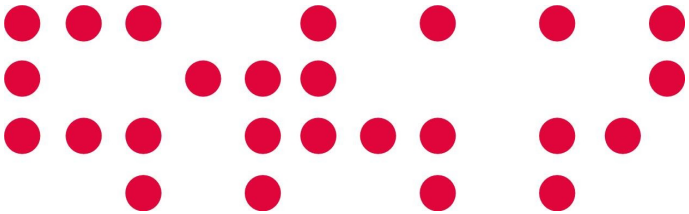
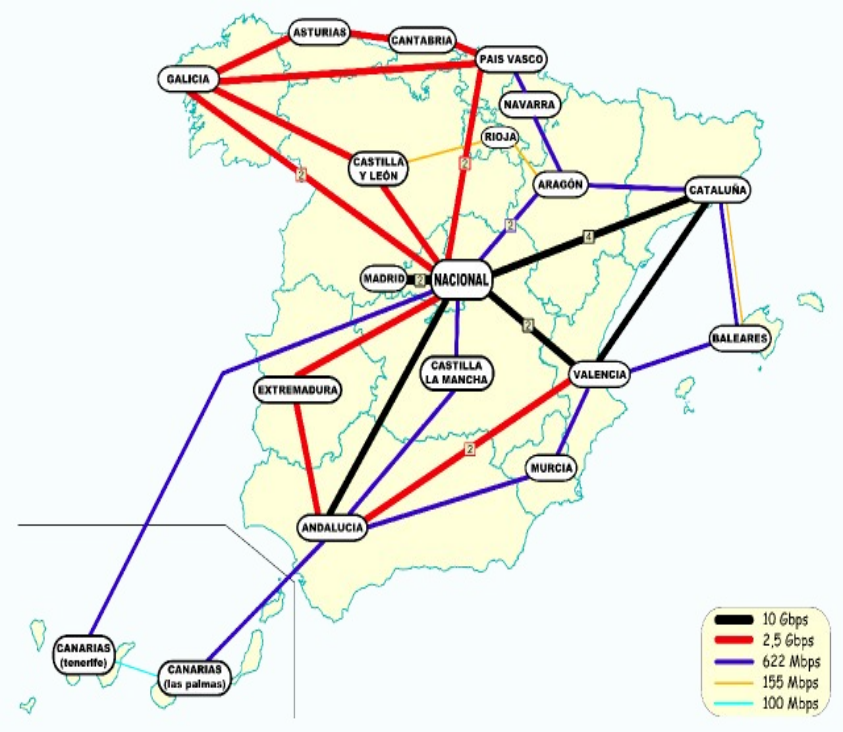
Creating a Malware analysis Laboratory

Francisco Jesus Monserrat Coll
IRIS-CERT / RedIRIS
FIRST TC /COLARIS ,
Montevideo UY. NOV 2008

Index



About RedIRIS



- Spanish Academic & Research Network
- Interconnect 250 Universities & Research centers
- Part of goverment company, red.es
- IRIS-CERT, CSIRT inside RedIRIS

definitions

- Malware: Software designed to infiltrate or damage a computer system without the owner's consent. It's a mix of malicious and software. http://en.wikipedia.org/wiki/Malware

- Refers to other similar terms, bot, virus, worm, spyware, adware , ….

definitions

- Laboratory: (or lab), a facility that provides controlled conditions in which scientific research , experiments and measurement may be performed , again http://en.wikipedia.org/wiki/Laboratory

What's a malware lab ?

- Systems /place that provides:

- Controlled environment : All the information must be recorded for later usage.
- Isolated: The malware must not be allowed to contact with any external source , but…
- Full simulation: The laboratory must provide all the resources needed by the malware.

Why we need a malware lab ?

- Analysis of unknown files
- Public information from antivirus & Security Companies is not complete
- Private information about the malware required a expensive paid service.
- Customized malware would target small organization.

How can we get the malware ?

- Unfortunately malware are quite easy to obtain:
  - Spamtrap
  - From honeypots
  - Received from another CSIRT or group
  - From our costumer, when handling an incident

Spamtrap

- Malware received by email.
- First used by worms, now used to propagate bots and bank keyloggers.
- Search for “http:/.*{,cmd|.exe|.scr} in you spam folder.

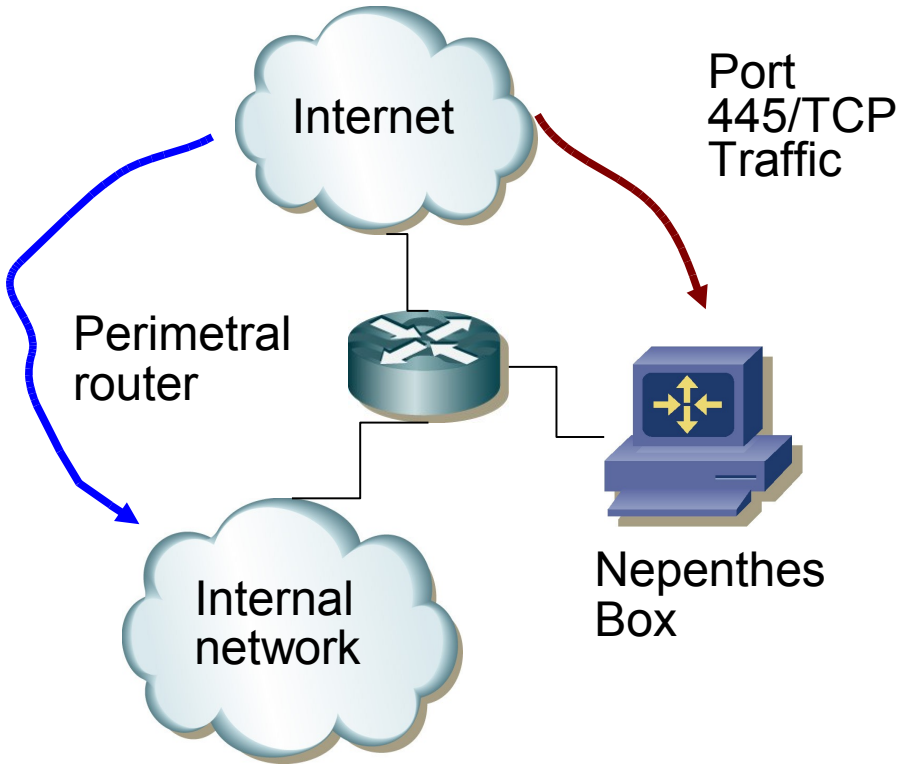
Malware from Honeynets

- Recovered from complete machines
- Automated capture systems.
  - Nepenthes, http://nepenthes.mwcollect.org
  - Vulnerable service simulation (Ex: MS-RPC)
- ...and the good news are...
  - Do NOT execute the buffer overflow code
  - Parse the attack and simulate an infected system
  - Download and store those interesting payloads

Malware from the web

- Web pages that contains malicious code:
  - Intruder change web pages
    - Add a obfuscated Javascript code
  - End users browsers will execute the code and download the binary
  - Used this year with the MS-SQL worms .
- Mpack & friends to control the web pages and users infected
- Also exploit at the document format level (PDF) at the end used to download a binary.

Getting binaries from outside



- Instead of blocking malicius trafic (ex 445/TCP) , redirect it to a nepenthes box
- Use DNAT in your nepenthes box to accept and simulate the victims
- ~10,000 files /day

From our costumer

- Perhaps the most difficult.
- Phone calls to help desk,
  - Why my computer is running slowly ?
- from outside:
  - Your computer is scanning me ....
- Or from you own sensors

- Freeware tool from MyNetWatchman
  - ***http://www.mynetwatchman.com/tools/sc***
- Analyzes the system and generates a plain-text report:
  - Processes running
  - Open files
    - DLL information (used by processes)
  - Network information
  - Running services
- Some worth tool to send your users to provide you that useful information

Other similar tools

- Hijack-it,
  - http://www.merijn.org/index.php
- Sysinternal tools
  - http://www.microsoft.com/technet/sysinternals
- Foundstone tools
  - http://www.foundstone.com/index.htm?subnav

- Group a lot of forensic tools together
- Allow automatic recovering of evidences from a suspected compromised machine.
- Can be expended to include more tools:
- See http://code.google.com/p/rapier/ and

  http://www.first.org/conference/2006/prog
- Good to teach users a recovery procedure.

The Lab

Lab elements

- Victim machines : In which the malware can be run.
- Support tools for building the lab
- Network simulation: One of more elements that simulate internet for the malware.
- Analysis tools : that can be used to analyze the malware.

- So how many machines do we need for our lab ?
- Hardware is not only expense, but
  - Difficult to maintain
  - Too much space ..
- Virtualization software can be used to reduce this cost.

Virtualization software

- run different virtual machines at the same time.
- Run unmodified version of most operating system
- Allow to have different , isolated networks for the machines.
- Machines can be connected to the real interfaces.
- Provide a “redo option” to restart the virtual machines

Virtualization Software

- VMWARE,
  - http://www.vmware.com , commercial, but with free & demo products
- Parallels
  - http://www.parallels.com , similar to vmware, has a Mac OS x86 version
- Qemu
  - http://fabrice.bellard.free.fr/qemu/ , opensource can run in diferent architectures
- Bochs
  - http://bochs.sf.net , first one , full software emulation

Virtualization problems

- Malware is incorporing code to detect virtualization environment.
  - Check for some special drivers
  - Check for some special devices.
  - Red Pill from Joanna Rutkowska
    - http://invisiblethings.org/papers/redpill.html
- Sometimes we need to try with another virtualization software or use real machines.

