Email Analyzer
Sistema di Analisi di Sicurezza Multi-Servizio per Email IMAP

Montesi Flavio

Giugno 2025

Indice

1 Introduzione

L'applicazione "Email Analyzer" è uno strumento avanzato progettato per analizzare le email scaricate da un server IMAP con un approccio multi-servizio alla sicurezza informatica. Questo sistema fornisce un'analisi completa e dettagliata di vari componenti delle email, inclusi header, link, allegati e verifiche di autenticazione.

2 Architettura del Sistema

2.1 Panoramica Generale

Il sistema Email Analyzer è progettato seguendo un'architettura modulare. Al centro dell'applicazione troviamo la classe `app.py`, che funge da entry point principale e gestisce tutti i parametri della command line interface, orchestrando l'intero processo di analisi.

La gestione delle connessioni IMAP e l'elaborazione delle email è affidata al modulo `email_core.py`, che si occupa di stabilire connessioni sicure con i server di posta e di processare i messaggi nel loro formato nativo. Questo approccio permette di preservare tutte le informazioni tecniche necessarie per l'analisi.

Per quanto riguarda l'analisi vera e propria, il sistema si avvale di moduli specializzati contenuti nella directory `/analyzers`, ciascuno focalizzato su un aspetto specifico della sicurezza email. L'interfacciamento con i servizi di sicurezza esterni è gestito dal modulo `connectors.py`, che implementa tutte le logiche di comunicazione con le API dei vari provider.

La generazione dei report, sia in formato JSON che HTML, è delegata ai moduli contenuti nella directory `/output`, mentre la configurazione centralizzata del sistema, incluse le chiavi API e i parametri operativi, è gestita dal file `config.py`.

2.2 Moduli di Analisi Specializzati

L'architettura modulare del sistema comprende quattro componenti di analisi specializzati, ciascuno progettato per affrontare specifici aspetti della sicurezza email.

L'**Header Analyzer** si occupa dell'analisi completa degli header email, implementando un sistema di categorizzazione automatica che permette di organizzare e interpretare efficacemente i metadati del messaggio.

Il **Link Analyzer** è responsabile dell'estrazione e della verifica di sicurezza di tutti i collegamenti presenti nell'email, utilizzando tecniche avanzate di pattern matching e normalizzazione degli URL.

L'**Attachment Analyzer** si concentra sull'analisi degli hash degli allegati e sul rilevamento di potenziali malware.

Il **DMARC Analyzer** verifica l'implementazione e l'efficacia dei protocolli di autenticazione DKIM, SPF e DMARC.

3 Funzionalità Principali

3.1 Download e Gestione Email

3.1.1 IMAP (Internet Message Access Protocol)

Definizione: IMAP rappresenta il protocollo standard per l'accesso alle email memorizzate su un server di posta elettronica. A differenza del protocollo POP3, che scarica i messaggi localmente rimuovendoli dal server, IMAP mantiene tutti i messaggi sul server permettendo l'accesso sincronizzato da più dispositivi.

Le caratteristiche principali di IMAP includono la sincronizzazione bidirezionale tra client e server, che garantisce che le modifiche apportate da un dispositivo siano visibili su tutti gli altri. Il protocollo offre inoltre supporto nativo per cartelle multiple e gerarchiche, permettendo un'organizzazione avanzata dei messaggi. Una funzionalità particolarmente utile è l'accesso parziale ai messaggi, che consente di scaricare selettivamente solo gli header, solo il corpo, o specifiche parti del messaggio, ottimizzando l'utilizzo della banda. Infine, IMAP gestisce efficacemente gli stati dei messaggi, tracciando informazioni come letto, non letto, eliminato e altre flag personalizzabili.

3.1.2 Implementazione nel Sistema

La gestione delle connessioni IMAP nel sistema è stata implementata con particolare attenzione alla sicurezza e all'affidabilità. Il sistema si connette esclusivamente a server IMAP utilizzando protocolli SSL/TLS, garantendo che tutte le comunicazioni siano crittografate. Le email vengono scaricate e preservate nel loro formato originale (.eml), mantenendo intatti tutti i metadati e le informazioni tecniche necessarie per l'analisi di sicurezza.

Il processo di connessione include l'autenticazione sicura con le credenziali utente fornite, il download selettivo basato sulla mailbox specificata, e la preservazione completa del formato originale dell'email. Il sistema implementa inoltre una gestione robusta degli errori, includendo la gestione di timeout di connessione, errori di autenticazione e problemi di rete temporanei.

3.2 Analisi degli Header

3.2.1 Header Email

Definizione: Gli header email costituiscono un insieme di metadati strutturati che precedono il corpo di un messaggio di posta elettronica. Questi metadati contengono informazioni tecniche fondamentali riguardanti il percorso seguito dal messaggio, i meccanismi di autenticazione utilizzati e le caratteristiche generali della comunicazione.

Gli header possono essere suddivisi in diverse categorie funzionali. Gli **header di routing** tracciano il percorso completo che il messaggio ha seguito attraverso i vari server di posta, includendo informazioni come Received e Return-Path. Gli **header di identificazione** servono a identificare univocamente mittente e destinatario, attraverso campi come From, To e Message-ID. Gli **header di autenticazione** contengono tutte le informazioni relative ai meccanismi di verifica dell'identità, come DKIM-Signature e Authentication-Results. Infine, gli **header di sicurezza** riportano i risultati delle analisi antispam e antimalware effettuate dai vari server durante il transito del messaggio.

