

# **A Cryptographic and Threat Modeling Approach for Privacy-Compliant Digital Forensics in Healthcare Systems.**

**Submitted by ,**

**Jyothi Avulareddygar**

**Student ID : 2436065**

# Introduction

- Healthcare is a key target for cybercrime and attackers are increasingly targeting electronic health records (EHRs), clinical applications that are cloud-hosted, and networked medical devices to exfiltrate or extort sensitive patient data [1]. Digital evidence from these systems is now crucial to the understanding of how incidents happen, to mitigate threats and to provide due diligence to regulators and patients [2].
- At the same time, health care information is especially private, including details about health, treatment and life circumstances, and is therefore regulated with special laws, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union [3].
- These regimes require that any processing of protected health information (PHI) – including for security investigations – be based on the principles of data minimisation, purpose limitation and minimum necessary use and disclosure of patient records [4].

# Problem Statement

Patient data is a uniquely sensitive asset, and it is one of the most targeted assets by cybercriminals, especially in the healthcare sector. If a security incident does happen, investigators are caught in a delicate balancing act: they want to find the digital evidence but they have to do so within the legal and ethical boundaries of safeguarding patients' privacy whose records aren't relevant to the security incident being investigated. In reality, in the process of conducting a forensic investigation, this tension can lead to an inadvertent secondary privacy violation, and thus to a clash with existing regulatory frameworks like HIPAA and GDPR. To tackle this gap, a paradigm change in the approach to digital forensics is needed that focuses on recovering meaningful evidence of the incident while also protecting the privacy of non-participating patients.



# Aim & Objective

To develop and evaluate a complete Privacy-Preserving Healthcare Forensics (PPHF) framework that integrates cryptographic privacy preserving principles into the digital forensic investigation process in healthcare environment and to validate the fundamental ideas of the developed framework by implementing a proof-of-concept prototype

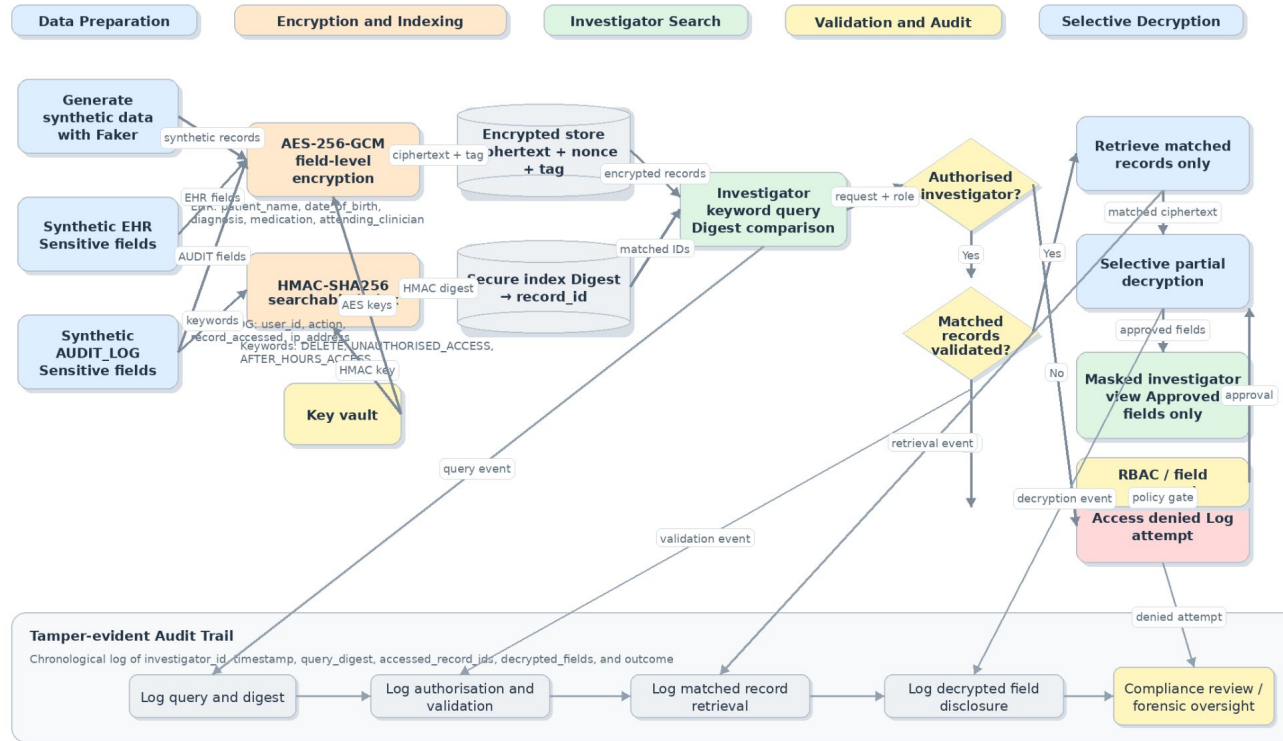
1. To examine the current state of the art in digital forensics and privacy-enhancing technologies, and the key challenges
2. To articulate the research philosophy, design approach and architectural principles that will guide the proposed PPHF framework.
3. To build a prototype of the core PPHF framework principles
4. To compare the proposed framework with known security and privacy threat models
5. To reflect on the trade-offs between forensic investigative capability and patient privacy protection, and to identify opportunities for future research and real-world deployment

