*electronics*

MDPI

*Article*

# RDBAlert: An AI-Driven Automated Tool for Effective Identification of Victims' Personal Information in Ransomware Data Breaches

**Juan Manuel Tejada-Triviño, Elvira Castillo-Fernández, Pedro García-Teodoro and José Antonio Gómez-Hernández ***

Network Engineering & Security Group, CITIC, University of Granada, 18014 Granada, Spain; juanma@darkeye.io (J.M.T.-T.); elviracastillo@ugr.es (E.C.-F.); pgteodor@ugr.es (P.G.-T.)
* Correspondence: jagomez@ugr.es

**Abstract**

Ransomware attacks are increasingly resulting in the public leakage of sensitive personal data, affecting both individuals and organizations worldwide. Aimed to inform victims when their personal information is compromised, this paper introduces *RDBAlert*, a rapid and efficient practical tool that automates the extraction of multimodal personal data from ransomware leak repositories, enabling victims to mitigate damage early and take necessary precautions to protect themselves from further harm. The comprehensive and modular nature of this novel tool contributes several notable features: *(i)* automation of ransomware data leak detection; *(ii)* analysis of information in multiple formats and languages by integrating well-known OCR, text/PDF, and image recognition, as well as multimodal currently available AI-related tools; *(iii)* user-friendly interface for quick and efficient analysis; and *(iv)* ability to gather forensic evidence for studying security incidents. In addition to the flexible nature of *RDBAlert*–as each module can be replaced or upgraded with potentially more effective solutions without impacting the overall service–experimental results show that it is highly effective at identifying personal information, which will contribute to the mitigation of ransomware attack consequences.

**Keywords:** ransomware; data breach; leak; LMM



Academic Editor: Cheonshik Kim

Received: 27 September 2025
Revised: 14 October 2025
Accepted: 22 October 2025
Published: 4 November 2025

**Citation:** Tejada-Triviño, J.M.; Castillo-Fernández, E.; García-Teodoro, P.; Gómez-Hernández, J.A. RDBAlert: An AI-Driven Automated Tool for Effective Identification of Victims' Personal Information in Ransomware Data Breaches. *Electronics* **2025**, *14*, 4327. https://doi.org/10.3390/electronics14214327



## 1. Introduction

Numerous studies affirm that ransomware continues to be one of the most significant cybersecurity threats worldwide [1–3]. In 2024 alone, a total of 5414 ransomware incidents were reported worldwide, representing an 11% increase compared to the previous year [4]. The ransomware economy has grown into a highly lucrative criminal enterprise, with industry estimates indicating that cybercriminals extorted over 1 billion USD from victims worldwide in 2023 [5].

Ransomware is a type of malicious software that restricts access to a system or its files, effectively locking users out until a ransom is paid [6]. Attackers typically demand payment in exchange for restoring access, with the threat that, without payment, the data will be permanently lost or publicly disclosed. Over time, ransomware has evolved from simple single-extortion tactics to more sophisticated triple-extortion schemes. In these cases, cybercriminals not only encrypt victims' data but also threaten to expose sensitive information on public data leak sites (DLSs), sell it on underground markets hosted on the dark web or *.onion* domains, and increasingly target organizations rather than individuals [7,8].

These attacks inflict a wide range of damage, including physical, financial, reputational, psychological, and social damage. In severe cases, the disruption and costs caused by such attacks can be so overwhelming that they threaten the very survival of the affected organizations, forcing some to cease operations permanently.

Given the relevance of the situation, the European General Data Protection Regulation (GDPR) (EU 2016/679) [9] introduced the requirement for organizations to notify (in 72 h) the relevant national supervisory authority in the event of a personal data breach (the European GDPR defines a "personal data breach" in Article 4(12) as *"a breach of security leading to accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed."* [10]. Other international regulatory frameworks such as that of the USA (California Consumer Privacy Act, CCPA) [11], Canada (Personal Information Protection and Electronic Documents Act, PIPEDA) [12], Japan (Act on the Protection of Personal Information, APPI) [13], or Brazil (General Data Protection Law, LGPD) [14] establish similar obligations regarding breach notification and the ethical management of data, thus strengthening transparency and ensuring greater control over personal information.

The aforementioned notification process offers several advantages, such as enabling organizations to plan ahead, implement processes for the timely detection and containment of breaches, gather forensic evidences for the breach [15], and assess the risk posed to individuals. In addition to the requirement of direct notification by affected organizations, there are some initiatives to proactively search, announce, and analyze DLSs to detect and notify potential victims of a data breach. Developing an automated tool to monitor, download, and analyze exfiltrated data from ransomware attacks is strategically crucial for contemporary cybersecurity efforts. This necessity stems from the growing prevalence of double and triple extortion tactics—previously discussed—which are now observed in approximately 80% of ransomware incidents [16].

Although further fundamental research remains important, the design and deployment of practical tools are essential to effectively counteract the current threat of ransomware. In this line, some proactive DLS search and analysis tools are available, but these approaches are not specific for ransomware-related incidents, while they are very limited by the heterogeneous nature of information in such repositories, which may include plain text, PDFs, images, and more. Primarily designed to address and overcome these limitations, the present work introduces *RDBAlert* (short for 'Ransomware-related Data Breach Alert'), a comprehensive, modular, and web-based alert system that

1. detects and maintains links of ransomware-related DLSs in an autonomous and dynamic way over time;
2. allows users (individuals or organizations) to introduce/provide potential stolen personal data, so that it
   (a) collects data from the DLS repositories;
   (b) automatically analyzes the associated documents by using AI-based tools to extract the personal information specified by the user.

The remainder of this paper is organized as follows. Section 2 reviews various approaches and advancements in the literature concerning ransomware and data breaches. In Section 3, we introduce *RDBAlert*, an innovative automated system that utilizes AI-driven public tools to analyze the diverse information typically found in data leak sites (DLSs), aiming to alert victims when personal data is detected in ransomware-related breaches. Section 4 assesses the performance of RDBAlert, highlighting its significant potential in combating ransomware and discussing some relevant misuse- and ethical-related aspects. Finally, Section 5 summarizes the work's contributions and suggests avenues for future research.

## 2. Background

As highlighted in reports such as [17], ransomware is a leading cause of data breaches at present. In fact, ransomware attacks involving data exfiltration have increased by 96% since 2019, becoming the prevailing trend [18,19]. Beyond the profits gained from ransom payments, academic research shows that cybercriminals often monetize stolen data by selling it on online marketplaces or exploiting it for various other malicious purposes. Each attack group typically chooses one or more of the following options as the destination for exfiltrated data [20]: *(i)* posting data on leak sites, *(ii)* posting or publicizing data leaks on underground forums, *(iii)* publicizing data leaks on social media, and *(iv)* selling or auctioning data.

Given the relevance of the situation, the existing regulatory frameworks impose stringent penalties for non-compliance with their provisions. The European Union's General Data Protection Regulation (GDPR) authorizes fines of up to EUR 20 million or 4% of the organization's annual global turnover, whichever is higher, in cases of severe violations, such as failure to notify breaches promptly or inadequate data protection measures [9]. In countries like Spain, the LOPDGDD complements these penalties with additional corrective measures, such as suspending data processing activities. In the United States, HIPAA enforces tiered penalties that can reach 1.5 million annually for repeated violations, while the CCPA allows fines of up to USD 7500 per intentional infringement. Under Canada's PIPEDA, administrative penalties of up to CAD 100,000 may be levied for failure to report breaches or negligent data handling. Japan's APPI imposes fines of up to JPY 1 million (for individuals) or JPY 100 million (for corporations) for improper management of personal information. In Brazil, the LGPD stipulates fines of up to 2% of a company's revenue, capped at 50 million BRL per violation, alongside measures such as prohibiting data processing activities.

The aforementioned provisions highlight the growing need for automated mechanisms to mitigate risks and promote proactive compliance in an increasingly stringent and globalized regulatory environment. In this context, several approaches have been conceived to prevent and respond to security incidents—particularly data breaches [21]. Moreover, since regulation usually requires notification of data breaches, there are a number of lists and repositories [22–26] mainly intended to help organizations tackle the threat of cyber crime and understand and mitigate it [27]. In particular, in January 2024, a massive data breach—dubbed the "Mother of All Breaches"—exposed more than 26 billion records, affecting major platforms such as Twitter, Adobe, Canva, LinkedIn, and Dropbox. Hundreds of thousands of personal records, including names, addresses, health information, and banking data, were compromised. As evident, the impact of data breaches is especially significant in the healthcare industry, where the sensitivity of personal data amplifies both the risks and consequences of such incidents [28,29].

Paying the ransom may seem like the only way to minimize disruption and regain control for organizations or individuals affected. However, from a societal standpoint, paying ransoms only fuels the business model of these criminal groups, making ransomware more profitable and widespread [5]. Cybercriminals are increasingly employing advanced technologies and sophisticated techniques to transfer and conceal their illicit profits. At this stage, disrupting the payment cycle becomes essential to prevent these funds from reaching the perpetrators and to hinder the expansion of their criminal operations.

In this overall context, the relevance of collaboration and sharing information on ransomware incidents is highlighted in [30]. This collaboration is often expressed in terms of tools intended to investigate security incidents and help mitigate them, like the aforementioned lists and repositories. This is the particular case of several intelligence platforms designed for the analysis of data breaches, like those in Table 1.

**Table 1.** Intelligence platforms for data breach searches, where the main features are provided from websites; tutorials and documentation are available for each of the tools.

| Platform | Description | Main Features |
| --- | --- | --- |
| AmILeaked | Similar to "Have I Been Pwned" below, this service allows users to verify whether their email or password has appeared in a known data breach. | - Strengths: simple UI for quick checks; personal alerts/monitoring options.<br>- Access: free checks + paid monitoring/business plans.<br>- Caveat: smaller footprint than large commercial engines. |
| DeHashed | A search engine specialized in compromised datasets. It enables users to check if their personal information, such as emails or passwords, has been exposed in data breaches. | - Strengths: powerful filtering, programmatic API, monitoring/alerts; commonly used in incident response.<br>- Access: freemium (limited free lookups) with paid tiers and API.<br>- Caveat: paywall for full results and commercial use; always handle PII legally. |
| GhostProject | An online dataset that allows users to search for compromised passwords and data using email addresses or usernames. It helps determine if personal information has been exposed in data leaks. | - Strengths: very large record counts (advertises billions of records); quick credential/password lookup.<br>- Access: web UI—model varies (some features gated).<br>- Caveat: provenance and legality of some indexed dumps is unclear; treat outputs carefully. |
| Have I Been Pwned? | A free service that allows users to check whether their email address or phone number has been involved in a known security breach. It provides alerts and advice to protect their accounts. | - Strengths: authoritative, transparent methodology, free checks, enterprise/notification APIs, strong privacy-aware APIs (password API uses k-anonymity).<br>- Access: free for casual queries; paid API/notify options for enterprise.<br>- Caveat: HIBP only indexes breaches it has validated/ingested (not every private dump). |
| IntelX | A data intelligence and search platform that allows users to explore a wide variety of sources, including data breaches, WHOIS records, documents, and more. IntelX aids in digital investigations and in retrieving hard-to-find information. | - Strengths: selectors-based search (email, IP, Bitcoin, IPFS), historical archive, darknet coverage—useful for deep OSINT and historical tracing.<br>- Access: free basic search; PRO features for darknet/history/advanced filters require subscription.<br>- Caveat: the breadth is powerful but may surface outdated/duplicated items; requires experienced filtering. |
| Leak Lookup | A service that provides access to multiple leaked datasets to check if personal information has been compromised. It is useful for security investigations and identity protection. | - Strengths: large record counts, domain monitoring and business-focused subscriptions.<br>- Access: account/login required; paid monitoring tiers.<br>- Caveat: commercial service—depth and freshness depend on subscription level. |
| Ransomwhere | A platform that tracks and aggregates data on ransom payments related to ransomware attacks. It helps understand the financial impact of these attacks and promotes transparency in cybersecurity. | - Strengths: focused ransomware dataset, downloadable data and reporting options for researchers.<br>- Access: web access; data export options on site.<br>- Caveat: narrow scope (ransomware-related)—not a general credential search engine. |
| Recon-ng | A complete web reconnaissance framework that includes modules to search for information in public datasets and third-party services, useful for security professionals and pentesters. | - Strengths: modular, scriptable, integrates with APIs/services for enrichment; great for structured investigations and pivoting.<br>- Access: open-source (run locally), requires modules/API keys for some external services.<br>- Caveat: not a raw breach dump index—it aggregates OSINT and can query breach services when configured. |
| Scylla | A search engine for compromised information that enables users to search across multiple leaked datasets simultaneously. It offers an API for custom integrations. | - Strengths: username/social-profile pivoting, Shodan integration, geolocation features—useful for profile enumeration.<br>- Access: open-source GitHub projects or custom installs.<br>- Caveat: name ambiguity—be sure you mean the OSINT tool, not ScyllaDB. Handle results ethically. |

To the best of our knowledge, there is no available comparative analysis of the aforementioned tools. Therefore, we include in the table a summary of their main general features, compiled after consulting the official websites, tutorials, and available documenta-

tion for each platform. In any case, current intelligence platforms for data breach search present two principal limitations:

- Most of them are not specifically focused on ransomware-related leaks, which constitutes a major issue as users whose information is exposed through that typology of attacks also have the right to know about it promptly.
- The information publicly exposed in DLSs is highly heterogeneous, encompassing a variety of formats such as text, PDFs, images, and more. As a result, personal data often remains inaccessible to most existing automated analysis tools.

This way, an integral, multi-format specialized analysis tool for ransomware data breaches alerts would allow for the detection of these leaks in their early stages, before compromised information is widely disseminated. This would enable the prompt initiation of mitigation measures, such as notifying affected entities, revoking compromised access credentials, or applying containment controls. Therefore, it would help reduce operational damage, financial loss, and reputational harm, which are key elements in effective cyber risk management.

