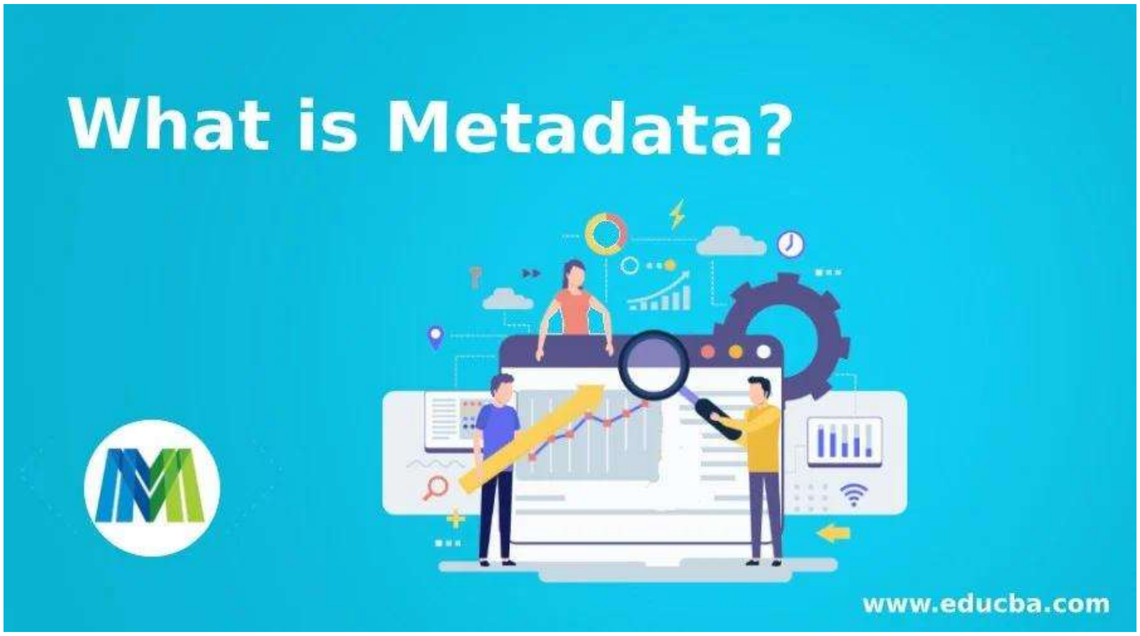
Day - 6 of CEH

The objective of **Day–6** is to understand **Metadata**, its importance in **Ethical Hacking**, and to gain hands-on exposure to **information gathering (reconnaissance) tools** such as **ExifTool, Metagoofil, theHarvester, and Shodan**.
This session focuses on how **hidden data inside files, documents, and online services** can unintentionally expose sensitive information and how ethical hackers analyze this data responsibly to improve security.

---

1. Metadata

Metadata refers to *data about data*. It is automatically generated information stored within digital files such as images, PDFs, Word documents, videos, and spreadsheets.



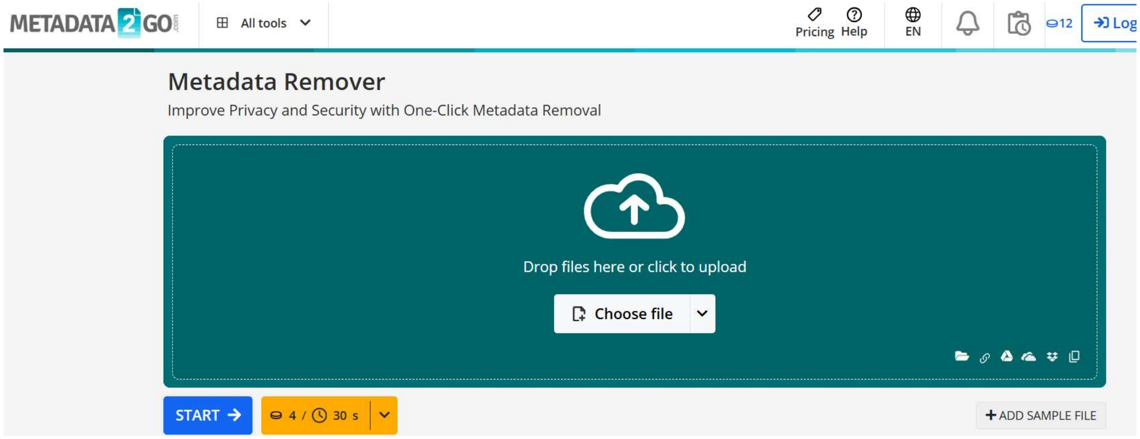
Examples of Metadata

- Author or creator name
- File creation and modification dates
- Software and operating system used
- GPS coordinates (in images taken by mobile phones or cameras)
- Device, camera, or smartphone model

Metadata Remover

A **Metadata Remover** is a tool or technique used to **remove hidden metadata information** from digital files such as images, PDFs, Word documents, and videos.

Removing metadata helps protect **privacy, confidentiality, and security** by preventing unintended information disclosure.



Metadata Remover

Improve Privacy and Security with One-Click Metadata Removal

What Metadata Removers Do

Metadata removers are used to:

- Delete author and creator information
- Remove GPS location data from images
- Erase device or camera details
- Clear software and system information
- Sanitize documents before public sharing

Importance in Ethical Hacking

Metadata plays a crucial role during the **reconnaissance phase** of ethical hacking because it can unintentionally disclose sensitive organizational details.

Metadata analysis helps to:

- Identify employee names and usernames
- Discover software versions and platforms in use
- Reveal internal directory paths and file structures
- Expose geographic location information
- Assist in ethical social engineering assessments

