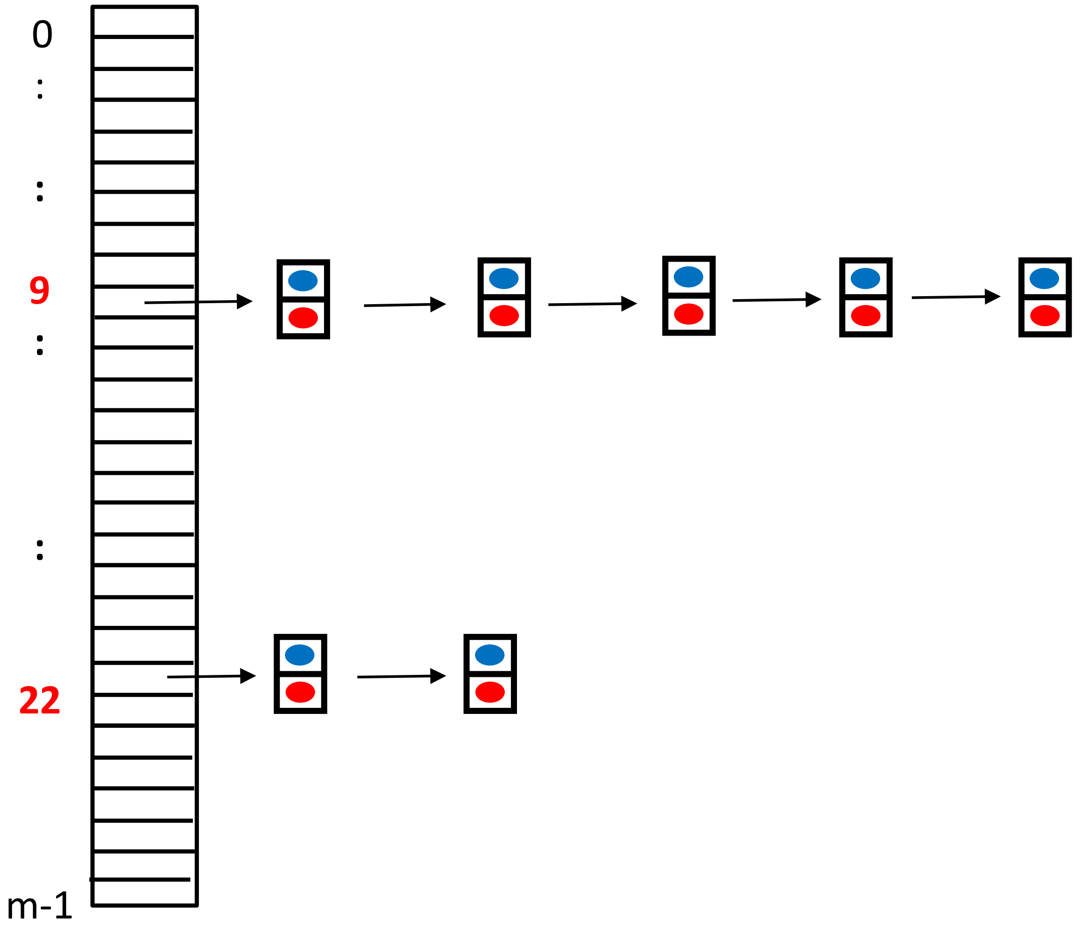
1.2 A mixture of jumbled incongruous things; a mess.

+ Example sentences + Synonyms

Good Hash



Bad Hash



Example

$$h : K \rightarrow \{0, 1, \ldots, m-1\}$$

Example: Suppose keys are McGill Student IDs, e.g. 260745918.

How many buckets to choose ? (~number of entries)

Good hash function? (rightmost 4 digits)

Bad hash function ? (leftmost 4 digits)

Performance of Hash Maps

- put(key, value)
- get(key)
- remove(key)

If load factor is less than 1 and if hash function is good, then operations are O(1) “in practice”.

