

Introduction to Malware analysis

Contents

5.1	Objectives	51
5.2	Laboratory environment	51
5.3	The malware: a trojan copy of a windows live messenger	52
5.4	Behavioural analysis	52
5.5	Network traffic analysis	55
5.5.1	CaptureBAT	56
5.5.2	Wireshark	57
5.5.3	Analysis	57
5.6	Code analysis	58

5.1 Objectives

In this session, we will introduce some common approach to analyse a malware, so that you can turn malicious executable inside out to understand their inner-workings (technique known as reverse engineering). Knowing how to analyse a malware can bring an element of control into an otherwise chaotic environment that exists around a security incident. It is also a critical aspect of modern forensic analysis as it is frequent for investigators to discover malware on compromised systems.

The approach followed in this session is reverse-engineering that has worked for many analysts; it involves two key phases: behavioural analysis and code analysis. During behavioural analysis, we examine how the malware interacts with its environment. The code analysis phase allows us to learn about the specific capabilities by examining the code of the compromised program.

5.2 Laboratory environment

When performing malware analysis, it is convenient to stay in a virtualisation environment. Such tools typically simulate the underlying hardware, allowing you to run multiple instances of virtual machines simultaneously. For instance, you could use Windows 7 as your base OS, while having a separate instance of Windows XP running in another window, and a Linux instance running in another one.

Each virtual machine behaves mostly as *real* physical systems, in that it has its own set of I/O peripherals, RAM, network settings, and so on. All these aspects of the virtual machine are indeed virtualised.

The convenience of a virtualised lab comes, in part, from the flexibility of having multiple instances of various operating systems available to you within a single physical system. Virtualisation software can even emulate a network, so that your lab does not need to be connected to a physical network at all. Yet, the virtual machines will be able to communicate with each other over the simulated network, blissfully unaware that the network is not “real”.

For this session we will use the VMware Player software. To do this, open the virtual machine called “malware” from the link provided in the presentation and start it. If the VMware Player asks, answer that “you copied the virtual machine”. The password of the administrator user is **si2012**.

5.3 The malware: a trojan copy of a windows live messenger

The malicious executable which we will learn from in this session is shown in Figure 5.1. It is a trojan copy of Windows Live Messenger - a fake instant messenger client that was being distributed to victims via email. Many such trojans have the capability of capturing the victims’ logon credentials, and may have other *undocumented* features.

In this lab, we will assume that we have got the trojan executable file and the *msnsettings.dat* file from the victim’s PC.

Let’s see what capabilities are built into this malicious executable. First, let’s introduce the tools and techniques that will help with the reverse-engineering process.

Note that in this example, as with the majority of malicious incidents you will probably encounter, we will be examining a compiled Windows executable for which we have no source code.

5.4 Behavioural analysis

Malware analysis typically starts examining a malicious executable with behavioural analysis as it is easier than code analysis and provides some hints for it.

When performing behavioural analysis, we are going to infect a laboratory system with the malware. Then we will observe how the malicious executable accesses the file system, the registry, and the network. As we learn about the program’s expectations of its runtime environment, we will slightly adjust our analysis to evoke additional behaviour from the program. We will also attempt to interact with the program to discover additional characteristics it may exhibit.

Let’s see this approach in action. Imagine you have a suspicious executable that you would like to analyse. You bring it into your lab, possible via a removable USB

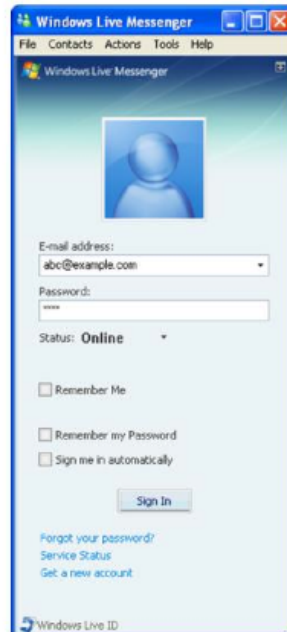


Figure 5.1: Windows live messenger snapshot.

disk and place it on the desktop of the virtual machine¹ you are about to infect. Now what?

First, take a snapshot of the state of the machine’s file system and the registry. This will allow you to quickly see what major changes have occurred on the system after you infect it.

To do this we will use the free tool called **RegShot** (<http://sourceforge.net/projects/regshot>). RegShot is already installed in your VM. To use it, enable the “Scan dir1” option, and in the corresponding window type “C:\”. This will allow the tool to scan the registry and the full C: drive. Click now “1st shot”.

After RegShot takes the first snapshot, launch the malicious executable. Interact with it a bit (e.g., try logging into it). Then kill the process, if you can. Next, click the “2nd shot” button in RegShot, and click the “Compare” button. You will see a report that describes the major changes to the system’s state. In this case, we see that, among other things, two files were added to the system.

