

Goals of this Lecture



Help you learn (or refresh your memory) about:

· Common data structures: linked lists and hash tables

Why? Deep motivation:

- · Common data structures serve as "high level building blocks"
- · A power programmer:
 - Rarely creates programs from scratch
 - · Often creates programs using high level building blocks

Why? Shallow motivation:

- · Provide background pertinent to Assignment 3
- ... esp. for those who have not taken COS 226

2

Symbol Table API



Maintain a collection of key/value pairs

- Each key is a string; each value is an int
- Unknown number of key-value pairs

Examples

- (student name, grade)
 - ("john smith", 84), ("jane doe", 93), ("bill clinton", 81)
- (baseball player, number)
 - ("Ruth", 3), ("Gehrig", 4), ("Mantle", 7)
- (variable name, value)
 - ("maxLength", 2000), ("i", 7), ("j", -10)

Agenda



Linked lists

Hash tables

Hash table issues

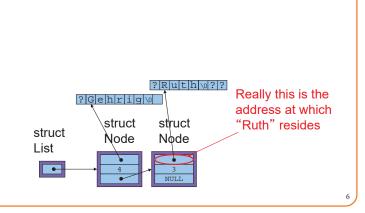
Linked List Data Structure



struct Node const char *key; int value; Your Assignment 3 struct Node *next; data structures will be more elaborate struct List { struct Node *first; Really this is the address at which struct struct "Ruth" resides struct Node Node List

Linked List Data Structure





Linked List Algorithms



Create

- Allocate List structure; set first to NULL
- Performance: O(1) ⇒ fast

Add (no check for duplicate key required)

- · Insert new node containing key/value pair at front of list
- Performance: O(1) ⇒ fast

Add (check for duplicate key required)

- · Traverse list to check for node with duplicate key
- · Insert new node containing key/value pair into list
- Performance: O(n) ⇒ slow

7

Linked List Algorithms



Search

- · Traverse the list, looking for given key
- · Stop when key found, or reach end
- · Performance: ???

8

▶ iClicker Question

Q: How fast is searching for a key in a linked list?

- A. Always fast O(1)
- B. Always slow O(n)
- C. On average, fast
- D. On average, slow

Linked List Algorithms



Search

- Traverse the list, looking for given key
- · Stop when key found, or reach end
- Performance: $O(n) \Rightarrow slow$

ree

- Free Node structures while traversing
- Free List structure
- Performance: O(n) ⇒ slow

10

Agenda



Linked lists

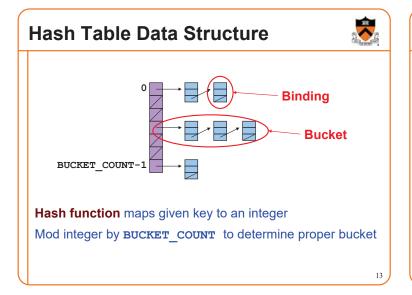
Hash tables

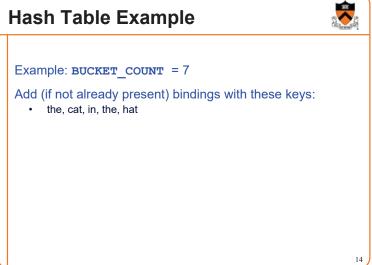
Hash table issues

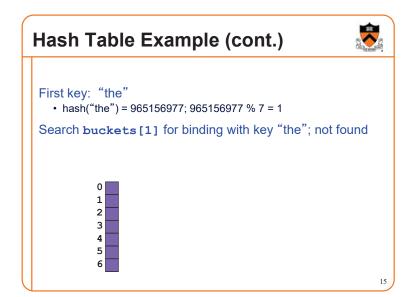
Hash Table Data Structure Array of linked lists

