

Lab Report on:

Reverse Engineering

SWE 4802: Software Maintenance Lab



Lutfun Nahar Lota

Assistant Professor

Department of Computer Science and Engineering

Islamic University of Technology

Submitted By:

Abrar Mahabub (200042103)

Mashrur Ahsan (200042115)

Nafisa Maliyat (200042133)

Shanta Maria (200042172)

Table of Contents

Table of Contents.....	1
IDA Pro.....	2
Features of IDA Pro.....	2
C Code.....	3
Analysis of our C code.....	3
Main function Identification.....	4
Cross-referencing Identification.....	5
The Functions window.....	7
Output window.....	7
Jump, Search, View options.....	8
Hex View-1.....	9
IDA View-A (Disassembled Code).....	10
Imports.....	10
Exports.....	11
PseudoCode Generation.....	12
Graph View of Code Flow.....	12

IDA Pro

IDA (Interactive Disassembler) is a comprehensive tool used for reverse engineering, which means understanding how a program works without having access to the source code. IDA offers a range of useful features such as disassembling binary files, displaying code flow visually through graphs and flowcharts, and running a debugger to see what each instruction is doing. It also allows use of Python scripts to automate the analysis process. In addition, it also works with many CPU types and platforms, making it suitable for analyzing different types of programs. The progress of any analysis is saved as .idb file so users can continue their work anytime.

Features of IDA Pro

Some prominent features of this tool are as follows:

- **Disassembly:** Converts executables into symbolic assembly code for many file formats and architectures
- **Decompiler (Pseudocode):** Generates human-readable C-like pseudocode to ease understanding of binary logic (cloud-based in limited versions, local in Pro)
- **Cross-references & Type Reconstruction:** Automatically identifies cross-links, functions, stack frames, variables, and reconstructed data types
- **Interactive Editing:** Analysts can override auto-analysis results—rename, re-flag, retype—for precision reverse-engineering
- **Programmability & Automation:** Support scripting via IDC and IDAPython, plus plugin support to extend capabilities.
- **Plugin Architecture:** Open SDK enables writing GUI enhancements and analysis tools; large ecosystem of community plugins
- **FLIRT Signature Recognition:** Automatically identifies standard library functions to simplify disassembly
- **idalib (Headless Mode):** Use IDA engine as a library for C++ or Python automation, running offline or in batch environments (Pro only)
- **Private Lumina Add-on:** Enterprise feature for sharing function metadata and analysis across teams (Pro only)
- **Teams Add-on:** Enables team collaboration, file version control, and syncing (Pro only)

C Code

This program is designed to determine whether a given year is a leap year. It accepts a numerical input representing a year and applies a sequence of conditional logic. Specifically, the program checks if the year is divisible by 4 but not by 100, except when it is also divisible by 400. If these conditions are satisfied, the year is classified as a leap year, which comprises 366 days. Otherwise, the year is identified as a common year with 365 days.

```
#include <stdio.h>
int main() {
    int year;
    printf("Enter a year: ");
    scanf("%d", &year);

    if (year % 400 == 0) {
        printf("%d is a leap year.", year);
    }

    else if (year % 100 == 0) {
        printf("%d is not a leap year.", year);
    }

    else if (year % 4 == 0) {
        printf("%d is a leap year.", year);
    }
    else {
        printf("%d is not a leap year.", year);
    }

    return 0;
}
```