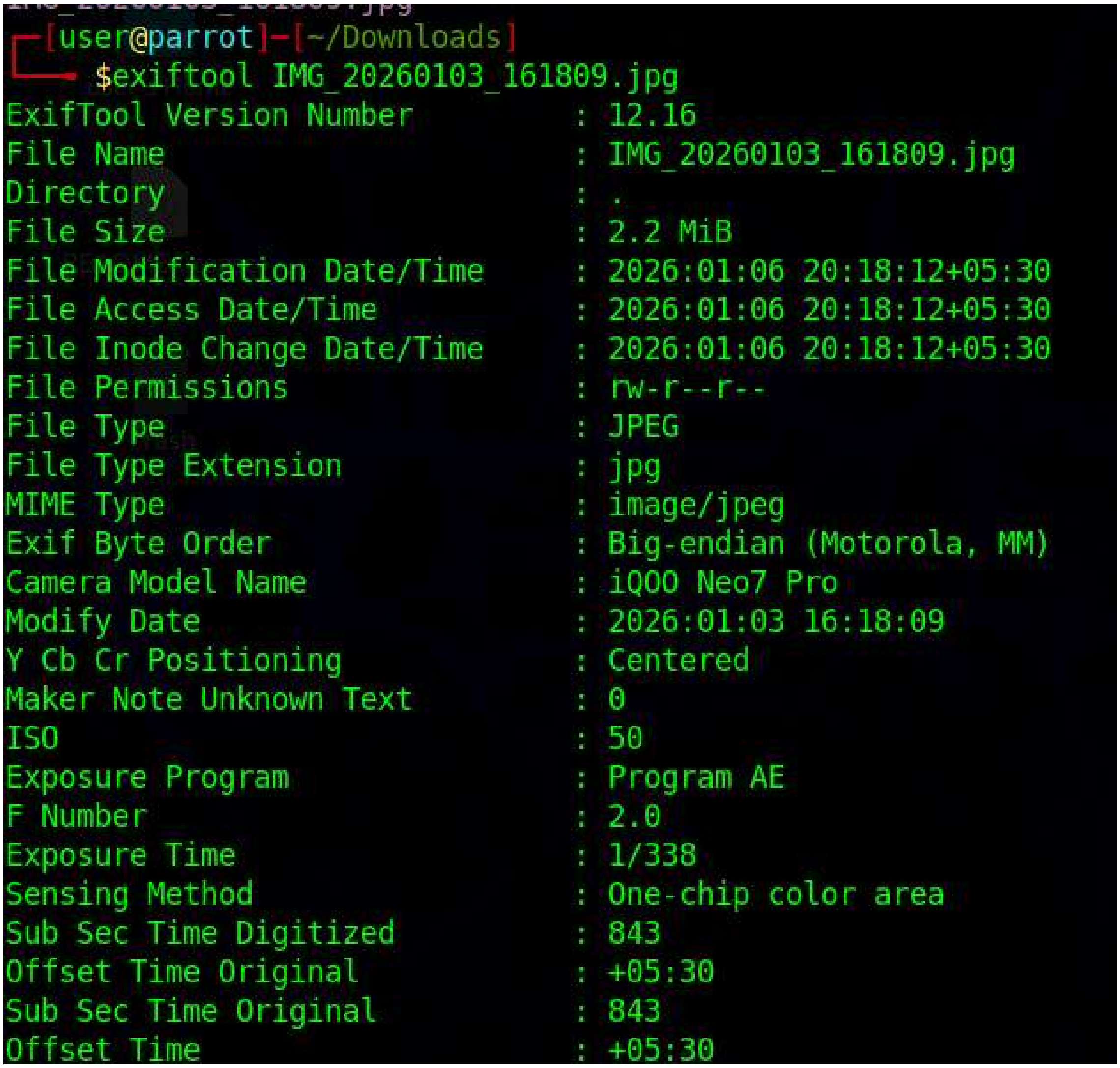
---

2. ExifTool

ExifTool is a powerful command-line utility used to read, write, and analyze metadata from various file formats including images, videos, PDFs, and documents.

Basic ExifTool Command

exiftool image.jpg



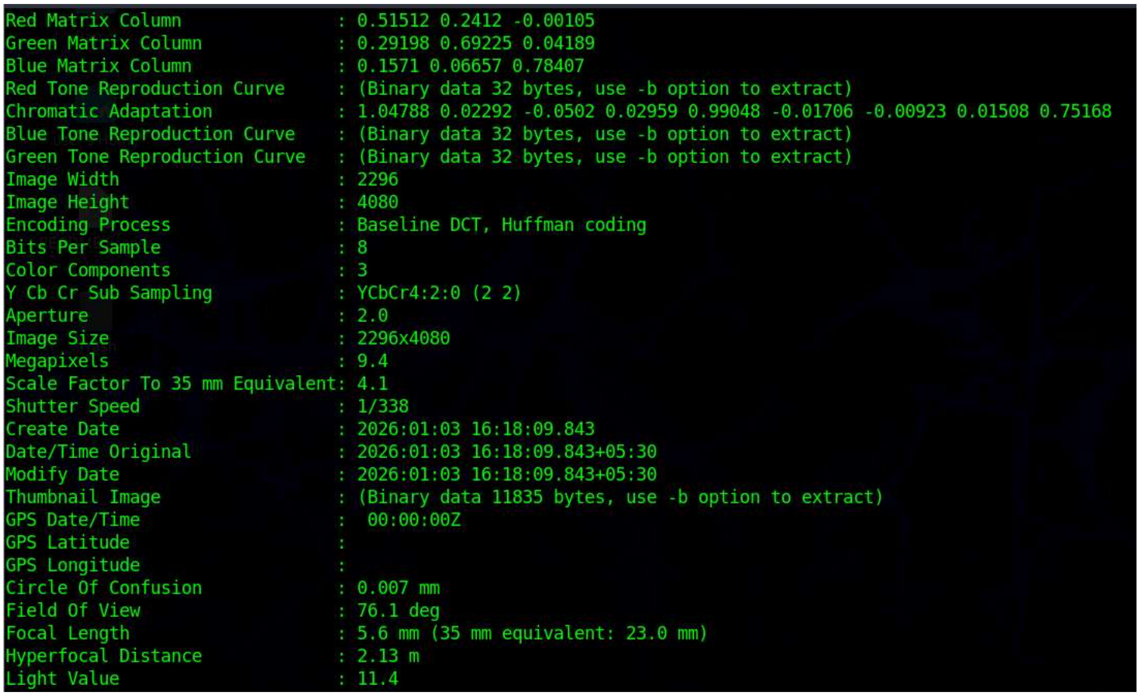
```
IMG_20260103_161809.jpg
┌─[user@parrot]─[~/Downloads]
└──╼ $exiftool IMG_20260103_161809.jpg
ExifTool Version Number         : 12.16
File Name                       : IMG_20260103_161809.jpg
Directory                       : .
File Size                       : 2.2 MiB
File Modification Date/Time     : 2026:01:06 20:18:12+05:30
File Access Date/Time           : 2026:01:06 20:18:12+05:30
File Inode Change Date/Time     : 2026:01:06 20:18:12+05:30
File Permissions                : rw-r--r--
File Type                       : JPEG
File Type Extension             : jpg
MIME Type                       : image/jpeg
Exif Byte Order                 : Big-endian (Motorola, MM)
Camera Model Name               : iQOO Neo7 Pro
Modify Date                     : 2026:01:03 16:18:09
Y Cb Cr Positioning             : Centered
Maker Note Unknown Text         : 0
ISO                             : 50
Exposure Program                : Program AE
F Number                        : 2.0
Exposure Time                   : 1/338
Sensing Method                  : One-chip color area
Sub Sec Time Digitized          : 843
Offset Time Original            : +05:30
Sub Sec Time Original           : 843
Offset Time                     : +05:30
```

Information Extracted

- Camera or mobile device model
- Date and time of image capture
- GPS latitude and longitude
- File size, format, and encoding details