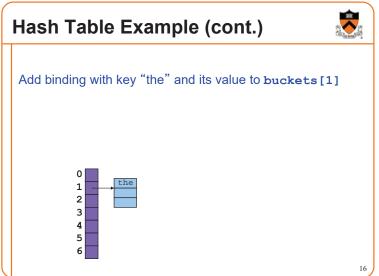


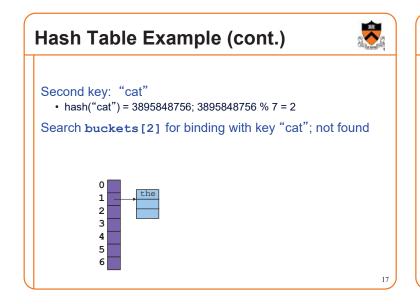
```
Really this is the
                                                            address at which
enum {BUCKET_COUNT = 1024};
                                                            "Ruth", resides
struct Binding
{  const char *key;
  int value;
                                              struct
                                                          struct
                                              Table
                                                          Binding
  struct Binding *next;
                                                                      struct
                                                          "Ruth"
                                                                      Binding
 struct Binding *buckets[BUCKET_COUNT];
      Your Assignment 3
                                              806 NULL
      data structures will
      be more elaborate
```

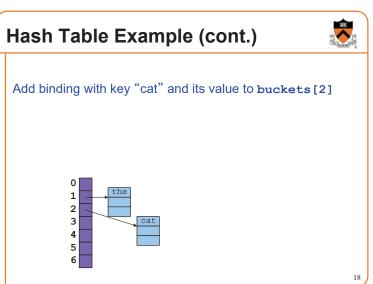


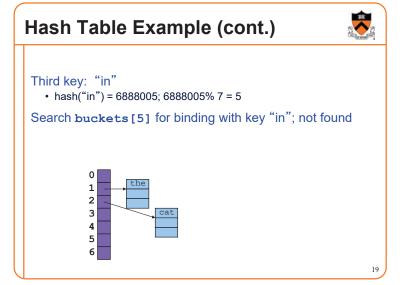


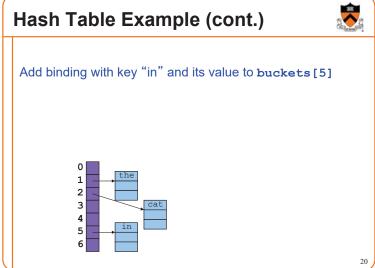


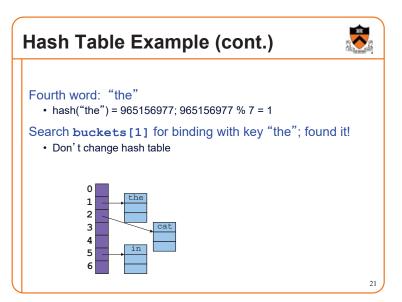


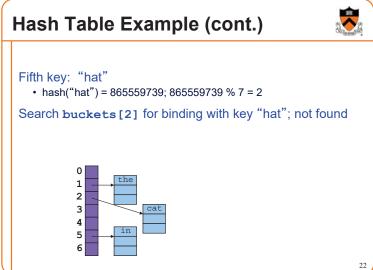


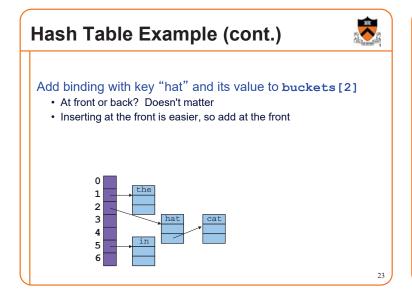


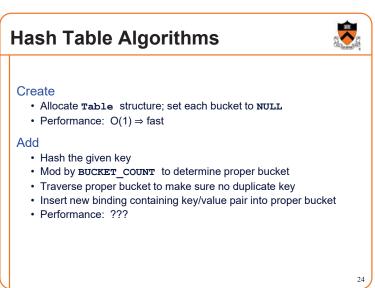












▶ iClicker Question

Q: How fast is adding a key to a hash table?

