Analysing malware samples

Pablo Collado Soto

**Setting up**

Before setting off on a quest to find malware lurking around, we need to define the scenario we’ll be working on. As stated in the coursework specification, our environment will be a virtual machine running Windows XP.

In order to make it feel like at home we have taken a series of steps so that all the necessary components and tools were available. We’ll briefly describe the entire process so that what we include in this report is as reproducible as possible.

Please note that in the ensuing discussion we’ll refer to the machine running Windows XP as the guest machine and the local machine as the host machine.

**Importing the VM**

The guest machine we are to work with is tremendously specific and so is very hard to exactly replicate. That’s why the way it’s been distributed is through a virtual machine image using the *\*.ova* format. As we are working with Oracle’s VirtualBox virtualization software we just need to open it up and import it as shown in the following screenshot:

**TODO: Get screenshot!**

**Installing the Guest Additions**

We shan’t forget Windows XP uses a graphical desktop environment. We’ll need to make the guest machine’s desktop as big as ours so that we can comfortably move around. To this end, we can install the so called *VirtualBox Guest Additions*which just bundle up a set of drivers designed to make using VMs a seamless experience. Installing it is as easy as going into the *Device* menu within the running VM and clicking on *Insert Guest Additions CD Image…* We can then navigate to Windows XP’s *My Computer* from the *Start* menu and double click on the disk unit to bootstrap the installation. After rebooting the VM we should be ready to go.

It’s likely that before the guest additions can be inserted into the machine we need to create an optical disk bay. We can do so from the settings of the VM before booting it up by just adding a new entry to the *IDE Controller*. We are attaching a screenshot showing how to accomplish that.

**TODO: Get screenshot!**

**Configuring the Network**

As our intention is to “detonate” malware within the machine we should mind the network configuration so that we don’t let any of the samples roam freely on the network. This can be easily accomplished by just clicking on the *Devices > Network > Connect Network Adapter* option within the running VM which will just toggle the NIC (Network Interface Card) as we please. This let’s us have an entirely functional network setup that can be enabled at will. As our host machine is a MacBook Pro we’ll use a *bridged adapter* configuration which will expose the guest VM as an independent machine to the LAN our host machine is connected to. This implies it’ll get it’s very own IP address within the LAN our host is attached to. We can also point out that the network topology we are logically implementing is a layer 2 bridge connecting the host and guest machines so that both still belong to the same network segment (i.e. LAN). If we were to have a Windows-based host we would go with a *NAT* configuration option as that’s less error prone and given the Windows XP machine is not to behave as a server we won’t perceive any functional difference.

Either way, as we intend on leveraging *WireShark’s* capabilities for capturing network traffic we need to install the suitable NIC driver. This will also force us to select the correct NIC when configuring the VM, thing we can do under the *Advanced* section within the *Network* section of the configuration. Given the driver we were provided we’ll choose the *Intel Pro/1000 MT Desktop (82540EM)* NIC so that everything remains compatible. The driver installer can be found in our course repository as usual. Please refer to the annex on the repository structure. Installation is just a matter of double-clicking the executable and accepting the different steps.

**Showing the hidden files**

As more recent Windows releases, we see that Windows XP already made use of the *NTFS* file system for its drives. It, as many others, has support for hidden files. These files are nonetheless not as hidden as we might expect. We can inspect them in a fairly easy way both through the graphical file explorer and from Windows’ CLI (*cmd*). In order to show hidden files on the file explorer we just need to go to the *Control* *Panel* (which we can do from the *Start* menu)and click on the *Folder Options* option. We are attaching a screenshot showing the option we are to tick.

**TODO: Add screenshot.**

We can also use good ol’ *cmd* to navigate the file hierarchy and inspect hidden files. As we know we’ll find a gold mine on *C:\WINDOWS\msagent\intl\MS\_PMAL\_Agent* we might as well show how we can list that hidden directory. We just need to navigate to it’s parent directory with cd *C:\WINDOWS\msagent\intl* to then run *dir /AH.* As seen in *dir’s* help which we can read through with *help dir*, we can control its output through a series of options. Hidden files and folders won’t be shown by default but we can force it to display them with the */AH* which only displays files with the *Hidden* attribute. Then, we can indeed see the *MS\_PMAL\_*Agent directory if we run *dir /*AH after the previous cd statement.

Either way, we need to be aware of the existence of hidden files and take measures so as to have them present and take them always into account.

**Stopping automatic updates**

The malware samples we have been provided are guaranteed to behave as expected if and only if we maintain the correct OS version. That’s why we need to prevent automatic updates from happening. We can easily do so by going to the *Control Panel* and clicking on *Automatic Updates*. We can them disable them form the next window that’ll pop up and that we are including in below.

**TODO: Add screenshot!**

This will guarantee nothing is meddled with as we carry out our analysis so that we can worry only about crucial and important aspects whilst being sure no “funny” behaviours are taking place behind the scenes.

**Regarding the firewall**

As before proceeding to the analysis we can’t be sure of there being any malicious use of the network we decided that shutting the firewall down before having any leads as to what the behaviour of the malware was a little bit of an overkill. We don’t depend on the network for using our tools either so there is no immediate need for doing so. We’ll keep in mind that it’s actually up in case it becomes something to consider at some point. Nonetheless, we can find the configuration under *Control Panel > Windows Firewall*. We are attaching a screenshot showing our current configuration.

**TODO: Add screenshot!**

Please note that we intend to capture network traffic from within the guest itself so there is no need to have any external machines capturing traffic. Should that need arise we’ll shut the firewall down as needed. Again, we don’t want to take any actions that might turn up being unnecessary as we strive to carry out an analysis that’s as little invasive as we can.