The two files that appeared on the system after we infected it are *pas.txt* and *msnsettings.dat*. Take a look at them using notepad.

It looks like *pas.txt* has captured the login credentials we used when logging into the malicious executable. That makes sense, because we received reports that this executable is a trojan copy of Windows Live Messenger.

The *msnsettings.dat* file looks like a configuration file of some sort.

¹The fake windows live messenger is already copied in your VM image

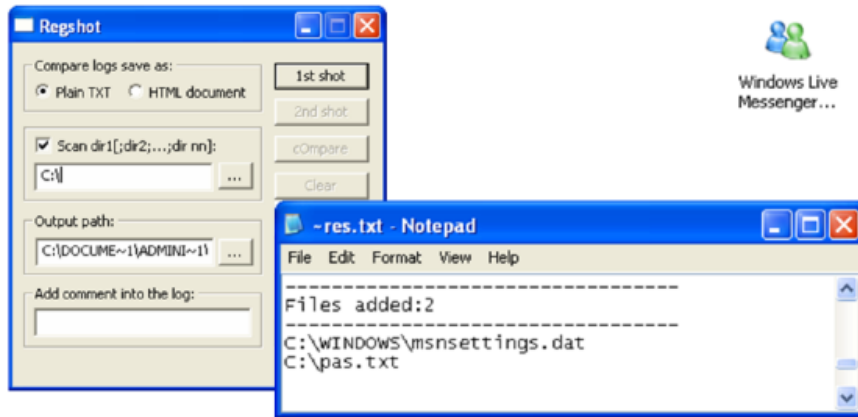


Figure 5.2: The RegShot tool and its output.

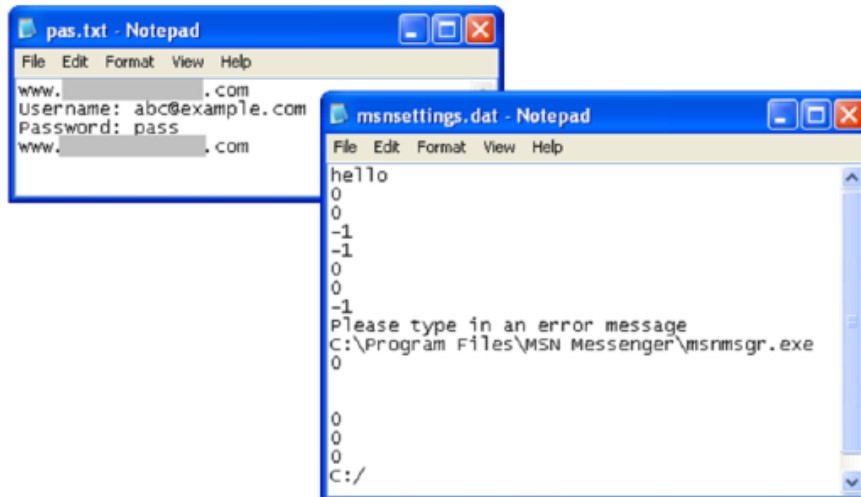
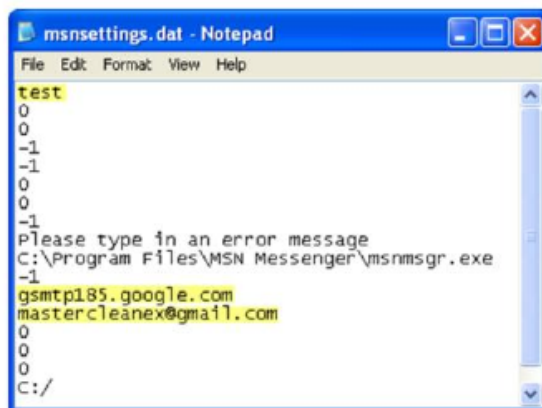


Figure 5.3: The two files created by the Windows live messenger malware.



```
msnsettings.dat - Notepad
File Edit Format View Help
test
0
0
-1
-1
0
0
-1
Please type in an error message
C:\Program Files\MSN Messenger\msnmsgr.exe
-1
gsmtpl85.google.com
mastercleanex@gmail.com
0
0
0
0
C:/
```

Several settings on the victim's system differ from the lab's.

Figure 5.4: Windows live messenger snapshot.

With the obtained information, reverse-engineering the malware can help you at incident response and forensic analysis. In our scenario, we have already discovered that Windows Live Messenger trojan makes use of the *msnsettings.dat* file. Now you know to look for it on the compromised system, even if you did not initially realise that this file was important.

