Privacy-Enhancing Technologies in Collaborative Healthcare Analysis

A Seminar Report

Submitted to the Bhivarabai Sawant Institute Of Technology and Research
in partial fulfillment of requirements for the award of degree

Bachelor of Engineering
in
Computer Engineering

Sonawane Anushka Prakash
TE-C, Roll No. 31

DEPARTMENT OF COMPUTER ENGINEERING
Bhivarabai Sawant Institute of Technology and ResearchWagholi, Pune

CERTIFICATE

This is to certify that the report entitled **Privacy-Enhancing Technologies in Collaborative Healthcare Analysis** submitted by **Sonawane Anushka Prakash (TE-C, Roll No. 31)**, to the Department of Computer Engineering in partial fulfillment of the B.E. degree in Computer Engineering is a bonafide record of the seminar work carried out by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

<table>
<tr><td>**Prof. Priyanka Patil**</td><td>**Prof. Sayali Suryawanshi**</td></tr>
<tr><td>(Seminar Guide)</td><td>(Seminar Coordinator)</td></tr>
<tr><td>Assistant Professor</td><td>Assistant Professor</td></tr>
<tr><td>Dept. of Computer Engineering</td><td>Dept. of Computer Engineering</td></tr>
<tr><td>BSITR, Wagholi, Pune</td><td>BSITR, Wagholi, Pune</td></tr>
</table>

Dr. G. M. Bhandari
Professor and Head, Dept. of Computer Engineering
BSITR, Wagholi, Pune

DECLARATION

I, **Sonawane Anushka Prakash**, hereby declare that the seminar report entitled **Privacy-Enhancing Technologies in Collaborative Healthcare Analysis**, submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Engineering of the Bhivarabai Sawant Institute of Technology and Research, Wagholi, is a bonafide work done by me under the supervision of **Prof. Priyanka Patil**.

This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data, idea, fact or source in my submission. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Wagholi   **Sonawane Anushka Prakash**

Abstract

Healthcare data is often fragmented across different institutions (hospitals, clinics, research centers), creating data silos. Privacy-enhancing technologies (PETs) play a funda mental role in collaborative healthcare analysis, enabling healthcare providers to improve care while protecting patient privacy. By providing a compliant framework for data sharing and research, PETs facilitate collaboration while adhering to stringent regulations like HIPAA and GDPR. This work conducts a comprehensive survey to investigate PETs in healthcare industry. It investigates the privacy requirements and challenges specific to healthcare, and the key enabling PETs are explored. A review of recent research trends that identify challenges, and AI related concerns is presented.

Acknowledgement

I take this opportunity to express my deepest sense of gratitude and sincere thanks to everyone who helped me to complete this work successfully. I express my sincere thanks to **Dr. G. M. Bhandari**, Head of Department, Computer Engineering, Bhivarabai Sawant Institute of Technology and Research, Wagholi, for providing all the necessary facilities and support. I would like to express my sincere gratitude to my seminar guide **Prof. Priyanka Patil** and coordinator **Prof. Sayli Suryawanshi**, Department of Computer Engineering, for her constant guidance and mentorship throughout the course. Finally, I thank my family and friends who contributed to the successful fulfilment of this seminar work.

Contents

Chapter 1

Introduction

Digitisation of health and patient data is significantly changing the clinical, operating, and business models which continue to impact the world of economy for the foreseeable future. In parallel, there is a significant increase in the volume of collected, stored, and processed data. According to The US International Data Corporation (IDC) prediction, that by 2025 the amount of data created will rise to around 163 zetta bytes (ZB) worldwide. Consequently, the vast number of connected devices and the increasing quantity of data fuelling an ever-growing number of applications, are collectively raising significant concerns in related security and privacy issues surrounding this data. While this change has improved patient care workflow and reduced costs, it also increases the probability of security and privacy breaches. In addition, one of the key barriers to widely adopting clinically-validated artificial intelligence (AI) applications is preserving patients' privacy.

