

MSRPC Heap Overflow – Part II

Dave Aitel



So a new approach is needed. As with any heap overflow, you get to choose a “where” and a “what” value, subject to certain constraints. If you choose a “what” value that is the address of a writable memory location, that particular allocation will succeed, and the next allocation/free may cause an access violation. This can be useful in many circumstances. For example, if you know that the codepath you are in will call a function pointer, and you know the location of your attack string in memory, you can reliably jump to your attack string.

However, there are also some advantages to having a “what” that points to a non-writable memory address. This will cause the very next instruction to cause an access violation exception, which, in certain cases you can control. I normally prefer the first method, since it doesn't rely on operating system and service pack version. It's possible that the first method is still possible in this exploit. To test, you'd have to see if there are any function pointers called after the heap corruption word-write is triggered. For example, eEye mentioned a function pointer in `rpcss`. However, with all the modifications to `rpcss` that have been happening recently, I don't feel comfortable relying on a magic number that is dependent on that module.

So let's back up a bit, to the point the corruption happens. We can do this by setting `0x01020304` as our “what”. Then I do all this:

1. Run the attack
2. Wait for break on access violation in heap allocator
3. Use debug->Open Run or Clear Run Trace
4. Modify the EAX register to be a writable address (`0x7ffdf080`)
5. Debug->Trace Into.

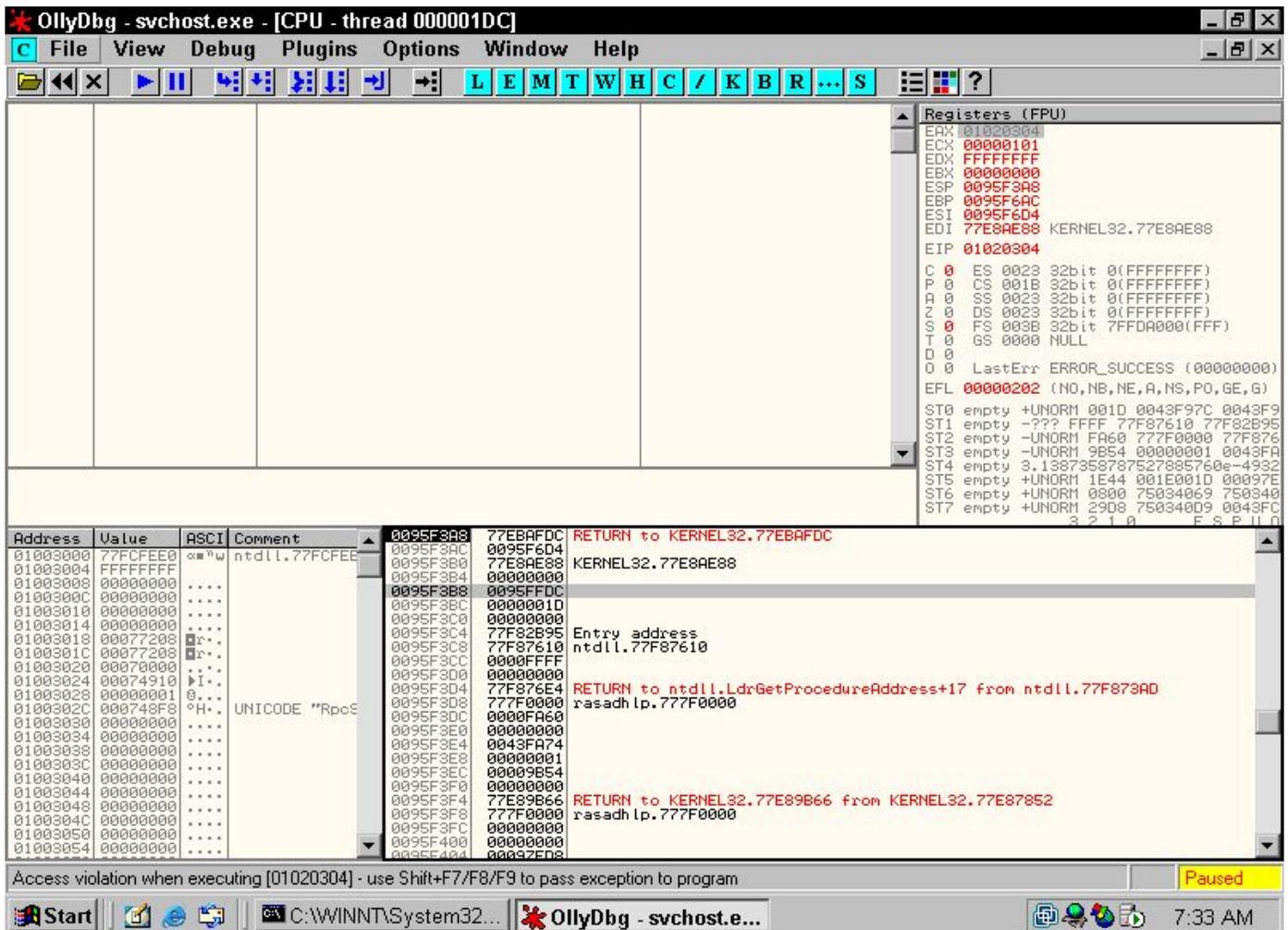
The results from that little test simply show that modifying EAX doesn't change the corruption enough to let the program continue where it needed to go. So I don't discover any magic function pointers. Finding function pointers via Ollydbg is really easy, as long as you have a breakpoint you can set just after the overwrite occurs. Perhaps we'll save that for another time though, since the correct way - trapping on every allocation - is a pain.