Once you have a copy of *msnsettings.dat*, you can open it to see whether it reveals additional details about the program. It should appear like the one reported in Fig. 5.3. It seems that the malware uses *msnsettings.dat* as a configuration file.

Now, let's take a look at the *msnsettings.dat* inside the *live-messenger-malware* folder on the desktop. This file proceeds from the victim of the fake messenger malware. It is shown in Fig. 5.4, in which we have highlighted some lines.

Let's close the fake messenger and **replace the *msnsettings.dat* in the Windows folder with the victim's one.**

In the first line there is a string "test", which we may be able to use later when trying to understand how the trojan processes the *msnsettings.dat* file. Another line, "gsmtpl85.google.com" specifies an SMTP mail server; this suggests that the malware has the ability to send email. The file also includes an email address, "mastercleanex@gmail.com". This may be the recipient of the information that the trojan might attempt to send out. Of course, these are just theories at this point. We will need to confirm or deny them during subsequent analysis steps.

5.5 Network traffic analysis

Our next step is to confirm if the fake Microsoft MSN sends some SMTP messages. To do this, we have two options: **CaptureBAT** and **Wireshark**. You can launch the tool you prefer, both options will provide the same conclusions, and afterwards execute the malware again with the victim's *msnsettings.dat*.

Warning: before capturing network traffic, flush the DNS cache (in windows

```
C:\>capturebat -c -n

Option: Capturing network packets
Option: Collecting modified files
Loaded kernel driver: CaptureProcessMonitor
Loaded kernel driver: CaptureRegistryMonitor
Loaded filter driver: CaptureFileMonitor

-----
process: created C:\WINDOWS\explorer.exe -> ...Windows Live Messenger.exe
file: Write ...Windows Live Messenger.exe -> C:\WINDOWS\msnsettings.dat
file: Write ...Windows Live Messenger.exe -> C:\pas.txt
```

Figure 5.5: CaptureBat output snapshot.

command line mode, execute “ipconfig /flushdns”; you can also flush all the network related caches by going in the Control Panel -> Network settings -> Network connections-> double click on the network connection, select the Support tab and then click on Repair’).

5.5.1 CaptureBAT

CaptureBat is a very useful and free tool available at <http://www.honeynet.org/projects/ols/capture-bat>.

CaptureBAT² records local processes’ interactions with their environment. CaptureBAT’s logs are quite easy to follow and understand because it comes with filters that eliminate the majority of standard, non-malicious activities from the logs. You can customise these filters to your liking, as they are text files located in the directory where you install CaptureBAT.

If you launch CaptureBAT³ with the “-c” parameter, it will capture any files deleted in the background, allowing you to look at and restore even those files that the Windows Recycle Bin cannot capture.

Launching CaptureBAT with the “-n” parameter tells the tool to capture network traffic, like a sniffer would, saving the result into a local .cap file.

Warning! Do not terminate CaptureBAT with “ctrl+c”, CaptureBAT dumps the data on the disk at the end of its execution. Use “enter” to terminate the CaptureBAT process.

Open the .cap file in WireShark. As you can see on the next figure, CaptureBAT confirmed our earlier findings about the malware⁴.

If you do not like using CaptureBAT, you could use Wireshark to capture the traffic sent by the malware.

²Install CaptureBat and reboot the system

³In command line mode, go to C:\Program Files\Capture

⁴If it does not show the DNS query we’re looking for, make sure that you have substituted the *msnsettings.dat* in the Windows folder with the one that came in the Desktop folder

No.	Time	Source	Destination	Protocol	Info
1	0.000000	00:0c:29:ca:2a:f2	ff:ff:ff:ff:ff:ff	ARP	who has 192.168.11.129? Tell 192.168.11.129
2	0.038277	00:0c:29:15:71:e1	00:0c:29:ca:2a:f2	ARP	192.168.11.129 is at 00:0c:29:15:71:e1
3	0.039663	192.168.11.128	192.168.11.129	DNS	Standard query A gsmtpl85.google.com
4	0.067967	192.168.11.129	192.168.11.128	ICMP	Destination unreachable (Port unreachable)

Frame 3 (79 bytes on wire, 79 bytes captured)
Ethernet II, Src: 00:0c:29:ca:2a:f2 (00:0c:29:ca:2a:f2), Dst: 00:0c:29:15:71:e1 (00:0c:29:15:71:e1)
Internet Protocol, Src: 192.168.11.128 (192.168.11.128), Dst: 192.168.11.129 (192.168.11.129)
User Datagram Protocol, Src Port: blackjack (1025), Dst Port: domain (53)
Domain Name System (query) <ul style="list-style-type: none"> Transaction ID: 0x1bad Flags: 0x0100 (standard query) Questions: 1 Answer RRs: 0 Authority RRs: 0 Additional RRs: 0 Queries <ul style="list-style-type: none"> gsmtpl85.google.com: type A, class IN

The hostname suggests SMTP, but use DNS resolution to confirm.