- Automatic clean up of the machines
- Dual boot:
  - Boot to windows, next boot with linux, infect the machine
  - Reboot
  - Boot to Linux, recover the installed files, registry changes, and rewrite the Windows section , next boot to windows
- http://www-128.ibm.com/developerworks/linux/library/l-oss

Network simulation

- Most of the malware need internet connection to their controlled system.
  - Internet connection
  - DNS connection
  - Irc, web, smtp, server
- The information sent by the malware must be recollected by the simulated network

Internet in a box

- Normal machines:
  - Has only a default route to internet
  - A DNS server
  - A mail server.

- We can configure a linux/Unix box that
  - Accept traffic like a router
  - Respond to the DNS queries
  - Accept traffic to some services

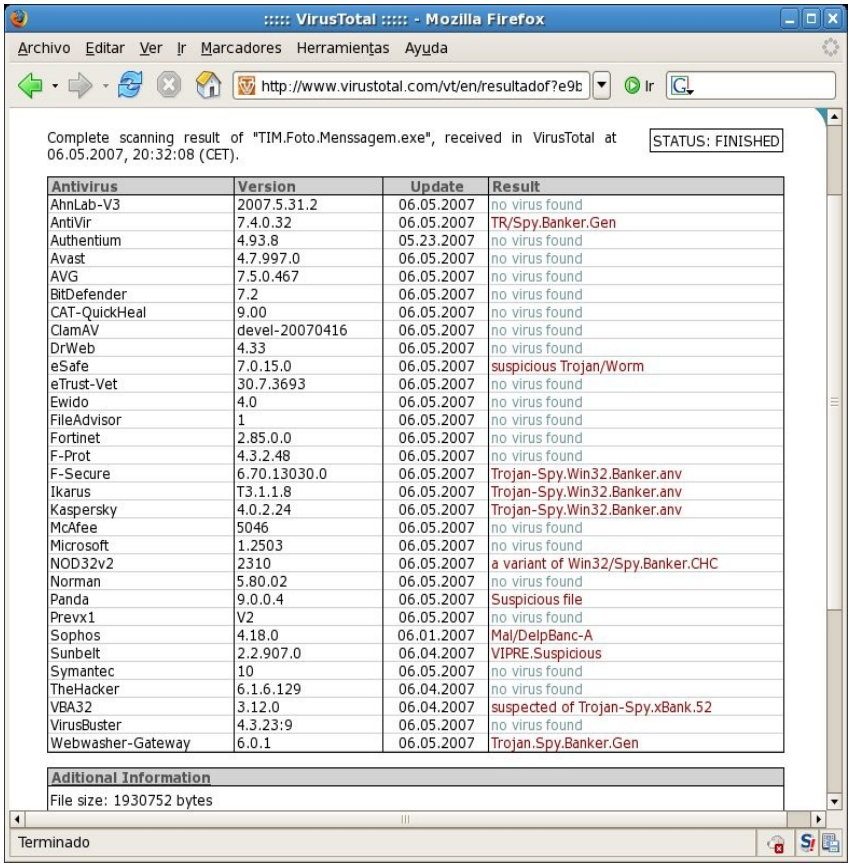
Analysis tools

- Different kind of tools:
  - Behavior analysis, executed in the test machine, to obtain information about the execution of the binary.
  - Static tools to perform static analysis of the file, (dissasembly) .
  - External tools:
    - Provide information about the tools

Remote analysis tools

- Outside the lab, public or private service providing automatic analysis of the files.
- Advantage:
  - Fast analysis of the file
  - Most of them keep the information for later referral.
- Disadvantages:
  - Moving to pay services
  - Sometimes don’t provide the required information
  - Malware can detect some of the systems

Remote: virustotal.com
http://www.virustotal.com



- Analyze a file against a battery of antivirus.
- Don't perform any analysis of the file
- Detection rate varies due to encryptatation techniques used to avoid antivirus

Why antivirus fails ?

- Most malware is encrypted / packed to avoid analysis.
  - UPX, http://upx.sf.net
  - Not possible to direclty perfom analysis based on pattern matching .
  - Cryptographic checksum (MD5, SHA1, fails)
- Malware change every few hours
  - Need to recover all the files and analyze it.

Building the lab ..

- Simple Lab
  - For analyzing behavior of Windows Malware
  - Can be run in your own laptop (with memory)

- Caution before executing the malware
  - Check that all the machines are in the correct network
  - Check that the lab is not connected to any other network.
  - Check that you are executing the malware in the correct machine

Lab hardware

- Intel based:
- Machine with
  - Memory for running three virtual machines ~2gb
  - Network interfaces
  - Disk space for storing virtual machines ~ 3Gb.

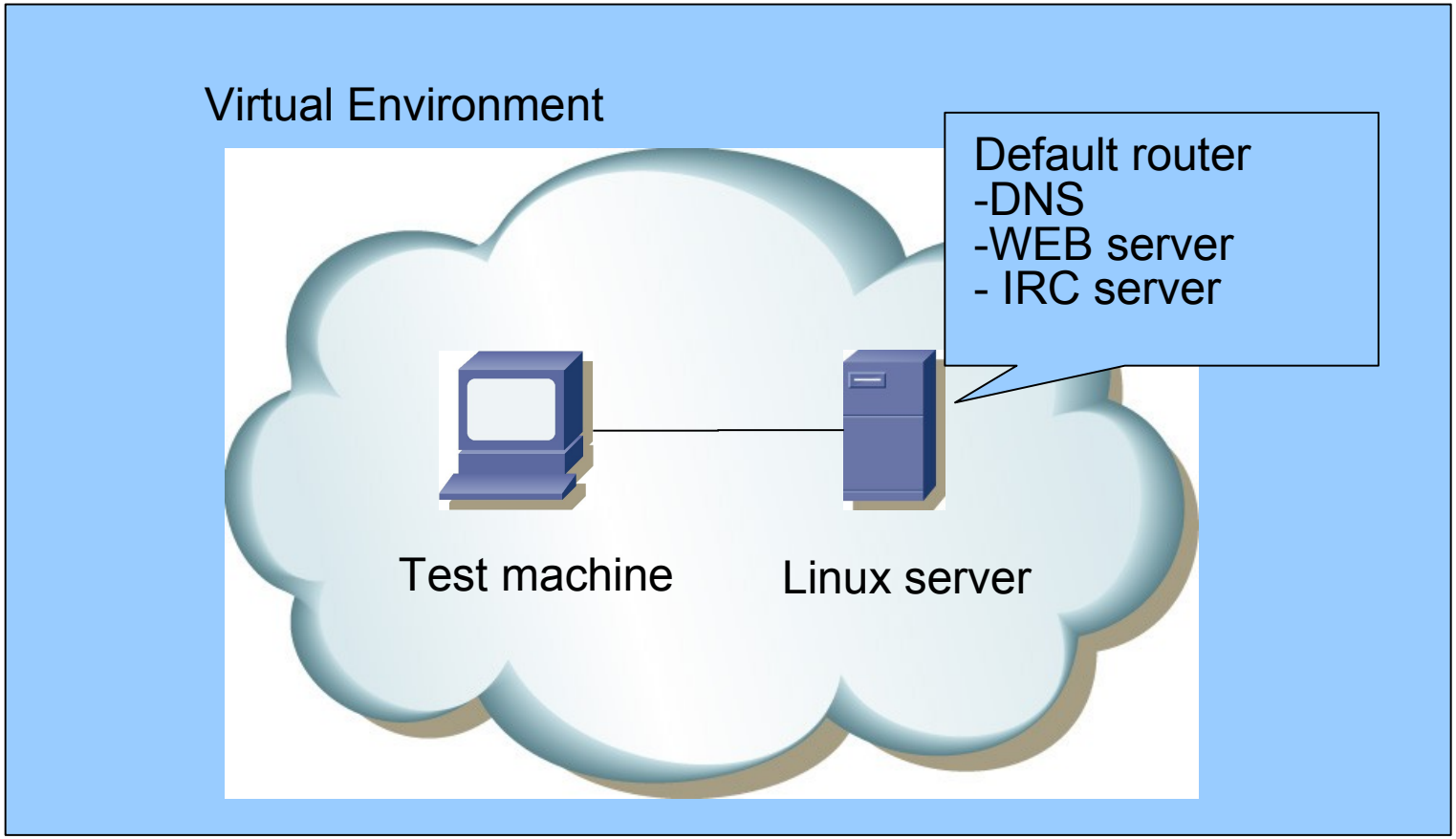
- Additional hardware/software
  - Emulator of other hardware
  - Real machines

Virtualization software

- Vmware is the most used :
  - Workstation , you can build the lab in your own laptop, or deskop , but requiere a licence.
  - Server , free , you can install the lab in a remote machine and handle the binaries remotely.

