Privacy Preserving Data Fusion Scheme for Heterogeneous Networks in IoT Devices

MCS

By

Danial Gohar

A thesis submitted to the faculty of Information Security Department, Military College of Signals, National University of Sciences and Technology, Rawalpindi in partial fulfilment of the requirements for the degree of MS in System Engineering

August 2022

Thesis Acceptance Certificate

Certified that final copy of MS thesis written by Mr. **Danial Gohar**, student of **MSSE-05**, Course Reg. No. **00000274183**, of **Military College of Signals** has been vetted by undersigned, found complete in all respect as per NUST Statutes / Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for the award of MS / MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the student have been also incorporated in the said thesis.

Signature: _____

Supervisor: **Assoc Prof Dr. Mian Muhammad Waseem Iqbal**

Dated: _____

Signature (HoD): _____

Dated: _____

Signature (Dean): _____

Dated: _____

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Danial Gohar
August 2022

Dedication

This thesis is dedicated to my Father Muhammad Munawar, Family, Teachers, and Friends for their unconditional love, endless support, and continuous encouragement.

Acknowledgement

I would like to convey my gratitude to my supervisor, Assoc Prof Dr. Mian Muhammad Waseem Iqbal, for his supervision and constant support. His invaluable help of constructive comments and suggestions throughout the experimental and thesis work are major contributions to the success of this research. Also, I would thank my committee members; Brig Syed Amer Ahsan Gilani, PhD and Asst Prof Waleed Bin Shahid for their support.

I am also extremely thankful to my colleagues, PhD scholar Khuwaja Mansoor-ul-hassan, PhD scholar Mehmood-ul-Hassan, Phd scholar Aiman Sultan, and Ms. Ain-ul-Zia who helped me out during my thesis work.

Abstract

Data Fusion at edge computing plays an important role in IoT infrastructure and a lot of research has already been carried out in this domain on privacy preservation of homogeneous data fusion. However, there still remains a dire need to design a secure and lightweight privacy preserving scheme for heterogeneous data fusion which should be dynamic and adaptive in nature that can address the issues with authentication of the devices connected to CSP for communication purposes and for fused data communication while providing privacy preservation. This research focused to take data from multiple sensors *i.e.* heterogeneous data and then use data fusion techniques to accurately identify the action needed to be taken autonomously by the underlying machine. The main objective of this study is to put forward a secure and efficient scheme to overcome these prevailing issues. This thesis has presented a novel PPFHI scheme for heterogeneous data fusion in IoT devices that can efficiently balance privacy and trust assessment while requiring little overhead in terms of computation, communication, and storage to enable distributed data fusion across the e-healthcare sector. Additionally, we have provided in-depth theoretical research, and the findings have shown that the PPFHI scheme is better compared to state-of-the-art schemes in many ways, including the accuracy of fusion outcomes.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

CTA	Cloud Trusted Authority
e-GW	Electronic Gateway
HP	Healthcare Platform
e-HC	Electronic Healthcare
PPFHI	Privacy Preserving Data Fusion in E-Healthcare IoT Devices
MITM	Man in the middle attack

Symbols

Pr_{CTA}, Pu_{CTA}	CTA's Private and Public Keys
S_1, S_2,..., S_n	n alternative Reputation stage / threshold stage
T_1,T_2,...	Equal-length time intervals

$T_{\beta,1}$,..., $T_{\beta,k}$	time units in T_β of k equal-lengths
ζ	each time interval's length
σ	each time unit's length
Cs_γ^β, Ci_γ^β	Congruous and in congruous secret values with respect to S_γ and T_β
HP_i	*ith* Health Platform
Pr_{HP_i}, Pu_{HP_i}	Private and Public keys of health platform
$RV_{HP_i}^\beta$	Health platform reputation stage
$RS_{HP_i}^\beta$	Health platform reputation value
$TS_{HP_i}^\beta$	Health Platform customised threshold
$Q_{HP_i}^\beta$	HP_i Query to CTA for private information
$Re_{HP_i}^\beta$	HP_i response
$Enc_{Pu_{CTA}}$(*)	Asymmetric encryption with Pu_{CTA} on *
$Sign_{Pr_{CTA}}$(*)	Digital signature with Pr_{CTA} on *
$Enc_{Pu_{HP_i}}$(*)	Asymmetric encryption with Pu_{HP_i} on *
$Sign_{Pr_{HP_i}}$(*)	Digital signature with Pr_{HP_i} on *
$Dsig_{HP_i}^\beta$	Digital signatures in $Q_{HP_i}^\beta$
$Dsig_{T,HP_i}^\beta$	Digital signatures in $RV_{HP_i}^\beta$
$PR_{HP_i}^\beta$	HP_i private reputation level set in T_β

$PL_{HP_i}^{\beta}$	HP_i private threshold limit in T_β
$hash_T(*)$	CTA's Hash Function value
$N\upsilon_T^{\beta}$	CTA's Nonce value
$Pr_{HP_i,1}^{\beta}$	n ordered elements in $PR_{HP_i}^{\beta}$
$Pr_{HP_i,n}^{\beta}$	n ordered elements in $PR_{HP_i}^{\beta}$
$FD_{HP_i}^{\beta,\alpha}$	Formatted Data content generated by HP_i for T_β
HA_{HP}(*)	Hash function shared between Registered health platforms
$N\upsilon_{HP_i}^{\beta,\alpha}$	Nonce value set calculated by HP_i corresponding to $FD_{HP_i}^{\beta,\alpha}$
$HV_{HP_i}^{\beta,\alpha}$	Hash Value set calculated by HP_i corresponding to $FD_{HP_i}^{\beta,\alpha}$
$H\upsilon_{HP_i,1}^{\beta,\alpha}$,...,$H\upsilon_{HP_i,n}^{\beta,\alpha}$	n ordered element in $HV_{HP_i}^{\beta,\alpha}$
$GDt_{HP_i}^{\beta,\alpha}$	Data generated by HP_i in $T_{\beta.\alpha}$
$\overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}}$	Hash calculated by HP_i in $T_{\overline{\beta},\overline{\alpha}}$ corresponding to $GDt_{HP_i}^{\beta,\alpha}$

Chapter 1

Introduction

1.1 Overview

This chapter covers the brief introduction of the research work and problem formulation of the thesis. It also states the applications, aims and methodology adapted to complete the research work. Lastly, it presents the organization and outline of the thesis.

