Dynamic Memory Allocation

Dynamic 1-Dimensional Arrays

Stefano Quer
Dipartimento di Automatica e Informatica
Politecnico di Torino

Local arrays with variable size

- Array dimensions in C **traditionally** had to be compile-time constants
  - It **was impossible** to declare local arrays of a size matching a variable value
  - In other words, it was **impossible** to write code such

```
scanf ("%d", &n);
...
int v[n];
```

A local variable is used to define the size of a local array

```
void f (int n) {
  int v[n];
  ...
}
```

A formal parameter is used to define the size of a local array

Local arrays with variable size

- The C standard ISO/IEC 9899 1999 (C9X) introduced Variable-Length Arrays (VLA's)
  - They allow the previous definitions
    - Local arrays may have sizes set by variables or other expressions, perhaps involving function parameters
  - In other words, it is now **possible** to write code such

```
scanf ("%d", &n);
...
int v[n];
```

A local variable is used to define the size of a local array

```
void f (int n) {
  int v[n];
  ...
}
```

A formal parameter is used to define the size of a local array

Local arrays with variable size

- However, we **will not** use this sort of constructs for many reasons
  - VLAs are a **subset** of what we can obtain with dynamic memory allocation
  - Run-time allocation is **unsafe**, as the object size is defined at run-time, and there is no proper checking strategy
  - VLAs are **local** objects, and, as such, they cannot be exported
    - They are automatically deallocated once the environment in which they have been created is abandoned, and they cannot be used outside that environment

Problem definition

- Dynamic memory allocation can be used to allocate arrays of the desired size at run-time
- We focus on 1D and 2D arrays
  - Multi-dimensional generalizations are possible and somehow straightforward
- The target is the following
  - How can we define and use and array whose size is known **only** at run-time?
  - We can use the **duality** array ←→ pointers !

Example

Allocate an array to store N integer values

```
int n, *v;

fprintf (stdout, "Introduce n: ");
scanf ("%d", &n);
```

Example

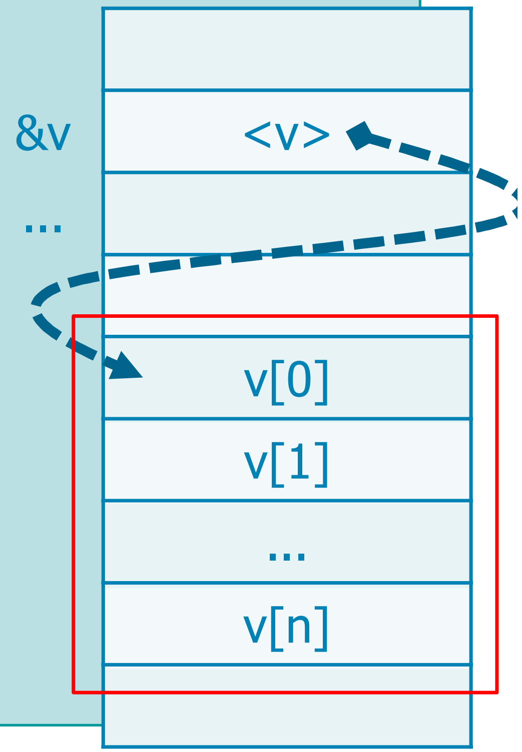
```
int n, *v;

fprintf (stdout, "Introduce n: ");
scanf ("%d", &n);
v = (int *) malloc (n * sizeof (int));
```

At this time, n must be known

As before, function **malloc** is normally used with the **sizeof** operator to allocate a proper quantity of memory. All objects stored in the chunk of memory reserved are usually of the same type.

EXPLICIT CAST that could be also IMPLICIT



Example

stdin-input

stdout-screen
stderr-errorscreen

```
int n, *v;

fprintf (stdout, "Introduce n: ");
scanf ("%d", &n);
v = (int *) malloc (n * sizeof (int));
if (v == NULL) {
  fprintf (stderr, "Memory allocation error.\n");
  exit (1);
}
for (i=0; i<n; i++) {
  fprintf (stdout, "v[%d]: ", i);
  scanf ("%d", &v[i]);
}
for (i=n-1; i>=0; i--) {
  fprintf (stdout, "v[%d]=%d\n", i, v[i]);
}
free (v);
```

Always check the result of a malloc

After allocation, we can use the array the standard way, using the **Array** or the **Pointer** notation

