ETHICAL HACKING · LESSON 2.1

Intro to OSINT

OSINT: the art of finding free information in an oversharing world.

Finding, verifying, and reporting what is already public — legally and ethically.

NINE MODULES · LECTURES · LABS · DISCUSSION

COURSE MAP

Nine Modules

01 OSINT Foundations
What it is, who uses it, and why

02 Legal & Ethical Boundaries
Scope, laws, terms of service, harm

03 OSINT Methodology
The cycle, pivoting, and planning

04 Collection Techniques
Works, images, metadata, DNS

05 OSINT Sources
Twelve categories of sources

06 Dark Web & Threat Intel
Underground sources, researched safely

07 Verification & Reporting
Confidence, evidence, and reports

08 OPSEC
Protecting the investigator

09 Defending Against OSINT
Shrinking your own footprint

BEFORE ANY LAB

Rules for Any Labs

01

Passive only

Look, don't touch. Active techniques only do not try and run scans, scripts or tools against the target.

02

Approved targets only

Yourself, classmates who consent, lab-provided or lab-owned targets, and purpose-built practice sites.

03

Findings stay in the lab

Don't post or share anything you find about real people outside of the lab.

When in doubt: stop, write down what you found, and check before going further.

HOW THE PROFESSIONALS DEFINE IT

Three Definitions

UNITED STATES · DHS

Prompt, and tied to a need

Intelligence derived from publicly available information, collected and disseminated promptly to address specific intelligence needs.

NATO

Public, plus limited distribution

Intelligence obtained from publicly available information and other unclassified data with limited public distribution or access.

EUROPEAN UNION

Collection plus analysis

Collecting and analysing information from open sources to generate actionable intelligence for national security, law enforcement, and business.

Strip out the institutional language and all three say the same thing: public information, analyzed to answer a specific question. With a Scope!

01

OSINT Foundations

What OSINT is, who uses it, and why it matters to attackers and defenders alike.

DEFINITION · THE INTS · HISTORY · USERS · CASE STUDIES · CAREERS

What Counts as OSINT

Open source are sources open to the public this is different from open-source software.

- Web pages: blogs, forums, and news
- Socials: photos, profiles, and video
- Public records: courts, property, filings
- Technical data: DNS, RDAP, certificate logs
- Documents: job ads, patents, manuals, resume
- Broadcasts: podcasts, and conference talks
- Maps: geo-location, imagery, and archived copies of pages
- Grey zone: breach data

Generally open source	Generally not open source
Public websites, social posts	Private accounts you can’t view
News, press releases, broadcasts	Data taken by hacking or stolen credentials
Court, property, and business records	Anything behind someone else’s login
Commercial data anyone can buy	Intercepted communications
DNS, WHOIS, certificate logs	Info gained by pretending to be someone else

Think in source categories, not tool names: tools change, categories last. The 2024 IC OSINT Strategy calls OSINT “the INT of first resort.”

Test: could an ordinary person lawfully see, request, or buy it without special access or deception?

Data, Information, Intelligence

The question we were asked: how exposed is Security-Inc to credential phishing?

Data

Raw facts

Six separate things we found. None of them mean anything alone.

- MX records point to Microsoft
- 14 IT staff listed on LinkedIn
- Contact page shows jane.doe@
- vpn.security-inc.com in cert logs
- Job ad: M365, Okta; new CFO named

Information

Organized, in context

The same facts, sorted and joined up. Now they describe a company.

- Runs Microsoft 365 and Okta
- Email format is first.last@
- Has an internet-facing VPN
- Has a brand-new CFO

Intelligence

Answers “so what?”

What it means for the question, and what someone should do next.

- Highly exposed to credential phishing
- Fake M365 and Okta pages would work
- New CFO is a likely BEC target
- Confidence: moderate
- Act: phishing-resistant MFA first

Test: Read your finding and ask “so what?” If you can’t answer, you have information, not intelligence.

The Intelligence Disciplines

OSINT
Open Source Intelligence

Publicly or commercially available information

e.g., news, social media, public records

HUMINT
Human Intelligence

People

e.g., interviews, informants, debriefings

SIGINT
Signals Intelligence

Intercepted signals and communications

e.g., intercepted radio or network traffic

GEOINT
Geospatial Intelligence

Imagery and geospatial data (includes IMINT)

e.g., satellite imagery, maps

MASINT
Measurement and Signature Intelligence

Technical measurements

e.g., radar, acoustic, chemical signatures

Related specialties

SOCMINT (social media) · FININT (financial) · CTI (cyber threat intelligence)

Where they overlap

Commercial satellite imagery is both OSINT and GEOINT. OSINT often tells the other disciplines where to look.

MODULE 1 · OSINT FOUNDATIONS

A Short History of OSINT

1941 · WWII

Listen and translate. Clerks transcribe Axis radio by hand.

1947 · CIA

Read everything. FBIS joins the CIA; BBC Monitoring indexes TV.

1995 · Web

Search it yourself. Usenet, mailing lists, first crawlers.

2004 · Google

Query, don't browse. Dorking and the GHDB arrive.

2014 · Social

Everyone is a sensor. Bellingcat verifies MH17 from posts.

2024 · AI

Automate, then verify. "The INT of first resort."

Volume

A few hundred broadcasts a day became billions of pages, posts, and records.

Speed

Weeks of translation and mail became minutes of search. Events get verified live.

Access

Governments only, then anyone with a browser. AI collects; humans still verify.

Unchanged since 1941: monitor, translate, analyze, report. What changed is the volume, the speed, and who is allowed to do the work.

Who Uses OSINT

Group	Goal	Example
Pentesters & red teams	Find weaknesses, with permission	Map staff and subdomains for an authorized test
Attackers	Money, access, espionage	Plan a BEC scam from public posts alone
Defenders (SOC, CTI, ASM)	Find exposure before attackers	Weekly check for lookalike domains and leaks
Journalists & researchers	Verify events, expose wrongdoing	Geolocate and chronolocate a viral video
Law enforcement	Investigate crimes, find people	Develop leads from a suspect's public posts
Businesses	Due diligence, fraud, brand safety	Vet a vendor's filings, lawsuits, and breaches
Volunteers & nonprofits	Missing persons, accountability	Trace Labs Search Parties, Bellingcat projects
Everyday people	Safety and trust	Vet a contractor or rental listing

OSINT is research on steroids. Humans like to share; we post our lives online, and all of it eventually gets saved.

MODULE 1 · OSINT FOUNDATIONS

Reconnaissance Comes First

Every model of how an attack unfolds starts in the same place: learning about the target.

RECON → WEAPONIZATION → DELIVERY → EXPLOITATION → INSTALLATION → C2 → ACTIONS ON OBJECTIVES

2011 · LOCKHEED

Cyber Kill Chain

Seven phases. Phase 1 is Reconnaissance: pick a target, research it, find weaknesses.

lockheedmartin.com

2013 · MITRE

MITRE ATT&CK

Reconnaissance is a whole tactic, TA0043, with 10 techniques — nearly all outside your network.

attack.mitre.org

2017 · PAUL POLS

Unified Kill Chain

18 phases merging ATT&CK and the Kill Chain. Recon opens the Initial Foothold stage.

unifiedkillchain.com

2009 · PTES

Intel Gathering

The pentest standard makes Intelligence Gathering its own phase, before any testing.

pentest-standard.org

Break the chain early: if attackers learn little in phase one, every later phase gets harder — and most of phase one happens outside your network.

MODULE 1 · WHY OSINT MATTERS

62%

of breaches involved the human element, up from 60% the year before.

OSINT is what makes phishing and pretexts convincing.

The weakest part of any security system is you.

“More than 90% of successful cyber attacks start with a phishing email.”

— CISA, Phishing Infographic

Sources: Verizon 2026 Data Breach Investigations Report · CISA

How It Reaches You

YOUR PHONE NUMBER

SIM swapping

Your number, carrier, birthday and old addresses are all for sale. One convincing call moves your number onto their SIM, and every SMS code follows it: bank, email, crypto, Instagram.

THE OLDEST ONE

The “African Prince”

Advance-fee fraud is older than the web. OSINT made it personal: it now arrives as a fake internship, a scholarship, a campus job, or a match who knows your major and your home town.

LEAKS THAT ARE YOURS

Breaches and stealer logs

University portals, loan servicers and people-data brokers all get breached. Cracked software and game cheats install infostealers that take saved passwords and live sessions.

None of this needs a nation-state. It needs your name, your number, and ten minutes of public information.

02

Legal & Ethical Boundaries

What you're allowed to do, what the law says, and what you should do even when something is legal.

AUTHORIZATION · SCOPE · LAWS · TERMS OF SERVICE · ETHICS

Authorization Comes First

An attacker and a pentester run the same searches. The difference is a signed agreement from someone with the authority to approve the work.

Why it matters for passive work

- OSINT findings usually feed active steps
- Collecting personal data needs a purpose
- The written scope is your proof

Who can the client authorize?

YES — **Systems the client owns**
Its domains, sites, and infrastructure, as the scope lists them.

NO — **Vendors and cloud providers**
Third parties need to give their own consent. Many publish testing policies.

NO — **Employees' personal lives**
A company can't consent for staff's personal accounts, homes, or families.