Although further fundamental research is still needed to combat the ransomware threat, there is an urgent societal need for the deployment of effective, practical tools. In this regard, to the best of our knowledge, only a few existing tools offer partial or narrowly focused security solutions. With the aim of addressing the current significant gap in cybersecurity tools capable of providing rapid and automated alerts for breaches disclosed by ransomware groups, the following section introduces *RDBAlert*, a comprehensive and novel automated system characterized by the following main characteristics:

- It features a flexible, modular architecture, allowing each module to be replaced or upgraded with potentially more effective solutions without affecting the overall service.
- Although some of the modules are specifically developed by the authors (e.g., the crawler and the Kibana-based visualization module), our solution leverages public AI-powered tools to analyze the variety of information typically found in ransomware-related DLSs. In particular, a number of well-known tools for OCR-based character recognition, image analysis, LLM-based text analysis, etc., are considered to confirm the existence of personal data in these data lakes.
- This core functionality is enhanced by a web-based interface that allows users to input their personal data for searches within DLSs.

## 3. RDBAlert: A Novel Tool to Identify Personal Data in Ransomware Leaks

The protection of personal data constitutes a critical dimension of cybersecurity governance. Exfiltrated documents usually contain sensitive information, such as social security numbers, medical histories, or financial records, whose unauthorized disclosure entails significant legal, economic, and even personal risks. The development of an automated tool capable of locating exfiltrated data, downloading and analyzing it, and detecting structured patterns of personally identifiable information (PII) would facilitate the generation of structured reports, thereby expediting compliance with international data protection frameworks while minimizing the impact on organizations and individuals.

The tool *RDBAlert* (which stands for 'Ransomware Data Breach Alert') introduced in this work integrates a set of AI-based powerful and critical functionalities intended to optimize the identification, management, and analysis of exfiltrated data in ransomware incidents. The integral modular approach used for RDBAlert enables scalability in handling increasing volumes of data. The resulting novel tool acts as a proactive resource for

researchers and incident response teams, enabling the early detection of security breaches and supporting the mitigation of risks associated with ransomware.

### *3.1. RDBAlert Architecture*

As illustrated in Figure 1, the system is composed of four main modules:

- A **crawler**, to automatically navigate platforms that host data leaks, such as forums on the dark web or *.onion* sites, employing anonymization technologies such as Tor or distributed proxies to ensure secure and ethical access to these environments.
- An **automated search and classification module**, which is responsible for decompressing the retrieved files and applying structured inspection techniques to identify PII within documents, databases, and other formats contained within them.
- A **flexible storage architecture and advanced query module**, in charge of identifying sensitive data along with contextual metadata such as the original file name, its hierarchical location within the folder structure, the originating platform, and timestamps. All of them are integrated into a NoSQL database designed to support multidimensional queries.
- A **reporting module**, for structured and actionable insights, designed to transform query results from the database into structured, visually accessible, and technically detailed outputs.

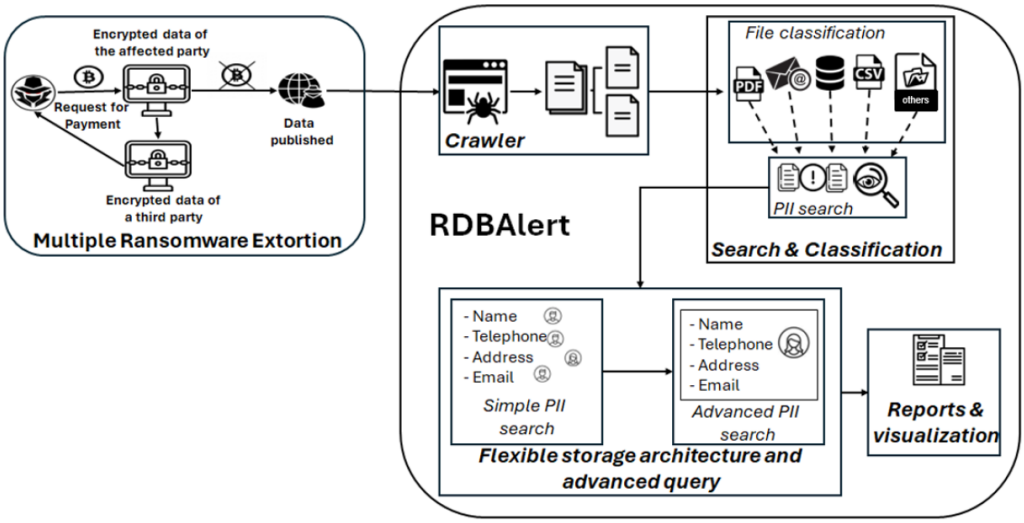



**Figure 1.** RDBAlert's functional architecture and modules.

The **crawler module** [31] operates exclusively on publicly available information without funding illicit activities. It not only identifies and retrieves exfiltrated data packages when accessible but is also enhanced through its integration with external intelligence sources and platforms such as *DarkOwl* [32] and *Recorded Future* [33]. These integrations enable the system to enrich tracking with up-to-date indicators of compromise (hereafter, IoC) and contextualize identified threats by correlating file metadata such as size, type, and creation dates with documented malicious activity patterns.

Additionally, the module indexes the provenance of the data, including specific URLs or locations within anonymous networks, facilitating robust forensic analysis and incident traceability. This synergy between automated data collection and intelligent contextualization positions the tool as a proactive resource for the early identification of breaches and the mitigation of ransomware-related risks.

Once the download is complete, the system activates the **automated search and classification module**. By leveraging regular expressions, machine learning models, and preconfigured patterns, the module detects sensitive elements such as identification numbers, credit card details, physical addresses, full names, and email addresses. When such data are identified, the corresponding file is stored within a hierarchical folder structure categorized according to its nature (e.g., PDF documents, XML, or databases). This approach optimizes subsequent review processes and ensures an organized and systematic handling of digital evidence.

The tool enhances its analytical capabilities through a **flexible storage architecture and advanced query module**. After decompression, classification, and the detection of PII, the identified sensitive data are integrated into a NoSQL database designed to support multidimensional queries. This repository employs a denormalized, document-oriented structure that associates each PII record (e.g., an identification number, address, or email) with descriptive attributes of the containing file, including its cryptographic hash, file format type, and contextual metadata about its origin in the dark web. This schema-free design enables not only simple queries—such as retrieving all PII linked to a specific document—but also complex, dynamic queries that correlate fragmented data points across diverse sources. For example, the system leverages native NoSQL indexing and aggregation capabilities to retrieve all occurrences of a given data element, presenting related details (e.g., associated names and addresses) together with comprehensive source metadata. This includes the data's traceable origins, from underground forums to data auction sites. By avoiding rigid table relationships, the NoSQL framework allows the tool to map the trajectory of exfiltrated information and assess its propagation without relying on traditional relational joins.

Furthermore, the NoSQL database leverages native indexing mechanisms, such as document-oriented indexing on exposure dates or PII types, to accelerate the identification of temporal or categorical patterns in incidents. This capability is further strengthened by event-driven audit logs, which record data modifications at scale and provide granular traceability, facilitating comprehensive forensic investigations. The integration of these results with external intelligence sources (IoCs or threat reports) further enriches contextual impact analyses, identifies recurring affected entities, and dynamically prioritizes responses to breaches. By systematically documenting violations through schemaless data models, the approach not only streamlines compliance with data protection regulations (e.g., GDPR) but also establishes an empirical foundation for analyzing trends in ransomware tactics.

RDBAlert consolidates its operational utility through a **reporting module**, which processes retrieved information including specific PII records, associated metadata, and provenance sources to generate dynamic reports tailored to the needs of various end-users, such as forensic investigators, incident response teams, or regulatory compliance auditors.

The reports follow a hierarchical structure, prioritizing clarity and actionability: from executive summaries highlighting critical findings (e.g., volumes of exposed data or affected entities) to technical sections detailing file access paths, breach correlations, and evidence of data reuse across multiple attacks.

Additionally, the module integrates interactive visualizations, including geographic distribution maps of victims, temporal exposure graphs, and correlation networks of compromised entities, facilitating the identification of complex patterns. Each report can be exported in standardized formats (PDF, CSV, or JSON) and customized through advanced filters such as date ranges, PII types, or source platforms, ensuring its usability in legal contexts, where precision and documentation are essential.

This component not only automates forensic documentation generation but also serves as a bridge between technical analysis and strategic decision-making, enabling

organizations to prioritize breach responses, optimize security controls, and demonstrate due diligence in audits.

### *3.2. RDBAlert Implementation*

The detailed RDBAlert's operational flow is shown in Figure 2. The tool is fully developed in Python, taking advantage of the language's modularity and extensibility. The architecture integrates open-source software with custom scripts to create a cohesive system, where the output of each module is passed sequentially to the next. This processing pipeline covers the entire lifecycle of handling ransomware-related leaks, from initial identification to advanced forensic analysis. The tool operates exclusively with data from open sources, ensuring transparency and adherence to ethical and legal standards. The following sections provide detailed insights into the implementation of the core components of the system.

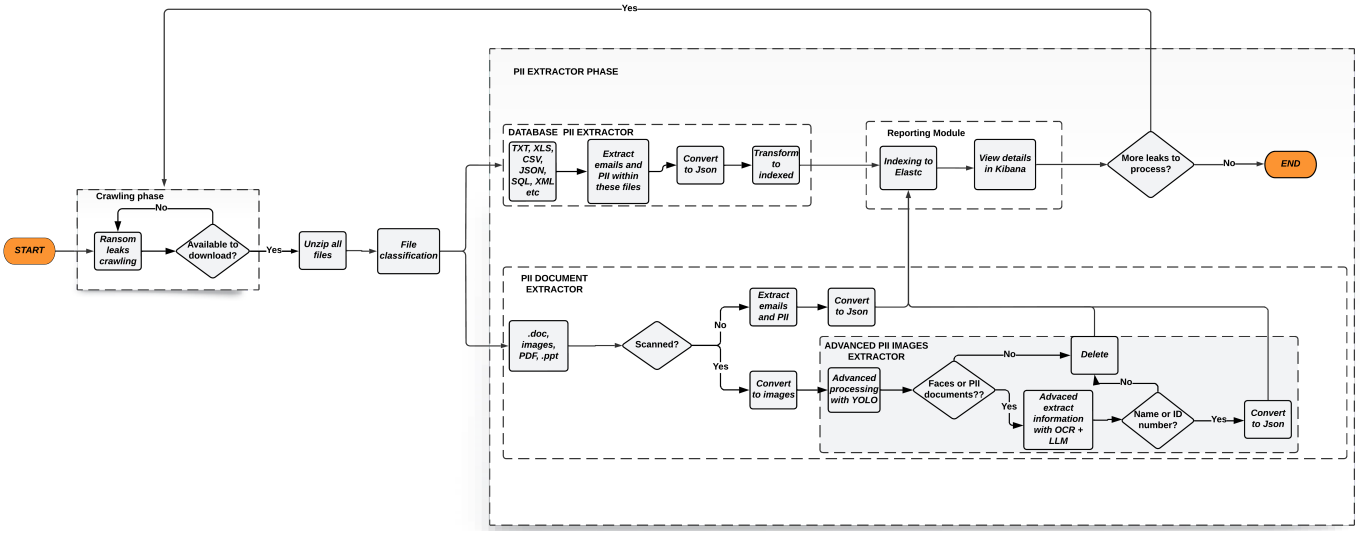



**Figure 2.** RDBAlert's operational flow.

#### 3.2.1. Crawler Module

Tracking ransomware groups and their activities on the dark web is crucial for threat intelligence and cybersecurity. This way, a principal component in RDBAlert is the *crawler module*, responsible for **tracking, identifying, and automatically downloading leaked data from specialized platforms**. To achieve this, we integrate open-source trackers such as *Ransomwatch*, *Breach House*, *Ransomware.live*, and *Ransomlook.io*, which monitor the emergence of new leaks in real time across dark web forums, underground markets, and *.onion* sites (see Table 2).

From an implementation standpoint, the crawler operates in a continuous *curl loop* through a TOR proxy and employs preconfigured regular expressions built from the distinctive signatures of each ransomware group to detect leaks. If an error or connection loss occurs, the system retries on an hourly schedule and recovers data in subsequent cycles. The design prioritizes resilience and continuity over raw speed to ensure comprehensive coverage of active ransomware seeds. For dark web coverage the crawler uses a curated set of seed URLs pointing to official ransomware blogs on the TOR network (*.onion* domains) and routes traffic through a proxy configured with *torsocks* on port 4001. Downloads are performed by emulating real browser behavior to avoid user-agent filters and suspicious headers.

**Table 2.** Tools for monitoring ransomware sites.

| Name | Description | Link |
| --- | --- | --- |
| **Breach house** | An advanced ransomware site monitoring system that collects information about active groups, recent attacks, and extortion trends on the dark web. Its focus is to provide real-time data to researchers and cybersecurity professionals. | https://breach.house (accessed on 14 October 2025) |
| **Ransomlook.io** | A web-based ransomware tracker that facilitates the observation of ransomware group activity in real time. It enables researchers and analysts to gain insights into new leaks and attack patterns on the dark web. | https://www.ransomlook.io/recent (accessed on 14 October 2025) |
| **Ransomware.live** | A ransomware monitoring portal that tracks the activities of major extortion groups. It is constantly updated and allows visualization of trends in the growth of these groups and their victims. | https://github.com/JMousqueton/ransomware.live (accessed on 14 October 2025) |
| **Ransomwatch** | An open-source tool that tracks and archives posts from ransomware groups on the dark web. It monitors multiple known group sites and provides real-time alerts on new data leaks. | https://github.com/joshhighet/ransomwatch (accessed on 14 October 2025) |

Once a leak is detected, the system initiates a **structured download process** utilizing a script based on `Aria2-Onion-Downloader` [34], optimized for efficient transfers over the *Tor* network. To mitigate bandwidth limitations inherent to this network, complementary solutions are implemented:

- `Torget` [35], a command-line tool that accelerates downloads through parallel connections and dynamic session management.
- `Torboost` [36], designed to optimize performance by employing multi-threaded execution and fine-tuned network parameter adjustments.

#### 3.2.2. Classification and Search Module

Once the full leak download is complete, a *classification and search module* is activated. This component starts by **recursively extracting files** stored in compressed formats (e.g., .zip, .7z, .tar, .gz, .rar), ensuring that all content is fully extracted, including files within nested archives. This step is essential for accessing the raw data and verifying its integrity prior to further analysis.

After extraction, a progressive and layered **filtering process** is applied, which is aimed to maximize operational efficiency while minimizing processing errors. Files are categorized into specific types, such as plain text, structured documents (e.g., DOCX), PDFs, images, and databases. This segmentation enables the use of analytical techniques tailored to the characteristics of each file format, enhancing the accuracy and efficiency of identifying relevant information.