- Only two machines:
  - One to simulate the net
  - Another to execute & analyze the tool

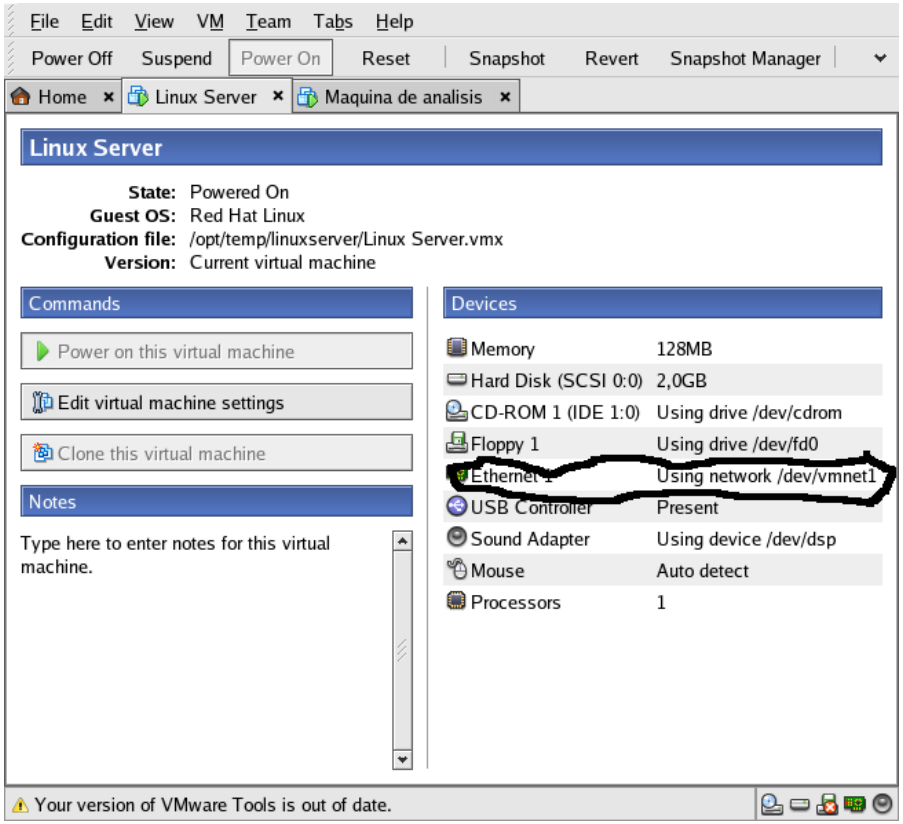
Example lab


Virtual Environment
Default router
-DNS
-WEB server
- IRC server
Test machine
Linux server

Configuring the lab

VMWare Network configuration

- Create or choose a non used VMWare *vmnet host-only* network for both the linux and windows hosts
- Kill or stop the dhcpd server on VMWare host
- Maybe a vmware-config reconfiguration could be necessary



Setting up our lab: Linux server (I)

- Used to perform simulated interaction between the *Malware* and external systems
- Provides common services needed by the Malware:
  - DNS server
  - Web server
  - IRC server
  - DHCP server (not needed)
- Use a free address range

- After booting the linux system you will have:
  - Fixed IP address ej. 192.168.100.10
  - DNS server configured to anwser with this IP address to all queries.
  - IRC servers configured in standard ports.
- Typical tools (tcpdump, ssh, netcat, etc) installed.
- Additional servers, FTP, HTTP, etc.

Configuring a DNS root server

```
// named.conf for the whole internet
options {
        directory "/var/named";
        dump-file "/var/named/data/cache_dump.db";
        statistics-file "/var/named/data/named_stats.txt";
};
controls {
        inet 127.0.0.1 allow { localhost; } keys { rndckey; };
};
zone "." IN {
        type master;
        file "fake-master";
        allow-update{ none;};
};
channel query_logging {
       file “/var/log/named_log”;
       version 3 size 10M;
       print-category yes;
       print-severity yes;
       print-time yes;
};
```

- Configuration file is *“/etc/named.conf”*
- Set up the root “.” zone to be answered by the DNS
- Logs all queries to one file

DNS: Universal resolver


```
$TTL 86400
@   IN SOA @   root(
                   42          ;serial
                   3H          ;refresh
                   15M             ;retry
                   1W          ;expiry
                   1D )            ; minimum

IN NS @

.  IN     A       LINUX_SERVER_IP
.  IN     MX 10   LINUX_SERVER_IP
```

- Configuration file is *"/var/named/fake-master"*
- Set up the corresponding fake DNS zone
- All queries will reply the same IP address

Network traffic

- Configure the default route of the windows machine to point to the Linux box
- You can use “DNAT” in the linux box to accept traffic destined to other IP address.
  - Iptables -t NAT -A PREROUTING -d 0.0.0.0/0 -i eth0 -j DNAT –to ipaddress
- Same thing can be done for port ranges

Windows machine

- Unpatched Windows machine.
  - To execute the malware
  - To analyze the malware
- Tools installed in the machine
  - Regshot http://regshot.blog.googlepages.com/regshot
  - LordPE http://scifi.pages.at/yoda9k/LordPE/info.htm
  - Binhex , from foundstone tools
  - Ollydbg , http://www.ollydbg.de http://ollydbg.ispana.es
  - Idapro , http://www.datarescue.com/idapro
  - ...

Analyzing a malware

- We are going to use a malware recovered from a nepenthes box.

  Obtain external information about the malware.

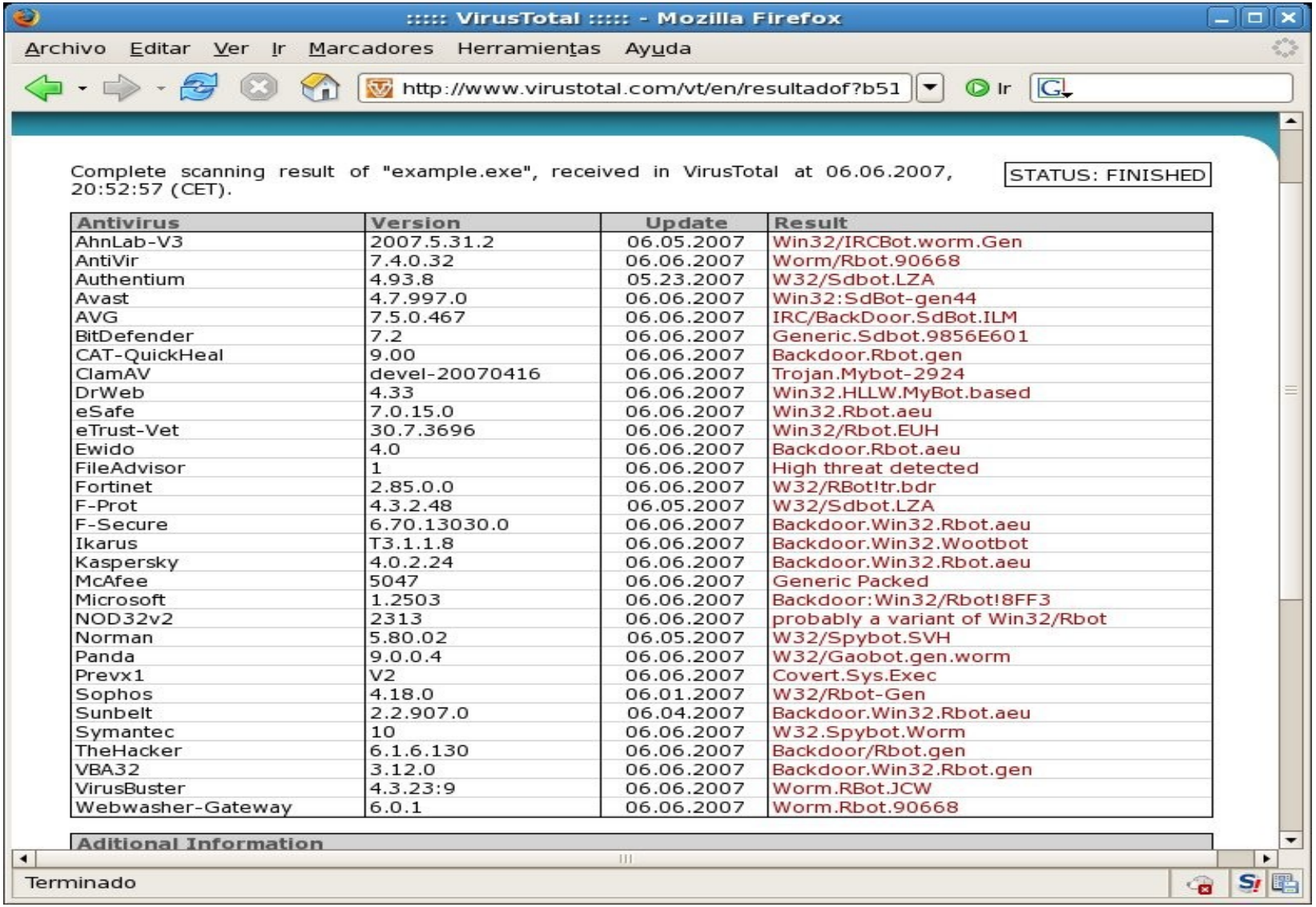
  Execute the malware in the lab to obtain some information