3.3 Analisi dei Link

3.3.1 Phishing

Definizione: Il phishing rappresenta una delle tecniche di ingegneria sociale più diffuse e pericolose nell'ambito della cybersecurity. Questa metodologia di attacco utilizza comunicazioni elettroniche fraudolente, principalmente email, per ingannare le vittime e rubare informazioni sensibili quali credenziali di accesso, dati bancari o informazioni personali. L'efficacia del phishing risiede nella capacità degli attaccanti di presentarsi come entità legittime e affidabili.

Le caratteristiche tipiche di un attacco phishing includono l'utilizzo di URL contraffatti che imitano fedelmente l'aspetto di siti web legittimi, spesso utilizzando domini molto simili a quelli originali attraverso tecniche di typosquatting. Gli attaccanti creano inoltre un senso di urgenza nelle loro comunicazioni, richiedendo azioni immediate come l'aggiornamento delle credenziali o la verifica dell'account. Per aumentare la credibilità del messaggio, vengono spesso copiati loghi, grafiche e elementi di design delle organizzazioni impersonate.

3.3.2 Sistema di Analisi Comprehensive

L'implementazione dell'analisi dei link nel sistema segue un approccio multi-servizio avanzato che massimizza l'accuratezza del rilevamento delle minacce. L'**estrazione intelligente** utilizza algoritmi per identificare URL nascosti sia nel contenuto HTML che nel testo puro, garantendo che nessun collegamento sfugga all'analisi.

Il processo di **normalizzazione** standardizza tutti gli URL rilevati per evitare che variazioni sintattiche dello stesso collegamento vengano considerate come entità separate.

Il **sistema di caching** rappresenta una delle innovazioni più significative dell'implementazione, riducendo il numero di chiamate API duplicate e migliorando le performance complessive del sistema. Il **risk scoring** calcola un punteggio di rischio più accurato applicando criteri di valutazione più severi per URL potenzialmente pericolosi, inclusi i servizi di abbreviazione e i collegamenti contenenti termini sospetti.

3.3.3 Riconoscimento di Link Pericolosi

Il sistema è sviluppato per rilevare i collegamenti potenzialmente pericolosi, andando oltre i tradizionali metodi di blacklist. Il **riconoscimento degli URL shortener** si basa su un database completo di oltre 40 servizi di abbreviazione URL, molti dei quali sono frequentemente utilizzati nelle campagne di phishing per nascondere la destinazione reale del collegamento.

La **detection più restrittiva** rappresenta un approccio dove anche un singolo rilevamento positivo su qualsiasi servizio di sicurezza può determinare un livello di rischio elevato, specialmente per determinate categorie di URL. Questo approccio più conservativo è giustificato dalla natura critica delle minacce email-based.

Logica di Classificazione del Rischio
Per massimizzare la sicurezza del sistema, la logica di classificazione del rischio è stata significativamente migliorata rispetto agli standard tradizionali. Per gli **URL standard**, il sistema considera un collegamento come non sicuro quando vengono rilevati due o più rilevamenti positivi sui servizi di sicurezza, oppure quando almeno il 2% dei motori di analisi riporta la presenza di minacce.

Tuttavia, per gli **URL shortener**, i criteri di valutazione diventano molto più severi: anche un solo rilevamento positivo su qualsiasi servizio è sufficiente per classificare il collegamento come non sicuro. Questa decisione è motivata dalla natura intrinsecamente opaca di questi servizi, che rendono impossibile una valutazione preliminare del rischio da parte dell'utente.

Gli **URL contenenti termini sospetti** vengono valutati con una severità intermedia ma comunque elevata: anche con un singolo rilevamento, questi collegamenti vengono classificati come potenzialmente pericolosi, richiedendo un'analisi più approfondita prima di essere considerati sicuri.

3.3.4 URL Shorteners e Sicurezza

I servizi di abbreviazione URL rappresentano una sfida particolare nel panorama della sicurezza informatica per diverse ragioni fondamentali. Innanzitutto, questi servizi nascondono completamente la destinazione effettiva del collegamento al destinatario dell'email, impedendo qualsiasi valutazione preliminare del rischio basata sul dominio di destinazione.

Questa opacità consente agli URL shortener di eludere facilmente i filtri di sicurezza tradizionali che si basano su blacklist di domini specifici, poiché il dominio visibile è quello del servizio di abbreviazione (spesso legittimo) piuttosto che quello della destinazione malevola. Non sorprende quindi che questi servizi siano diventati strumenti ampiamente utilizzati nelle campagne di phishing proprio per la loro capacità di mascherare URL malevoli.

L'impossibilità per l'utente finale di effettuare una valutazione immediata del rischio rappresenta un ulteriore elemento di vulnerabilità, costringendo a fare affidamento esclusivamente sui risultati dell'analisi automatizzata. Per queste ragioni, il sistema applica automaticamente una valutazione del rischio più severa a tutti gli URL shortener, assegnando un punteggio di rischio base elevato indipendentemente dai risultati specifici dei servizi di sicurezza consultati.

