INSTITUTO SUPERIOR TÉCNICO

Departamento de Engenharia Informática

Forensics Cyber Security

MEIC / METI 2018-2019 – 1st Semester

Digital Forensics Report

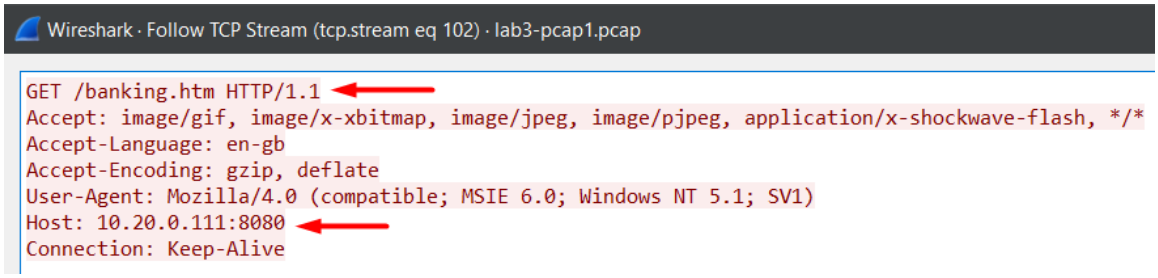
Authors

GROUP I – Network trace analysis

1

1.1 Indicate the complete URL of the original web request that led to the client being compromised

http://10.20.0.111:8080/banking.htm



1.2 What file type was requested in the final web request to the malicious server?

GIF

1.3 What is the number of the first frame that indicates that the client has been compromised?

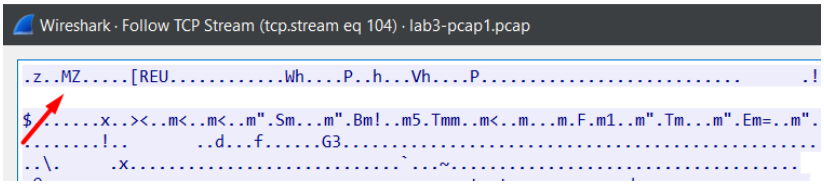
4722

Upon analyzing the existing TCP streams in the packet capture we found TCP stream 104 which contained what we thought to be the packets containing the virus that would compromise the client. Its first frame was the one referenced above, 4722.

1.4 At one point, the malicious server sends a malicious file to the client. What type of file it is?

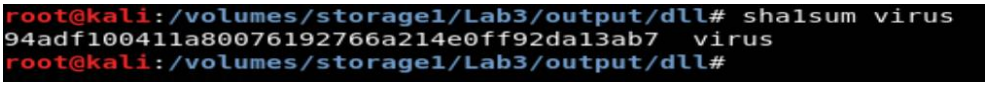
Windows Executable

After analyzing the TCP stream 104 we figured its start was the magic number corresponding to a windows executable file.



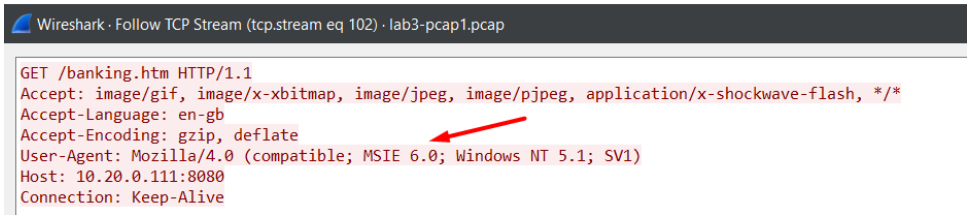
1.5 What is the SHA1 hash of the malicious file?

94adf100411a80076192766a214e0ff92da13ab7



1.6 What vulnerable software has been exploited (in the following format, Firefox 3.5, Firefox 3.6, Word 2010, IE7, Safari2, Chrome2, AdobeReader, IE9)?

MSIE 6.0



1.7 When the capture ends, is the client still connected to the malicious attacker?

Yes, because there is no packet that indicates the end of the connection, i.e., no FIN or RST packet.

1.8 Can you indicate the corresponding CVE security bulletin that covers the vulnerability that was exploited here (answer in form of CVE-$year-$number)?