As a result, the system compiles a comprehensive dataset that includes internal communications, invoices, and even confidential documents from the compromised entity/user. Nevertheless, the primary focus remains on detecting PII, which usually represents only a small portion of the overall dataset in large-scale data leaks (e.g., 1 GB within a 10 TB leak).

To avoid retaining excessive non-essential data—since PII detection is the primary objective—all extracted files are processed by the *PII Search submodule*. This submodule

performs a multi-stage search within each file. If PII is detected, the file is retained; otherwise, it is deleted to optimize disk space usage.

**PII identification** in extracted files is carried out using differentiated strategies, each tailored to the structural and semantic characteristics of the specific file type. The methodological approaches applied to each category are as follows:

1. Plain Text Files: The analysis of PII in plain text files (e.g., .txt, .log) relies on regular expressions (*regex*) designed in accordance with international standards. For instance, email addresses are detected using patterns based on the RFC 2822 specification [37], which defines the syntactic structure of valid email formats.
   Each identified email address is normalized into a JSON format that includes specifically contextual metadata, a 100-character snippet before and after the match. This context allows for inference of the email's usage within the text (e.g., in forms or internal communications). To prevent redundancy in the database, an SHA-1 hash is generated for each record, ensuring uniqueness by cryptographically encoding both the content and its surrounding context.
2. PDF Files: PDFs are classified into two categories based on their content structure:
   - Digitally generated PDFs (containing embedded text) are processed similarly to plain text files. Regex-based PII detection is applied, and the results are structured as JSON entities for standardized representation.
   - Scanned PDFs (either partially or entirely composed of images) require a computer vision-based workflow. When textual content is not directly accessible, the system leverages YOLO ('You Only Look Once' [38])—see paragraph "Multimodal recognition tools" below—a convolutional neural network (CNN) architecture optimized for real-time object detection. YOLO is used to identify visual elements such as faces, passports, or regional identity documents within the scanned content. Any PDF containing such elements is converted into individual image files (in PNG or JPEG format) and then forwarded for advanced processing.
3. Documents and Images: For file formats such as DOC, JPG, or PNG, the same approach used for scanned PDFs is applied. YOLO is employed to scan these files and detect three critical categories: faces, passports, and identification documents. Images that yield positive matches are tagged and prepared for the text extraction phase.
4. Databases and Structured Formats: For files such as CSV, XML, SQL, or JSON, the analysis follows a dual-strategy approach:
   - *Header-based detection*: If column names (e.g., 'email', 'ID number', 'phone number') indicate the presence of PII, the corresponding fields are automatically extracted.
   - *Heuristic search*: When headers are ambiguous or absent, the first 30 lines of the file are scanned using targeted *regex* patterns. This statistical sampling enables the system to infer sensitive data patterns within unlabeled columns.

In all cases, the extracted data is normalized into JSON format, preserving metadata such as row and column indices as well as the original file format. To ensure compatibility with NoSQL or hierarchical databases (e.g., MongoDB), the system employs custom adapters that convert relational database structures into standardized JSON-based schemas. This approach enables seamless integration of relational data models into NoSQL environments by encapsulating tabular information within nested, document-oriented formats.

**Multimodal Recognition Tools**

Files containing detected sensitive information, such as images, scanned PDFs, or labeled documents, are specifically analyzed here using two typologies of tools: **image and optical character recognition** and **face detection**.

Table 3 shows several available traditional and AI-based multimodal tools for OCR and PDF/image analysis. Traditional recognition/detection systems like OCR-based tools are prone to various errors, especially when dealing with low-quality images, handwritten text, unusual fonts, or complex document layouts. As a result, there is an increasing demand for AI-driven tools, especially LMM-based (Large Multimodal Model) approaches [39]. Moreover, due to the sensitive nature of many processed documents, commercial multimodal AI models are generally not a viable option because of concerns regarding data privacy, security, and vendor lock-in.

**Table 3.** Traditional and AI-based multimodal tools for OCR and text/PDF/image analysis.

| Name | Description | Link |
| --- | --- | --- |
| **ABBYY FineReader** | A leading OCR software offering high text recognition accuracy. It supports a wide range of languages and document formats, allowing scanned PDFs to be converted into searchable and editable documents. | https://pdf.abbyy.com/es (accessed on 14 October 2025) |
| **Amazon Textract** | Amazon Textract is a powerful machine learning service provided by AWS that allows you to automatically extract text, forms, and tables from scanned documents and images. | https://aws.amazon.com/textract/?nc1=h_ls (accessed on 14 October 2025) |
| **Detectron2** | An object detection and recognition system developed by Facebook AI Research. It is flexible with high performance for segmentation and object detection tasks. | https://github.com/facebookresearch/detectron2 (accessed on 14 October 2025) |
| **EasyOCR** | An easy-to-use OCR library supporting over 80 languages. It is based on PyTorch and ideal for projects requiring fast implementation. | https://github.com/JaidedAI/EasyOCR (accessed on 14 October 2025) |
| **Google Cloud Vision API** | A cloud service offering OCR and image analysis capabilities. It can extract text from images and PDFs and is scalable for large volumes of data. | https://cloud.google.com/vision?hl=en (accessed on 14 October 2025) |
| **H2O-VL Mississippi-2B** | A vision–language model (VLM) optimized for text recognition and document-oriented visual question answering (VQA). It is designed for advanced multimodal reasoning over documents and images. | https://h2o.ai/platform/mississippi (accessed on 14 October 2025) |
| **InternVL2-5-MPO** | A powerful multimodal AI model with enhanced OCR and scene text recognition capabilities. It provides high accuracy in extracting structured and unstructured text from complex visual documents. | https://internvl.github.io/blog/2024-12-20-InternVL-2.5-MPO (accessed on 14 October 2025) |
| **Keras OCR** | Provides a set of tools to build OCR systems using Keras and TensorFlow. It includes pre-trained models and simplifies the training of custom models. | https://keras-ocr.readthedocs.io/en/latest (accessed on 14 October 2025) |
| **LLaVA (Large Language and Vision Assistant)** | A state-of-the-art model that integrates vision and language understanding, allowing for document interpretation, OCR, and VQA. It is widely used for AI-driven document processing and multimodal learning. | https://llava-vl.github.io (accessed on 14 October 2025) |
| **Mediapipe (Google)** | In addition to OCR, it offers advanced image analysis capabilities, including object detection, content tagging, facial recognition, and logo detection. | https://github.com/google-ai-edge/mediapipe (accessed on 14 October 2025) |
| **Microsoft Azure Computer Vision** | A unified service that offers innovative computer vision capabilities. Give your apps the ability to analyze images, read text, and detect faces with prebuilt image tagging, text extraction with optical character recognition (OCR), and responsible facial recognition. | https://azure.microsoft.com/en-us/products/ai-services/ai-vision (accessed on 14 October 2025) |

**Table 3.** *Cont.*

| Name | Description | Link |
| --- | --- | --- |
| **MiniCPM** | A large multimodal model (LMM) designed for document and image understanding. MiniCPM incorporates OCR capabilities and can perform complex reasoning over textual and visual data. | https://github.com/OpenBMB/MiniCPM-o (accessed on 14 October 2025) |
| **MiniMonkey** | A multimodal model trained for visionlanguage tasks, including OCR, document understanding, and scene text recognition. It integrates with various AI pipelines to extract and analyze information from images. | https://huggingface.co/mx262/MiniMonkey (accessed on 14 October 2025) |
| **Tesseract OCR** | An open-source OCR engine developed by Google. Despite being free, it is highly powerful and supports over 100 languages. It can be integrated into custom applications and is ideal for developers. | https://github.com/tesseract-ocr/tesseract (accessed on 14 October 2025) |
| **YOLO (You Only Look Once)** | A real-time object detection algorithm that utilizes convolutional neural networks. It is highly efficient and used to identify and locate multiple objects within an image with high precision. | https://docs.ultralytics.com (accessed on 14 October 2025) |

From the above, in RDBAlert, we have utilized *YOLOv11* [38] for face detection and personal identification document recognition in diverse environments, and *MiniCPM* [40] for OCR recognition specifically adapted to the morphological and contextual characteristics of personal identification documents.

The implemented workflow follows a hybrid sequential approach (see Figure 2):

1. Hierarchical Detection Phase: YOLOv11 follows a two-level detection strategy. First, it identifies the entire document within the image using global bounding boxes. Then, it segments the critical subregions (such as the name field, identification number, and photograph) using normalized relative coordinates. This process is supported by a spatial reference system that preserves the geometric proportions of the document, regardless of its orientation or scale.
2. Advanced Text Extraction Phase: Each detected subregion is processed through MiniCPM, which applies the following:
   (a) *Spatial text alignment*: by estimating homographies based on key points, perspective distortions inherent to documents captured at non-orthogonal angles are corrected. This transformation converts skewed regions into normalized frontal views, facilitating precise character recognition.
   (b) *Multimodal contextual recognition*: where MiniCPM combines visual embeddings (extracted via convolutional layers) with linguistic embeddings, enabling the resolution of ambiguities in deformed or partially occluded characters.
3. Rule-Based Syntactic Post-Processing: Extracted text undergoes structured validation, for which syntax- and semantics-specific constraints are applied based on the document type. For instance,
   - For identification numbers (ID cards), the system verifies the expected length (8–10 digits), the presence of control letters (in alphanumeric systems), and consistency with regional prefixes.
   - For proper names, recognized strings are cross-checked against standardized lexical databases to discard OCR-generated artifacts.

This methodological framework not only enhances the accuracy of digitizing heterogeneous documents but also reduces false positives through the integration of contextual knowledge. Empirical validation showed a 22.4% improvement in extraction accuracy compared to approaches that rely solely on conventional OCR models, particularly in

documents affected by compression artifacts or uneven lighting conditions. The synergy between geometric detection (YOLOv11) and semantic understanding (MiniCPM) establishes a replicable paradigm for the automated processing of sensitive data in forensic contexts.

#### 3.2.3. Flexible Storage and Advanced Query Module

The detected PII and its associated metadata are **indexed in *Elasticsearch***, a search and analytics engine built on *Apache Lucene* (https://lucene.apache.org), which is chosen here for its horizontal scalability and advanced full-text query capabilities (see Table 4 for other NoSQL databases to provide scalability, flexibility, and performance for modern applications, which make them essential for handling large-scale, real-time, and semi-structured data).

**Table 4.** Examples of open-source NoSQL databases.

| Name | Description | Link |
| --- | --- | --- |
| **Apache Cassandra** | A high-performance, distributed NoSQL database designed for handling massive amounts of structured and semi-structured data across multiple nodes. It provides high availability, fault tolerance, and linear scalability. | https://cassandra.apache.org/_/index.html (accessed on 14 October 2025) |
| **Apache HBase** | A column-family NoSQL database built on top of Hadoop, designed for processing large amounts of sparse data. It is well-suited for analytical workloads and time-series data storage. | https://hbase.apache.org/ (accessed on 14 October 2025) |
| **Elasticsearch** | A distributed search and analytics engine optimized for full-text search, log analysis, and real-time data indexing. It enables scalable querying across large datasets and integrates well with the ELK stack (Elasticsearch, Logstash, Kibana). | https://www.elastic.co (accessed on 14 October 2025) |
| **Neo4j** | A graph database that specializes in handling highly connected data. It is widely used for applications that require complex relationship modeling, such as recommendation systems and fraud detection. | https://neo4j.com (accessed on 14 October 2025) |
| **Redis** | An in-memory key-value store known for its speed and efficiency. It is commonly used for caching, real-time data processing, and message queuing in high-performance applications. | https://redis.io (accessed on 14 October 2025) |

The index structure is defined through custom mappings, which include the following:

- Text fields: For raw PII (e.g., email: "john.doe@example.com"), analyzed using specialized tokenizers (e.g., domain-based segmentation for email addresses).
- Geospatial fields: For documents linked to locations (extracted from file metadata or textual references).
- Nested fields: To store 100-character contextual windows around each PII, allowing for searches based on surrounding phrases.
- Embedding vectors: Generated using Sentence-BERT [41] for textual PII, allowing semantic searches (e.g., matching misspelled names through similarity analysis).

Complex queries, such as tracking an identification number across multiple leaks, are implemented using *Elasticsearch's Domain-Specific Language* (DSL). This combines time-range filters, aggregations by affected entities, and boolean logic clauses. To ensure high performance on terabyte-scale datasets, sharding techniques are employed—dividing the index into five primary shards—along with replication, using three replicas per shard. This setup is complemented by a hot-warm architecture to efficiently manage varying workloads.

This Elasticsearch integration not only enables millisecond-level ad hoc searches but also supports real-time monitoring of data exposures via *Kibana* dashboards, correlating PII findings with external threat intelligence feeds.

#### 3.2.4. Reporting Module

The *reporting module* of the RDBAlert tool constitutes the final layer of its analytical architecture, designed to **convert raw data into actionable insights through dynamic visualizations and adaptable narrative structures**. Implemented on the Kibana platform version 8.12 (https://www.elastic.co/kibana) (accessed on 14 October 2025), this module integrates directly with Elasticsearch, utilizing its native query and real-time aggregation capabilities to generate interactive visual representations, detailed forensic reports, and geospatial analyses.

Kibana serves as a unified visualization engine, enabling the creation of customizable dashboards that synthesize key metrics derived from indexed PII. Notable functionalities include the following:

- *Visualization of spatio-temporal correlations*: Heatmaps overlaid with geospatial layers, where PII instances are plotted based on coordinates extracted from metadata (e.g., physical addresses in documents).
- *Multivariable statistical graphs*: Temporal histograms showing data exposure trends, network diagrams linking compromised entities, and stacked bar charts breaking down PII types (e.g., email addresses vs. identification numbers).
- *Advanced aggregations*: Complex query processing, such as tracking the recurrence of a phone number across multiple leaks or calculating the percentage distribution of confidential documents by economic sector.

For automated report generation, the module employs configurable templates that combine static elements (such as institutional logos and methodological frameworks) with dynamic data extracted from Elasticsearch. Reports support multiple export formats (PDF, CSV, XLSX) and include dedicated sections for the following:

- Executive summaries, highlighting macro trends such as the total volume of exposed PII per region or affected entity.
- Detailed forensic analyses, incorporating screenshots of visualizations, contextual excerpts from original documents, and links to source files on the dark web.
- Mitigation recommendations, derived from identified patterns (e.g., correlations between PII types and recurrent attack vectors).