What an OSINT Scope Covers

1 Objective
The question the client needs answered. It limits everything else.

2 Organizations
Legal entities, subsidiaries, domains, brands, IP ranges, locations.

3 People
All staff or specific roles. Former employees and contractors?

4 Personal accounts
Limited at most. Family members and minors: out of scope.

5 Techniques
Passive, semi-passive, active, social engineering, sock puppets?

6 Sources
Breach data, dark web platforms, paid data brokers: allowed or not?

7 Data handling
Storage, encryption, deletion date, and what goes in the report.

8 Escalation
Contacts, and what to do if you find a breach, a crime, or danger.

9 Time window
Start and end dates. Data outside the window may be out of scope.

Levels of Personal Information

Level	What it covers	Collect?	Goes in the report?
1 · Professional	Name, job title, employer, work email and phone, public bio	Yes	Yes
2 · Public personal	Usernames, personal accounts and site, photos they published	If the scope allows it	As roles and counts, not names
3 · Identifying	Home address, personal phone, date of birth, family, daily routine	Only if the objective needs it	Restricted appendix at most
4 · Sensitive	Health, religion, sexuality, finances, immigration status, minors	No	Never

You may have to look at a level to answer the question. That is not the same as putting it in the report — collect narrow, report narrower.

The Computer Fraud and Abuse Act

18 U.S.C. § 1030 · 1986

The CFAA

The main US federal computer crime law. It covers access “without authorization” and “exceeding authorized access,” and lets companies sue.

Supreme Court · 6–3

Van Buren (2021)

“Gates-up-or-down”: were you allowed into the system, or that part of it? Your motive doesn't change the answer.

May 19, 2022

DOJ policy (2022)

Prosecutors should decline good-faith security research. It doesn't block lawsuits or state charges.

LOWER RISK

Reading information you're allowed to see, even for a purpose the site owner doesn't like.

STILL COVERED

Going through a gate you weren't given: someone else's login, an admin panel, a restricted database.

Privacy and Harassment Laws

Law	What it covers	What it means for OSINT
GDPR (EU)	Any information about an identifiable person	Public personal data is still personal data
CCPA / CPRA (California)	Personal information businesses hold about residents	Excludes “publicly available” information; other states differ
FCRA (US)	Data used for job, housing, credit, or insurance decisions	Don't use people-search data to screen people
Voter data rules	Public voter records with use limits	California bans personal or commercial use and harassment
Stalking and doxxing laws	18 U.S.C. § 2261A and a growing number of state laws	Compiling public info to harass or intimidate can be a crime

Not legal advice. Laws vary by state and country, so ask a lawyer before real engagements.

Read the Terms of Service

What to look for

Acceptable use
The platform's list of what you may not do

Automated access
Whether scrapers, bots, and scripts are allowed

Account rules
Real-name and one-account rules affect sock puppets

APIs and research programs
The approved path for bulk collection

Prohibitions hint at what's possible

“No collecting data using automated means”
Suggests: data can be collected at scale

“No copying cookies to access others' accounts”
Suggests: session cookies can grant account access

“No using the service to track someone's location”
Suggests: location data exists somewhere in the platform

Possible is not permitted. If the terms say no, change your method or source, or get permission.

MODULE 2 · WAR STORY

The Campus Directory and the Campaign Email

AS TOLD FROM MEMORY · AN EXAMPLE OF WHAT NOT TO DO

He never broke into anything

Adam was running a campaign and wanted to reach everyone at his school. He scripted the public search box on Facebook and on the school site — type one letter, take the twenty names it returns, repeat until he had enumerated the student and faculty roll into his own mailing list. Then he borrowed a friend's login on a lab machine and sent the blast from a trusted internal address.

Facebook not the school spotted the scraping and warned him. He complied within minutes, terrified of losing the account his whole social life ran on. Adam faced no consequences. His friend nearly got expelled. His defence: faculty send mass emails all the time, so why couldn't he.

"Nothing was hacked" is not a defence	Authorization cannot be borrowed	Role permission does not transfer	Bulk enumeration is loud

Public ≠ permitted: automating a public feature is still collection at scale, and you cannot borrow someone else's authorization.

Scraping Cases

Case	What happened	Outcome
hiQ Labs v. LinkedIn (2017–2022)	Scraped public profiles. Breached LinkedIn's terms through scraping and fake profiles.	Dec 2022: $500,000, permanent injunction, data deleted
Meta v. Bright Data (2023–2024)	Sold Facebook and Instagram data scraped while logged off	Jan 2024: Meta's terms didn't cover logged-off scraping
X Corp. v. Bright Data (2023–2024)	Scraped and sold public data from X	May 2024: X's contract claims preempted by copyright law

LOW RISK

Viewing while logged out

CONTESTED

Scraping while logged out

TERMS APPLY

Collecting while logged in

HIGH RISK

Using fake accounts

MODULE 2 · THE CORE RULE

Public is not the same as permitted.

Legal is the floor, not the ceiling. What matters is how you got it and what you do with it.

The Mosaic Effect

Five public facts about “Alex” (fictional)

Fitness app	Runs a 5K loop weekdays at 6 a.m.
Dog photo	The dog's name is Biscuit
Local news	Won a garden contest on Maple Street
Work profile	Works night shifts at a hospital
Old post	Birthday post names mom, Linda Carter



Combined, they reveal

Where Alex lives
Contest street plus where the run starts and ends

When the home is empty
Running schedule plus night shifts

Security question answers
Pet's name and mother's maiden name

No single fact is sensitive. The compilation is.

MODULE 2 · LEGAL & ETHICAL BOUNDARIES

Check, Download, Use, Buy

Action	Example	Position
Check	Looking up an email in Have I Been Pwned	Generally fine: you learn that exposure happened
Search (licensed tool)	Threat intel platform shows exposed accounts	OK with authorization; handle as sensitive
Download	Grabbing a breach dump from a forum	stolen data, illegal content, malware
Use	Logging in with a leaked password	Unauthorized access without explicit permission
Buy	Paying a seller for a dataset	Never: funds criminals, serious legal risk (Grey Hat area)

No plaintext passwords

Report “credential exposed for user X on date Y” in notes, tickets, chat, and reports.

Encrypt and delete

Store results encrypted, limit access, and delete them on schedule.

Only what you need

Stealer logs hold victims' personal data. Look only at what the objective needs.

Ethical Principles

Minimization
Collect only what the objective needs, and keep it only as long as needed.

Proportionality
Match how deep you dig to how important the purpose is.

Sensitive data
Health, religion, sexuality, home address: almost never belongs in a report.

Minors
Don't collect, include, or pivot from a minor’s information.

Responsible disclosure
Found an exposure? Stop, don't use it unless authorized, document it, and report it privately.

Berkeley Protocol
Professional, methodological, and ethical principles for open source investigations (2020).

Remember: Investigating someone is not the same as exposing them. Doxxing causes real harm.

Discussion: What Would You Do?

SCENARIO

You're doing OSINT on Security-Inc for an authorized pentest. You find the CEO's teenager's Instagram, which shows his Desktop with his personal IP.

Think about: family and minors, home addresses, minimization, and when to stop and ask.

1 Is it in scope?
What does the scope sheet say about family?

2 Does it go in the report?
What does the objective actually need?

3 What would you do next?
Keep going, leave it, or stop and ask?

Not legal advice. Laws vary by state and country.

03

OSINT Methodology

A repeatable way to turn a question into verified, defensible findings.

THE OSINT CYCLE · PIVOTING · PASSIVE VS. ACTIVE · PLANNING

MODULE 3 · OSINT METHODOLOGY

The OSINT Cycle

1 Planning & Direction

Define the decision, the questions, the scope, and the time budget.

2 Collection

Gather raw data that can answer the requirements. Record provenance.

3 Processing

De-duplicate, normalize, extract metadata, translate, and tag.

4 Analysis

Connect, corroborate, find gaps, and state a confidence level.

5 Reporting

Deliver findings the audience can act on, securely and on time.

6 Review

Did it answer the question? What new questions came up?

Remember: Planning is the most skipped phase and the most important. Review creates new requirements, and the cycle repeats.

Writing Intelligence Requirements

Quality	Vague	Better
Specific	"Find info on Security-Inc"	"What email address format does Security-Inc use?"
Answerable	"What's Security-Inc's internal network map?"	"What internet-facing hosts and subdomains does Security-Inc have?"
Tied to a decision	"Who works there?"	"Which IT and help desk staff are publicly identifiable, for the authorized phishing simulation?"
Bounded	"Everything about the CEO"	"What does the CEO publicly post that an attacker could use for a pretext?"

Before collecting anything, write down:

- Scope and exclusions
- Allowed methods
- Time budget
- Definition of done
- OPSEC level

MODULE 3 · OSINT METHODOLOGY

Plan Before You Search

A Scope Sheet
Targets, exclusions, methods, data handling

B Requirements
Question, why, done-when rule, time box

C Collection Plan
Each requirement mapped to sources

D Source Checklist
Tick every source category you considered

E Pivot Log
From, method, to, confidence, in scope?

F Parking Lot
Interesting leads: pursue, drop, or ask

G Time Budget
Budget vs. used for each requirement

H Findings
Finding, confidence, pivots, gaps

Start here: fill out the scope sheet before you open a browser. The findings table bridges your notes and the final report.