Figure 5.6: WireShark output snapshot.

5.5.2 Wireshark

Wireshark is a full-feature network sniffer (<http://www.wireshark.org>). Instead of using CaptureBAT, you can directly open Wireshark, and run the sniffer over the network interface using the appropriate filters.

5.5.3 Analysis

As you can see on the dump reported in the Fig. 5.6, the sniffer shows that the infected system has issued a DNS query, attempting to resolve the hostname “gsmtpl85.google.com”. The SMTP in the hostname suggests that the malware is looking for a mail server to connect to, reinforcing our earlier theory of how the trojan might use this hostname.

We will therefore try to intercept any e-mail that the fake messenger will possibly send to that SMTP server. To this end, we will

1. Redirect DNS queries to localhost: you can do this in the properties of the Windows Network Connection in the Control Panel; Go to TCP/IP properties and set 127.0.0.1 as the primary DNS server;
2. Capture any DNS request with **Fake DNS**. Fake DNS is a DNS server that you can configure to answer any DNS query with a single IP address of your choice⁵. Which IP address should you use? The easiest option is to use again “localhost”. This will redirect any DNS query to localhost; therefore, a possible connection to “gsmtpl85.google.com” on port 25 will be redirected to “localhost:25”. Put Fake DNS to “Listen” mode.

⁵An alternative way, is to set up name resolution by inserting an entry for the hostname into the ‘hosts’ file on the infected system; Fake DNS is a faster alternative

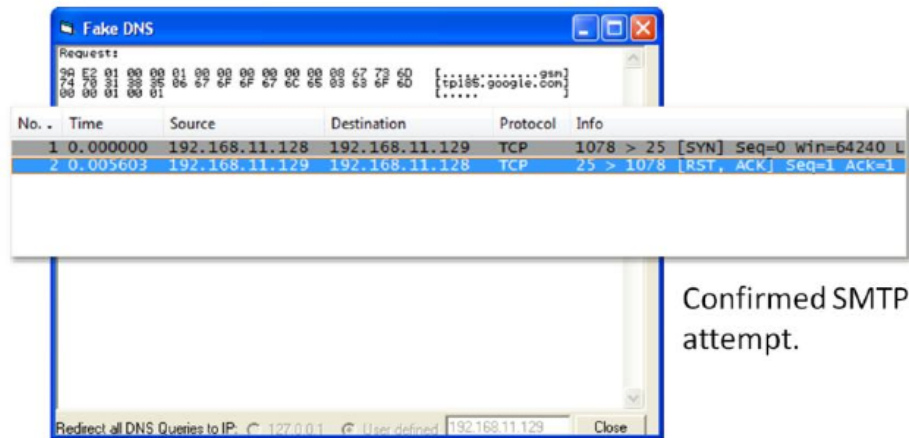


Figure 5.7: Fake DNS output snapshot.

- Set up an SMTP listener on port 25, capturing any mail will be sent. An easy way to do this is to use the **Mailpot** tool. Mailpot pretends to be a mail server, happily accepting SMTP messages from clients, but not sending them out. Instead, it stores the messages locally for your review. To use Mailpot, run it on the host to which you have redirected the SMTP requests (localhost) using Fake DNS and put it to 'Listen' mode.

Now that you have set up the DNS server as localhost (in Windows network connection settings), are replying the DNS request with localhost and are listening the SMTP connection on port 25, you are ready to try to login with the fake MSN and see what happens.

Now you should see both the DNS request and response in the Fake DNS window, and in the Mailpot window there should be the email that has been sent by the fake MSN and captured by Mailpot.

If you open the mail⁶, you can see the contents of the message that the trojan is mailing to the attacker. As highlighted in the Fig. 5.8, the message includes the victim's Messenger username and password.

5.6 Code analysis

Behavioural analysis can be insightful and relatively fast. However, it will rarely tell you everything you need to know about malware of moderate and advanced complexity. That is where code analysis can be of help. It can help to reinforce

⁶If the mail does not open (by double clicking on it), you have to (1) close Mailpot, change the permissions of the C:\mailpot folder to writeable (remove the thick from "Read-only") and (2) start Mailpot again; you may have to repeat this procedure a couple of times; if it doesn't work yet, you can reboot the system and try again (remember to reactivate the Fake DNS)

