WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Dark Reaper – Automated OSINT Tool

Viraj Kadam[1], Prof. J. R. Mahajan[2]

[1]Student & [2]Asst. Prof. of Department of Computer Engineering,
Adsul's Technical Campus, Chas, Ahilyanagar, Maharashtra, Savitribai Phule Pune University, Pune

A B S T R A C T

This research proposes DarkReaper, an automated Open-Source Intelligence (OSINT) tool designed to streamline investigations across both the surface and dark web. The tool features a modular, command-line interface (CLI) architecture that performs targeted reconnaissance on identifiers like Phone numbers, Email address, Images, Ip address, and Usernames. By leveraging free APIs, Python tools and Scraping techniques instead of paid services, DarkReaper provides a cost-effective solution. It standardizes all collected intelligence into structured JSON format for ease of analysis. The project aims to enhance accessibility for cybersecurity professionals, students, and researchers, offering a unified platform for comprehensive intelligence gathering.

Keywords: OSINT, Dark Web, Cybersecurity, Data Gathering, Python Automation

1. INTRODUCTION

The internet is an extensive resource for gathering intelligence, but manual OSINT investigation often requires the use of multiple tools, platforms, and manual searches. While existing solutions such as SpiderFoot provide valuable capabilities, many require premium subscriptions or API keys, making them inaccessible to students or budget-limited professionals.Existing OSINT tools present practitioners with significant challenges. Comprehensive platforms like SpiderFoot offer robust capabilities but require substantial financial investment. Additionally, most accessible OSINT tools provide limited coverage of dark web sources, despite the critical importance of dark web intelligence in modern cybersecurity investigations.

This project, DarkReaper, aims to unify various OSINT capabilities into a single CLI-based platform. The system will integrate surface and dark web sources, perform targeted lookups, and store results in JSON format. By addressing the accessibility gap in comprehensive OSINT tools, DarkReaper enables broader participation in cybersecurity research while maintaining professional-grade functionality.

2. OBJECTIVE

- To develop a modular, command-line based OSINT tool capable of performing reconnaissance on multiple input types (Email address, Phone number, IP address, Username, Image).
- To integrate both surface web and dark web data sources for comprehensive intelligence gathering.
- To implement automated data collection using free APIs, Python libraries and Web scraping techniques, eliminating dependency on paid services.
- To standardize all collected intelligence into structured JSON format for systematic analysis and reporting.
- To provide an accessible, open-source solution for cybersecurity professionals, students, and researchers.

3. LITERATURE SURVEY

The field of Open-Source Intelligence (OSINT) has evolved through general-purpose platforms, unified automation efforts, and domain-specific research. A review of existing work highlights both the advancements and the gaps that DarkReaper seeks to address.

3.1 General-Purpose OSINT Frameworks

The OSINT landscape features a range of mature frameworks designed for broad reconnaissance. Tools like SpiderFoot (2025) are modular frameworks offering broad coverage across hundreds of data sources and are often used as benchmarks in security investigations. However, as noted by Nordin, Yusoff, and Ahmad (2022), their heavy dependency on API keys and web services creates significant barriers to access, reproducibility, and consistent use for students and resource-constrained practitioners. Pastor-Galindo, Nespoli, Gómez Mármol, and Martínez Pérez (2020) further contextualized this,

describing OSINT as an underexploited "goldmine" whose potential is hindered by challenges like data veracity, source volatility, and fragmentation—issues that existing tools attempt, yet often fail, to overcome.

3.2 Unified and Automation Platforms

Recent work has focused on creating streamlined, unified OSINT platforms. For example, Răfailă, Gurzău, Grumăzescu, and Bica (2023) developed MTAFinder, which aggregates disparate sources into a single interface to improve efficiency, though it raises challenges in handling source heterogeneity. Similarly, Shin and Jung (2024) proposed a maintainable OSINT framework that emphasizes modularity, task queues, and caching—principles that directly influenced DarkReaper's design.