Time Boxes and Parking Lots

OSINT is addictive: every result suggests three more searches. Discipline keeps you on the question and inside scope.

Before every pivot, finish this sentence:

”This helps answer requirement Requiremnet_ because ___.”

Can't finish it? Stop, or park the lead.

Time-box each requirement

E.g., 45 minutes. When time is up, decide consciously.

Use a parking lot

Write off-target leads down. Review them at the end.

Set stopping rules first

"Stop at three independent confirming examples."

Watch for scope creep

Family members, home addresses, unrelated accounts.

Pivoting

Use what you know to find what's connected. Each new selector starts a mini-cycle.

Domain
`security-inc.com`
from the business filing

→

Email
`contact@sec-inc.com`
contact page reveals the format

→

Person
`Jane Doe`
PDF author field, then a profile: IT Manager

→

Technology
`firewall vendor`
profile skills, confirmed by a job posting

Log every pivot

Number, method, confidence, scope decision, evidence.

Weakest link wins

A conclusion is only as confident as its weakest link.

Record dead ends

Most branches die. Note why, so no one repeats them.

Passive, Semi-Passive, Active

PASSIVE

Look from the street

Query third parties that already hold the data.

Search engines, WHOIS/RDAP, CT logs, scan databases, archives

Target sees it: no

SEMI-PASSIVE

Walk in like a customer

Normal-looking traffic to the target's own systems.

Browsing the website, reading robots.txt, standard DNS lookups

Target sees it: in logs, among normal visitors

ACTIVE · NEEDS AUTHORIZATION

Jiggle the doorknobs

Directly probe the target's systems or people.

Port and vulnerability scans, directory brute forcing, pretext calls

Target sees it: yes, logged and alerted

The rule: passive by default. Active only with written authorization that covers the technique and the targets.

Strong and Weak Selectors

Selector	Common pivots	Strength
Email address	Username, breach exposure, accounts, email format	Strong (unique)
PGP key / contact ID	Other forums or sites using the same key or ID	Strong (unique)
Crypto wallet	Transactions, linked addresses, exchanges	Strong (unique)
Domain	Subdomains, IPs, email format, registrar, archives	Strong (unique)
Username	Same handle elsewhere, old posts, photos	Weak alone (shared)
IP address	Hosting provider, ASN, co-hosted domains	Weak on shared hosting
Real name	Profiles, publications, records, news	Weakest alone

Rule of thumb: require one strong link, or several independent weak links, before you connect two things.

Who Sees Your Activity?

Activity	Type	Target sees it?	Authorization?
Search engine dorking	Passive	No	No
Certificate Transparency search	Passive	No	No
Scan database search	Passive	No	No (follow ToS)
Viewing a public profile	Passive	Sometimes ("who viewed")	Check scope
Browsing the target's website	Semi-passive	Possibly (server logs)	Normally no
Port or vulnerability scan	Active	Yes (IDS alerts)	Yes
Directory brute force	Active	Yes (many 404s)	Yes
Emailing or calling staff	Active	Yes (the person)	Yes

MITRE ATT&CK Reconnaissance (TA0043) labels these too, e.g., T1593 Search Open Websites/Domains and T1595 Active Scanning.

04

Collection Techniques

The hands-on how-to: precise queries, profiles, images, documents, and infrastructure.

DORKING · USERNAMES · IMAGES · METADATA · INFRASTRUCTURE · AUTOMATION

Search Operators That Still Work

Operator	What it does	Example
" "	Exact phrase match	"quarterly revenue report"
-	Exclude a term	jaguar speed -car
site:	Restrict to one domain or subdomain	site:security-inc.com
filetype: / ext:	Restrict to a file extension	filetype:pdf
intitle:	Term appears in the page title	intitle:"index of"
inurl:	Term appears in the URL	inurl:login
intext:	Term appears in the body text	intext:"internal use only"
before: / after:	Date range	after:2024-01-01 before:2025-01-01
OR and ( )	Match either term; group terms	(filetype:pdf OR filetype:docx)

Bing bonus: ip:93.184.216.34 finds other sites hosted on the same IP address, which Google can't do.

Building a Dork

```
# 1. Anchor to the target
site:security-inc.com
# 2. Add a content type
site:security-inc.com filetype:pdf
# 3. Add a topic keyword
site:security-inc.com filetype:pdf intext:confidential
# 4. Exclude noise
site:security-inc.com filetype:pdf intext:confidential -template
# 5. Bound the time window if it helps
site:security-inc.com filetype:pdf intext:confidential
after:2024-01-01
```

DON'T RELY ON

Retired operators

`cache: related: info: link: + ~`

Old page? Use the Wayback Machine instead of cache:.

Borrow from the GHDB

Copy proven patterns from the Google Hacking Database, then scope them with site:security-inc.com.

MODULE 4 · COLLECTION TECHNIQUES

Reading a Profile

01

Stated facts

What the person typed: name, employer, location, links. Treat these as claims.

"First day at the new job!"

02

Structural facts

What the platform records: account age, posting cadence, followers, tags. Harder to fake.

Account created years ago; posts weekly.

03

Incidental facts

What leaked in by accident: signs, reflections, badges, screens, weather. Often the most valuable.

A lanyard shows the building name.

Ethics checkpoint: pattern-of-life analysis rebuilds someone's routine from timestamped posts. In this class, only run it on yourself, a consenting classmate, or a fictional persona.

MODULE 4 · COLLECTION TECHNIQUES

Username Enumeration

WhatsMyName

Web app · no install

Community dataset behind many tools; free app at whatsmyname.app.

Sherlock

400+ sites

Command-line hunt across social networks, with CSV and JSON output.

github.com/sherlock-project/sherlock

Maigret

3,000+ sites

Widest coverage, HTML/PDF reports, and pivots on what it finds.

github.com/soxoj/maigret

```
# Sherlock
pipx install sherlock-project
sherlock user123 --csv
# Maigret, with an HTML report
pip install maigret
maigret user --html
```

Hits are candidates

Two people can share a handle, and some sites falsely report a match. Confirm with photos, bios, and linked accounts first.

MODULE 4 · COLLECTION TECHNIQUES

Four Questions for Every Image

01

Where did it come from?

Reverse image search on several engines; follow the oldest copy.

02

What does the file say?

EXIF metadata: device, timestamps, editing software, GPS.

03

Where was it taken?

Geolocation: reason from clues, then confirm on maps.

04

When was it taken?

Chronolocation: shadows, SunCalc, and weather records.

```
# Just the GPS tags
exiftool -gpslatitude -gpslongitude -gpsposition photo.jpg
# Plain decimal degrees to paste into a map
exiftool -n -gpslatitude -gpslongitude photo.jpg
```

Platforms strip EXIF

Instagram, Facebook, X, and TikTok remove it on upload. Original files on websites, forums, and file shares often keep GPS.

MODULE 4 · COLLECTION TECHNIQUES

Geolocation and Chronolocation

Where: geolocation

01 **Inventory the clues:** signs, plates, buildings, plants

02 **Narrow the region first:** language, driving side, climate

03 **Search the specifics:** business names, phone formats

04 **Confirm on the map:** satellite and street-level imagery

05 **Query map data:** Overpass Turbo finds feature combinations

When: chronolocation

Shadows point away from the sun; their length shows its height.

SunCalc models the sun for any place and date. Slide the time until the shadows match.

Weather records confirm the date.

BEFORE YOU REPORT

Get two independent lines that agree, and rule out look-alike places like chain stores.

Document Metadata Workflow

```
# 1. Find what the domain published
site:security-inc.com filetype:pdf
# 2. Gather at scale (metagoofil downloads; it no longer analyzes)
metagoofil -d security-inc.com -t pdf,docx,xlsx -l 100 -n 25 -o ./security-inc-docs -w
# 3. Analyze with exiftool
exiftool -r -csv -Author -Creator -Producer ./security-inc-docs/ > security-inc-metadata.csv
# 4. Sanitize before publishing (the defender's job)
exiftool -all= -overwrite_original brochure.pdf
```

AUTHOR

Real employee names and internal usernames

CREATOR / PRODUCER

Exact software and version: patch-level hints

FILE PATHS

Usernames, server names, and share names

DATES

Activity timelines and document age

Passive Infrastructure Recon Commands

```
# Registration data (RDAP replaced WHOIS for gTLDs in January 2025)
curl -s -L https://rdap.org/domain/example.com | jq .
# DNS: mail provider and SaaS verification tokens
dig security-inc.com MX +short
dig security-inc.com TXT +short
# Subdomains from Certificate Transparency logs and aggregators
curl -s "https://crt.sh/?q=%25.security-inc.com&output=json" | jq -r '.[].name_value' | sort -u
subfinder -d security-inc.com -silent
# Exposed services: Shodan did the scanning, not you
shodan search hostname:security-inc.com
# Forgotten URLs from the Wayback Machine
echo security-inc.com | waybackurls
```

Man jq -r and dig!