Ok, so back to plan B. Which was setting a invalid address (0x0102034) as our “what” value, and seeing what sorts of magic we can do from there, assuming our address will be non-writable.

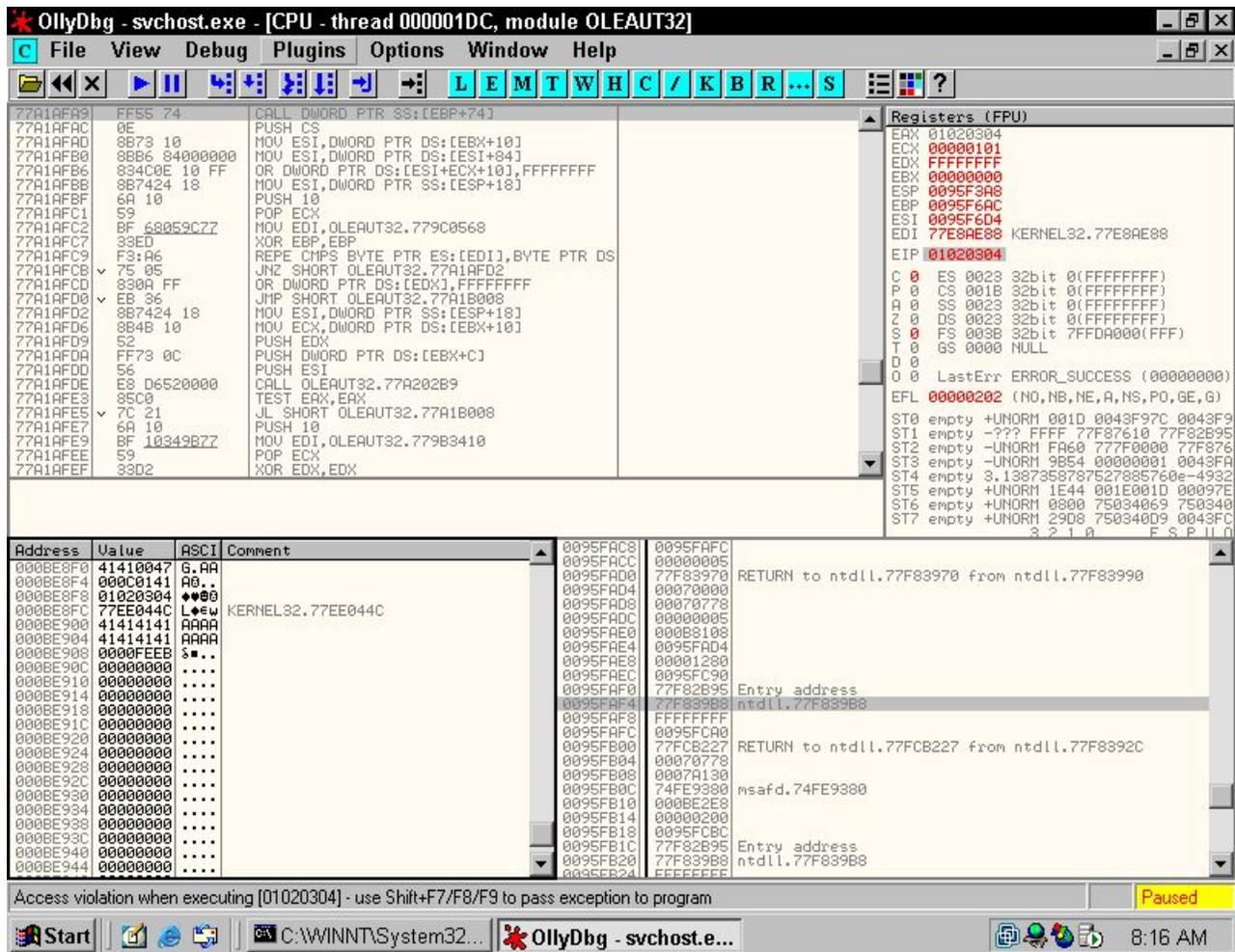
Steps:

