

Introduction to Programming in C
Prof. Satyadev Nandakumar
Department of Computer Science and Engineering
Indian Institute of Technology, Kanpur

Lecture - 54

(Refer Slide Time: 00:09)



In this lecture, let us look at one particular part of the C compiler which is very important, namely the preprocessor. Technically speaking, the preprocessor is the step before compilation. So, let us understand this in detail.

(Refer Slide Time: 00:18)

A slide with a white background and a black border. The title "The C preprocessor" is at the top in red. Below it are several colored boxes containing text and code examples.

The C preprocessor

We have used statements such as	You may have also seen the use of #define in C programs
<code>#include <stdio.h></code> <code>#include "list.h"</code>	<code>#define PI 3.1416</code> <code>#define MAX 9999</code>

Lines in a C program that start with the hash character # are viewed as macros by the C preprocessor and are transformed by it.

The C preprocessor implements the macro language used to transform C programs **BEFORE** they are compiled.

1. As part of the compilation, first the C preprocessor runs and transforms macros.
2. The resulting file, including the transformed macros is compiled by the C compiler.

We use statement such as hash include within angular brackets `stdio.h`; also hash include within quotes `list dot h`. And you may have seen C code which looks like this. You say hash define `PI` to be `3.1416`; hash define `MAX` to be `9999`, something like this. So lines in a C program, that is star with a hash symbol are called macros. And they are viewed and they are processed by the C preprocessor. Now, the C preprocessor implements what is known as a macro language; part of C. And it is used to transform C programs before they are compiled. So, C preprocessor is the step just before compilation. We do not explicitly call the C preprocessor. But when you write `gcc`, some file name, `gcc` the first step is the preprocessor step. So, as part of the compilation, first the C preprocessor runs and then transforms the macros. The resulting file, including the transformed macros is compiled by the C compiler.

(Refer Slide Time: 01:33)

Header files

- A header file is a file containing C declarations and macro definitions to be shared between several source files.
- header files are included in your program using C preprocessing directive `#include`. For example,

```
#include <stdio.h>  
#include "list.h"
```

Header files serve two purposes.

1. System header files declare interfaces to parts of the operating system (system calls, libraries).
2. Your own header files : contain declarations for interfaces between the source files of your program.

<http://gcc.gnu.org/onlinedocs/gcc-3.0.1/>

So, let us look at header files. A header file is a file containing C declarations, macro definitions etcetera to be shared between several source files. Header files are included in your program using C preprocessor directives “hash include”. For example, we have seen include within angular brackets `stdio dot h` and within quotes `list dot h`. So, header files serves two purposes that we have seen. First is that it could be a system header files. This declares interfaces to part of the operating system including system calls, C libraries and so on. Or you could have your own header files, which you have written to contain declarations of your code of the functions in your code.

(Refer Slide Time: 02:31)

Header files

What does including the header file do?

Including a header file produces the same results as copying the header file into each source file and at the exact place where the corresponding `#include` command was written.

Advantages

1. Related declarations appear in only one place.
2. Single file to change, modify etc.. Modifications are automatically seen by all files that include it.

<http://gcc.gnu.org/onlinedocs/gcc-3.0.1/>

Now, what happens exactly when you include a header file in a C file? Including a header file produces the same results as copying the header file into each source files. So, when you say `include stdio.h`, it is essentially taking the contents of the `stdio.h` file and copy, pasting in to your source code. So, this happens at exactly the place where the corresponding hash include command was written. The advantages of having these header file and hash include is that related declarations appear only in one place. And if you want to change a particular function or declaration of a function, you just have to change it in a single file. And all files which include that header file will automatically see the change.

(Refer Slide Time: 03:24)

Include Syntax: #include

- `#include <file>`: used for system header files. It searches for a file named *file* in a standard list of system directories.
- `#include "file"`: used for header files of your own program.
 - It searches for a file named *file* first in the directory containing the current file, then in the same directories used for `<file>`.

The argument of `#include` behaves like a string constant.

Comments are not recognized, and macro names are not expanded. Thus, `#include <x/*y>` specifies inclusion of a system header file named `x/*y`.

<http://gcc.gnu.org/onlinedocs/gcc-3.0.1/>

So, here is something that we have mentioned in the previous video. If the difference between angular bracket in the double quotes; so the angular bracket is usually used in system header files and it searches for the file named let say `stdio.h` in a standard list of system directories. If you say within double quotes, on the other hand like `list dot h`, it searches for this `list dot h` first in the current directory. If it is not found in the current directory, then it goes again in to the standard list of that. Now, the argument for hash include; whether you include it in an angular bracket or in a double quotes, it behaves like a string constant and it is literally put there. So, if you have like comments, the comments are not recognized as just comments. If you have a star symbol, for example, it will be just put exactly like a star symbol. So, it is just treated as a string constant and no interpretation is done.

