Aegis

Un framework per la condivisione sicura di immagini

Marzia De Maina, Chiara Sivieri

Alma Mater Studiorum - Università di Bologna

Gennaio 2026

Il Problema: la vulnerabilità dei contenuti

Il Contesto attuale

La diffusione non autorizzata di immagini private (*Revenge Porn*) o documenti sensibili è una minaccia critica. Le piattaforme attuali proteggono i dati **in transito** (HTTPS, E2EE), ma non offrono garanzie sul **contenuto decifrato**.

Il **"Gap Forense":** una volta che l'immagine appare sullo schermo del destinatario, essa esce dal controllo del mittente.

- Uno screenshot o una foto allo schermo bypassano la crittografia.
- Non esiste un legame indissolubile che provi *chi* ha fatto trapelare il file.

La Soluzione Aegis

Introdurre un livello di **watermarking invisibile** che sopravvive alla ridistribuzione, trasformando l'immagine stessa in una prova forense che identifica mittente e destinatario.

Scopo e Obiettivi del Progetto

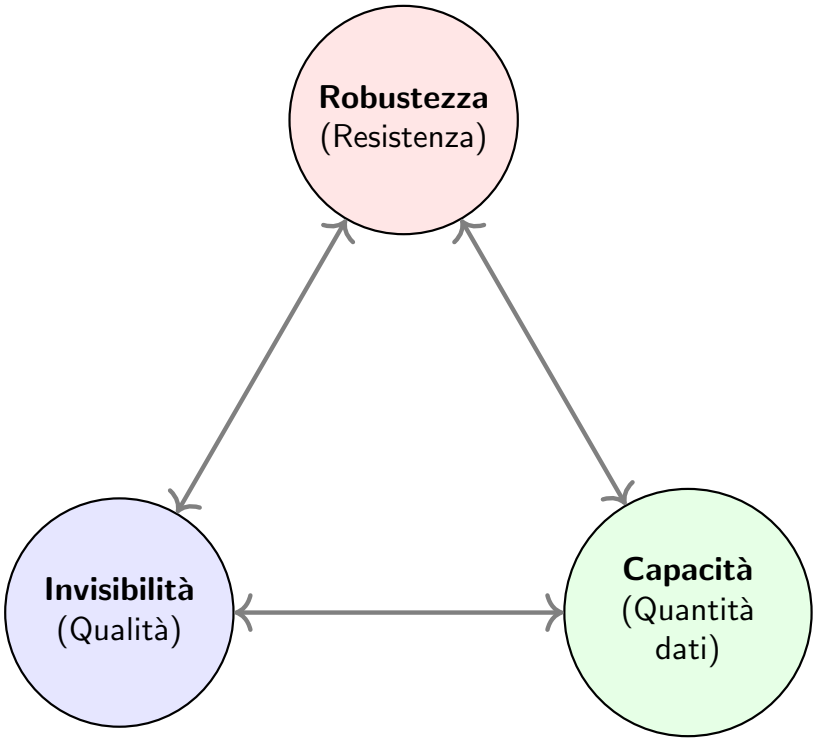
Aegis non vuole impedire la copia (impossibile per definizione), ma garantire la **non ripudiabilità** dell'azione.

- **Invisibilità (Impercepibilità):** il watermark non deve degradare la qualità dell'immagine né destare sospetti nell'osservatore.
- **Robustezza (Resilienza):** la firma deve rimanere leggibile anche dopo compressioni JPEG aggressive, ridimensionamenti, ritagli (cropping) e screenshot.
- **Tracciabilità (Attribution):** il sistema deve estrarre un payload contenente ID Mittente e ID Ricevente in modo univoco.

Metodologia: abbiamo implementato e confrontato tre algoritmi distinti (LSB, DCT, Spread Spectrum) per sviluppare una soluzione ibrida ottimale.

Il Triangolo Impossibile del Watermarking

In letteratura, ogni algoritmo deve bilanciare tre forze contrastanti. Migliorarne una peggiora inevitabilmente le altre.



- **Robustezza:** capacità di resistere agli attacchi (es. compressione).
- **Invisibilità:** assenza di artefatti visivi o rumore.
- **Capacità:** numero di bit inseribili (es. un ID lungo richiede più spazio).

Aegis sacrifica parzialmente la capacità per massimizzare robustezza e invisibilità.