(v+i)

*(v+i)

Example

Same solution, using pointer arithmetic

```
int n, *v, *p;

fprintf (stdout, "Introduce n: ");
scanf ("%d", &n);
v = (int *) malloc (n * sizeof (int));
if (v == NULL) {
  fprintf (stderr, "Memory allocation error.\n");
  exit (1);
}
for (i=0, p=v; i<n; i++, p++) {
  fprintf (stdout, "v[%d]: ", i);
  scanf ("%d", p);
}
          n-1
for (i=0, p--; i>=0; i--, p--) {
  fprintf (stdout, "v[%d]: ", i, *p);
}
free (v);
```

After allocation we can use the array the standard way, using the **Array** or the **Pointer** notation

Example

Same solution, using calloc

```
int n, *v;

fprintf (stdout, "Introduce n: ");
scanf ("%d", &n);
v = calloc (n, sizeof (int));
if (v == NULL) {
  fprintf (stderr, "Memory allocation error.\n");
  exit (1);
}
for (i=0; i<n; i++) {
  fprintf (stdout, "v[%d]: ", i);
  scanf ("%d", &v[i]);
}
for (i=n-1; i>=0; i--) {
  fprintf (stdout, "v[%d]=%d\n", i, v[i]);
}
free (v);
```

We can also use calloc

but useless since then you also initialize

If we do not waste time (initializing the array twice)

Observations

- The typical application which can benefits from dynamic array allocation i the following one
  - Example
    - A file include a list of integers
    - Read the list
    - Save it in another file in reverse order
    - Input file
      - 2 4 6 8 10 12
    - Output file
      - 12 10 8 6 4 2

Observations

- Without dynamic memory allocation, we could
  - Statically allocate the array of size N

```
#define N 100
...
int v[N];
```

    - Read the file, and if the file
      - Has less than N values, terminate the process
      - Has more than N values, stop the program, go back to the editor phase, increase N, recompile the program, and re-run it until the program ends

Observations

- With dynamic memory allocation, we can
  1. Dynamically allocate the array of size N

```
#define N 100
...
int *v;
v = malloc (N * sizeof (int));
if (v==NULL) {...}
```

  - Read the file, and if the file
    - Has less than N values, terminate the process
    - Has more than N values, re-allocate the array

Avoid starting with a malloc of size "1" and then reallocate of size "+1" when reading a new value
This is tremendously inefficient.

Observations

2. Read the file a first time to count-up the number of values inside
   - Allocate the array of the correct size
3. Specify the number of elements on the first row of the file
   - Read this number
   - Allocate the array of the proper size
4. Use a more “dynamic“ data structure
   - Do not use dynamic arrays but some other data structures, e.g., lists

Example

```
#define N 1000

int *v1, *v2;

v1 = malloc (N * sizeof (int));
if (v1 == NULL) { ... }
...
v2 = realloc (v1, 2 * N * sizeof (int));
if (v2 == NULL) {
  fprintf (stderr, "Memory allocation error.\n");
  free (v1);
  exit (1);
}
...
free (v2);
```

Allocation strategy: Double the number of elements at each new allocation

Common errors

```
char v[10];
char *p = malloc (10 * sizeof (char));
```

- sizeof (v)
  - The size of the array (in bytes), i.e., a set of 10 characters each one of 1 byte, that is, 10
- sizeof (p)
  - The size of the pointer p, i.e., 4 or 8 bytes on modern hardware architectures (with 32 or 64 bits)

Modularity

- One of the main problems with dynamic memory allocation is how to export objects
  - How can we make dynamically allocated variables visible from outside the environment in which they have been allocated?

Example

Allocation function

```
void array_create (int *ptr, int n) {
  ptr = (int *) malloc (n * sizeof (int));
  if (ptr == NULL)  { ... }
  return;
}
```

Here I want to allocate the array (and maybe read it from stdin)

Caller (user or client)

```
int n, *v=NULL;

scanf ("%d", &n);
array_create (v, n);
```

Here I want to use it

because it is passed by value

Unfortunately, v is **NULL** here

Modular Allocation

solutions

- To rectify this problem there are at least three possible solutions
  1. Define variables, i.e., pointers, as **global** objects
     - This is the simplest solution, but …
     - Global variables must be avoided as long as possible
       - We will discuss this option (advantages and disadvantages) in the modulary (ADT) section
     - We will avoid this approach as long as possible