Automation Frameworks

theHarvester

Fast first pass: emails, subdomains and hosts from many sources. github.com/laramies/theHarvester

```
theHarvester -d security-inc.com -l 500 -b duckduckgo
```

SpiderFoot

Broad sweep that correlates results; pick the passive use case. spiderfoot.net

```
spiderfoot -s security-inc.com -u passive -o json
```

Recon-ng

Console with a module marketplace and a workspace database for repeatable reports. github.com/lanmaster53/recon-ng

```
recon-ng -w security-inc
```

Maltego CE

Graph view: transforms expand entities into a map of relationships. maltego.com · free tier, capped, non-commercial

Remember: frameworks collect candidates; humans verify and decide what matters. API keys multiply results, so guard them like passwords, and point every tool only at authorized targets.

05

OSINT Sources

Where information lives, what each source reveals, and how attackers and defenders use it.

SEARCH · SOCIAL · ORGS · CODE · INFRASTRUCTURE · ARCHIVES · RECORDS · IMAGERY · BREACHES

MODULE 5 · OSINT SOURCES

Twelve Source Categories

01
Search Engines
Different indexes

02
Social Media
People self-document

03
Organizations
Sites, docs, job posts

04
Code Repos
Secrets and identities

05
Infrastructure
DNS, certs, scans

06
Archives
Deleted isn't gone

07
Public Records
Authoritative, hard to fake

08
Data Brokers
Pre-aggregated people

09
Imagery & Maps
Where and when

10
Breach Data
Check, never use

11
News & Genealogy
Context and family

12
Directories
Keep your kit current

Direct and Indirect Sources

THE TARGET PUBLISHED IT

Direct sources

Their own website, job ads, filings, press releases, official accounts and repos. Authoritative but curated. They chose what to say, and they can often see who reads it.

SOMEONE ELSE PUBLISHED IT

Indirect sources

Vendor case studies, conference talks, podcasts, partner sites, reviews, court records, résumés, local news, regulators. Uncurated, rarely scrubbed, and outside the target's view.

UNEXPECTED PLACES THAT PAY OFF

- A vendor case study naming the exact firewall
- A conference talk with an internal architecture slide
- A podcast interview dating a cloud migration
- A résumé listing internal system names and versions
- A supplier press release naming the client
- Employee reviews naming teams, tools, and reorgs

Search Engines

Engine	Own index?	Strengths for OSINT	Cost
Google · google.com	Yes	Largest index; most operators; strong file indexing	Free
Bing · bing.com	Yes	Different index; ip: finds sites on the same server	Free
DuckDuckGo · duckduckgo.com	Mostly Bing	Privacy: no search history tied to you	Free
Yandex · yandex.com	Yes	Russian and Eastern European content; image matching	Free
Baidu · baidu.com	Yes	Chinese-language web	Free
Brave Search · search.brave.com	Yes	Fully independent Western index since 2023	Free
Mojeek · mojeek.com	Yes	Small UK index; surfaces pages big engines bury	Free

Tips: A search engine is one company's index, not the whole web, so run the same query on several. Google's cache: operator is gone (2024): use web archives. AI search summaries are not sources; open the original page.

Social Media Platforms

Platform	What's public	Watch for
LinkedIn	Employees, titles, skills (tech stack), org charts	Profile views notify, unless private mode
Facebook	Friends, workplace, check-ins, groups, Marketplace	Graph Search shut down in 2019
Instagram	Photos, location tags, tagged photos, followers	Story viewers are visible to the owner
X	Posts, replies, networks, bio, join date	Free API ended in 2023
TikTok, YouTube	Video backgrounds, audio, exact upload dates	Videos get deleted; capture early
Reddit	Pseudonymous posts; local subreddits hint at location	Paid API and Pushshift loss (2023)
Discord	Almost nothing without joining a server	Joining is an interaction, not passive
Telegram	Public channels via t.me/s/ with no account	Private chats are not OSINT sources
Bluesky	Posts, likes, and follows as public records	Free firehose; deleted posts may persist

What Tips Off the Target?

Can the target notice?

Action	Notices?
Viewing a LinkedIn profile (not private mode)	Yes
Viewing an Instagram or Facebook story	Yes
Liking, following, friending, or joining	Yes
Viewing a public Strava activity	Usually no
Viewing a public X post or Reddit comment	Generally no
Reading a public channel via t.me/s/	No

Non-social apps leak too

Strava
Routes often start and end at home or work

Venmo
Payments and friend lists public by default

Google Maps reviews
A reviewer's profile maps where they go

Takeaway: Photos reveal more than captions: badges, screens, signs, and landmarks.

Organization-Owned Sources

Company Website

Team pages, locations, clients. One contact email reveals the format for everyone.

Hosted Documents

Metadata leaks usernames, software versions, and internal file paths.

Job Postings

A technology inventory: security tools, cloud, migrations in progress.

Control Files

robots.txt lists paths the owner hid; security.txt shows how to report bugs.

ToS & Privacy Policies

Vendor and subprocessor lists; prohibitions hint at what's possible.

Press & Case Studies

Launches, acquisitions, and vendor pages on how Security-Inc uses their product.

Takeaway: "You may not collect profile data at scale" means it can be done. A ToS prohibition is a lead, not permission: possible is not permitted.

MODULE 5 · OSINT SOURCES

Code Repositories

28.6M

new hardcoded secrets added to public GitHub commits in 2025

Source: GitGuardian, State of Secrets Sprawl 2026

Rule: Finding a secret is never permission to use it. Document, report, rotate.

Code and config files

Hardcoded keys, .env and CI files, internal hostnames

Commit metadata

Author name and email; .patch on a commit URL shows them

History

Secrets deleted later still sit in earlier commits

Beyond GitHub

Gists, paste sites, Docker image layers, Postman workspaces

DNS Records Reveal Vendors

Record	What it holds	What it reveals
A / AAAA	IPv4 / IPv6 address	Hosting provider or cloud region
MX	Mail servers	Email provider: Google Workspace, Microsoft 365, Proofpoint
TXT	Text records	SPF/DMARC; verification tokens name SaaS tools in use
CNAME	Alias to another name	CDNs, help desks, marketing platforms, status pages
NS	Name servers	Who runs the organization's DNS
SOA	Zone authority info	Admin email and zone serial
PTR	Reverse DNS	Naming conventions on IP ranges

The TXT trick: every token like google-site-verification or atlassian-domain-verification marks a SaaS tool the org onboarded. Simple lookups are effectively passive; subdomain brute forcing and zone transfers are active.

Scan Engines and Certificate Logs

Certificate Transparency

Search %.security-inc.com on crt.sh: public certificates expose vpn., dev., and staging. hosts.

Passive Subdomains & DNS History

subfinder (github.com/projectdiscovery), amass (owasp.org/www-project-amass) and theHarvester; virustotal.com and dnsdumpster.com show past records.

Scan search engine	Focus
Shodan · shodan.io	Internet-connected devices, service banners, IoT, industrial systems, webcams
Censys · search.censys.io	Hosts, open ports, and certificates
ZoomEye, FOFA · zoomeye.ai · fofa.info	Devices and services, with strong Asia coverage
GreyNoise · greynoise.io	Who is scanning the internet (mostly defensive)

Why it's passive: these engines already did the scanning, so searching sends nothing to the target.

Ownership, Fingerprints, and Traps

WHO OWNS IT

IP Ownership

Regional internet registries (ARIN, RIPE NCC, APNIC, LACNIC, AFRINIC) and bgp.he.net show who holds an IP range and ASN.

WHAT IT RUNS

Technology Fingerprinting

builtwith.com shows current and past site technologies. wappalyzer.com inspects pages you load, so it is semi-passive.

OPSEC TRAP

URL Scanners

virustotal.com and urlscan.io show screenshots and related infrastructure. Submitted scans can be public: choose private or unlisted.

LEGAL LINE

Exposed Cloud Storage

grayhatwarfare.com indexes open S3 and Azure buckets. Finding one is not permission to download from it: note it, report it, stop.

Archives and Historical Data

Wayback Machine

Decades of snapshots at web.archive.org. The CDX API lists every capture for a domain; waybackurls and gau automate it.

archive.today

archive.ph (also .today / .is): on-demand captures with a screenshot, handles JavaScript-heavy pages, and often holds snapshots Wayback lacks.

Common Crawl

commoncrawl.org: a free, massive crawl dataset with a searchable URL index, built for research at scale.

Historical Imagery

Google Earth Pro (earth.google.com) has a time slider, and Street View has “See more dates”, to show what a place used to look like.

Remember: "Deleted" rarely means gone. Reading an archive is passive; browsing the live site is not. Archives have gaps, so record capture timestamps and corroborate.

Public Records

Category	Main sources	What it reveals
Property	County assessor and recorder · netronline.com	Owner, mailing address, purchase date and price
Federal courts	courtlistener.com (RECAP), then pacer.uscourts.gov	Lawsuits, bankruptcies, criminal cases
Businesses	Secretary of State sites · opencorporates.com	Officers, registered agents, addresses
Public companies	SEC EDGAR · sec.gov/edgar (10-K, 8-K Item 1.05)	Executives, risk factors, cyber incidents
Nonprofits	ProPublica · projects.propublica.org/nonprofits	Executives, salaries, finances
Campaign finance	fec.gov	Donor name, city, employer, occupation
Patents	ppubs.uspto.gov · patents.google.com	Inventors, assignees, R&D focus
Wireless devices	fcc.gov/oet/ea/fccid · fccid.io	Internal photos, manuals, test reports
Aircraft, vessels	registry.faa.gov · adsbexchange.com · marinetraffic.com	Owners, flight paths, ship positions
Voter records	State election offices	Name, address, party; use is restricted

MODULE 5 · OSINT SOURCES

People-Search Sites and Data Brokers

People-Search Sites

whitepages.com, spokeo.com, beenverified.com. Anyone can search; quality varies widely.