3.4 Analisi degli Allegati

3.4.1 Hash Crittografici

Definizione: Gli hash crittografici rappresentano funzioni matematiche fondamentali nella sicurezza informatica, progettate per convertire dati di qualsiasi dimensione in una stringa di caratteri di lunghezza fissa. Queste funzioni sono utilizzate principalmente per verificare l'integrità dei file e per identificare univocamente contenuti digitali, rappresentando un pilastro fondamentale nei sistemi di rilevamento malware.

Tra gli algoritmi più comunemente utilizzati troviamo l'**MD5**, che genera hash di 128 bit ed è particolarmente apprezzato per la sua velocità di calcolo, sebbene sia ormai considerato vulnerabile a attacchi di collisione. L'**SHA1** produce hash di 160 bit ed era molto diffuso in passato, ma è ora deprecato per applicazioni crittografiche critiche a causa di vulnerabilità note. L'**SHA256**, che genera hash di 256 bit, rappresenta attualmente lo standard per applicazioni di sicurezza, offrendo un ottimo equilibrio tra sicurezza e performance.

3.4.2 Malware via Email

Definizione: Il malware distribuito tramite email rappresenta uno dei vettori di attacco più persistenti e pericolosi nel panorama delle minacce informatiche. Questo tipo di software dannoso viene tipicamente distribuito attraverso allegati email che contengono direttamente il payload malevolo, oppure attraverso collegamenti che dirigono la vittima verso siti web che effettuano il download automatico di componenti dannosi.

Le tipologie più comuni di malware email-based includono i **Trojan**, software che si nascondono dietro l'apparenza di applicazioni legittime per eludere i sospetti dell'utente e ottenere accesso non autorizzato al sistema. I **Ransomware** rappresentano una categoria particolarmente pericolosa, in quanto crittografano i file dell'utente richiedendo il pagamento di un riscatto per il ripristino dei dati.

Gli **Stealer** sono progettati specificamente per rubare credenziali, informazioni sensibili e dati personali, spesso operando in modo silenzioso per lunghi periodi. I **Botnet Agent** trasformano il sistema infetto in parte di una rete controllata da cybercriminali, utilizzandolo per attività dannose coordinate come attacchi DDoS o spam.

3.4.3 Implementazione nel Sistema

Il sistema implementa un approccio multi-layer per l'analisi degli allegati, calcolando simultaneamente hash multipli (MD5, SHA1, SHA256) per ogni file allegato per massimizzare le possibilità di identificazione. Questi hash vengono quindi verificati contro database specializzati di malware come VirusTotal e MalwareBazaar, permettendo l'identificazione non solo della presenza di malware, ma anche della specifica famiglia di appartenenza.

Il sistema include inoltre funzionalità avanzate come l'identificazione automatica della famiglia malware e un sistema di quarantena automatica per file che risultano sospetti dall'analisi, prevenendo potenziali danni al sistema ospitante.

3.5 Verifica Autenticazione Email

3.5.1 SPF (Sender Policy Framework)

Definizione: Meccanismo di autenticazione email che consente ai proprietari di domini di specificare quali server di posta sono autorizzati a inviare email per conto del loro dominio.

Funzionamento:

- Pubblicazione di record DNS TXT contenenti le policy SPF
- Verifica da parte del server ricevente dell'IP mittente contro il record SPF
- Risultati possibili: Pass, Fail, SoftFail, Neutral, None, TempError, PermError

Esempio record SPF:

```
v=spf1 ip4:192.168.1.0/24 include:_spf.google.com -all
```

Listing 1: Esempio record SPF

3.5.2 DKIM (DomainKeys Identified Mail)

Definizione: Sistema di autenticazione che utilizza crittografia a chiave pubblica per verificare che un'email non sia stata alterata durante il transito e che provenga effettivamente dal dominio dichiarato.

Componenti chiave:

- **Chiave privata:** Utilizzata dal server mittente per firmare l'email
- **Chiave pubblica:** Pubblicata nei record DNS per la verifica
- **Selettore:** Identificatore che specifica quale chiave utilizzare
- **Firma digitale:** Hash crittografico di header e corpo specifici

Struttura firma DKIM:

```
DKIM-Signature: v=1; a=rsa-sha256; d=example.com; s=selector1;
    h=from:to:subject:date; bh=hash_corpo; b=firma_crittografata
```

Listing 2: Struttura firma DKIM

3.5.3 DMARC (Domain-based Message Authentication, Reporting & Conformance)

Definizione: Policy di autenticazione che si basa su SPF e DKIM per determinare l'autenticità di un'email e specificare le azioni da intraprendere per i messaggi che falliscono l'autenticazione.

Policy DMARC:

- **none:** Solo monitoraggio, nessuna azione
- **quarantine:** Messaggio considerato sospetto (spam folder)
- **reject:** Messaggio rifiutato completamente

Allineamento:

- **Strict:** Il dominio deve corrispondere esattamente
- **Relaxed:** Sottodomini accettati

3.5.4 Email Spoofing

Definizione: Falsificazione dell'indirizzo mittente di un'email per far sembrare che provenga da una fonte diversa da quella reale.