Use Case

ExifTool is commonly used to analyze images downloaded from websites, emails, or social media platforms to identify **metadata leakage** that may reveal sensitive personal or organizational information.



```
Red Matrix Column               : 0.51512 0.2412 -0.00105
Green Matrix Column             : 0.29198 0.69225 0.04189
Blue Matrix Column              : 0.1571 0.06657 0.78407
Red Tone Reproduction Curve     : (Binary data 32 bytes, use -b option to extract)
Chromatic Adaptation            : 1.04788 0.02292 -0.0502 0.02959 0.99048 -0.01706 -0.00923 0.01508 0.75168
Blue Tone Reproduction Curve    : (Binary data 32 bytes, use -b option to extract)
Green Tone Reproduction Curve   : (Binary data 32 bytes, use -b option to extract)
Image Width                     : 2296
Image Height                    : 4080
Encoding Process                : Baseline DCT, Huffman coding
Bits Per Sample                 : 8
Color Components                : 3
Y Cb Cr Sub Sampling            : YCbCr4:2:0 (2 2)
Aperture                        : 2.0
Image Size                      : 2296x4080
Megapixels                      : 9.4
Scale Factor To 35 mm Equivalent: 4.1
Shutter Speed                   : 1/338
Create Date                     : 2026:01:03 16:18:09.843
Date/Time Original              : 2026:01:03 16:18:09.843+05:30
Modify Date                     : 2026:01:03 16:18:09.843+05:30
Thumbnail Image                 : (Binary data 11835 bytes, use -b option to extract)
GPS Date/Time                   :  00:00:00Z
GPS Latitude                    :
GPS Longitude                   :
Circle Of Confusion             : 0.007 mm
Field Of View                   : 76.1 deg
Focal Length                    : 5.6 mm (35 mm equivalent: 23.0 mm)
Hyperfocal Distance             : 2.13 m
Light Value                     : 11.4
```

3. gofile.io – Image Download Process

gofile.io is a free file-sharing platform that allows users to upload and download files without mandatory registration.

Steps to Download an Image

1. Open a web browser and visit **gofile.io**
2. Access the shared download link
3. Select the required image file
4. Click on **Download**
5. Save the image to the local system
6. Analyze the downloaded image using **ExifTool**

Purpose

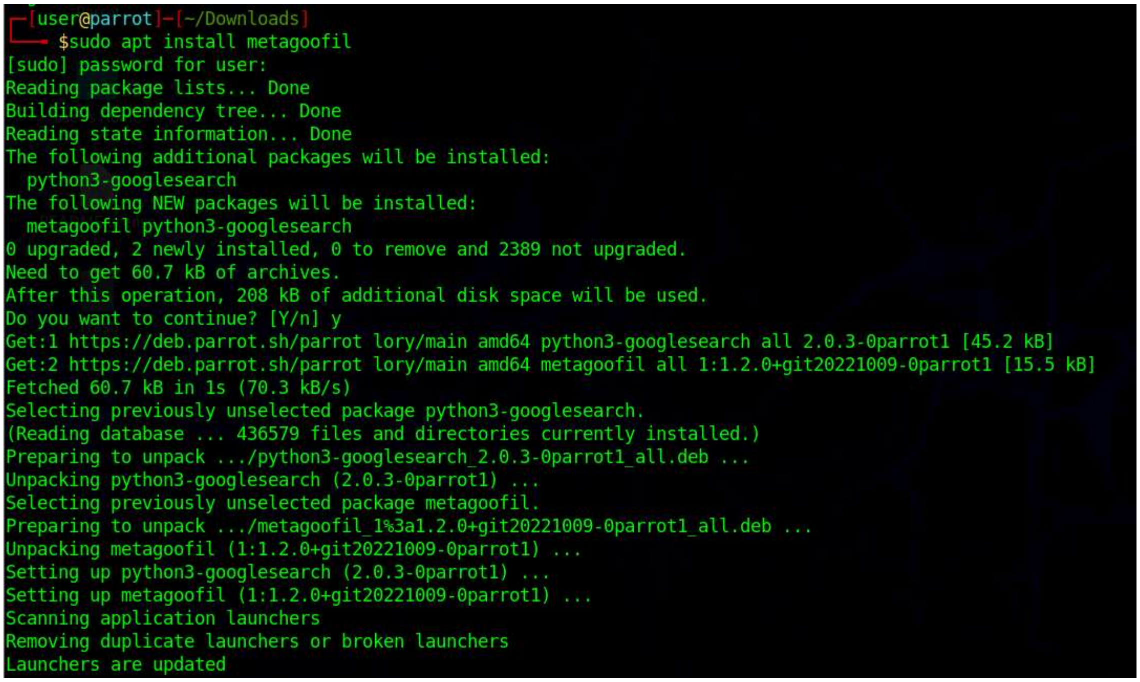
The downloaded images are analyzed to check for **metadata exposure**, such as GPS location, device details, and timestamps, which could pose privacy or security risks.

4. Metagoofil

Metagoofil is a Kali Linux–based reconnaissance tool used to extract metadata from **publicly available documents** belonging to a specific domain.

Installing Metagoofil

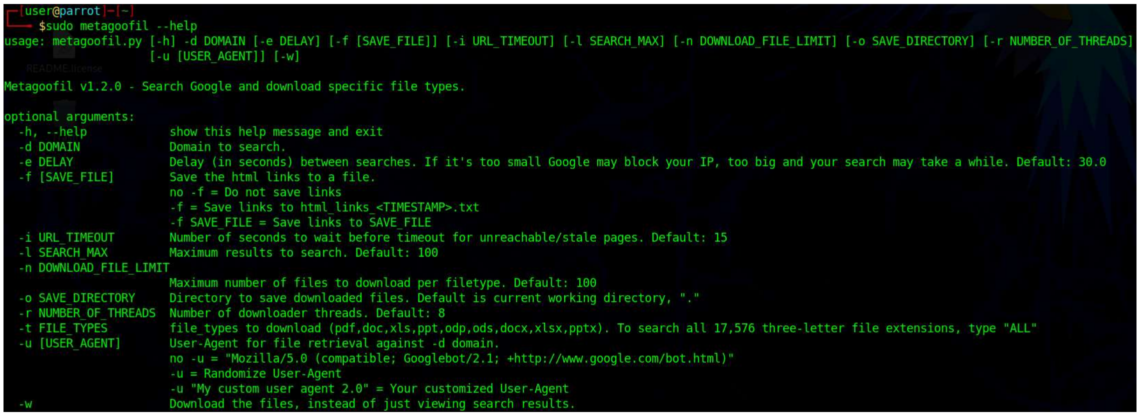
sudo apt install metagoofil



```
┌─[user@parrot]─[~/Downloads]
└──╼ $sudo apt install metagoofil
[sudo] password for user:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  python3-googlesearch
The following NEW packages will be installed:
  metagoofil python3-googlesearch
0 upgraded, 2 newly installed, 0 to remove and 2389 not upgraded.
Need to get 60.7 kB of archives.
After this operation, 208 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 https://deb.parrot.sh/parrot lory/main amd64 python3-googlesearch all 2.0.3-0parrot1 [45.2 kB]
Get:2 https://deb.parrot.sh/parrot lory/main amd64 metagoofil all 1:1.2.0+git20221009-0parrot1 [15.5 kB]
Fetched 60.7 kB in 1s (70.3 kB/s)
Selecting previously unselected package python3-googlesearch.
(Reading database ... 436579 files and directories currently installed.)
Preparing to unpack .../python3-googlesearch_2.0.3-0parrot1_all.deb ...
Unpacking python3-googlesearch (2.0.3-0parrot1) ...
Selecting previously unselected package metagoofil.
Preparing to unpack .../metagoofil_1%3a1.2.0+git20221009-0parrot1_all.deb ...
Unpacking metagoofil (1:1.2.0+git20221009-0parrot1) ...
Setting up python3-googlesearch (2.0.3-0parrot1) ...
Setting up metagoofil (1:1.2.0+git20221009-0parrot1) ...
Scanning application launchers
Removing duplicate launchers or broken launchers
Launchers are updated
```