Based on Statista , the healthcare industry is considered one of the most vulnerable to cyber-crime. In 2023, it remained the most targeted by cyber-attacks in the US, resulting in data compromises. The number of data compromise incidents in the US increased more than twice compared to 2022. In addition, between 2016 and 2022 the highest number of reported breached records was registered in 2022, totalling 51.4 million. These findings reveal a critical need for healthcare providers to adopt a more proactive and comprehensive cybersecurity strategy to protect against growing cyber threats.

In light of these issues, several techniques have emerged to mitigate privacy threats, known as PETs. These technologies that are designed to achieve data sharing and privacy preservation are gaining an expanding interest. In the past decade, there is a significant interest in using PETs technologies to enhance privacy in the healthcare industry. A comprehensive literature review was conducted to provide researchers with an overview of PETs in healthcare and identify research opportunities. It is based on data from three major academic databases (SCOPUS, IEEE, and ScienceDirect) illustrating the increasing adoption of PETs in the healthcare environment.

This survey addresses three guiding research questions:

Which key PETs are currently deployed in a healthcare system, and how do they function?

What privacy requirements and challenges define their application in this domain?

What barriers hinder their widespread adoption, particularly regarding data utility in collaborative analysis?

The main contributions can be summarized as:

A comprehensive literature review has been conducted to focus on recent research studies using PETs in healthcare systems and investigations of the privacy requirements and challenges in the healthcare industry.

This work investigates key enabling PETs, including federated learning, differential privacy, homomorphic encryption, synthetic data generation, multi-party computation (MPC), etc., and analyzed how they affect the data utility in collaborative analysis.

Key recent research trends in the protection of healthcare data analysis were addressed, specifically highlighting privacy protection schemes within AI models utilizing healthcare data, and their impact on data utility.

Chapter 2

Literature Review

Numerous comprehensive surveys have examined various aspects of Privacy-Enhancing Technologies (PETs) in different fields like AI and IoT. These surveys offer a detailed understanding of PETs' evolution and current trends. By summarizing key insights from these reviews, existing gaps are identified, which provide a basis for the contributions of this paper.

- **Categorization and Classification:** Cha et al. conducted a survey on PETs in IoT applications using a newly proposed categorization. The work in [2, 6] focused on classifying PETs into different classes. Kaaniche et al. classified PETs into three groups and eight categories based on the main entity involved in the privacy-preserving decision. These groups are user-side, server-side, and channel-side techniques.

- **Frameworks and Protocols:** Kunz et al. presented a framework for PETs application in IoT communications. It identified stakeholder and GDPR requirements and evaluated the framework's design based on these requirements to demonstrate its ability to support data minimization and data protection by design. In a similar context, Haddad proposed a privacy-preserving handover protocol using blockchain and Zero-Knowledge Proof (ZKP) to enhance privacy and security in 5G networks. Terhorst et al. proposed privacy evaluation protocols (PEPs) for assessing Soft-Biometric PETs.

- **Application-Specific Studies:** A study by Li et al. (2022) developed a privacy-preserving smart healthcare system for Alzheimer's disease (AD) detection using a differential privacy-based mechanism and a federated learning-based framework. The experimental results showed that the system, named ADDETECTOR, achieved high accuracy while ensuring strong security protection.

- **Challenges and Guidelines:** Garrido et al. studied PETs challenges and provided guidelines synthesized from expert interviews and a literature review focused on automotive use cases. Soykan et al. (2022) deeply analyzed PETs, including secure multiparty computation, homomorphic encryption, differential privacy, and confidential computing, in the context of collaborative machine learning. This study provided guidelines for selecting PETs for collaborative machine learning.

8

Chapter 3

Motivation

The core motivation for this study is the critical need to address the privacy and security challenges posed by the increasing digitization of healthcare data. The authors highlight the following key points:

[label=3.]

1. **The Problem of Data Silos and Privacy Risks:** Healthcare data is often fragmented across different institutions, creating "data silos" that prevent effective collaboration and analysis. At the same time, the vast volume of this data and its transfer through connected devices (like IoT) and cloud computing raise significant privacy and security concerns.

2. **Healthcare as a Primary Target for Cybercrime:** The document specifically identifies the healthcare industry as one of the most vulnerable to cyber-attacks. The significant increase in data compromises and breached patient records in recent years underscores the urgent need for a more comprehensive cybersecurity strategy.