The push for automation extends to AI integration. Browne, Chen, and Lavoie (2024) systematically reviewed AI-based OSINT automation, finding benefits in entity resolution and clustering, but also warning against bias and the "black box" problem, motivating DarkReaper's explainable, post-processed AI analysis.

3.3 Domain-Specific Intelligence

Research also focuses on niche OSINT domains. For dark web intelligence, Gopireddy (2020) surveyed methods for crawling hidden services and extracting threat indicators, while Vignesh and Patidar (2024) discussed modern methodologies for investigating leaked data on dark web forums. These studies justify the need for a dedicated dark-web module while also emphasizing ethical limitations of direct crawling—hence DarkReaper's API-based approach.

In image analysis, Gangwar and Pathania (2018) demonstrated the value of EXIF metadata for digital image authentication but noted its fragility. For mobile and email intelligence, Okmi, Por, Ang, and Ku (2024) outlined practical approaches to phone and email lookup that align with DarkReaper's modular structure.

3.4 Synthesis and Identified Gaps

Across the reviewed literature, three consistent shortcomings are observed:

- **Accessibility and Cost:** Many tools rely on paid APIs or large infrastructures.
- **Reproducibility:** OSINT workflows often lack deterministic, exportable results.
- **Integrated Dark-Web Depth:** Dark web capabilities are often treated as peripheral add-ons rather than robust modules.

Consequently, no single free tool currently integrates surface- and dark-web OSINT into a fully modular, API-key-free CLI system. DarkReaper is positioned to fill this gap by emphasizing a low-API model, CLI reproducibility, and structured JSON output for ethical, unified, and scalable OSINT gathering.

4. PROBLEM STATEMENT

Investigators often face challenges in accessing comprehensive intelligence tools that are both free and easy to use. Existing solutions are fragmented, require complex setup, or most modules depend on paid APIs. Additionally, few tools integrate dark web searches alongside surface web OSINT in one package. There is a need for a free, modular, CLI-based OSINT tool capable of providing more accurate and automated results from both open and dark web sources.

5. PROPOSED SYSTEM

The proposed DarkReaper provides a modular framework for OSINT investigation using surface web, dark web (via clearnet access), and future AI-assisted analysis.

A. System Workflow

1. The user launches the CLI interface and selects a module for investigation (phone lookup, email lookup, username lookup, iP lookup, or image analysis).
2. The selected module performs intelligence gathering from multiple sources through scraping,Python tools and API calls.
3. For dark web intelligence, DarkReaper integrates with free dark web search engines (Ahmia) to retrieve information from hidden services without direct .onion crawling.
4. All collected results are standardized and stored in structured JSON format.
5. **Future Scope:** AI handler will process raw data by summarizing, filtering, and highlighting key insights.

B. Tools and Libraries

1. **Programming Language:** Python 3.x
2. **Scraping & Data Handling:** Requests, BeautifulSoup, JSON
3. **Browser Automation**: Playwright
4. **Dark Web Intelligence:** Ahmia
5. **Image Analysis:** EXIF metadata extractor, Tesseract OCR, MobileNet SSD
6. **Utility Modules:** Validation scripts, social media finders