LSB – Least Significant Bit

La tecnica più semplice: sostituzione del bit meno significativo dei pixel con i bit del messaggio.

Funzionamento

Un pixel RGB (8 bit per canale) ha valori da 0 a 255. Modificare l'ultimo bit cambia il valore di ± 1, variazione impercettibile all'occhio umano.

$$10110101 \rightarrow 10110100$$

Perché lo abbiamo scartato come soluzione principale? Sebbene offra alta capacità, è estremamente **fragile**.

- Basta una compressione JPEG (che arrotonda i valori dei pixel) per distruggere completamente l'informazione nascosta.
- Utile solo per verifica di integrità (checksum), non per tracciamento robusto.

DCT – Discrete Cosine Transform

Tecnica che opera nel dominio delle frequenze, simulando il comportamento della compressione JPEG standard.

Procedura di inserimento:

1. L'immagine viene divisa in blocchi 8×8.
2. Si applica la trasformata DCT per separare le basse frequenze (colore medio) dalle alte frequenze (dettagli fini/rumore).
3. Il watermark viene inserito sommando un valore ai **coefficienti di media frequenza**.

Vantaggio Strategico

Le medie frequenze sono percettivamente importanti, quindi gli algoritmi di compressione (come JPEG) tendono a preservarle. Questo rende il watermark **persistente** anche dopo il salvataggio dell'immagine.

SS – Spread Spectrum

Ispirato alle telecomunicazioni militari anti-interferenza. Il segnale (watermark) viene trattato come se fosse rumore bianco e distribuito sull'intera immagine.

Modello Matematico:

$$I_w(x, y) = I(x, y) + \alpha \cdot W(x, y)$$

Dove W è un pattern pseudo-casuale generato da una *chiave segreta*.

- **Sicurezza:** senza la chiave (seed), è matematicamente impossibile distinguere il watermark dal rumore termico del sensore fotografico.
- **Estrazione:** avviene calcolando la correlazione tra l'immagine sospetta e il pattern originale. Se la correlazione supera una soglia, la firma è presente.
- **Limiti:** richiede la chiave originale per la verifica (non-blind).

Confronto e Scelta Progettuale

Abbiamo analizzato i trade-off per determinare la tecnica migliore per Aegis:

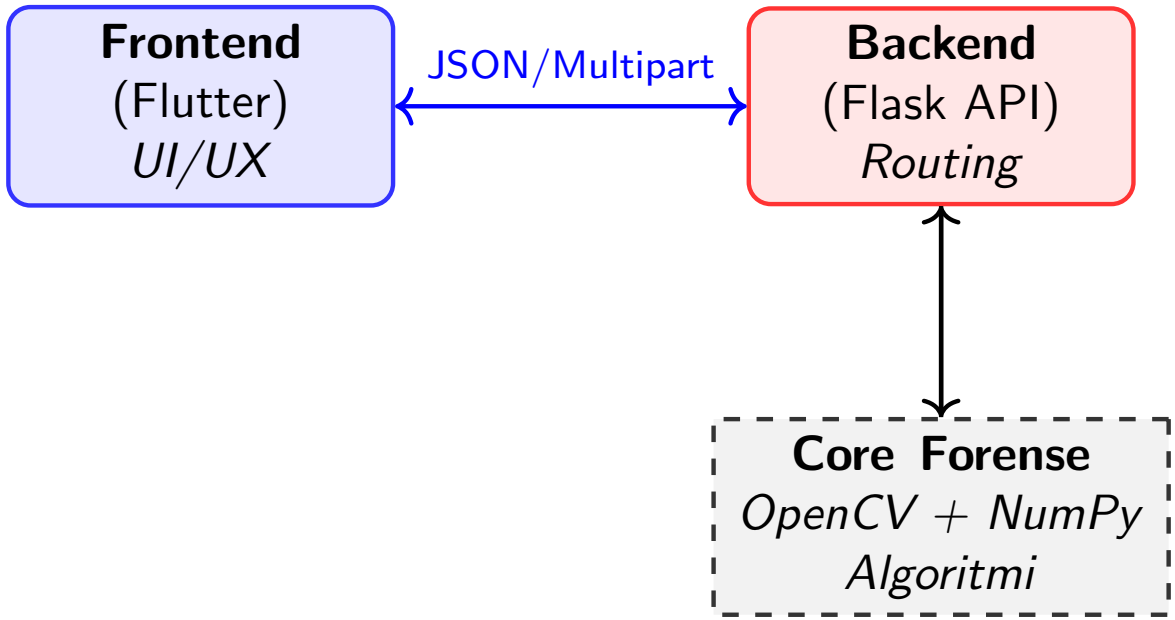
Tecnica	**Robustezza**	**Visibilità**	**Complessità**
LSB	Bassa	Eccellente	Bassa
DCT	Media	Buona	Media
Spread Spectrum	Alta	Eccellente	Alta