Tecniche comuni:

- Manipolazione header FROM
- Uso di domini simili (lookalike domains)
- Display name spoofing
- Reply-To manipulation

4 Servizi di Sicurezza Integrati

4.1 Servizi Primari

4.1.1 VirusTotal

Definizione: VirusTotal rappresenta uno dei servizi online più autorevoli e completi per l'analisi di sicurezza, offrendo gratuitamente l'accesso a una piattaforma che integra oltre 70 motori antivirus e servizi specializzati nel rilevamento di malware e minacce informatiche.

L'implementazione di VirusTotal nel sistema sfrutta appieno le capacità di **scansione multi-engine**, permettendo l'analisi simultanea di file, URL e indirizzi IP attraverso più di 60 motori antivirus differenti. Questa diversificazione massimizza le probabilità di rilevamento anche delle minacce più sofisticate o recenti.

L'integrazione tecnica utilizza **context manager** per una gestione ottimale delle sessioni di comunicazione con le API, implementando una **retry logic** intelligente che gestisce automaticamente i fallimenti temporanei e un sistema di **cache intelligente** che riduce il numero di chiamate API duplicate. Il servizio fornisce inoltre accesso a un vasto **database storico** di minacce conosciute, permettendo l'identificazione di campagne di attacco già documentate. L'**integrazione API** è completamente automatizzata e rispetta i limiti di chiamate imposti dal servizio.

4.1.2 PhishTank

Definizione: PhishTank costituisce un database collaborativo e gratuito di URL di phishing, gestito da OpenDNS (ora parte di Cisco) e alimentato da una community globale di esperti di sicurezza che verificano e catalogano le minacce in tempo reale.

L'implementazione nel sistema sfrutta la natura **community-driven** del database, beneficiando di aggiornamenti costanti e in tempo reale provenienti da analisti di sicurezza distribuiti globalmente. Il servizio offre il vantaggio significativo di essere completamente **gratuito** e di non richiedere chiavi API per l'accesso base, semplificando notevolmente l'integrazione.

La **verifica manuale** effettuata dalla community garantisce un'alta qualità dei dati, riducendo significativamente i falsi positivi. Inoltre, PhishTank offre una **categorizzazione avanzata** che classifica gli attacchi di phishing per tipo di target (banche, social media, servizi cloud, etc.), fornendo contesto aggiuntivo per l'analisi delle minacce.

4.2 Servizi Specializzati

4.2.1 URLScan.io

Definizione: URLScan.io rappresenta un servizio di analisi comportamentale avanzata che va oltre il semplice controllo delle blacklist, eseguendo una scansione completa e dinamica delle pagine web in un ambiente sandbox controllato e sicuro.

Il servizio eccelle nell'**analisi comportamentale** dei siti web, simulando la visita di un utente reale e monitorando tutte le attività che si verificano durante il caricamento della pagina. Questa capacità include la generazione di **screenshot della pagina renderizzata**, che permette agli analisti di verificare visivamente l'aspetto del sito e identificare tentativi di impersonificazione di brand legittimi.

L'**analisi del traffico di rete** monitora tutte le comunicazioni che avvengono durante il caricamento della pagina, identificando connessioni sospette, download automatici e redirect nascosti. Il servizio è inoltre in grado di rilevare le **tecnologie utilizzate** dal sito web e di identificare la presenza di **JavaScript malevolo** o offuscato.

Mentre l'utilizzo base del servizio non richiede una **API key**, l'accesso alle funzionalità avanzate e a rate limits più elevati è disponibile per utenti registrati. L'esecuzione di tutte le analisi avviene in un **ambiente sandbox** completamente isolato, garantendo la sicurezza dell'infrastruttura di analisi.

4.2.2 MalwareBazaar

Definizione: MalwareBazaar è un database pubblico e gratuito di campioni malware mantenuto da abuse.ch, un'organizzazione non-profit svizzera specializzata nella raccolta e condivisione di intelligence sulle minacce informatiche.

Il servizio si distingue come **database specializzato** esclusivamente nel malware, offrendo funzionalità avanzate di **hash lookup** e **identificazione della famiglia** malware. Questa specializzazione permette una classificazione molto più dettagliata e accurata rispetto ai servizi generalisti.

Il database è **completamente gratuito** e viene costantemente aggiornato con nuovi campioni provenienti da ricercatori di sicurezza, organizzazioni e sistemi automatizzati di rilevamento. Oltre alla semplice identificazione, il servizio fornisce **metadati dettagliati** sui campioni, inclusi informazioni sulla classificazione delle famiglie malware e **timeline di scoperta** che aiutano a comprendere l'evoluzione delle minacce.

L'**integrazione API** è semplice e diretta, senza limiti particolarmente stringenti che potrebbero impedire l'utilizzo in contesti di analisi intensiva.

4.2.3 AbuseIPDB

AbuseIPDB rappresenta una risorsa fondamentale per la valutazione della reputazione degli indirizzi IP, specializzandosi nella raccolta e nell'analisi di report di abuso provenienti da una vasta community di professionisti della sicurezza informatica.

Il servizio eccelle nella **valutazione della reputazione IP** fornendo un **confidence score** che quantifica la probabilità che un determinato indirizzo IP sia coinvolto in attività malevole. Le funzionalità di **geolocalizzazione** e **identificazione ISP** forniscono contesto geografico e infrastrutturale utile per l'analisi delle minacce.