1.2 Cloud Computing

Considering the huge storage requirements of data these days, the sole solution lies in the usage of cloud computing for day-to-day tasks, thus making it a necessary requirement. Even with the addition of more cloud based companies in the market, the issues related to the security have not been addressed appropriately. Cloud Computing boasts many advantages like reliable backups, storage ease, convenience with easy accessibility and simple software / hardware maintenance [1]. Multi-tenant customers, are storing

crucially sensitive data on cloud. This suggests that the CSP and its administrators have taken over custody of the data. The confidentiality and integrity of data can be vulnerable to certain important threats since the owner of the data no longer has control over it. If the confidentiality of the data is violated, this could result in catastrophic damage. Due to these issues, the businesses are always concerned about their data being outsourced to the cloud and where it might be a potential victim of some sort of data breach [1, 2].

As reported by McAfee, about 3.1 million external attacks were carried out in 2020 on the cloud user accounts. Most of these attacks focused on the vulnerabilities, which involved stolen credentials, IOT, SQL injection attacks along with XSS and malicious file inclusion [3]. Another way to visualize the prevalent threat is being realizing that the medical record of patients, which is to be kept confidential at all cost but when outsourced to a cloud and becomes vulnerable to cloud attacks. Another such example would be the criminal records being utilized by the Law Enforcement Agencies.

Internet of Things (IoT) has been omnipresent in recent years due to the developments in sensor and computer technology and commercial uses from e-healthcare, smart farming, to autonomous vehicles. As a result, the number of network nodes connecting to the Internet is increasing rapidly, which then produces an enormous volume of data that must be processed and evaluated in a timely manner. The widespread use of cloud computing and the quick expansion of file sharing over the cloud have compelled academics working on newer ways to put their data on a cloud server with a reasonable level of trust. A survey [4] conducted in 2019 indicates that 150,000,000 estimated unique visitors visit just Dropbox in a month. These staggering stats are enough to justify the

criticality of the privacy and security of the clients. Figure 1.1 shows the approximate number of users that the most popular cloud service provider is handling.

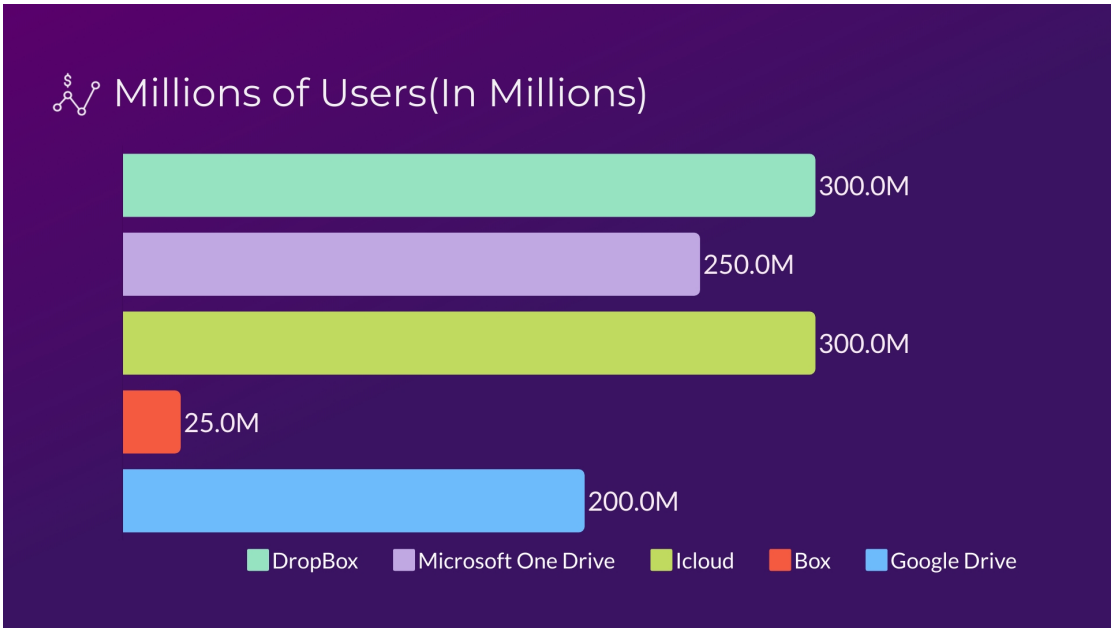

Figure 1.1: Number of Users per Cloud Service Provider

1.3 Research Motivation

According to Gartner [5], 2020 promises to be a big year for IoT industry. The market is expected to show a growth of 389$ billion, which is 21% more than 2019. The number of IoT endpoints is expected to grow to about 5.8 Billion. Gartner [1] has also mentioned that the two fastest growing industries which will benefit from this development are automotive and healthcare. The report highlights that healthcare is expected to see a growth rate of 29% in 2020. Figure 1.2 highlights the IoT Endpoint Market by different segments for the duration of year 2018 to 2020.

The combination of cloud and IoT poses numerous opportunities as well as compli-

[1] https://www.gartner.com/en/information-technology/insights/internet-of-things

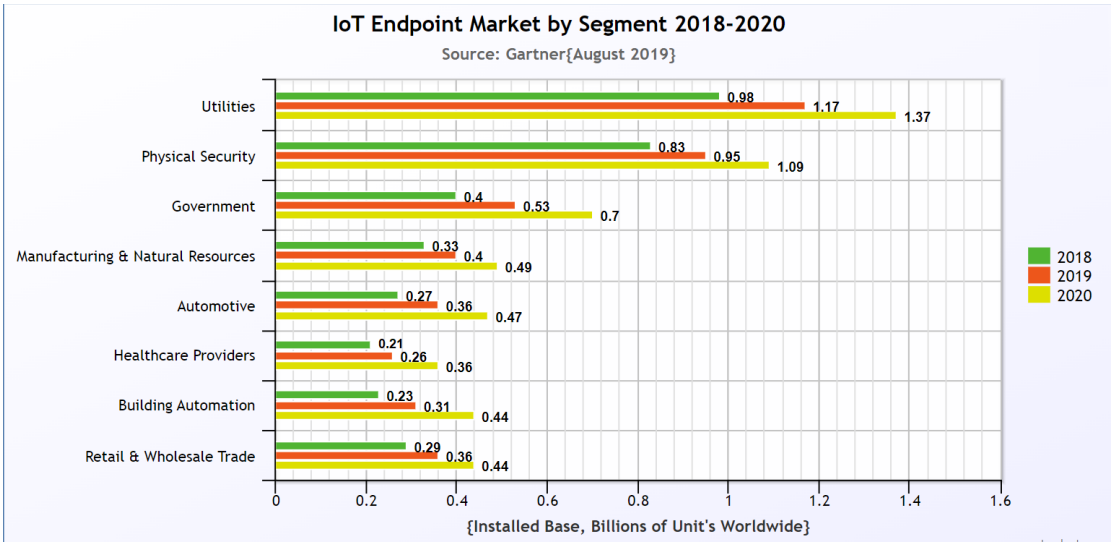

Figure 1.2: IoT Endpoint Market By Segment

cations. The cloud-IoT model represents a significant development in information and communication technologies (ICT). Sectors like Agricultural, e-healthcare, energy, smart cities, and environmental protection are all influenced by this paradigm. The cloud has made it possible to store and display data over the internet and operate devices just about anywhere in the world instantly [6]. The use of cloud in IoT has assisted in the creation and implementation of scalable systems. The two technologies support each other by providing a platform for progress. As seen, the incorporation has resulted in a variety of applications, the majority of which influence daily activities. IoTs will benefit from the cloud by widening its reach to work with real-world objects in a more complex and dispersed manner. In most situations, Cloud will serve as a bridge between devices and applications, abstracting all the complexities and functionalities used to execute the services [7].