Help Command

sudo metagoofil --help



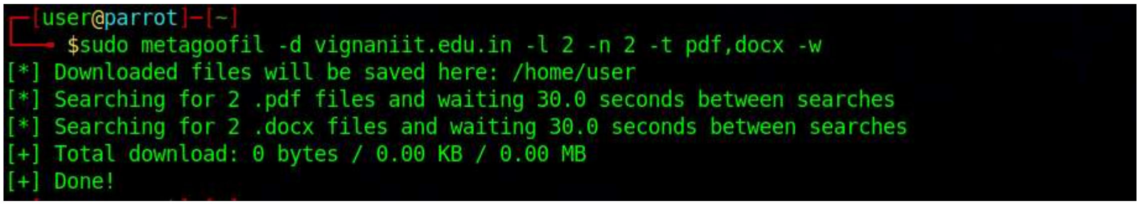
```
┌─[user@parrot]─[~]
└──╼ $sudo metagoofil --help
usage: metagoofil.py [-h] -d DOMAIN [-e DELAY] [-f [SAVE_FILE]] [-i URL_TIMEOUT] [-l SEARCH_MAX] [-n DOWNLOAD_FILE_LIMIT] [-o SAVE_DIRECTORY] [-r NUMBER_OF_THREADS]
                     [-u [USER_AGENT]] [-w]

Metagoofil v1.2.0 - Search Google and download specific file types.

optional arguments:
  -h, --help            show this help message and exit
  -d DOMAIN             Domain to search.
  -e DELAY              Delay (in seconds) between searches. If it's too small Google may block your IP, too big and your search may take a while. Default: 30.0
  -f [SAVE_FILE]        Save the html links to a file.
                        no -f = Do not save links
                        -f = Save links to html_links_<TIMESTAMP>.txt
                        -f SAVE_FILE = Save links to SAVE_FILE
  -i URL_TIMEOUT        Number of seconds to wait before timeout for unreachable/stale pages. Default: 15
  -l SEARCH_MAX         Maximum results to search. Default: 100
  -n DOWNLOAD_FILE_LIMIT
                        Maximum number of files to download per filetype. Default: 100
  -o SAVE_DIRECTORY     Directory to save downloaded files. Default is current working directory, "."
  -r NUMBER_OF_THREADS  Number of downloader threads. Default: 8
  -t FILE_TYPES         file_types to download (pdf,doc,xls,ppt,odp,ods,docx,xlsx,pptx). To search all 17,576 three-letter file extensions, type "ALL"
  -u [USER_AGENT]       User-Agent for file retrieval against -d domain.
                        no -u = "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
                        -u = Randomize User-Agent
                        -u "My custom user agent 2.0" = Your customized User-Agent
  -w                    Download the files, instead of just viewing search results.
```

Targeted Command Used

sudo metagoofil -d vignaniit.edu.in -l 10 -n 10 -t pdf,xls,xlsx -w



```
┌─[user@parrot]─[~]
└──╼ $sudo metagoofil -d vignaniit.edu.in -l 2 -n 2 -t pdf,docx -w
[*] Downloaded files will be saved here: /home/user
[*] Searching for 2 .pdf files and waiting 30.0 seconds between searches
[*] Searching for 2 .docx files and waiting 30.0 seconds between searches
[+] Total download: 0 bytes / 0.00 KB / 0.00 MB
[+] Done!
```

Explanation

- -d → Specifies the target domain
- -l 10 → Limits the number of search engine results
- -n 10 → Number of files to download
- -t → File types to search (pdf, xls, xlsx)
- -w → Generates an HTML report

Information Collected

- Author and employee names
- Usernames and email IDs
- Software and application details
- Internal file paths and document structure

Insecam

:contentReference[oaicite:0]{index=0} is a publicly accessible website that indexes **open and unsecured IP cameras** available on the internet.
It is commonly referenced in **cyber security awareness and OSINT studies** to demonstrate the risks of misconfigured devices.

Purpose of Insecam

Insecam is used for **educational and awareness purposes** to show how improperly secured cameras can expose live video feeds to anyone on the internet.

Its main goals are to:

- Highlight the importance of securing IoT and IP camera devices
- Demonstrate real-world impacts of weak or default credentials
- Promote cyber security awareness among users and organizations

---

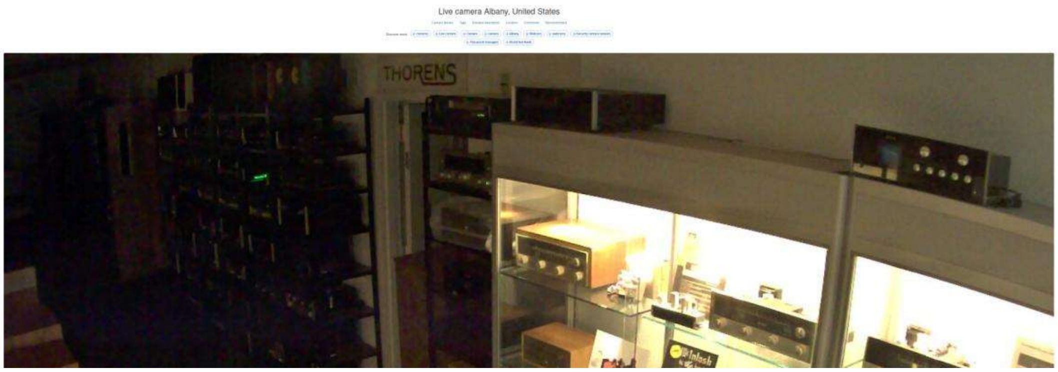
Type of Devices Shown

The platform may list publicly exposed:

- Home surveillance cameras
- Office and shop security cameras
- Baby monitors
- Traffic and public area cameras

These devices are visible **only because they are misconfigured or left unsecured**.