**Communicating with the outside world: shared folders**

In order to speed our work process and in an effort to let us work as much as possible from our host machine we have decided to leverage the power of shared folders to share files and folders between the host and guest seamlessly. In order to do so we just need to go to *Devices > Shared Folders > Shared Folder Settings…* on the running VM. We’ll then add a new folder by clicking on the marked icon.

**TODO: Add screenshot!**

We just need to point it to the folder we want to share from our host machine and select the *Auto-mount* and *Make Permanent* options so that it’s added to the VM whenever we boot it up. We’ll see that the shared folder shows up as a networked drive within *My Computer* in Windows XP and that anything we copy to it on one end will appear on the other and vice versa. This will let us move files from one machine to the other in no time thus speeding up the analysis process by a great amount.

**Getting the tools**

A malware analyst is only as good as his/her tools so we need to get our hands on a few handy programs that will aide us in our analysis. We mustn’t fall into the trap of thinking that he more tools we employ the better and more thorough our analysis will be. Whilst it is indeed true that different tools may provide different insight, we should always trust our own judgement as the tools are submitted to us and not the opposite. Thus, once we have concluding evidence on a piece of malware, we won’t continue running it through several tools for the sake of doing it if we believe this is not necessary. That’ll let us employ less tools in a more efficient way so that the overall outcome of our work is as solid as we want it to be. We are including a complete relation of them as well as a short description so that the reader can have an idea of what we are looking for when we use each of them. Note the ones marked with \* were installed by us and their setups can be found within the aforementioned repository. The rest were already present in the imported VM.

**TODO: Add the list at the end! Tentative list:**

**Wireshark**

**InetSim**

**IDAPro**

**Cutter**

**OllyDbg**

**Process Explorer**

**Process Monitor**

**TCPView**

**RegShot**

**PEview**

**PEiD**

**Strings**

**RDG Packte Detector**

**Resource Hacker**

**TODO: Explain what other VM we are using for analysing stuff. We are just using or Mac…**

**Better safe than sorry**

Playing around with malware can be risky. We shouldn’t forget that these pieces of code are designed to, in the least, access sensitive and in the worst to wreck the machine they run on. This implies we might get to a point our VM is “broken” beyond repair so we had better get a *snapshot* of the machine before proceeding so that we have a correctly configured fallback point that we can always fall back to. We can easily get this snapshot by hitting *Cmd + T* on macOS and naming it to something suitable such as *“Backup State”*. We can also opt to add a description to it but we believe it’s not necessary given the title.

Once saved we can always start the machine from the saved state from the VirtualBox manager as seen in the following picture.

**TODO: Add screenshot!**

**Final notes**

In order to run the VM we just need to launch it from within *VirtualBox’s* manager. The user-password pair is *Administrator-AVictim.*

With that we are ready to get our hands dirty. We’ll include the analysed malware samples as we found them so that the chronology becomes another valuable asset showing how the analysis progressed with time.

**The first malware sample**

Malware will want to run undetected as long as it possibly can. This should make us consider the different ways programs can be started to try and see whether they are being weaponized to launch malicious software.

One such example are services. As seen on page 152 of [PMAL] these services are not run as standalone process. Instead, the *Windows Service Manager* will schedule them for execution according to their configuration. The tasks run in this way will remain in the background in the same way *daemons* do. The term *daemon* refers to programs that are not designed for the user to interact with. They carry out a task such as providing an HTTP server accepting external connections and their behaviour can be configured through text-based configuration files in most cases but not by direct interaction.

In order to inspect the existing services on a machine we can just *Right-click* on *My Computer* and select the *Manage* option. We can then navigate through *Services and Applications > Services* to find a list of available services. If we look at the names and descriptions we’ll soon stumble upon the *GrayPigeon\_Hacker.com.cn* service with description ª“∏Î◊”∑˛ŒÒ∂À≥Ã–Ú°£‘∂≥Ãº‡øÿπ‹¿Ì.

**TODO: Add services list screenshot!**

It certainly doesn’t look like a legitimate entry so we can click on it to gather more information about it. In doing so we’ll see that it is launching the *C:\WINDOWS\Hacker.com.cn.exe* file when it’s invoked so a sensible next step would be to analyze said file.

We’ll also try to gather some more information about the service itself to see whether it’s being started automatically upon boot, what type of service it is… This type of information can be found on the *Windows Registry* which we can access by going to *Start > Run* and typing *regedit*. We then have to navigate to *HKEY\_LOCAL\_MACHINE\SYSTEM\CurrentControlSet\Services* and look for our service. Please note we’ll contract *HKEY\_LOCAL\_MACHINE* to *HKLM* from now on. Once we get to *HKLM\SYSTEM\CurrentControlSet\Services\GrayPigeon\_Hacker.com.cn* we’ll see the different entries configuring this service.

**TODO: Add registry screenshot!**

They are quite unreadable, so we believe it’s useful to discuss the *sc* command. As seen in its help (which we can access with *sc /?*) it’s a program that communicates with the *NT Service Controller* *and Services* so that it can get information about services. We’ll be looking into the service’s configuration, so we’ll want to be using the *qc* option. Then, by running *sc qc “GrayPigeon\_Hacker.com.cn”* we’ll get the same information as before but in a more comfortable format.

**TODO: Add sc screenshot!**

We see how the service is set to be run automatically at boot (its *START\_TYPE* is *AUTO\_START. See* [*this*](https://docs.microsoft.com/en-us/windows/win32/services/automatically-starting-services) *link for more information*) and it’s of the *WIN32\_OWN\_PROCESS* type which means it’s code is an *EXE* file as stated in page 153 of [PMAL], something we already knew before!

1. **Where is the malware located?**

Either way, we have found that **the malware itself is contained in the aforementioned *EXE* file on *C:\WINDOWS\Hacker.com.cn.exe****.* We’ll try to copy it to our shared folder so that we can analyze it on our host machine which is way faster than the VM so that scrolling and combing the file is not that cumbersome.