The number of network nodes connecting to the Internet is increasing rapidly generating enormous volume of data which must be processed and evaluated in a timely manner. For handling such huge volume of data, a probable solution is linked to the data fusion.

Data fusion [8] has recently drawn significant interest in IoT as a probable solution for this data processing. Data fusion applies to the idea, methods, and resources utilized to integrate related knowledge from different sources to produce stronger choices or behavior would be feasible if all of these data sources were used collectively. All data must be sent from data sources to the cloud server in order to complete the cloud-based data fusion process leading to distressing concerns such as privacy-leaking, huge latency between data capturing and computing, and excessive consumption of bandwidth input.

1.4 Applications

There are numerous applications where the raw data from different IoT based sensors is being processed using machine learning and deep learning into meaningful information which can help to make decisions with the intervention of any human. The use in healthcare can be considered as one of such examples. The demand in this sector is to protect data privacy and access privacy. In [9], the authors have presented a survey of heterogeneous data fusion for healthcare monitoring.

Another such example is vehicular networks [10] where the whole autonomous structure is built from tiniest of decisions. The authors have provided a lightweight scheme which aims to solving the issue of road safety in driver-less cars.

Similarly, authors in [11], have proposed a multi-sensor data fusion technology for smart homes using the widely available wearable intelligent technologies, sensor fusion technology, and artificial intelligence.

1.5 Problem Statement

Data Fusion at edge computing plays a vital role in IoT infrastructure and a lot of research has already been carried out in this domain on privacy preservation of homogeneous data fusion, but it is a dire need to design a secure and lightweight privacy preserving scheme for heterogeneous data fusion which should be dynamic and adaptive in nature. To address the issues with authentication of the devices connected to CSP for communication purposes and for fused data communication while providing privacy preservation, there needs to be a mechanism which provides high security as well as performs efficiently. If a device from healthcare is establishing a connection with the cloud server than there must be a proper mechanism for authentication of the device while it attempts to establish a connection with another entity. Mainly, the focus is to take data from multiple sensors *i.e.* heterogeneous data and then use data fusion techniques to accurately identify the action needed to be taken autonomously by the underlying machine. The main focus of this study is to put forward a secure and efficient scheme to overcome these prevailing issues. This thesis explores the possibility of providing privacy preservation and authentication mechanism while using data fusion in the field of IoT.

1.6 Aims and Objectives

The proposed scheme aims for the following performance and security objectives:

- Comprehensive study of existing authentication schemes of data fusion in the

field of IoT.

- Proposal of an enhanced privacy-preserving scheme / framework for heterogeneous data fusion in IoT devices.

- Analysis of the proposed scheme in terms of security and efficiency.

1.7 Research Methodology

The research work starts from literature review of the existing techniques being used for heterogeneous IoT devices using data fusion. The literature review is done from various academic sources. This research then narrows down to the privacy preservation and authentication of IoT devices connecting to the cloud while listing down the drawbacks of existing schemes and formulates the problem. Then, it discusses the construction of scheme in detail and covers thoroughly the literature, design and implementation part of the thesis. A novel scheme is then presented for privacy preserving data fusion in e-healthcare IoT devices. A formal analysis is carried to show the efficiency in terms of security and performance of the proposed mechanism. In the end, a road map for future research areas are discussed and study is concluded.

1.8 Thesis Outline

In summary, the thesis breakdown is as follows:

- **Chapter 1: Introduction** presents the overview of healthcare sector and data

fusion. It also discusses some application areas, puts forward the problem statement, explains the research aims, methodology, and lastly, summarizes the research's objectives.

- **Chapter 2: Literature Review** discusses the existing literature on data fusion schemes for IoT devices along with their limitations. It also explains some of the applications of data fusion, the need for privacy preserving protocols and draws a comparative analysis of some existing schemes.

- **Chapter 3: Proposed Work** puts forward the network model, explains the threat model and design goals. Some formal definitions are presented. It also introduces a scheme for privacy preserving data fusion in e-healthcare IoT devices with detailed discussion of every phase.

- **Chapter 4: Performance Analysis** covers the formal analysis of the proposed scheme in terms of performance, security and efficiency. It presents a comparative analysis with different schemes with respect to storage, communication and computation overhead.

- **Chapter 5: Conclusion and Future Work** concludes the thesis and discusses the directions that can be explored in the future research.

Chapter 2

Literature Review

2.1 Overview

In this chapter, different existing schemes and protocols, their merits and demerits are explained in detail as well as comprehensive analysis of their security and performance is carried out. Different challenges that are being faced in a data fusion domain are discussed and various application areas are explored.

2.2 Related Work

The Internet of Things (IoT) seeks to constructs a world that allows for the interconnection and convergence of things in both the real and virtual realms. It is anticipated that as an evolving technology, it would link all devices and enable them to share data. To carry out sensing and perception of various data, several sensors/devices are typically deployed and because of the multi-source heterogeneity and vast volume of sensory

data, it is not practical to relay all data, as this leads to wastage of network bandwidth and system resources. As a result, data fusion has arisen as an effective strategy for extracting critical information from a vast volume of data collected to enhance data accuracy and promote decision-making.

Data fusion will, for instance, shrink the size and dimensions of data, minimize the volume of data, and extracts valuable information from it. It assists in removing data imperfections and overcoming the complicity of sensed data from various sensors. In simple words, Data fusion is the method of combining different data sources to provide information that is more reliable, accurate, and beneficial than any single data source. Based on the processing level at which fusion happens, data fusion processes are often graded as low, moderate, or strong.

A review of techniques associated with data fusion is provided in [12]. These are majorly based on the following:

1. Classification constructed on the relationships among various data sources.
2. Classification constructed on the nature of types of input / output data.
3. Classification constructed by the JDL defined different fusion levels.
4. Classification contracted on multiple architecture types.