Additionally, the reporting module includes cohort analysis functions, enabling the segmentation of victims by exposure periods or data criticality levels. The experimentation section below provides sample cases demonstrating these capabilities.

In summary, this module not only democratizes access to technical insights through visual abstractions but also acts as a bridge between granular forensic analysis and strategic cybersecurity decision-making. It enhances organizational capabilities to anticipate and mitigate breaches resulting from ransomware attacks.

### *3.3. RDBAlert's Operation*

As described, *RDBAlert* is designed to monitor and analyze data leaks announced by ransomware groups, with the aim of identifying compromised personal information. Regarding the practical operation of the tool, there are two primary methods for accessing and analyzing the indexed data to determine whether specific information has been exposed in a data leak:

- First, as illustrated in Figure 3, a dedicated web service (available at https://ransomdbalert.com) enables users to check whether their email address or other data points appear in the leaks indexed by the system. This service offers a user-friendly interface for performing quick and easy checks.
- Alternatively, users can conduct a more detailed analysis locally by downloading the relevant leak data and following the setup instructions available in the RDBAlert GitHub repository [42]. Once an Elasticsearch instance is running and the leaked data has been indexed, users can perform detailed queries directly within Elasticsearch to analyze and retrieve relevant information.

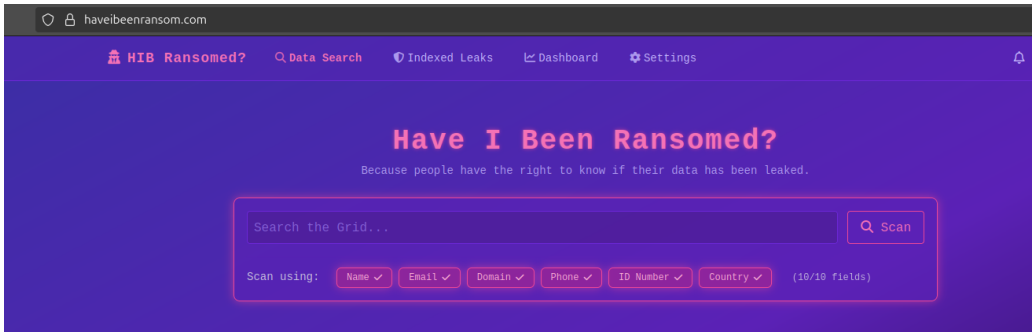



**Figure 3.** *'Have I Been Ransomed?'* service available at the website https://ransomdbalert.com.

#### 3.3.1. Ransomware Data Leak Monitoring

Regardless of the operation method used, the system relies on a specialized, standalone crawler module located at https://breach.house (accessed on 14 October 2025) and developed by the authors some time ago as a separate tool (see Figure 4), configured with a curated list of known ransomware group blog URLs primarily hosted on the *Tor* network (*.onion* resources). The crawler systematically monitors these sites for new posts or updates related to victim organizations. When a new post potentially announcing a data leak is detected, the crawler extracts key metadata, such as the victim's name, the posting date, and any provided links—often pointing to proof packs or placeholders for the full data release. This information is then cataloged and displayed through the web interface, enabling continuous tracking of ongoing incidents.

Ransomware operations often employ a double extortion strategy, in which data exfiltrated before encryption is used as leverage—threatening public release if the ransom is not paid within a specified timeframe, typically between two weeks and one month. As a result, obtaining the initial link from a leak site does not necessarily guarantee immediate access to the leaked data.

To manage this delay and automate data acquisition, a secondary component of the system monitors the status of detected leaks. It periodically checks whether the ransom payment deadline has passed and verifies if the data has been published by the ransomware group. Publication methods can vary, including direct downloads from file servers or distribution via *BitTorrent* links. Once the deadline has expired and the data is confirmed to be publicly accessible (e.g., through a direct download link or a torrent file -magnet-), the system proceeds with the automated retrieval and analysis of the leaked content. It

makes use of a robust download manager, such as `aria2c`, capable of handling various protocols including HTTP/HTTPS and BitTorrent, even over the Tor network.

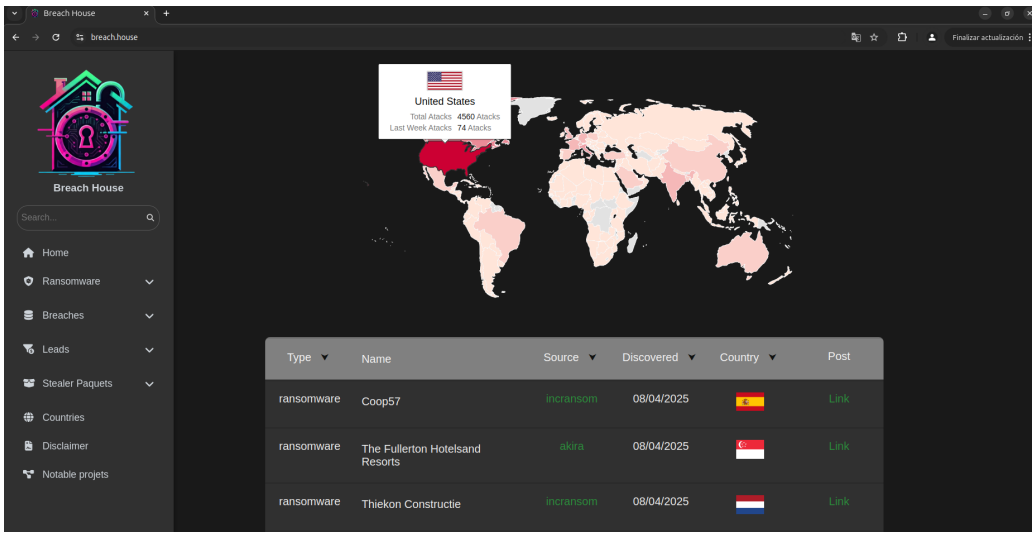



**Figure 4.** RDBAlert's crawler interface at https://breach.house (accessed on 14 October 2025).

Finally, the downloaded data files are then available for subsequent analysis stages.

### 3.3.2. Data Leak Analysis

Within the Elasticsearch index, specific fields enable targeted searches (see Figure 5):

- To search for email addresses, phone numbers, names, or other information directly associated with an email, the `emails` field is deployed.
- For broader contextual information linked to an email, the field `email_context` exists.
- To search for data related to specific domains, the `domain` field is specified.

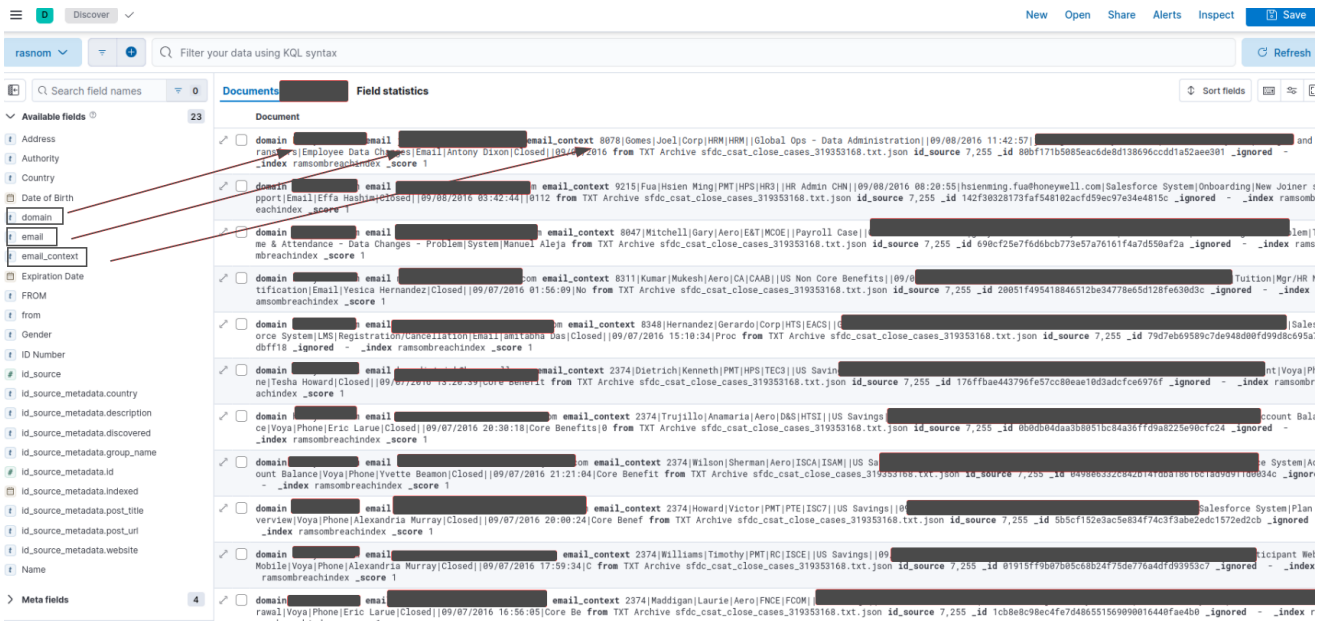



**Figure 5.** Examples of PII findings.

For advanced searches—particularly those involving leaked images of identity documents such as ID cards or passports— users can query specific metadata fields extracted during the indexing process. As shown in Figure 6, these fields include `ID number`, `country`,

`Address`, `Authority`, `date of birth`, `expiration date`, `Gender`, and `name` (Figure 7). Additionally, it is important to highlight that MiniCPM can extract text content regardless of the language, enabling a comprehensive analysis of the identified documents.

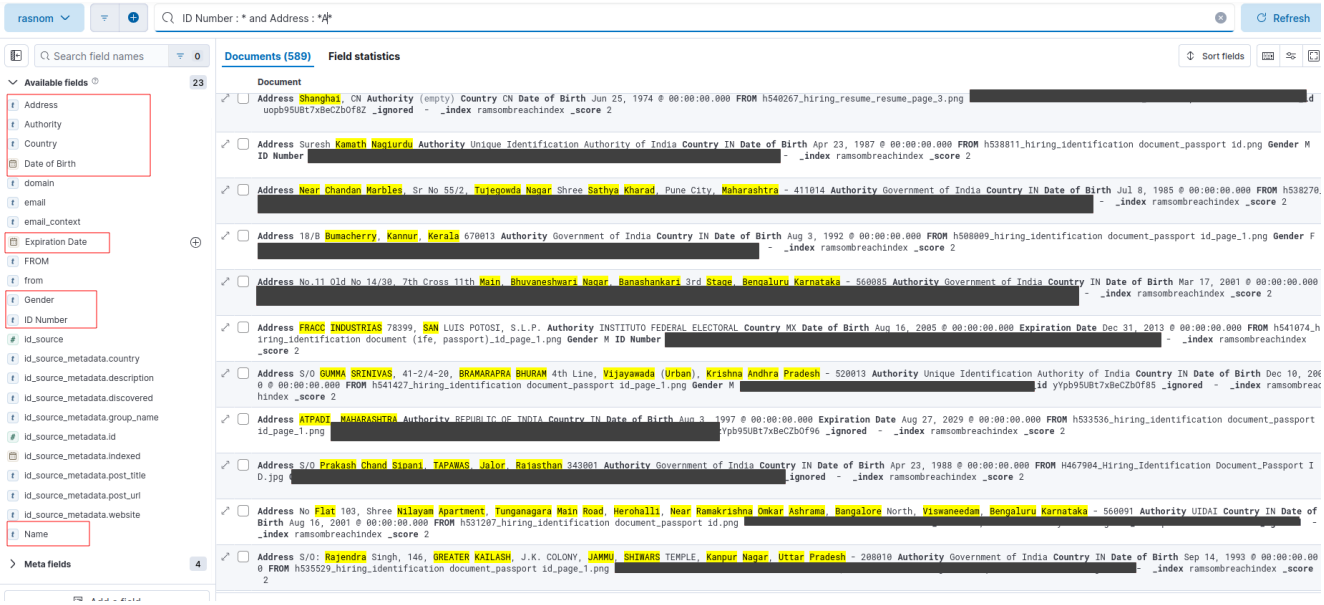



**Figure 6.** ID/passport-related PII findings.

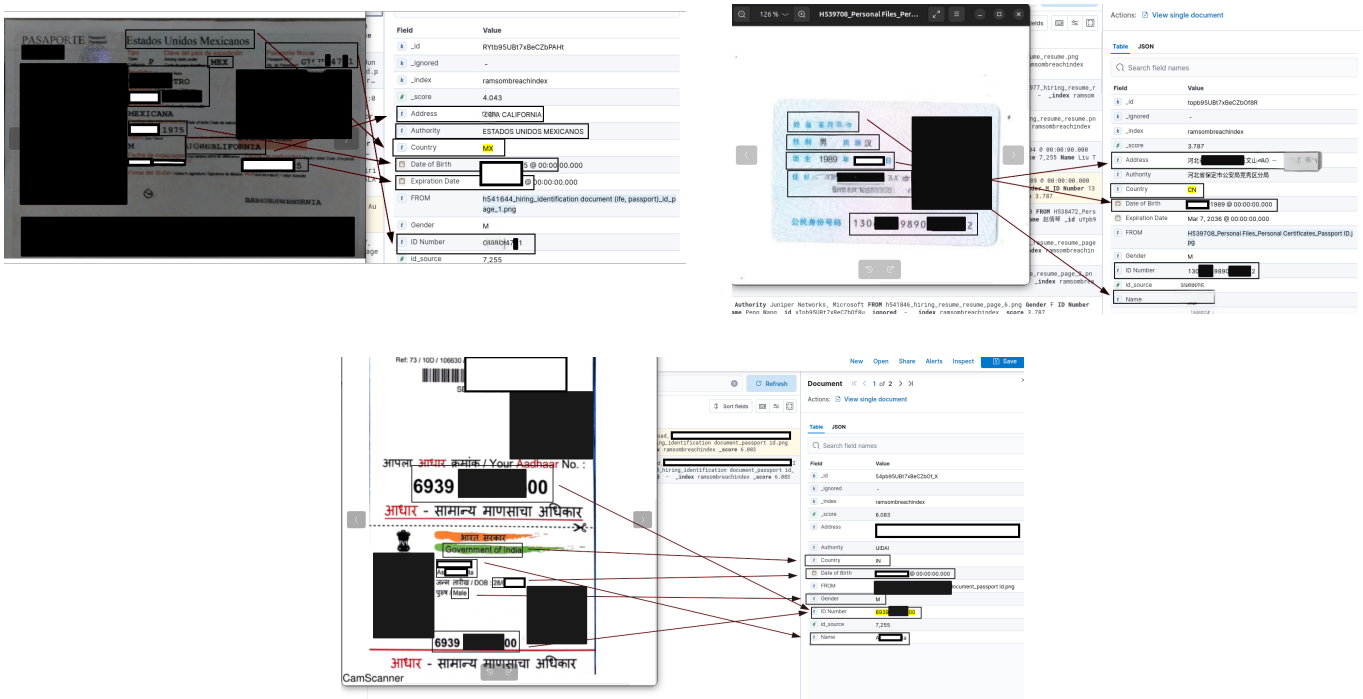



**Figure 7.** Examples of ID/passport PII findings in the *PhilHealth* dataset—see Section 4—(note the successful identification results achieved despite the poor quality of some analyzed images).