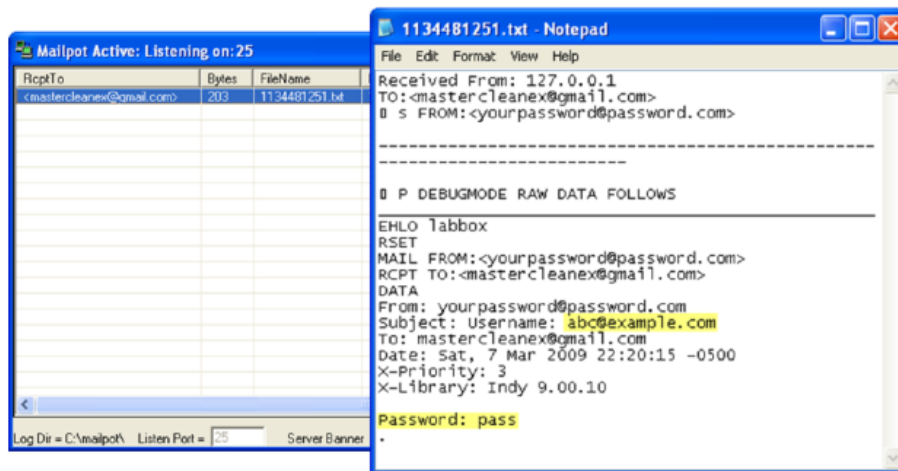


Figure 5.8: Mailpot output snapshot.

your behavioural findings, and can shine a light on additional properties of the malware that you may not have discovered behaviourally.

Code analysis can be tricky and time-consuming, because in the world of malware you almost never have the luxury of seeing the source code of the program you are analysing. Instead, you need to reverse-engineer the compiled executable's functionality by examining its code at the assembly level. A debugger and a disassembler can help you in this task. A disassembler converts the malware's instructions from their binary form into the human-readable assembly form. A debugger lets you step into the most interesting parts of the code, interacting with it and observing the effects of its instructions to understand their purpose.

We will use **OllyDbg** to perform code analysis. It is free, and includes both a disassembler and a debugger. You can download OllyDbg from: <http://www.ollydbg.de/>

A good way to start analysing the malware's code often involves looking at the strings embedded in its executable. To do this with OllyDbg, first load the malicious executable into OllyDbg via File -> Open. Then, right-click on the code you will see in the disassembler window, and select Search for -> All referenced text strings.

OllyDbg will then bring up a new window that will show the strings it discovered, as you can see on this slide. Notice that we have seen some of these strings during behavioural analysis! Some of them look like contents of the default *msnsettings.dat* file that our malware creates when infecting the system.

The reason we may be interested in looking at the embedded strings is because the string listing might include a reference to a malicious characteristic or a behavioural trait that we would like to understand. In this case, consider the following screenshot. We got here by highlighting one of the instances of *msnsettings.dat* strings, if you press Enter, OllyDbg shows us how the program makes use of this

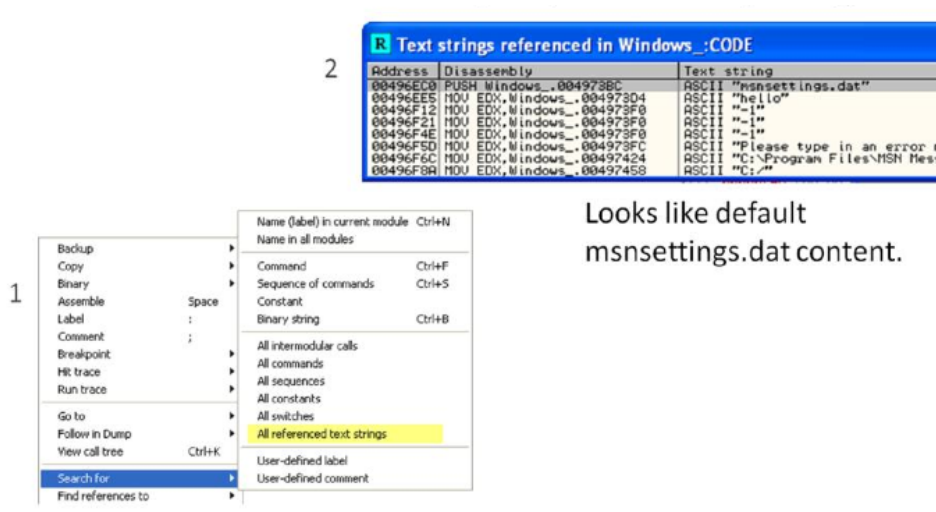


Figure 5.9: A snapshot from OllyDbg.

string.

If we wanted to pursue this path of analysis further, we could now set a breakpoint on this command, run the trojan in the debugger, and see what it does. But, we are not going to investigate this particular aspect of the malicious program, because we will focus on another, more interesting technique.

You may recall that the version of *msnsetting.dat* on the victim's system was slightly different from the version that the trojan created on our laboratory system when we first ran it. Specifically, in our case, the file contained the string "hello", while the victim's version had the string 'test' instead. What's that about?

