Intro to OSINT: arcane Geoguessr techniques to learn everything about everything

Byron Stewart

October 24, 2024

UtahSec

So what’s OSINT?

- **O**pen **S**ource **INT**elligence *(hence the name)*
- is the practice of gathering public, open-source information.

Outline

1. Introduction
2. Metadata and EXIF
   - Applications of Metadata
   - Filesystem Metadata
3. Google Dorking
   - Vulnerable Webservices
4. Shodan and Mass Scanning
   - Shodan Filtering
   - Conclusion

What and Where is Metadata

- Metadata is data that describes data.

What and Where is Metadata

- Metadata is data that describes data.
- We can roughly divide metadata between that which exists *inside* a file and *outside a file*

What and Where is Metadata

- Metadata is data that describes data.
- We can roughly divide metadata between that which exists *inside* a file and *outside a file*
- We are going to look at EXIF, an internal metadata container for images, and filesystem attributes, which are stored externally.

EXIF

EXIF (the Exchangeable image file format) is a ubiquitous tagging system for images and audio.

EXIF

EXIF (the Exchangeable image file format) is a ubiquitous tagging system for images and audio.

- Photos taken on cameras and some phones store the settings in which they were taken to make editing easier.

EXIF

EXIF (the Exchangeable image file format) is a ubiquitous tagging system for images and audio.

- Photos taken on cameras and some phones store the settings in which they were taken to make editing easier.
- Common tags include ISO/aperture, camera model, the date photo was taken, and location of photo (!)
- Because this data is *inside* the file, it will be retained when transferring them.

How to read EXIF data

- `exiftool` is a common CLI tool for reading EXIF

How to read EXIF data

- `exiftool` is a common CLI tool for reading EXIF (good for grepping across a lot of files) ■
- The default image viewer on your computer probably also has an EXIF viewer in the properties view.

How to read EXIF data

- `exiftool` is a common CLI tool for reading EXIF (good for grepping across a lot of files) ■
- The default image viewer on your computer probably also has an EXIF viewer in the properties view.
- *it's also very visible at the beginning of the file in a hex editor*■

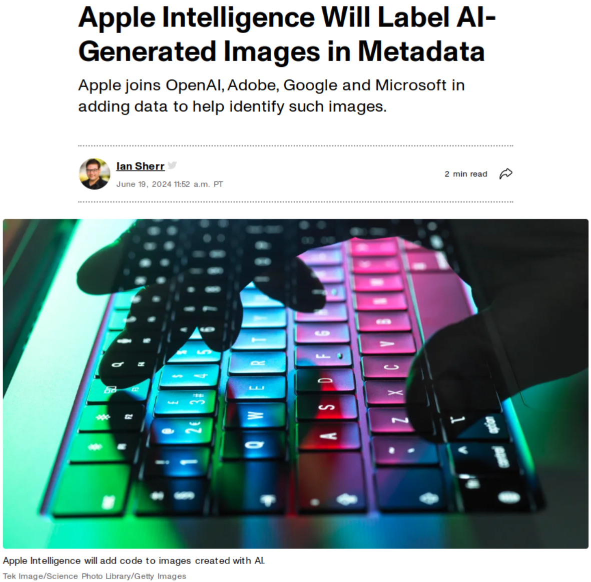
Practical Demonstration

- Most social media sites strip EXIF in the compression step when uploading images.

Practical Demonstration

- Most social media sites strip EXIF in the compression step when uploading images.
- But what about personal sites for people unaware of metadata stripping? ■

Apple's AI Tagging



Apple Intelligence Will Label AI-Generated Images in Metadata

Apple joins OpenAI, Adobe, Google and Microsoft in adding data to help identify such images.

Ian Sherr

June 19, 2024 11:52 a.m. PT

2 min read

Apple Intelligence will add code to images created with AI.

Tek Image/Science Photo Library/Getty Images

Filesystem Attributes

- You already know the common ones: date modified, accessed, and created.

Filesystem Attributes

- You already know the common ones: date modified, accessed, and created.
- Most filesystems allow users to specify arbitrary file attributes.
  Perhaps author, or file format version, or character encoding.

Filesystem Attributes

- You already know the common ones: date modified, accessed, and created.
- Most filesystems allow users to specify arbitrary file attributes. Perhaps author, or file format version, or character encoding.
- This information is stored within the filesystem, and cannot be extracted from the data inside a file.