3. **The Barrier to AI Adoption:** A key motivation is the challenge of integrating AI into healthcare. The authors note that preserving patient privacy is a major barrier to the widespread adoption of clinically-validated AI applications, even though these technologies have the potential to improve patient care.

4. **Addressing Gaps in Existing Research:** The authors also motivate their work by identifying a gap in the existing literature. While many surveys on Privacy-Enhancing Technologies (PETs) exist for various fields (such as IoT and general AI), there is a lack of a comprehensive, focused survey on the use of PETs specifically for collaborative healthcare analysis.

Therefore, the study's central motivation is to systematically investigate the application of PETs to address these specific challenges. The authors frame their research around three guiding questions, which reflect their primary goals:

[label=3.]

1. To identify which key PETs are currently used in healthcare and how they function.

2. To define the privacy requirements and challenges specific to this domain.

3. To identify the barriers that hinder the widespread adoption of PETs, particularly concerning the trade-off between privacy and data utility in collaborative analysis.

Chapter 4

Objective and Scope

Objective

The primary objective of this seminae report is to provide a comprehensive survey on the use of **Privacy-Enhancing Technologies (PETs) in the healthcare industry.**

[label=4.]

1. To conduct a comprehensive survey to investigate Privacy-Enhancing Technologies (PETs) in the healthcare industry.

2. To investigate the privacy requirements and challenges that are specific to the healthcare domain.

3. To identify and explore key enabling PETs, such as Federated Learning, Differential Privacy, and Homomorphic Encryption.

4. To present a review of recent research trends and their related challenges, particularly those concerning AI.

5. To investigate the barriers that hinder the widespread adoption of PETs, especially regarding the trade-off between privacy and data utility in collaborative analysis.

Scope

Which key PETs are currently deployed in healthcare and how do they function? The paper investigates enabling PETs such as Federated Learning, Differential Privacy, Homomorphic Encryption, and Anonymization, and analyzes how they affect data utility in collaborative analysis.

What privacy requirements and challenges are specific to this domain? The study investigates the unique privacy challenges and requirements within the healthcare field.

What barriers hinder their widespread adoption? The research explores the challenges that prevent the widespread use of PETs, particularly the trade-off between privacy protection and the utility of the data for collaborative analysis.

Additionally, the paper presents a review of recent research trends, identifying challenges and AI-related concerns.

Chapter 5

Methodology

5.1 Methodology: A Systematic Literature Review

This seminar report is based on a systematic and comprehensive literature review designed to examine the state of Privacy-Enhancing Technologies (PETs) within the context of collaborative healthcare analysis. The methodology provides a structured foundation for understanding how these technologies function, their specific application challenges, and their impact on data utility.

5.2 Research Design

The study adopts a descriptive and analytical design, focusing on synthesizing knowledge from a wide array of academic sources. It aims to provide a high-level overview of PETs in healthcare while also identifying key research gaps and future directions. Emphasis is placed on peer-reviewed sources from the past several years to ensure the findings are current and representative of the latest advancements in the field.

5.3 Data Collection

The data collection process was rigorous, centered on a systematic search across major academic databases.

[label=3.]

1. **Academic Literature:** The primary data sources were peer-reviewed articles from three major databases: SCOPUS, IEEE, and ScienceDirect. These platforms were chosen for their extensive coverage of research at the intersection of cryptography, computer science, and health informatics. The search was conducted using a combination of keywords, including 'privacy enhancing technologies', 'PETs in healthcare', and related terms, to identify relevant studies.

2. **Inclusion Criteria:** To ensure the review's relevance and currency, the search was limited to scholarly articles published between 2018 and 2024. The focus was on papers that applied PETs and offered empirical or insightful analysis related to their use in a healthcare context.

5.4 Data Analysis and Synthesis

The collected material was systematically organized and analyzed to address the core objectives of the report. The process involved identifying key themes and categorizing the findings into the following areas:

[label=4.]

1. **Key PETs:** An analysis of the function and application of primary technologies, such as Federated Learning, Differential Privacy, and Homomorphic Encryption.

2. **Privacy Requirements:** An investigation into the specific legal, ethical, and technical privacy challenges and requirements unique to the healthcare domain.

