Jr PenTester-2--Content Discovery

WHAT I LEARNED

what a favicon is and how it can lead to which framework was/is being used

sometimes the sitemap.xml will reveal something about a web app, and going to a frameworks home page can reveal default credentials that can allow someone to log in as admin

i learned two new osint sources: wappalyze and that S3 buckets can be a source

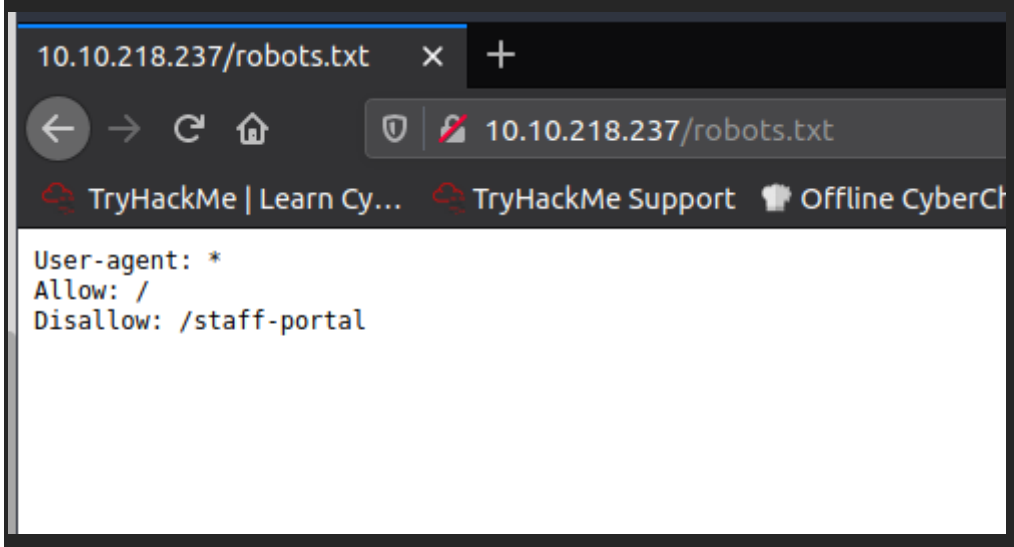
I also reviewed using gobuster and dirb, but also I was reminded of ffuf!!!

Task 1 What Is Content Discovery?

Content discovery: manual, automatic, osint (open source intel)

Task 2 Manual Discovery - Robots.txt

What is the directory in the robots.txt that isn't allowed to be viewed by web crawlers?



Task 3 Manual Discovery - Favicon

The favicon is a small icon displayed in the browser's address bar or tab used for branding a website.

Sometimes when frameworks are used to build a website, a favicon that is part of the installation gets leftover, and if the website developer doesn't replace this with a custom one, this can give us a clue on what framework is in use.