(Refer Slide Time: 04:30)

Example

Suppose there is the following header file named header.h

The file p.c includes this header file as follows.

After C preprocessor processes the file, the C compiler will see the input which would be same as p.c written as follows:

```
header.h
char *error="Overflow";

p.c
int x;
#include "header.h"
void main() {
    printf("%s", test());
}
```

```
int x;
char *error="Overflow";
void main() {
    printf("%s", test());
}
```

1. Included files are not limited to declarations and macro definitions; those are merely the typical uses.
2. Any fragment of a C program can be included from another file.

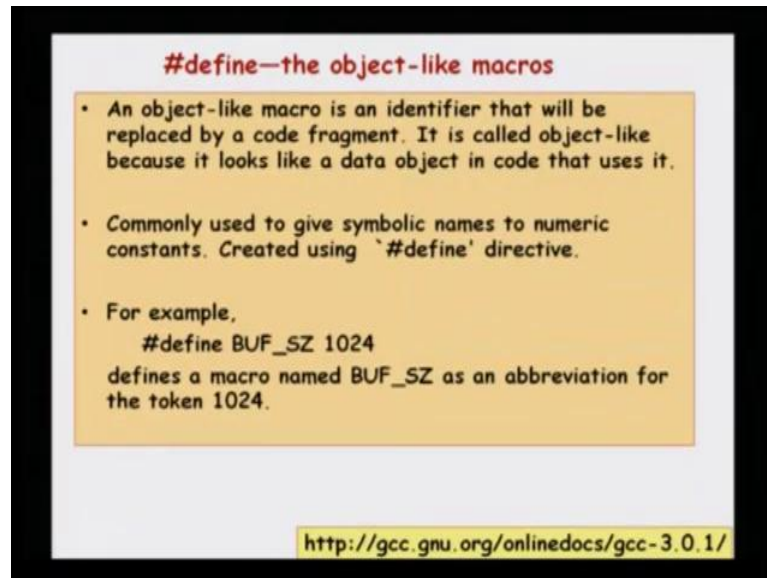
<http://gcc.gnu.org/onlinedocs/gcc-3.0.1/>

Now, let us look at a very special case that could happen in the header file. Typically, you would not do this. So, suppose you have; within the header.h you have a text character star error equal to overflow. Typically, you do not initialize variables in a header file. But, let us say that in a particular header.h, we have this character star error equal to overflow.

Now, in p.c I write this very peculiar thing. I write int x and then in the middle of the code I say hash include header.h. Till now we have always used hash include headers in the beginning of the file. But, suppose what happens if I do it in the middle? Now, after the C preprocessor processes the file, what happens is that whatever text is there in header.h is copy, pasted at that position. So, for example, this code will be transformed by the C preprocessor to look like this. It will say int x and the header.h contain the single line; character star error equal to overflow. So, that text will come here. Now, this transformed text is what the C compiler sees and it will compiled in and produce the object code or the executive.

So, included files are not limited to declarations and macro definitions, these are merely the typical uses. You can put any text there. And when you include that header file, the text will be copy, pasted into the position. Typically though you would want to avoid this, you would want only declarations in a header file.

(Refer Slide Time: 06:13)



#define—the object-like macros

- An object-like macro is an identifier that will be replaced by a code fragment. It is called object-like because it looks like a data object in code that uses it.
- Commonly used to give symbolic names to numeric constants. Created using `#define` directive.
- For example,

```
#define BUF_SZ 1024
```

defines a macro named `BUF_SZ` as an abbreviation for the token `1024`.

<http://gcc.gnu.org/onlinedocs/gcc-3.0.1/>

Now, let us look at some other features that the C preprocessor gives. We have seen in some code, this use of hash define. So, hash define is used for what are known as object-like macros. An object-like macro is basically an identifier and it will be replaced by some code text. It is called object-like because it looks like an object. So, its most common use is to give symbolic names to numeric constants. Suppose, you want to write a program in which the maximum array size is let us say thousand and twenty four, instead of putting thousand and twenty four in several places, the typical usage in a program would be to say hash define Buf size; so buffer size to be thousand and twenty four. So, you have used hash define to define this identifier Buf size and Buf size will be assigned the text 1024. So, this says that I am defining a macro named Buf size. And it is an abbreviation, the short form for the token 1024.