L'accesso alle **funzionalità complete** richiede una chiave API, ma il valore aggiunto giustifica questo requisito. Il database è alimentato da **report della community** di sicurezza informatica, garantendo informazioni aggiornate e relevant. Il servizio mantiene inoltre **dati storici** dettagliati delle segnalazioni e delle attività sospette associate a ciascun IP, permettendo analisi longitudinali delle minacce.

5 Sistema di Risk Scoring

5.1 Metodologia di Risk Scoring Migliorata

Il sistema usa un algoritmo di risk scoring che rappresenta un'evoluzione significativa rispetto ai metodi tradizionali di valutazione delle minacce. Questo algoritmo combina i risultati di tutti i servizi di sicurezza disponibili.

Il sistema utilizza una **scala standardizzata da 0 a 10**, dove il valore 10 rappresenta il massimo rischio possibile. Questa standardizzazione facilita l'interpretazione dei risultati e la definizione di policy di sicurezza basate su soglie numeriche chiare.

Un aspetto innovativo dell'implementazione è il **peso differenziato** assegnato a ciascun servizio di sicurezza: i servizi considerati più affidabili o specializzati ricevono un peso maggiore nella valutazione finale, mentre servizi meno specializzati o con tassi di falsi positivi più elevati influenzano meno il punteggio complessivo.

Il sistema introduce inoltre **punteggi base** per determinate categorie di URL: i servizi di abbreviazione URL e gli URL contenenti termini sospetti ricevono automaticamente un punteggio di rischio base anche prima che inizi l'analisi vera e propria. Questo approccio riflette la natura intrinsecamente rischiosa di queste categorie.

La **valutazione progressiva** garantisce che anche un singolo rilevamento positivo abbia un impatto significativo sul punteggio finale, abbandonando l'approccio tradizionale che spesso ignorava rilevamenti isolati. Infine, il sistema genera automaticamente **raccomandazioni standardizzate** (SAFE, WARNING, CAUTION, BLOCK) basate sul punteggio finale, facilitando la presa di decisioni operative.

5.1.1 Logica di Valutazione del Risk Score

Il sistema valuta il rischio secondo questa logica:

1. **Analisi preliminare dell'URL**:
   - Se l'URL è un servizio di abbreviazione (shortener): +2.0 punti base di rischio
   - Se contiene termini sospetti (login, verify, account, ecc.): +1.5 punti base
2. **Integrazione risultati VirusTotal**:
   - 1 rilevamento positivo: +2 punti (precedentemente +1)
   - 2 rilevamenti positivi: +3 punti (precedentemente +2)
   - 3+ rilevamenti: scala progressivamente fino a +5 punti
   - Se la percentuale di rilevamenti supera il 5%: +1 punto aggiuntivo
3. **Valutazione PhishTank**: fino a +4 punti (se verificato come phishing: +5)
4. **Valutazione URLVoid**: fino a +4 punti (anche con un singolo rilevamento: +1.5)
5. **Casi speciali**:
   - URL shortener senza risultati dai servizi: almeno +3 punti di rischio base
   - URL shortener con qualsiasi rilevamento positivo: almeno +4 punti di rischio

5.1.2 Esempio di Valutazione

Un link può essere valutato secondo questi passaggi (esempio concettuale):

1. L'URL è `bit.ly/abc123` → URL shortener identificato → +2.0 punti di rischio base
2. VirusTotal riporta 1 rilevamento positivo su 80 scanner → +2.0 punti
3. Google Safe Browsing non rileva minacce → +0 punti
4. PhishTank non ha informazioni → +0 punti
5. URLVoid non rileva problemi → +0 punti
6. Caso speciale: URL shortener con almeno un rilevamento positivo → minimo +4 punti
7. Punteggio finale: 4.0/10, classificato come "Medium Risk"

Questa metodologia garantisce che anche piccoli segnali di potenziale rischio vengano adeguatamente considerati, soprattutto quando combinati con fattori di rischio come l'utilizzo di servizi di abbreviazione URL.

6 Ottimizzazioni e Performance

6.1 Sistema di Caching

Il sistema implementa un sofisticato meccanismo di caching progettato per ottimizzare significativamente le performance dell'applicazione riducendo al minimo le chiamate API ridondanti. Il **caching basato su file** salva persistentemente tutti i risultati delle API in file JSON strutturati, permettendo il riutilizzo delle informazioni anche tra sessioni diverse di analisi.

La **generazione delle chiavi di cache** utilizza hash MD5 degli URL e degli indirizzi IP per creare identificatori univoci che garantiscono l'accesso rapido ai dati salvati senza possibilità di collisioni. Il sistema include una funzionalità di **scadenza automatica** che rimuove i dati di cache dopo 7 giorni (periodo configurabile), garantendo che le informazioni rimangano aggiornate rispetto all'evoluzione delle minacce.

Un ulteriore livello di ottimizzazione è rappresentato dalla **deduplicazione in memoria**, che elimina le richieste duplicate prima ancora che vengano inviate alle API esterne, riducendo il carico sui servizi esterni e migliorando i tempi di risposta.

6.2 Gestione Errori e Resilienza

Il sistema implementa un framework comprehensive di gestione degli errori progettato per garantire la massima resilienza operativa. La **retry logic** utilizza un algoritmo di exponential backoff per gestire automaticamente gli errori transitori, aumentando progressivamente gli intervalli tra i tentativi per evitare di sovraccaricare servizi temporaneamente non disponibili.