Marketing Brokers

acxiom.com, epsilon.com. Businesses buy interests, demographics and purchasing data.

Investigator Databases

Accurint (lexisnexis.com), TLOxp (tlo.com), CLEAR (thomsonreuters.com). Vetted customers with a permissible purpose only.

Leads, not facts

Identities get merged and errors copy across sites. The FTC's $5.8M TruthFinder/Instant Checkmate settlement (2023) shows the accuracy problem.

Legal limits

FCRA: people-search sites generally can't be used for employment, housing, or credit decisions. California's DROP sends one deletion request to every registered broker.

Email pivots: hunter.io for company formats; gravatar.com and epieos.com for linked profiles. In class, only look up yourself.

MODULE 5 · OSINT SOURCES

Images, Maps, and Geospatial Sources

Question	Start with
Where did this image first appear?	tineye.com (sort by Oldest), lens.google.com, yandex.com, bing.com
Where was this street photo taken?	Google Street View, yandex.com/maps, Baidu, Naver, mapillary.com
Did this building exist in 2019?	earth.google.com history, Street View “See more dates”
What changed at this site last week?	browser.dataspace.copernicus.eu (Sentinel-2, 10 m)
Where are these features together?	overpass-turbo.eu · bellingcat.com OpenStreetMap search
Is the claimed time of day right?	suncalc.org (sun position and shadows)
Is the claimed weather right?	meteostat.net, historical weather records
Is this video a repost?	invid-project.eu keyframes + reverse image search

Off-limits in class: face-search engines. They exist, but raise serious legal and ethical concerns.

Breach and Leak Data

Types of leaked data

Type	What it is	Where you see it
Data breach	Data stolen from one organization	Breach forums, HIBP, vendor notices
Combolist	Credential pairs compiled for stuffing	Telegram channels, paste sites
Stealer logs	Passwords and cookies from infected PCs	Telegram log shops, Hudson Rock
Paste leaks	Samples or ads for larger dumps	Pastebin clones, Intelligence X
Ransomware leaks	Files published when a victim won't pay	Leak sites, ransomware.live

CHECK

Yes: your own or authorized accounts

DOWNLOAD

No: legal risk, malware risk, stolen data

USE

Never: that is unauthorized access

Lookup tools: haveibeenpwned.com · hudsonrock.com (stealer logs) · socradar.io · dehashed.com · intelx.io. Never copy plaintext passwords into notes or reports.

News, Genealogy, and Tool Directories

News & Archives

Local news has the most detail. chroniclingamerica.loc.gov, news.google.com/newspapers, newspapers.com, prnewswire.com. Example: a 2013 story names the plant

Academic & Talks

scholar.google.com, core.ac.uk, university theses, slideshare.net, conference talks and podcasts. Example: a DEF CON slide shows the speaker’s internal network diagram.

Genealogy

familysearch.org, ancestry.com, findagrave.com, legacy.com obituaries. Example: an obituary lists surviving family, home town, and a mother’s maiden name.

Curated Directories

osintframework.com, bellingcat.com/toolkit, github.com/jivoi/awesome-osint, github.com/cipher387, weekinosint.com.

Evaluate a New Tool

Is it maintained? Who sees my data? What's the OPSEC impact? Is it legal? What does it cost? Can I verify it? Is the software safe?

Security questions: genealogy answers "What is your mother's maiden name?"

Putting It Together

01
Job posting
Security-Inc runs Microsoft 365 and a named firewall

→

02
MX and TXT
Confirm Microsoft 365 plus several SaaS tools

→

03
crt.sh, Shodan
vpn.security-inc.com exists; Shodan shows the VPN product

→

04
Staff lists
LinkedIn IT staff; contact page confirms email format

AN ATTACKER NOW KNOWS

Who to phish, what the email should look like, and which login page to fake.

A DEFENDER NOW KNOWS

Exactly what to fix and who to train.

06

Dark Web & Threat Intel

Where the underground lives, how analysts watch it safely, and why stolen credentials matter so much.

TOR BASICS · FINDING SOURCES · STEALER LOGS · FLARE · SAFE RESEARCH

Surface, Deep, and Dark Web

LAYER 1

Surface web

Pages that search engines index.

Examples: news sites, Wikipedia, public social media posts

Reach it: any browser, any search engine

LAYER 2

Deep web

Anything not indexed, usually behind a login, paywall, or form.

Examples: your email inbox, online banking, a school LMS Learning Management System

Reach it: a browser plus an account or the exact URL

LAYER 3

Dark web

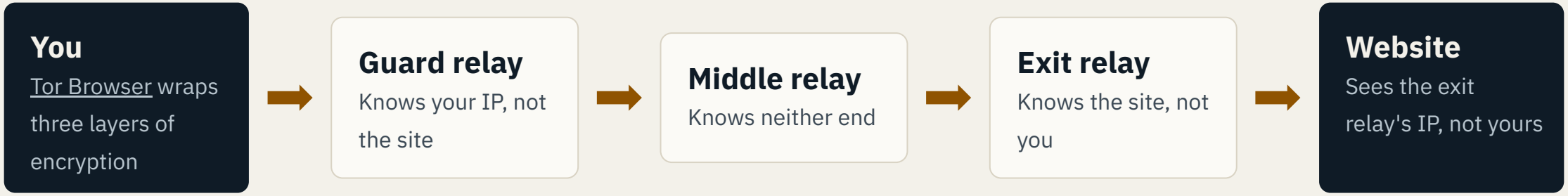
Sites on overlay networks built to hide who runs them and who visits.

Examples: Tor .onion sites, I2P .i2p sites

Reach it: Tor Browser or an I2P router

Takeaway: The dark web is a small slice of the deep web, and much of the criminal underground runs on clear-web forums and Telegram.

How Tor Works



Onion services

Servers reachable only inside Tor. Introduction and rendezvous points connect you, so neither the visitor nor the server learns the other's IP.

v3 onion addresses

56 characters that contain the site's public key. No DNS or WHOIS to look up, and lookalike clones are easy to make, so verification is critical.

Legitimate Uses of the Dark Web

Tor protects real people in real danger.

- Getting around censorship
- Whistleblowing through SecureDrop
- Privacy from tracking
- Activists in hostile environments
- Security research and law enforcement

Organizations with onion sites

Organization	Why it runs one
Facebook	Reachable where Facebook is blocked
ProPublica	First major news outlet on Tor
The New York Times	Readers in censored countries
BBC	Access where BBC content is blocked
CIA	Read the site and send tips privately
DuckDuckGo	Private search from inside Tor

Get their addresses from the organization's own site or Tor Browser's ".onion available" button, never from a link list.

Myth check: Using Tor is legal in the US and most countries. What you do with it can be illegal.

The Cybercrime Supply Chain

01

Malware developers

Build infostealers and rent them out as a service.

02

Infostealer operators

Spread the malware and collect the stolen logs.

03

Log shops and channels

Sell logs and account access on markets and Telegram.

04

Initial access brokers

Turn stolen logins into network access and sell it.

05

Ransomware affiliates

Rent the ransomware and carry out the attacks.

06

RaaS and leak sites

Encrypt and steal data, then threaten to publish it.

Where defenders fit: stealer-log hits, access listings, and leak-site posts are early warnings.

Finding Underground Sources

There’s no Google for the dark web — the onion search engines are thin and stale, so analysts mostly pivot from mentions elsewhere.

Onion sites

Ahmia (ahmia.fi) searches onion sites from a normal browser; Torch, Haystak and Tor66 run inside Tor. Coverage is thin, stale and full of clones verify; every hit. Skip “hidden wiki” link lists.

Forums and markets

Many run on the clear web behind vetting. Security news, vendor reports, court filings, and takedowns name the ones that matter.

Telegram

site:t.me dorks, Telegram's global search, t.me/s/ web previews that need no account, forwarded messages, and TGStat.

Ransomware leak sites

Ransom notes, trackers like ransomware.live and ransomlook.io, CISA #StopRansomware advisories, and security news.

Verifying What You Find There

Before you trust a lead

- Is it a v3 address (56 characters)?
- Do two independent sources confirm it?
- Does a PGP-signed mirror list verify?
- Is the lead recent?
- Could it be a clone or a re-posted old leak?
- Is it logged with UTC time, source, and confidence?

Takedowns confirm what was real

Operation	Target	Date
Operation Cronos	LockBit ransomware; its leak site was taken over	Feb 2024
Cookie Monster	Genesis Market, which sold stolen cookies and logins	Apr 2023
Operation Magnus	RedLine and META infostealers	Oct 2024
Lumma disruption	Lumma infostealer; about 2,300 domains seized	May 2025

Court filings explain how each operation worked, and analysts watch where users migrate next.

