



International Journal of Research Publication and Reviews

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

Digital Shadow Detector: An AI-Driven Framework for Detecting and Assessing Unintended Digital Footprints

Mayuri Vhanmane, Amarja Jagtap, Manas Chavan, Harshada Jadhav, Rushika Burande

PG Students, Department of Computer science

ABSTRACT

The rapid growth of social media, cloud services, mobile applications, online transactions, and connected devices has increased the amount of information that individuals unintentionally leave across the Internet. This collection of publicly observable, residual, and indirectly associated information is referred to in this paper as a digital shadow. A digital shadow can include exposed usernames, old profiles, public metadata, breached-account indicators, reused aliases, public documents, and relationships between seemingly unrelated digital identities. This paper proposes Digital Shadow Detector (DSD), a defensive cybersecurity framework that combines lawful open-source intelligence (OSINT), entity resolution, risk scoring, and optional machine-learning classification to discover and prioritize an individual's publicly exposed digital footprint. The framework is designed around privacy-by-design principles: it uses only authorized or publicly accessible data, avoids credential attacks, and presents evidence and remediation guidance rather than exploiting discovered information. A prototype architecture is described with an Android/web client, Python backend, search and collection layer, normalization engine, risk-analysis module, database, and reporting layer. Evaluation is proposed using precision, recall, F1-score, duplicate-resolution accuracy, scan latency, and risk-prioritization quality. The work provides a practical foundation for personal cybersecurity, digital-footprint awareness, and defensive security assessments.

Keywords: Digital Shadow, Digital Footprint, OSINT, Cyber Security, Entity Resolution, Risk Scoring, Machine Learning, Privacy, Threat Intelligence

Introduction

Modern users maintain accounts on many platforms and routinely interact with websites, applications, payment services, professional networks, forums, and cloud systems. Even when a user deletes an account or stops using a service, fragments of information may remain publicly searchable or may be reproduced by other sources. These fragments form part of a digital shadow—the observable trail produced by a person's online activities, interactions, and associated identifiers.

The security significance of a digital shadow is that individual pieces of low-sensitivity information can become more revealing when correlated. For example, a public username may connect a professional profile to an old forum account, while a public document may reveal an email address or organizational relationship. The risk is therefore not limited to one exposed record; it also arises from correlation and aggregation.

Existing privacy tools commonly focus on isolated tasks such as password monitoring, breach alerts, or social-media privacy settings. A unified system that discovers, correlates, explains, and prioritizes a person's publicly exposed footprint can provide a more useful defensive view. Digital Shadow Detector addresses this gap.

Problem Statement

Users generally lack a centralized, understandable method to identify what information about their digital identity is publicly discoverable, how separate records may be related, and which exposures deserve attention first. Manual searching is time-consuming, inconsistent, and prone to duplicate or false associations. The proposed system aims to automate lawful discovery and risk assessment while minimizing unnecessary collection of sensitive information.

Objectives

Discover publicly accessible digital-footprint indicators associated with an authorized subject.

Normalize usernames, domains, emails, profile references, documents, and other indicators into a common representation.

Correlate records using confidence-based entity resolution rather than assuming that matching identifiers represent the same person.

Calculate explainable risk scores based on exposure type, confidence, sensitivity, and potential impact.

Use machine learning, where appropriate, to classify findings and reduce false positives.

Generate a human-readable report containing evidence, confidence, risk level, and recommended remediation.

Apply privacy-by-design controls, including authorization, data minimization, secure storage, and retention limits.

Research Questions

RQ1: Can multiple public digital-footprint indicators be correlated into a useful identity graph without relying on a single identifier?

RQ2: Can explainable risk scoring prioritize the findings that require the most urgent remediation?

RQ3: Can machine-learning classification reduce false positives in digital-shadow detection compared with simple rule matching?

RQ4: Can the framework provide useful results while limiting collection and retention of unnecessary personal information?