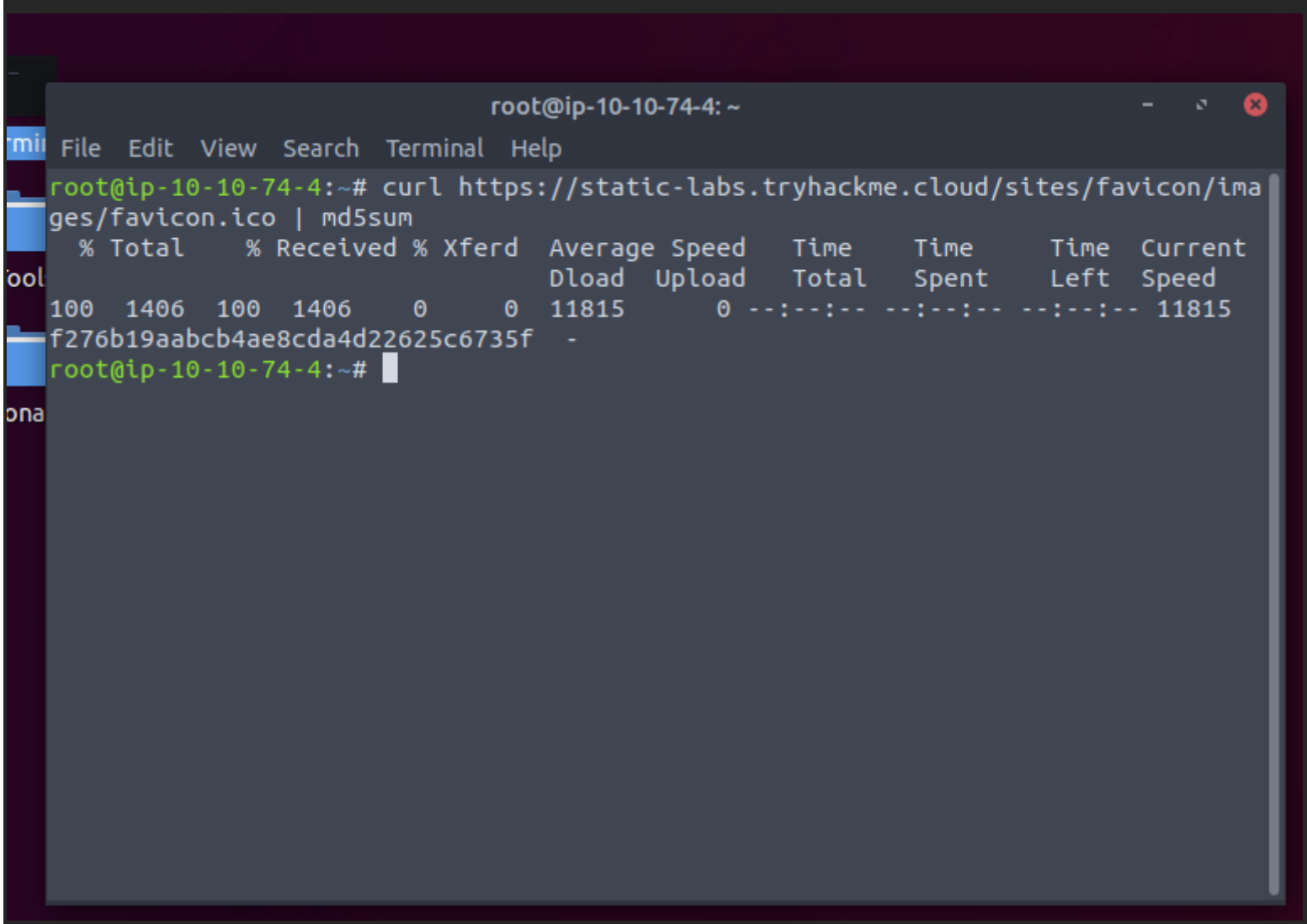
EXERCISE:

Viewing the page source you'll see line six contains a link to the images/favicon.ico file.

https://wiki.owasp.org/index.php/OWASP_favicon_database.

If you run the following command on the AttackBox, it will download the favicon and get its md5 hash value which you can then lookup on the

user@machine$ curl https://static-labs.tryhackme.cloud/sites/favicon/images/favicon.ico | md5sum