Even though one might think copying the file would have been easy we had to fiddle with the settings a bit. It turns out the *Hacker.com.cn.exe* was marked as a *System File* which made it dodge our current folder options. If we navigate to *C:\WINDOWS* through a command line we can indeed run *dir /AH* once more and we’ll eventually see the *Hacker.com.cn.exe* appear as it should. If we run *attrib* on it to see its attributes we’ll see the following output:

**TODO: Add attrib screenshot**

The *SHR* output indicates it’s a *System, Hidden and Read-only* file. This can be seen on attrib’s help with *help* attrib. If we visit the folder view options once more we’ll have to uncheck the option hiding system files as seen in the following image.

**TODO: Add new folder config**

In doing so the file will show up in the graphical explorer as well. We just want to point out that we could have copied the file to our shared folder with the *copy* command without a problem and we could have skipped tweaking the folder settings again. This shows the power of text-based interfaces.

If we try to copy the file we’ll stumble upon the following error:

**TODO: Add error screenshot**

Nonetheless, the service that launched the file is already stopped, that is, it’s already stopped as we could see in some of the screenshots above. Then we’ll just disable it so that it doesn’t start at boot and try to copy the file once it’s not run in the first place. Once we did that the file copied without a problem. Note that if we are to use the *copy* command to copy the file over, we need to run *attrib -s -h C:\WINDOWS\Hacker.com.cn.exe* beforehand so that it stops being a system and hidden file. We can then run *copy C:\WINDOWS\Hacker.com.cn.exe Z:\MW\_sample\_1* to create the *MW\_sample\_1* in our shared folder.

Now we can run the *file* command on the sample from our *macOS* machine to find the following output:

**TODO: Add file type screenshot.**

Taking this information into account we can be sure the file is indeed a **32-bit Portable Executable file**. Now, this *PE* format is the

As seen in page 14 of [PMAL] this file format is the one employed in Windows executables, *DLLs* (Dynamic Link Libraries) and object code (the one we get when running the compiler on source code before passing it through the linker). It basically tells the Windows program loader all the information it needs for loading and running the program and it also contains the executable compiled program code itself. Information included in this header are the needed library functions, and space/memory requirements among others. As this information is provided in a known format, we can easily spot this type of file.

We can now try to inspect the file using *PEviewer* so that we can clearly inspect the different sections composing the file.

**TODO: Add PEview screenshot**

All the sections present in the analysis seem to be normal except *4s.love* and *Silvana*. If we look at the contents shown in the above image, we’ll see a bunch of text on the HEX view so we can expect some information to be there. We cannot know much more at the moment, but these 2 sections are worth to keep an eye on for later.

We can continue our static analysis (note we haven’t analyzed the running file yet) by looking into whether the file itself is packed. As stated in page 13 of [PMAL] we see that packed programs “pack” or compress the “real” or main program so that it cannot be statically analyzed. Then, an unwrapper is added on top of it so that it can be unpacked at run time and become the program it was supposed to be in memory. In other words, upon execution the unwrapper will kick in, decompress the packed code and run it as the OS itself would have done if it hadn’t been packed.

The above implies we can only really statically analyze the unpacker which isn’t of much use to us. This approach severely limits the information we can gather through a static analysis and almost always obliges us to perform a dynamic analysis of some sort.

Detecting this type of program can be done through specialized programs such as *PEiD* or by inspecting the imported functions, which can be easily done with programs such as *Dependency Walker*. A very large red flag signaling packed programs is the brevity of the imported function list. If it’s very concise we are almost certainly dealing with a packed sample as stated in page 21 of [PMAL]. After running the aforementioned tools on our sample we can conclude it’s **not packed** as it has a great deal of imported functions and it’s not deemed as packed by *PEiD* itself. We could run it through more “pack detectors” but as we are already aware of 2 factors telling us it’s **not** packed we won’t go down that rabbit hole: we need to be efficient when analyzing samples and we can’t devote time to hypothesis not backed by any evidence. We are attaching the output of both programs below.

**TODO: Attach PEiD screenshot**

**TODO: Attach Dpendency Walker screenshot**

We should end by pointing out that the approach to take for analyzing these programs, should we encounter them, is to dynamically analyze them as the program’s memory image will be the same no matter if it’s packed or not. At the end of the day, the malware sample needs to execute as a normal program so by inspecting its memory usage and attaching debuggers to it we can get a clearer idea so as to its purpose. This in fact constitutes another indicator of a program being packed. We will see that comparing the sizes of the executable and the memory image (program in execution) are uneven with the image being larger. We mustn’t forget the executable code was packed!

If we inspect the sample’s dependencies again as seen in the above screenshot we’ll soon discover that among the imports we find:

* *Wininet.dll:* This DLL contains high-level network functionality as it implements application layer protocols such as FTP, HTTP or DNS among others. We can expect our program to make use of the network to some extent. This implies we’ll analyze the traffic it produces through wireshark.
* *Advapi32.*dll: This DLL let’s the process access core Windows functionality such as the registry. As we found this piece of malware being launched from a service, we can expect the use of this library for providing that means of persistence.
* *Kernel32.*dll: This DLL is very commonly imported by process. We’ll see nonetheless that it might be being used to inject code into another process by means of the *WriteProcessMemory* function as the *IEXPLORER.EXE* string is present in the program itself. We’ll discuss this later.
* *User32.*dll & Gdi32.dll: Having these 2 DLLs present indicates this program makes use of some sort of GUI (Graphical User Interface). It can exist even if the current user can’t see it, we shouldn’t forget Windows is a multi-user system!