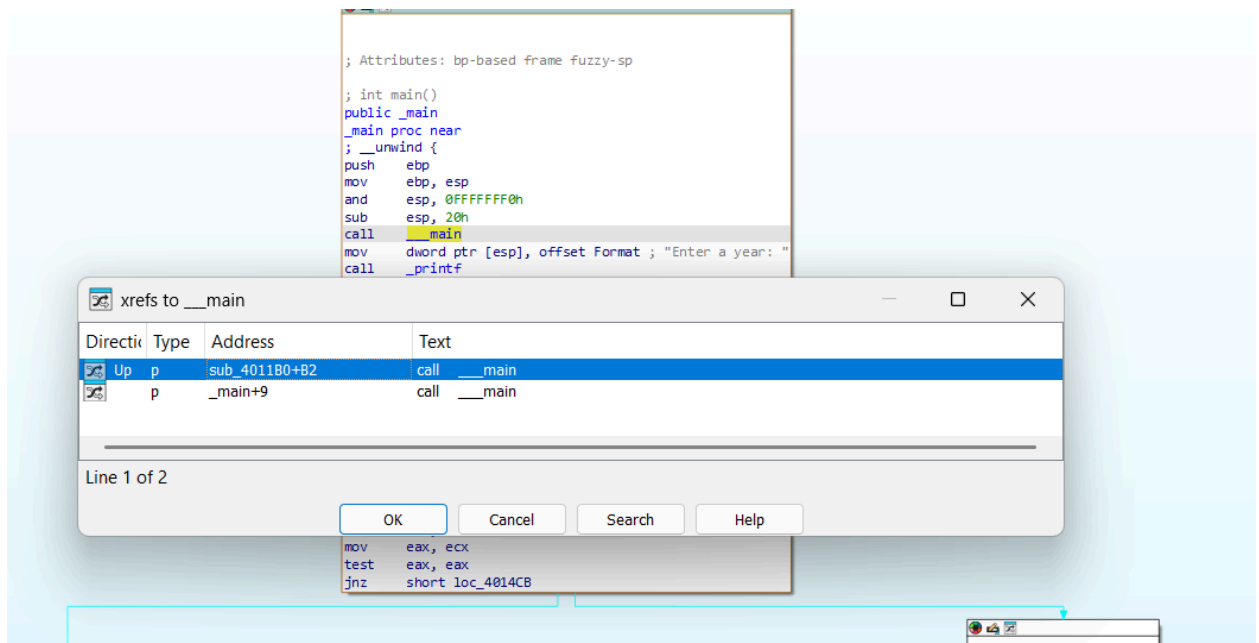
```

.text:00401460 .main proc near ; CODE XREF: sub_4011B0+D3↑p
.text:00401460 ; __unwind {
.text:00401460 push ebp
.text:00401461 mov ebp, esp
.text:00401463 and esp, 0FFFFFFF0h
.text:00401466 sub esp, 20h
.text:00401469 call __main
.text:0040146E mov dword ptr [esp], offset Format ; "Enter a year: "
.text:00401475 call __printf
.text:0040147A lea eax, [esp+1Ch]
.text:0040147E mov [esp+4], eax
.text:00401482 mov dword ptr [esp], offset aD ; "%d"
.text:00401489 call __scanf
.text:0040148E mov ecx, [esp+1Ch]
.text:00401492 mov edx, 51EB851Fh
.text:00401497 mov eax, ecx
.text:00401499 imul edx, ecx
.text:0040149B sar edx, 7
.text:0040149E mov eax, ecx
.text:004014A0 sar eax, 1Fh
.text:004014A3 sub edx, eax
.text:004014A5 mov eax, edx
.text:004014A7 imul eax, 190h
.text:004014AD sub ecx, eax
.text:004014AF mov eax, ecx
.text:004014B1 test eax, eax
.text:004014B3 jnz short loc_4014CB
.text:004014B5 mov eax, [esp+1Ch]
.text:004014B9 mov [esp+4], eax
.text:004014BD mov dword ptr [esp], offset aDIsALeapYear ; "%d is a leap year."
.text:004014C4 call __printf
.text:004014C9 imo short loc_40153A
00000869: 00401469: _main+9 (Synchronized with Hex View-1)

```

Cross-referencing Identification

- In IDA view-A, we will hover the cursor on the function we wish to investigate.
- By pressing X, we can view the cross-references in the window.
- In the window, it will display the location of the function being called.



The screenshot shows a debugger window with assembly code. A red arrow points to a line of code, and a blue arrow points to a jump instruction. A window titled 'xrefs to _printf' is open, displaying a list of references to the `_printf` function. The list includes entries for `_main+15`, `_main+64`, `_main+9E`, `_main+BF`, and `_main+D5`, all of which are calls to `_printf`. The window also shows the address and text of each reference.

Direction	Type	Address	Text
Up	p	_main+15	call _printf
Up	p	_main+64	call _printf
Up	p	_main+9E	call _printf
Up	p	_main+BF	call _printf
Up	p	_main+D5	call _printf

Line 2 of 5

OK Cancel Search Help

- Double-clicking any entry in the list to navigate directly to the usage.