Furthermore, metadata related to the source of the ransomware leak (e.g., the responsible group or the announcement date) can be accessed through the `id_source_metadata` field, as illustrated in Figure 8. This enables users to contextualize their findings within the specific breach event.

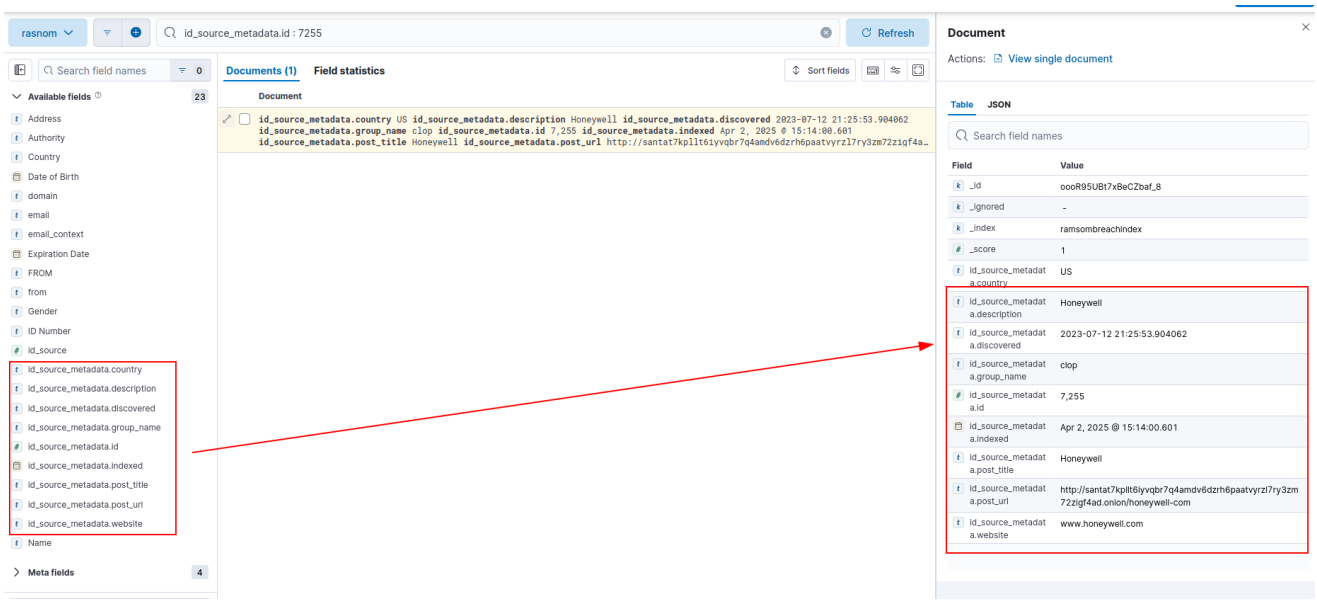



**Figure 8.** Metadata related to the ransomware leak.

## 4. Experimental Results

At this point, to assess the RDBAlert's PII extraction and analysis capabilities, we selected and downloaded five datasets of exfiltrated data corresponding to the following targeted organizations:

- *Honeywell*—In May 2023, this Fortune 100 company—specializing in aerospace and energy equipment—experienced a data exfiltration incident involving 233.45 GB of data, including 22.2 GB of personally identifiable information (PII) [43].
- *Pension Benefit Information (pbInfo)*—In May 2024, the *Cl0p* ransomware group targeted this US provider specializing in population data management solutions, exfiltrating 33.21 GB of confidential information, including data on pension and insurance beneficiaries. [44].
- *Pioneers Electronics*—In July 2023, the same group also exfiltrated 114.14 GB of data from the company Pioneer Electronics [45].
- *Philippine Health Insurance (PhilHealth)*—In September 2023, the Philippine Health Insurance Corporation released a comprehensive report on a data breach incident that compromised the personal information of 42 million individuals [46]. This incident stands as one of the most significant cases of mass data exfiltration to date, with an unprecedented volume of personal records compromised.
- *Targa Viasat Spain*—In 2024, the *Medusa* ransomware group targeted Targa Viasat Spain, a company specialized in satellite communications and vehicle tracking solutions, exfiltrating 87.56 GB of sensitive data [47].

### *4.1. Initial Training Stage*

Before analyzing the aforementioned datasets to obtain results related to exfiltration, it is important to highlight the need to properly train the YOLO tool for accurate detection. To achieve this, two specialized datasets were used, carefully selected for their representativeness and diversity across a wide range of scenarios:

1. Face Detection: For this task, the WIDER FACE dataset, available at http://shuoyang1213.me/WIDERFACE (accessed on 14 October 2014) was utilized. This dataset is a benchmark in the scientific community, comprising 32,203 images with 393,703 manually annotated faces. The dataset covers the following:

- Variable scales (from close-up portraits to dense crowds).
- Extreme lighting conditions (overexposure and shadows).
- Partial occlusions (accessories and hair).

Each annotation includes precisely adjusted bounding box coordinates, along with difficulty labels (easy, medium, hard), allowing the model to distinguish between straightforward and challenging cases during training.

2. Personal Identification Document Detection: The dataset used to train the detection of personal identification documents in RDBAlert was developed using an innovative approach based on real-world data extracted from historical ransomware leaks, ensuring its relevance to real-world scenarios. This corpus includes images of identity documents (such as ID cards, passports, and residence permits) sourced directly from previously processed leaks. To ensure diversity, the dataset covers variations in the following:
   - Capture quality: Ranging from high-resolution scans to photographs taken with mobile devices under suboptimal conditions (e.g., blurriness and reflections).
   - Regional formats: Documents issued in different countries, incorporating variations in design, color schemes, and security features (e.g., holograms and microtext).
   - Exposure contexts: Documents that are partially obscured, folded, or overlaid with other objects in the image.

A critical aspect of the training process was the use of images previously detected by RDBAlert's facial recognition module. These images, which contained faces associated with documents in leak contexts, formed the foundation for dataset generation.

The implemented workflow adopts a sequential hybrid approach, wherein YOLOv11 and MiniCPM operate synergistically:

1. Hierarchical Detection Phase: YOLOv11 executes a two-tiered detection strategy: initially, it identifies the entire document within the image using global bounding boxes; subsequently, it segments critical subregions (such as the name area, identification number, and photograph), employing normalized relative coordinates. This process is supported by a spatial reference system that preserves the document's geometric proportions, irrespective of its orientation or scale.
2. Advanced Textual Extraction Phase: Each detected subregion is processed through MiniCPM, which applies the following:
   - *Spatial Text Alignment*: By estimating homographies based on key points, it corrects the perspective distortion inherent in documents captured at non-orthogonal angles. This transformation converts skewed regions into normalized frontal views, facilitating accurate character recognition.
   - *Multimodal Contextual Recognition*: MiniCPM integrates visual embeddings (extracted via convolutional layers) with linguistic embeddings, enabling the resolution of ambiguities in deformed or partially occluded characters.
3. Rule-Based Syntactic Postprocessing: The extracted texts undergo structured validation, wherein specific syntactic and semantic constraints are applied based on the document type. For instance,
   - For identification numbers (e.g., IDs), the system verifies the expected length (8–10 digits), the presence of control letters (in alphanumeric systems), and consistency with regional prefixes.
   - For proper names, recognized strings are cross-referenced with normalized lexical databases to eliminate OCR artifacts.

To evaluate the capability of the YOLOv11 model in the initial detection of documents potentially containing PII, a dataset comprising 8000 photographs was processed. This dataset specifically included 500 images of curriculum vitaes (CVs) featuring visible faces, 500 identification cards (ID cards), and 500 passports, totaling 1500 documents confirmed to contain PII.

Employing a YOLOv11 model trained on the `WIDERFACE` dataset and applying an inference confidence threshold of 0.60, a total of 2736 items were detected. Notably, these detections included 100% of the documents known to contain PII (the 1500 CVs with faces, ID cards, and passports). The consistency of this outcome was verified by repeating the experiment five times, yielding identical results on each occasion.

Subsequently, the inference confidence threshold was increased to 0.65. With this more restrictive threshold, the total number of detections decreased to 1283. As in the previous case, the experiment was replicated five times with consistent results. While no false negatives were observed among the detected items at this threshold, it was noted that the model failed to detect some documents that did, in fact, contain PII.

Considering that the detections from YOLOv11 are intended for a subsequent filtering and analysis phase using MiniCPM, it is concluded that operating with the 0.60 confidence threshold is preferable. Although this configuration generates a higher number of detections (potential false positives—detected items that do not contain relevant PII), it ensures the capture of all documents containing PII. This approach minimizes the risk of omitting sensitive documents (false negatives relative to the entire set of PII), which is crucial for the subsequent processing stage.

In a similar line, to evaluate the effectiveness of MiniCPM in detecting personal identification documents, a total of 650 digitized PDF files related to the *Honeywell* data breach were analyzed. Among these, only 35 documents lacked an associated identification number (IDNumber). A manual review determined that in 32 of these cases, the absence of an IDNumber was justified, as the documents were primarily Chinese resumes, resignation letters, or recommendation letters—types that typically do not include personal identifiers. However, in the remaining three cases, which involved Indian documents, the identification numbers were clearly present but had been placed within the address field instead of the designated IDNumber field. This raises a point of discussion as to whether these instances should be classified as model errors or formatting inconsistencies.

To enhance the model's training, *Aadhaar* [48] (Indian identity documents, known as *Aadhaar*, constitute the world's largest biometric identification system. As of February 2020, over 90% of India's population had been issued an Aadhaar number, amounting to more than 1.26 billion individuals) documents from *Honeywell* were utilized, explicitly indicating to the model which elements represented the IDNumber. This approach aimed to ensure accurate identification of IDNumbers, preventing their misclassification into other fields such as the address.

Although the core of *RDBAlert* does not lie in the specific analysis tools used—such as YOLO and MiniCPM in this case—it is crucial to train these models appropriately to ensure strong performance. Based on the analyzed data, MiniCPM achieved an accuracy rate of approximately 99.54% in detecting personal identification numbers. This high level of precision indicates that MiniCPM is an effective tool for the automated identification of personal data in digitized documents.

The quantitative evaluation of the integrated YOLOv11 and MiniCPM pipeline substantiates its efficacy and underlying design strategy. The object detection module, YOLOv11, was evaluated on a substantial dataset of 87,982 samples. With a deliberately lowered confidence threshold of 0.40, the model achieved an exceptionally high recall of 0.994, resulting in only 4 false negatives, albeit at the cost of a higher number of false

positives (195), which yielded a precision of 0.782 and an F1-score of 0.875. This strategic trade-off prioritizes the comprehensive capture of all potential documents of interest (minimizing false negatives) over precision, operating on the premise that the subsequent MiniCPM module is highly effective at filtering false positives during textual analysis. This premise is validated by MiniCPM's performance, which on a subset of 886 samples demonstrated near-perfect precision (0.998) and recall (0.990), culminating in an F1-score of 0.994. The cascaded architecture thus leverages the strengths of each model: YOLOv11's high-recall detection ensures minimal data loss, while MiniCPM's high-precision OCR rectifies initial false detections. Consequently, the end-to-end system achieves a robust sample-weighted F1-score of 0.876, empirically validating the chosen design paradigm for a forensic tool where missing critical data (false negatives) is a more significant failure than a preliminary false alarm.

### *4.2. Exfiltrated Data's Analysis Results*

The forensic examination of the exfiltrated datasets across the five case studies was extended beyond a purely quantitative statistical description to incorporate tactical intelligence on adversary behavior. This section presents a multi-faceted analysis that transitions from a statistical baseline to an investigation of the operational patterns characterizing specific ransomware groups. First, a statistical overview establishes the scale and composition of the exfiltrated data. Subsequently, we perform a correlation analysis to identify distinct exfiltration TTPs associated with different threat actors. Finally, we extrapolate these findings to reveal geographical targeting preferences, thereby positioning the case-specific evidence within the broader context of the global ransomware ecosystem and its strategic trends.

#### 4.2.1. Statistical Overview of Exfiltrated Data Across Case Studies

Once the system is properly tuned, some general figures obtained by RDBAlert from the analysis of the aforementioned five exfiltrated datasets are shown in Table 5:

- In the *Honeywell* case, RDBAlert identified 1714 PDF files containing PII (payrolls, contracts, tax records) and 2542 databases with 165,343 internal emails and 24,893 external emails. The data includes employee numbers, demographic information (name, birth date, federal taxpayer registry), salary details (daily and monthly), and corporate information (email and company).
- Regarding *pbInfo*, we identified 3.9 GB of data related to PII, distributed across 783 databases, 702 PDF files, and 72 text documents. The automated analysis revealed the exposure of 79,554 full names, 8674 internal emails, 72,695 external emails, and 354 identification documents (ID cards/passports).
- In the *Pioneers* case, the tool accurately correlated the 1361 PDF files analyzed with the 1352 JSON files generated. Notably, 93% of the files contained exportable PII, highlighting the prevalence of clerical documents as a critical attack vector for data exfiltration.
- For the *Philippine Health Insurance* dataset, 3.35 TB of data was identified, including 58.9 GB containing PII. The automated analysis revealed 5098 text files, 7082 PDF documents, 3476 Word files, 8059 image files, and 9523 databases containing PII. The data exposure involved 51,083,051 full names, 4201 internal emails, 304,104 external emails, and 3729 identification documents (ID cards/passports). All of this highlights the critical vulnerability of healthcare systems, which continue to be prime targets for large-scale, coordinated cyberattacks.
- Finally, in the *Viasat* case, we identified 11.3 GB of PII distributed across 3173 databases, 1136 PDF files, and 98 text documents. The automated analysis revealed the exposure of 138,749 full names, 38,715 internal emails, and 327 identification documents (ID

cards/passports). Additionally, the data included sensitive corporate information, such as client contracts, vehicle tracking logs, and financial records, underscoring the critical nature of operational data in such attacks.