What a Stealer Log Contains

Why it's worse than a typical breach

Typical data breach	Stealer log
One website	Every site saved in the browser
Passwords often hashed	Passwords in plaintext
Often months or years old	Often days or hours old
No session data	Live session cookies
Shows an email was exposed	Shows exact login URLs: VPN, SSO, admin panels

HOW THEY ARE SOLD

Telegram channels and Discord servers run vending bots: a live feed posts new infections, you search by domain or URL, pay in crypto, and the bot delivers instantly. Freshness is the product — an hour-old log still has working session cookies.

MFA BYPASS

Pass-the-cookie

A stolen session cookie says "this browser already logged in," so the site may skip the password and MFA prompt.
MITRE ATT&CK T1539 (steal) and T1550.004 (use).

Respond by revoking sessions, not just resetting passwords.

Infostealers grab saved passwords, cookies, autofill data, crypto wallets, and system details in seconds, and they're rented out as a service.

MODULE 6 · CASE STUDY: SNOWFLAKE 2024

~165

organizations were notified after attackers logged into Snowflake customer accounts with credentials stolen by infostealers.

The accounts had no MFA · some credentials were up to four years old · at least 79.7% had prior exposure · no Snowflake platform vulnerability was exploited

Source: Mandiant and Snowflake (UNC5537 investigation)

MODULE 6 · FEATURED RESOURCE

Flare: Dark Web Intelligence

Search an archive of the underground instead of visiting it — flare.io.

Flare monitors	Coverage Flare cites
Cybercriminal forums	160+ on Tor, I2P, and the clear web
Telegram	127,000+ channels
Ransomware leak sites	100+
Stealer logs	159M+ archived; 1M+ new each week
Leaked credentials	20B+

Real-time alerts
Mentions of your domains, brands, or executives

Stealer-log detection
Corporate credentials in stealer logs and combolists

Automated remediation
Resets passwords and revokes sessions via Entra ID or Okta

Free trial
Shows your real exposure; use only an authorized domain

MODULE 6 · FEATURED RESOURCE

Darkroom: A Free Dark Web Lab

A simulated dark web investigation from Flare Academy, with no dark web required.

Detail	Darkroom
Launched	August 2026
Cost	Free: course, community, and certification
Access	flare.io/darkroom, then the Flare Discord
Time	About 4 hours, self-paced
Finish	Certification, badge, and CPE credits

What you do in the lab

- Investigate simulated forum threads
- Analyze stealer logs and credential databases
- Triage breaches in an alert console
- Trace cryptocurrency transactions
- Examine ransomware operations
- Negotiate with AI-simulated threat actors

Flare Academy (free)

Expert-led training, 40+ hours of replays, research, and a Discord community with Flare's researchers.

Free and Paid Platforms

Free and freemium

Tool	What it's for
Hudson Rock · hudsonrock.com	Stealer-log lookups; free monitoring
SOCRadar · socradar.io	Free domain-level exposure scan
Have I Been Pwned · haveibeenpwned.com	Breaches and stealer logs by email
Intelligence X · intelx.io	Leaks, pastes, Tor and I2P; limited free tier
ransomware.live · ransomlook.io	Track leak sites without visiting them

Paid enterprise

Flare (free trial)	Recorded Future
Flashpoint	Intel 471
KELA	DarkOwl Vision
Cybersixgill (Bitsight)	ZeroFox
CrowdStrike Falcon Intelligence Recon	SpyCloud

Compare coverage, history, freshness, context, integrations, collection ethics, and cost.

Safe Research Practices

FIRST

Authorization

- Written scope, allowed actions, data handling, and escalation contacts
- Legal counsel reviews the plan
- Passive viewing only, unless approved

ISOLATE

Environment

- Whonix: Gateway and Workstation VMs
- Tails: amnesic USB system, now part of the Tor Project
- Tor Browser on Safest: JavaScript off everywhere

TEMP

Only with written sign-off

- Download files
- Buy anything
- Engage threat actors
- Use your real identity
- Log in with stolen credentials
- Store plaintext passwords

In this class: intelligence platforms, the Darkroom lab, and instructor-led onion demos only — none of the TEMP column, ever.

If You Encounter Illegal Content

Child sexual abuse material

01 **Stop immediately** and close the page.

02 **Don't** screenshot, save, copy, forward, or show it to anyone.

03 **Report it** to the NCMEC CyberTipline: report.cybertip.org or 1-800-843-5678, available 24/7.

04 **Tell** your instructor or supervisor and follow your organization's policy.

Protect your wellbeing

Disturbing content affects everyone who sees it. That's normal, not a weakness.

- Prefer text-first tools; blur or hide images
- Limit session length and take breaks
- Decompress before going to sleep
- Talk to peers, a supervisor, or a counselor
- It's okay to step away from a task

07

Verification & Reporting

Turn raw collection into findings people can trust, act on, and defend.

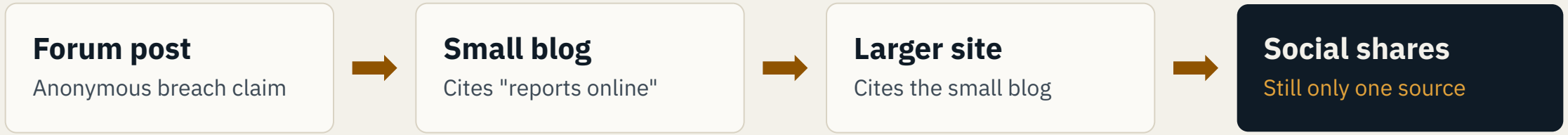
SOURCES · IDENTITY · IMAGES · CONFIDENCE · EVIDENCE · REPORTS

MODULE 7 · VERIFICATION & REPORTING

Evaluating a Source

Ask	Look for	Red flags
Who published it?	A named person or org with a track record	Anonymous or brand-new accounts
When?	A clear date; still current	No date; old content shared as new
Why?	Motive: inform, sell, persuade, deceive	Emotional language, an agenda, a sale
How would they know?	Direct access: they were there or own it	Second-hand claims, no explanation
Where else?	Independent sources that confirm it	One source, or many echoing one post

CIRCULAR REPORTING: FOUR SOURCES, ONE ORIGIN



Same Name, Different Person

STRONG

Email address, phone number, account ID, a photo that matches exactly across accounts

MEDIUM

Uncommon name plus city, employer plus job title, the same profile photo, writing style

WEAK

Common name, first name only, a city, a hobby, a generic username like mike123

Fictional example: Jordan Lee, IT admin at Security-Inc

Account found	Decision
LinkedIn: name, employer, role	Link (medium)
GitHub: commit email in company format	Link (strong)
Instagram: same profile photo	Link after image check
Reddit "jordanlee": other city, other job	Don't link

Rule: require two independent medium-or-strong identifiers before linking accounts, and record why you linked them.

Image Verification Workflow

01

Find the earliest version

Reverse search on several engines; TinEye can sort by oldest.

02

Check the metadata

exiftool shows device, capture time, and GPS. Platforms strip it, and it can be edited.

03

Confirm the location

Match signs, buildings, and terrain to maps and street-level imagery.

04

Confirm the time

Shadows (SunCalc), historical weather, seasonal and event clues.

05

Examine the content

Lighting, warped lines, repeated patterns. Forensic filters are hints, not verdicts.

06

Investigate the uploader

Who posted it first? Could they plausibly have been there?

Tool: the free InVID-WeVerify plugin bundles keyframes, metadata, forensics, and reverse search.

MODULE 7 · WHEN VERIFICATION FAILS

Good intentions don’t replace verification.

BOSTON · APRIL 2013

After the Marathon bombing, Reddit and Twitter users crowdsourced the FBI’s photos and named Sunil Tripathi, a Brown student who had been missing for a month. His family was flooded with abuse. He had died before the bombing ever happened, and Reddit’s general manager later apologised publicly.

CHARLOTTESVILLE · AUGUST 2017

After the Unite the Right rally, online crowds identified a marcher from photos and named Kyle Quinn, an engineering professor at the University of Arkansas, because he resembled a man wearing an “Arkansas Engineering” shirt. He was 1,100 miles away. He and his wife left their home after the threats started.

BOTH WERE WRONG WITHIN HOURS, AND NO CORRECTION REACHES EVERYONE WHO SAW IT

Estimative Language

Probability	Set 1	Set 2
01–05%	almost no chance	remote
05–20%	very unlikely	highly improbable
20–45%	unlikely	improbable
45–55%	roughly even chance	roughly even odds
55–80%	likely	probable
80–95%	very likely	highly probable
95–99%	almost certain	nearly certain

ICD 203, US Intelligence Community analytic standards. Pick one set and don't mix them.

Likelihood

About the event: how probable is it that this is true?

Confidence

About your evidence and reasoning: high, moderate, or low.

Keep them apart

Don't put both in one sentence. Estimates never reach 0% or 100%.

MODULE 7 · THE ADMIRALTY CODE

Grading Sources and Information

Grade	Source reliability
A	Completely reliable
B	Usually reliable
C	Fairly reliable
D	Not usually reliable
E	Unreliable
F	Reliability cannot be judged

Grade	Information credibility
1	Confirmed by other sources
2	Probably true
3	Possibly true
4	Doubtful
5	Improbable
6	Truth cannot be judged

Grade the source (A–F) and the information (1–6) separately, then combine: B2 means a usually reliable source saying something probably true.

