FORENSICS CYBER-SECURITY

MEIC, METI

2021/2022 1st Semester

1st Exam November 26, 2021 Duration: 2h00

- Use a pen only; no extra material is allowed, such as calculator, scratch paper, etc.
- Write your answers in the free space after each question.
- The exam can be answered in Portuguese or in English.
- Identify all sheets; **unidentified pages will not be graded!**

I. (1 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 4 points)

1. "Strict adherence to the Kruse model ensures that all the evidence collected by a forensic analyst will be considered admissible in court." Do you agree with this statement? Justify your answer.

2. When collecting forensic data from a hardware device, volatility is an essential aspect to consider:

   a. What are the two factors that mostly affect the level of data volatility?

   b. Consider two forensic tools t1 and t2. Both of them can collect the following information of the runtime state of a process: (1) stack memory dump, (2) the operating system's buffer cache that caches blocks of the files opened by the process, (3) a snapshot of the CPU registers. However, these tools collect this information in a different order as shown below (left is collected first). In principle, which tool would you select to analyze a process and why?

   $$t1 : (1) \rightarrow (2) \rightarrow (3) \qquad t2 : (3) \rightarrow (1) \rightarrow (2)$$

3. The following investigative scenario was conducted in two steps, separated in time. For each step, indicate if you agree with the presented statements. Justify your responses:

   STEP 1: A small bank was the victim of a cyber-attack leading to the exfiltration of highly-sensitive data from their online banking system. The forensic analyst, a member of the bank's IT staff, was called to collect digital evidence of the event. This evidence will serve two purposes: (1) produce a forensic auditing report proving that the attack took place, and (2) help police authorities to further investigate the author of the attack. The online banking system runs on a set of servers and it follows a classic 3-tier web service architecture. The forensic analyst did not shut down any of the system's servers. He collected only the following digital artifacts: (1) web server logs, (2) application logs, (3) NetFlow logs, and (4) IDS event logs. The collected log samples span 24 hours before the attack. He then computed the MD5 hashes of all log samples and stashed several copies of the log samples.

   a. "The forensic analyst acted legally because he was authorized by the bank."

   b. "The forensic analyst should have created bit-stream images of the servers' hard drives."

   STEP 2: After analyzing the logs, the forensic analyst found evidence that an exploited vulnerability in the banking system led to the data exfiltration. Records in the web server logs indicate the source of the attack coming from a residential IP address (i.e., assigned from an ISP to a homeowner and associated with a single owner and location.) The bank decided to press charges to the police who then initiated a criminal investigation. The IP address points to the residence of Abby Chapman. The police searched her house with a second forensic analyst – a police officer – who found two pieces of equipment: a Windows workstation and a mobile phone. The workstation was powered and logged into Abby's account. The mobile phone was locked (i.e., password protected) and the screen was wiped. The officer realized that this specific phone model was being sold on eBay and was rooted.

   c. "The correct way to preserve evidence from the Windows workstation is to reboot the computer and create a bit-stream image of the hard drive."

   d. "There is not much we can do to extract data from from the mobile phone."

II. (1 + 1 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 1 = 8 points)

1. Alice and Bob want to exchange a covert secret message M inside an image and agree to employ 2-bit LSB encoding. To make the encoding more robust against bit corruption attacks, the algorithm first splits the secret message M into chunks c_i of 4 bits each and appends one extra parity bit y_i to each chunk. The value of y_i is 1, if chunk c_i has an even number of bits set to 1; otherwise $y_i = 0$. For instance, if chunk $c_0 =$ `1011`, then $y_0 = 0$, hence the encoded sequence will be $c_0y_0 =$ `10110`. If M contains multiple n chunks, the algorithm encodes each c_iy_i sequentially, i.e., $M' = \langle c_0y_0c_1y_1...c_{n-1}y_{n-1}\rangle$. The size of M is always a multiple of 4 bits. A password P can be provided to select the pixels embedding the covert bits of M':
   - the first pixel (p_0) to encode bits is given by: $p_0 = P$ *mod* 7;
   - the next pixels are found by: $p_i = p_{i-1} + 1$;
   - pixel numbering begins in 0, i.e., the index of the first pixel of an image is 0.