IoT networks are susceptible to several issues related to connectivity and privacy. Earlier, the focus of the authors was towards homogeneous integration of data in IoT which in simple terms means that there was a single data source from where the raw data was being collected and later on turned in to something useful. But in recent times, the

idea of heterogeneous data fusion has been in discussion and seems to have the potential to be used in major industries as well. Some existing research works have been intending to give an outline of the data fusion efforts. Bostrom *et al.* [13] stated the definition as "Information fusion is the analysis of efficient methods for automatically or semi-automatically translating information from various sources and points in time into a representation that provides effective support for human or automated decision making". This fusion of data helps in cutting the size and proportions of data, reduce the volume of data, and derive valuable information from it. It assists in removing data imperfections and overcoming the complicity of sensed data from various sensors. Even with the advancement in the field of data fusion in IoT, there are still some major challenges that question the practicality of this system.

There are some papers in which authors have provided an overview about the efforts in the field of data fusion. In [14], Lee et al. published a review on the fusion techniques but the limitation of this research work lies in the specifics because there is no discussion on the design of these techniques for the IoT big data platform. In [15], the authors have focused on the data fusion for smart environments. Although, their research seems promising but the fact that they have not discussed about security and privacy issues in IoT raises concerns. The authors in [16, 17] have investigated various methods of data fusion.

2.3 Applications of Data Fusion

Considering the applications of data fusion, the authors in [18–20] have concluded that IoT enables the different types of objects to be sensed and even controlled without the intervention of human and this offers a great opportunity towards integration of this technology with the existing network systems. This would help in providing accuracy, high efficiency, and economic benefits. With the development in this field, the application areas keep on growing.

2.4 Privacy Preservation

Privacy preservation has come out to be primer before the implementation of this technology in the real world. A strong guarantee is necessary to not only improve the accuracy of the resultant fused data but also to provide the users with sense of security about their data [21, 22]. A spate of privacy-preserving techniques for achieving secrecy and untraceability in cooperative vehicular safety (CVS) applications have been developed in recent years. For fog computing aided CVS applications, the authors in [23] have developed a hierarchical pseudonym management system that protects privacy. When compared to earlier techniques, this scheme can significantly improve vehicle location privacy while also lowering communication overhead. Xu et al. [24] developed a novel privacy preserving data aggregation scheme capable of combining the data classification while preserving the privacy of vehicular sensing networks. The scheme is resistant against sensing data link attack and is able to provide efficiency, accuracy, and scalability. Despite the fact that the preceding schemes contain many outstanding concepts,

using them with trust evaluation techniques in CVS applications is still not possible.

2.4.1 Privacy Preservation in Vehicular Networks

In CVS applications, privacy protection and trust evaluation have opposing needs, and a suitable balance between them is required [25]. Only a few privacy-preserving trust evaluation systems for cooperative vehicular safety applications have been presented in recent years, group signature, leveraging pseudonym, threshold cryptography, partially blind signing, blockchain technologies and homomorphic cryptography [26, 27]. Several authors [25, 28, 29] pointed out that it is easier for an attacker to link each data provider's reputation score in reputation certificate-based trust evaluation schemes. Therefore, the authors carried out the conversion of the specific reputation thresholds to a some fuzzy threshold levels, but the schemes do not achieve strong privacy preservation because the adversary can still link each vehicle's data via the vehicle's threshold. The authors in [30] provided a BTMPP scheme to counter the drawbacks identified earlier. The authors suggest that aggregating the Bloom Filter (BF)-based Private Set Intersection (PSI) and reputation certificates is able to provide the users with a strong guarantee about trust evaluation and privacy preservation but at a cost of computational, storage, and communication overheads and complexities.

2.4.2 Privacy Preservation in Healthcare

Considering the privacy preservation in the field of healthcare, several authors have presented novel schemes which can be implemented in the real-world scenarios. Wang et

al. [31] has presented a forward privacy preservation for IoT-based healthcare systems. The proposed model uses Searchable Encryption [32] technique with forward privacy. The search query in this scheme is designed in such a way that it becomes difficult for the adversary to distinguish between data and also renders it very difficult to find a relationship which could help in breaking the system. The authors have also provided a formal security analysis to prove that the scheme does provide forward secrecy.

In [33], the authors have presented a chaos-based encryption system for the privacy protection of patient's data. The system works by using a probabilistic cryptosystem for hiding the medical key-frames which are extracted from wireless capsule endoscopy procedure. Leakage of information is contained against various attacks. Only authorized users have the option to decrypt the encrypted data of a patient. The authors claim that their system is providing excellent performance compared to existing systems.

Another important work was done by Wei et al. [34]. Their cryptosystem was able to hide the human face by blurring certain regions for the use in multimedia social networks and only authorized users were able to blur those regions.

A relatively newer approach towards privacy preservation is the use of blockchain for healthcare in IoT. The authors in [35] have used this technology to propose a solution which is able to protect the health related IoT data involving a security mechanism which is capable of protecting the privacy and also provide data integrity of a patient using blockchain.

Recently Xin Su *et al.* [50] also has addressed privacy leakage problems through the data fusion process and has created a centralized data fusion system with integrated

K-anonymous and non-interactive differential privacy applications as a standard for privacy. Under the protection of privacy, a multiparty data fusion algorithm is introduced. This approach focuses on Attribute Security. The authors introduced the idea of participation in the phase of incorporation, together with the real condition of redundancy, to prevent some sort of disruptive activities. To achieve better data protection and privacy, They seeks to ensure the secure and effective processing of distributed data within the DaaS architecture and the Shared Data Fusion. The authors are anticipating and resisting a class of malicious behaviors, that is, by maliciously raising the attribute score to minimize the probability of private data and the correctness of the experimental algorithm as the authors demonstrated that their algorithm is effective, but in malicious behavior monitoring misjudgment rate is high and the focus of this paper is only under the semi-honest model and if we want such an algorithm to run under a malicious framework then a better monitoring and retribution model should be designed to address other suspicious activities and promote safety. Also, this model is still vulnerable to internal attacks, and to resist such attacks a privacy framework such as m-privacy should be introduced in the algorithm.

2.5 Data Fusion

Data fusion is defined as an amalgamation of multiple sources to get meaningful information. This simply means higher quality, less expensive, and more useful information. In recent years, data fusion in the IoT has received a lot of attention. However, some problems and difficulties exist, such as data leakage and power usage. IoT allows a large

range of sensors and computers to communicate in real time and make our lives easier. By 2025, the number of sensors is projected to reach upto approximately 50 billion. Various devices are used to perceive/collect valuable data and, through data fusion and analytics offer a deeper view of the surroundings. knowledge-based methods, evidence reasoning methods and probability-based methods are the three basic categories of data fusion methods. Data Fusion can be categorized as follows:

- Probability based methods (PBM)
- Evidence reasoning methods (EBM)
- Knowledge-based methods (KBM)

Knowledge-based approach (KBM) allows the fusion center to extract information from imprecise big data without having to obtain density or distribution functions.

Intelligent aggregation methods and machine learning are example of knowledge-based methods.