A1
Reliable source, confirmed elsewhere crt.sh and Shodan both show vpn.security-inc.com

B2
Usually reliable, probably true the Security-Inc careers page asks for firewall skills

D3
Not usually reliable, possibly true a people-search site lists an employee address

F3
Cannot be judged, possibly true a new forum account claims a database sale

MODULE 7 · DOCUMENTATION & EVIDENCE

Capture as You Go

Document everything!!!!

Online content disappears. Capture it the moment you find it.

- Minimum: URL, full-page screenshot, HTML source
- Timestamps in UTC, clock synced
- SHA-256 hash at collection
- Log every step, even empty searches
- Never edit originals; work on copies

OPSEC: public archives are public. A new capture of an obscure page can tip off the target.

```
# Linux: hash every capture, save the list
sha256sum * > ../../01_log/web.sha256
# Later: prove nothing changed
sha256sum -c ../../01_log/web.sha256
# macOS
shasum -a 256 capture.png
# Windows PowerShell
Get-FileHash .\capture.png -Algorithm SHA256
# Current time in UTC
date -u +"%Y-%m-%dT%H:%M:%SZ"
```

Report Structure

1 · Executive summary
Bottom line up front: the answer, rating, and top actions.

2 · Scope & authorization
Who authorized it, what's in and out, and the rules.

3 · Methodology
Collection window, sources, tools, grading keys, limits.

4 · Summary of findings
One table: ID, finding, risk, confidence.

5 · Detailed findings
One section per finding, from a template.

6 · Attack path
How an attacker could chain the findings.

7 · Recommendations
Prioritized actions with owners and effort.

8 · Gaps & limitations
What you couldn't determine, and what wasn't tested.

Every finding = evidence + source grade + confidence + impact + recommendation

Rating Risk

	Low impact	Medium impact	High impact
High likelihood	Medium	High	Critical
Medium likelihood	Low	Medium	High
Low likelihood	Informational	Low	Medium

Likelihood

How easily could an attacker find and use this?

Impact

What's the worst realistic outcome?

Be consistent

CVSS scores software flaws, not exposure. Define your scale in the report.

Protecting the Report

An OSINT report is a roadmap for attackers. Internal work follows your policy; external follows the SOW.

01

Label it

- TLP:RED named recipients
- TLP:AMBER need-to-know
- TLP:GREEN their community
- TLP:CLEAR share freely

02

Protect it

- Encrypt at rest and in transit
- Password via another channel
- Named, versioned recipients
- Real redaction, not boxes

03

Minimize it

- Roles and counts over names
- Names in a restricted annex
- No plaintext passwords
- No family, minors, addresses

04

Store it

- Encrypted drive in the VM
- Or media kept off that box
- Not personal cloud storage
- Delete on the SOW date

State what you didn't do: "We did not test the key" protects you and shows how far a finding goes.

08

OPSEC

Protecting the investigator: your identity, your client, and your findings.

FUNDAMENTALS · RESEARCH ENVIRONMENT · SOCK PUPPETS · COMMON MISTAKES

MODULE 8 · OPSEC

What OPSEC Protects

Operations security: find the information that would hurt you if an adversary learned it, then deny them the clues.

1966 → 1988

Operation PURPLE DRAGON found that routine, unclassified US activity was warning the enemy in Vietnam. NSDD-298 made OPSEC a formal program in 1988.

Your identity

Your name, employer, and personal accounts.

Your client

Revealing your firm can reveal who hired you.

Your methods

Research accounts, tools, and infrastructure.

Your findings

A ready-made roadmap for real attackers.

Why it matters: a target who notices you can delete posts, change behavior, feed you false information, or retaliate.

MODULE 8 · OPSEC

The Basic Three

Before any of the advanced tradecraft, three habits do most of the work.

01

Use a VPN

Your home IP is registered to your address and your name, and campus and corporate ranges name the organisation. A VPN puts someone else's IP in the logs. Turn the kill switch on and test for DNS, WebRTC and IPv6 leaks at ipleak.net.

02

Use a hardened browser

Brave (brave.com), or Firefox locked down, blocks trackers and fingerprinting by default and needs no account. Keep a separate profile for research so cookies and logins never mix with your own.

03

Don't post anything online

Mostly joking. The real rule: limit what you post, and treat anything you put somewhere you don't own or control as permanent and public. You can delete your post. You cannot delete the screenshot.

None of this is exotic. Most OPSEC failures are ordinary mistakes, not sophisticated attacks.

The Five-Step OPSEC Process

Step	The question	Security-Inc assessment example
1. Identify critical information	What would hurt if it leaked?	That the assessment exists; your firm's name
2. Analyze threats	Who might notice or want it?	Employees, IT staff not told about the test, real attackers
3. Analyze vulnerabilities	How could it leak?	Firm-branded LinkedIn views; office IP in web logs
4. Assess risk	How likely, and how bad?	LinkedIn views: high likelihood, high impact, fix now
5. Apply countermeasures	What reduces the risk?	Private mode, isolated network, private scans, clean metadata

The steps form a loop: revisit them when the target, platform, or risk changes. Most countermeasures are discipline, not expensive tools.

Match OPSEC to the Threat

Low: notices likes, follows, and profile views

Medium: reviews logs, watches scan sites, plants canaries

High: hunts researchers, sets traps, retaliates

Investigation	Threat tier	Reasonable baseline
Finding a company's email format for an authorized phishing test	Low–Medium	Separate browser profile, logged out of personal accounts, LinkedIn private mode
Full external assessment of a company with a mature security team	Medium	Dedicated VM, network that doesn't trace to your employer, approved research accounts
Monitoring a ransomware group's public channels (authorized)	High	Isolated, disposable environment; hardened browser; strict persona separation; legal sign-off
Investigating a harassment campaign against a client	Medium–High	Strong separation from your personal identity; attention to personal safety

MODULE 8 · OPSEC

Build in Layers

LAYER	WHAT CAN LEAK	COUNTERMEASURE
Host computer	Malware, personal files and accounts	Keep it updated; research inside a VM
Virtual machine	Shared clipboard and shared folders	Disable sharing; use clean snapshots
Operating system	Hostname, username, time zone	Generic names; time zone fits your cover
Browser	Cookies, fingerprint, extensions, sync	Dedicated profiles; hardened settings
Accounts	Linked phone numbers, reused usernames	One purpose per account
Network	Home or corporate IP, DNS queries	VPN, Tor, or a non-attributable network

One weak layer undoes the rest. A VM isolates and resets, but it does not hide your IP.

MODULE 8 · OPSEC

VPN, Tor, and Leak Tests

Compare	VPN	Tor
Website sees	The VPN server's IP	A Tor exit relay's IP (publicly listed)
Who sees your traffic	The VPN provider, so you must trust them	Exit relay sees non-HTTPS traffic, not who you are
Speed and access	Fast; some sites flag VPN ranges	Slow; many sites block Tor or show CAPTCHAs
Best for	Everyday OSINT	High-risk research and .onion sites

TEST FIRST

Check for DNS, WebRTC, and IPv6 leaks with ipleak.net and BrowserLeaks, and turn on the VPN kill switch.

NEVER

Research from corporate or campus networks. Their IP ranges are publicly registered to the organization.

MODULE 8 · OPSEC TOOL

Gluetun: A VPN Container for Your Tools

Your collection tools each have their own network settings, and most have none at all. A system VPN covers your browser, but a container can still leak around it. Gluetun puts the VPN inside the container, so the tool physically cannot reach the internet any other way.

Kill switch

If the VPN drops, tools go offline instead of leaking your IP.

DNS over TLS

Lookups skip your ISP's DNS servers.

Built-in proxies

HTTP and SOCKS5, so one browser profile can use the tunnel.

One per persona

Separate containers keep identities off a shared exit IP.

```
# In docker-compose.yml: the tool reaches the internet only through the VPN
network_mode: "service:gluetun"
# Check the exit IP from inside the tunnel (Gluetun wiki)
docker run --rm --network=container:gluetun alpine:3.22 sh -c "apk add wget && wget -qO- https://ipinfo.io"
```

Official sources only: github.com/passteque/gluetun (moved from qdm12 in 2026) and the qmcgaw/gluetun image.

MODULE 8 · OPSEC

Sock Puppet Rules

A sock puppet is a fictional research account for viewing login-only content.

No real people

Personas are fictional. Never clone a real profile.

No real photos

Stolen photos harm the person pictured.

Observe only

No likes, comments, or messages. Interacting is social engineering.

Compartmentalize

Separate environment, email, phone, and IP for each persona.

Log and retire

Record purpose, approval, and activity. Burn it when exposed.

Alternatives first

Logged-out views, archives, official APIs, or client access.

LinkedIn and Facebook ban fake accounts. In this class, plan personas on paper only.

MODULE 8 · OPSEC

Common OPSEC Mistakes

“Who viewed” features

Profile views, story viewer lists, and read receipts. Use logged-out views or private mode.

Accidental taps

A stray like, follow, or request; synced contacts. Research on desktop and slow down.

Your IP in their logs

Corporate and campus IPs name your organization. Use a non-attributable network.

Traps and canaries

IP-logger links, canary files, fake keys. Don't click, don't open, never test found credentials.

Public submissions

Scanners like urlscan.io are public by default. Use private scans and hash lookups.