3. **Barriers to Adoption:** An examination of the factors hindering the widespread implementation of PETs, particularly the trade-off between data privacy and utility.

Key findings from each category were then synthesized to provide a comprehensive overview. The final interpretation of the literature aims to provide researchers with a clear understanding of the current landscape of PETs in healthcare and highlight areas for future research and development.

5.5 Validation

Systematic Approach: The validation is built into the systematic nature of the literature review itself. By methodically searching and selecting studies from reputable academic databases (SCOPUS, IEEE, ScienceDirect), the report ensures that its conclusions are based on credible, peer-reviewed research.

Focus on Peer-Reviewed Sources: The methodology explicitly prioritizes peer-reviewed studies published within a specific timeframe (2018–2024). This serves as a fundamental validation step, as peer review is the standard for ensuring the quality, accuracy, and reliability of academic work.

Cross-Verification of Findings: Although not explicitly stated as "cross-verification," the review's process of synthesizing insights from numerous comprehensive surveys effectively serves this purpose. By comparing and contrasting findings from different papers, the report confirms common themes and identifies research gaps, thereby validating the consensus and highlighting areas of disagreement in the field.

Empirical Insights: The methodology also focuses on selecting papers that provide "empirical insights". This means the report's conclusions are not just theoretical but are supported by real-world data or experimental results from the cited studies, adding another layer of validation.

Transparency and Traceability: The report provides a summary table of related works, including their contributions and limitations. This transparency allows the reader to trace the origins of the report's claims and understand the context and constraints of the studies that support them, which is a key part of academic validation.

Chapter 6

Conclusion

The digitization of healthcare data, while offering immense benefits for patient care and research, has introduced significant privacy and security challenges. The fragmented nature of this data and the increasing vulnerability of the healthcare industry to cyber-attacks necessitate a proactive approach to data protection. This report has demonstrated that Privacy-Enhancing Technologies (PETs) play a fundamental role in addressing these challenges by providing a compliant framework for secure data sharing and collaborative analysis.

This comprehensive survey has investigated the critical privacy requirements and challenges specific to the healthcare domain. It has provided an overview of key enabling PETs, including:

- **Federated Learning (FL):** Which allows for the training of machine learning models on decentralized data without exposing raw information.

- **Homomorphic Encryption (HE):** Which enables computations on encrypted data, preserving privacy throughout the analysis process.

- **Anonymization and Data Minimization:** Which are foundational principles and techniques for reducing the risk of re-identification.

While PETs offer a powerful solution, their widespread adoption is hindered by obstacles such as the trade-off between privacy and data utility, as well as the high computational and communication costs associated with some technologies. Addressing these challenges will require further research to enhance the performance and practical application of these technologies. Ultimately, the successful integration of PETs is crucial for building a secure, trustworthy, and collaborative digital healthcare ecosystem.

Chapter 7

References

Kaaniche, N.; Laurent, M.; Belguith, S. *Privacy enhancing technologies for solving the privacy-personalization paradox:* Taxonomy and survey. J. Netw. Comput. Appl. 2020, 171, 102807.

Kunz, I.; Stephanow, P.; Banse, C. *An Edge Framework for the Application of Privacy Enhancing Technologies in IoT Communications.* In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6.

Thapa, C.; Camtepe, S. *Precision health data: Requirements, challenges and existing techniques for data security and privacy.*Comput. Biol. Med. 2021, 129, 104130.

Xiong, J.; Chen, J.; Lin, J.; Jiao, D.; Liu, H. *Enhancing privacy-preserving machine learning with self-learnable activation functions in fully homomorphic encryption.* J. Inf. Secur. Appl. 2024, 86, 103887.

Li, J.; Meng, Y.; Ma, L.; Du, S.; Zhu, H.; Pei, Q.; Shen, X. *A Federated Learning Based Privacy-Preserving Smart Healthcare System.* IEEE Trans. Ind. Inform. 2022, 18, 2021–2031

Rai, H.M.; Shukla, K.K.; Tightiz, L.; Padmanaban, S. *Enhancing data security and privacy in energy applications: Integrating IoT and blockchain technologies.* Heliyon 2024, 10, e38917.