   Answer the following questions, justifying them properly:

   a. Considering that Alice wants to send M = `1001 0111` (binary) using a shared password P = 12 (decimal), indicate which pixels of the following greyscale 16-bit bitmap image (the hex dump) will be modified by the algorithm and what will be the pixels' new values.

```
00: 5bae 09e0 237f 1684 8203 ab85 61a0 0a15
10: 464a 5de2 672e 3f40 6c36 48ef 5e2c d53f
20: 7d3a 0995 c8d0 ae3f 2f33 64cd 09fe a1b4
30: 14be 8629 8fe8 b576 f1ac d993 70c7 3f13
40: .... .... .... .... .... .... .... ....
```

   b. If Mallory intercepts the stego image and adds noise as described next, will the secret message M be corrupted? If so, will Bob be able to detect that M has been corrupted?

      For every pixel p_i such that p_i *mod* $7 = 1$, set to 0 the least significant bit of p_i's value.

2. Explain the purpose of the following command and indicate one limitation of this technique.

```
dc3dd if=/dev/fmem of=file.dd bs=1MB count=1024 hash=md5 log=/var/l.log
```

3. File `windisk.dd` contains a bitstream image of a partition formatted to the NTFS file system. The fls command executed below shows a trimmed list of deleted files (* marked) located in that partition. When compared to analyzing Ext$x$ file systems, the identifier that precedes the name of each file has a different format, namely x-`128`-y, where `128` corresponds to the attribute number in the NTFS specification that identifies the $DATA file attribute.

```
sherlock@forensic:~/$ fls -o 2048 -Frd windisk.dd
-/r *  40-128-1: Users/Albert/Documents/Credit.pdf
-/r *  40-128-2: Users/Albert/Documents/Credit.pdf:Zone.Identifier
-/r * 236-128-1: Users/Albert/Documents/ManProj/Project.docx
-/r * 239-128-1: Users/Albert/Documents/ManProj/manhattan_project.zip
-/r * 220-128-1: Users/Albert/Pictures/Tails/GemoTailUniversal.jpg
-/r *  12-128-1: Windows/Prefetch/EXPLORER.EXE-A80E4F97.pf
...
```

Answer the following questions:

a. In NTFS, what is the most important data structure that stores file matadata?

b. Suggest a possible meaning for the value of x. Justify your answer.

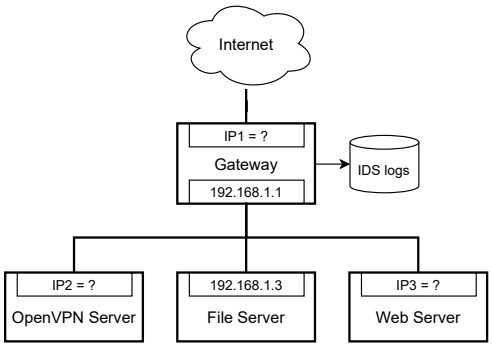
c. Suggest a possible meaning for the value y. Justify your answer.

d. What is the forensic value of file `Windows/Prefetch/EXPLORER.EXE-A80E4F97.pf`?

4. A storage partition contains multiple deleted image files: JPEG or KTTY. The files are not fragmented. However, when running a properly configured, single-pass structure-based file carver, some recovered files were still corrupted. Can you suggest an explanation for this?
   * JPEG: "`0xFF 0xD8`" header and "`0xFF 0xD9`" footer
   * KTTY: "`0xA1 0xD8`" header and "`0xFF 0xD8`" footer

5. Your job is to locate a rogue wireless access point. Indicate two ways by which the inspection of the MAC addresses in the 802.11 frames can help you to track the device.

6. A forensic investigator was called to analyze a data breach in a company whose network and server infrastructure is represented below: there is a gateway, a VPN server, a file server, and a web server. Files were stolen from the file server and posted on an public wiki (IP:34.5.63.32).