An aside on metadata modification

- You don't have to rely on your operating system to modify metadata

An aside on metadata modification

- You don't have to rely on your operating system to modify metadata
- You can programmatically edit it yourself with the command line!

An aside on metadata modification

- You don’t have to rely on your operating system to modify metadata
- You can programmatically edit it yourself with the command line!
- *useful for stripping unwanted metadata, transferring between incompatible file systems, and forgery ;)*

An aside on metadata modification

- You don't have to rely on your operating system to modify metadata
- You can programmatically edit it yourself with the command line!
- *useful for stripping unwanted metadata, transferring between incompatible file systems, and forgery ;)*

- On MacOS and other Unix-likes: ■
- `touch -a -m -t 202409110130.01 Foobar.pdf`

An aside on metadata modification

- You don’t have to rely on your operating system to modify metadata
- You can programmatically edit it yourself with the command line!
- *useful for stripping unwanted metadata, transferring between incompatible file systems, and forgery ;)*

- On MacOS and other Unix-likes: ■
- `touch -a -m -t 202409110130.01 Foobar.pdf`
- Windows (should be something like):
- `set metadata [<drive>:][<path>]<metadata.cab>`

1 Introduction

2 Metadata and EXIF

3 Google Dorking

Vulnerable Webservices

4 Shodan and Mass Scanning

What is Google Dorking?

- Advanced Googling.

What is Google Dorking?

- Advanced Googling. Yes, really

What is Google Dorking?

- Advanced Googling. Yes, really
- Search engines support advanced filters like `site=`, `type=`, `inurl=`, etc. They can find specific types of data or specified patterns within them.

What is Google Dorking?

- Advanced Googling. Yes, really
- Search engines support advanced filters like `site=`, `type=`, `inurl=`, etc. They can find specific types of data or specified patterns within them.
- It turns out that misconfigured or underconfigured websites share common patterns between them...that are indexed by Google>:)

Discovering Cool Stuff with Dorking ■

- Open directories typically begin with ”index of”, right?

Discovering Cool Stuff with Dorking ■

- Open directories typically begin with ”index of”, right?
  So `intitle:"index of"` will show open directories

Discovering Cool Stuff with Dorking ■

- Open directories typically begin with ”index of”, right?
  So `intitle:"index of"` will show open directories
  - Though external crawlers like https://odcrawler.xyz/ might be better for looking for specific files

Discovering Cool Stuff with Dorking

- Open directories typically begin with ”index of”, right?
  So `intitle:"index of"` will show open directories
  - Though external crawlers like https://odcrawler.xyz/ might be better for looking for specific files
- Use `inurl:/wp-content/` to find WordPress sites that are incorrectly configured *(because this location really shouldn't be an open directory)*

Discovering Cool Stuff with Dorking ■

- Open directories typically begin with ”index of”, right?
  So `intitle:"index of"` will show open directories
  - Though external crawlers like https://odcrawler.xyz/ might be better for looking for specific files
- Use `inurl:/wp-content/` to find WordPress sites that are incorrectly configured*(because this location really shouldn’t be an open directory)*

Discovering Cool Stuff with Dorking ■

- Open directories typically begin with ”index of”, right?
  So `intitle:"index of"` will show open directories
  - Though external crawlers like https://odcrawler.xyz/ might be better for looking for specific files
- Use `inurl:/wp-content/` to find WordPress sites that are incorrectly configured*(because this location really shouldn't be an open directory)*
- `inurl:ViewerFrame?Mode=` gives webcams with public URLs

Discovering Cool Stuff with Dorking

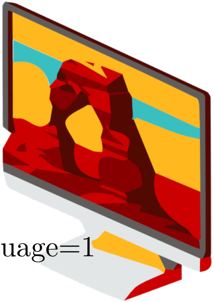
- Open directories typically begin with ”index of”, right?
  So `intitle:"index of"` will show open directories
  - Though external crawlers like https://odcrawler.xyz/ might be better for looking for specific files
- Use `inurl:/wp-content/` to find WordPress sites that are incorrectly configured *(because this location really shouldn't be an open directory)*
- `inurl:ViewerFrame?Mode=` gives webcams with public URLs
- AngelSecurityTeam/SearchCAM on Github has a long list of common webcam dorks