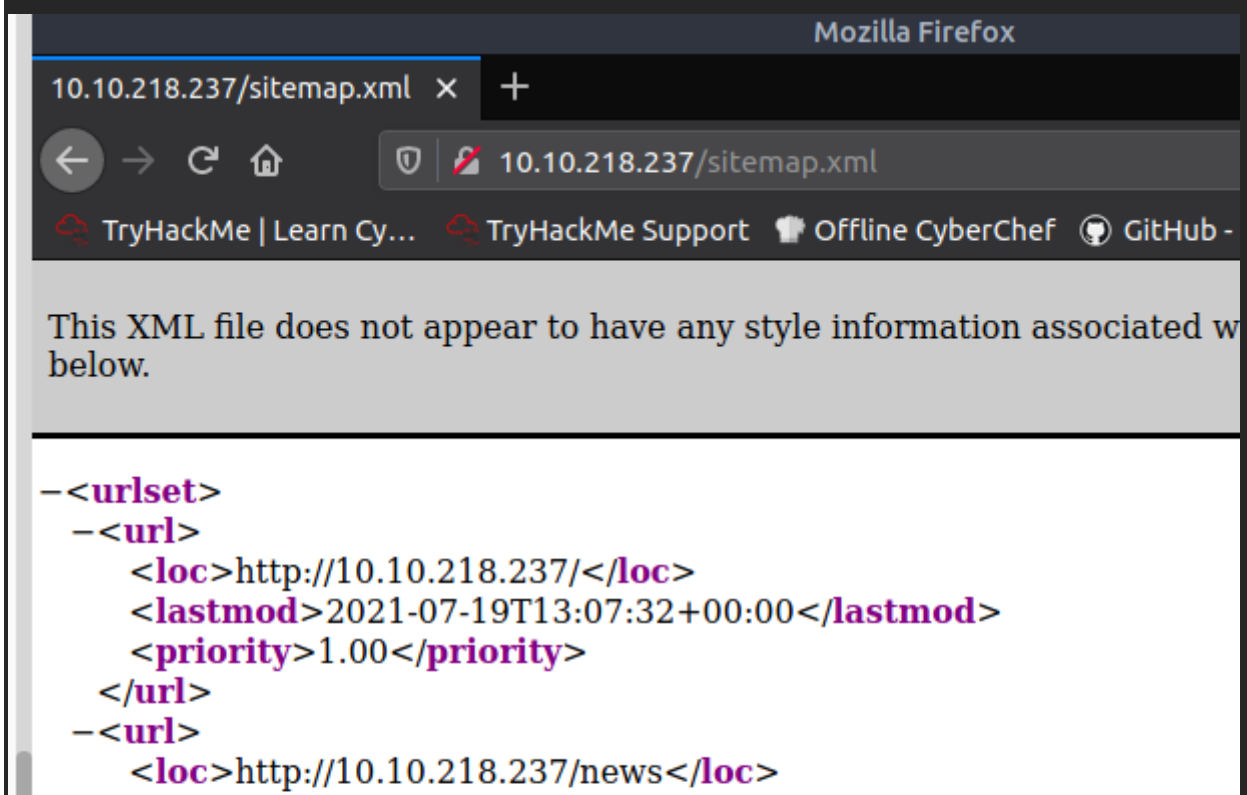
credientalsWhat framework did the favicon belong to?

```
f276b19aabcb4ae8cda4d22625c6735f:cgiirc (0.5.9)
```

```
140e3eb3e173bfb8d15778a578a213aa:bmpx (0.40.14)
4f12cccd3c42a4a478f067337fe92794:cacti (0.8.7b)
c0533ae5d0ed638ba3fb3485d8250a28:CakePHP (1.1.x)
66b3119d379aee26ba668fef49188dd3:cakephp (1.2.x-1.3x)
09f5ea65a2d31da8976b9b9fd2bf853c:caudium (1.4.12)
f276b19aabcb4ae8cda4d22625c6735f:cgiirc (0.5.9)
a18421fbf34123c03fb8b3082e9d33c8:chora2 (2.0.2)
23426658f03969934b758b7eb9e8f602:chronicle (2.9) theme-steve
75069c2c6701b2be250c05ec494b1b31:chronicle (2.9) theme-blog.
```

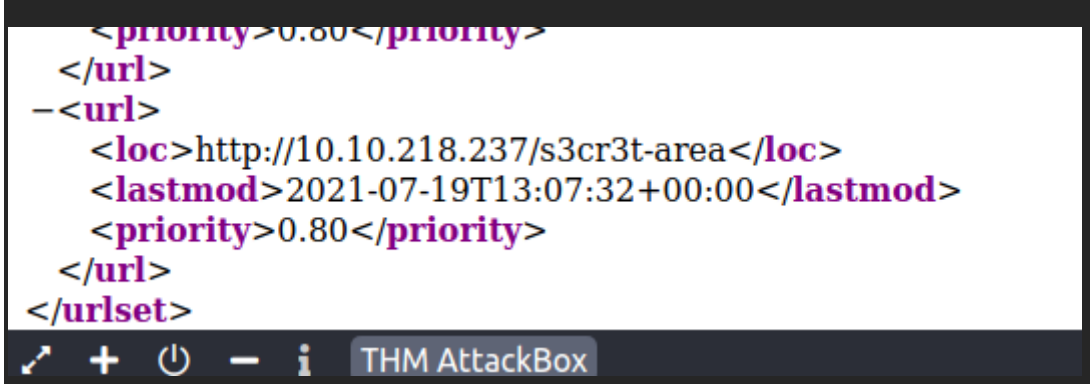
Task 4 Manual Discovery - Sitemap.xml

Unlike the robots.txt file, which restricts what search engine crawlers can look at, the sitemap.xml file gives a list of every file the website owner wishes to be listed on a search engine.



What is the path of the secret area that can be found in the sitemap.xml file?

http://10.10.218.237/s3cr3t-area



Task 5 Manual Discovery - HTTP Headers

HTTP Headers

When we make requests to the web server, the server returns various HTTP headers. These headers can sometimes contain useful information such as the webserver software and possibly the programming/scripting language in use. In the below example, we can see the webserver is NGINX version 1.18.0 and runs PHP version 7.4.3. Using this information, we could find vulnerable versions of software being used. Try running the below curl command against the web server, where the **-v** switch enables verbose mode, which will output the headers (there might be something interesting!).