Soluzione Finale: Aegis Combo (Dual Layer)

Per ottenere il massimo da entrambi i mondi, abbiamo implementato un sistema a strati: **DCT** come base robusta + **LSB** come strato superficiale per il controllo integrità.

Architettura del Sistema

Il sistema implementa un'architettura Client-Server RESTful per separare l'interfaccia utente dalla logica computazionale pesante.



- **Frontend (Dart):** gestisce l'upload sicuro e la visualizzazione dei risultati. Cross-platform (Android, iOS).
- **Backend (Python):** esegue le trasformazioni matriciali (DCT/SVD) necessarie per il watermarking, gestendo le chiavi crittografiche lato server.

Dettaglio Algoritmo "Aegis Combo"

L'algoritmo implementa una strategia di **doppio embedding** (Dual Layer) per massimizzare le chance di recupero.

1. **Stratificazione (Embedding a Cascata):** il sistema applica prima il watermark **DCT** (nel dominio delle frequenze) per garantire la robustezza. Successivamente, sull'immagine risultante, applica il watermark **LSB**.
   - *Risultato:* l'immagine contiene due copie della firma a livelli di profondità diversi.

2. **Estrazione Adattiva (Fallback Mechanism):** In fase di analisi, l'algoritmo tenta l'approccio più veloce:
   - **Step 1:** controlla LSB. Se integro (immagine originale), estrae subito.
   - **Step 2:** se LSB è corrotto (compressione/attacco), attiva l'estrazione DCT per recuperare la firma "profonda".

3. **Optimization:** uso del **Canale Verde** per la DCT (miglior rapporto segnale/rumore sui sensori Bayer).

Risultato Forense e Verifica

Quando un'immagine sospetta viene analizzata, il backend restituisce un oggetto JSON strutturato che funge da evidenza digitale.

```
{
  "status": "DETECTED",
  "timestamp": "2026-01-30T10:45:00Z",
  "evidence": {
    "technique": "Combo_DCT_LSB",
    "sender_id": "USR_0001",
    "receiver_id": "USR_0002",
    "confidence_score": 0.98
  },
  "raw_watermark": "USR_0001<->USR_0002###AegisSig"
}
```

Analisi del Payload:

- **Evidence:** identifica univocamente chi ha inviato e chi ha ricevuto.
- **Confidence Score:** indica l'affidabilità statistica del rilevamento.
- **Raw Watermark:** la stringa grezza estratta dai bit dell'immagine.

Testing e Validazione Sperimentale

Il sistema è stato sottoposto a stress test per verificarne i limiti operativi.

Verifica Funzionale

Test del flusso completo (Handshake):

- Upload immagine mittente.
- Watermarking lato server.
- Download destinatario.
- Re-upload per verifica.

Stress Test Forense

- **Compressione JPEG (Q=70%):** firma recuperata (grazie al livello DCT).
- **Integrità File:** verifica immediata (grazie al livello LSB).
- **Cropping:** resistenza parziale basata sulla ridondanza dei blocchi DCT.

Sviluppi Futuri

Aegis è un Proof-of-Concept che pone le basi per una cybersecurity etica e attiva.

Roadmap verso "AIgis"

- **Autenticazione Forte:** implementazione di firme digitali asimmetriche (RSA) per impedire lo spoofing dell'ID utente.
- **AI Detection:** integrazione pre-invio di modelli ML per rilevare deepfake.
- **Web Crawler:** bot automatico per scansionare il web alla ricerca di immagini leakate, utilizzando la firma Aegis come "impronta digitale".

Conclusione: è possibile bilanciare privacy e tracciabilità in un'unica piattaforma.

DEMO

Grazie per l'attenzione

Q&A