Taking this functionality into account we need to try and analyze the sample in a dynamic way to try and capture some network traffic hoping to find new evidence such as the URLs it’s trying to connect to for instance. Note the above information belongs to Table 1-1 from page 17 on [PMAL].

Analyzing the network traffic whilst the malware is running we can see something like the following:

**TODO: Attach wireshark screenshot**

Seeing the different DNS queries as well as the HTTP traffic we find the malware is talking to domains *yutao318525.3322.org* and *hillaryklinton.com*. This implies the malware is either receiving instructions or uploading data to these sites which makes us consider a model where the malware is a client and the server is a C&C center at one of these IP addresses (183.236.2.18 & 34.102.136.180 respectively). Note the communication with *yutao318525.3322.org* is accepted on the server’s port 8000 and the one with *hillaryclinton.com* is carried out on port 80, the standard for HTTP.

We should also note that the HTTP method we are using is POST, that is, we are uploading something to the *hillaryklinton.com* site.

Now that we are concerned with a dynamic analysis, we should mention that this process doesn’t show up on the *ProcessExplorer* program. If you recall our discussion about the imported libraries you’ll see how the *kernel32.dll* provides access to the *WriteProcessMemory* function. This can be used to inject code directly (page 257 on [PMAL]) into another process so that it runs the malicious code without the malware itself being a process. That way, the malware covertly executes on behalf of another task. This is exactly what’s happening here. The sample under study is injecting malicious code onto the *IEXPLORER.EXE* process as we can deduce by inspecting the *strings* section on the *cutter* disassembler when analyzing the program:

**TODO: Add cutter screenshot showing IEXPLORER**

We effectively find the *IEXPLORER.EXE* flag on it.

Having taken a look at the *strings* of the sample motivates a new question: is the malware obfuscated? In order to answer we must first define what the term stands for.

In an effort to hamper the malware analyst’s efforts a malware author can bloat the code with unnecessary functions, twisted logic, arbitrary jumps and the like so that our task becomes harder and we feel “lost” within the disassembled code. This also applies to the constant strings that are common to every program. Instead of including them in plain ASCII they can be encoded by means of schemes such as *base64* or any other the malware writer devises. This will make our life more difficult and we might even find ourselves forced to debug the program and inspect memory and register contents at runtime as these strings will be eventually decoded at some point so that can be used normally. All in all, the malware programmer is trying to hide how the malware operates through obfuscation by throwing us off the correct track. We mustn’t be fooled by this technique and always be mindful of what code we analyze so as to know what is going on and how to effectively counteract it. This discussion is based on page 13 of [PMAL].

Now, this particular sample sits somewhere in the middle of the definition. It’s true that some strings are readable like *IEXPLORER.EXE* but it’s also true that the URLs we know the program is connecting to **cannot** be found within its strings. We also find a bunch of rubbish (unreadable strings) in this section as well so our answer has to be that this program is **obfuscated or hidden to a certain degree**, that is, some parts are readable whilst others aren’t. We are attaching a screenshot showing some unreadable strings to complement the previous one.

**TODO: Attach cutter unreadable strings**

As we discussed before, this piece of malware is indeed leveraging network capabilities. We can only infer that from a static analysis through the import of *wininet.dll* but once we perform a dynamic analysis with wireshark we find that connections are being made to sites that, unless being made by the malware itself, cannot be explained. What’s more we are supposing the sample is injecting its code to *IEXPLORER.EXE* and this process also has access to networking DLLs such as *wininet.dll* as we can confirm with *ProcessExplorer*.

**TODO: Insert IEXPLORER process explorer screenshot**

We see that some of the networking DLLs are attached in the image above.

We would also like to delve a little bit deeper into the persistence mechanism used by the malware. As we saw before it’s starting itself through a service. Then, as seen in page 140 of [PMAL] this can be achieved by writing entries to the *run* subkey in the registry. We mustn’t forget the malware imported *advapi32.dll* which lets it modify the registry. What’s more, if we look for the word *run* within the sample’s strings, we’ll see the following:

**TODO: Add run subkey cutter strings screenshot**

Knowing the malware is using this mechanism for persistence we can then take actions to counteract it when deleting the malware.

**Summary of #1**

After all the above we can then summarize the samples behavior as follows since it’s executed:

1. The malicious executable is moved to *C:\WINDOWS\Hacker.com.cn.exe* and it’s marked as hidden and a system file so as to dodge detection.
2. It sets up a service that runs automatically on boot so that it’s persistent.
3. When running it injects malicious code into the *IEXPLORER.EXE* process which will serve as the host.
4. Then it tries to make POST requests to *hillaryklinton.com* and it also establishes a cover channel over a TCP connection with *yutao318525.3322.org* as detected by our *WireShark* analysis.

We might be leaving something behind but as the disassembled code was not easy to analyze we cannot state any more than this with confidence. It’s true that some DLLs might be getting replaced and that we are missing out on some other functionality, but this is the overall working of the sample. Given the program injection it carries out we can classify this malware as a trojan. If we run it through [*VirusTotal*](https://www.virustotal.com/gui/) we’ll see it’s detected as such by 68 antivirus:

**TODO: Add VirusTotal screenshot**

**Sample #1 removal**

As we did when we wanted to copy the original executable to another place for analysis we can just disable a service so that the malware itself doesn’t start at boot. We then only have to delete the original executable to get rid of the entire thing. Nonetheless, we prefer to delete the service rather than just disabling it so we can just run *sc*, the same tool we used to query the service’s information, to delete it. We can do so by just hitting *Run > cmd* and executing the following:

**TODO: Add service deletion screenshot**

We can just then run the following to remove the executable and finish the sample off. Note that now that we know the malware is injecting code into *IEXPLORER.EXE* we know what was causing the issue that wouldn’t let us copy the file at the beginning of the analysis. We can just kill the Internet Explorer process from *ProcessExplorer* or the Task Manager and delete the function without a problem. The fact the killing the browser lets us delete the file is a clear indicator that the malicious code is being injected into the otherwise legitimate *IEXPLORER.EXE.*

Either way, after killing the IEXPLORER.EXE process we can just run the following to finish the removal, and therefor analysis, of this first malware sample. Please note we need to mark the file as not *hidden, read-only* or belonging to the *system* before deletion, which we do with the *attrib* command.