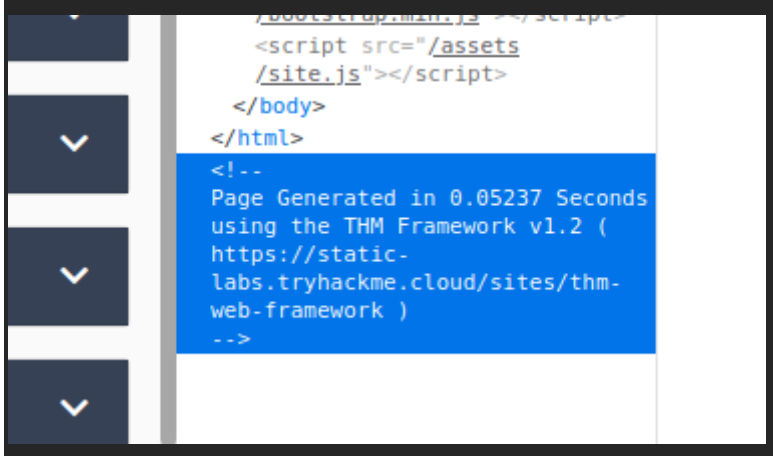
What is the flag value from the X-FLAG header?

--seems like this is just a made up header for ctf!!!

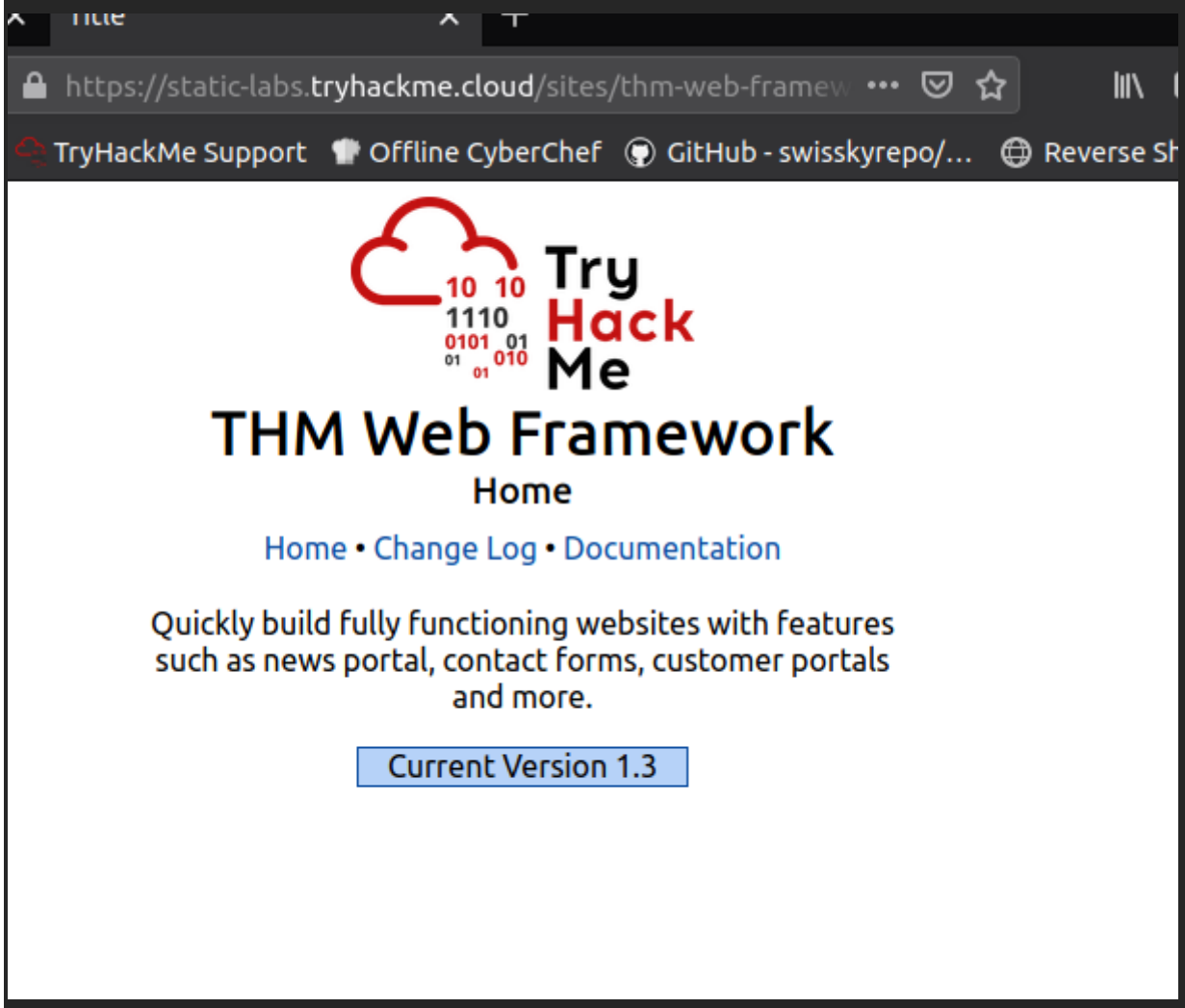
```
>
< HTTP/1.1 200 OK
< Server: nginx/1.18.0 (Ubuntu)
< Date: Wed, 26 Jan 2022 09:22:18 GMT
< Content-Type: text/html; charset=UTF
< Transfer-Encoding: chunked
< Connection: keep-alive
< X-FLAG: THM{HEADER_FLAG}
<
<!--
```