The screenshot shows the 'xrefs to _printf' window with the entry for `_main+D5` selected. The list of references is the same as in the previous screenshot, but the entry for `_main+D5` is highlighted. The window also shows the address and text of each reference.

Direction	Type	Address	Text
Up	p	_main+15	call _printf
Up	p	_main+64	call _printf
Up	p	_main+9E	call _printf
Up	p	_main+BF	call _printf
Up	p	_main+D5	call _printf

Line 5 of 5

OK Cancel Search Help

The screenshot shows the assembly code window with the entry for `_main+D5` selected. The code at `loc_401526` is displayed, showing the usage of `_printf`. The code is as follows:

```

loc_401526:
mov     eax, [esp+1Ch]
mov     [esp+4], eax
mov     dword ptr [esp], offset aDIsNotALeapYea ; "%d is not a leap year."
call    _printf

```

The Functions window

Functions																
Function name	Segment	Start	Length	Locals	Arguments	R	F	L	M	O	S	B	T	=	X	
TopLevelExceptionFilter	.text	00401000	000001A5	0000001C	00000004	R	T	.	.	
sub_401180	.text	00401180	000000E7	0000001C		B	.	.	.	
__mingw32_init_mainargs	.text	004012A0	0000003F	0000003C		R	
__mainCRTStartup	.text	004012E0	00000015	0000001C		
__WinMainCRTStartup	.text	00401300	00000015	0000001C		
__atexit	.text	00401320	00000006	00000000		R	T	.	.	
__onexit	.text	00401330	00000006	00000000	00000004	R	T	.	.	
__gcc_register_frame	.text	00401340	000000E1	0000001C		R	B	.	.	.	
__gcc_deregister_frame	.text	00401430	0000002E	0000001C		R	B	T	.	.	
__main	.text	00401460	000000E1	00000024		R	B	T	.	.	
__setargv	.text	00401550	0000039F	0000005C		R	B	.	.	.	
__cpu_features_init	.text	004018F0	00000107	00000208		R	
__do_global_dtors	.text	00401A00	0000002C	00000000		R	T	.	.	
__do_global_ctors	.text	00401A30	00000046	0000001C		R	
__main	.text	00401A80	0000001C	00000000		R	
TlsCallback_1	.text	00401AA0	00000043	0000001C	0000000C	R	
__dyn_tls_init(xxx)	.text	00401AF0	00000083	0000001C	0000000C	R	
__tregdtor	.text	00401B80	00000003	00000000		R	
sub_401B90	.text	00401B90	00000060	0000001C		R	
__w64_mingwthr_add_key_dtor	.text	00401BF0	00000082	0000001C	00000008	R	
__w64_mingwthr_remove_key_dtor	.text	00401C80	00000098	0000001C	00000004	R	
__mingw_TLSCallback	.text	00401D20	00000097	0000001C	00000008	R	
sub_401D30	.text	00401D30	00000044	0000001C	0000000F	R	

Here we can see several columns in the “Functions” window. Here is what the columns represent:

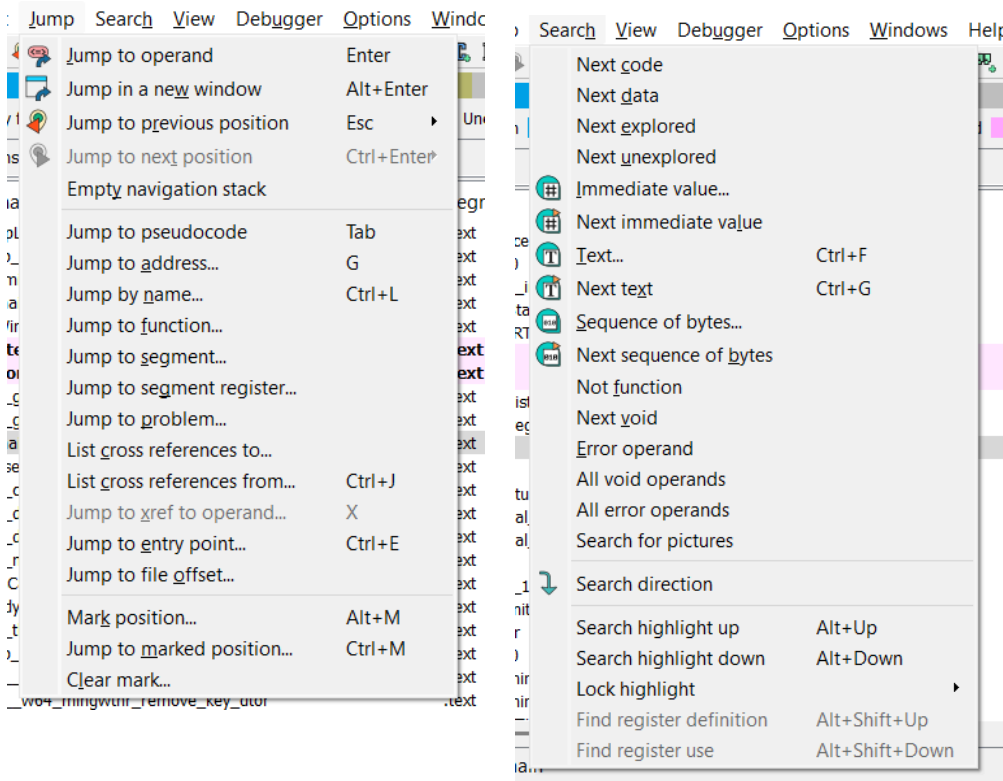
- The first column Function names
- Second column signifies the segment
- Third column shows the start Memory locations, fourth column shows the whole length
- Arguments column shows the number of arguments the functions that in
- Rest of the columns represent different status(s) like referred, library, manual etc.