We should also point out that the IEXPLORER.exe process doesn’t automatically start at boot once the malware is removed. This implies this piece of malware is also starting it upon execution so that it can inject itself into it.

**TODO: Add deletion screenshot**

**Second malware sample**

Now we’ll take a look at the different processes displayed in *ProcessExplorer* to look for anything unusual that could point us into the right direction to find another piece of malware. When inspecting the DLLs associated with all the processes we’ll see that one of *winlogon.exe’s* DLLs does **not** live under *C:\WINDOWS\system32* or *C:\WINDOWS\WinSxS*: *C:\WINDOWS\msagent\intl\MS\_PMAL\_Agent\BinaryCollection\Chapter\_11L\msgina32.dll.*

**TODO: Add process explorer screenshot**

This DLL doesn’t look like something that’s legit. We’ll just go to the *Chapter\_11L* folder and try to see what its contents can tell us.

As we could expect, all the contents under *MS\_PMAL\_Agent* are hidden. As we had already configured our system so as to show these files we could nonetheless traverse the directories without any major problems.

Upon reaching the folder we find two executable files: *Lab11-01.exe* and *Lab11-03.exe*. After getting both of them out of the VM for an easier handling, we’ll perform a basic *strings* on them to see if there is anything of use.

Given the DLL that brought us to this folder is *msgina32.dll* we’ll filter the output with *grep* to look for it. We can easily do so with a pipe redirection (|) to chain both commands. With that in mind we can see the following:

**TODO: Add strings output**

As the *msgina32.dll* itself is contained in the *Lab11-01.exe* executable we can be pretty safe that that’s the one we are to delve deeper into to find more useful information.

Taking into account page 235 on [PMAL] we begin to see a lot of elements fall into place. We found this page thanks to the book’s index and we were utterly surprised when we discovered that *msgina.dll* is the name of a legit DLL loaded by the *winlogon* process. This set us on the track of a GINA hijacker. As explained in the book, this type of malware leverages the *Graphical Identification and Authentication* system built into Windows XP to recover user credentials. What we’ll see is that the original *msgina.dll* is replaced with a modified version that will in most cases just forward *winlogon’s* calls to the real functions also exports (*winlogon* needs to be able to call them). In other words, the malicious DLL positions itself in between *winlogon* and *msgnina.dll* so that its operation is mostly transparent. The catch is that it’ll run some malicious code before calling the real *msgina.dll* code in some cases. As the GINA system manages user authentication this DLL will have access to the user credentials at some point. Don’t forget they’re a parameter that’s to be passed to the real *msgina.dll* and, as all the information from *winlogon* to *msgina.dll* flows through *msgina32.dll* it will at some point know the user’s login credentials. What it’ll then do is log them out to a file that we can later retrieve. Note that this file tries to conceal itself by having a name resembling that of a driver (\*.sys) so that we don’t expect it to contain stolen credentials in the first place. This strategy is just exploiting a system Microfost put in place to provide support third party GINA modules!

**TODO: Add GINA hijack diagram**

If the above were to be true we can dig a little deeper to try and find new indicators of this type of malware. As we explained before, the malicious DLL needs to eventually call the real *msgina.dll* functions so it needs to include them. These are all prepended by *Wlx* so we can try to look at the strings output again but filtering for this type of function names. Doing so revealed the following:

**TODO: Add strings wlx screenshot**

We need not lose sight of the structure of the attack. We expect to find these function calls in *msgina32.dll*, not on the executable itself. The above output then implies the executable must contain the DLL in some sort of way. We’ll then try to extract it from the executable itself, something we can easily do thanks to the *ResourceHacker* tool as seen below.

**TODO: Add resource hacker screenshot**

We’ll then compare both the extracted file and *msgina32.dll* which is the one that was already loaded by *winlogon*. We can do so thanks to the diff tool and, as expected, the files are identical. Note we saved the resource as *Likely\_msgina32.dll*.

**TODO: Add diff output**

As we can be pretty confident the executable is just dropping the DLL to the same folder it lives in we can postpone the dynamic analysis for now. We know what to look for in the disassembly: a call for writing the payload the executable carries and another for altering the registry value configuring the DLL to be loaded by *winlogon.* As shown in page 235 of [PMAL] we expect the key to be changed to be *HKLM\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Winlogon\GinaDLL.* We can indeed confirm this string is contained in the executable:

**TODO: Add strings filtered for the regentry**

We’ll now look at the disassembly of the program just to be sure our guess is what’s actually going on. It’s worth noting this program is **not** **packed** as we’ll see by getting access to the code through disassembly. There’s no need to run it through packet detectors like *PEiD* or the like. If we look at the dependencies, we’ll nonetheless see how it only imports *kernel32.dll* and *advapi32.dll*. The latter will be used to modify the registry whilst the former will allow the “dropping” of the malicious *msgina32.dll* to the file system.

**TODO: Add dependency walker screenshot.**

Even though there are not that many entries we indeed see the ones the program will need so in this case not having a lot of imports is not denoting the program is packed, it’s just carrying out its purpose with surgical precision.