Metadata in your files

Photo GPS, document authors, fake redactions. Check with exiftool before sharing.

Mistakes are permanent. Logs, archives, and screenshots don't forget.

OPSEC Failures

2012

John McAfee

The antivirus founder was on the run in Belize in 2012 over a neighbour's death. A Vice journalist travelling with him posted an iPhone photo with the GPS tags intact, placing him in Guatemala.

Lesson: your files can expose the people you are protecting.

2012

Higinio Ochoa (w0rmer)

Ochoa belonged to CabinCr3w, an Anonymous-linked crew that leaked data from US police sites during the AntiSec campaign. After a defacement he posted a taunting photo of his girlfriend; its EXIF GPS pointed to a house in Australia.

Lesson: one uncleaned photo was enough. He served 27 months.

2011–2013

Ross Ulbricht

Ulbricht built and ran Silk Road, the first large dark web drug market, as "Dread Pirate Roberts". An IRS investigator found the earliest Silk Road posts by "altoid", who months later posted his real Gmail address.

Lesson: reused handles link identities for years.

One moment of carelessness was enough each time, and OSINT is how investigators found it.

09

Defending Against OSINT

Flip the perspective: find and shrink what you and your organization expose before someone else uses it.

SELF-AUDIT · PERSONAL EXPOSURE · DATA BROKERS · ATTACK SURFACE · MONITORING

Minimize, Compartmentalize, Harden

01

Minimize

Share less, and make less of what you share public.

Try: private accounts, "Limit past posts," posting after you leave.

02

Compartmentalize

Keep school, work, gaming, and personal life from being easily linked.

Try: different usernames, email aliases, separate phone numbers.

03

Harden

Assume some data will leak, and make it useless to an attacker.

Try: a password manager, passkeys, made-up security answers.

Goal: You can't reach zero exposure. Be a harder, less rewarding target than the next person.

MODULE 9 · DEFENDING AGAINST OSINT

Audit Your Own Footprint

List your identifiers first: names, usernames, emails, phones, addresses, photos.

01 **Search engines**

02 **Usernames**

03 **Email addresses**

04 **Phone numbers**

05 **Social media, logged out**

06 **Photos and backgrounds**

07 **Data brokers**

08 **Breaches and stealer logs**

09 **Public records**

10 **Old accounts**

11 **What others post**

12 **Think like an attacker**

Reduce Personal Exposure

Lock down platforms

Run each platform's privacy checkup. Private accounts and "Limit past posts" go a long way.

Stop leaking location

Turn off camera location tags. Post trips and events after you leave, not while you're there.

Check the background

Badges, keys, screens, whiteboards, mail, house numbers, and school logos give you away.

Break the pivot chain

Use different usernames for different parts of your life, and a unique email alias per service.

Make leaks useless

Unique passwords from a manager, passkeys where possible, and made-up security answers.

Clean up old accounts

Delete, don't just deactivate. Google "Results about you" removes Search results, not the source.

Make Leaked Data Useless

MFA method	Strength	Why
Passkeys or security keys	Strongest	Only work on the real site
Authenticator app codes	Good	Can be typed into a fake site
SMS codes	Weakest	SIM swapping and phishing

Watch for infostealers

Cracked software and game cheats are common bait. Stolen session cookies can skip MFA entirely.

Fake security answers

"What city were you born in?"

`violet-harbor-2291`

Store the made-up answer in your password manager.

Unique passwords

Reuse turns one breach into many takeovers through credential stuffing.

First: Protect your email account. It's the master key for resetting everything else.

MODULE 9 · DEFENDING AGAINST OSINT

Data Broker Opt-Outs ARE FAKE!

70%

of listings gone within a week with manual opt-outs, the most effective method (Consumer Reports, 2024)

68% · 65%

removed after four months by the best paid services (Optery, EasyOptOuts);.

1 request

through California's DROP reaches every registered broker; brokers must process them from Aug 1, 2026

Search for yourself → **Copy each listing URL** → **Opt out with an alias** → **Confirm, then re-check**

Do it yourself: manual opt-outs beat every paid service, they are free, and you avoid handing a new company the exact data you are trying to erase.

Why it matters: In July 2020, a gunman went to federal judge Esther Salas's home. Her son, Daniel Anderl, was killed when he answered the door.

MODULE 9 · DEFENDING AGAINST OSINT

Shrink an Organization's Attack Surface

Assess with sign-off

Get written approval and a scope, then run the OSINT cycle on your own organization.

Know your assets

Keep an asset inventory and monitor it continuously (EASM). Shadow IT piles up.

Scrub documents

Remove metadata before publishing, and use real redaction, not black boxes.

Careful job postings

Name categories, not your exact security stack. Skip "our sole security engineer."

Train employees

No badges, screens, or whiteboards in photos. Teach staff to spot fake recruiters.

Protect executives

Broker removal for families, no real-time travel posts, call-back checks on payments.

Rule: Every finding needs a risk rating, a fix, and an owner. Findings without owners don't get fixed.

Leaked Secrets and Dangling DNS

```
# Scan a local repo's full history
gitleaks git -v path/to/repo
# Live secrets only, in repos you own
trufflehog github --org=security-inc
--results=verified
```

A secret leaked? Rotate first.

1. Revoke or rotate it immediately
2. Check logs for any use since exposure
3. Clean history; copies may already exist
4. Fix the root cause (hooks, secrets manager)

How subdomain takeover works

1. promo.security-inc.com points (CNAME) to a hosted page
2. The campaign ends and the page is deleted
3. Nobody deletes the DNS record: it's dangling
4. **An attacker signs up for that hostname on the service**
5. They now control content on your subdomain

```
dig CNAME promo.security-inc.com +short
```

Prevent: delete the DNS record in the same change as the thing it points to — the bucket, app, or landing page.

Watching for TypoSquatting

Typosquatting: registering a domain that looks like yours to catch people who mistype, misread, or don't look closely.

Trick	Lookalike of security-inc.com
Omission (typosquat)	securty-inc.com
Addition (typosquat)	securityy-inc.com
Transposition (typosquat)	secuirty-inc.com
Replacement (typosquat)	secutity-inc.com
Homoglyph (IDN homograph)	security-lnc.com
Hyphenation	securityinc.com
TLD swap	security-inc.co
Dictionary (combosquatting)	security-inc-login.com

```
# Only lookalikes that are registered
dnstwist --registered security-inc.com
# Compare pages to your site (isolated VM)
dnstwist --registered --lsh security-inc.com
```

When one shows up

Investigate: website, mail servers, copied branding?

Block it in email and web filters

Report it to the registrar and host

Warn employees or customers

Prevent: register the most obvious typos yourself

Early-Warning Tools

Tool	Cost	What it watches
Google Alerts · google.com/alerts	Free	New search results for names, domains, products
Have I Been Pwned · haveibeenpwned.com	Free; paid for larger domains	Known breaches (domains must be verified)
Hudson Rock · hudsonrock.com	Free tier (3 of each identifier)	Infostealer logs tied to your identities
SOCRadar Dark Web Report · socradar.io	Free domain scan	Breach records, forums, stealer logs
Google "Results about you" · myactivity.google.com/results-about-you	Free	Your phone, address, email, or ID numbers in Search
certspotter (open source) · sslmate.com/certspotter	Free	New certificates issued for your domains
Flare · flare.io	Paid, with a free trial	Forums, Telegram, leak sites, stealer logs, with alerts

Stealer-log hit? Reset passwords, revoke all sessions, reimage the device, and review recent logins.

Build a Monitoring Routine

When	What
Daily (automated)	Google Alerts, HIBP domain alerts, stealer-log alerts, dnstwist job, CT monitoring
Weekly	Triage every alert and follow up on anything new
Monthly	Search for brand and executive impersonation; review job postings and press releases
Quarterly	Full OSINT self-assessment, DNS zone review, executive data broker re-check
After events	Breach news, layoffs, mergers, launches, executive changes, phishing incidents



WRAP-UP

Key Takeaways

01 **OSINT** is public information plus analysis to answer a specific question.

02 Attackers and defenders use the same techniques; **intent and authorization** make the difference.

03 **Public is not permitted:** stay in scope, follow the law, collect only what you need.

04 **Define the question first,** then collect, process, analyze, and report.

05 **Stay passive by default;** active recon requires written authorization.

06 **Think in source categories,** not tool names.

07 **Verify before you trust,** and state your confidence.

08 **Document everything** as you go.

09 **Protect yourself** while investigating, and shrink your own footprint.

10 **Tools change;** the method lasts.

WRAP-UP

Where to Practice Legally

Flare Academy

flare.io/academy — free Darkroom lab: a simulated dark web investigation with a certification.

Trace Labs

tracelabs.org — Search Party CTFs on real missing-persons cases. Advanced, with strict rules.

TryHackMe

tryhackme.com — OSINT rooms in a practice environment built for it.

GeoGuessr

geoguessr.com — builds the visual clue-reading behind geolocation.

Bellingcat

bellingcat.com — investigation challenges and the Online Investigations Toolkit.

Quiztime

quiztime.github.io — community geolocation and verification puzzles.

Exit question: What's one thing about your own online footprint you'll change after today?

UP NEXT

2.2 Web Hacking

Where the recon findings from this lesson turn into authorized testing.