T-401-ICYB
Open Source Intelligence (OSINT)

Stephan Schiffel

stephans@ru.is

Reykjavik University, Iceland

01.12.2025

Outline

1. What is OSINT?
2. Tools and Techniques
3. OPSEC
4. Lab today

What is OSINT?

What is OSINT?

Definition

OSINT = **O**pen **S**ource **Int**elligence.
The practice of collecting, analyzing, and making decisions based on information that is **publicly available** and **legally accessible**.

Key Distinction: OSINT vs. Classified Intelligence

- Does **not** involve hacking, spying, or stealing restricted data.
- Relies entirely on data found in the public domain.
- *"The information is out there; the skill lies in aggregating and analyzing it."*

"Open Source" in Intelligence vs. Computer Science

- **Open Source Software (OSS):**
  - Source code available for modification and redistribution (e.g., Linux, Python).
- **Open Source Intelligence (OSINT):**
  - Refers to the **overt nature** of the data source.
  - The source is unclassified and accessible to the public.

Data Vectors: Where does it come from?

OSINT is not limited to Google Search. It encompasses:

1. **The Internet (Surface & Deep Web):**
   - Google Search, etc.
   - Social Media (Twitter/X, LinkedIn, Instagram, Facebook).
   - Discussion Boards (Reddit, HackerNews).
   - Domain registrations (Whois data).
2. **Government & Public Records:**
   - Court filings, property records, census data.
   - FCC licenses, patent databases.
   - Fincancial Records, Annual Reports.
3. **Grey Literature:**
   - Technical reports, whitepapers, conference proceedings.
   - **CS Relevance:** Analyzing metadata in PDFs or GitHub commit history.
4. **Mass Media:**
   - News broadcasts, print media, radio.

Applications in Cybersecurity (Defensive)

Red Team: Penetration Testing

- **Reconnaissance Phase:** Gathering info before touching a server.
- Mapping network infrastructure via public DNS records.
- Identifying employees for social engineering tests.

Blue Team: Defense

- Monitoring ”paste sites” (e.g., Pastebin) for leaked credentials.
- Tracking threat actors on dark web forums.
- Scanning for accidental public code repository leaks.

Applications in Other Sectors

- **Law Enforcement & Intelligence:**
  - Counter-terrorism and tracking criminal networks without needing warrants for private data.
- **Business Intelligence:**
  - Competitive analysis (Mergers & Acquisitions due diligence).
  - Supply chain verification.
- **Journalism:**
  - Fact-checking and geolocation of events.
  - Verifying war zone footage using satellite imagery and landmarks.

The "Dark Side"

How malicious actors (Black Hats) utilize OSINT against targets:

- **Target Profiling:**
  - Using LinkedIn to identify SysAdmins and their tech stack (e.g., *"Expert in AWS"* implies the company uses AWS).
- **Social Engineering:**
  - Crafting Spear-Phishing emails based on hobbies or recent events posted on social media.
- **Doxing (doc dropping):**
  - Aggregating disparate data points to reveal a user's real-world identity and address.
  - Typically in order to intimidate, harass, or endanger a target by exposing their identity and address.

Tools and Techniques

Advanced Search Techniques (”Dorking”)

Google Dorks (Search Operators)

Using commands to filter results for specific data types.

- `site:linkedin.com "project manager"`
  (Search only inside specific domains)
- `filetype:pdf "confidential"`
  (Find specific file types)
- `intitle:"index of"`
  (Find unprotected server directories)
- `cache:example.com`
  (View Google’s saved version of a site)

Alternative Search Engines: e.g., DuckDuckGo

- Useful for unbiased results (avoids ”filter bubbles”).
- Does not track search history.

Tools

- Find online accounts by username, email, etc: Sherlock, Epieos, ...
- Reverse Image Search: Yandex Images, Google Lens, TinEye
- Use image metadata (EXIF): camera model, **time, date, GPS coordinates**.
- Whois Lookup: domain registration details (owner, registration date)
- DNSDumpster: finds (hidden) subdomains
- The Wayback Machine: view deleted/old versions of web pages
- Google Earth: timeline slider to view locations over the years
- SunCalc: check whether the shadow in an image matches the time and location
- ...

Tool to select tools: OSINT Framework

OPSEC

What is OPSEC?

Definition

OPSEC (Operational Security) is the process of protecting individual pieces of data that could be grouped together to give away critical information (like your identity or location).

The Risk:

- Every website logs your IP address, device type, and "Referrer"
- **Reciprocal Surveillance:** If you investigate a sophisticated target, they check their server logs. They can see who is looking at them and where they came from.

Protecting Hardware & Software

Isolating the research environment to protect the host machine from malware and trackers.

- Virtual Machines (VMs)
- Tails OS:
    - An amnesic Operating System that runs from a USB stick.
    - ”Forgets” everything immediately upon shutdown (leaves no forensic trace on hardware).
    - Forces all traffic through Tor for anonymity.

Hiding Your Location

The goal is to dissociate your traffic from your home (or work) ISP.

- **VPN (Virtual Private Network):**
  - Encrypts traffic and routes it through a remote server.
  - *Pro:* Fast and easy. *Con:* Must trust the provider's logging policy.
- **Tor (The Onion Router):**
  - Routes traffic through multiple random volunteer nodes globally.
  - *Pro:* High anonymity. *Con:* Slow; often blocked by websites.
- **Public Wi-Fi (Attribution Management):**
  - Conducting high-risk searches from a library or cafe.
  - Even if the IP is traced, it leads to a public location, not your home.

Hide your Identity

Websites use ”Browser Fingerprinting” (screen resolution, installed fonts, battery level) to track unique devices even without cookies.

Countermeasures

1. User Agent Spoofing
2. Script Blockers, e.g., uBlock Origin, NoScript.
3. **Dedicated Research Browser:** Use a browser that has **never** logged into your personal accounts (and limits tracking).

Assume a Fake Identity

If you investigate a target on LinkedIn, LinkedIn will tell the target, "John Smith viewed your profile." To avoid this, analysts use **Sock Puppets**, fake online identities created for research purposes.

Anatomy of a Sock Puppet:

- **The Name:** Use a fake name.
- **The Face:** Use AI-generated faces (e.g., https://thispersondoesnotexist.com/) to avoid Reverse Image Detection.
- **The History:** Accounts must be "aged." Join groups and like posts weeks before using the account for investigation.
- **Verification:** Use "Burner Phones" (Prepaid SIMs) or VOIP (Google Voice) for SMS verification.

Behavioral OPSEC (Human Factors)

Technology does not matter if human error occurs.

- **Avoid Cross-Contamination:** NEVER log into a personal account (Gmail, Facebook) inside your investigation VM. One cookie can link your real identity to your sock puppet.
- **Physical Separation:** Do not conduct investigations on your personal devices.
- **Copy/Paste Discipline:** Ensure you do not accidentally paste a personal URL or password into a research window.

Summary

OSINT

The internet creates a massive amount of ”noise.”
OSINT is the process of filtering that noise to find the ”signal.”

OPSEC in OSINT

Preemptively mask your trail to evade identification.

Takeaway

- Be mindful of your digital footprint.
- Information you publish can be aggregated to attack you (spear fishing, identity theft, etc).
- Information employees publish can be aggregated to form a picture of an organization’s security posture.

Lab today

Lab today

- Lab 6: Open-Source Intelligence
- Select a target (company, organization) in Iceland
- Find everything you can about them **using legal means**
- Plan an attack (**but don’t do it!**) again only with **legal means**