CVE-2010-0249

Upon searching the given options, we narrowed them upon two that pertained to MSIE 6.0, then we searched what each one did and found out that CVE-2010-1127 was a Denial of Service attack and CVE-2010-0249 was Execute Code. With that in mind we had our answer.

1.9 From the capture, it is clear that the attacker gets a certain form of access (i.e. the interface), what (type of) access does the attacker "get" on the client?

The access is being able to execute arbitrary code as described in the CVE security bulletin (CVE-2010-0249).

FILE	MD5
banking.htmTysdAWdqQEBybyCGKQkGJyVuQsNWvmIFg.gif	df3e567d6f16d040326c7a0ea29a4f41
banking.htm?UOjiXfyAbAISuH	432fa43eddeddd9e72f370fb6d964cc9
exploit.js	4d58adce41a006a3a9326b6e756691a2
Virus	216c9d064601b4d2066f4573ac20d656

2

2.1 What is the frame that indicates that something strange might be going on?

8

Upon analyzing the capture, we saw an uncommon amount of ICMP frames coming from the same IP and with a destination IP that was incrementing. We thought this IP might be scanning the network for IP's that would answer him.

6	5.595503	10.20.0.110	192.168.10.11	SSH	90	Client: Encrypted packet (len=36)
7	5.696960	192.168.10.11	10.20.0.110	TCP	54	22 → 1280 [ACK] Seq=69 Ack=37 Win=65535 Len=0
8	15.019198	10.20.0.110	192.168.10.1	ICMP	42	Echo (ping) request id=0x692d, seq=0/0, ttl=43 (no response found!)
9	15.026311	10.20.0.110	192.168.10.4	ICMP	42	Echo (ping) request id=0x0bd0, seq=0/0, ttl=40 (no response found!)
10	15.026432	10.20.0.110	192.168.10.5	ICMP	42	Echo (ping) request id=0x9494, seq=0/0, ttl=50 (no response found!)
11	15.026504	10.20.0.110	192.168.10.6	ICMP	42	Echo (ping) request id=0x00ca, seq=0/0, ttl=41 (no response found!)
12	15.026569	10.20.0.110	192.168.10.7	ICMP	42	Echo (ping) request id=0x1c97, seq=0/0, ttl=44 (no response found!)
13	15.026632	10.20.0.110	192.168.10.8	ICMP	42	Echo (ping) request id=0x3c91, seq=0/0, ttl=54 (no response found!)
14	15.026696	10.20.0.110	192.168.10.9	ICMP	42	Echo (ping) request id=0xafb5, seq=0/0, ttl=51 (no response found!)
15	15.026758	10.20.0.110	192.168.10.10	ICMP	42	Echo (ping) request id=0x58ef, seq=0/0, ttl=38 (reply in 18)

2.2 What tool is generating this traffic?

Nmap

2.3 What does this frame constitute the beginning of?

Ping Scan

2.4 A switch was removed from the command to improve the speed, what was this switch (just the letters, case-sensitive)?

On the first attack, around 4-5 minutes of capture the attack scanned for ports first using TCP then using UDP. On the second attack around 11 minutes of capture the attack scanned for ports only using TCP. The switch that was removed was the UDP Scan one (-sU).

2.5 What switch was added to the final scan (case-sensitive)?

-A

When analyzing the last packets on the capture we found that on the last HTTP packets the user-agent field was Nmap Scripting Engine. We searched what it was and what I did and found that it did the following tasks:

We designed NSE to be versatile, with the following tasks in mind:

Network discovery

> This is Nmap's bread and butter. Examples include looking up whois data based on the target domain, querying ARIN, RIPE, or APNIC for the target IP to determine ownership, performing identd lookups on open ports, SNMP queries, and listing available NFS/SMB/RPC shares and services.

More sophisticated version detection

> The Nmap version detection system (Chapter 7, *Service and Application Version Detection*) is able to recognize thousands of different services through its probe and regular expression signature based matching system, but it cannot recognize everything. For example, identifying the Skype v2 service requires two independent probes, which version detection isn't flexible enough to handle. Nmap could also recognize more SNMP services if it tried a few hundred different community names by brute force. Neither of these tasks are well suited to traditional Nmap version detection, but both are easily accomplished with NSE. For these reasons, version detection now calls NSE by default to handle some tricky services. This is described in the section called "Version Detection Using NSE".