Security Risks Demonstrated

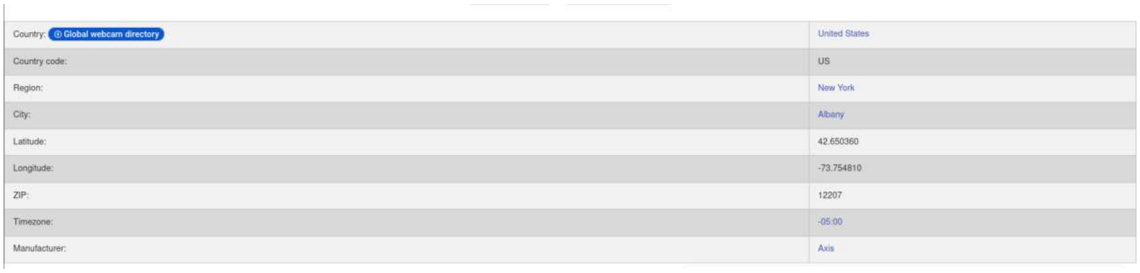


Live camera Albany, United States

Insecam clearly shows the dangers of poor security practices, such as:

- Using default usernames and passwords
- Not enabling authentication on IP cameras
- Exposing cameras directly to the internet
- Lack of firmware updates and security patches

Such issues can lead to **privacy violations and unauthorized monitoring**.



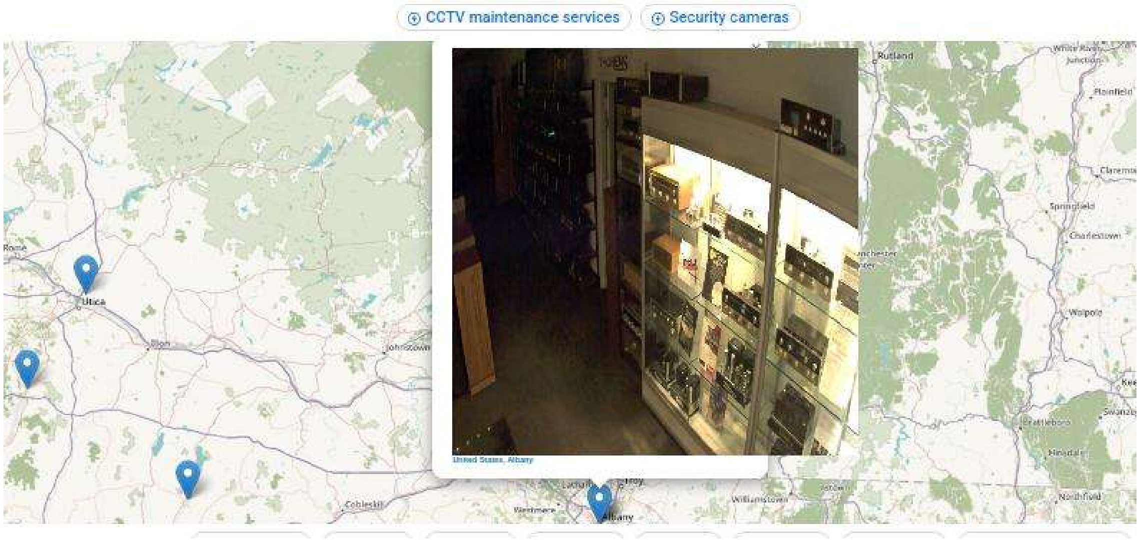
Country: ⊙ Global webcam directory	United States
Country code:	US
Region:	New York
City:	Albany
Latitude:	42.650360
Longitude:	-73.754810
ZIP:	12207
Timezone:	-05:00
Manufacturer:	Axis

Importance in Ethical Hacking & OSINT

From an ethical hacking perspective, Insecam helps learners understand:

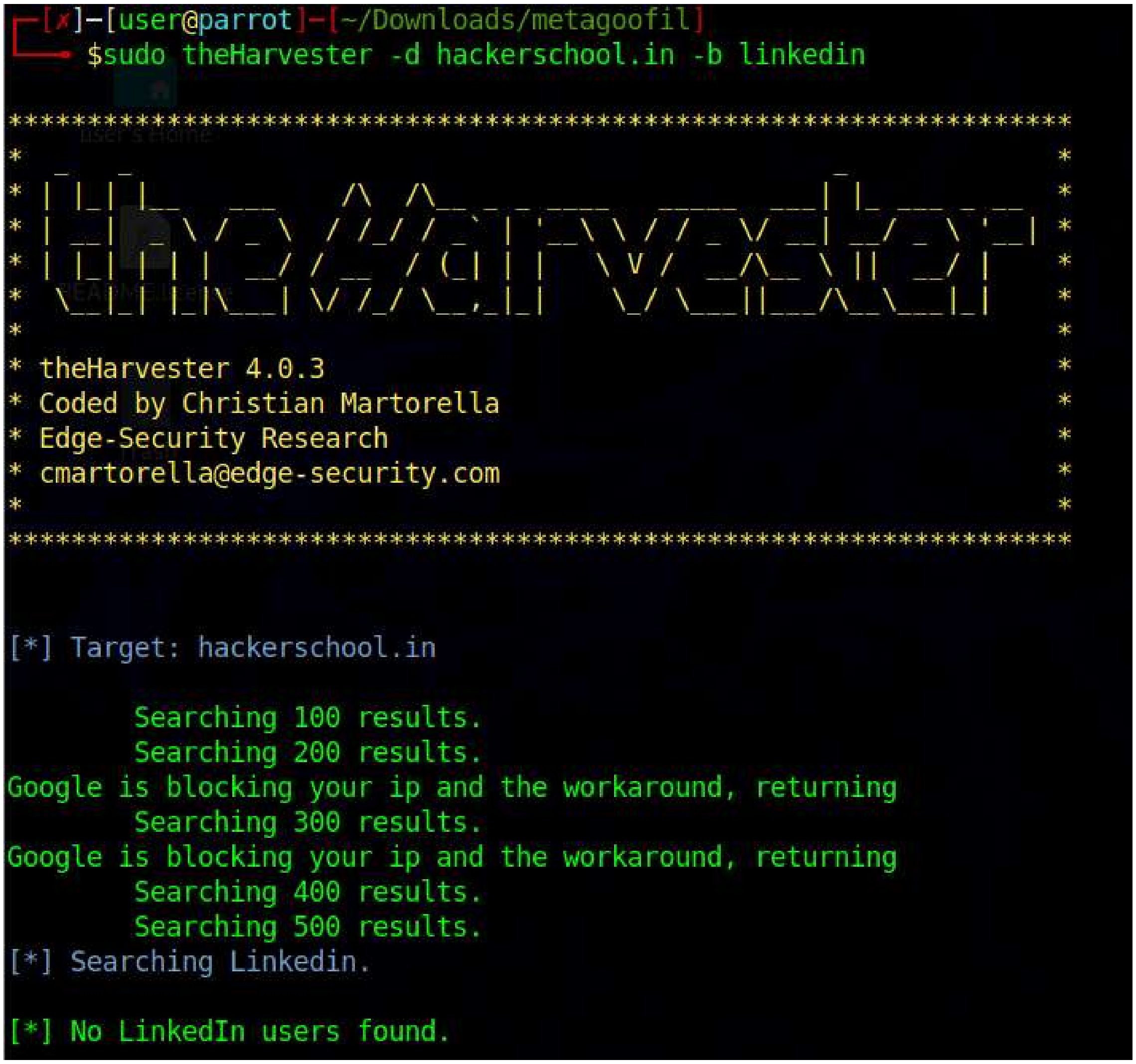
- How exposed IoT devices can be discovered through OSINT
- The consequences of misconfiguration in real-world systems
- Why device hardening and access control are critical

It is mainly used in **defensive security learning**, not for exploitation.