Task 6 Manual Discovery - Framework Stack

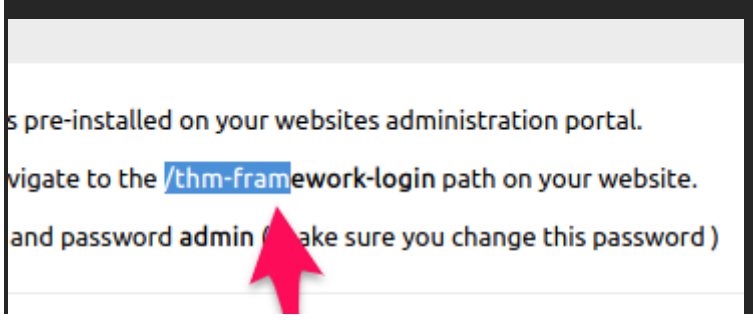
Looking at the page source of our Acme IT Support website (http://10.10.218.237), you'll see a comment at the end of every page with a page load time and also a link to the framework's website, which is https://static-labs.tryhackme.cloud/sites/thm-web-framework. Let's take a look at that website.



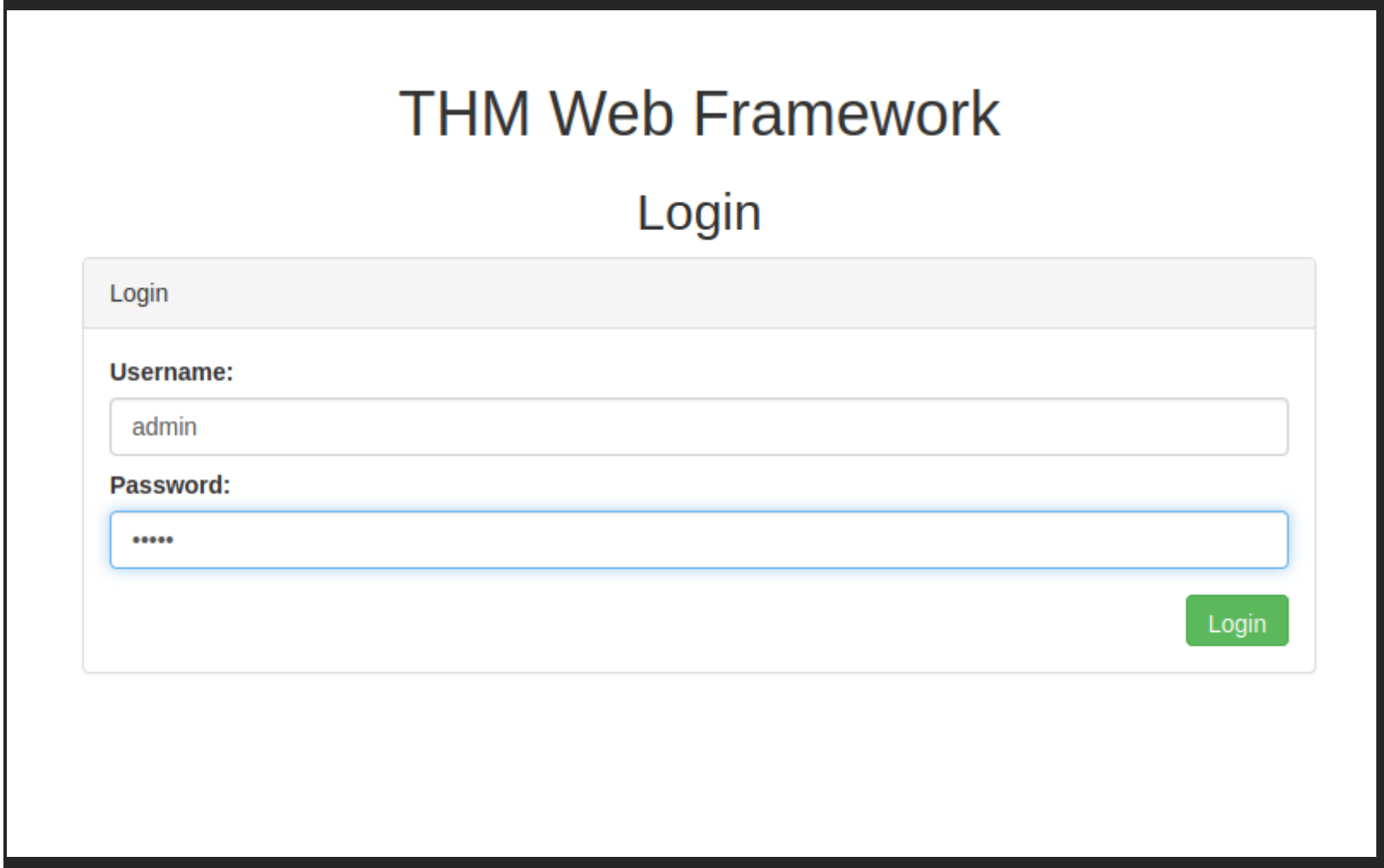
Let's take a look at that website. Viewing the documentation page gives us the path of the framework's administration portal, which gives us a flag if viewed on the Acme IT Support website.Let's take a look at that website. Viewing the documentation page gives us the path of the framework's administration portal, which gives us a flag if viewed on the Acme IT Support website.