The string "test" is not visible anywhere within the body of the malicious executable when it's not running. That's probably because the trojan loads this string from *msnsettings.dat* during run time. To understand how the trojan uses the string "test", we will search for it in the memory of the running trojan.

Once we will have located the string in the trojan's memory (described further), we will set an access breakpoint there. A breakpoint is a condition that tells the debugger when to pause the normal execution of the debugged program. Once the execution is paused, the debugger will give us a chance to review the debugged program's run time environment to understand what it is doing. This is probably the most useful feature of a debugger in the context of malware reverse-engineering.

To make use of this technique, load the malicious program into OllyDbg, then run it into the debugger. Once the trojan is running, press "Alt+M" to bring up the memory map in OllyDbg. This shows the listing of the memory segments mapped and used by the currently-debugged executable. To search the executable's memory for a particular string, select the first line in the Memory Map window, and press "Ctrl+B"; then, enter your string as ASCII text field in the dialog box. Then press Enter.

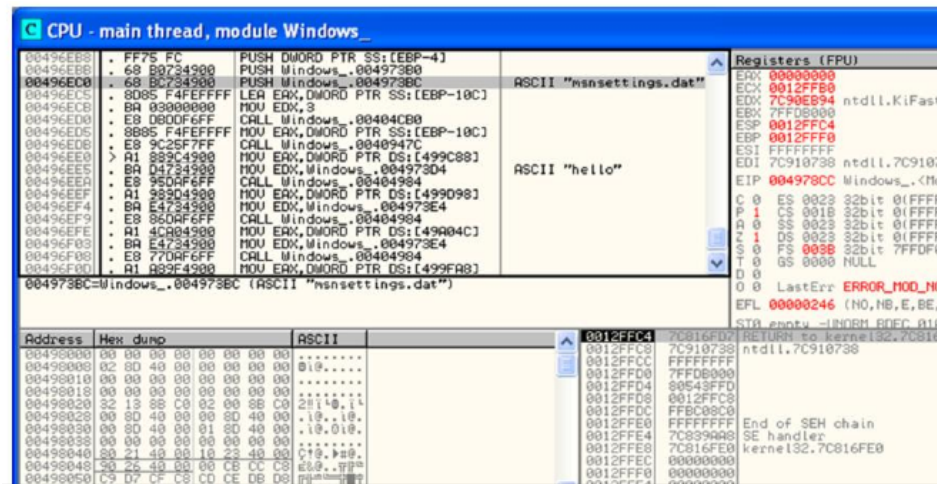


Figure 5.10: A snapshot from OllyDbg.

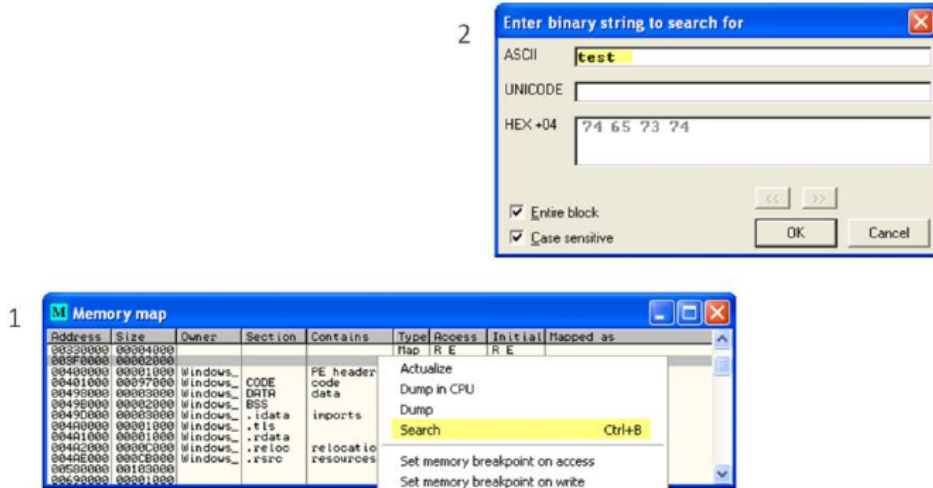


Figure 5.11: A snapshot from OllyDbg.