Vulnerability detection

> When a new vulnerability is discovered, you often want to scan your networks quickly to identify vulnerable systems before the bad guys do. While Nmap isn't a comprehensive vulnerability scanner, NSE is powerful enough to handle even demanding vulnerability checks. When the Heartbleed bug affected hundreds of thousands of systems worldwide, Nmap's developers responded with the `ssl-heartbleed` detection script within 2 days. Many vulnerability detection scripts are already available and we plan to distribute more as they are written.

Backdoor detection

> Many attackers and some automated worms leave backdoors to enable later reentry. Some of these can be detected by Nmap's regular expression based version detection, but more complex worms and backdoors require NSE's advanced capabilities to reliably detect. NSE has been used to detect the Double Pulsar NSA backdoor in SMB and backdoored versions of UnrealIRCd, vsftpd, and ProFTPd.

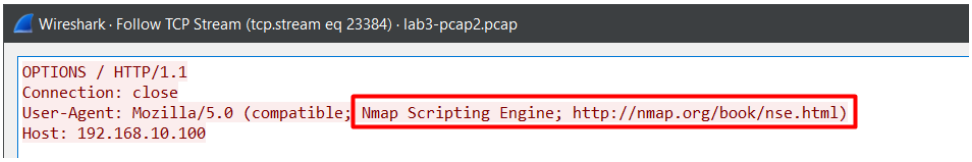
Vulnerability exploitation

> As a general scripting language, NSE can even be used to exploit vulnerabilities rather than just find them. The capability to add custom exploit scripts may be valuable for some people (particularly penetration testers), though we aren't planning to turn Nmap into an exploitation framework such as Metasploit.

These listed items were our initial goals, and we expect Nmap users to come up with even more inventive uses for NSE.

```
-A: Enable OS detection, version detection, script scanning, and traceroute
--datadir <dirname>: Specify custom Nmap data file location
--send-eth/--send-ip: Send using raw ethernet frames or IP packets
--privileged: Assume that the user is fully privileged
--unprivileged: Assume the user lacks raw socket privileges
```

With this in mind and a list of switches and general explanation of what they did we reached the conclusion that it should be the switch -A.



GROUP II – Website fingerprinting over Tor

1 Indicate: (a) the selected feature set, and (b) the accuracy of your classifier using this feature set. Give an account of other features you have tried before converging into this feature set. Explain how you've modified the source code in order to compute your proposed feature set.

O objetivo de alcançar 70% de accuracy foi atingido, mais precisamente foi atingido uma accuracy de 71,7%. Para isso foram utilizadas as seguintes features:

Num_pos: que é o numero de celulas na direção +1.

Num_total: que é o numero de células.

Time_neg: Que é calculado vendo quais as posições na direção -1 e calculando a diferença de times[i]-times[i-1]

Time_pos: Que é calculado vendo quais as posições na direção +1 e calculando a diferença de times[i]-times[i-1]

Time_total: Que é o tempo total de todas as células que é dado pelo times da ultima posição.

Max_diff: Que é a maior diferença entre duas células consecutivas

Per-neg: que é o numero de celulas na direção +1 a divider pelo numero total de celulas.

Per_time_neg: que é a soma das diferenças de tempo na direção +1 a dividir pelo tempo total

Per_num: que é o numero de celulas na direção -1 a dividir pelo numero de celulas na direção +1

Seq_pos_dois: que é o numero de vezes que houve sequencias de exatamente dois +1.

Seq_pos_tres: que é o numero de vezes que houve sequencias de exatamente tres+1

Seq_pos_quatro: que é o numero de vezes que houve sequencias de exatamente quatro +1

Seq_pos_cinco: que é o numero de vezes que houve sequencias de exatamente cinco +1

Seq_pos_seis: que é o numero de vezes que houve sequencias de exatamente seis +1

Seq_pos_sete: que é o numero de vezes que houve sequencias de exatamente sete +1

Media_indices: que é a soma dos indices das posições com a direção +1 a dividir pelo numero de vezes que +1 ocorre

Media_indices_metade: que é a soma dos indices das posições com a direção +1 a dividir pelo numero de vezes que +1 ocorre (mas apenas para metade do vetor)

Media_indices_quarto: que é a soma dos indices das posições com a direção +1 a dividir pelo numero de vezes que +1 ocorre (mas so para um quarto do vetor)

Desvio: que é o desvio para a media_indices

Desvio_metade: que é o desvio para a media_indices_metade

Media_ind_igual: que é a soma dos indices de posição com diff==0 (times[i]-times[i-1] a dividir pelo numero de vezes que esta diferença ocorre.