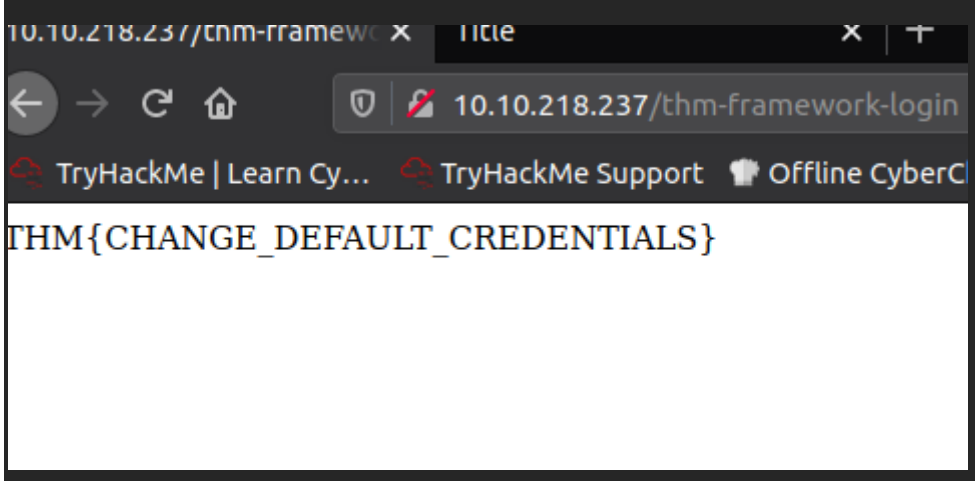


credientals



What is the flag from the framework's administration portal?





Task 7 OSINT - Google Hacking / Dorking

There are also external resources available that can help in discovering information about your target website; these resources are often referred to as OSINT or (Open-Source Intelligence) as they're freely available tools that collect information:

Filter	Example	Description
site	site:tryhackme.com	returns results only from the specified website address
inurl	inurl:admin	returns results that have the specified word in the URL
filetype	filetype:pdf	returns results which are a particular file extension
intitle	intitle:admin	returns results that contain the specified word in the title

More information about google hacking can be found here: https://en.wikipedia.org/wiki/Google_hacking

What Google dork operator can be used to only show results from a particular site?

site:

Task 8 OSINT - Wappalyzer

Wappalyzer (https://www.wappalyzer.com/) is an online tool and browser extension that helps identify what technologies a website uses, such as frameworks, Content Management Systems (CMS), payment processors and much more, and it can even find version numbers as well.