1. Modify the kernel32.UnhangedExceptionFilter to null out the jump if it discovers the program is being debugged (Look for a ZwQuery... then a cmp jmp of some kind)
2. Run the attack (with 0x01020304 as “what”)
3. Wait for break on access violation in heap allocator
4. Shift-F9 to continue and send the exception to the program's exception handler routines. Normally this would then result in a “This program could not handle the exception and will terminate” message. As a side note, you can get OUT of that message by hitting f7 again and letting it catch the exception. That way you can try a few different things before it exits.

The program will die executing instruction 0x01020304, but this gives us the unique perspective of what exactly the stack and registers will look like when the exception is handled by our function pointer, assuming we send the function pointer into a .text page somewhere.



At this point I look both up and down on the stack. I right click anything that looks like a heap address and click "Follow in Dump." Of course, if it doesn't point at a valid memory address, Ollydbg won't offer a Follow in Dump option. This can take a long time. Someday I'll have a bevy of Ollydbg plug-ins to take care of this sort of thing, but for now, I just do it manually. It would also help if Ollydbg's \$tring recognition" feature had an option to detect very short strings.



In any case, there are a few pointers to memory we control on the stack. For example, we get two bytes at ebp+0x74, and we get a lot more at ebp+0xb4. For now, I'll just use the ebp+0x74 location. To reach it, we have several options:

- Call [ebp+0x74]
- Jmp [ebp+0x74]
- Find the esp offset and use that as above
- find a code sequence that does something similar to any of the above, like pop; pop; jmp [esp+0xXX]