- A. Always fast
- B. Usually fast, but depends on how many keys are in the table
- C. Usually fast, but depends on how many keys hash to the same bucket
- D. Usually slow
- E. Always slow

Hash Table Algorithms



Search

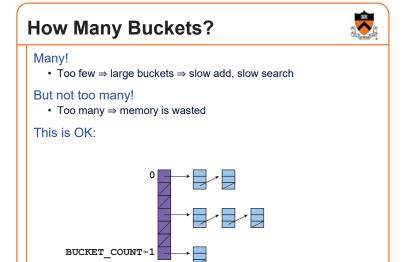
- · Hash the given key
- Mod by BUCKET COUNT to determine proper bucket
- · Traverse proper bucket, looking for binding with given key
- · Stop when key found, or reach end
- Performance: Usually O(1) ⇒ fast

Free

- · Traverse each bucket, freeing bindings
- Free Table structure
- Performance: O(n) ⇒ slow

26

Agenda Linked lists Hash tables Hash table issues



What Hash Function? Should distribute bindings across the buckets well Distribute bindings over the range 0, 1, ..., BUCKET_COUNT-1 Distribute bindings evenly to avoid very long buckets This is not so good: What would be the worst possible hash function?

How to Hash Strings?

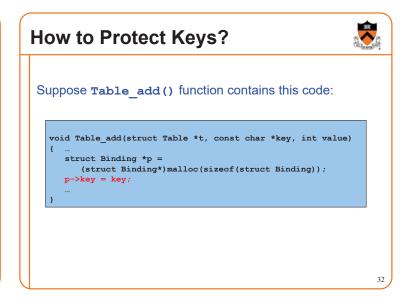


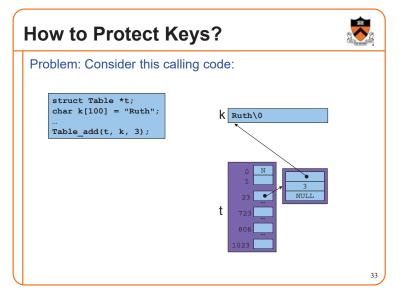
Simple hash schemes don't distribute the keys evenly

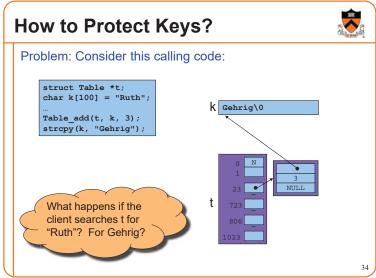
- Number of characters, mod **BUCKET_COUNT**
- Sum the numeric codes of all characters, mod BUCKET COUNT
- ...

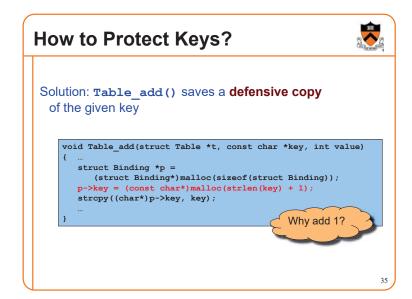
A reasonably good hash function:

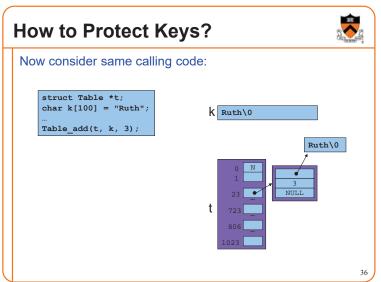
- Weighted sum of characters $\boldsymbol{s}_{\boldsymbol{i}}$ in the string \boldsymbol{s}
 - (Σ aⁱs_i) mod BUCKET_COUNT
- Best if a and BUCKET COUNT are relatively prime
 - E.g., a = 65599, BUCKET COUNT = 1024