I used it on https://www.wappalyzer.com/lookup/tryhackme.com

Tryhackme.com

Website technology lookup

Website URL, technology, keyword or email address

Technology stack

Programming languages

Node.js

JavaScript graphics

Paths.js

Web frameworks

Express

Video players

Asciinema

VideoJS

UI frameworks

Bootstrap

animate.css

Website profile

Company information PLUS

Company name

TryHackMe

Industry

Computer & Network Security

About

TryHackMe takes the pain out of learning and teaching Cybersecurity. Our platform makes it a comfortable experience to learn by designing prebuilt courses which include virtual machines (VM) hosted in the cloud ready to be deployed. This avoids the h... Show more

Locations

London
London, GB

Company size

11-50 employees

Company type

Public Company

Company founded

2018

Task 9 OSINT - Wayback Machine

Wayback Machine

The Wayback Machine (https://archive.org/web/) is a historical archive of websites that dates back to the late 90s. You can search a domain name, and it will show you all the times the service scraped the web page and saved the contents. This service can help uncover old pages that may still be active on the current website.

tryhackme.com back in 2018, DEC

Task 10 OSINT - GitHub

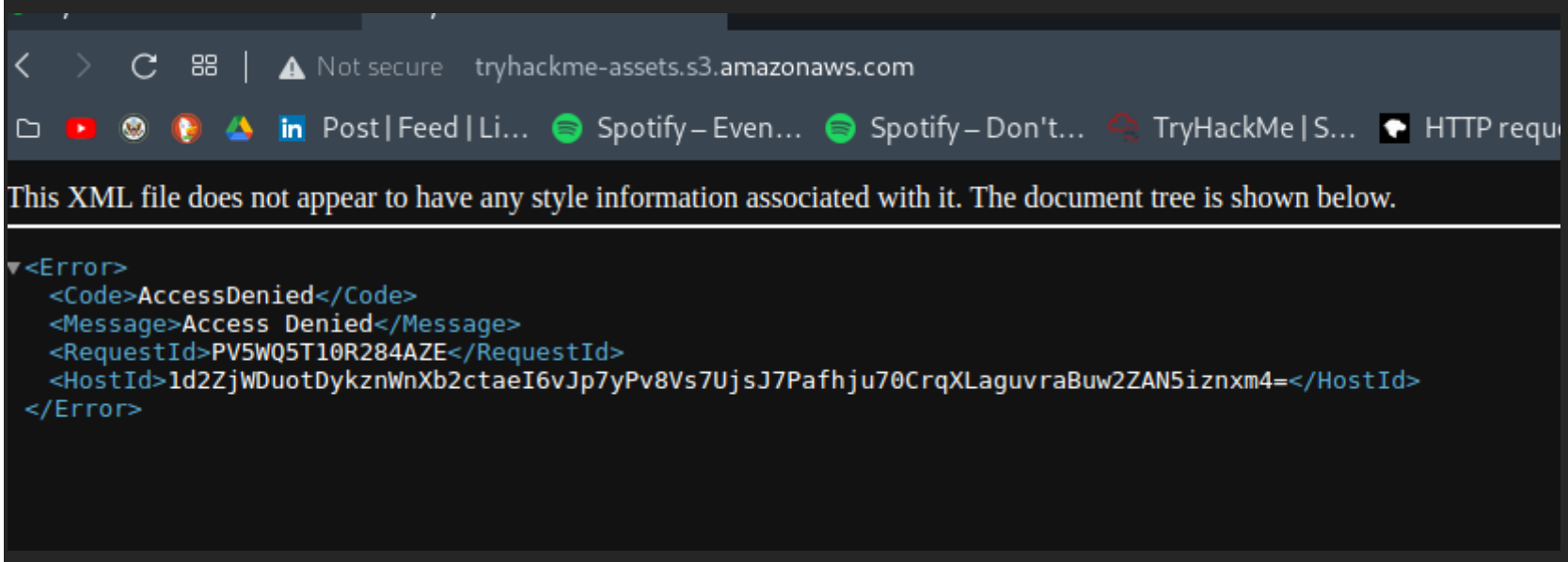
What is Git?

a version control system

Task 11 OSINT - S3 Buckets

Task 11 OSINT S3 Buckets

S3 Buckets are a storage service provided by Amazon AWS, allowing people to save files and even static website content in the cloud accessible over HTTP and HTTPS. The owner of the files can set access permissions to either make files public, private and even writable. Sometimes these access permissions are incorrectly set and inadvertently allow access to files that shouldn't be available to the public. The format of the S3 buckets is http(s)://**{name}.**s3.amazonaws.com where {name} is decided by the owner, such as tryhackme-assets.s3.amazonaws.com.