Output window

```
Output
Please check the Edit/Plugins menu for more information.
Using FLIRT signature: SEH for vc7-14
Propagating type information...
Function argument information has been propagated
The initial autoanalysis has been finished.
Search completed
Search completed
Search completed
```

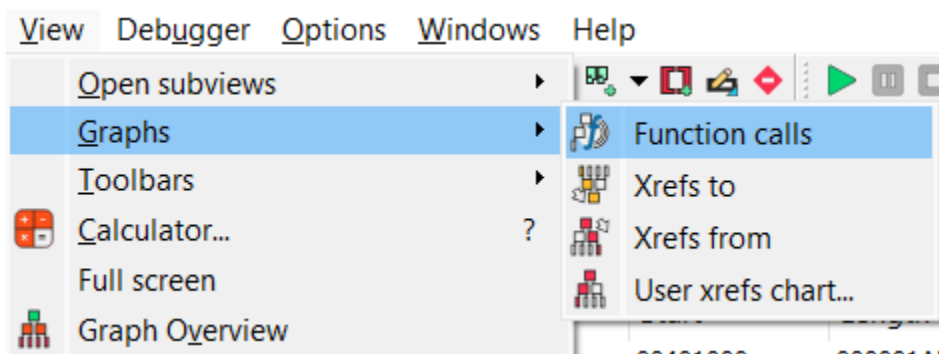
The output window shows all the logs during our analysis of the code. It acts as a logger. It shows the sequence of operations and logs confirmation messages accordingly.

Jump, Search, View options



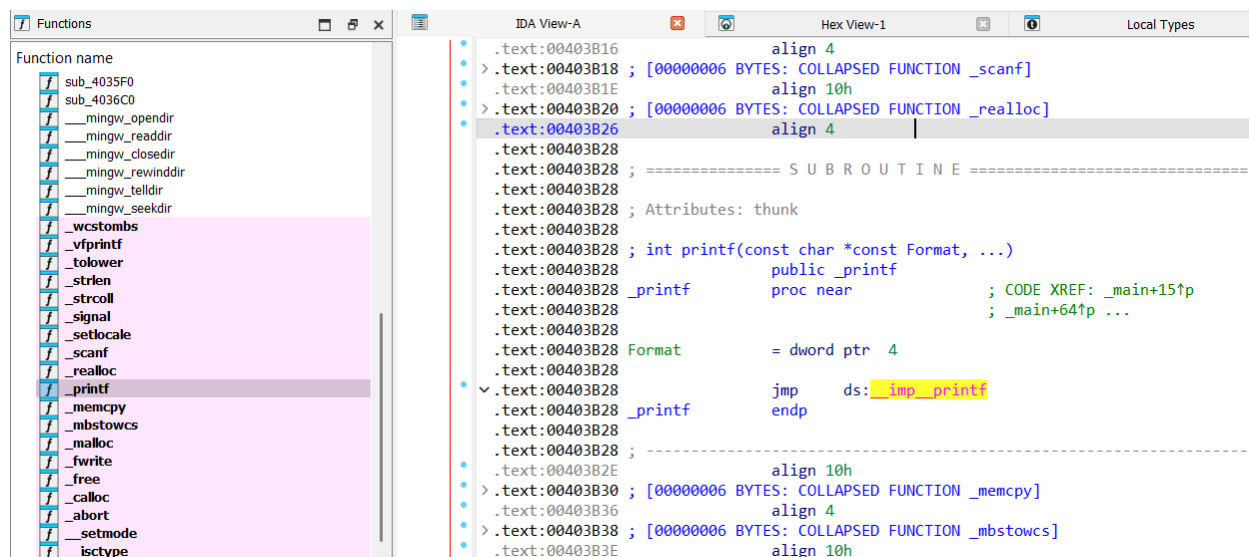
Here we can see several columns in the “Functions” window. Here is what the columns represent:

- Jump option lets one jump directly to an operand, address, memory location, segment, function etc.
- The search operation lets one search for some data in code, memory locations
- The view option lets one look at function calls, cross references, references from or to, user charts



If we click on a function name in the Functions panel, the respective address in the Hex View-1 panel would be shown. In the above example, we can see that upon clicking the `main` function we can see it is stored at `0x00401050`.

IDA View-A (Disassembled Code)



The code in the screenshot shows a tiny helper function called `_printf`. It's **not** the real `printf` function.

Instead, this helper just **jumps to the real `printf`**, which lives in a system file like a DLL (for example, `msvcrt.dll`).

This is how programs usually work when they use common functions like `printf` — they don't include the whole code for `printf` inside the program. They just **link to it** and call it when needed.

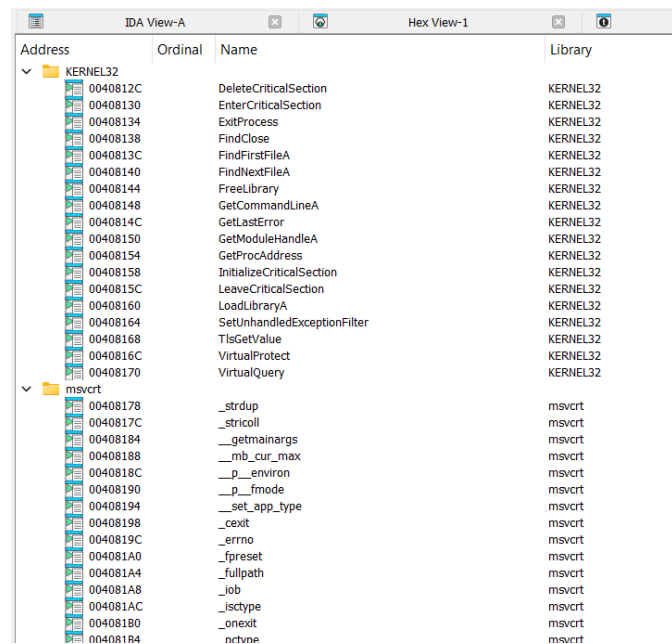
So, this `_printf` function is like a call to the `printf` function in the system library.

Imports

The imported functions can be viewed and analysed in the Imports tab. It shows all functions the program uses from external libraries in this case:

- `printf`
- `scanf`

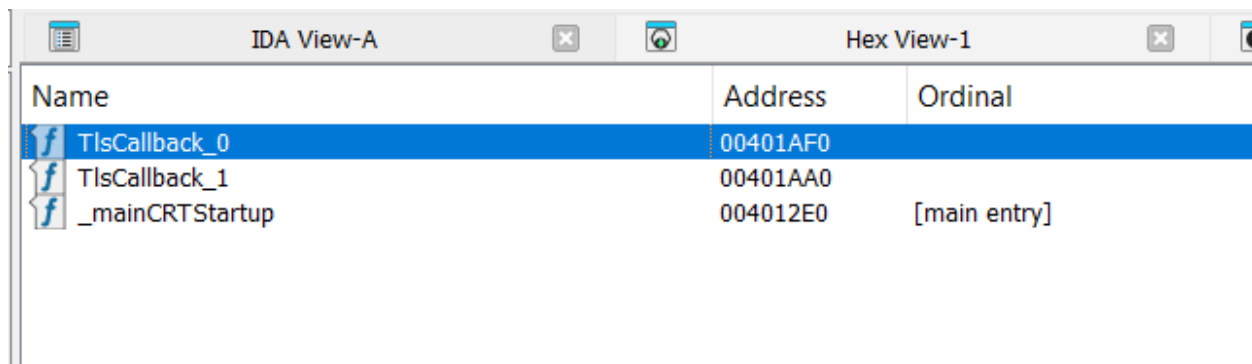
The functions are linked from (like msvcrt.dll or kernel32.dll). The addresses or pointers used to jump to those functions. You can double-click any import to see where it's used in the disassembly.