Evidence reasoning method (EBM) instigated the concepts of belief and plausibility to reflect ambiguity in the real world and allow inference in dynamic contexts, where belief reflects the degree of belief with which a specific piece of evidence supporting a particular event and plausibility refers to the degree of belief with which a particular piece of evidence fails to contradict a particular event. In addition, it implements a mass function to reflect conviction distribution. It does, however, have a difficult time calculating mass functions, which limits its implementations.

To deal with data imperfection, the density function and probability distribution was

introduced, which can articulate the dependence among random variables. Bayesian inference, belief propagation, state-space models and Markov models are some common examples of probability-based data fusion methods.

Data fusion approaches have been extensively used in multi-sensory settings to combine and aggregate data from many sensors, nonetheless, similar techniques may also be used in other areas, such as text processing. By combining data from numerous distant sources, data fusion in multi-sensory settings aims to reduce detection error probability and increase dependability. Three categories of data fusion approaches have been established that are not mutually exclusive:

1. Data association
2. State estimation
3. Decision fusion

The purpose of this study, rather than providing a comprehensive evaluation of all the research, is to highlight the crucial processes involved in the data fusion framework and to look at the most popular approaches for each stage due to the vast amount of published papers on data fusion.

2.5.1 Classification of Data Fusion Techniques

Data fusion is a multi-disciplinary area which includes several fields. We classify the data fusion techniques based on the relationship among different data sources. A proposed classification criteria based on the relationships of the data sources is provide by

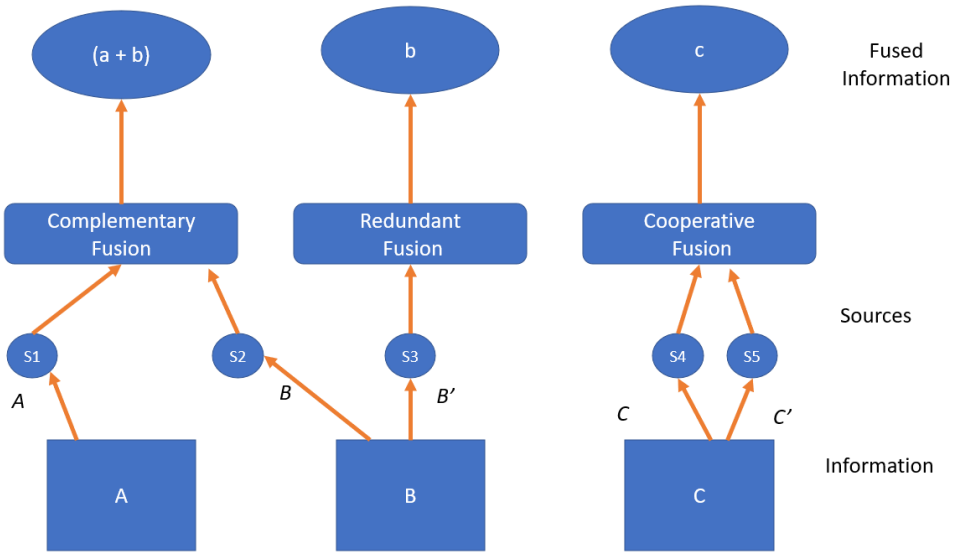

Figure 2.1: Categorization of Data Fusion Techniques

Durrant-Whyte [36]. Figure 2.1 shows various techniques for Data Fusion.

The criteria include:

- **Complementary:** The information given by the input sources reflects various aspects of the scene and may thus be utilized to generate more comprehensive global information.
- **Redundant:** Two or multiple input sources give knowledge about similar objective and may be combined to enhance certainty.
- **Cooperative:** The supplied data is integrated to create new data that is usually more complicated than the earlier (original) data.

2.5.2 Classification Model of Dasarathy

Dasarathy's classification model [51] is among the most well-known data fusion classification systems as shown in figure 2.2. It consists of the following elements:

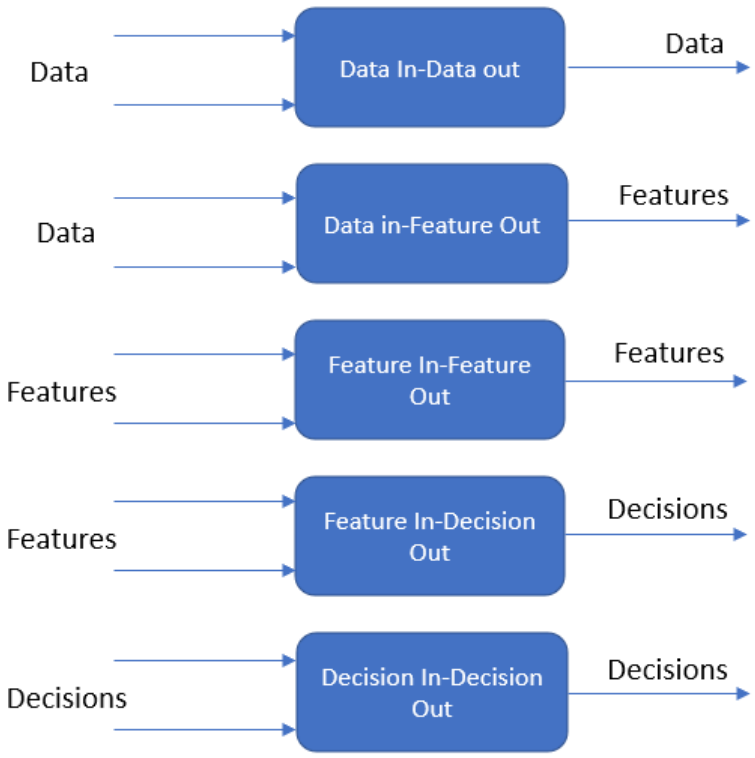

Figure 2.2: Classification Model of Dasarathy

- **Data in-Data Out:** The most fundamental or basic data fusion strategy that is taken into account in categorization. Raw data is input and output in this form of data fusion process, and the outcomes are usually more dependable or accurate. At this level, data fusion takes place as soon as the data from the sensors is collected. At this level, the algorithms are constructed on signal and image processing methods.

- **Data In-Feature Out:** Technique uses raw data from sources to derive attributes or traits that define an object in the environment at this level.

- **Feature In-Feature Out:** All the data *i.e.* input / output of the data fusion procedure are features at this level. At out-turn, the data fusion process focuses on a group of features in order to improve, modify, or create new ones. Feature fusion,

information fusion, symbolic fusion and intermediate level fusion are all terms used to describe this phenomenon.

- **Feature In-Decision Out:** This level receives a group of features as input and outputs a set of choices. This category of classification includes the majority of classification systems that conclude based on sensor inputs.

- **Decision In-Decision Out:** Decision fusion is another name for this form of categorization. It combines input judgments to produce better or new ones.