S3 buckets can be discovered in many ways, such as finding the URLs in the website's page source, GitHub repositories, or even automating the process. One common automation method is by using the company name followed by common terms such as **{name}**-assets, **{name}**-www, **{name}**-public, **{name}**-private, etc.

What URL format do Amazon S3 buckets end in?

.s3.amazonaws.com

Task 12 Automated Discovery

Automation Tools

Although there are many different content discovery tools available, all with their features and flaws, we're going to cover three which are preinstalled on our attack box, ffuf, dirb and gobuster.

On the AttackBox execute the following three commands, targeting the Acme IT Support website and see what results you get.

FFuF Results

```
 :: Threads          : 40
 :: Matcher          : Response status: 200,204,301,302,307,401,403,405
________________________________________________

assets                  [Status: 301, Size: 178, Words: 6, Lines: 8]
contact                 [Status: 200, Size: 3108, Words: 747, Lines: 65]
customers               [Status: 302, Size: 0, Words: 1, Lines: 1]
development.log         [Status: 200, Size: 27, Words: 5, Lines: 1]
monthly                 [Status: 200, Size: 28, Words: 4, Lines: 1]
news                    [Status: 200, Size: 2538, Words: 518, Lines: 51]
private                 [Status: 301, Size: 178, Words: 6, Lines: 8]
robots.txt              [Status: 200, Size: 46, Words: 4, Lines: 3]
sitemap.xml             [Status: 200, Size: 1391, Words: 260, Lines: 43]
:: Progress: [4655/4655] :: Job [1/1] :: 3959 req/sec :: Duration: [0:00:01] :: Errors: 0 ::
root@ip-10-10-74-4:~#
```

dirb

```
root@ip-10-10-74-4:~# dirb http://10.10.218.237/ /usr/share/wordlists/SecLists/Discovery/Web-Content/common.txt

-----------------
DIRB v2.22
By The Dark Raver
-----------------

START_TIME: Wed Jan 26 09:52:34 2022
URL_BASE: http://10.10.218.237/
WORDLIST_FILES: /usr/share/wordlists/SecLists/Discovery/Web-Content/common.txt

-----------------

GENERATED WORDS: 4654

---- Scanning URL: http://10.10.218.237/ ----
==> DIRECTORY: http://10.10.218.237/assets/
+ http://10.10.218.237/contact (CODE:200|SIZE:3108)
+ http://10.10.218.237/customers (CODE:302|SIZE:0)
+ http://10.10.218.237/development.log (CODE:200|SIZE:27)
+ http://10.10.218.237/monthly (CODE:200|SIZE:28)
+ http://10.10.218.237/news (CODE:200|SIZE:2538)
==> DIRECTORY: http://10.10.218.237/private/
+ http://10.10.218.237/robots.txt (CODE:200|SIZE:46)
+ http://10.10.218.237/sitemap.xml (CODE:200|SIZE:1391)

---- Entering directory: http://10.10.218.237/assets/ ----
==> DIRECTORY: http://10.10.218.237/assets/avatars/

---- Entering directory: http://10.10.218.237/private/ ----
+ http://10.10.218.237/private/index.php (CODE:200|SIZE:49)

---- Entering directory: http://10.10.218.237/assets/avatars/ ----

-----------------
END_TIME: Wed Jan 26 09:52:52 2022
DOWNLOADED: 18616 - FOUND: 8
```

gobuster

```
root@ip-10-10-74-4:~# gobuster dir --url http://10.10.218.237/ -w /usr/share/wordlists/SecLists/Discovery/Web-Content/common.txt
===============================================================
Gobuster v3.0.1
by OJ Reeves (@TheColonial) & Christian Mehlmauer (@_FireFart_)
===============================================================
[+] Url:            http://10.10.218.237/
[+] Threads:        10
[+] Wordlist:       /usr/share/wordlists/SecLists/Discovery/Web-Content/common.txt
[+] Status codes:   200,204,301,302,307,401,403
[+] User Agent:     gobuster/3.0.1
[+] Timeout:        10s
===============================================================
2022/01/26 09:54:20 Starting gobuster
===============================================================
/assets (Status: 301)
/contact (Status: 200)
/customers (Status: 302)
/development.log (Status: 200)
/monthly (Status: 200)
/news (Status: 200)
/private (Status: 301)
/robots.txt (Status: 200)
/sitemap.xml (Status: 200)
===============================================================
```

```
2022/01/26 09:54:22 Finished
===============================================================
```

What is the name of the directory beginning "/mo...." that was discovered?

/monthly

What is the name of the log file that was discovered?

/development.log