Address	Ordinal	Name	Library
0040812C		DeleteCriticalSection	KERNEL32
00408130		EnterCriticalSection	KERNEL32
00408134		ExitProcess	KERNEL32
00408138		FindClose	KERNEL32
0040813C		FindFirstFileA	KERNEL32
00408140		FindNextFileA	KERNEL32
00408144		FreeLibrary	KERNEL32
00408148		GetCommandLineA	KERNEL32
0040814C		GetLastError	KERNEL32
00408150		GetModuleHandleA	KERNEL32
00408154		GetProcAddress	KERNEL32
00408158		InitializeCriticalSection	KERNEL32
0040815C		LeaveCriticalSection	KERNEL32
00408160		LoadLibraryA	KERNEL32
00408164		SetUnhandledExceptionFilter	KERNEL32
00408168		TlsGetValue	KERNEL32
0040816C		VirtualProtect	KERNEL32
00408170		VirtualQuery	KERNEL32
00408178		_strdup	msvcrt
0040817C		_strcoll	msvcrt
00408184		__getmainargs	msvcrt
00408188		__mb_cur_max	msvcrt
0040818C		__p__environ	msvcrt
00408190		__p__fmode	msvcrt
00408194		__set_app_type	msvcrt
00408198		_cexit	msvcrt
0040819C		_errno	msvcrt
004081A0		_fpreset	msvcrt
004081A4		_fullpath	msvcrt
004081A8		_job	msvcrt
004081AC		_isctype	msvcrt
004081B0		_onexit	msvcrt
004081B4		nctvne	msvcrt

Exports

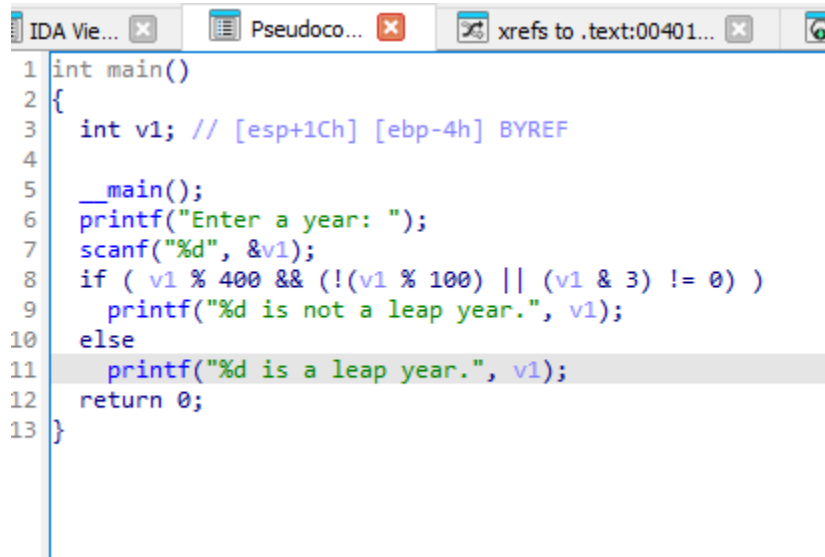
Shows all the functions that we export from our code for use in other programs. We have no export function in our code as shown except main:



Name	Address	Ordinal
TlsCallback_0	00401AF0	
TlsCallback_1	00401AA0	
_mainCRTStartup	004012E0	[main entry]

`_mainCRTStartup` is stored at the Address shown (004012E0) that is the main function of our C program.