(Refer Slide Time: 07:34)

#define—the object-like macros

- If somewhere *after* the `#define BUF_SZ 1024` directive there is a C statement say:
`char *str = calloc (BUF_SZ, sizeof(char));`
- Then the C preprocessor will recognize and *expand* the macro `BUF_SZ`.
- The C compiler will see the same input as it would if you had written
- `char *str = calloc (1024, sizeof(char));`

• By convention, macro names are written in upper case. Programs are easier to read when it is possible to tell at a glance which names are macros.

<http://gcc.gnu.org/onlinedocs/gcc-3.0.1/>

Now, if somewhere in your code if you say hash define Buf size thousand and twenty four, in all places after that, suppose you say like character star string equal to malloc or calloc Buf size, size of character, what the preprocessor will do is that it will look at this string constant. It is the name of a macro. And it will replace it with thousand and twenty four which is value of the macro. So, the transformed text will look like this; character star string equal to malloc or calloc thousand and twenty four, size of char.

Now, by convention macro names are return in upper case, so that somebody who reach the code will be aware that code this could be a macro; because if I write it in lower case, there are chances that somebody would think, that it is a variable name and look for the variable. So, writing it in capital letters is a way of indicating to the user that this is actually a macro. So, please look at in a header file for example.

(Refer Slide Time: 08:44)

E.g.,

- The C preprocessor scans your program sequentially.
- Macro definitions take effect at the place you write them.
- E.g., the following input to the C preprocessor

```
foo = X;
#define X 4
bar = X;
```

- produces

```
foo = X;
bar = 4;
```

So, the C preprocessor scans through a program sequentially. This is an important thing to understand. And macro definitions take effect at the place you write them. So, let us look at a very specific example to illustrate this point. So, suppose you write foo equal to capital X, after that you have a line say hash define X four and then bar equal to X. What will the C preprocessor do? It will look through the file and say foo equal to X. fine. It does not know what X is. It will not transform that line. Then it sees the hash define X is four. Now, it knows that X is a macro and it has the value four. And then it sees bar equal to X, but now X is a macro. The preprocessor knows about this. So, it will replace X with four. So, that transformed text will be foo equals X bar equal to four. It is natural to imagine that I would have four equal to four. But, that is not what happens; because the way the source code was written, the hash define happened after four equal to X. So, anything that happens before the macros was defined is not changed.

(Refer Slide Time: 10:13)

**A typical project management problem:
#ifndef**

- Suppose we have created a file list.h and list.c.
- There is a file p1.c that needs list functions and creates a new set of functions. So we create p1.h and include "list.h" in p1.h.
- There is a file p2.c that needs list functions and some of the functions created by p1.c. So we create p2.h and include both "list.h" and "p1.h" in p2.h.
- When we compile p2.c, we include list.h twice, once from list.h and another from p1.h. Structure definitions are re-defined—this is a problem.

```
graph TD; list_h[list.h] --> p1_h[p1.h]; list_h --> p2_h[p2.h]; p1_h --> p2_h;
```

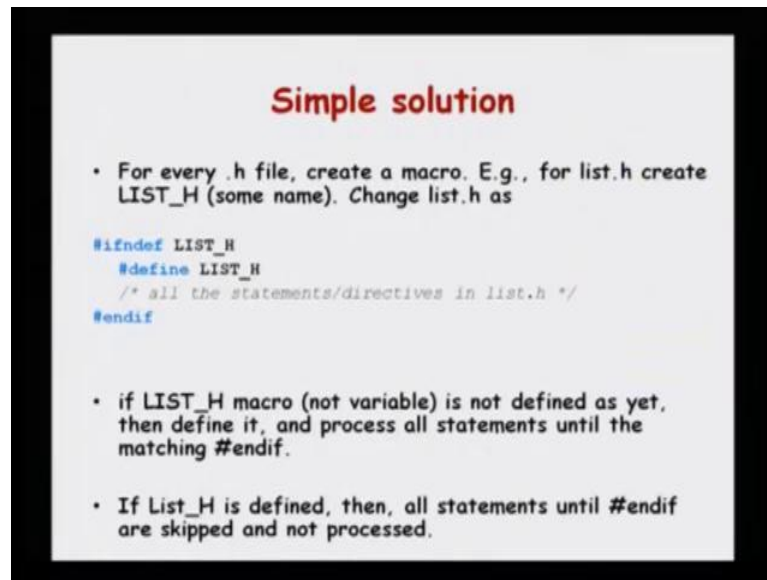
Now, let us conclude this discussion of preprocessor with a very typical project management problem. And we will see a third macro, that is, third operation that is done by the C preprocessor. This is something called hash ifndef. This is used typically when you have multiple files in your project and you need to include multiple header files into a single a source file.