Desvio_igual: desvio para a media_ind_igual.

Media_seq_dois: soma dos indices de sequencias com exatamente dois +1 a dividir numero de vezes que ocorre

Media_seq_tres: soma dos indices de sequencias com exatamente tres +1 a dividir numero de vezes que ocorre

Media_seq_quatro: soma dos indices de sequencias com exatamente quatro +1 a dividir numero de vezes que ocorre

Num_igual: numero de vezes que passa para um numero igual em celulas consecutivas (+1 +1 ou -1 -1)

Num_muda: numero de vezes que passa para um numero diferente em celulas consecutivas (+1 -1 ou -1 +1)

Pos_pos: numero de vezes que passou de +1 para +1 em celulas consecutivas

Pos_neg: numero de vezes que passou de +1 para -1 em celulas consecutivas

Time_pos_pos: somas dos tempos das celulas que passaram de +1 para +1

Time_pos_neg: somas dos tempos das celulas que passaram de +1 para -1

Time_equal: numero de vezes que em duas celulas consecutivas não houve alteração do tempo ou seja, times[i]-times[i-1]==0

Para além destas features foram testadas as seguintes features:

Num_neg: numero de celulas na direção -1, nao foi considerada influente

Min_diff: a menor diferença entre celulas consecutivas, numa primeira fase percebeu-se que seria sempre 0 e quando foi testado para o valor minimo diferente de 0, nao teve alterações significativas

Max_pos e max_neg: a diferença maxima so para celulas na direção +1 e só para celulas -1 (nao significativo)

Byt_time e time_byt: time total/ size total e o contrário (não significativo)

Também foram testadas sequencias com direções negativas (nao significativo)

2

2.1 Can you spot a trend in the accuracy of the classifier as k increases? Explain the variations observed.

k	1	2	5	10	15
ACC	0,895	0.825	0,716	0,614	0,523

Table 1

Como se pode ver pela tabela 1 o aumento do número de K prejudice a accuracy do classificador, sendo para k=1 o valor que tem maior accuracy (0,895) isto acontece quando os dados possuem muito ruido e aí quando se procura entre vários Ks irá encontrar valores de ruido que prejudicam a accuracy. Sendo k=1, o classificador irá so se preocurar com a mais próxima havendo menos hipotese que um elemento errado consiga se intrometer na busca. Para um valor maior por exemplo k=15 a probabilidade de haver ruido entre estes 15 é consideravelmente maior produzindo assim uma accuracy menor.

2.2 Can you spot any trend in the TPR/FPR trade-off alongside the variation of k and nm? Justify.

nm%		1	2	5	10	15
0.1	**TPR:**	0,981	0,842	0,704	0,589	0,488
	FPR:	0,371	0,171	0,057	0,027	0,011
0.2	**TPR:**	0,972	0,819	0,773		
	FRP:	0,277	0,102	0,049		
0.4	**TPR:**	0,895				
	FPR:	0,224				
0.6	**TPR:**	0,751				
	FPR:	0,143				
0.8	**TPR:**	0,702		0,644		
	FPR:	0,089		0,012		

Table 2

Devido a um erro durante a execução dos testes os restantes resultados não foram apurados. Porém deu para perceber que os TPR tinham tendência a diminuir com o aumento do k e com o aumento do nm. Da mesma forma, os valores de FPR também têm tendência a diminuir com o aumento de k e de nm.

2.3 According to your results in Table 2, which configuration leads to a more successful attack? Do you think an attack with such a TPR / FPR trade-off may be effective in practice? Justify.

Na nossa opinião apesar do numero de TPR ser bastante alto o numero de FPR também é consideravelmente alto para valores baixos de k sendo por isso um risco usar este metodo. Para dados com pouco ruido seria um bom metodo a executar porém na pratica seria mais benefico o uso de um ataque closed-world