Metadata Audit

During my final year of my bachelors, I was required to perform a security check on an external companies publicly facing documents that were indexable. This means that those are documents that one could gather though search engines like google and duck duck go.

For my client in particular, I went the extra mile of searching their website for any potentially harmful metadata. Metadata is data about data. One would find information such as a camera type (for images) that were used to take pictures, and even directories where files were saved that an attacker could leverage if they were to gain such information.

Once the exercise was over, I wrote a comprehensive list of findings in a report to my client. Due to privacy concerns and the requirement to keep silent on my client, I will not share that report and redact any mention of my client in the script below with [client website].

Script

```
#!/bin/bash

# this is a bash script that was made in a kali linux virtual machine
# for the best results, place this script in an empty directory and run it
# all matadata will be in the metdata directory


# creating the files for data gathering

sudo touch main_page.txt && sudo chmod 777 main_page.txt

sudo touch about_page.txt && sudo chmod 777 about_page.txt

sudo touch product_page.txt && sudo chmod 777 product_page.txt

sudo touch value_page.txt && sudo chmod 777 value_page.txt

sudo touch contact_page.txt && sudo chmod 777 contact_page.txt

sudo touch ultimate_file.txt && sudo chmod 777 unltimate_file.txt


# gathing of all page source data into the created files
```

```
sudo curl [client website]/ > main_page.txt
sudo curl [client website]/about > about_page
sudo curl [client website]/services > product_page
sudo curl [client website]/gallery > value_page
sudo curl [client website]/contact > contact_page
sudo cat about_page >> ultimate_file.txt
sudo cat about_page >> ultimate_file.txt
sudo cat contact_page >> ultimate_file.txt
sudo cat main_page >> ultimate_file.txt
sudo cat product_page >> ultimate_file.txt
sudo cat value_page >> ultimate_file.txt

# making diretories for metadata txt files
sudo mkdir metdata && sudo chmod 777 metdata

# gathering metadata from jps
sudo mkdir jpg && sudo chmod 777 jpg
sudo touch jpglinks.txt && sudo chmod 777 jpglinks.txt
sudo touch metadata/jpgmetadata.txt && sudo chmod 777 metadata/jpgmetadata.txt
sudo grep -Eo 'https?://[^"]+' ultimate_file.txt | sudo grep -Ei '\.(jpg|jpeg)$' > jpglinks.txt
sudo wget -P jpg -i jpglinks.txt -A jpg,jpeg
sudo exiftool jpg >> metadata/jpgmetadata.txt

# gathering metadata from pngs
```

```
sudo mkdir png && sudo chmod 777 png

sudo touch pnglinks.txt && sudo chmod 777 pnglinks.txt

sudo touch metadata/pngmetadata.txt && sudo chmod 777 metadata/pngmetadata.txt

sudo grep -Eo 'https?://[^"]+' ultimate_file.txt | sudo grep -Ei '\.(png)$' > pnglinks.txt

sudo wget -P png -i pnglinks.txt -A png

sudo exiftool png >> metadata/pngmetadata.txt


# gathering metadata from txt files

sudo mkdir txt && sudo chmod 777 txt

sudo touch txtlinks.txt && sudo chmod 777 txtlinks.txt

sudo touch metadata/txtmetadata.txt && sudo chmod 777 metadata/txtmetadata.txt

sudo grep -Eo 'https?://[^"]+' ultimate_file.txt | sudo grep -Ei '\.(txt)$' > txtlinks.txt

sudo wget -P txt -i txtlinks.txt -A txt

sudo exiftool txt >> metadata/txtmetadata.txt


# gathering metadata from ico files

sudo mkdir ico && sudo chmod 777 ico

sudo touch icolinks.txt && sudo chmod 777 icolinks.txt

sudo touch metadata/icometadata.txt && sudo chmod 777 metadata/icometadata.txt

sudo grep -Eo 'https?://[^"]+' ultimate_file.txt | sudo grep -Ei '\.(ico)$' > icolinks.txt

sudo wget -P ico -i jpglinks.txt -A ico

sudo exiftool ico >> metadata/icometadata.txt

# gething metadata from html
```

```
sudo touch metadata/htmlmatadata.txt && sudo chmod 777 metadata/htmlmetadata.txt

cat ultimate_file.txt | grep '<meta' > metadata/htmlmetadata.txt
```

Additional info on script

The script describes the process that was used to gather files and metadata extraction. The curl tools were used to get the source file info on each webpage. All pages source files were compiled together in the ultimate_file.txt file. From there I could run the command: sudo grep -Eo 'https?://[^"]+' ultimate_file.txt | sudo grep -Ei '\.(file_type)$'

file_type is any file extension. I ran this command with every known file extension. The only files that were found were jpg, png, ico, and txt files.

To get the metadata on the html, the <meta tags were single out and placed into a text file.

The tool of choice to extract metadata from files was exiftool.