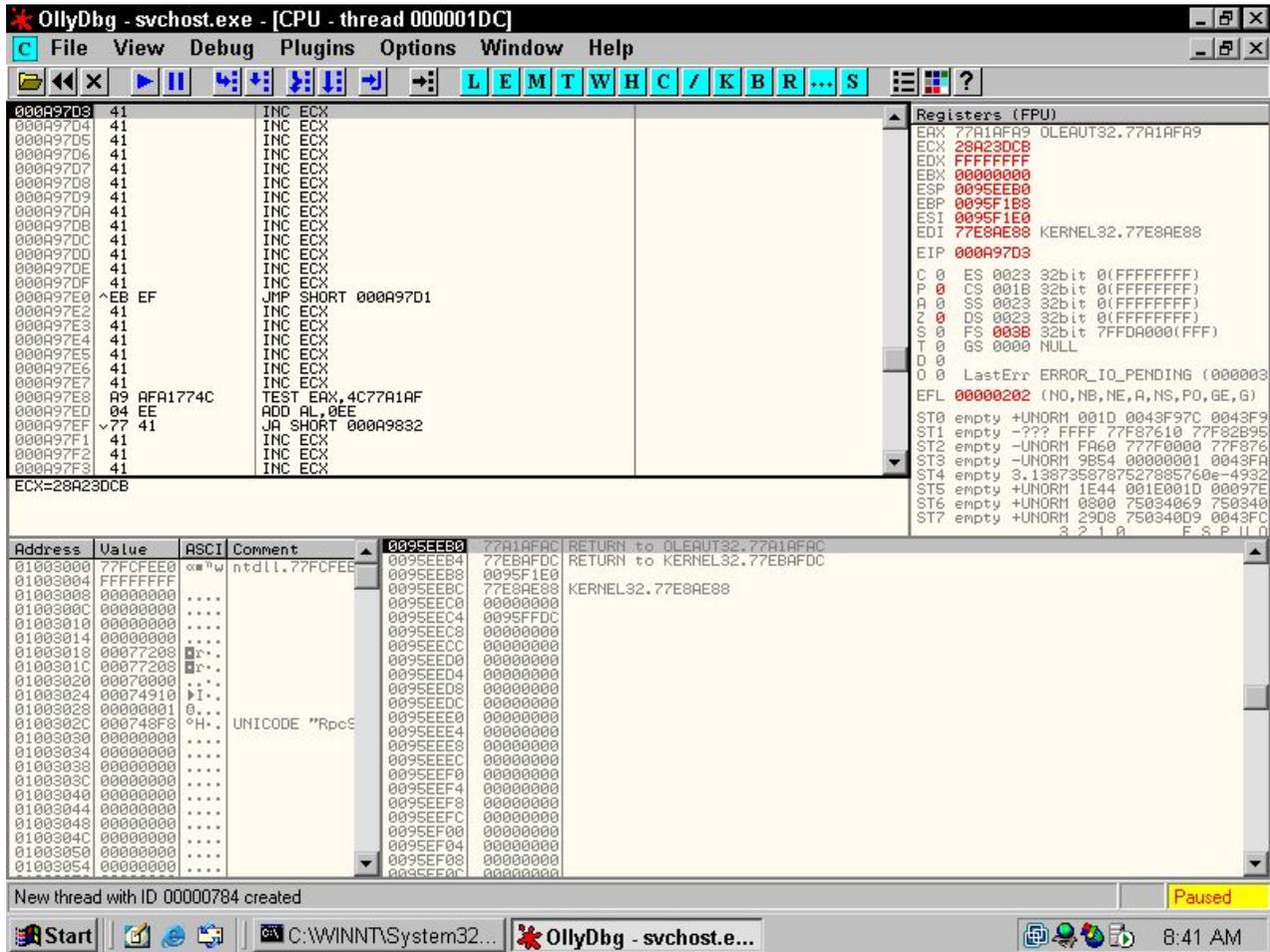
I normally use “View->executable modules and click on any module to view it and then hit space to assemble a command” to find the hex bytes of call [ebp+0x74]. This corresponds to: ff 55 74.

I used view->memory, and then Control-B for search to find this sequence in

OLEAUT32.0x77a1afa9. This is a text page, so it's not writable, which is what we need. Note that for the search, you'll have to click "case sensitive".

So now we'll replace the "what" address with the address in OLEAUT32. And we'll put a "\xeb\xfe" at the end of attackstr (actually, at the location of "what" - 8). Hopefully this will get executed and after the exception the code will jump to our string and run our infinite loop. When the exception occurs, don't forget to look in the title-bar for the thread-ID.

What should happen is that your exception will occur (“Can't write to memory in OLEAUT32”) and then you hit shift-f9, and the exception is handled by . . . you! But you've chosen to just loop. So VMWare will churn and become sluggish and horrible. Hit F12 to break out of that and the right click in the disassembly and go to Thread->(thread ID you memorized) and you'll your code executing. It should look like this.



Well, so now we reliably have our shellcode executing. We move to the next stage, which is “What to do with our shellcode”.

It's here that we want to discover any filters being placed on our shellcode, look at how big our shellcode can be, or deal with any other issues our shellcode is going to have to deal with.