La filosofia di **graceful degradation** permette al sistema di continuare l'analisi anche quando alcuni servizi non sono disponibili, fornendo comunque risultati parziali ma utili basati sui servizi funzionanti. Il **quota management** monitora intelligentemente l'utilizzo delle quote API, rallentando automaticamente il ritmo delle richieste quando ci si avvicina ai limiti imposti dai provider.

6.3 Batch Processing

Per rispettare efficacemente i rate limits imposti dai servizi esterni, il sistema organizza tutte le operazioni in lotti ottimizzati. L'elaborazione degli URL avviene in gruppi di 5 unità, dimensione che rappresenta un compromesso ottimale tra efficienza e rispetto dei vincoli API.

Il sistema implementa pause automatiche tra i batch, calcolate dinamicamente in base ai rate limits specifici di ciascun servizio. Il monitoraggio dei rate limits avviene in tempo reale, permettendo al sistema di adattare automaticamente il proprio comportamento per evitare interruzioni. Infine, un sistema di prioritizzazione assicura che le richieste critiche per la sicurezza vengano elaborate con precedenza rispetto a quelle meno urgenti.

7 Implementazione Tecnica

7.1 Requisiti Sistema

- Python 3.8+
- Librerie principali:
  - `imaplib`: Connessioni IMAP
  - `requests`: Chiamate API HTTP
  - `dnspython`: Verifiche DNS per DMARC
  - `aiohttp`: Chiamate asincrone (VirusTotal)
  - `hashlib`: Calcolo hash file
  - `email`: Parsing email RFC-compliant

7.2 Testing e Validazione

Per verificare l'efficacia del sistema di analisi, sono stati condotti test utilizzando **GoPhish**, una piattaforma open source specializzata nella creazione di campagne di phishing simulate per scopi di formazione e testing di sicurezza. GoPhish ha permesso di generare email di phishing controllate con caratteristiche specifiche, consentendo di validare la capacità del sistema di rilevare correttamente:

- URL di phishing con domini contraffatti
- Messaggi con tecniche di spoofing degli header
- Link abbreviati (URL shortener) utilizzati per nascondere destinazioni malevole
- Email con contenuti sospetti progettati per eludere i filtri tradizionali

Questi test hanno confermato l'efficacia dell'approccio multi-servizio nell'identificazione delle minacce e hanno permesso di calibrare accuratamente gli algoritmi di risk scoring.

7.3 Configurazione

7.3.1 File di Configurazione

```
# Configurazione API Keys
VIRUSTOTAL_API_KEY = "your_virustotal_api_key_here"
URLSCAN_API_KEY = "your_urlscan_api_key_here"
ABUSEIPDB_API_KEY = "your_abuseipdb_api_key_here"
URLVOID_API_KEY = "your_urlvoid_api_key_here"
GOOGLE_SAFE_BROWSING_API_KEY = "your_google_api_key_here"

# URLs dei servizi
VIRUSTOTAL_BASE_URL = "https://www.virustotal.com/vtapi/v2/"
URLSCAN_BASE_URL = "https://urlscan.io/api/v1/"
PHISHTANK_URL = "http://checkurl.phishtank.com/checkurl/"
MALWAREBAZAAR_URL = "https://mb-api.abuse.ch/api/v1/"

# Configurazione analisi sicurezza
SECURITY_ANALYSIS_CONFIG = {
    'enable_multiple_services': True,
    'virustotal_enabled': True,
    'urlscan_enabled': True,
    'phishtank_enabled': True,
    'malwarebazaar_enabled': True,
    'abuseipdb_enhanced': True,
    'batch_size': 5,
    'request_delay': 1.0,
    'max_retries': 3
}
```

Listing 3: Configurazione chiavi API

7.4 Utilizzo

7.4.1 Interfaccia Command Line

```
# Analisi completa con investigazione
python app.py -s imap.server.com -u user@example.com -p password -m INBOX -o emails --complete
    --investigate

# Analisi file locali
python app.py -f emails/ -i -o results.html

# Solo analisi header
python app.py -f emails/ --header -o headers.json
```

Listing 4: Esempi utilizzo CLI

8 Output e Visualizzazione

8.1 Formato JSON

Il sistema genera output JSON strutturato con le seguenti sezioni principali:

- **Information**: Metadata scansione (timestamp, file analizzato)
- **Headers**: Analisi completa header con investigazione IP
- **Links**: Analisi URL con risk scoring multi-servizio
- **Digests**: Hash file con verifica malware
- **Authentication**: Risultati DKIM, SPF, DMARC

8.2 Spiegazione Dettagliata Campi JSON

8.2.1 Sezione Information

La sezione `Information` contiene i metadati della scansione:

- **Filename**: Path completo del file email analizzato
- **Generated**: Timestamp di quando è stata eseguita l'analisi
- **Scan_Type**: Tipo di scansione (completa, header-only, etc.)
- **Investigation_Mode**: Indica se è stata eseguita l'investigazione con servizi esterni

8.2.2 Sezione Headers - Data

Contiene i metadati estratti dall'email:

- **from**: Indirizzo mittente estratto dall'header FROM
- **to**: Destinatario principale dall'header TO
- **subject**: Oggetto dell'email decodificato
- **date**: Data di invio parsata dall'header DATE
- **message-id**: Identificatore univoco dell'email
- **received**: Catena completa header RECEIVED per tracciamento percorso
- **content-type**: Tipo MIME del contenuto email
- **x-spam-status**: Risultato analisi antispam del server
- **dkim-signature**: Firma digitale DKIM se presente
- **authentication-results**: Risultati verifica SPF/DKIM/DMARC

8.2.3 Sezione Headers - Investigation

Risultati dell'investigazione di sicurezza:

- **X-Sender-Ip**:
  - `IP`: Indirizzo IP estratto dall'header RECEIVED
  - `Virustotal`: Link diretto per verificare IP su VirusTotal
  - `Abuseipdb`: Link per controllo reputazione su AbuseIPDB
  - `Safety`: Valutazione sicurezza (Safe/Suspicious/Malicious)
  - `Positives`: Numero motori antivirus che segnalano l'IP
  - `Country`: Geolocalizzazione IP
  - `ISP`: Provider internet dell'IP
  - `Abuse_Confidence`: Percentuale confidenza abuso (0-100%)
- **Blacklist_Check**:
  - `Blacklist_Status`: Stato nelle blacklist principali
  - `Listed_In`: Elenco blacklist che contengono l'IP
  - `Spamhaus_SBL`: Presenza in Spamhaus SBL
  - `Spamcop`: Presenza in SpamCop blacklist
- **Spoof_Check**:
  - `Reply-To`: Indirizzo Reply-To se diverso da FROM
  - `From`: Indirizzo FROM originale
  - `Conclusion`: Valutazione possibile spoofing

8.2.4 Sezione Links - Data

Elenco numerato di tutti i link estratti:

- **Numerazione**: Ogni link ha un ID numerico progressivo
- **URL**: Link completo estratto dal contenuto email
- **Type**: Tipo di link (HTTP, HTTPS, MAILTO, FTP)
- **Domain**: Dominio estratto dall'URL
- **Path**: Percorso specifico dell'URL

8.2.5 Sezione Links - Investigation

Analisi di sicurezza per ogni link:

- **Comprehensive_Analysis**:
  - `Risk_Score`: Punteggio rischio 0-10 calcolato
  - `Recommendation`: Raccomandazione (SAFE/WARNING/CAUTION/BLOCK)
  - `Total_Services`: Numero servizi che hanno analizzato l'URL
  - `Malicious_Count`: Servizi che lo segnalano come malevolo
- **Service_Results**:
  - `VirusTotal`: Risultati VirusTotal (positives/total, permalink)
  - `PhishTank`: Presenza nel database phishing
  - `URLScan`: Risultati analisi comportamentale
  - `URLVoid`: Aggregazione motori reputazione
  - `Google_Safe_Browsing`: Risultato Google Safe Browsing
- **Fallback_Links**:
  - `Virustotal`: Link diretto ricerca manuale
  - `Urlscan`: Link ricerca URLScan.io
  - `Google`: Link ricerca Google Safe Browsing

8.2.6 Sezione Digests - Data

Hash calcolati per il file email:

- **File_MD5**: Hash MD5 del file .eml completo
- **File_SHA1**: Hash SHA1 del file .eml completo
- **File_SHA256**: Hash SHA256 del file .eml completo
- **Content_MD5**: Hash MD5 del solo contenuto email
- **Content_SHA1**: Hash SHA1 del solo contenuto email
- **Content_SHA256**: Hash SHA256 del solo contenuto email
- **Attachment_Hashes**: Array di hash per ogni allegato

8.2.7 Sezione Digests - Investigation

Verifica sicurezza degli hash:

- **Per ogni hash**:
  - `Virustotal`: Link diretto controllo hash su VirusTotal
  - `MalwareBazaar`: Risultato controllo su MalwareBazaar
  - `Is_Malware`: Boolean indicante se riconosciuto come malware
  - `Malware_Family`: Famiglia malware identificata
  - `First_Seen`: Data prima identificazione
  - `Detection_Ratio`: Rapporto detection motori antivirus

8.2.8 Sezione Authentication (DMARC/DKIM)

Risultati verifica protocolli autenticazione:

- **DMARC_Policy**:
  - `Policy`: Politica DMARC del dominio (none/quarantine/reject)
  - `Percentage`: Percentuale email sottoposte a policy
  - `Subdomain_Policy`: Policy per sottodomini
  - `DKIM_Alignment`: Modalità allineamento DKIM
  - `SPF_Alignment`: Modalità allineamento SPF
- **DKIM_Analysis**:
  - `Signature_Valid`: Validità strutturale firma DKIM
  - `Algorithm`: Algoritmo crittografico utilizzato
  - `Domain`: Dominio che ha firmato l'email
  - `Selector`: Selettore DKIM utilizzato
  - `Headers_Signed`: Lista header protetti dalla firma
  - `Body_Hash`: Hash del corpo email nella firma
- **SPF_Results**:
  - `Result`: Risultato verifica SPF (pass/fail/softfail/neutral)
  - `IP_Authorized`: Se l'IP mittente è autorizzato dal record SPF
  - `SPF_Record`: Record SPF del dominio mittente