  Examine with a dissasembler to obtain more information

Virustotal.com

MINISTERIO DE INDUSTRIA, TURISMO Y COMERCIO

red.es



Red IRIS

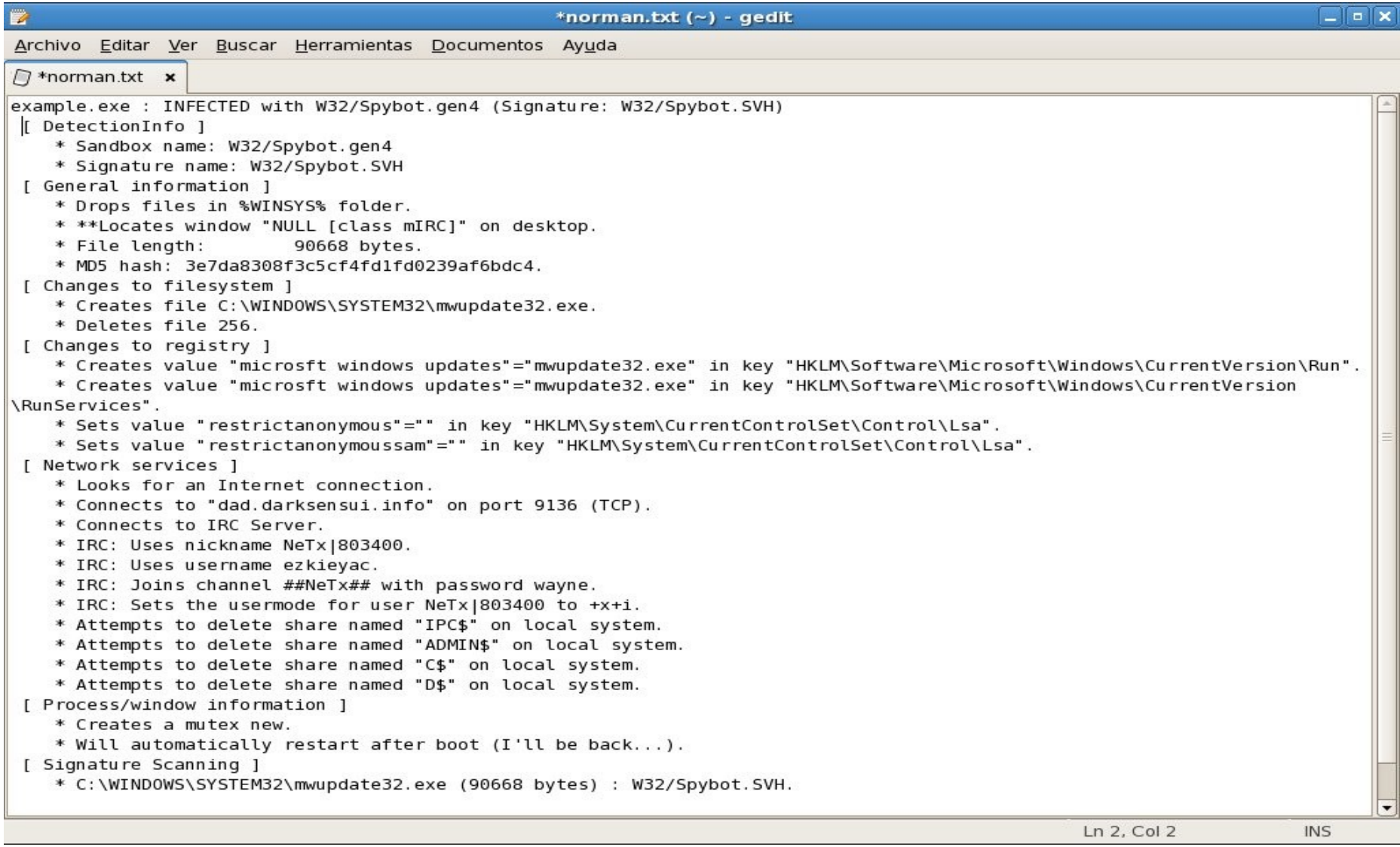
Remote: Norman Sandbox

- First remote malware analysis tool
  - http://www.norman.com/microsites/nsic/en-us
- Two level model.
  - Free, small report by email.
  - Paid service: detailed information