# Methodology

1. In the project, synthetic data is produced using the Faker library.
2. AES-256-GCM encryption for sensitive fields in both EHR (i.e patient\_name, date\_of\_birth, diagnosis, medication, attending\_clinician) and AUDIT\_LOG (i.e user\_id, action, record\_accessed, ip\_address)
3. HMAC -SHA256 Searchable indexing is used to allow secure searching without exposing keywords (i.e DELETE, UNAUTHORISED\_ACCESS, AFTER\_HOURS\_ACCESS)
4. Investigator keyword query, Investigator search for suspicious activities without directly accessing all records
5. Retrieve only the matched records from encrypted store
6. Decrypt only approved fields and Masked Investigator View

# End to End WorkFlow

## Privacy-Preserving Healthcare Forensics Workflow with Audit Validation

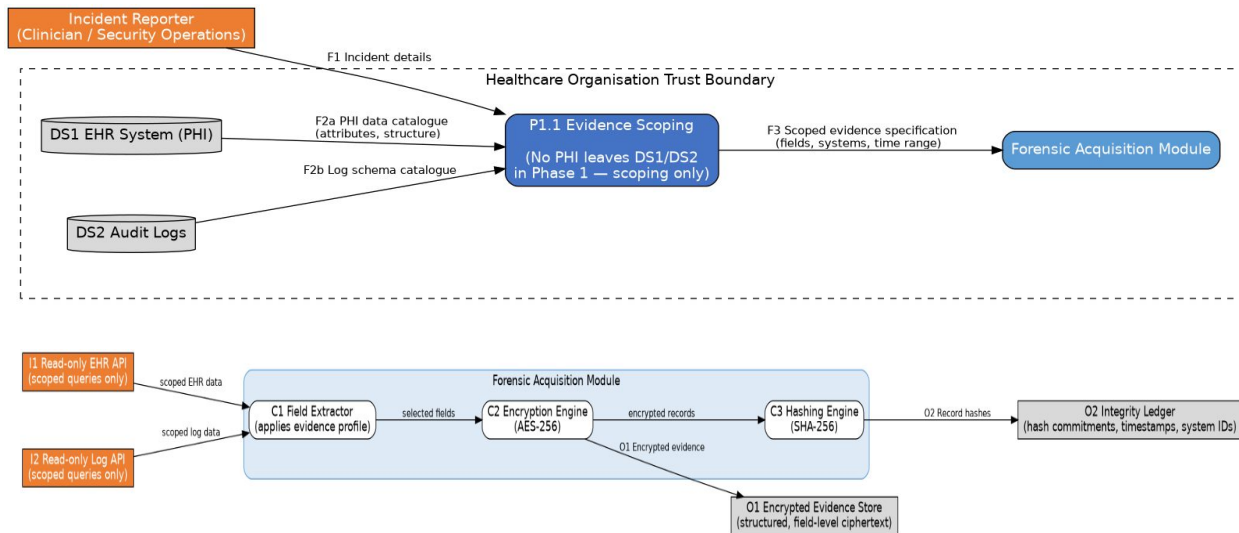


# Methodology

- **Research Philosophy - Positivism** : For the purpose of this research, positivistic approach is use as PPHF framework uses the field-level encryption that can be used to measure data minimisation, that Searchable Symmetric Encryption (SSE) - a way to enable privacy-preserving forensic search, to meets the requirements of HIPAA, GDPR, and the NIST Privacy Framework, are all empirically testable.
- **Research Approach - Deductive Reasoning** :The research involves moving from theory to conclusion, which means that the outcome of the proposed design should be tested to determine whether it meets conditions that are based on previous knowledge and understanding.
- **Research Strategy - Design Science Research** : Design Science Research (DSR) is a paradigm of research in which the main scholarly product is a useful, original design artefact, e.g., a framework, model, prototype, method, in the form of a designed entity, to address a class of real world problems, and the artefact is evaluated for utility and performance in a rigorous manner.

# Framework Design

- The PPHF framework outlines four phases of healthcare digital forensics, which are aligned with functions within NIST Privacy Framework and threat categories from LINDDUN.
- Phase 1 is the investigative authority's scope of the incident and the minimum set of data fields and systems needed to support the forensic objectives .
- Phase 2 focuses on the NIST Protect-P function to make sure that evidence collected from production systems is cryptographically protected at time of collection.
- Phase 3 provides the forensic investigators with the ability to analyse encrypted evidence without decrypting PHI fields.
- In Phase 4, selective partial decryption allows to further operationalize by making only those fields visible in plaintext that are explicitly authorised.



# Implementation

- The prototype is written in Python, and uses well developed public cryptographic libraries.