Dasarathy's categorization makes a significant addition by specifying the abstraction level as an input or output, so giving a framework for categorizing various approaches or techniques.

2.5.3 Classification Based on Abstraction Levels

This type of classification was provided by Luo *et al.* [37] and has following levels:

- **Signal Level:** It addresses signals taken from multiple sensors.

- **Pixel Level:** It works at image level and provides improvement in image processing related tasks.

- **Characteristic:** It makes use of characteristics collected from pictures or signals

- **Symbol:** The information is presented as symbols at this level, which is sometimes referred to as the decision level.

There are other data fusion classification but we restrict to the above mentioned as others are out of the scope of this study.

2.6 Data Fusion in IoT Healthcare Systems

IoT is a cutting-edge technical advancement that will make every city smart. IoT has recently become one of the top technologies all around the world. This technology's success originates from its diversified character, since it combines multiple heterogeneous systems to function together. This varied character resulted in several difficult challenges. Aside from all of these elements, data fusion is one of the most significant advancements in any autonomous system, because it increases system functioning by adopting fusion algorithms [38]. IoT enables the construction of smart spaces by transforming current surroundings into sensor-rich data-centric cyber-physical systems with an increasing level of automation, resulting in Industry 4.0. This trend, when implemented in commercial / industrial contexts, is altering many parts of our daily lives, including how individuals access and receive healthcare services. As we mshift towards Healthcare Industry 4.0, the underlying IoT infrastructure of Smart Healthcare spaces continue growing and tend to become more complex, marking it critical to ensure that proper processing of massive amounts of collected data is carried out to provide crucial input and decisions in accordance with existing requirements [39].

Intelligent, low-power, wireless networking medical equipment form the core of Smart Healthcare. These devices continuously monitor, process, collect, and safeguard weight, body position, sleep quality, movement, body temperature, blood oxygen saturation,

blood pressure, heart rate, exhaustion levels, blood oxygen, and some other bio-metric data. This given rise to the a relatively newer technology in the field of medicine, termed as Internet of Medical Things (IoMT) [40].

The use of sensors in various equipment e-g wheelchairs, beds, and ventilator etc makes this a "medical things". The staff in a hospital is able to view and receive the valuable bio-metric information on their personal computers or mobile devices remotely using a wired or a wireless connection and can also perform their duties. By using this newer approach, the doctors can take decisions immediately. The information is received from multiple sensors because a single sensor cannot provide enough conclusive evidence for any particular case. Therefore, a mechanism is needed to take the raw data from these sensors and turn them into something useful which can ease the processes in the medical field. This is where the concept of data fusion comes into play. The raw sensors data in the field of medical is of heterogeneous nature. Data fusion helps to make efficient and timely decisions based on the data collected from these sensors. This helps to pin point the exact cause, effect, and treatment of a patient.

In [41], the authors have proposed a privacy enhancing technique termed as Data Fusion Strategy (PDFS). The scheme has four components which include:

1. Homomorphic Encryption based data fusion
2. Contract design
3. Task completion assessment
4. Sensitive task classification

The authors claim that under COVID-19 application environments based on IoMT, their scheme is able to provide better privacy protection for data fusion.

In [42], the authors have proposed another privacy preservation and incentive-based data fusion technique for the implementation of fair incentives and privacy security of patients in the process of health data collection.

For the device oriented anonymous privacy protection in fog-aided IoMT, Guan *et al.* [43] has presented a privacy protection and authentication-based data fusion scheme.

2.7 Privacy Preservation and Data Fusion in IoT

Data Fusion deals with sensitive user data and once the data is fused it becomes more sensitive in nature. Data fusion at edge is vulnerable to attacks as discussed above, we need privacy preservation techniques to safeguard user location, preference and sensitive fused data which is valuable to network and servers. Different techniques for Privacy Preservation are as follows:

- Differential Privacy
- K - Anonymity
- Homographic Encryption for access control

Wang *et al.* [52] reviewed data fusion in CPSS systems. After analyzing they proposed that tensors should be used to represent CPSS data fusion as starting from CPSS definition to different data fusion systems are reviewed and explained to achieve a better

understanding. Furthermore, the authors proposed a series of CPSS device tensor-based fusion methods. The architecture of data fusion mechanisms is also examined and a suggestion for a thorough data fusion mechanism for CPSS is provided because all existing fusion methods are CP, CS, or SP-specific, and lack of standardization, stable, effective, and productive fusion process for CPSS. Though it is an extensive review in terms of CPSSs and data fusion, but Privacy and security are not discussed, and the energy consumption is still to be considered in the future.

Costel *et al.* [44] proposed two data fusion methods for the SPIDER peer-to-peer overlay network to concatenate the way data is transmitted as Client-server approach are obsolete as intelligent objects created by sensors are prone to malfunction the focus of the author is on ring-based and chain-based data fusion, which is evaluated for efficiency and fault-tolerance concerning the size of the overlay network. Two case scenarios are discussed and evaluated about the proposed algorithms and achieved the experimental results which prove two important aspects. First, local fault recovery is easy, and the second is that the ring-based fusion method is the efficient one. Overall, it's a good approach but it does not support heavy network load and does not work in crowded case scenarios and to overcome this problem fusion methods must be improved.

Yang *et al.* [45] addressed the issue of transferring fused data to the cloud server for fusion, which erases concerns such as privacy-leaking, excessive bandwidth consumption, and high latency, thus Offer the idea of edge temporal data fusion by way of an algorithm architecture, "GPTDF," that functions at the edge of delivering online sequential prediction service. This algorithm provides real-time forecasts which are more effective and reliable based on their demonstrations used on archived traffic data sets

from the Caltrans (PeMS). By experimenting it is validated that GPTDF provides real-time predictions that are more effective and reliable at the network edge. The authors simply consider the fusion of homogeneous data sets at the edge, while heterogeneous temporal data is not addressed. The fusion method can rely on implementations, as well as how background knowledge such as the edge server relative positions in the fusion phase is still a question.

The use of data fusion in intelligent transportation was primarily addressed by Faouzi *et al.* [46]. Even though there are several survey articles on multisensory data fusion, none of them offer a detailed analysis of IoT data fusion.