**Table 5.** Results obtained with RDBAlert from the exfiltration cases analyzed.

| Data Exfiltration | | Data Size | | #Total Files Found and Analyzed by Type with PII | | | | | #Total PII Found and Analyzed by Data Type | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Company** | **Country** | **Leaked Data** | **PII** | **.txt** | **PDF** | **.doc** | **Images** | **Data-Bases** | **#Names** | **#Internal Emails** | **#External Emails** | **#ID Cards/ Passports** |
| *Honeywell* | USA | 233.45 GB | 22.2 GB | 159 | 1714 | 13 | 145 | 2542 | 115,961 | 165,343 | 24,893 | 653 |
| *pbInfo* | USA | 33.21 GB | 3.9 GB | 72 | 702 | 36 | 23 | 783 | 79,554 | 8674 | 72,695 | 354 |
| *Pioneer Electronics* | Japan | 114.14 GB | 14.6 GB | 233 | 1361 | 84 | 106 | 1352 | 195,443 | 104,875 | 95,471 | 582 |
| *PhilHealth* | Philippines | 3.35 TB | 58.9 GB | 5098 | 7082 | 3476 | 8059 | 9523 | 51,083,051 | 4201 | 304,104 | 3729 |
| *Viasat* | Spain | 97.66 GB | 11.3 GB | 98 | 1136 | 141 | 63 | 3173 | 138,749 | 38,715 | 88,451 | 327 |

The following are some additional relevant aggregate figures obtained by RDBAlert.

With respect to the analysis of email domains from **Honeywell**'s internal communications, the tool processed a total of 1,220,345 emails. Internal company domains accounted for 82.44%, while external domains made up 17.56% of all detected email addresses (see Figure 9). Among the external domains, *gmail.com* was the most prevalent, representing 20.82%, followed by *kcnsc.doe.gov*—a domain managed by Honeywell– which accounted for 7.95% of external emails.

Regarding the unique identifiers detected (e.g., Social Security numbers, local ID numbers, driver's licenses, and passports), RDBAlert identified a total of 1239 unique identity documents. Of these, 63.68% were associated with individuals from India, 11.79% from China, 9.12% from Mexico, 5.35% from the United States, 2.04% from Korea, and the remaining 8.02% from other countries (see Figure 10).

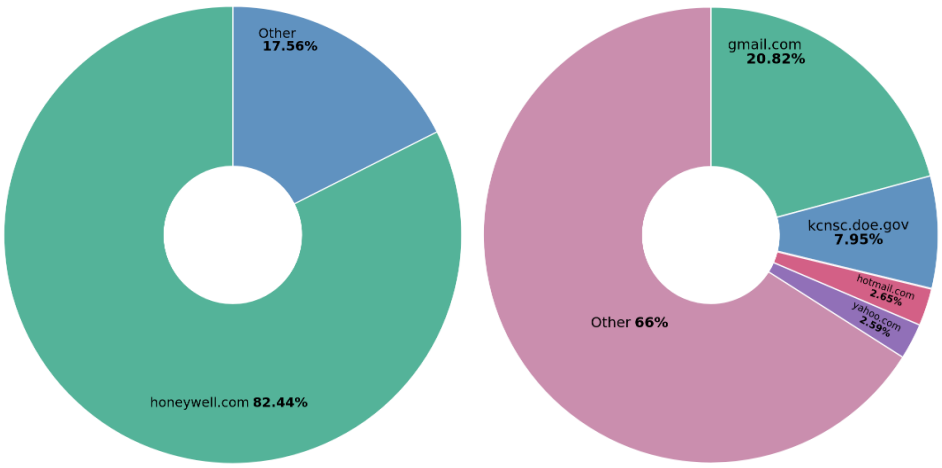



**Figure 9.** *Honeywell* internal emails vs. external emails.

For **pbInfo** (see Figure 11), the tool analyzed the distribution of email domains, revealing that internal company domains accounted for 93.46%, while external domains made up 6.54%. Among the external domains, *gmail.com* was the most common, representing 5.38%, followed by *mercer.com* at 3.46%. Other notable external domains included *yahoo.com* (1.05%), *aol.com* (0.59%), and *csc.com* (0.51%).

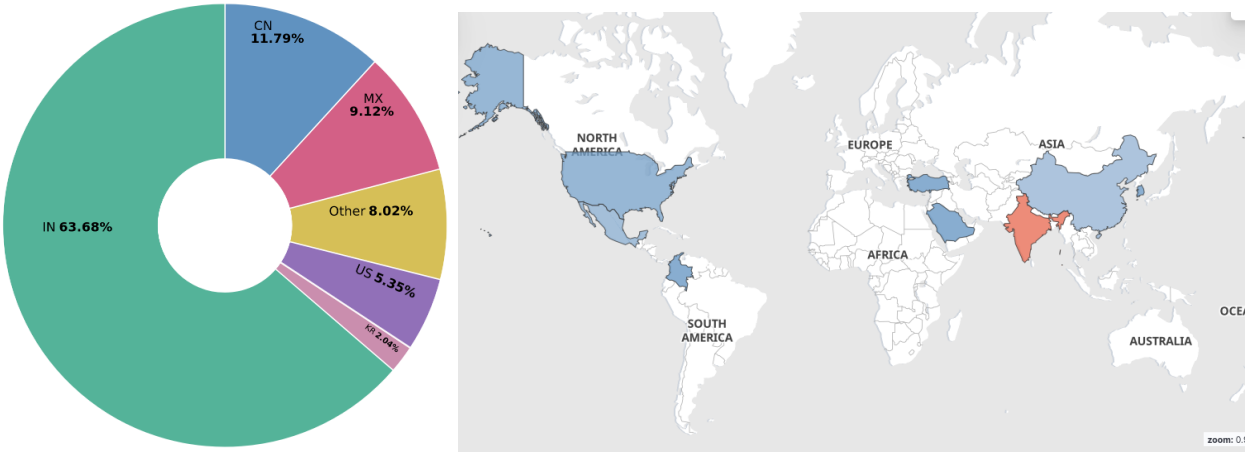
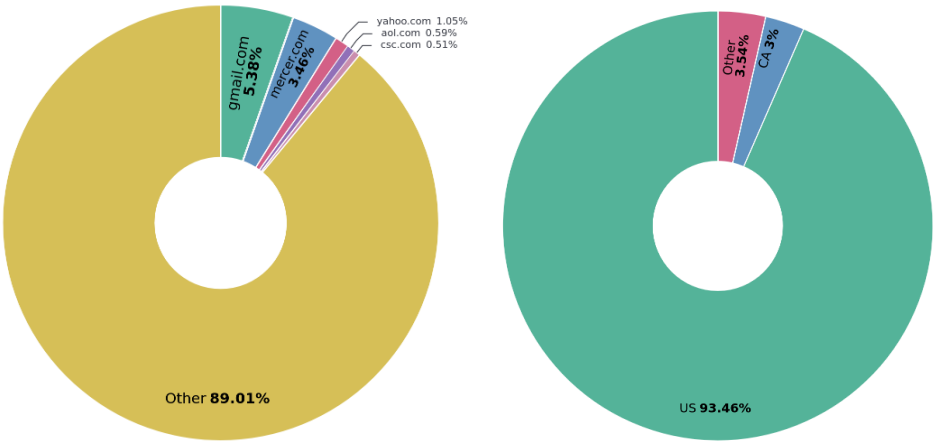



**Figure 10.** *Honeywell* ID number information.




(**a**) Email related analysis. (**b**) PII by country.

**Figure 11.** Analysis performed for *pbInfo* dataset.

Regarding the unique identity documents identified, the analysis revealed a significant dominance by the United States, accounting for 93.46% of the total. Canada followed with 3%, and the remaining 3.54% were from other countries.

In the case of **Pioneer Electronics**, email domain analysis (Figure 12) revealed *shipamerican.com* as the most prominent domain, accounting for 10.02% of detected emails. Other significant external domains included *pioneer-usa.com* (9.23%), *post.pioneer.co.jp* (3.88%), and *mail.com* (2.35%). Remaining domains accounted for 73.01%.

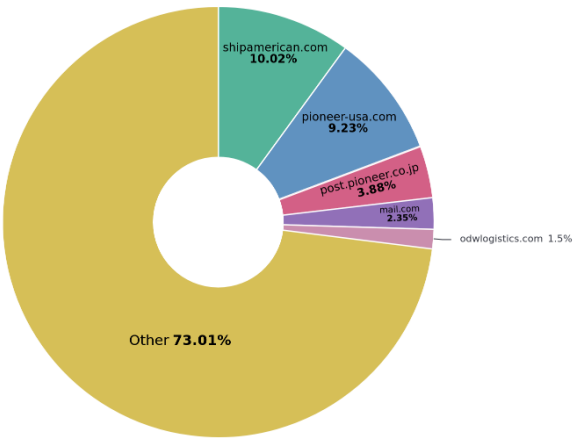



**Figure 12.** *Pioneer* internal emails vs. external emails.

The analysis of identity documents and passports showed an exclusive identification of the United States, representing 100% of the total detected documents (Figure 13).

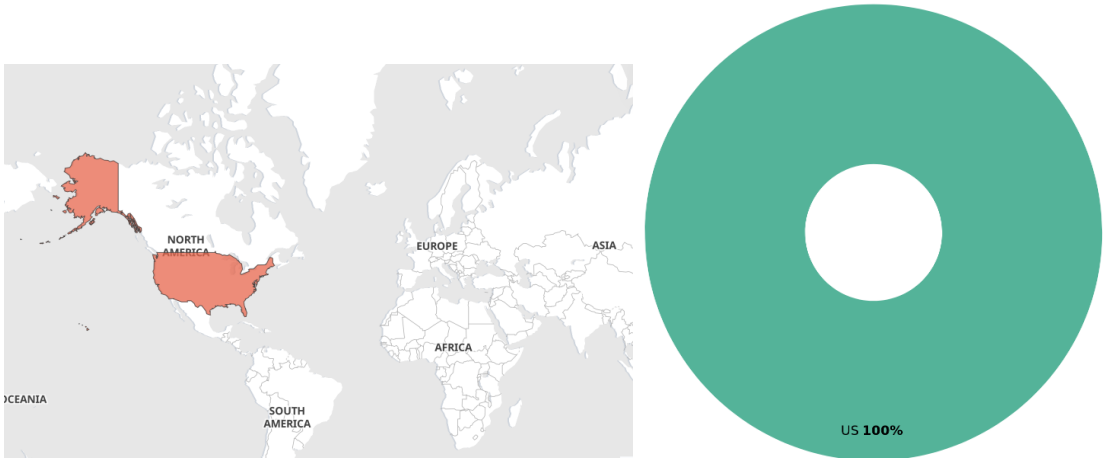



**Figure 13.** *Pioneer* ID number information.

Regarding the analysis of email domains in the **PhilHealth** case (see Figure 14), the majority of emails were associated with *yahoo.com*, accounting for 48.92% and suggesting a strong dependence on consumer-grade email services. *Gmail.com* followed at 15.64%, while the official government domain *philhealth.gov.ph* represented only 1.18%, reflecting limited adoption of institutional accounts. Other identified domains included *yahoo.com.ph* with 2.49%, and the remaining 30.22% of emails corresponded to a broad set of external personal or organizational providers, further highlighting the use of non-corporate communication channels in sensitive operations.

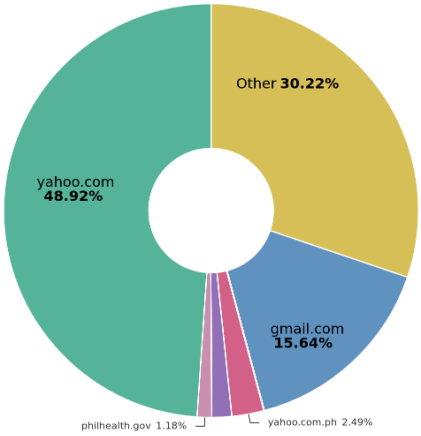



**Figure 14.** *PhilHealth* internal emails vs. external emails.

With respect to identity documents (see Figure 15), our analysis revealed that an overwhelming 92.83% of the exposed identification records—such as ID cards and passports—originated from the Philippines, emphasizing the predominantly local impact of the breach. The United States accounted for 5.54%, while Vietnam, Austria, and Australia contributed smaller portions of 0.41%, 0.08%, and 0.08%, respectively. The remaining 1.06% of documents were linked to individuals from other countries, suggesting a degree of international exposure, possibly related to overseas employment or cases of dual citizenship.

Finally, the email domain analysis performed for **Viasat** (see Figure 16) identified *grupodetector.com* as the dominant internal domain, comprising 40.58% of emails, reflecting the acquisition of *Grupo Detector* by Viasat. The remaining 59.42% of the emails corresponded to other internal domains.