Now, we don't have much room, in Win32 terms, for our shellcode. We have something just over 500 bytes. My Win32 shellcode that will put a lock hold on a process so no other thread gets a chance to crash it is something just over 900 bytes. We could squeeze it down in many ways, since we know exactly what

version of Windows we're running on. But this solution is onerous in terms of maintaining thousands of shellcodes and optimizing them all.

CANVAS does have another solution, which I like to think is more elegant. This is a tiny shellcode that will search all of memory for another shellcode, which it then executes. So my exploit is now this:

1. Shove the large shellcode into memory somewhere. (CANVAS includes a function to do this)
2. Run the heap overflow request with tiny searchcode
3. Run the dcedump request to trigger the magic

We have to do some trickery to get back far enough into our buffer to have the 150 bytes or so we need for our “tiny” shellcode. A small stub of “call back, pop ebx, sub 400 from ebx, jmp ebx” gets us back into where we need to be. And we have to make sure that neither of our shellcodes is getting corrupted as it is stored in process memory. This may take some wiggling of various offsets, but is generally quite easy.

This method works to get our tiny search shellcode into memory and executed and then our much larger stable shellcode executed. From there, the rest is gravy.

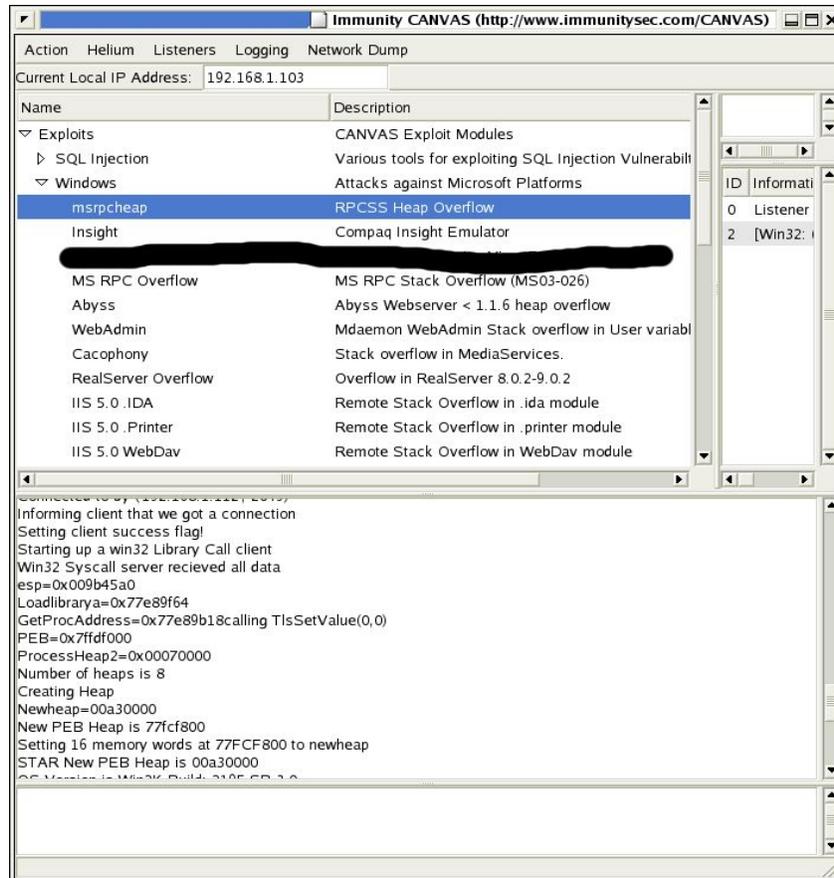


Illustration 1 CANVAS successfully attacks a target via the MSRPC Heap Overflow

Now, of course, it remains to be seen how reliable this technique is. It works very well against my SP3 Unpatched box. But it's possible that after some testing on different machines, it may have to change, as the pattern of allocations in RPCSS becomes better understood. In fact, there is actually a tree of decisions that I made in order to accomplish this exploit, and I will have to walk down the entire tree to find out if there are any more reliable solutions to the complex equation that is this exploit.

In fact, like many heap overflows, this one has corrupted vital parts of the heap that I need to repair before I can do complex things like run a command shell. But the hard part is over. My shellcode stub is communicating with me and I have full control of the program.