Existing practices to IoT data fusion and handling have depended heavily on a Cloud system, in which gathered data (in raw form) by edge sensors is sent to a Cloud that acts as the primary processing facility. However, according to [47], such a vertical off-loading model continues to neglect or underrate the (increasing) computing capability of edge devices, which are required to allow data analytics and processing, including data fusion. As a result, rather of being conducted on the spot, certain very simple data fusion procedures are shifted to a Cloud server through a potentially crowded public network, causing a number of issues. For starters, not processing easy jobs locally increases average response time greatly owing to network latency and limited capacity. Second, transferring potentially sensitive data via a public network raises security issues, while employing extra data protection techniques to address the issue increases technical and network effort. In light of the foregoing, the authors suggest a tiered automated data fusion structure for Smart Healthcare settings in which individual components perform data fusion across several data sources based on contextual information and compu-

tational capabilities. Lower-level components perform limited data fusion and convey aggregated information to higher-level elements, which can then move freshly statistics further up the hierarchy after fusing the collected information. The suggested framework is based on Complex Event Processing technology, which tries to find complex event trends in a sequence of atomic events and might be used to implement data fusion in remote IoT networks [48, 49].

2.8 Comparative Analysis

A comparative analysis of different schemes along with their limitations is given in the table 2.1 below:

Table 2.1: Data Fusion Authentication schemes for IoT Devices

Scheme	**Based on**	**Salient Features**	**Limitations**
Xin Su *et al.* [50]	centralized data fusion system with integrated K-anonymous and attribute based mechanism	the scheme is anticipating and resisting a class of malicious behaviors raising the attribute score to minimize the probability of private data and the correctness of the experimental algorithm	the scheme is still vulnerable to internal attacks with focus on semi-honest model

Costel *et al.* [44]	SPIDER peer-to-peer overlay network, ring-based and chain-based data fusion	two case scenarios are discussed and evaluated about the proposed algorithms and achieved the experimental results which prove two important aspects. First, local fault recovery is easy, and the second is that the ring-based fusion method is the efficient one	doesn't work in crowded case scenarios, limited dataset and vulnerable against privacy attacks
Yang *et al.* [45]	edge temporal data fusion by way of an algorithm architecture, "GPTDF," that functions at the edge of delivering online sequential prediction service	provides real-time forecasts which are more effective and reliable based on their demonstrations used on archived traffic data sets from the Caltrans (PeMS), GPTDF provides real-time predictions that are more effective and reliable at the network edge	only the fusion of homogeneous data sets at the edge are considered, no mention of heterogeneous temporal data, vulnerable against privacy attacks

Wang *et al.* [52]	Cyber physical social system (CPSS) device tensor-based fusion methods	a suggestion for a thorough data fusion mechanism for CPSS is provided because all existing fusion methods are CP, CS, or SP-specific, and lack of standardization, stable, effective, and productive fusion process for CPSS	no mention or analysis with respect to security or privacy
End of Table			

2.9 Summary

This chapter discussed some basic terminologies for Data Fusion in detail and some classification techniques for data fusion. It presented various areas of applications and their existing works with their merits as well as demerits. Chapter 3 will present the proposed work in electronic health care domain.

Chapter 3

Proposed Work

3.1 Overview

In this chapter, we will describe the network model for framework for heterogeneous data fusion in IoT devices in healthcare sector. A detailed description of entities and their communication flow will be presented for better understanding of data fusion in e-healthcare. Security goals and assumptions will be discussed. We will present our proposed mutual authentication protocol with all phases discussed in detail. The research includes the following contributions:

- Comprehensive study of existing authentication schemes of data fusion in the field of IoT.
- Proposal of an enhanced privacy-preserving scheme / framework for heterogeneous data fusion in IoT devices.
- Analysis of the proposed scheme in terms of security and efficiency.

A formal analysis is discussed in chapter 4 in terms of security, performance and efficiency.

3.2 System Model

The healthcare framework of PPFHI technique is discussed in this section.

3.2.1 Network Model

The network model of framework for heterogeneous data fusion in e-healthcare sector consists of three major entities:

1. **Cloud Trusted Authority (CTA):** The PPFHI technique utilizes a centralized Cloud trusted authority that is supposed to have sufficient processing power. CTA is primarily in charge of registering healthcare platforms and maintaining information for each platform in the data-set. It also has a clock that divides time into intervals of equal length, each of which consists of a number of equal-length time units. In addition, the Secret Information *i.e.* private threshold levels and private reputation level, is generated and distributed regularly to each healthcare platform on request.

2. **E-Gateway unit (e-GW):** This system incorporates an e-gateway (e-GW) that is hooked to the CTA and serves as a bridge for interaction between the HP and CTA.

3. **Health platform (HP):** PPFHI technique accommodates huge numbers of IoT

devices related to healthcare platform in which each one is equipped with multiple sensors and On-device unit(ODU) which is able to communicate with infrastructures as well as other platforms ODU's Hp2Hp and Hp2CTA wireless scenario. Though each health platform can either be data receiver or data provider. When data is broadcast from a platform it is know as data provider and when received then it is acknowledged as data receiver.

The framework is shown in the figure 3.1.