Norman Sandbox

MINISTERIO DE INDUSTRIA, TURISMO Y COMERCIO

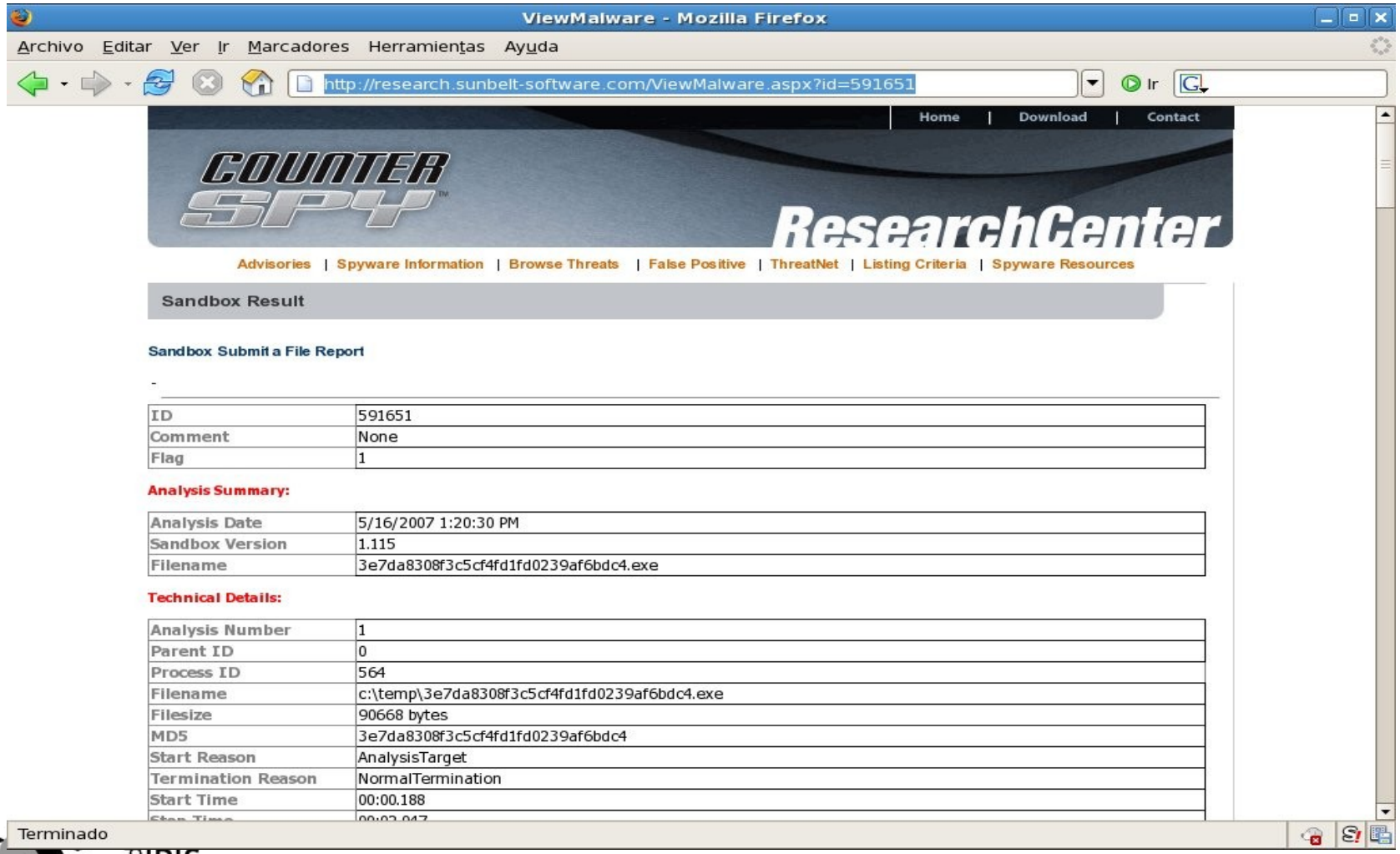
red.es



RedIRIS

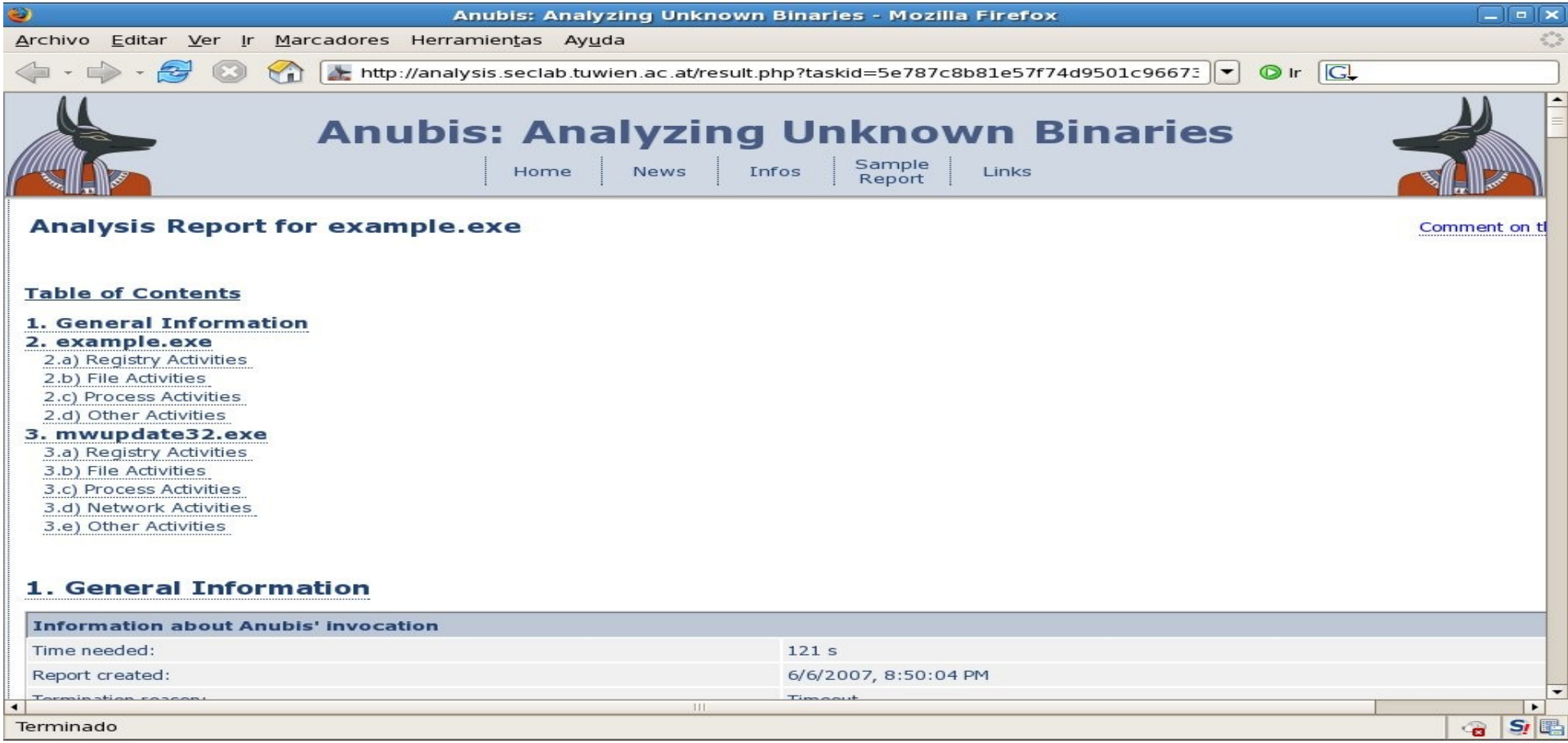
Remote: cwsandbox

http://research.sunbelt-software.com/ViewMalware.aspx?id=591651



Remote: Anubis

http://analysis.seclab.tuwien.ac.at/result.php?taskid=5e787c8b81e57f74d9501c966734d74d&refresh=1&embedded=1



How this tools works ?

- Use a virtual machine to execute the malware.
  - Perform automatic check
    - Windows registry
    - File system changes
    - Network activity
  - DLL hoocks
    - Replace operating system API
    - Malware calls the API
    - The new dll log the call and execute the windows API

Capturing the traffic:

- BEFORE launching the "malware" we need to launch *tcpdump* in the Linux VM box to record the traffic
- *Tcpdump -n -s 2000 -w /tmp/capture*
- *Useful information to get:*
  - *Host that it is used by the botnet*
  - Ports being used to connect to services

- Using Regshot we can
- check the changes when
- running a file:
- Change file path to c:\
- First “shot”
- Execute the file
- Second “shot” and compare

RegShot example

- ----------
- Values added:4
- ----------
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows\CurrentVersion\Run\microsft windows updates: "mwupdate32.exe"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows\CurrentVersion\RunServices\microsft windows updates: "mwupdate32.exe"
- HKEY_USERS\S-1-5-21-1409082233-1078081533-725345543-1004\Software\Microsoft\Windows\CurrentVersion\Explorer\UserAssist\{75048700-EF1F-11D0-9888-006097DEACF9}\Count\HRZR_EHACNGU:P:\znyjner\fcrpvzragf\rknzcyr.rkr: 01 00 00 00 06 00 00 00 D0 AF D0 A4 45 20 C6 01
- HKEY_USERS\S-1-5-21-1409082233-1078081533-725345543-1004\Software\Microsoft\Windows\ShellNoRoam\MUICache\C:\malware\speciments\example.exe: "example"

tcpdump example

- 01:25:42.120500 **IP 192.168.150.254.1029 > 192.168.150.2.domain: 24256+ A? dad.darksensui.info.** (37)
- 0x0000: 0050 5601 0203 000c 29d5 7e15 0800 4500 .PV.....).~...E.
- 0x0010: 0041 282c 0000 8011 642e c0a8 96fe c0a8 .A(,....d.....
- 0x0020: 9602 0405 0035 002d 9d6e 5ec0 0100 0001 .....5.-.n^.....
- 0x0030: 0000 0000 0000 0364 6164 0a64 6172 6b73 .....dad.darks
- 0x0040: 656e 7375 6904 696e 666f 0000 0100 01 ensui.info.....
- 01:25:42.253265 IP 192.168.150.2.domain > 192.168.150.254.1029: 24256* 1/1/0 A
- 192.168.151.2 (65)
- 0x0000: 000c 29d5 7e15 0050 5601 0203 0800 4500 ..).~..PV.....E.
- 0x0010: 005d 018a 4000 4011 8ab4 c0a8 9602 c0a8 .]..@.@.....
- 0x0020: 96fe 0035 0405 0049 87c5 5ec0 8580 0001 ...5...I..^.....
- 0x0030: 0001 0001 0000 0364 6164 0a64 6172 6b73 .....dad.darks
- 0x0040: 656e 7375 6904 696e 666f 0000 0100 01c0 ensui.info.....
- 0x0050: 0c00 0100 0100 0151 8000 04c0 a897 0200 .....Q.....
- 0x0060: 0002 0001 0001 5180 0001 00 .....Q....
- 01:25:42.334090 I**P 192.168.150.254.1107 > 192.168.151.2.9136:** S 4021988678:4021988678(0) win 64240 <mss 1460,nop,nop,sackOK>
- 0x0000: 0050 5601 0203 000c 29d5 7e15 0800 4500 .PV.....).~...E.
- 0x0010: 0030 282d 4000 8006 2349 c0a8 96fe c0a8 .0(-@...#I.....
- 0x0020: 9702 0453 23b0 efba ad46 0000 0000 7002 ...S#....F....p.
- 0x0030: faf0 13d8 0000 0204 05b4 0101 0402 .....

IRC information (I)


```
0x0040:  6554 787c 3836 3032 3434 0d0a              eTx|860244..
01:54:25.624472 IP 192.168.150.254.1077 > 192.168.150.2.9136: P
   71:181(110) ack
1864 win 64009
        0x0000:  0050 5601 0203 000c 29d5 7e15 0800 4500
   .PV.....).~...E.
        0x0010:  0096 27be 4000 8006 2452 c0a8 96fe c0a8  ..'.@...
   $R......
        0x0020:  9602 0435 23b0 62f8 5e01 96e5 0a1a
   5018  ...5#.b.^.....P.
        0x0030:  fa09 273e 0000 4d4f 4445 204e 6554
   787c  ..'>..MODE.NeTx|
        0x0040:  3836 3032 3434 202b 782b 690d 0a4a 4f49
   860244.+x+i..JOI
        0x0050:  4e20 2323 4e65 5478 2323 2077 6179 6e65
   N.##NeTx##.wayne
        0x0060:  0d0a 5553 4552 484f 5354 204e 6554 787c
   ..USERHOST.NeTx|
        0x0070:  3836 3032 3434 0d0a 4d4f 4445 204e 6554
   860244..MODE.NeT
        0x0080:  787c 3836 3032 3434 202b 782b 690d 0a4a  x|
   860244 +x+i..J
        0x0090:  4f49 4e20 2323 4e65 5478 2323 2077 6179
```

IRC information (II)

- Which is the hardcoded name of the bot:
  - dad.darksensui.info
- Port used for connections: 9136
- IRC channel and password: ##NeTX## wayne
- This is enough to connect to the IRC channel and listen to the bots, but what is the password for managing the “bots” ?

Two approach

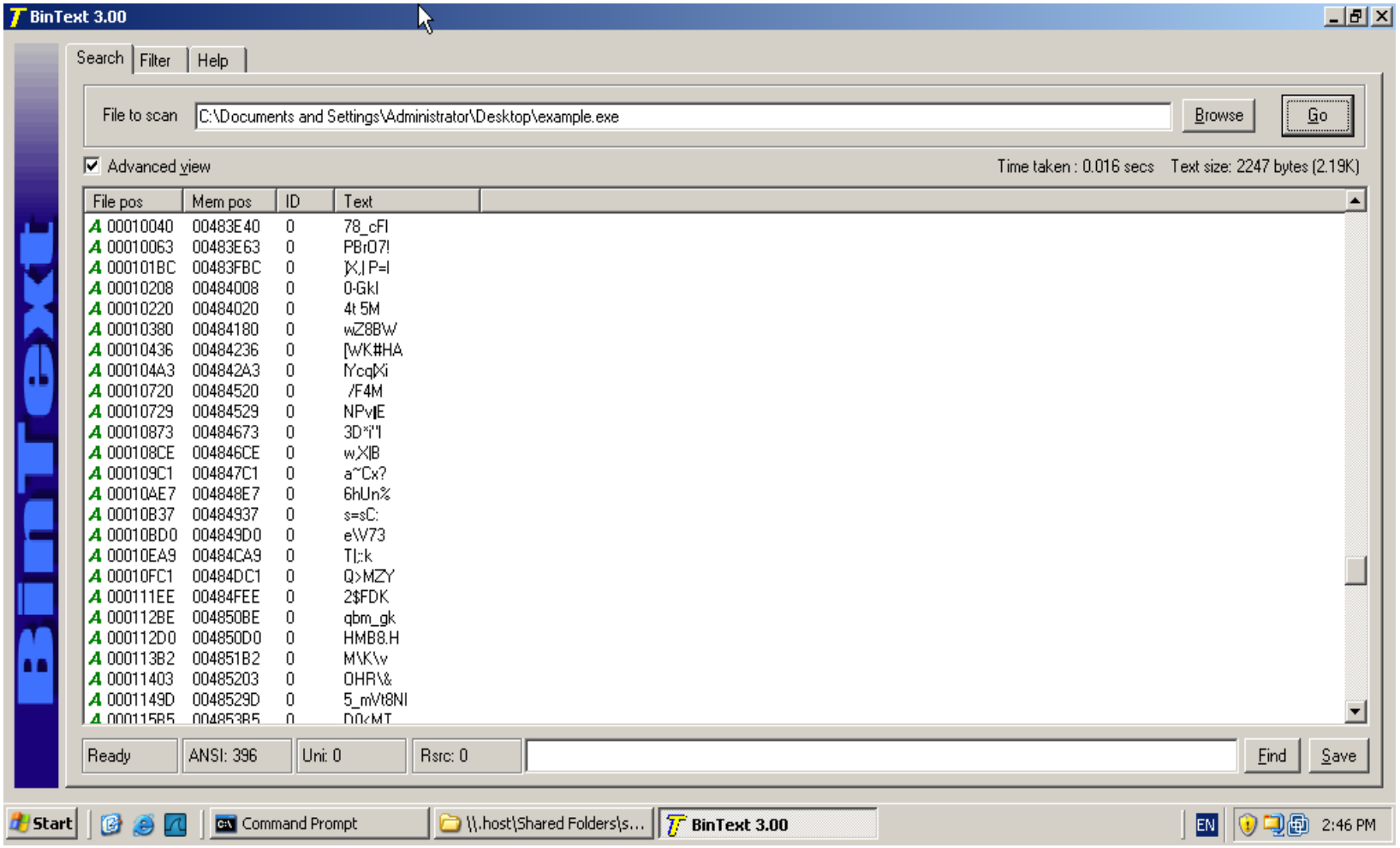
- Connect to the botnet and simulate be a client with a irc client
- Wait until the owner of the bots connects and type the password .
- Problems:
  - Are you allowed to do this ?
  - What happens if they detect you ?