C. Architecture

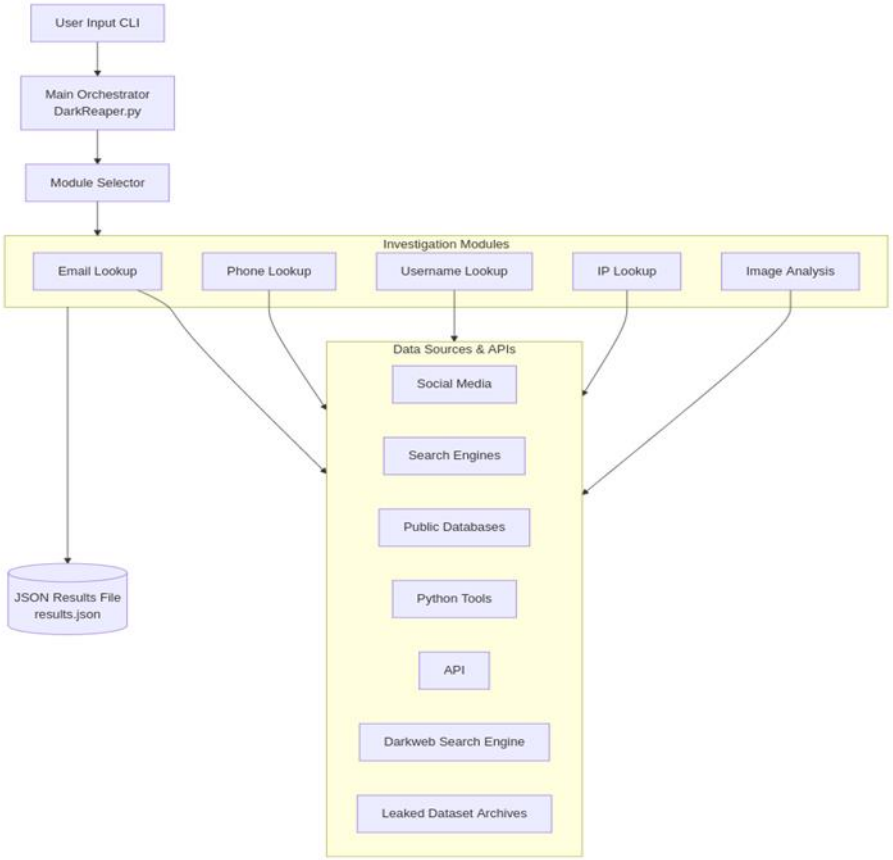

Figure 1. *DarkReaper System Architecture*

The architecture consists of:

- **CLI Interface:** User interaction and module selection
- **Core Modules:** Phone, Email, Username, IP, and Image intelligence modules
- **Data Processing:** JSON standardization and result correlation
- **Storage Layer:** Structured output and caching

6. WORKING OF EXISTING SYSTEM

The existing system utilizes **SpiderFoot**, an open-source automated OSINT framework used for reconnaissance and data collection. It automates the process of gathering intelligence from multiple public sources related to phone number, IP addresses, Email address, Usernames, and more.

SpiderFoot operates by executing multiple **modules**, each designed to query different data sources such as search engines, WHOIS databases, social media platforms, and threat intelligence APIs. These modules function independently to collect specific types of information, which are later correlated and presented in detailed reports or through its web interface.

Current Workflow:

- The user provides a target such as a Phone number, IP address, Email ID, Usernames, etc.
- SpiderFoot automatically executes a set of reconnaissance modules.
- Each module gathers relevant information from open-source data.
- The collected data is stored in a local database or displayed on the dashboard.
- The user manually analyzes, filters, and exports the results for further investigation.

Limitations:

- Requires manual selection and configuration of modules.
- Many data sources need individual API keys.
- Integration with other OSINT tools is limited.
- Lacks AI-based result interpretation or automation across multiple tools.

7. RESEARCH METHODOLOGY

A. Problem Identification and Literature Review

Comprehensive study of existing OSINT tools identifies key limitations: heavy API dependency, lack of modular integration, and minimal dark web support.

B. Requirement Analysis

Functional Requirements:

- Perform OSINT on multiple input types
- Integrate surface web and dark web sources
- Store data in structured JSON format
- Allow modular execution

Non-Functional Requirements:

- Platform independent (Linux-based)
- Lightweight, open-source, and free
- Extensible for future AI integration

C. System Design

Modular architecture with independent Python scripts (phone_lookup.py, email_lookup.py, ip_lookup.py, username_lookup.py, image_search.py). CLI interface for user interaction. JSON-based storage for structured reporting.

D. Planned Implementation

Python implementation integrating scraping techniques with open-source libraries. Error handling and retry logic for robustness. Integration of external utilities like Sherlock, H8mail, and Onioff. Testing and deployment in subsequent phases (PS2).