5. theHarvester

theHarvester is an **Open-Source Intelligence (OSINT)** tool used for gathering publicly available information related to a domain.



```
┌─[✗]─[user@parrot]─[~/Downloads/metagoofil]
└──╼ $sudo theHarvester -d hackerschool.in -b linkedin

*******************************************************************
*  _   _                                            _             *
* | |_| |__   ___    /\  /\__ _ _ ____   _____  ___| |_ ___ _ __  *
* | __|  _ \ / _ \  / /_/ / _` | '__\ \ / / _ \/ __| __/ _ \ '__| *
* | |_| | | |  __/ / __  / (_| | |   \ V /  __/\__ \ ||  __/ |    *
*  \__|_| |_|\___| \/ /_/ \__,_|_|    \_/ \___||___/\__\___|_|    *
*                                                                 *
* theHarvester 4.0.3                                              *
* Coded by Christian Martorella                                   *
* Edge-Security Research                                          *
* cmartorella@edge-security.com                                   *
*                                                                 *
*******************************************************************


[*] Target: hackerschool.in

        Searching 100 results.
        Searching 200 results.
Google is blocking your ip and the workaround, returning
        Searching 300 results.
Google is blocking your ip and the workaround, returning
        Searching 400 results.
        Searching 500 results.
[*] Searching Linkedin.

[*] No LinkedIn users found.
```

Usage

- Performs passive reconnaissance
- Collects data without directly interacting with the target
- Uses public sources such as search engines and databases

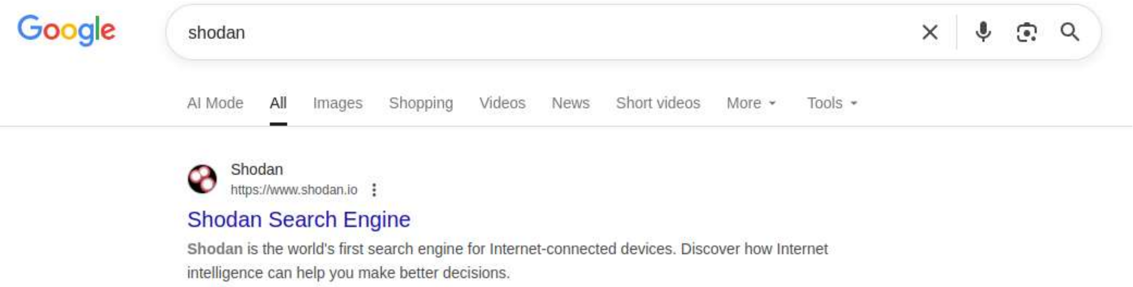
Collected Data

- Email addresses
- Subdomains
- IP addresses
- Hostnames

This information helps in understanding the **digital footprint** of an organization.

6. Shodan

Shodan is a specialized search engine designed to discover **Internet-connected devices and services**.



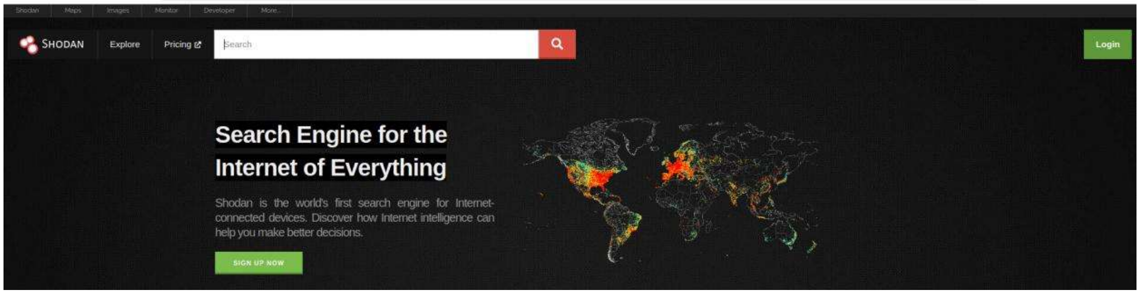
Shodan
https://www.shodan.io

Shodan Search Engine

Shodan is the world's first search engine for Internet-connected devices. Discover how Internet intelligence can help you make better decisions.

Login Process

1. Visit **shodan.io**
2. Click **Sign Up / Login**
3. Register using an email address
4. Log in to access advanced search filters

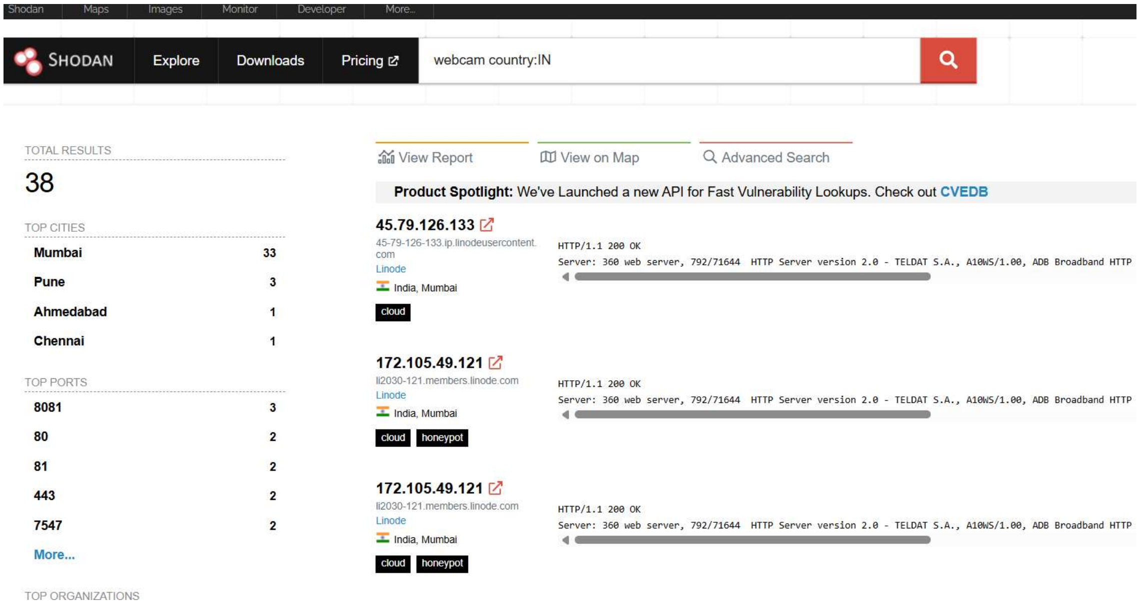


Search Engine for the Internet of Everything

Shodan is the world's first search engine for Internet-connected devices. Discover how Internet intelligence can help you make better decisions.