* The website's DNS name is "www.chickenking.org"
* The gateway has a single public IP address
* The gateway implements NAT
* The gateway has two port forwarding rules:

  R1: UDP port 1194 → 193.168.1.14

  R2: TCP port 80 → 193.168.1.8
* The gateway runs an IDS monitoring the internal LAN

Answer the following questions. Justify all your answers.

a. Identify the missing IP addresses in the network interfaces of the figure using the output of the command executed by the investigator: `nslookup www.chickenking.org`

```
Name: www.chickenking.org
Address: 207.148.248.143
```

b. What source IP address would the wiki server observe, if the documents had been posted by someone accessing the file server from home (IP:x.x.x.x) through the VPN server?

c. The IDS log contained numerous alerts triggered by the snort rule represented below. Are these logs helpful to help the investigator determine the cause of the breach? Justify.

```
alert tcp any any -> any any (flags:S; msg:"NMAP TCP SYN!";)
```

7. A company needs to prepare for future network forensic audits. It maintains a web server infrastructure serving nearly 100 million requests per day. Which monitoring device would you recommend and why? You have two options: a packet sniffer or a NetFlow logger.

III. (2 + 0.5 + 0.5 + 0.5 + 0.5 + 1 + 0.5 + 0.5 + 0.5 + 1 + 0.5 = 8 points)

1. For each of the following statements, indicate whether it is true (T) or false (F). Each correct answer is awarded 0.25 points; each wrong answer is penalized by subtracting 0.10 points.

   a. ____: Spoofers use open relays to hide the MAC address, the IP address, and the email address of the real sender.

   b. ____: The Google search engine provides an advanced command to specifically locate Internet webcams and automatically detect if they are unprotected.

   c. ____: On the Internet, it is hard to be anonymous because individuals tend to use the same password across many different services.

   d. ____: Dynamic analysis is a better technique to analyze malware than static analysis.

   e. ____: The IMEI of an Android device cannot be modified by rooting the device.

   f. ____: Smudge attacks are advanced hardware-based physical techniques for the extraction of data from mobile devices.

   g. ____: Repackaging attacks is a widely used practice to deploy Trojan horses in servers.

   h. ____: Multi-tenancy and geographical data dispersion raise difficult obstacles to the extraction of forensic data from the cloud.

2. Indicate two types of useful information that can be extracted from email headers.

3. The police are investigating an Onion Service (OS) that sells illicit goods on the Dark Web. The OS hides a website behind the onion address: `http://33y6fjyhs3phzfjj.onion/`. The goal of the police is to deanonymize the real IP address of the website. A police investigator started by directly accessing the website through the Tor browser at that specific onion address. Answer the questions below. Justify them. (Unjustified answers get 0 points.)

   a. What is the destination IP address of the packets sent from the investigator's computer when he submits HTTP requests to the destination website from the Tor browser?

   b. After a few interactions with the website, the investigator learned that it was implemented in PHP and managed to identify a vulnerability. He confirmed this vulnerability by crafting the following URL and sending the exploit to the website. What is the expected result of this request? By what name is this type of vulnerability known?

      ```
      http://33y6fjyhs3phzfjj.onion/?parameter=value;system('ls -l');
      ```

   c. How can this vulnerability help us deanonymize the real IP address of the website?

4. The architecture of early rootkits was based on system-call interposition. Give an example of what a rootkit may be attempting to do by intercepting the `open()` system call.

5. Consider the following scenario and answer the questions bellow:

   A cluster hosts 10 servers that provide an online service on the Internet. Each server has an independent, statically-assigned public IP address. All these servers are configured with the same software stack, but 5 run inside virtual machines (Group A) and the other 5 on bare metal, i.e., on physical machines (Group B). At some point, the network administrator detected abnormal behavior from the servers in group B, where each server occasionally attempted to contact remote destinations on various IP addresses. Servers in group A, however, did not manifest this behavior. By consulting the latest vulnerability reports, he saw that a zero-day vulnerability had been discovered affecting the OS that was being used in the cluster.

   Answer the following questions:

   a. Suggest a hypothesis that explains the behavior of the servers in group B.

   b. Suggest a hypothesis that explains why the servers in group A are not behaving in the same way as the servers in group B.

   c. Describe your next steps to test your hypothesis for the behavior of both server groups?

6. Bonnie and Clyde want to acquire two assets transacted online in Bitcoin: A1 and A2, respectively. The BTC price of A1 is worth 0.1 units and A2 is worth 0.2 units. Bonnie has 0.8 BTC in her account and Clyde 0.4 BTC in his. However, they want to buy these goods while creating reasonable doubt about who bought each asset. Respond to the following questions:

   a. How can they proceed if they are explicitly concerned about defeating transaction graph analysis using shared-spending heuristic?

   b. How does shared-spending heuristic compare to the fresh-change-address heuristic?