Note we can use a different hash function if performance is poor.

Java HashMap class

- HashMap<K, V>

- In constructor, you can specify initial number of buckets, and maximum load factor

- How is hash function specified ?

ASIDE: Java HashSet class

- HashSet<E>

- In constructor, you can specify initial number of buckets, and maximum load factor

- How is hash function specified ?

Cryptographic Hashing

e.g. h: String $\rightarrow$ {128 bits}

online tool for computing md5 hash of a string

We want a hash function h( ) such that:

- small changes in the key give very different hash values
- given a hash value, we infer almost nothing about the key.

Example Application (Sketch): Password Authentication

Keys are user names (String)

Values are passwords (String)

{ (user names, passwords) } defines a map.

Password Authentication (unsecure)

Suppose the {(user name, password)} map is stored in a plain text file on the web server where user logs in.

What does the user do to log in?

What does the web server do?

What could a mischievous hacker do?

Password Authentication (secure)

Suppose the {(user name, h(password)) } map is stored in a file on the web server.

What does the user do?

What does the web server do ?

What could a mischievous hacker do?

Announcements

- Discussion board - search before you ask

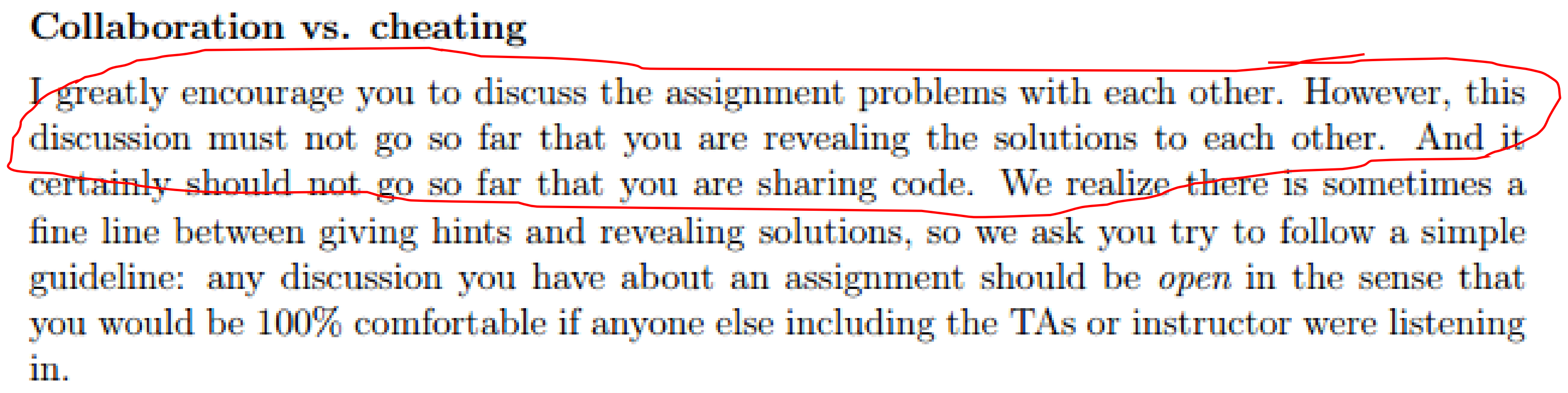
- Facebook
  - What happened last night was not good.

  Is it *plagiarism* ?

  (taking someone else's work/ideas and claiming it is your own)

  No, but handing in answers that you got from facebook is plagiarism.

From the Course Outline.



Collaboration vs. cheating

I greatly encourage you to discuss the assignment problems with each other. However, this discussion must not go so far that you are revealing the solutions to each other. And it certainly should not go so far that you are sharing code. We realize there is sometimes a fine line between giving hints and revealing solutions, so we ask you try to follow a simple guideline: any discussion you have about an assignment should be *open* in the sense that you would be 100% comfortable if anyone else including the TAs or instructor were listening in.