Modular Allocation

2. Use the **return statement** to return the variables, i.e., pointers, from the function
   - This is simple enough, but …
   - Unfortunately in C only one value can be returned
     - Even if we can return a C structure including more pointers this can be seen as an awkward solution to solve easy cases

Example

Allocation function

solution 2

```
int *array_create (int n) {
  int *ptr;
  ptr = (int *) malloc (n * sizeof (int));
  if (ptr == NULL)  { ... }
  return ptr;
}
```

Here I want to allocate the array (and maybe read it from stdin)

Caller
(user or client)

Here I want to use it

```
int n, *v=NULL;

scanf ("%d", &n);
v = array_create (n);
```

i.e. copying the value of ptr to v

V is not **NULL** here

Modular Allocation

3. Pass the variables, i.e., pointers, to the function as a **parameter by reference**
   - This is the most complex solution, but …
   - It is also the most general one as we can pass and receive back more than one pointer

Example

Allocation function

solution 3

more complex function: ptr to ptr to int

```
void array_create (int **ptr, int n) {
  *ptr = (int *) malloc (n * sizeof (int));
  if (*ptr == NULL)  { ... }
  return;
}
```

Here I want to allocate the array (and maybe read it from stdin)

Here I want to use it

Caller
(user or client)

```
int n, *v=NULL;

scanf ("%d", &n);
array_create (&v, n);
```

V is generally not **NULL** here

Example

Allocation function

another version

```
void array_create (int **ptr, int n) {
  int *lptr;
  lptr = (int *) malloc (n * sizeof (int));
  if (lptr == NULL)  { ... }
  *ptr = lptr;
  return;
}
```

Here I want to allocate the array (and maybe read it from stdin)

Here I want to use it

Caller (user or client)

```
int n, *v=NULL;

scanf ("%d", &n);
array_create (&v, n);
```

V is generally not **NULL** here

String allocation

- Dynamic strings can be allocated as other dynamic arrays
- However, it is necessary to remind that a string has a termination character '\0'
  - Therefore, it is necessary to **always** reserve space for that character
- Alternatively, we can use the **strdup** function

generally the same, but the special termination character

Example

```
char str[100+1];
char *v;

scanf ("%s", str);
v = malloc ((strlen (str) + 1) * sizeof (char));
if (v == NULL) { ... }
strcpy (v, str);
...
free (v);
```

This +1 may worth several hours of **useless** debugging effort

Notice that **str** may/must have more elements than required, **v** has the tightest possible size

```
char str[100+1];
char *v;

scanf ("%s", str);
v = strdup (str);
...
free (v);
```

With **strdup**

only for strings

General array allocation

- The previous code snippets can be generalized to any arrays
  - Arrays of structures including
    - Static fields
    - Dynamic fields
    - Etc.

Example

```
#define N 100
...

struct student {
  char last_name[N], first_name[N];
  int register_number;
  float average;
};
...
int n;
struct student *v;
...
v = (struct student *)
    malloc (n * sizeof (struct student));
if (v == NULL) { ... }
...
free (v);
```

We can allocate dynamic arrays with static arrays inside

We allocate

We use the structure v

We free it

Example

```
#define N 100
...
struct student {
  char *last_name, *first_name;
  int register_number;
  float average;
};
...
char ln[N], fn[N];
int n;
struct student *v;
...
v = (struct student *)
    malloc (n * sizeof (struct student));
if (v == NULL) { ... }
...
```

We can allocate dynamic arrays with dynamic array fields

But these dynamic array **must** be allocated … We need to allocate the last_name and first_name fields for each element in v

Example

```
for (i=0; i<n; i++) {
  scanf ("%s%s%d%d", ln, fn, &rn, &a);
  last_name = malloc ((strlen(ln)+1)*sizeof(char));
  if (last_name==NULL)  {...}
  first_name = malloc ((strlen(fn)+1)*sizeof(char));
  if (last_name==NULL)  {...}
  strcpy (v[i].last_name, ln);
  strcpy (v[i].first_name, fn);
  v[i].register_number = rn;
  v[i].average = a;
}
...
for (i=0; i<n; i++) {
  free (v[i].last_name); free (v[i].first_name);
}
free (v);
...
```

free everything in opposite order of definition

We allocate the inner fields

We use the structure v

We free it (up-side down)