We are using *cutter* to analyze the code. In doing so we’ll begin reading at *entry0*. After getting the command arguments we’ll *call fcn.* *0x004011d0* which gets us to the following:

**TODO: Add cutter screenshot A**

First of all, we’ll *call fcn.0x00401080.* Within it we’ll call the *FindResource()* and *LoadResource()* functions among others. This subroutine is then in charge of extracting the *msgina32.dll* from the executable itself. After several other calls we’ll eventually jump to another subroutine with *call fcn.00401320* which will execute the following:

**TODO: Add cutter screenshot B**

This will eventually call the *WriteFile()* function which makes us be pretty sure this is in fact unpacking the payload DLL.

If we go back to *fcn.0x004011d0* we’ll eventually reach a *call section… text* which will take us to *fcn.00401000*, which we include below:

**TODO: Add cutter screenshot C**

We can easily see the calls to *RegCreateKeyExA()* and *RegSetValueExA()* which give away how the executable is configuring the register to force the loading of the malicious version of *msgina.dll*.

In light of how explicit information was on the executable we must conclude it was **not obfuscated**. Even though it’s true that some of the strings displayed by and unfiltered *strings* are not readable the fact that we were able to get all the above with a somewhat simple static analysis prevents us from even considering this to be an **obfuscated** sample.

What’s more, this executable did not import any network related libraries so it’s **not leveraging** network capabilities at all.

After breaking down this file we can be certain it’s an **installer**. It’s only concerned with setting up the system in such a way that the malicious DLL it carries can collect the credentials it needs but once it’s run it’s not necessary. That is, once the registry is set up correctly and the *msgina32.dll* is installed on the system there is no more need for this executable: it can be deleted. This makes it hard to just state the malware itself is the executable as once it’s executed it becomes a 2-file combo. Nonetheless, it’s distributed as this standalone executable so if we had to choose one or the other, we would have to say the malware itself is *Lab11-01.exe* itself. Nonetheless we believe it’s more accurate the say that once it becomes “alive”, the malware is integrated by both the executable and malicious DLL as the one ultimately retrieving the credentials is the DLL itself.

Anyway, we can now analyze the *msgina32.dll* file to take a look at its contents.

Upon opening it up in cutter we find that whenever a legit *msgina.dll* function is called it’ll be transparently run. We have structures like the one below which will just call *fcn.00401000* to silently call the original functions. *fcn.00401000* will just call *GetProcAddress()* to get the location for the real function. It’ll be returned on *EAX* and used for an unconditional jump as seen. Jumping unconditionally will return control to the original DLL function caller as no return address will be pushed onto the stack: that of the original caller will still be there. In this case the original caller is *winlogon*. This is based off of page 569 of [PMAL].

**TODO: Add cutter screenshot d**

Anyway, we’ll find that this is the case for all the common methods **except** for *WlxLoggedOutSAS*(). The body for this one is quite larger as seen in the following screenshot. Please note that we had already found the function logging user credentials and we renamed it to *logCredentials* manually so as to make the screenshot clearer:

**TODO: Add cutter screenshot e**

After calling the real *WlxLoggedOutSAS()* the user credentials will be logged to a location specified within *logCredentials* itself.

If we now inspect *logCredentials* we’ll see the following:

**TODO: Add cutter screenshot f**

We have already added a red box around the important calls. The thing to take out is that the filename *msutil32.sys* is provided as an argument before opening a file. Then, this must be the file these credentials are being logged to! As the DLL is imported by *winlogon* which sits on *C:\WINDOWS\system32* we can expect to find it just there. We should point out this string did show up when running *strings* on the sample.

As the *WlxLoggedOutSAS()* method is a *hook* that runs when the user logs out then the credentials will be stolen right there and then for any user.

In any case, if we get the file and take a look at its contents we’ll see the following:

**TODO: Add stolen credentials screenshot**

It’s a regular text file with our credentials! Each entry corresponds to one logout. We can check that if we delete it it’ll be created once more after we log out, provided the malicious malware is still running.

Note the format agrees with the following string included in the DLL and referenced from *logCredentials*. It’s printing out 4 strings (%s): our username, our domain name, our current password and the old one. This is the information we indeed see above. The format string is included in the next picture

**TODO: Add cutter format string screenshot**

The way to get rid of this nasty program is to just delete the associated registry entry we discussed at the beginning of the analysis. It’s **not** necessary to download a legitimate copy of *msgina.dll* as it was already being included by *winlogon.exe*. We just need to navigate through the registry and either delete or rename the *GinaDLL* key. We chose to rename it to *Foo* so that we could test things out should something not work as expected

In order to test our hypothesis we just need to delete the *msutil32.sys* file and logout to see if it’s created again after deleting the GinaDLL registry entry.

As expected, this wasn’t the case: we just need to delete/rename the registry key. What’s more, upon deleting or renaming said key we’ll see a change to the login screens style: we shouldn’t forget we are altering the path the calls to GINA modules take so we can expect small changes such as this to happen.

**TODO: Add clean registry screenshot**

**Summary**

Summing up, we see how *Lab11-01.exe* is a common **PE executable** (as shown by the file command) installed a malicious DLL that got between *winlogon* and *msgina.dll* in such a way that it retrieved user login credentials upon user logout. It did so by adding the *GinaDLL* entry in the registry and pointing it to the malicious DLL. Then, upon execution of *WlxLoggedOutSAS()* the credentials were being written to *C:\WINDOWS\system32\msutil32.sys* which was nothing more than a text file. Deleting or renaming said registry entry is enough to take care of the entire problem.

**Sample number 3**

Once we have discovered the *C:\WINDOWS\msagent\intl\MS\_PMAL\_Agent\BinaryCollection* directory we can just dive deeper into it to find a complete collection of rather suspicious executables.

After browsing around them we settled on analyzing the *Lab12-03.exe* file next living under the *Chapter\_12L* directory belonging to the aforementioned one.