So, let us discuss what is the problem with the particular example? Suppose, we have a list dot h and the list dot c So, I have this header file list dot h. Now, I have a program p 1. c; that needs the list functions and also creates a bunch of new functions. So, its declarations will be included in p1.h. Now in p1.h, I would say include list dot h. So, this is okay. I will have a corresponding p1.c, which will we just say include p 1.h. Now, suppose that I have another file p 2.c; the p 2.c needs some functions in list dot h and some functions n should p1.h. So now, there, when I write p 2.h, I will say include p 1. h and include list dot h. Now, what happens is that when we compile p 2.c, list dot h gets included twice. First, because it indirectly includes list dot h, and second because it includes p1.h which itself includes list dot h. So, list dot h code will be a copy, pasted twice in p 2.h.

So, for example, this is the problem, because if list dot h contains a structure definition, it will be included twice and the compiler will produce an error. So, this is the standard problem in large projects; where you want to include a file, but you do not want to

include it multiple times. So, in this particular example I want to improve list dot h, but I want to avoid the possibility that list dot h is included multiple times leading to compiler errors.

(Refer Slide Time: 12:33)



Simple solution

- For every .h file, create a macro. E.g., for list.h create LIST_H (some name). Change list.h as

```
#ifndef LIST_H
#define LIST_H
/* all the statements/directives in list.h */
#endif
```

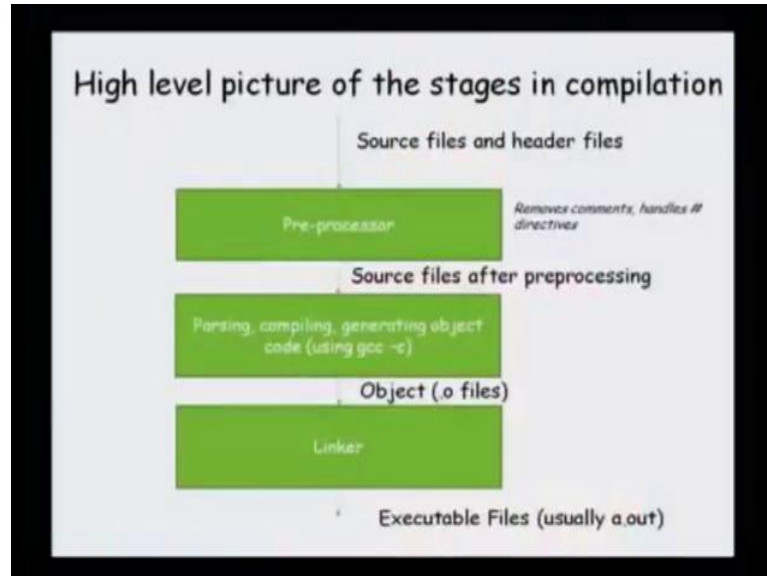
- if LIST_H macro (not variable) is not defined as yet, then define it, and process all statements until the matching #endif.
- If List_H is defined, then, all statements until #endif are skipped and not processed.

So, how do we solve this? So, this is a facility provided by the C preprocessor. You can say you can use this hash ifndef. So, it is saying that if this macro is not defined, then do something. So in list dot h, I will write the following: hash ifndef. This is a macro that I will define. So, usually if a file is list dot h the macro will be named in capital letters; List underscore H. If this macro is not defined, then hash define list dot h. So, this says define list dot h from me and then all the remaining statements in list dot h. And then it will be enclosed in an end if.

So, now, what happens is that suppose list dot h is included for the first time in p 1. h, then list dot h is not defined. So, it will define list dot h and then include list dot h in p1. h. Now, p 2. h includes p 1. h and list dot h. So, now when list dot h is included for the second time, the C preprocessor will look at this statement; ifndef List underscore H. That has been defined because p 2 p 1.h has already defined. It calls it to be defined. So, it says that List underscore H is defined, so I will skip the entire thing until endif. So, this is one way to say that. So, if List_H macro is not defined as yet, then define it and process all statements until the matching endif. If it is already defined, this happens when you are trying to include it for the second time, then all statements until the endif are

skipped. So, you do not copy, paste it for the second time. So, this is the standard way to avoid including one file multiple types.

(Refer Slide Time: 14:40)



So, the high level picture of the stages in compilation. You have; we take up. So, the high level picture of the stages in compilation. You have source files and then it run through this preprocessor, it produces the transformed files. And then after compilation using gcc minus c, it produces object files. And after the object files are done, they are linked to produce the executable files. So when you press gcc, some source file, internally it first runs the C preprocessor, then it runs the compiler and then it is runs the linker. And gcc provides facilities to stop the compilation at any stage. And so for example, we have seen in the previous video that you can stop the compilation just after the compilation itself by using gcc minus c. So, it will produce dot o files. And several dot o files can be later linked to produce the a dot out file.