- We need to revert to reverse engineering tools

Packed files

- Most the malware is encrypted / compressed
  - Most times with more than one layer
  - With different compressor at the same time
- The result file is difficult to analyze with an static disassembler and the “strings” commands show no information .
- Fortunately most of the bots code can be saved uncompressed to the disk when the bot is running

Looking at the strings with bintext

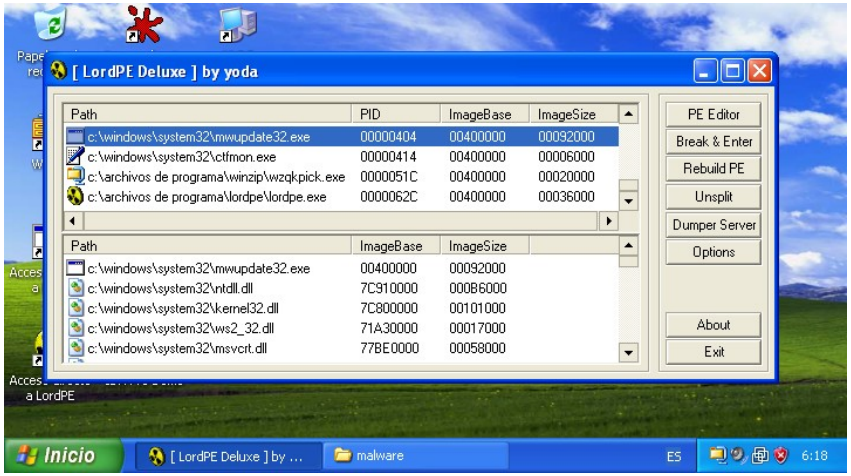


Running and encrypted file

- Normally the bot is compiled without any encryption and the miscreant uses external tools (like upx) to generate the file.
- When the file is run, the program decrypt itself in memory and the normal program is executed.
- There are some tools to dump the program memory and write unencrypted file.
  - LordPE , PeDump ...
  - Ollydbg dump plugin

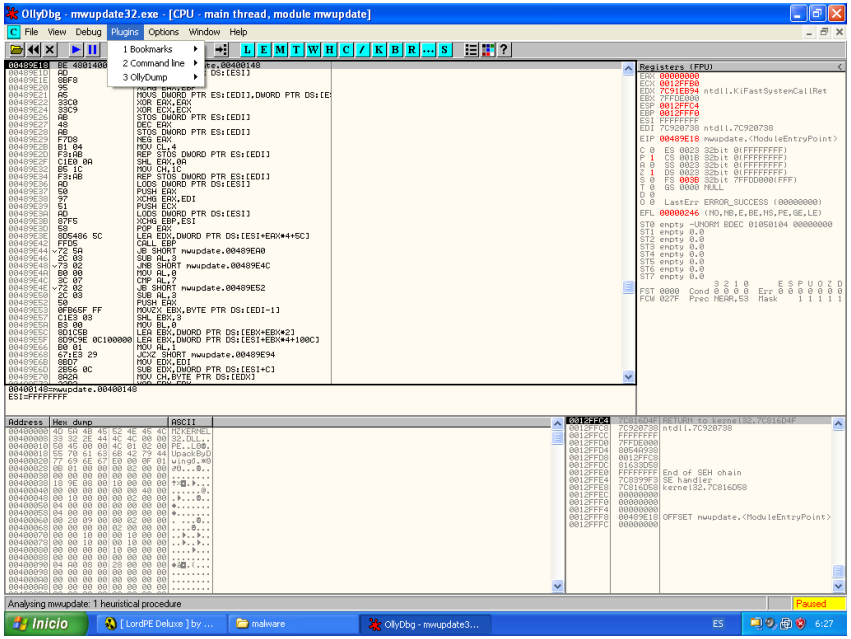
Using Lord PE

- Execute the malware.
- Launch Lord PE and select the process to dump.
- Righ click in the process and choose full dump.
- Save the file
- That's all

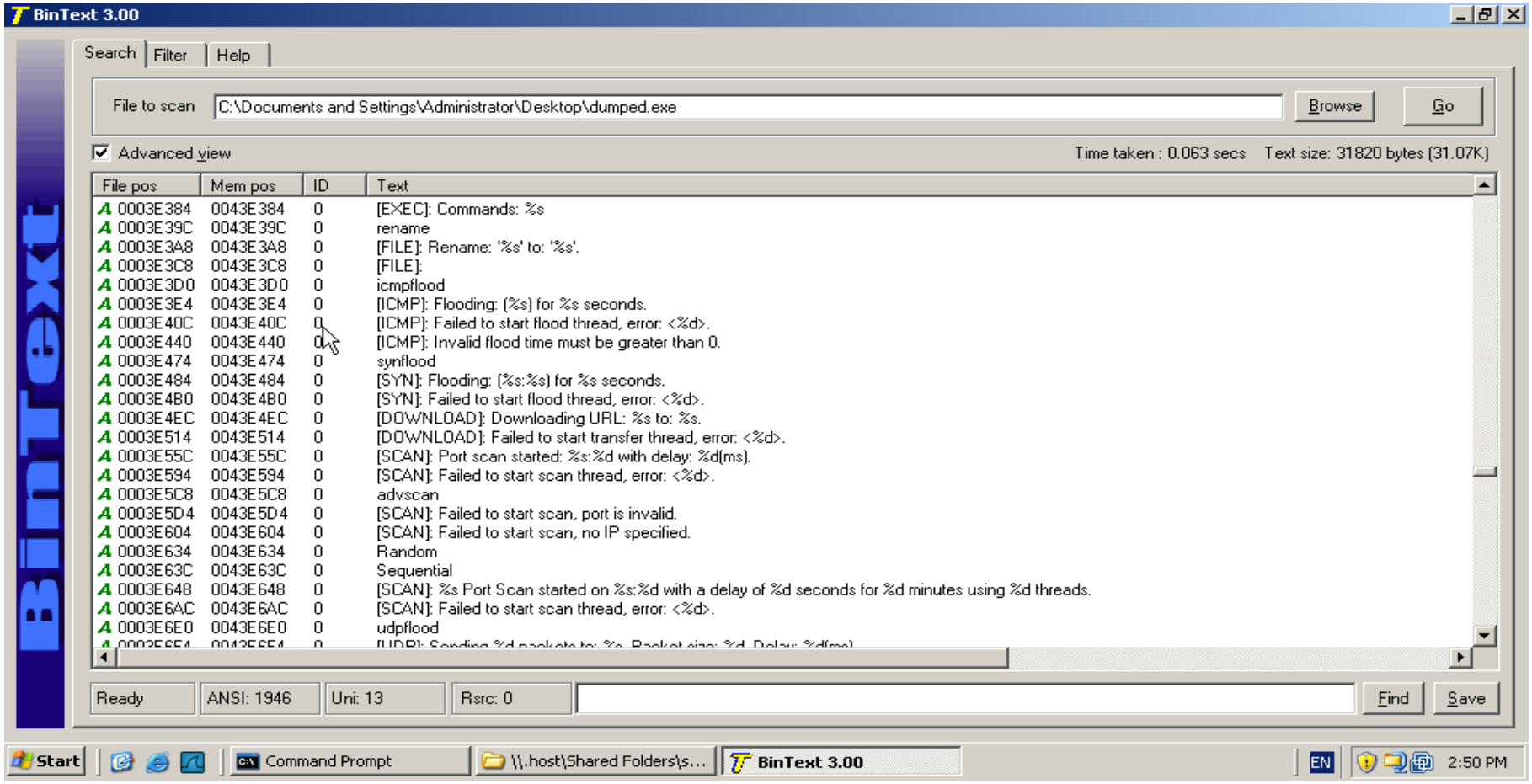


Using Ollydbg dump plugin

- Attach to the process.
- Launch Ollydump plugin
- Save the file ..