We’ll begin by running a common static analysis on it. In doing so we’ll launch *DependencyWalker* against it to get the library functions it’s importing. Even though not a whole bunch of them show up we can conclude several things:

* The sample is **not** packed. Despite the low number of imports, they convey a whole lot of information about the functionality of the sample as we’ll see later. Refer to the screenshot below for a highlighted list of crucial imported functions. What’s more, we are also including a screenshot of *PEiD’s* output showing it’s been compiled with Microsoft Visual C++ which is not by any means a packer.
* The process **won’t use any network capabilities** as no networking-related DLL is imported. We are also sure it’s not hijacking or injecting itself into any other projects as we don’t import the *WriteProcessMemory()* function from *kernel32.dll* so we won’t have a tangent access to networking recourses either.
* The file itself is a PE executable file as shown by the *file* command. We also ran *PEview* on the sample but as nothing seems to be out of the ordinary we believe it’s of no use including any screenshots whatsoever: we are up against a very common executable file with regards to its format.

**TODO: Add dependency walker screenshot**

**TODO: Add PEiD screenshot.**

We would like to draw special attention towards the *SetWindowsHookExA()* function shown in the *Dependency Walker* screenshot. As stated in the Microsoft Developer Network (MSDN) [Documentation](https://docs.microsoft.com/en-gb/windows/win32/api/winuser/nf-winuser-setwindowshookexa?redirectedfrom=MSDN), the *SetWindowsHooExkA()* function will provide a callback, a function that is to be called upon the occurrence of a given type of event. In other words, our sample will just “subscribe” to a certain occurrence and will be notified by the OS itself when it happens so that it can take the appropriate actions. Upon notification, the callback of *hook* function will be called.

We then need to delve deeper into the code to try and see what the callback function is and what type of event our sample is subscribing to. To do that, we’ll employ *cutter*.

Inspecting the documentation we linked above we see the first parameter to the *SetWindowsHookExA()* function determines the type of event it’ll listen for. As the parameters are passed in a reverse order (the last one that’s referenced is the first positional argument) due to the stacks LIFO (Last In First Out) nature we can look up the value of the *idHook* parameter (13 == 0xD) shown in the next instruction on the linked documentation. In doing so we’ll see that the value stands for *WH\_KEYBOARD\_LL* which “Installs a hook procedure that monitors low-level keyboard input events” according to the site.

**TODO: Add cutter screenshot a**

We then see how this program is concerned with keyboard input which makes us believe it’s some form of a keylogger right from the beginning. Having this lead early on will let us carry on a much more precise analysis because we now know what we are looking for. We can also see in the screenshot above that the callback is provided as the second parameter. In our case it’ll be the *keyboardHook* subroutine. Please note we have renamed the subroutines involved in the malware’s functionality to facilitate our work. This is the case for *main*, *keboardHook* and *writeLog*. The former two can be seen in the screenshot above.

We would also like to draw your attention to the while loop we have shown in the image that will call *GetMessageA()* continuously. As seen in the [MSDN documentation](https://docs.microsoft.com/en-us/windows/win32/api/winuser/nf-winuser-getmessage), this function will block the caller (i.e. the malware) until a message related to one of the events it’s subscribed to is received. Upon said reception the hook function we have defined will be called to handle it. If we were not to call this function we would effectively subscribe to an event to then not ask for any updates and thus, the malware’s operation wouldn’t be the expected one.

The above is all there is to the *main()* function it will just wait in that loop indefinitely. We should then proceed to analyze the *keyboardHook* subroutine to see what it’s executing.

We’ll begin by inspecting the disassembled code attached in the image below. We can see that the method is pretty straightforward. As we stated before it’ll be called on keyboard events and it’ll receive 3 parameters as seen in the [MSDN](https://docs.microsoft.com/en-us/previous-versions/windows/desktop/legacy/ms644985(v=vs.85)). The first one is of little concern to us and we only need to know that, as seen in the documentation, if the valid is not 0 we should abort immediately and return. Whatever *CallNextHookEx()* returns. The second parameter one is telling the hook what happened with the key and we’ll only delve deeper into the function if we have a key that’s down no matter if we were pressing ALT at the same time or not. We find the values shown in the disassembly in [these](https://docs.microsoft.com/en-gb/windows/win32/inputdev/wm-syskeydown?redirectedfrom=MSDN) [two](https://docs.microsoft.com/en-gb/windows/win32/inputdev/wm-keydown?redirectedfrom=MSDN) sites of the MSDN. The third parameter is a [*KBDLLHOOKSTRUCT*](https://docs.microsoft.com/en-gb/windows/win32/api/winuser/ns-winuser-kbdllhookstruct?redirectedfrom=MSDN) containing information about the pressed key. The filed we are the most interested in is the first one which contains the [*Virtual Key Code*](https://docs.microsoft.com/en-us/windows/win32/inputdev/virtual-key-codes) associated with the event. Note that the parameter is passed by referenced to the hook, that is, we are passed a memory address. Also note the *vkCode* struct member is the first one, that is, it’s at offset 0 from the passed pointer. That’s why on the marked lines we see that this value is read into ECX and then pushed so that the next function in line, *writeLog* can use it. We have renamed the interesting parts so that they are easier to spot. We should also mention that the call to *CallNextHookEx()* is “not” mandatory but tremendously encouraged as seen [here](https://docs.microsoft.com/en-us/windows/win32/api/winuser/nf-winuser-callnexthookex).

**TODO: Add cutter b screenthot**

As discussed, the *writeLog* function will receive a single parameter (*capturedChar*, we renamed it) that’s the key associated with the keyboard event. We could just try to write it to a log file, but we would then incur into some problems.