- Synthetic EHR records are created to mimic the format of actual clinical data, but without any patient information, thus avoiding ethical and regulatory limitations of using PHI.
- To ensure sensitive fields of EHRs are encrypted for confidentiality and integrity, AES-GCM is used at the field level, leaving less sensitive metadata in cleartext for investigation
- Each encrypted record is hashed with SHA-256 and added to a simple integrity ledger, to ensure the integrity of the chain-of-custody guarantees
- The prototype constructs a simple encrypted index that maps the hashed keywords to lists of log identifiers to demonstrate SSE-like behaviour without the need for an external (complex) library
- The prototype uses application logic to represent attributes and policies and then enforces them prior to the decryption

# Evaluation

- The evaluation design provides operationalization of the research questions in the form of explicit evaluation questions (EQs) that are associated with constructs in LINDDUN and NIST.
- LINDDUN categorizes the privacy threat analysis into 7 categories: Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of information, Unawareness and Non-compliance.
- The framework design is primarily oriented towards the Identify-P, Control-P and Protect-P functions, with the prototype showing specific behaviours in each of these functions.
  - Identify-P is realised in Phase 1, with investigators creating scoped evidence profiles
  - Control-P is realized in Phase 4 by attribute-based policy checks and selective partial decryption, which means that only actors who are authorized with the right attributes such as role, case and legal ones are able to decrypt certain fields.
  - Protect-P is implemented in Phases 2 and 3, with field encryption via AES, SSE-like queries and hash-chained audit logs .

## Evaluation (PPHF component vs NIST function)

| PPHF Component                      | NIST Function(s)      | Prototype Evidence                                                        |
|-------------------------------------|-----------------------|---------------------------------------------------------------------------|
| Phase 1: Evidence Scoping           | Identify-P            | Scoped profiles built over known EHR/log schemas; no full-database dumps. |
| Phase 2: Field-level AES + Hashing  | Protect-P             | AES-GCM encryption of PHI fields; SHA-256 hash chain for integrity.       |
| Phase 3: SSE Index and Search       | Protect-P, Control-P  | Token-based search limiting exposure to matching identifiers and records. |
| Phase 4: Attribute-based Decryption | Control-P             | is_authorized policy enforcement; partial decryption only.                |
| Audit Chain                         | Identify-P, Protect-P | Logged acquisition/search/decryption events with hash-linked records.     |