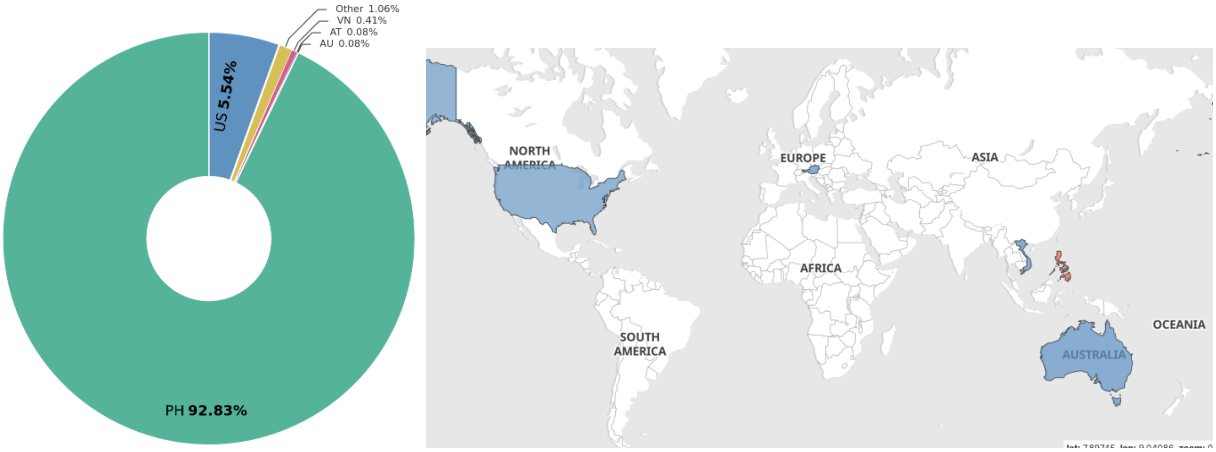



**Figure 15.** *PhilHealth* ID number information.

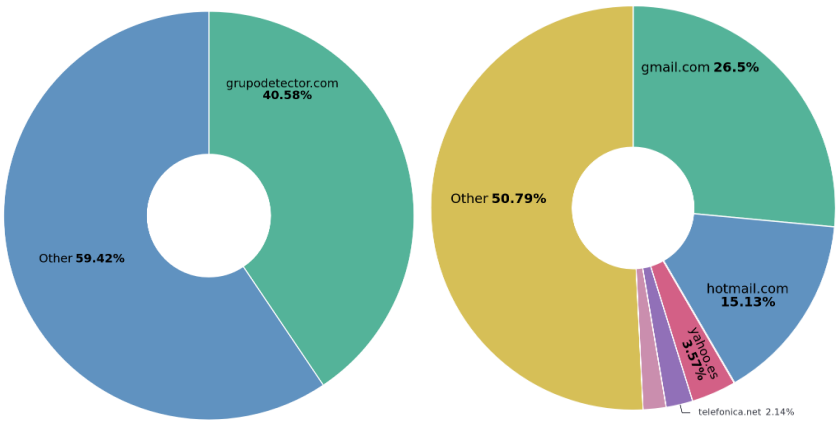



**Figure 16.** *Viasat* internal emails vs. external emails.

Regarding external email domains, *gmail.com* was predominant, accounting for 26.5%, followed by *hotmail.com* with 15.13%. Additional notable external domains were *yahoo.es* at 3.57% and *telefonica.net* at 2.14%. Other domains accounted for the remaining 50.79%.

With respect to the analysis of unique identity documents (see Figure 17), a significant majority are from Spain, accounting for 74.42% of the total. France followed with 11.63%, Italy represented 6.98%, Romania accounted for 4.65%, and other countries covered the remaining 2.33%.

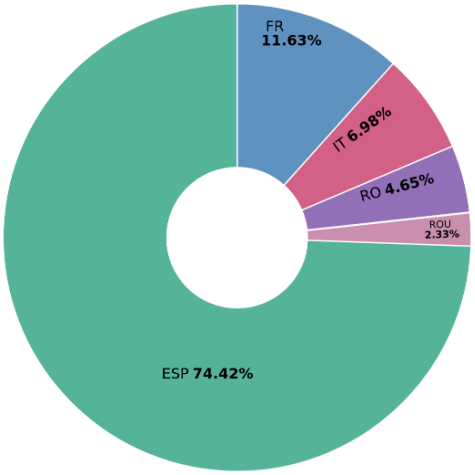



**Figure 17.** *Viasat* ID number information.

4.2.2. Correlation of Exfiltration Patterns with Specific Ransomware Groups

The detailed analysis of the exfiltrated data across the five case studies reveals distinct patterns that allow for correlation with the known Tactics, Techniques, and Procedures (TTPs) of specific ransomware groups. This correlation is based on the volume and type of data prioritized, the nature of the target entities, and the geographical composition of the exfiltrated personally identifiable information (PII):

- *Honeywell*: Correlation with groups specializing in industrial espionage and double extortion (e.g., BlackCat/ALPHV). The exfiltration of 233.45 GB, including 2542 databases (see Figure 9) containing internal emails (82.44% of the total) and sensitive corporate documents (payrolls, contracts, tax information), points to a sophisticated group with an interest in trade secrets and corporate intelligence. Groups like *BlackCat (ALPHV)* are known for conducting thorough analyses of stolen data prior to encryption to enable more effective extortion. The geographical diversity of the identity documents (with significant percentages from India, China, and Mexico) suggests that the target was a multinational corporation with a global workforce, a common target for such actors (see Figure 10).
- *pbInfo and Viasat*: Correlation with mid-scale Ransomware-as-a-Service (RaaS) groups (e.g., Phobos and Snatch). These cases present smaller but highly specific data volumes (3.9 GB and 11.3 GB of PII, respectively). The attack on *pbInfo* (see Figure 11) shows a strong focus on the U.S. (93.46% of IDs) and a high proportion of internal communications (89.01% of emails). Conversely, the attack on Spain-based *Viasat* (see Figure 17) reveals a clear regional target (74.42% of Spanish IDs) that includes critical operational information (vehicle tracking logs and financial records). These patterns are consistent with RaaS groups that enable their affiliates to select lower-profile targets with valuable data, using standardized tools for exfiltration and subsequent blackmail.
- *Pioneer Electronics*: Correlation with pure extortion groups (e.g., Ragnar Locker and Babuk). The profile of this attack, with a substantial data volume (114.14 GB) and an almost exclusive focus on employees and operations within the United States (100% of identity documents), correlates with groups seeking direct financial impact and streamlined extortion (see Figure 13). The nature of the data (93% of files containing exportable PII, clerical documents) indicates a broad-spectrum exfiltration without a highly specialized search. Groups like *Ragnar Locker* have demonstrated a similar pattern, attacking large corporations to exfiltrate data and then threaten its publication to cause reputational and financial harm.
- *PhilHealth*: Correlation with high-impact ransomware groups (e.g., Clop and LockBit). The massive volume of exfiltrated data (3.35 TB, with 58.9 GB of PII), affecting over 51 million individuals, is characteristic of large-scale attacks against the healthcare sector, a primary target for groups like *Clop* and *LockBit*. The data composition—with an overwhelming majority of Philippine identification documents (92.83%) and the prevalence of external (see Figure 15) and personal email domains (e.g., *yahoo.com* at 48.92% (see Figure 14)—indicates the exfiltration of patient and employee records on a national scale. This pattern aligns with these groups' strategy of maximizing pressure and extortion payments by threatening to expose highly sensitive health data of a vast population.

In summary, as synthesized in Table 6, the quantitative evidence obtained through *RDBAlert* not only characterizes the data breach but also allows for inferences regarding the modus operandi and potential perpetrators. This correlation between exfiltration metrics and known ransomware TTPs provides crucial strategic value for cyber threat intelligence, enabling a more targeted and effective defensive posture and response.

**Table 6.** Correlation between data exfiltration patterns and ransomware groups [20,49].

| Case Study | Key Patterns and Observed TTPs | Correlated Ransomware Group(s) |
|---|---|---|
| *PhilHealth* | Massive data volume; healthcare sector target; predominantly local PII; extensive use of personal email. | **Clop/LockBit** (specialized in high-impact, large-volume targets). |
| *Honeywell* | Sensitive corporate data (internal emails and finances); multinational target; geographically diverse PII. | **BlackCat/ALPHV** (focus on industrial espionage and double extortion). |
| *Pioneer Electronics* | Highly concentrated PII (U.S.); broad-spectrum exfiltration without a specific technical focus. | **Ragnar Locker/Babuk** (direct financial extortion and attacks on large corporations). |
| *pbInfo/Viasat* | Moderate data volume; specific regional or sectoral target; operational data and localized PII. | **Phobos/Snatch** (mid-scale RaaS for more focused attacks). |

#### 4.2.3. Geographical Targeting Trends of Major Ransomware Groups

To contextualize our case studies within the broader threat landscape, we monitored data exfiltration attacks perpetrated between February 2023 and September 2025, and we analyzed the geographical preferences of prominent ransomware groups. As shown in Table 7, there is a clear and overwhelming focus on targeting organizations in the United States. Groups like *Play* and *Medusa* exhibit strong geographical segmentation, almost entirely avoiding countries like France and Japan, while *Akira* shows emerging activity in Brazil. The targeting of *pbInfo* is highly consistent with the *Play* group's strategy. Conversely, the European focus of the *Viasat* breach and the Philippine-centric nature of the *PhilHealth* incident fall outside the primary targeting patterns of these major ransomware groups, suggesting that alternative threat actors, such as *APTs* or local groups, were responsible. This analysis demonstrates that the exfiltrated data itself can reveal implicit attributes of the attacking entity.

**Table 7.** Number of victim organizations per country attributed to major ransomware groups, based on *RDBAlert* data collected between 11 February 2023 and 27 September 2025.

| Group\ Country | USA | France | Japan | Brazil | Canada | United Kingdom | Germany | Other |
|---|---|---|---|---|---|---|---|---|
| *Clop* | 5 | 1 | 1 | 0 | 0 | 0 | 0 | 2 |
| *Akira* | 33 | 0 | 0 | 4 | 3 | 2 | 2 | 18 |
| *Play* | 467 | 0 | 0 | 0 | 50 | 6 | 12 | 35 |
| *Medusa* | 110 | 0 | 0 | 0 | 17 | 11 | 2 | 51 |

### *4.3. Computational Results*

In addition to the analysis previously conducted on data leaks, we also performed experiments to estimate the associated costs in terms of processing time and CPU/RAM utilization. These experiments were carried out on a machine with the following specifications: Ubuntu 24.04.2 LTS x86_64 operating system, kernel version 6.14.0-24-generic, AMD Ryzen 9 7900X (24 cores) @ 5.867 GHz processor, NVIDIA GeForce RTX 3090 GPU with 24 GB VRAM, 64 GB 5600 MT/s RAM, 4 TB M.2 SSD, and a 40 TB SATA RAID HDD storage system.

As shown in Table 8, the resource consumption results obtained, together with the inspected and filtered data previously discussed in Section 4, support the validation of *RDBAlert* as a practical and efficient security tool.

**Table 8.** Estimation of resource consumption by *RDBAlert*.

| Data Lake | Process | Resources Involved |
|---|---|---|
| *Honeywell* (233.45 GB) | Complete pipeline | - Time: 4 h 12 m<br>- CPU: average: 30%, range: 12–54%<br>- RAM: average: 15 GB, peak: 29 GB |
| | YOLO | - Time: 754,098 images/21 m (→598.49 img/s)<br>- 12,915 images filtered<br>- GPU: 80%<br>- VRAM: 1 GB |
| | MiniCPM | - Time: 12,915 images/1 h 12 m (→2.99 img/s)<br>- 653 images filtered<br>- GPU: 98%<br>- VRAM: 21 GB |
| *PhilHealth* (3.35 TB) | Complete pipeline | - Time: 7 h 39 m<br>- CPU: average: 40%, range: 12–63%<br>- RAM: average: 20 GB, peak: 29.5 GB |
| | YOLO | - Time: 2,084,578 images/1 h 2 m (→554.32 img/s)<br>- 35,478 images filtered<br>- GPU: 80%<br>- VRAM: 1 GB |
| | MiniCPM | - Time: 35,478 images/2 h 49 mm (→3.49 img/s)<br>- 3729 images filtered<br>- GPU: 98%<br>- VRAM: 21 GB |

### *4.4. Discussion and Ethical Considerations*

As with many current tools and technologies, we are aware of *RDBAlert*'s potential dual use. Hence, despite the large technical power of our tool, it is also necessary to discuss the associated large ethical risk due to the access to sensitive personal information. The core question here is who benefits and who is harmed when that power is exercised, and under what controls?

In what follows, some key considerations are discussed from a theoretical point of view:

- Privacy concerns. Data may be "public" but people rarely expect their aggregated records to be indexed, correlated, and queried at scale. Hence, availability does not mean consent.
  On the other hand, aggregation and linking increase identifiability; that is, many datasets that are harmless in isolation can become sensitive when combined.
- Harm and discrimination. Exposed sensitive attributes like health, ethnicity, or criminal history can enable discrimination, doxxing, extortion, or reputational and financial harm.
  Moreover, automated profiling can produce biased or incorrect inferences leading to unfair decisions (credit, hiring, insurance).
- Misuse and dual-use. Legitimate uses (research and incident response) sit beside illegitimate ones (stalking, fraud, targeted harassment). Easy access lowers the barrier for abuse.

- Transparency and accountability. Subject individuals typically have no visibility into who queried their data or why. Lack of auditability erodes trust.
- Legal/compliance exposure. Jurisdictions differ (GDPR, CCPA, ePrivacy, sector rules). Indexing or republishing sensitive data can create legal liabilities even if data was public.
- Security risks. The tool itself becomes a high-value target. A breach of the tool or its logs multiplies harm by exposing query histories and aggregated views.

Some technical controls to mitigate the aforementioned aspects are as follows: enforcement of least privilege and strong access policies, implementation of query and rate limits, provision of only aggregated data, evaluation of risk assessments before allowing data outputs, encryption of data, prevention of raw PII exports, and record of data provenance and timestamps. Ultimately, the ethical deployment of the technology discussed here depends less on its technical aspects and more on robust frameworks of governance, accountability, and the adoption of a cautious, privacy-first operational approach—all of which should be carefully evaluated before any real-world deployment.

## 5. Conclusions and Future Work

Despite the deployment of various early detection and response mechanisms, ransomware remains a major threat today, leading to the exfiltration of personal data all over the world. Due to the multi-extortion tactics commonly used by attackers, data are often exposed on public repositories or in data leaks. Current regulations require organizations to promptly notify breaches in order to alert affected victims and implement potential remedies to reduce impact.

This work introduces *RDBAlert*, a novel and comprehensive tool designed to automate the detection of ransomware data leaks and facilitate subsequent multimodal and language-independent analysis to extract personally identifiable information (PII). The tool's modular design leverages a combination of AI-powered components, incorporating existing solutions such as OCR-based character recognition, image analysis, and LLM-based text analysis to confirm the presence of personal data in ransomware data leaks. This core functionality is complemented by a web-based interface that enables users to conveniently input their personal data for searches within DLSs, while intuitively visualizing statistics and results through clear and interactive graphical representations.

The experimentation conducted demonstrates the effectiveness of the developed tool in quickly and reliably identifying personal data, marking a significant step forward in the fight against the impact of ransomware incidents. The two most important benefits are as follows: *(i)* users can determine whether their personal data are exposed and, if so, what data they are; and *(ii)* forensic analysis of such incidents is enhanced.