Search Filters Used

- port:20 → Identifies FTP services
- country:IN → Filters results from India



webcam country:IN

TOTAL RESULTS

38

TOP CITIES

Mumbai	33
Pune	3
Ahmedabad	1
Chennai	1

TOP PORTS

8081	3
80	2
81	2
443	2
7547	2

More...

TOP ORGANIZATIONS

Product Spotlight: We've Launched a new API for Fast Vulnerability Lookups. Check out **CVEDB**

45.79.126.133
45-79-126-133.ip.linodeusercontent.com
Linode
India, Mumbai
cloud

```
HTTP/1.1 200 OK
Server: 360 web server, 792/71644  HTTP Server version 2.0 - TELDAT S.A., A10WS/1.00, ADB Broadband HTTP
```

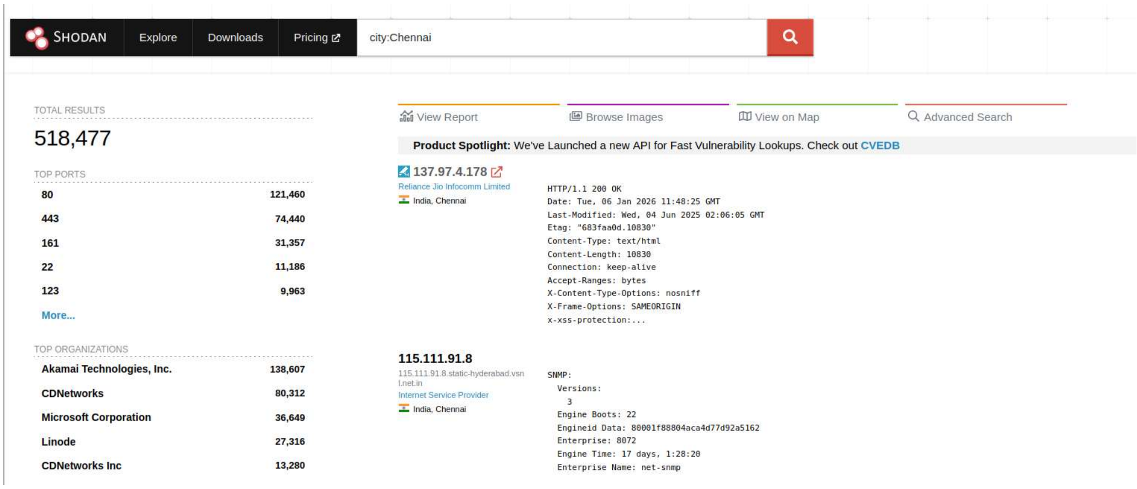
172.105.49.121
li2030-121.members.linode.com
Linode
India, Mumbai
cloud honeypot

```
HTTP/1.1 200 OK
Server: 360 web server, 792/71644  HTTP Server version 2.0 - TELDAT S.A., A10WS/1.00, ADB Broadband HTTP
```

172.105.49.121
li2030-121.members.linode.com
Linode
India, Mumbai
cloud honeypot

```
HTTP/1.1 200 OK
Server: 360 web server, 792/71644  HTTP Server version 2.0 - TELDAT S.A., A10WS/1.00, ADB Broadband HTTP
```

- apache → Finds servers running Apache
- city:chennai → Devices located in Chennai



city:Chennai

TOTAL RESULTS

518,477

TOP PORTS

80	121,460
443	74,440
161	31,357
22	11,186
123	9,963

More...

TOP ORGANIZATIONS

Akamai Technologies, Inc.	138,607
CDNetworks	80,312
Microsoft Corporation	36,649
Linode	27,316
CDNetworks Inc	13,280

Product Spotlight: We've Launched a new API for Fast Vulnerability Lookups. Check out **CVEDB**

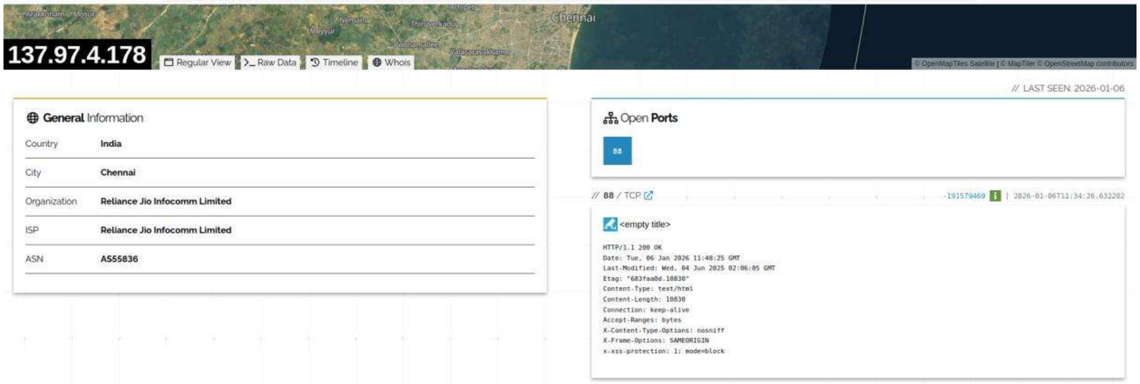
137.97.4.178
Reliance Jio Infocomm Limited
India, Chennai

```
HTTP/1.1 200 OK
Date: Tue, 06 Jan 2026 11:48:25 GMT
Last-Modified: Wed, 04 Jun 2025 02:06:05 GMT
Etag: "683faa0d.10830"
Content-Type: text/html
Content-Length: 10830
Connection: keep-alive
Accept-Ranges: bytes
X-Content-Type-Options: nosniff
X-Frame-Options: SAMEORIGIN
x-xss-protection:...
```

115.111.91.8
115.111.91.8.static-hyderabad.vsnl.net.in
Internet Service Provider
India, Chennai

```
SNMP:
  Versions:
    3
  Engine Boots: 22
  Engineid Data: 80001f88804aca4d77d92a5162
  Enterprise: 8072
  Engine Time: 17 days, 1:28:20
  Enterprise Name: net-snmp
```

- camera → Internet-connected cameras


Chennai
Regular View
Raw Data
Timeline
Whois
© OpenMapTiles Satellite | © MapTiler © OpenStreetMap contributors

137.97.4.178

// LAST SEEN: 2026-01-06

General Information

Country	India
City	Chennai
Organization	Reliance Jio Infocomm Limited
ISP	Reliance Jio Infocomm Limited
ASN	AS55836