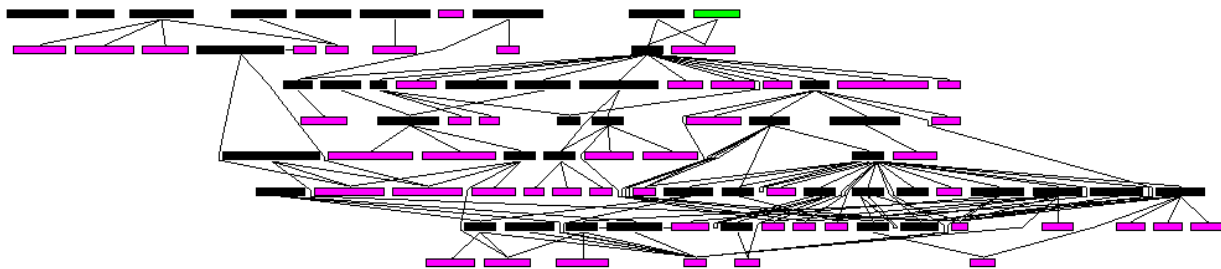
PseudoCode Generation



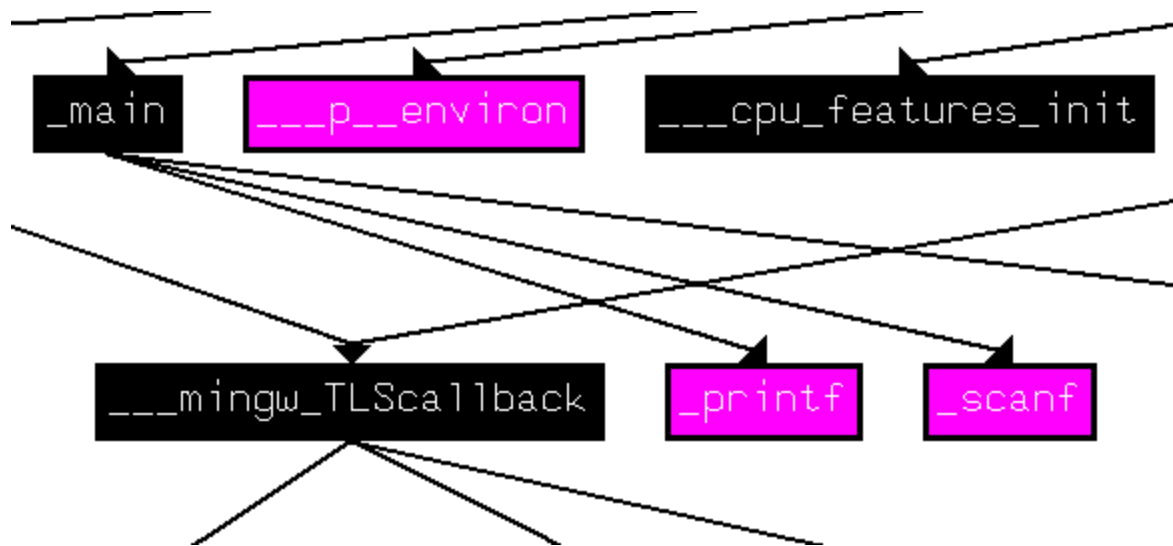
```
1 int main()
2 {
3     int v1; // [esp+1Ch] [ebp-4h] BYREF
4
5     __main();
6     printf("Enter a year: ");
7     scanf("%d", &v1);
8     if ( v1 % 400 && !(v1 % 100) || (v1 & 3) != 0 )
9         printf("%d is not a leap year.", v1);
10    else
11        printf("%d is a leap year.", v1);
12    return 0;
13 }
```

IDA tool also provides an option to view the pseudocode of a binary through its **Hex-Rays Decompiler plugin**. The pseudocode, although not exactly the original source code, is much more readable and helps to quickly figure out what the program is doing — like how functions work, how data is being used, and what the overall logic looks like. The pseudocode helps analysts understand the logic and control flow of the program without having to interpret complex assembly instructions directly.

Graph View of Code Flow



The Graph View in IDA provides a clear visual representation of a function's control flow, displaying how the program executes from start to finish. It shows the function calls and branches in a top-to-bottom format, making it easier to follow the logic and structure. It also allows zooming in and out to explore the flow between different blocks and understand how one part of the function leads to another.



If the view is expanded, it shows `printf` and `scanf` function calls from main. This provides an understanding that the main function is directly calling these two functions, as well as other built-in functions. The view provides a clearer picture of the flow of execution within the program.