Confidential Agentic Email Search: Evidence and Trust Boundaries

Stavros Kousidis

Master thesis · Published: 2026-10-25

Research question

Which retrieval and agent operations can safely cross the trust boundary of an encrypted email archive, and how do deployment choices affect confidentiality, evidence quality, and search performance?

Description

AI agents can reconstruct discussions across email archives. But **where may decryption, indexing, retrieval, and reasoning occur without disclosing confidential information?** This thesis examines evidence-grounded search over potentially OpenPGP-encrypted email, with KMail/Akonadi as an integration case study. Original messages remain the source of truth.

Scope and evaluation

Build a read-only prototype combining lexical and semantic **hybrid retrieval**, optional answer synthesis, and citations to original messages. Distinguish **index completeness**, relevance, and evidential support; expose missing or conflicting evidence.

Compare at least two trust architectures: **local** processing (including decryption and inference), **on-premises** processing (organizational infrastructure becomes trusted), and a **remote agent** restricted to a local policy-enforcing search broker.

Encryption boundary: Private keys remain inside the trusted cryptographic component. Decrypted messages, attachments, embeddings, and summaries stay local by default. Transferring any derived content requires explicit authorization and leakage analysis; even metadata and identifiers may disclose information. A remote agent cannot freely analyze confidential plaintext while learning nothing about it.

Threat-model prompt injection in emails, excessive agent privileges, unauthorized tool calls, and repeated-query leakage. Using synthetic or authorized test data, compare retrieval quality (Recall@k, nDCG), citation correctness, disclosure attempts, latency, and scalability.

Expected outcome

A threat model, policy-controlled retrieval prototype, and experimental comparison of achievable functionality and security guarantees. No new AI model or completed KMail plugin is required.

Mandatory open-source publication

Open-source publication is a condition for taking this topic. Before registration, students must agree to release thesis-developed code, tests, and reproducibility tools under an agreed, compatible OSI-approved license by project completion. Private development is permitted, but confidential messages, cryptographic keys, and credentials must never be published. **If you cannot accept this condition, please choose another topic.** See the publication and licensing policy for explanations of OSI, commercial use, and third-party licenses.

Starting points

- RFC 9580: OpenPGP.
- OWASP RAG security.

Related research

The Bachelor's topic investigates index integrity. Both topics are independently feasible.

Contact: stavros.kousidis@th-koeln.de
Code publication: open source by default; exceptions require prior agreement (policy).