Open **Ports**

88

// 88 / TCP   -191579469 | 2026-01-06T11:34:26.632282

<empty title>

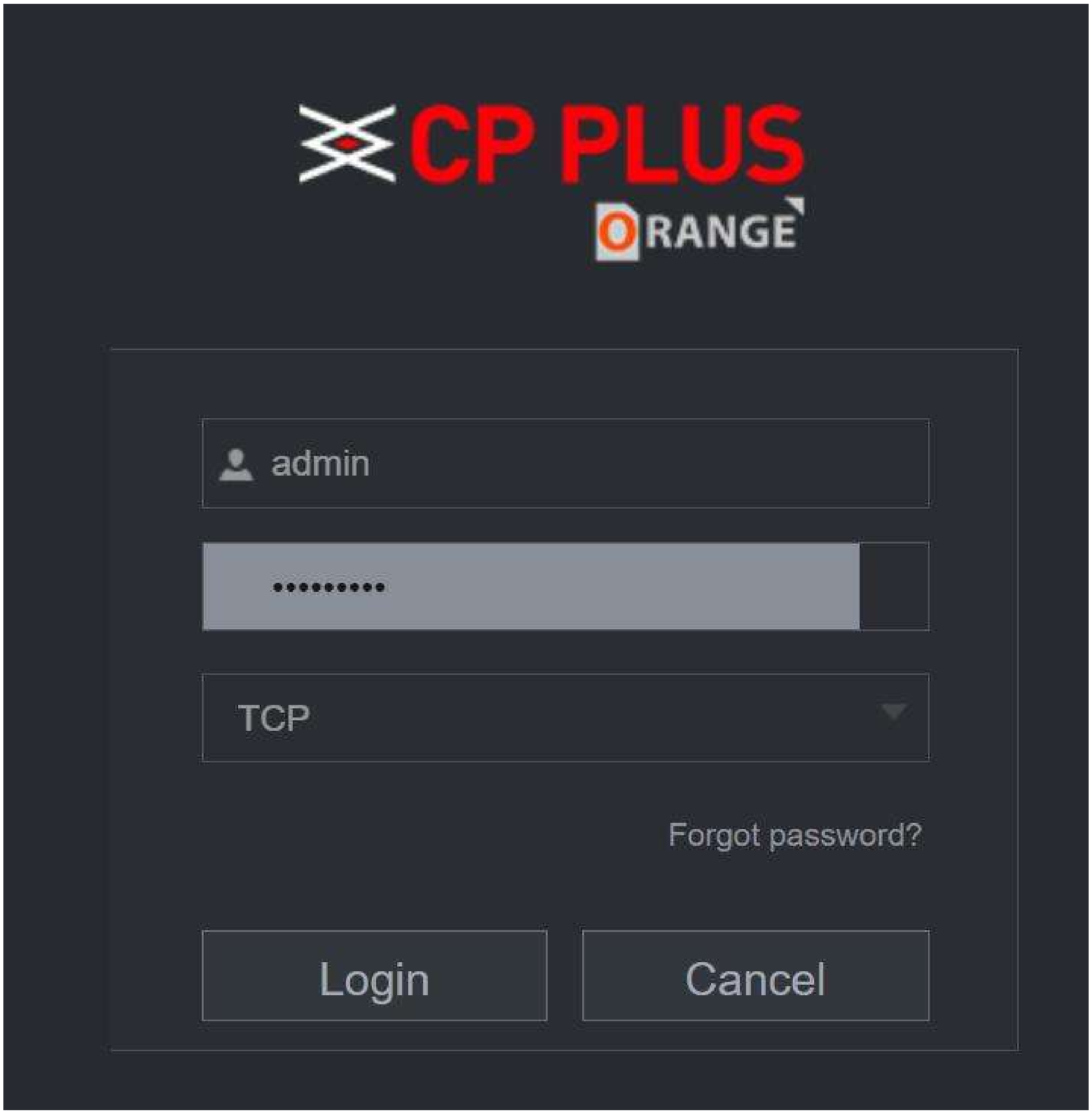
```
HTTP/1.1 200 OK
Date: Tue, 06 Jan 2026 11:48:25 GMT
Last-Modified: Wed, 04 Jun 2025 02:06:05 GMT
Etag: "683faa0d.10830"
Content-Type: text/html
Content-Length: 10830
Connection: keep-alive
Accept-Ranges: bytes
X-Content-Type-Options: nosniff
X-Frame-Options: SAMEORIGIN
x-xss-protection: 1; mode=block
```

CP PLUS ORANGE – Camera Login Interface

The above image shows the **CP PLUS ORANGE IP Camera login interface**, which is used to access and manage live camera feeds and settings.

Description

- The login screen requires a **username** and **password**
- Default user shown is **admin**
- Supports connection over **TCP protocol**
- Provides options such as **Login**, **Cancel**, and **Forgot password**
- Used for configuring and viewing surveillance cameras



admin

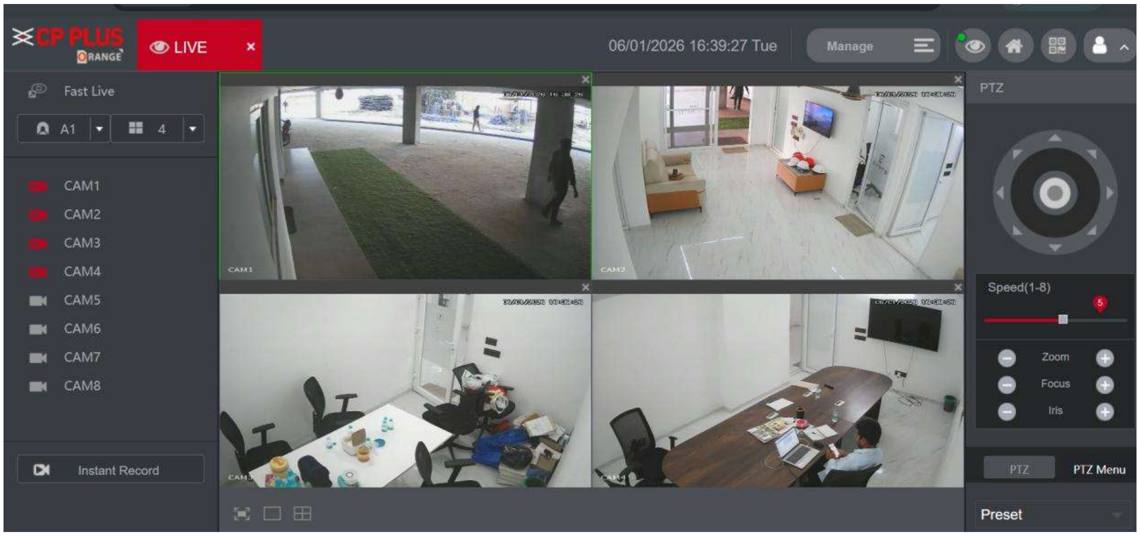
••••••••••

TCP

Purpose

Shodan helps ethical hackers and security teams to:

- Identify exposed services
- Detect misconfigured systems
- Discover vulnerable or unsecured devices



---

7. Key Learnings (Day–6)

- Gained a clear understanding of **Metadata** and its security implications
- Learned how metadata can unintentionally expose sensitive information
- Practiced metadata extraction using **ExifTool**
- Downloaded and analyzed files from **gofile.io**
- Installed and executed **Metagoofil** to extract document metadata
- Performed passive reconnaissance using **theHarvester**
- Used **Shodan search filters** to identify exposed services and devices
- Understood the importance of **ethical and legal boundaries** in information gathering