First of all, we wouldn’t know the program the key was pressed on. We could try to get the program’s name to get a bit more context so as to where the key was pressed. Then, if the key we pressed is NOT a printable character we need to find a way of representing it with a string. That is, if we press BACKSPACE, we should write BACKSPACE to the log file, not the associated code as that’s not meaningful.

**TODO: Add cutter c screenshot**

Taking the above into account we can take a look at the *writeLog* function we are attaching below and, as we read through it, we’ll notice several aspects:

* At the very beginning there is a call to *CreateFileA()* which is passed the “practical*malwareanalysis.log”* parameter. This already appeared in the *strings* output and hints that the file everything is being logged to is indeed “*practicalmalwareanalysis.org”*. The file descriptor for that file is stored into *iVar1* in the screenshot below.
* We then see a couple of calls to *GetForegroundWindow()* and *GetWindowTextA()* respectively. This will first get the window the key was pressed on and then it’s title so that we can add it to the log. Note the later call to *WriteFile()*.
* We’ve got several other calls to *WriteFile()* and a lot of jumps. Let’s make use of *cutter’s* decompiler to try and get a clearer view of the functions structure.

The above is pointed out in the following screenshot:

**TODO: Add cutter screenshot d**

If we pay closer attention to the switch statement that appears, we’ll see how it handles the “string conversion” of the non-printable keys. In order to do so we’ll try to walk through what happens if the backspace key is pressed on the disassembled code.

According to the site we linked before, the virtual key code for BACKSPACE is 0x08. Now, according to the following screenshot we’ll subtract 8 from it and use it as an offset to index address 0x00401220. As the MOV instruction uses the CL operand as the destination we’ll only move the byte addressed by the base + offset combination, nothing more. Our offset is then 0x8 – 0x8 = 0x0; the first byte in the table.

**TODO: Add cutter screenshot e**

**TODO: Add cutter screenshot f**

The table image we are showing above has been obtained by showing the contents of that region as data and not instructions. We can then see that the first few elements are: 0x0, 0x1, 0x12, 0x12, 0x12, 0x2, 0x12, 0x12, 0x3, 0x4, 0x12… As we said before our offset is 0 so that value 0x0, the first element in the table, will be loaded into CL. We’ll use that value, multiplied by 4, as an offset into the *switch’s* table that’s located at address 0x401441as seen in the following screenshot.

**TODO: Add cutter screenshot g**

If we take a look at the switch table attached in the image below, we’ll see that the first entry corresponds to entry 0. We would like to point out that we have discovered exactly how a *switch* is implemented in assembly: a table with the addresses to the different cases where we use the input parameter (multiplied by 4 as we are dealing with 32-bit (4 byte) addresses) as an offset within it to jump to the appropriate case.

**TODO: Add cutter screenshot h**

Either way, we see how the first entry corresponds to *case 0* and if we take a look at it, we’ll find the following:

**TODO: Add cutter screenshot i**

We should mention that the cases are not in order in the disassembled code. This is most likely due to alignment requirements so that the program runs faster by making memory accesses to it more efficient, but we can’t be sure.

The way keys are recorder into the log is a little bit different. Instead of going through the *switch* we’ll instead write the passed parameter directly after adding 0x20 to it. We’ll see the reason for that value and how it makes the keylogger only record keystrokes as lowercase keys rather than upper case ones. We shouldn’t forget we also know if the SHIFT key is being pressed so we can infer when the user is typing with caps. To check the above we can refer to the decompiled version of the function again as it’ll make our life way easier:

**TODO: Add cutter screenshot j**

We find that if the virtual key code is between 0x41 and 0x5A included (the ASCII and virtual key codes for ‘A’ and ‘Z’, respectively) we’ll just increment the number by 0x20 and write it to the file. Then, if we capture the ‘A’ key we’ll get 0x41 + 0x20 = 0x61 = ‘a’, the ASCII code for lower case A. This is what we’ll ultimately see written to the file! We’ll check this is true when attaching some screenshots showing the malware’s generated logs.

Either way, we have broken down the code and we know that the program will start, set a hook for keyboard events and log the pressed keys to the “practicalmalwareanalysis.log” file living in the same directory as the executable. In order to test this, we can just run the program. And open up a notepad to write stuff into. Notice that the log file will be appended to if it exists or created if it doesn’t. We started with a blank one to generate the following screenshot:

**TODO: Add working malware screenshot**

Now that we know the malware is running, we can just open up *ProcessExplorer* and we’ll effectively see it appear as a process. To stop the keylogger we just need to kill that process. As it’s taking no measures to ensure it’s persistent, such as dropping keys into the registry, it will only start if clicked on by the user. Then, killing the process and removing the executable from the system ensures the malware’s complete removal. Even though it may not be such a serious thread with regards to how it won’t start automatically it’s carrying out its keylogging functionality perfectly fine.

We would like to mention that if you open up several keylogger processes they’ll begin fighting among themselves and the log will be somewhat incomplete as there is no synchronization mechanism controlling that the file access is done in an orderly manner.

Summing up, we see how the malware is contained in a PE executable that implements a keylogger leveraging Windows hook functions. Once the executable is run all the keyboard input will be logged out to a text file in the same folder as the executable. Removal is easy however as the malware is not trying to start automatically. We can check that by looking for the “practicalmalwareanalysis.log” file across all the filesystem. As we only found the one we knew existed we can be sure no other instances were running camouflaged within other processes. We can just kill all its process instances and delete the executable to completely remove it. We also saw that it wasn’t packed or obfuscated in any way despite finding some unreadable strings as all the crucial information was free for us to read and interpret. No networking capabilities are employed either as no networking libraries are included. Through the study of this piece of malware we have become familiar with Windows hook functions as well as with the assembly language constructs commonly used within this operating system. We are now ready to analyse our last malware sample!

**Malware sample 4**