Acknowledging that the current version of *RDBAlert* is still a prototype, several strategic directions are proposed to further enhance and extend the tool, including the following:

- Conduction of temporal analyses to map the evolution of a data breach, identifying critical time windows during which data was accessed or modified.
- Integration of geolocation techniques into indexed documents to associate data leaks with specific regions or infrastructures while linking with external threat intelligence feeds (such as indicators of compromise or malicious file hashes) would enrich the analytical context.
- Data masking and end-to-end encryption, ensuring that identified PII is not exposed during queries or exports.
- Consideration of a security architecture for the system to strengthen confidence and ensure legal compliance.

**Author Contributions:** Conceptualization, J.M.T.-T. and P.G.-T.; methodology, J.M.T.-T. and P.G.-T.; software, J.M.T.-T.; validation, J.M.T.-T., P.G.-T., E.C.-F. and J.A.G.-H.; investigation, J.M.T.-T., P.G.-T. and E.C.-F.; data curation, J.M.T.-T.; writing—original draft preparation, J.M.T.-T., E.C.-F. and P.G.-T.; writing—review and editing, J.M.T.-T., P.G.-T., E.C.-F. and J.A.G.-H.; visualization, J.M.T.-T.; supervision, P.G.-T. and J.A.G.-H.; project administration, J.M.T.-T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by project C025/24 INCIBE-UGR, funded with European NextGeneration Funds.

**Data Availability Statement:** The original data presented in the study are available at http://7aqabivkwmpvjkyefonf3gpy5gsubopqni7kcirsrq3pflckxq5zz4id.onion; https://t.me/Osintcorp_chat; http://santat7kpllt6iyvqbr7q4amdv6dzrh6paatvyrzl7ry3zm72zigf4ad.onion and http://xfv4jzckytb4g3ckwemcny3ihv4i5p4lqzdpi624cxisu35my5fwi5qd.onion (all accessed on 14 October 2025). To access URLs in the .onion domain a Tor-enabled browser is required, which can be downloaded from the TOR project website (https://www.torproject.org/download/, accessed on 14 October 2025).

**Conflicts of Interest:** The authors declare no conflicts of interest.

**DURC Statement**: The current research is primarily intended for legitimate purposes, such as supporting law enforcement efforts or facilitating the early detection of personal data exposure within corporate environments—ultimately benefiting individual users and society as a whole. Authors are aware of *RDBAlert*'s potential dual use and thus we confirm that all necessary precautions have been taken to prevent potential misuse. As an ethical responsibility, authors strictly adhere to relevant national and international laws about dual use research of concern (DURC). Authors also advocate for responsible deployment, ethical considerations, regulatory compliance, and transparent reporting to mitigate misuse risks and foster beneficial outcomes.

## References

1. Sophos. The State of Ransomware 2024. Report, 2025. Available online: https://www.sophos.com/en-us/content/state-of-ransomware (accessed on 14 October 2025).
2. Cisco Umbrella. From Trojan Takeovers to Ransomware Roulette. Cisco Cyber Threat Trends Report, 2024. Available online: https://umbrella.cisco.com/info/cyber-threat-trends-report?utm_medium=search-paid&utm_source=google&utm_campaign=UMB_EMEA_EU_EN_GS_Nonbrand_Security_T1&utm_content=DNS-FY24-Q4-Content-Ebook-Cyber-Threat-Trends-Report&_bt=712318013894&_bk=latest%20cybersecurity%20trends&_bm=p&_bn=g&_bg=158065449997&gad_source=1&gclid=Cj0KCQiA_NC9BhCkARIsABSnSTb55LcWHiMIvNpFjTWsYi9lii816iuEfAPYavGB3EXZL_U8nzlgEx4aAhMaEALw_wcB (accessed on 14 October 2025).
3. Cyber Management Alliance. Top 10 Biggest Cyber Attacks of 2024 & 25 Other Attacks to Know About! Available online: https://www.cm-alliance.com/cybersecurity-blog/top-10-biggest-cyber-attacks-of-2024-25-other-attacks-to-know-about (accessed on 14 October 2025).
4. Check Point. Ransomware Annual Report 2024. Report 2024. Available online: https://cyberint.com/blog/research/ransomware-annual-report-2024/#:~:text=In%202024%2C%20the%20ransomware%20landscape,the%20remainder%20of%20the%20year (accessed on 14 October 2025).
5. Home Office. *Ransomware Legislative Proposals: Reducing Payments to Cyber Criminals and Increasing Incident Reporting*; Government Consultation, 2025. Available online: https://assets.publishing.service.gov.uk/media/67864097c6428e013188175a/Consultation-Document-Proposals-v2.pdf (accessed on 14 October 2025).
6. Hassan, N.A. *Ransomware Revealed. A Beginner's Guide to Protecting and Recovering from Ransomware Attacks*; Apress: New York, NY, USA, 2019; ISBN 978-1484242544.
7. Aggarwal, M. Ransomware Attack: An Evolving Targeted Threat. In Proceedings of the 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 6–8 July 2023; pp. 1–7. [CrossRef]
8. MacColl, J.; Husch, P.; Mott, G.; Sullivan, J.; Nurse, J.R.C.; Turner, S.; Pattnaik, N. *The Scourge of Ransomware. Victim Insights on Harms to Individuals, Organisations and Society*; Royal United Services Institute: London, UK, 2024. Available online: https://www.rusi.org/explore-our-research/publications/occasional-papers/ransomware-victim-insights-harms-individuals-organisations-and-society (accessed on 14 October 2025).

9. EU. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation). Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng (accessed on 14 October 2025).
10. EDPB. Guidelines 9/2022 on Personal Data Breach Notification Under GDPR. Version 2.0, 2023. Available online: https://www.edpb.europa.eu/system/files/2023-04/edpb_guidelines_202209_personal_data_breach_notification_v2.0_en.pdf (accessed on 14 October 2025).
11. State of California. California Consumer Privacy Act (CCPA). Available online: https://oag.ca.gov/privacy/ccpa (accessed on 14 October 2025).
12. Government of Canada. Personal Information Protection and Electronic Documents Act (S.C. 2000, c. 5). Available online: https://laws-lois.justice.gc.ca/eng/acts/p-8.6 (accessed on 14 October 2025).
13. Government of Japan. Act on the Protection of Personal Information (Act No. 57 of 2003). Available online: https://www.cas.go.jp/jp/seisaku/hourei/data/APPI.pdf (accessed on 14 October 2025).
14. Government of Brazil. General Personal Data Protection Act (LGPD). Available online: https://lgpd-brazil.info (accessed on 14 October 2025).
15. Roth, J. *Data Exfiltration in Ransomware Attacks: Digital Forensics Primer for Lawyers*; Kroll: New York, NY, USA, 2025. Available online: https://www.kroll.com/en/insights/publications/cyber/data-exfiltration-ransomware-attacks (accessed on 14 October 2025).
16. Fuentes, M.; Hacquebord, F.; Hilt, S.; Kenefick, I.; Kropotov, V.; McArdle, R.; Mercês, F.; Sancho, D. *Modern Ransomware's Double Extortion Tactics and How to Protect Enterprises Against Them*; Trend Micro, 2021. Available online: https://documents.trendmicro.com/assets/white_papers/wp-modern-ransomwares-double-extortion-tactics.pdf (accessed on 14 October 2025).
17. Imperva. *More Lessons Learned from Analyzing 100 Data Breaches*; Whitepaper, 2022. Available online: https://www.imperva.com/resources/whitepapers/More-Lessons-Learned-from-Analyzing-100-Data-Breaches_WP.pdf (accessed on 14 October 2025).
18. ArticWolf. *Artic Wolf 2025 Ransomware Report*; Arctic Wolf Networks Inc.: Eden Prairie, MN, USA, 2025. Available online: https://cybersecurity.arcticwolf.com/2025-Threat-Report-v1.html (accessed on 14 October 2025).
19. Price, A. *Data-Leak Site Emergence Continues to Increase*; Cyjax, London, UK, August 2024. Available online: https://www.cyjax.com/resources/blog/data-leak-site-emergence-continues-to-increase (accessed on 14 October 2025).
20. Center for Internet Security. *Ransomware: The Data Exfiltration and Double Extortion Trends*; Part 3; Center for Internet Security: East Greenbush, NY, USA. Available online: https://www.cisecurity.org/insights/blog/ransomware-the-data-exfiltration-and-double-extortion-trends (accessed on 14 October 2025).
21. Fisher, W.; Craft, R.E.; Ekstrom, M.; Sexton, J.; Sweetnam, J. Data Confidentiality: Detect, Respond to, and Recover from Data Breaches. NIST Special Publication 1800-29, February 2024. Available online: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1800-29.pdf (accessed on 14 October 2025).
22. Wikipedia. List of Data Breaches. Available online: https://en.wikipedia.org/wiki/List_of_data_breaches (accessed on 14 October 2025).
23. Hylender, C.D.; Langlois, P.; Pinto, A.; Widup, S. 2025 Data Breach Investigations Report. Verizon report, 2025. Available online: https://www.verizon.com/business/resources/reports/dbir (accessed on 14 October 2025).
24. Breachsense. The Most Recent Data Breaches in 2025. Report, 2025. Available online: https://www.breachsense.com/breaches (accessed on 14 October 2025).
25. Drapkin, A. Data Breaches That Have Happened in 2022, 2023, 2024 and 2025 so far. Tech.co report. Available online: https://tech.co/news/data-breaches-updated-list (accessed on 14 October 2025).
26. Bonta, R. Search Data Security Breaches. Available online: https://oag.ca.gov/privacy/databreach/list (accessed on 24 February 2025).
27. America's Cyber Defense Agency. StopRansomware: BianLian Ransomware Group. Available online: https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-136a (accessed on 14 October 2025).
28. Rubin, K. Ransomware and Healthcare: Why Hackers Target the Industry and How to Combat Attacks. Available online: https://www.linkedin.com/pulse/ransomware-healthcare-why-hackers-target-industry-how-kevin-rubin--efzhc (accessed on 14 October 2025).
29. HC3. New Threat Brief on Ransomware and Healthcare. Available online: https://dhinsights.org/news/new-threat-brief-on-ransomware-and-healthcare (accessed on 14 October 2025).
30. Autoriteit Persoonsgegevens. Report Data Breaches 2023. Report April 2024. Available online: https://www.autoriteitpersoonsgegevens.nl/en/system/files?file=2024-10/Report%20data%20breaches%202023.pdf (accessed on 14 October 2025).

31. Dalvi, A.; Kulkarni, P.; Kore, A.; Bhirud, S.G. Dark Web Crawling for Cybersecurity: Insights into Vulnerabilities and Ransomware Discussions. In Proceedings of the 2nd International Conference for Innovation in Technology (INOCON), Bangalore, India, 3–5 March 2023; pp. 1–6. [CrossRef]
32. Dalvi, A.; Bhirud, S. Dark web monitoring as an emerging cybersecurity strategy for businesses. *Int. J. Inf. Eng. Electron. Bus. (IJIEEB)* **2024**, *16*, 54–67. [CrossRef]
33. Recorded Future. Get Ahead of Present and Future Attacks with Recorded Future. Available online: https://www.recordedfuture.com (accessed on 14 October 2025).
34. GitHub. Aria2-Onion-Downloader. Available online: https://github.com/sn0b4ll/aria2-onion-downloader (accessed on 14 October 2025).
35. GitHub. Torget. Available online: https://github.com/pmorissette/torget (accessed on 14 October 2025).
36. GitHub. Torboost. Available online: https://github.com/tasooshi/torboost (accessed on 14 October 2025).
37. Network Working Group. Internet Message Format. RFC 2822. Available online: https://www.rfc-editor.org/rfc/rfc2822.html (accessed on 14 October 2025).
38. Ultralitics. Ultralitics Yolo11. Available online: https://docs.ultralytics.com/models/yolo11 (accessed on 14 October 2025).
39. Huang, D.; Yan, C.; Li, Q.; Peng, X. From Large Language Models to Large Multimodal Models: A Literature Review. *Appl. Sci.* **2024**, *14*, 5068. [CrossRef]
40. OpenBMB. MiniCPM: A Multimodal Large Language Model. 2024. Available online: https://github.com/OpenBMB/MiniCPM-o (accessed on 14 October 2025).
41. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 3982–3992. [CrossRef]
42. GitHub. The Source Code of RansomDBAlert. Available online: https://github.com/juanmill4/RansomDBAlert/tree/main (accessed on 14 October 2025).
43. Petkauskas, V. *Honeywell Hack Exposed Nearly 120 K People*; Cybernews, November 2023. Available online: https://cybernews.com/news/honeywell-breach-exposed-thousands (accessed on 14 October 2025).
44. Petkauskas, V. *Breach of Death Auditor PBI Exposes Details of 370,000 People*; Cybernews, July 2023. Available online: https://cybernews.com/news/pbi-data-breach-moveit (accessed on 14 October 2025).
45. Schappert, S. *TomTom, Pioneer Electronics, Autozone Latest to Navigate MOVEit Attacks*; Cybernews, November 2023. Available online: https://cybernews.com/news/tomtom-navigates-moveit-attacks-clop (accessed on 14 October 2025).
46. Insurance Business. PhilHealth Hack Potentially Exposes 42 Million People. Available online: https://www.insurancebusinessmag.com/asia/news/cyber/philhealth-hack-potentially-exposes-42-million-people-496453.aspx (accessed on 14 October 2025).
47. EuropaPress. Targa Viasat Suffers Cyberattack Compromising Nearly 100 GB of Financial and Personal Documents. Available online: https://www.europapress.es/motor/sector-00644/noticia-targa-viasat-sufre-ciberataque-compromete-casi-100-gb-documentos-financieros-personales-20240704185418.html (accessed on 14 October 2025).
48. UIDAI. Unique Identification Authority of India. Available online: https://www.uidai.gov.in/en/about-uidai/unique-identification-authority-of-india.html (accessed on 14 October 2025).
49. Cevallos-Salas, D.; Estrada-Jiménez, J.; Guamán, D.S.; Urquiza-Aguiar, L. Ransomware dynamics: Mitigating personal data exfiltration through the SCIRAS lens. *Comput. Secur.* **2025**, *157*, 104583. [CrossRef]