Existing System and Limitations

Current approaches include search-engine queries, individual platform privacy checks, breach-notification services, and manual OSINT investigations. These approaches are valuable but fragmented. Search results can be noisy; usernames can be shared by unrelated people; breach indicators may not explain the broader footprint; and manual correlation does not scale well.

Proposed System

Digital Shadow Detector is proposed as a modular defensive platform. The user first provides an authorized identity seed, such as a username or email address. The collection layer obtains permitted public information through compliant APIs or search interfaces. A normalization layer converts heterogeneous results into standardized indicators. The entity-resolution engine assigns confidence to possible relationships, while the risk engine evaluates exposure characteristics. Findings are stored with provenance and displayed through a dashboard and report generator.

System Architecture

[Android / Web Client]

|

v

[Authentication + Authorization]

|

v

[Python REST API / FastAPI]

|

+-----+-----+-----+

| | |

v v v

[OSINT/API] [Normalization] [Policy/Rate Limits]

| |

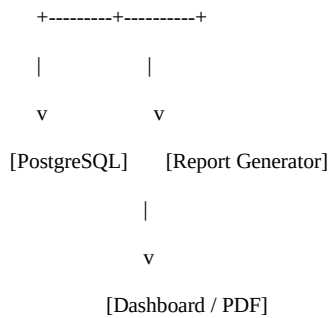
+-----> [Entity Resolution]

|

v

[Risk Engine]

|



Functional Modules

Methodology

Authorization and seed validation: verify that the scan is for the user or an explicitly authorized subject.

Collection: query permitted public sources using rate-limited connectors.

Extraction: identify candidate indicators such as usernames, domains, public emails, profile URLs, and document metadata.

Normalization: canonicalize values and remove obvious duplicates.

Correlation: compare indicators using exact and fuzzy matching features.

Classification: identify finding categories such as public profile, metadata exposure, reused identifier, or possible breach indicator.

Risk scoring: calculate a transparent score and explain the contributing factors.

Verification: flag low-confidence relationships for human review instead of presenting them as facts.

Reporting: display findings with evidence, confidence, date observed, risk rationale, and remediation.

Risk Scoring Model

A simple explainable score can be defined as $R = 0.30S + 0.25I + 0.20C + 0.15E + 0.10T$, where S is sensitivity, I is potential impact, C is correlation confidence, E is exposure breadth, and T is temporal relevance. Each factor can be normalized to 0–100. The weights should be calibrated using an evaluation dataset rather than treated as universal truths.

Example categories: 0–24 = informational, 25–49 = low, 50–74 = moderate, and 75–100 = high. These labels are operational priorities for the application, not claims that a user has been compromised.

Entity Resolution

A key research challenge is avoiding false associations. The same username may belong to multiple people. Therefore, DSD should combine independent signals. Candidate similarity may be calculated from normalized username similarity, domain context, profile-name similarity, public bio overlap, and other non-sensitive attributes. A record should be linked only when the combined confidence exceeds a configurable threshold. Borderline cases should remain unlinked until reviewed.

Machine Learning Component

An optional supervised classifier can categorize findings or estimate whether two records are likely to represent the same entity. Suitable baseline models include logistic regression, random forest, or gradient-boosted trees. Features should be privacy-conscious and interpretable. The dataset should contain both positive and negative identity-pair examples, with careful attention to class imbalance. Performance should be reported using precision, recall, F1-score, and a confusion matrix.

The ML model should support—not replace—human review. Training data must be lawfully obtained and de-identified where possible. The system should also record model version and confidence so that results are reproducible.

Database Design

Proposed Technology Stack

Frontend: Android application using Kotlin and Jetpack Compose, or a responsive web dashboard.

Backend: Python with FastAPI.

Database: PostgreSQL for structured findings and audit records.

Background processing: Celery/RQ or an equivalent task queue for scans.

ML: Python, scikit-learn, and pandas.

Authentication: OAuth/OIDC or Firebase Authentication.

Deployment: Docker-based services with HTTPS and secrets management.

Reporting: HTML/PDF report generation.