# Evaluation (LINDDUN)

| PPHF Phase                       | Linkability                                                    | Identifiability                                   |
|----------------------------------|----------------------------------------------------------------|---------------------------------------------------|
| Phase 1: Evidence Identification | Partially mitigated (scoped profiles, but identifiers present) | Residual (identifiers used in scoping)            |
| Phase 2: Encrypted Acquisition   | Mitigated for PHI (encrypted fields)                           | Strongly mitigated (identifying fields encrypted) |
| Phase 3: SSE Analysis            | Partially mitigated (encrypted tokens, pattern leakage)        | Mitigated (no direct identifiers exposed)         |
| Phase 4: Selective Decryption    | Residual, but bounded by policy                                | Residual, limited to authorised fields            |

The PPHF audit and chain-of-custody layer adds to non-repudiation, documenting hash-chained events of acquisition, so investigators can't credibly claim that they didn't access specific records or perform certain searches.

PPHF disclosure risk is systematically mitigated by encrypting PHI fields, avoiding full decryption during analysis .Unawareness is dealt with indirectly: while patients are not explicitly told to do so in the prototype, the fact that there are documented decryption policies and comprehensive audit logging, . The target of non-compliance is most direct: Phase 1 scoping and Phase 4 decryption policies are technical policies to enforce “minimum necessary” as per HIPAA guidance and GDPR data minimisation principle.

# Conclusions

This study found that there is a basic problem with the current practice of investigations in health care settings: the common practice of exposing medical records of uninvolved patients in security incident investigations. Consequently, a new investigative paradigm was devised to balance the competing interests of evidence gathering and privacy protection of patients. The proposed framework is effective in mitigating the risk of an unauthorised disclosure of information relating to non-involved patients, meets statutory privacy requirements and maintains the effectiveness of the investigative process. However, so far the solution has been tested only with simulated patient data. The framework still needs to be validated with an independent security test and with some pilot projects in clinical environments.

# References

1. Adedigba, A.P., Adewumi, A. and Aibinu, A.M. (2022) 'Securing electronic health records against insider threats: A supervised machine learning approach', *Smart Health*, 26, p. 100354. doi: 10.1016/j.smhl.2022.100354.
2. Al-Dhaqm, A. *et al.* (2022) 'An extended digital forensic readiness and maturity model', *Forensic Science International: Digital Investigation*, 41, p. 301379. doi: 10.1016/j.fsidi.2022.301379.
3. Al-Dhaqm, A. *et al.* (2024) 'Digital forensics readiness framework (DFRF) to secure database systems', *Engineering, Technology and Applied Science Research*, 14(2), pp. 13732–13740. doi: 10.48084/etasr.7116.
4. Al-Hawawreh, M. *et al.* (2024) 'Privacy preservation of electronic health records in the modern era: A systematic survey', *ACM Computing Surveys*, 56(10), pp. 1–36. doi: 10.1145/3653297.
5. Al-mugern, R. *et al.* (2024) 'A cloud forensics framework to identify, gather, and analyze cloud computing incidents', *Engineering, Technology and Applied Science Research*, 14(3), pp. 14483–14491. doi: 10.48084/etasr.7185.
6. Alexei, A. (2022) 'Implementing design science research method to develop a cyber security framework for HEIs in Moldova', in *Proceedings of the 11th International Conference on Electronics, Communications and Computing (IC|ECCO-2021)*. Chişinău: Technical University of Moldova, pp. 1–8.
7. Awolaye, J. *et al.* (2025) 'Forensic analysis frameworks for encrypted cloud storage investigations', *International Journal of Computer Applications*, 187(17), pp. 8–19. doi: 10.5120/ijca2025925241.
8. Carmi, L., Zohar, M. and Riva, G.M. (2023) 'The European General Data Protection Regulation (GDPR) in mHealth: Theoretical and practical aspects for practitioners use', *Medicine, Science and the Law*, 63(1), pp. 22–31. doi: 10.1177/00258024221118411.
9. Carboni, A. *et al.* (2023) 'Privacy by design in systems for assisted living, personalised care and wellbeing: A stakeholder analysis', *Frontiers in Digital Health*, 4, p. 934609. doi: 10.3389/fdgth.2022.934609.
10. Chaddad, A., Wu, Y. and Desrosiers, C. (2024) 'Federated learning for healthcare applications', *IEEE Internet of Things Journal*, 11(5), pp. 7339–7358. doi: 10.1109/JIOT.2023.3337017.