8.2.9 Sezione Risk_Assessment

Valutazione complessiva del rischio:

- **Overall_Risk_Score**: Punteggio rischio complessivo 0-10
- **Risk_Factors**: Array dei fattori di rischio identificati
- **Recommendation**: Raccomandazione finale di sicurezza
- **Confidence_Level**: Livello confidenza nell'analisi (0-100%)
- **Categories**: Categorie di minacce identificate (phishing, malware, spam)

9 Limitazioni e Considerazioni

9.1 Limitazioni Tecniche

Come ogni sistema che dipende da servizi esterni, l'Email Analyzer presenta alcune limitazioni intrinseche che devono essere considerate durante l'implementazione. I **rate limits** imposti dai provider di servizi di sicurezza rappresentano un vincolo fondamentale che influenza la velocità di elaborazione, specialmente quando si analizzano grandi volumi di email in tempi ristretti.

La **latenza** dell'analisi comprehensive è inevitabilmente superiore rispetto a sistemi più semplici, poiché la consultazione di servizi multipli richiede tempo aggiuntivo. Questa caratteristica deve essere bilanciata con i benefici in termini di accuratezza e completezza dell'analisi.

La **dipendenza dalla connettività di rete** rappresenta un altro aspetto critico: il sistema richiede connessione internet stabile per accedere ai servizi di investigazione esterni, limitandone l'utilizzo in ambienti completamente isolati. Infine, l'utilizzo di servizi multipli può potenzialmente aumentare il rischio di **falsi positivi**, richiedendo una calibrazione attenta degli algoritmi di risk scoring.

10 Conclusioni

Il progetto ha raggiunto tutti gli obiettivi prefissati, usando un'approccio alla sicurezza email attraverso la **diversificazione dei servizi** che ha ridotto la dipendenza da una singola piattaforma di controllo. Questa trasformazione ha portato a un **miglioramento della coverage** con un incremento del detection rate per tutte le categorie di minacce analizzate.

L'**ottimizzazione delle performance** attraverso l'implementazione di sistemi di caching e batch processing, ha reso il sistema più efficace. La **resilienza del sistema** è stata migliorata attraverso meccanismi di graceful degradation e error handling che garantiscono continuità operativa anche in condizioni avverse.

Infine, il miglioramento dell'**usabilità** attraverso output HTML interattivi e una CLI user-friendly ha reso il sistema accessibile a tutti.

11 Bibliografia

11.1 Documentazione Tecnica e Standard

- RFC 3501 - Internet Message Access Protocol (IMAP4) - https://tools.ietf.org/html/rfc3501
- RFC 6376 - DomainKeys Identified Mail (DKIM) - https://tools.ietf.org/html/rfc6376
- RFC 7208 - Sender Policy Framework (SPF) - https://tools.ietf.org/html/rfc7208
- RFC 7489 - Domain-based Message Authentication (DMARC) - https://tools.ietf.org/html/rfc7489

11.2 Strumenti di Test

- GoPhish - Piattaforma open source per simulazioni di phishing
  https://getgophish.com/

11.3 Risorse di Wikipedia

- Wikipedia - "Internet Message Access Protocol"
  https://it.wikipedia.org/wiki/Internet_Message_Access_Protocol
- Wikipedia - "Email Authentication"
  https://it.wikipedia.org/wiki/Email_authentication
- Wikipedia - "Phishing"
  https://it.wikipedia.org/wiki/Phishing
- Wikipedia - "DomainKeys Identified Mail"
  https://it.wikipedia.org/wiki/DomainKeys_Identified_Mail
- Wikipedia - "Sender Policy Framework"
  https://it.wikipedia.org/wiki/Sender_Policy_Framework
- Wikipedia - "DMARC"
  https://it.wikipedia.org/wiki/DMARC
- Wikipedia - "Email spoofing"
  https://it.wikipedia.org/wiki/Email_spoofing
- Wikipedia - "Malware"
  https://it.wikipedia.org/wiki/Malware
- Wikipedia - "Cryptographic hash function"
  https://it.wikipedia.org/wiki/Cryptographic_hash_function
- Wikipedia - "URL shortening"
  https://it.wikipedia.org/wiki/URL_shortening

12 Sitografia

12.1 Servizi di Sicurezza Utilizzati

- VirusTotal - Analisi multi-engine di file e URL
  https://www.virustotal.com/
- PhishTank - Database collaborativo di phishing
  https://www.phishtank.com/
- URLScan.io - Analisi comportamentale di siti web
  https://urlscan.io/
- MalwareBazaar - Database malware di abuse.ch
  https://bazaar.abuse.ch/
- AbuseIPDB - Database reputazione IP
  https://www.abuseipdb.com/

12.2 Documentazione e Tutorial Python

- Python Official Documentation
  https://docs.python.org/3/
- Real Python - Tutorial e Guide
  https://realpython.com/
- Stack Overflow - Community di sviluppatori
  https://stackoverflow.com/
- W3Schools Python Tutorial
  https://www.w3schools.com/python/

12.3 Forum e Community

- Reddit - r/cybersecurity, r/Python, r/netsec
  https://www.reddit.com/r/cybersecurity/
- GitHub - Repository di progetti open source
  https://github.com/