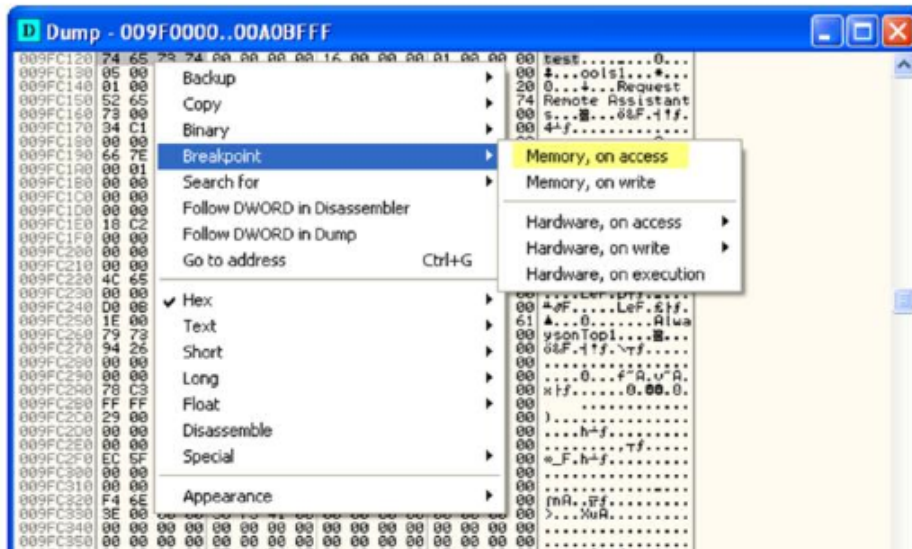


Figure 5.12: A snapshot from OllyDbg.

It is possible that your string will be located in several memory areas. The one you are interested in won't necessarily be the first one. To repeat your search, click on the memory map window, then press "Ctrl+L" (do not forget to click on the memory map window!).

In the case of our example, we will need to perform the initial search via "Ctrl+B". This will find us an instance of 'test' that is not promising. We will repeat the search by pressing "Ctrl+L" once.

Now that we have located the exact string "test" in the trojan's memory, we can set a breakpoint there. In this case we will be setting a memory access breakpoint, so that OllyDbg pauses the program's execution whenever it attempts to access this particular memory area. Effectively, this will allow us to catch the trojan while it is attempting to use the "test" string; we will then be able to see how it makes use of the string.

To set the breakpoint, highlight the exact characters of the string "test", then right-click and click "Breakpoint" -> "Memory, on access"⁷.

The trojan will continue to run. Now we can either wait for it to try using the string, or attempt interacting with the program with the hope it will use the string.

We can try interacting with the trojan by typing some text into its first field, the one labeled "E- mail address". If you type any character there after setting our memory breakpoint, you will immediately trigger the breakpoint, as you can see on the next figure.

We entered a character into the field (picked a letter at random: "g"). Right

⁷Note that this kind of breakpoint will not appear in the list of breakpoints of OllyDbg; don't worry, it is OK

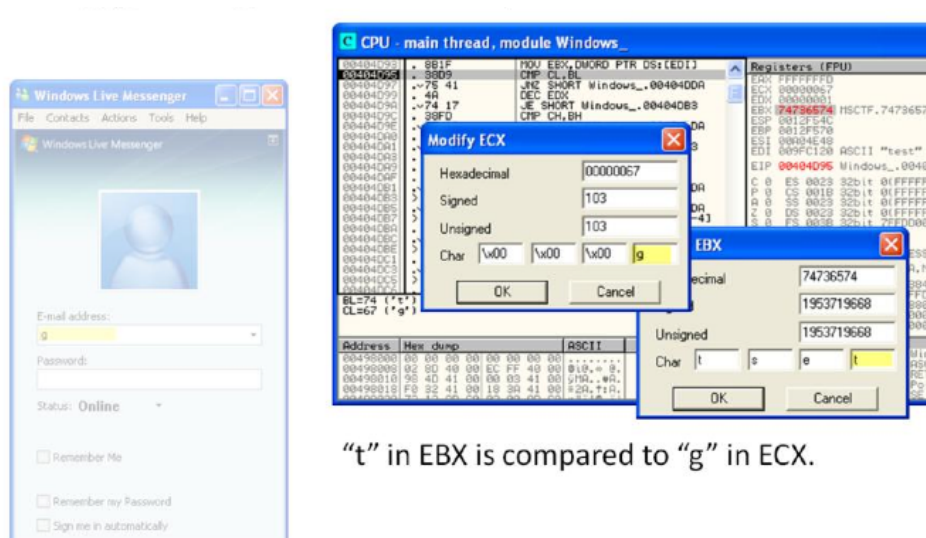


Figure 5.13: A snapshot from OllyDbg.

away, OllyDbg comes to the foreground, because we just triggered an attempt by the trojan to somehow use the string “test”. You can now interact with the code, looking at its environment, and even running it as slowly as one instruction at a time.

To execute one instruction, press F8. To examine the run-time environment of the program, look at its registers in the top right corner of the OllyDbg window. A register is a specialised and very fast location on the CPU that can store data.