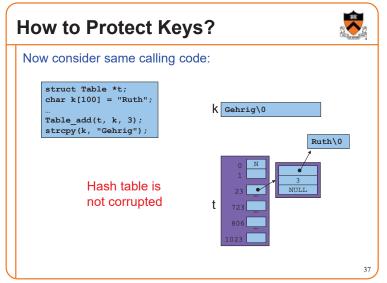


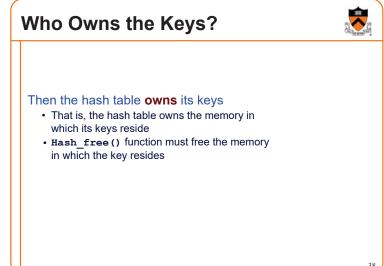


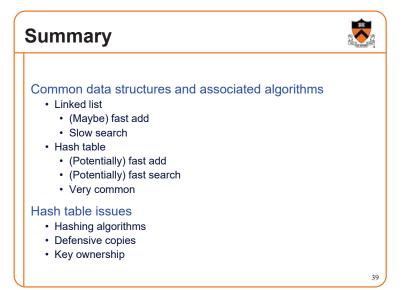


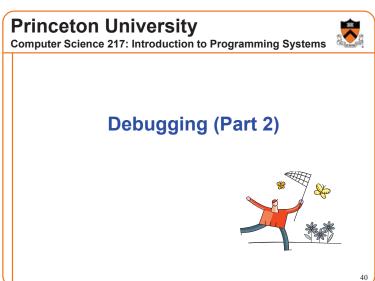




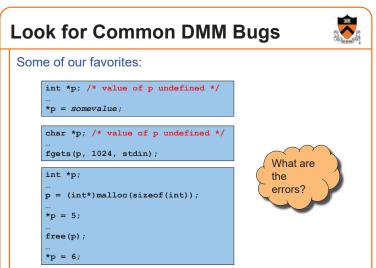


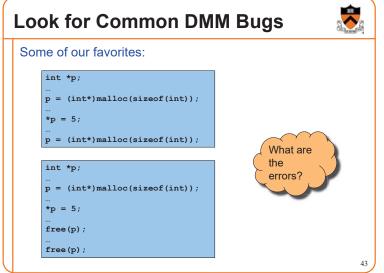














Diagnose Seg Faults Using GDB



Segmentation fault => make it happen in gdb

- Then issue the gdb where command
- · Output will lead you to the line that caused the fault
 - But that line may not be where the error resides!

Agenda



- (9) Look for common DMM bugs
- (10) Diagnose seg faults using gdb
- (11) Manually inspect malloc calls
- (12) Hard-code malloc calls
- (13) Comment-out free calls
- (14) Use Meminfo
- (15) Use Valgrind

46

45

Manually Inspect Malloc Calls



Manually inspect each call of malloc()

· Make sure it allocates enough memory

Do the same for calloc() and realloc()

Manually Inspect Malloc Calls

p = (long double*)malloc(sizeof(p));



```
Some of our favorites:
```

```
char *s1 = "hello, world";
char *s2;
s2 = (char*)malloc(strlen(s1));
strcpy(s2, s1);

char *s1 = "Hello";
char *s2;
s2 = (char*)malloc(sizeof(s1));
strcpy(s2, s1);

long double *p;
p = (long double*)malloc(sizeof(long double*));

long double *p;
```

Agenda



- (9) Look for common DMM bugs
- (10) Diagnose seg faults using gdb
- (11) Manually inspect malloc calls
- (12) Hard-code malloc calls
- (13) Comment-out free calls
- (14) Use Meminfo
- (15) Use Valgrind

Hard-Code Malloc Calls



Temporarily change each call of malloc() to request a large number of bytes

- Say, 10000 bytes
- If the error disappears, then at least one of your calls is requesting too few bytes

Then incrementally restore each call of malloc() to its previous form

• When the error reappears, you might have found the culprit

Do the same for calloc() and realloc()

50

Agenda



- (9) Look for common DMM bugs
- (10) Diagnose seg faults using gdb
- (11) Manually inspect malloc calls
- (12) Hard-code malloc calls
- (13) Comment-out free calls
- (14) Use Meminfo
- (15) Use Valgrind

Comment-Out Free Calls



Temporarily comment-out every call of free ()

- · If the error disappears, then program is
 - · Freeing memory too soon, or
 - Freeing memory that already has been freed, or
 - Freeing memory that should not be freed,
 - Etc.