Bintext II


BinText 3.00
Search
Filter
Help
File to scan
C:\Documents and Settings\Administrator\Desktop\dumped.exe
Browse
Go
Advanced view
Time taken : 0.063 secs
Text size: 31820 bytes (31.07K)
File pos
Mem pos
ID
Text
0003E384 0043E384 0 [EXEC]: Commands: %s
0003E39C 0043E39C 0 rename
0003E3A8 0043E3A8 0 [FILE]: Rename: '%s' to: '%s'.
0003E3C8 0043E3C8 0 [FILE]:
0003E3D0 0043E3D0 0 icmpflood
0003E3E4 0043E3E4 0 [ICMP]: Flooding: (%s) for %s seconds.
0003E40C 0043E40C 0 [ICMP]: Failed to start flood thread, error: <%d>.
0003E440 0043E440 0 [ICMP]: Invalid flood time must be greater than 0.
0003E474 0043E474 0 synflood
0003E484 0043E484 0 [SYN]: Flooding: (%s:%s) for %s seconds.
0003E4B0 0043E4B0 0 [SYN]: Failed to start flood thread, error: <%d>.
0003E4EC 0043E4EC 0 [DOWNLOAD]: Downloading URL: %s to: %s.
0003E514 0043E514 0 [DOWNLOAD]: Failed to start transfer thread, error: <%d>.
0003E55C 0043E55C 0 [SCAN]: Port scan started: %s:%d with delay: %d(ms).
0003E594 0043E594 0 [SCAN]: Failed to start scan thread, error: <%d>.
0003E5C8 0043E5C8 0 advscan
0003E5D4 0043E5D4 0 [SCAN]: Failed to start scan, port is invalid.
0003E604 0043E604 0 [SCAN]: Failed to start scan, no IP specified.
0003E634 0043E634 0 Random
0003E63C 0043E63C 0 Sequential
0003E648 0043E648 0 [SCAN]: %s Port Scan started on %s:%d with a delay of %d seconds for %d minutes using %d threads.
0003E6AC 0043E6AC 0 [SCAN]: Failed to start scan thread, error: <%d>.
0003E6E0 0043E6E0 0 udpflood
Ready
ANSI: 1946
Uni: 13
Rsrc: 0
Find
Save
Start
Command Prompt
\\.host\Shared Folders\s...
BinText 3.00
EN
2:50 PM

Reading disassembly code

- After dumping the file this should be “readable”, you can start searching for strings
-
- Most of the times the file is not executable, because some information is missing.
- But you can disassembly the malware and analyze it.

- Typical C function call:
  - Printf ("hello %s\n" , somename);
- Somename is a *char ;-)
- Subtitute %s by the string in somename and print it

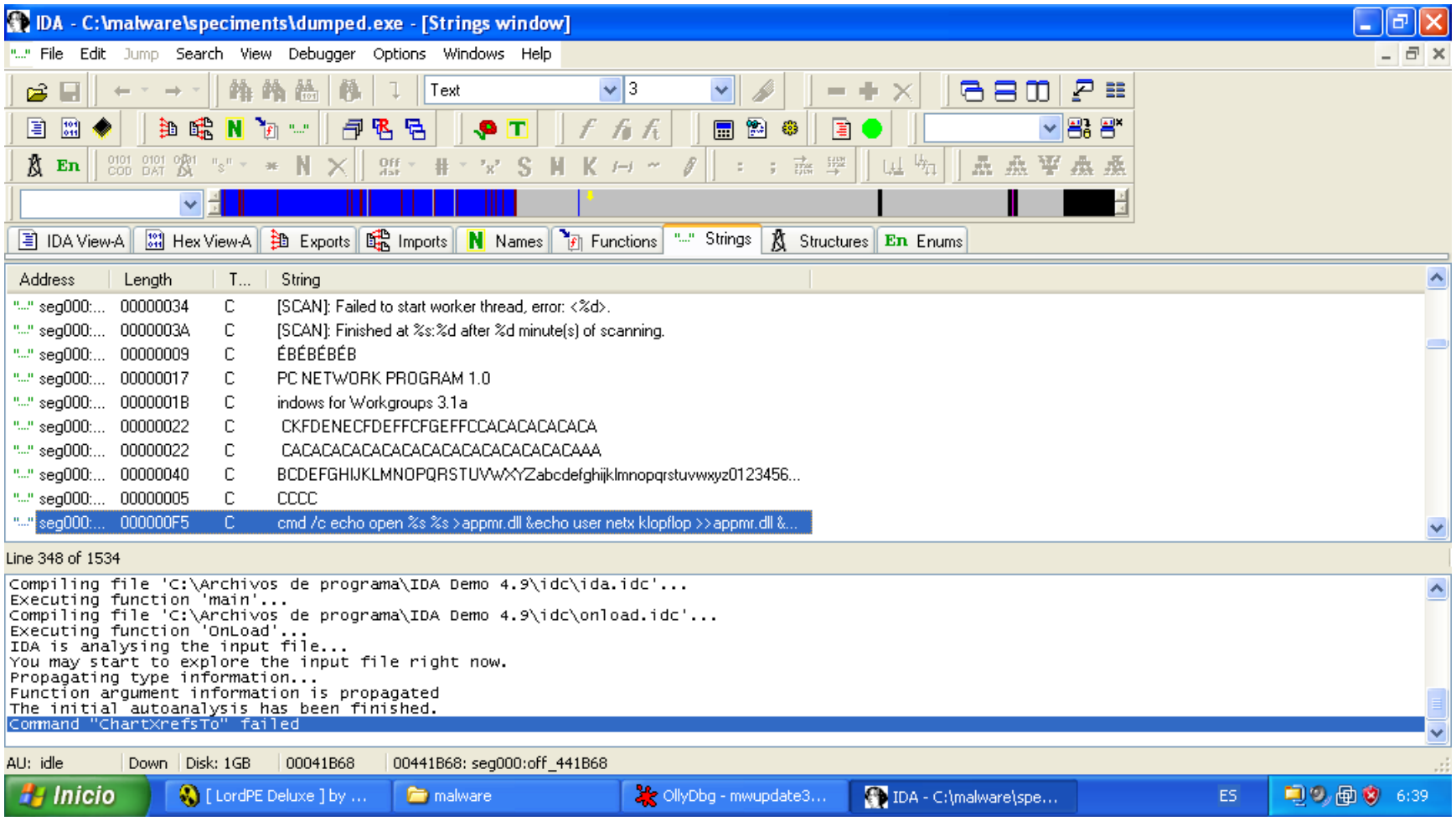
It's translated into asm as:

1. Push reference to somename in the stack
2. Push reference to "hello %s\n" in the stack
3. Call/execute printf function

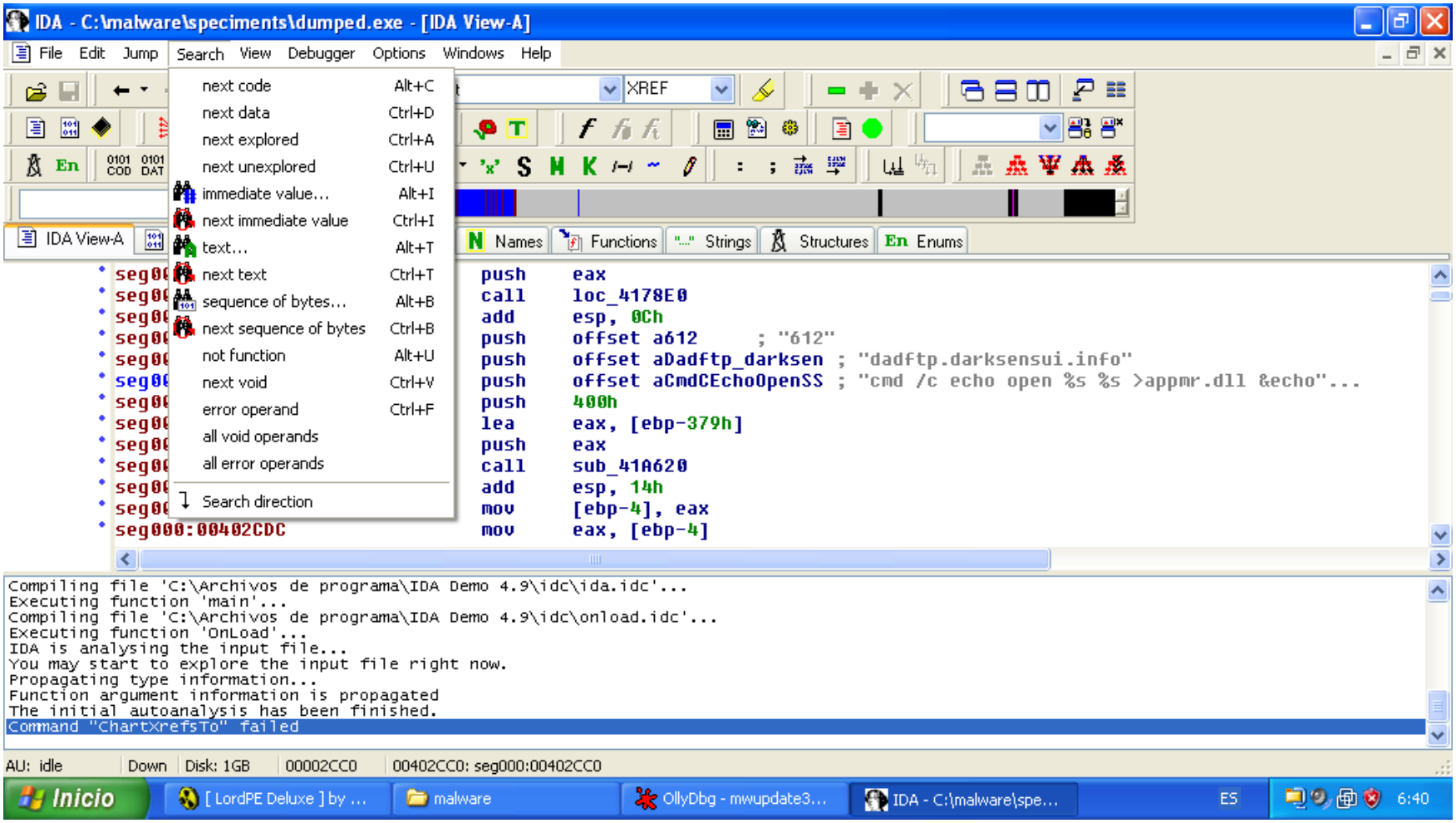
Note: the right to left order

- http://www.datarescue.com/idabase
- Commercial tools there is a freeware version that can be analyze only x86 binaries.
- Time-limited version available in the web
- There is a lot of plug-ins that help with the disassembly.