What is going on in this part of the code? Do not worry if you do not understand much of the assembly code you see there: this is just an introduction to malware analysis, so we will walk you through the most important parts. OllyDbg has highlighted the instruction that will be executed next by the program, “CMP CL, BL”. This compares contents of two registers, CL and BL. CL points to the lowest byte of ECX; BL points to the lowest byte of EBX, so it is an efficient way of comparing parts of ECX and EBX registers.

Double-click the registers to see their contents. ECX contains the character we entered, “g”. EBX contains the string that our input is being compared to, “test” (it is stored backwards).

Press F9 to continue executing the trojan. Delete the “g” character you have entered previously. This time, let the program match the first character of the “test” string, and see how it compares the second character. To do this, enter “ta” in the “E-mail address” box. If you keep triggering the breakpoint, press F9 to continue. You want to pause right after you have had a chance to type “ta”.

Press F8 to execute one instruction after you have triggered the breakpoint, just like you did previously. Keep on pressing F8 carefully, once at a time, until you reach “CMP CH,BH” some lines below. This time, if you look at contents of ECX

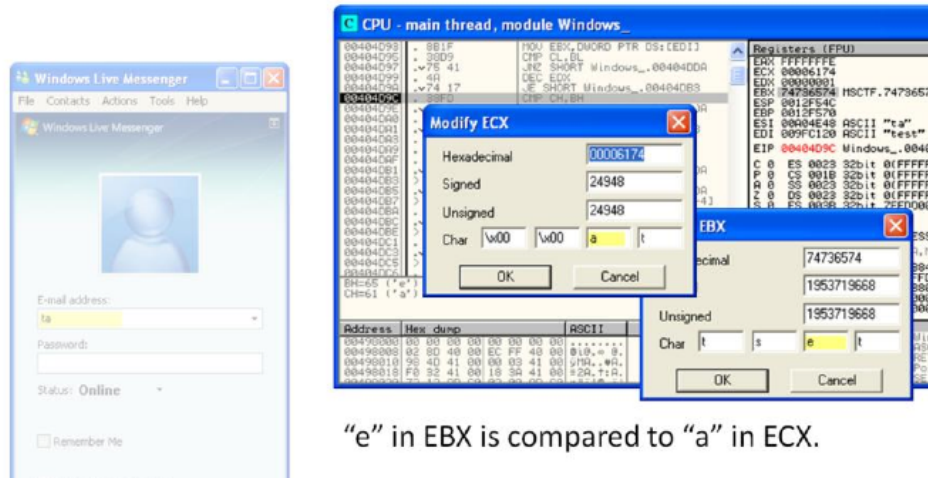


Figure 5.14: A snapshot from OllyDbg.

and EBX registers, you will notice that the trojan is comparing the character 'a' that we entered to the character 'e' that it seems to expect. That is because the CH register points to the second lowest byte of ECX; the BH register points to the second lowest byte of EBX.

So, the trojan seems to be looking for the string "test" in the "E-mail address" field. Exit the debugger, launch the trojan by itself, and enter "test" to see what happens.

Voila! When you enter "test", the trojan brings you to a brand new screen that seems to allow you to configure the trojan's operation. As you can see on this slide, the configuration options let you define the passphrase to activate this string, the address where the trojan will send captured login credentials, etc.

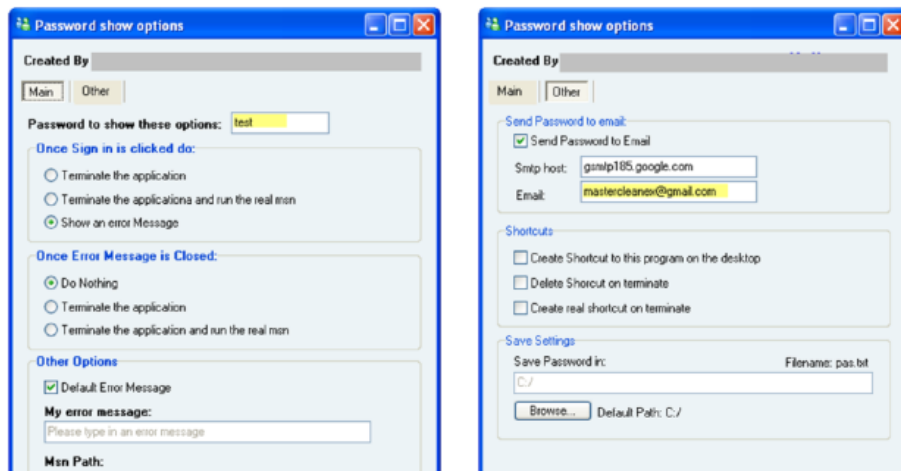


Figure 5.15: The configuration panels of the malware.