8. CONCLUSION

This research proposes DarkReaper, an Automated Intelligence Gathering Tool that combines surface and dark web OSINT techniques into a free, modular CLI system. The tool addresses significant gaps in current OSINT capabilities by providing a unified, accessible platform for comprehensive intelligence gathering.

DarkReaper tackles economic and technical barriers limiting access to professional-grade OSINT tools. By integrating surface web and dark web intelligence sources within a modular architecture, it overcomes fragmentation issues in current workflows. The standardized JSON output facilitates systematic analysis, while the API-mediated approach to dark web intelligence maintains ethical compliance.

Future enhancements will focus on AI integration, expanded intelligence domains, and improved visualization. The open-source nature encourages community contributions and continuous improvement. DarkReaper has potential applications in cybersecurity, law enforcement, and academic research, contributing to advancing cybersecurity education and practice.

Acknowledgement

I would like to thank the researchers and publishers for making their resources available. We are also grateful to our guide and reviewers for their valuable suggestions, and thank the college authorities for providing the required infrastructure and support.

References

1. SpiderFoot. (2025). *SpiderFoot – Open-source automated OSINT framework.* Retrieved August 27, 2025, from https://www.spiderfoot.net/
2. Pastor-Galindo, J., Nespoli, P., Gómez Mármol, F., & Martínez Pérez, G. (2020). The not yet exploited goldmine of OSINT: Opportunities, open challenges, and future trends. *IEEE Access*, *8*, 10282–10304. https://doi.org/10.1109/ACCESS.2020.2965257
3. Gopireddy, H. (2020). Dark web monitoring: Extracting and analyzing threat intelligence. *International Journal of Advanced Research in Computer Science*, *11*(3), 23–29. Retrieved from https://www.ijarcs.info/index.php/Ijarcs/article/view/7166
4. Vignesh, M., & Patidar, V. (2024). OSINT-based threat intelligence: Investigating leaked data on the dark web. *International Journal of Information Security and Privacy*, *18*(1), 1–13. https://doi.org/10.4018/IJISP.20240101.oa4
5. Browne, O., Chen, A., & Lavoie, N. (2024). A systematic review on research utilizing artificial intelligence for OSINT automation. *Journal of Defense Analytics and Logistics*, *8*(1), 101–120. https://doi.org/10.1108/JDAL-10-2023-0024
6. Nordin, M. R. N. M., Yusoff, R. N., & Ahmad, N. N. (2022). A review of open source intelligence (OSINT) tools and techniques for cybersecurity. *International Journal of Advanced Computer Science and Applications*, *13*(1), 147–156. https://doi.org/10.14569/IJACSA.2022.0130118
7. Răfailă, C., Gurzău, F., Grumăzescu, C., & Bica, I. (2023). MTAFinder – Unified OSINT platform for efficient data gathering. In *2023 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1–6). IEEE. https://doi.org/10.1109/ECAI58194.2023.10193922
8. Shin, S. M., & Jung, K. H. (2024). Framework of OSINT automation tool. *Journal of Information Technology Services*, *23*(2), 19–30. https://doi.org/10.9716/KITS.2024.23.2.019
9. Huang, Y.-T., et al. (2022). Open source intelligence for malicious behavior discovery and interpretation. *IEEE Transactions on Dependable and Secure Computing*, *19*(2), 776–789. https://doi.org/10.1109/TDSC.2020.3003928
10. Gangwar, S., & Pathania, T. S. (2018). Authentication of digital image using EXIF metadata. *International Journal of Computer Applications*, *181*(21), 1–4. https://doi.org/10.5120/ijca2018917964
11. Okmi, M., Por, L. Y., Ang, T. F., & Ku, C. S. (2024). Mobile phone data: A survey of techniques, features, and applications. *Sensors*, *24*(2), 584. https://doi.org/10.3390/s24020584