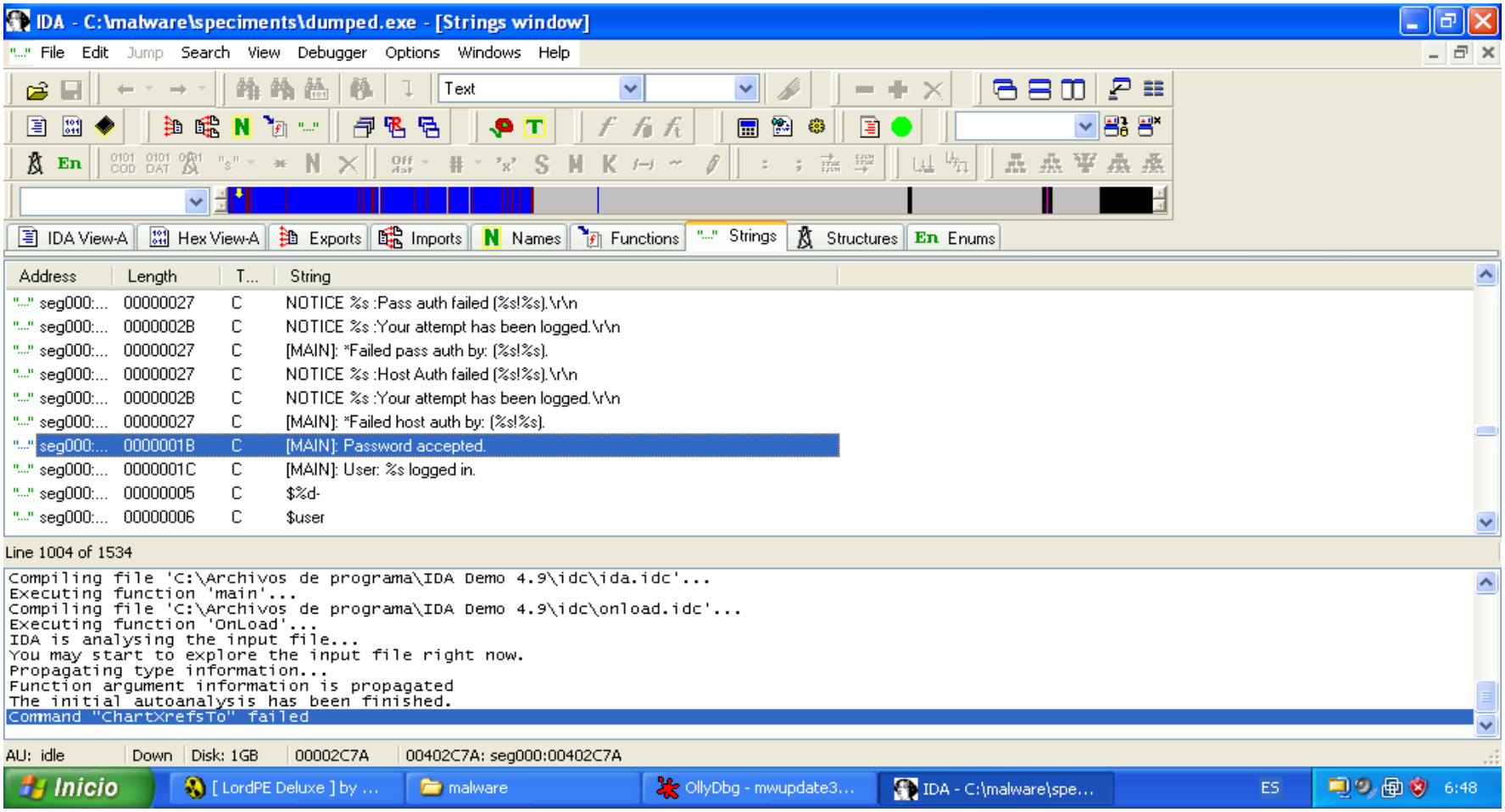
Where the malware comes from ?



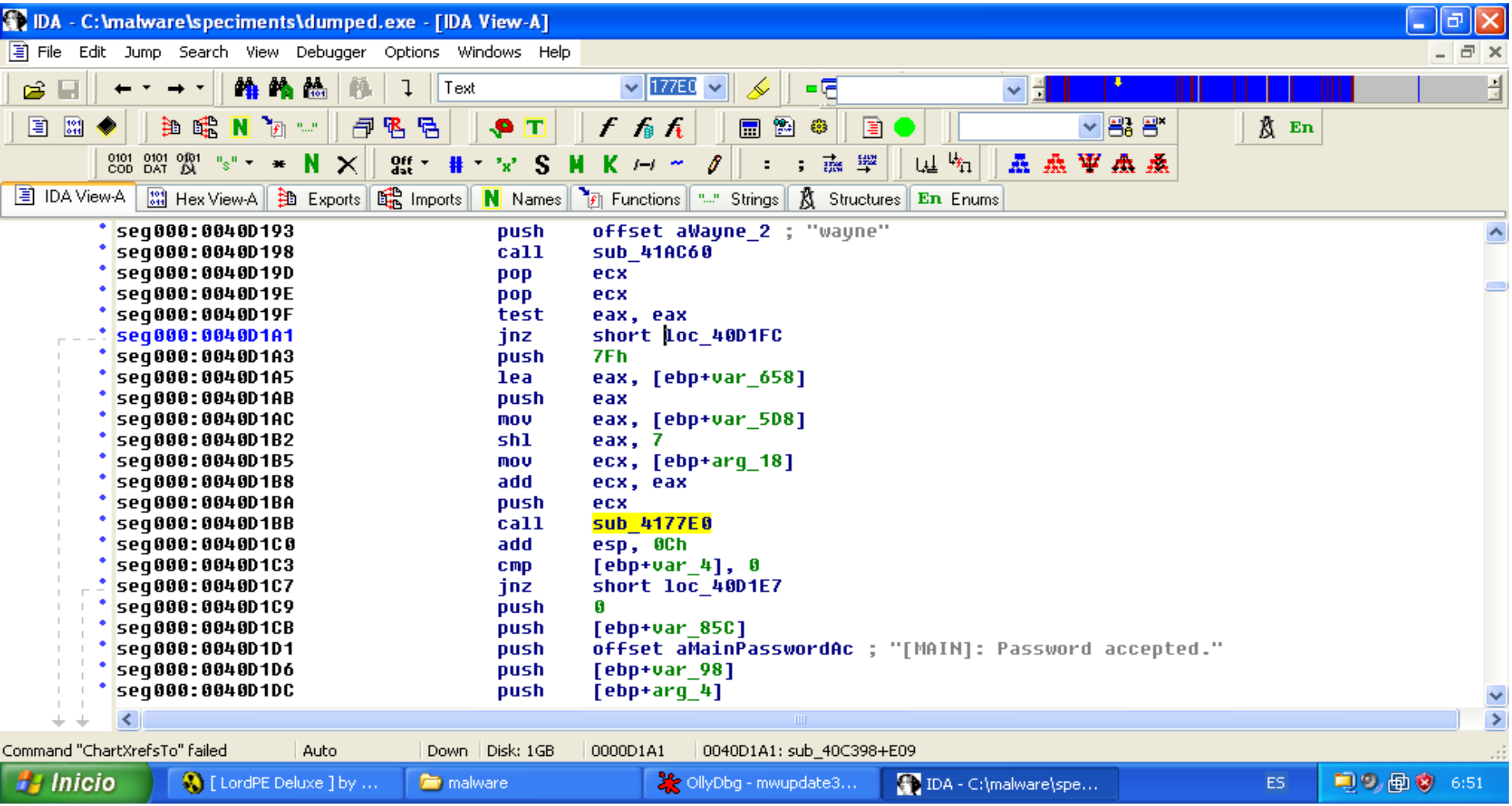
Where the malware comes from ? (II)



Finding the password



Finding the password



Problems

- Malware can detect that it's running virtual machines
  - Try with different virtualization software
  - Use real machines connected to the virtual lab
  - Malware can't be not be easily analized
    - Different encryptation levels
    - Ofuscate strings & password
      - Better behavior analysis
      - Disassembly

- Artifact Analysis, Kevin Hole (CERT/CC), http://www.first.org/resources/papers/con html and all the FIRST.org conference PAPERS (see
- SANS Malware Course , http://WWW.sans.org
- Nepenthes, http://www.mwcollect.org

- Google….

Conclusions ...

- To solve most of the Security incidents we need to get information about the malware.
- A minimal lab can be build with few resources.
- Most of the tools and information can be found free in Internet

Edificio Bronce
Plaza Manuel Gómez Moreno s/n
28020 Madrid. España

Tel.: 91 212 76 20 / 25
Fax: 91 212 76
www.red.es