Algorithm / Pseudocode

INPUT: authorized seed S

1. Validate authorization and input.
 2. `sources <- collect_public_results(S)`
 3. `indicators <- extract_indicators(sources)`
 4. `indicators <- normalize_and_deduplicate(indicators)`
 5. `candidates <- generate_candidate_relationships(indicators)`
 6. FOR each candidate:
 - `confidence <- resolve_entity(candidate)`
 - IF `confidence >= threshold`:
 - `create_relationship(candidate, confidence)`
 7. `findings <- classify_and_score(indicators, relationships)`
 8. `attach_source_and_timestamp(findings)`
 9. rank findings by risk score and confidence
 10. generate remediation report
- OUTPUT: prioritized digital-shadow report

Security and Privacy Considerations

Only scan subjects for whom the operator has authorization.

Do not attempt password guessing, credential stuffing, bypassing access controls, or unauthorized scraping.

Use source terms and APIs in accordance with applicable terms and laws.

Minimize storage of sensitive data and encrypt data at rest and in transit.

Use role-based access control and maintain tamper-resistant audit logs.

Apply rate limiting and source-specific request limits.

Provide retention and deletion policies for scan results.

Clearly distinguish observed facts, inferred relationships, and uncertain matches.

Evaluation Plan

Expected Results

The proposed framework is expected to produce a consolidated inventory of publicly discoverable indicators, reduce duplicate records through normalization, improve correlation quality through confidence scoring, and make remediation more actionable through prioritized findings. These are research expectations rather than experimentally verified results. Actual accuracy and performance must be established using a controlled dataset and reproducible experiments.

Advantages

Centralized view of fragmented digital-footprint information.

Explainable risk scoring instead of opaque alerts.

Confidence-based correlation to reduce false identity matches.

Modular design allowing additional lawful data sources.

Useful for personal cybersecurity awareness and defensive assessments.

Privacy-by-design architecture with authorization and data minimization.

Limitations

Public data can be incomplete, outdated, or inaccurate.

Shared usernames can create false associations.

Source availability and API policies may change.

Machine-learning performance depends on the quality and representativeness of labeled data.

A digital-shadow scan cannot prove that an account or system has been compromised.

Legal and platform-policy requirements vary by jurisdiction and service.

Future Scope

Privacy-preserving identity graphs and stronger false-positive controls.

Natural-language explanations of why a finding was flagged.

Continuous monitoring with user-controlled alerts.

Integration with password-manager and security-awareness workflows.

Federated or privacy-preserving ML for model improvement.

Graph-based anomaly detection for unusual relationships.

Automated remediation checklists and user verification workflows.

Conclusion

Digital Shadow Detector presents a defensive approach to understanding the information that an individual may unintentionally expose through online activity. The proposed architecture combines public-data discovery, normalization, entity resolution, risk scoring, optional machine learning, and explainable reporting. Its central research contribution is treating digital-shadow analysis as a correlation and risk-prioritization problem rather than simply a search problem. Future implementation and evaluation should focus on measurable accuracy, reproducibility, privacy protection, and strict authorization boundaries.

References

- National Institute of Standards and Technology (NIST), Cybersecurity Framework (CSF) 2.0, 2024.
- National Institute of Standards and Technology (NIST), Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management.
- OWASP Foundation, OWASP Application Security Verification Standard (ASVS).
- ENISA, Threat Landscape and cybersecurity risk-management publications.
- Clarke, R., "Information Technology and Dataveillance," Communications of the ACM, 1988.
- Solove, D. J., "A Taxonomy of Privacy," University of Pennsylvania Law Review, 2006.
- Bishop, M., Computer Security: Art and Science, Addison-Wesley.
- Stallings, W., Effective Cybersecurity: A Guide to Using Best Practices and Standards.

Appendix A – Sample Finding

Appendix A – Sample Finding

Finding Type: Public profile association

Observed Indicator: Example username

Confidence: 0.86

Risk Factors: public visibility + cross-source association

Evidence: source reference and observation timestamp

Recommended Action: review profile visibility, remove unnecessary personal information, and avoid identifier reuse.

— End of Research Paper —