Discovering Cool Stuff with Dorking ■

- Open directories typically begin with ”index of”, right?
  So `intitle:"index of"` will show open directories
  - Though external crawlers like https://odcrawler.xyz/ might be better for looking for specific files
- Use `inurl:/wp-content/` to find WordPress sites that are incorrectly configured*(because this location really shouldn’t be an open directory)*
- `inurl:ViewerFrame?Mode=` gives webcams with public URLs
- AngelSecurityTeam/SearchCAM on Github has a long list of common webcam dorks
- Bonus: `filetype:log intext:password`

Discovering Cool Stuff with Dorking ■

- Open directories typically begin with ”index of”, right?
  So `intitle:"index of"` will show open directories
  - Though external crawlers like https://odcrawler.xyz/ might be better for looking for specific files
- Use `inurl:/wp-content/` to find WordPress sites that are incorrectly configured *(because this location really shouldn't be an open directory)*
- `inurl:ViewerFrame?Mode=` gives webcams with public URLs
- AngelSecurityTeam/SearchCAM on Github has a long list of common webcam dorks
- Bonus: `filetype:log intext:password`



- *list in case I get barraged by CAPTCHAs:*
- http://61.211.241.239/CgiStart?page=Single&Mode=Motion&Language=1
- http://194.94.76.134/control/userimage.html

Manual Banner Scraping

- Last time we talked about banners, which describe information about the service running on a given port.

- Also, we can `whois` any given IP to see who it is registered to. For example, 204.79.197.200 ■
- **To gather information about a webserver, grabbing its banners is helpful.**

Manual Banner Scraping

- Last time we talked about banners, which describe information about the service running on a given port.
- We can `ping` the domain name to lookup its IP address and `nmap` it for any open ports
- Also, we can `whois` any given IP to see who it is registered to. For example, 204.79.197.200 ■
- **To gather information about a webserver, grabbing its banners is helpful.**

So you want to scan a whole lot of ports at once...

- But it's difficult to scan the entire internet (or at least the IPv4 space)

So you want to scan a whole lot of ports at once...

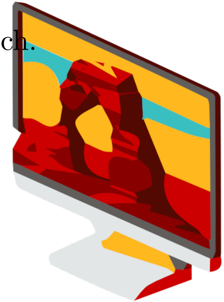
- But it's difficult to scan the entire internet (or at least the IPv4 space)
- *and it might be dubiously legal depending on your intent*

So you want to scan a whole lot of ports at once...

- But it’s difficult to scan the entire internet (or at least the IPv4 space)
- *and it might be dubiously legal depending on your intent*
- So what do you do?

So you want to scan a whole lot of ports at once...

- But it's difficult to scan the entire internet (or at least the IPv4 space)
- *and it might be dubiously legal depending on your intent*
- So what do you do?
- **Shodan** is a massive banner repository for IoT and security research.



More Dorking with Shodan ■

- The naive search *'webcams'* gives us IP addresses with open webcams

More Dorking with Shodan ■

- The naive search *'webcams'* gives us IP addresses with open webcams
- We can specifically look for mySQL databases with the filter `product:MySQL`

More Dorking with Shodan ■

- The naive search *'webcams'* gives us IP addresses with open webcams
- We can specifically look for mySQL databases with the filter `product:MySQL`
- `screenshot.label:ics`

More Dorking with Shodan ■

- The naive search *'webcams'* gives us IP addresses with open webcams
- We can specifically look for mySQL databases with the filter `product:MySQL`
- `screenshot.label:ics`
- A major aim of Shodan is to index by vulnerability.
  If we searched for `vuln:ms17-010`, we would get websites vulnerable to the 2017 Eternal Blue exploit...

  But this feature isn't available without a business or academic account.

Conclusions

We looked at

- different forms of metadata and how they reveal information

Conclusions

We looked at

- different forms of metadata and how they reveal information
- Google Dorking, a way to find information about potentially vulnerable websites

Conclusions

We looked at

- different forms of metadata and how they reveal information
- Google Dorking, a way to find information about potentially vulnerable websites
- Shodan, a massive database of the internet.

Conclusions

We looked at

- different forms of metadata and how they reveal information
- Google Dorking, a way to find information about potentially vulnerable websites
- Shodan, a massive database of the internet.

Any questions?