Then incrementally "comment-in" each call of free ()

• When the error reappears, you might have found the culprit

52

Agenda



- (9) Look for common DMM bugs
- (10) Diagnose seg faults using gdb
- (11) Manually inspect malloc calls
- (12) Hard-code malloc calls
- (13) Comment-out free calls
- (14) Use Meminfo
- (15) Use Valgrind

Use Meminfo



Use the **Meminfo** tool

- Simple tool
- · Initial version written by Dondero
- Current version written by COS 217 alumnus RJ Liljestrom
- · Reports errors after program execution
 - Memory leaks
 - · Some memory corruption
- User-friendly output

Appendix 1 provides example buggy programs

Appendix 2 provides Meminfo analyses

53

Agenda



- (9) Look for common DMM bugs
- (10) Diagnose seg faults using gdb
- (11) Manually inspect malloc calls
- (12) Hard-code malloc calls
- (13) Comment-out free calls
- (14) Use Meminfo
- (15) Use Valgrind

Use Valgrind



Use the Valgrind tool

- Complex tool
- Written by multiple developers, worldwide
 - See www.valgrind.org
- · Reports errors during program execution
 - · Memory leaks
 - · Multiple frees
 - · Dereferences of dangling pointers
 - · Memory corruption
- · Comprehensive output
 - But not always user-friendly

56

Use Valgrind



Appendix 1 provides example buggy programs

Appendix 3 provides Valgrind analyses

57

Summary



Strategies and tools for debugging the DMM aspects of your code:

- Look for common DMM bugs
- · Diagnose seg faults using gdb
- · Manually inspect malloc calls
- · Hard-code malloc calls
- · Comment-out free calls
- · Use Meminfo
- · Use Valgrind

58

Appendix 1: Buggy Programs



leak.c

```
1. #include <stdio.h>
 2. #include <stdlib.h>
 3. int main(void)
 4. { int *pi;
      pi = (int*)malloc(sizeof(int));
       *pi = 5;
 6.
     printf("%d\n", *pi);
      pi = (int*)malloc(sizeof(int));
 8.
       *pi = 6;
      printf("%d\n", *pi);
10.
11.
       free(pi);
       return 0;
12.
13. }
```

Memory leak:

Memory allocated at line 5 is leaked

Appendix 1: Buggy Programs



doublefree.c

```
1. #include <stdio.h>
2. #include <stdlib.h>
3. int main(void)
4. { int *pi;
5. pi = (int*)malloc(sizeof(int));
6. *pi = 5;
7. printf("%d\n", *pi);
8. free(pi);
9. free(pi);
10. return 0;
11. }
```

Multiple free:

Memory allocated at line 5 is freed twice

Appendix 1: Buggy Programs



danglingptr.c

```
1. #include <stdio.h>
 2. #include <stdlib.h>
 3. int main (void)
 4. { int *pi;
      pi = (int*)malloc(sizeof(int));
5.
      *pi = 5;
6.
7.
     printf("%d\n", *pi);
      free(pi);
     printf("%d\n", *pi);
 9.
10.
      return 0;
11. }
```

Dereference of dangling pointer:

Memory accessed at line 9 already was freed

61

Appendix 1: Buggy Programs



```
toosmall.c
```

```
1. #include <stdio.h>
2. #include <stdlib.h>
3. int main(void)
4. { int *pi;
5. pi = (int*)malloc(1);
6. *pi = 5;
7. printf("%d\n", *pi);
8. free(pi);
9. return 0;
10. }
```

Memory corruption:

Too little memory is allocated at line 5 Line 6 corrupts memory

62

Appendix 2: Meminfo



Meminfo can detect memory leaks:

```
$ gcc217m leak.c -o leak
$ ./leak
5
6
$ 1s
. . . leak.c leak meminfo30462.out
$ meminforeport meminfo30462.out
Errors:

** 4 un-freed bytes (1 block) allocated at leak.c:5
Summary Statistics:
Maximum bytes allocated at once: 8
Total number of allocated bytes: 8
Statistics by Line:
Bytes Location
-4 leak.c:11
4 leak.c:11
4 leak.c:5
4 TOTAL
Statistics by Compilation Unit:
4 leak.c
4 TOTAL
```

Appendix 2: Meminfo



Meminfo can detect memory corruption:

Appendix 2: Meminfo



63

Meminfo caveats:

- Don't mix .o files built with gcc217 and gcc217m
- meminfo*.out files can be large
 - · Should delete frequently
- Programs built with gcc217m run slower than those built with
 - Don't build with gcc217m when doing timing tests

Appendix 3: Valgrind



Valgrind can detect memory leaks:

```
$ gcc217 leak.c -o leak
$ valgrind ./leak
==31921== Memcheck, a memory error detector
==31921== Copyright (C) 2002-2012, and GNU GPL'd, by Julian Seward et al.
==31921== Command: leak
==31921== Command: leak
==31921== 5
6
==31921==
==31921== in use at exit: 4 bytes in 1 blocks
==31921== total heap usage: 2 allocs, 1 frees, 8 bytes allocated
==31921==
==31921== LEAK SUMMARY:
==31921== indirectly lost: 4 bytes in 1 blocks
==31921== indirectly lost: 0 bytes in 0 blocks
==31921== still reachable: 0 bytes in 0 blocks
==31921== FROR SUMMARY: 0 errors from 0 contexts (suppressed: 6 from 6)
```

Appendix 3: Valgrind



Valgrind can detect memory leaks:

```
$ valgrind --leak-check=full ./leak
=476== Memcheck, a memory error detector
=476== Copyright (C) 2002-2012, and GNU GPL'd, by Julian Seward et al.
=476== Using Valgrind-3.8.1 and LibVEX; rerun with -h for copyright info
=476== Command: leak
=476== 5

6
=476== = in use at exit: 4 bytes in 1 blocks
=476== total heap usage: 2 allocs, 1 frees, 8 bytes allocated
=476== 4 bytes in 1 blocks are definitely lost in loss record 1 of 1
=476== at 0x4A069EE: malloc (vg replace_malloc.c:270)
=476== tox4A069EE: malloc (vg replace_malloc.c:270)
=476== car6== definitely lost: 4 bytes in 1 blocks
=476== car6== by 0x400565: main (leak.c:5)
=476== car6== possibly lost: 0 bytes in 0 blocks
=476== possibly lost: 0 bytes in 0 blocks
=476== still reachable: 0 bytes in 0 blocks
```

Appendix 3: Valgrind



Valgrind can detect multiple frees:

Appendix 3: Valgrind



Valgrind can detect dereferences of dangling pointers:

Appendix 3: Valgrind



Valgrind can detect memory corruption:

Continued on next slide

Appendix 3: Valgrind



Valgrind can detect memory corruption (cont.):

Continued from previous slide

```
==436==
==436== HEAP SUMMARY:
==436== in use at exit: 0 bytes in 0 blocks
==436== total heap usage: 1 allocs, 1 frees, 1 bytes allocated
==436==
==436== All heap blocks were freed -- no leaks are possible
==436==
==436== For counts of detected and suppressed errors, rerun with: -v
==436== ERROR SUMMARY: 2 errors from 2 contexts (suppressed: 6 from 6)
```

Appendix 3: Valgrind



70

Valgrind caveats:

- Not intended for programmers who are new to C
 - Messages may be cryptic
- Suggestion:
 - · Observe line numbers referenced by messages
 - Study code at those lines
 - · Infer meanings of messages