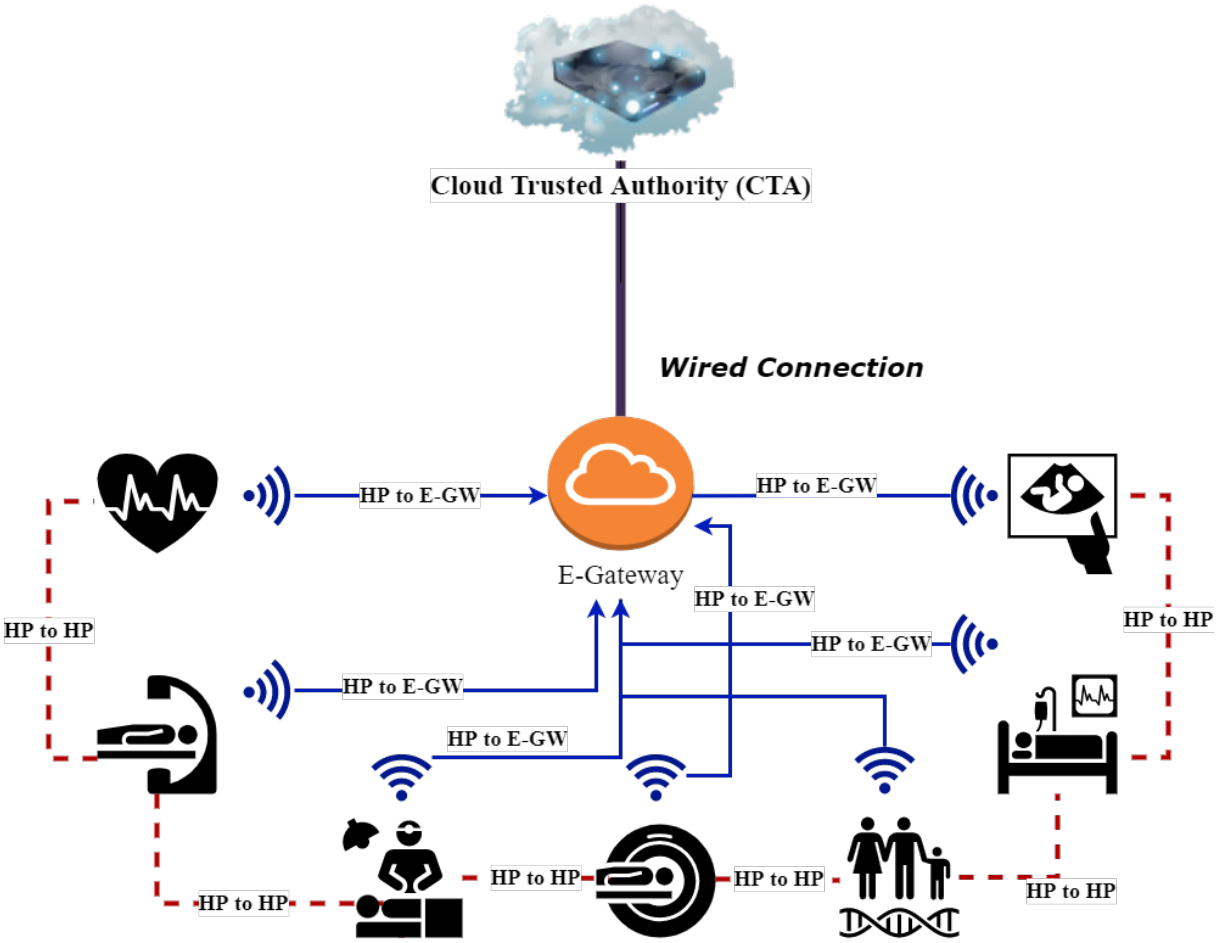

Figure 3.1: Communication flow between CTA and HP through e-GW

Communication flow is carried out via three channels as follows:

- between CTA to e-GW

- between HP to e-GW
- between HP to HP

All these entities carrying out communication need to ensure confidentiality as well as integrity of the data exchanged during these communications.

3.2.2 Threat Model

A threat model is established in which illegal access to the system is the principal goal of the attacker. Since the e-healthcare industry communicates over a public channel, an attacker can readily intercept the data. The attacker can perform one and / or many of the following actions:

- Capture / sniffing of data packets
- Carry out data (message) modification
- Save old intercepted packets so you can start a communication later by pretending to be an authentic entity
- Launch an MITM attack to intercept a session that is already underway and participate actively in undergoing communication

3.2.3 Design Goals

Following design goals are to be met in this research:

1. **Privacy Preservation:** The privacy of the data being exchanged during the communication of the devices and cloud server must always be in encrypted form to ensure confidentiality and it must also be tamper-proof.

2. **Device Authentication:** The devices trying to connect to the system must be properly authentication with a state-of-the-art authentication mechanism before the start of any kind of communication. This includes both new and old devices.

3. **Data Fusion:** The data outsourced to the cloud must be fused using machine learning techniques, so that data coming from multiple sources can provide a meaningful.

4. **Practicality:** The proposed scheme should be secure and practical in-terms of efficiency and must provide ease of use.

3.2.4 Security Assumptions

The following assumptions are made in this research:

- All the devices that are to be deployed for monitoring purposes possess state-of-the-art cryptographic properties so that the device itself is secure before deployment in the infrastructure prior to installation of the devices. The healthcare devices are registered in an off-line manner. This ensures that the privacy of devices is kept intact all the time during the registration process.

- The confidential information initialized during the requesting stage is always encrypted using the public key of the cloud server, while the response from the

devices is encrypted using the public key of each individual device. This ensures that if a certain device were found to be infected, its effects on the system would be negligible and would ensure that the infected is contained to that particular device.

- The data exchange between the devices would frequently generate requests and responses with different time intervals. All of the devices would process the data of a patient individually to ensure that the data of one patient is not mixed with another patient's information.

- The Cloud server is a semi-trusted authority *i.e.* it is trusted to store the data being provided to it only while at the same time it is curious as well and is interested in learning about the data.

- This research does not examine the privacy of the CTA or the infrastructures and accessing a HP's TM to disclose its privacy. The healthcare platform's privacy will not be compromised as the initial stage does not involve the platform.

3.2.5 Formal Definitions

We propose two definitions for the benefit of a subsequent explanation of the structured notation involved in the PPFHI scheme.

1. **Definition I:** A supplier of data In the opinion of the data recipient HP_j, HP_i will be regarded as trustworthy if only when HP_i is a registered platform and HP_i does not have a reputation lower than that Minimum level and reputation of HP_j's threshold.

2. **Definition II:** A data chunk sent by a data provider in view of a data receiver HP_j, HP_i is considered confident only where:

   (a) HP_i is a licensed platform *i.e.* the data confirms to the authenticity of the source of data

   (b) data confirms to the integrity

   (c) information confirms to the timeliness

   (d) HP_i's reputation levels shall not be less than the minimum of the level of threshold of HP_j and its degree of reputation

3.3 Proposed Model

This section put forwards the proposed work *i.e.* Privacy Preserving Data Fusion in E-Healthcare IoT Devices (PPFHI). The proposed scheme consists of various stages as follows:

- Scheme Startup
- HP-2-CTA Communication
- Healthcare Platform Registration
- Private Information Query
- HP-2-HP Information Sharing

3.3.1 Initialization of Cloud Trusted Authority (CTA)

The initialization of Cloud Trusted Authority (CTA) is presented in figure 3.2. After the clock is set, the key pair is generated and threshold / reputation levels are defined. The nonces are generated after division of time intervals is carried out. Hash collision is checked after generation of congruous and incongruous private values.

The phase is carried out in the following steps:

- After installing PPFHI technique for a certain healthcare platform safety application, the CTA generates it's public and private keys Pr_{CTA}, Pu_{CTA} respectively by setting it's clock, where Pr_{CTA} is kept secretly by CTA.

- It then defines a set of numbers n {where n $\in$ (2,3,..)} substitute threshold and reputation levels S_1. S_2,...,S_n, where $S_1 = 0 < S_2 = \frac{1}{n} < .... < S_n = \frac{n-1}{n}$.

- The CTA splits a sequence of time equal lengths T1, T2,... whereby each time interval is represented as the length of the interval denoted by ε.

- For each T_β {where $\beta \in$ (1,2,...)}, the CTA first generates a random nonce value $N\upsilon_T^\beta$ (which is only known to the CTA), and then creates a congruous secret value Cs_γ^β and incongruous secret value Ci_γ^β for each $S\gamma \in (S_1, S_2,....S_n)$ as

$$\begin{cases} Cs_\gamma^\beta = hash_T(S_\gamma || N\upsilon_T^\beta) \\ Ci_\gamma^\beta = hash_T(\text{-}S_\gamma || N\upsilon_T^\beta) \end{cases} \tag{3.3.1}$$

where $hash_T(*)$ denotes CTA hash function for generating incongruous and congruous secret values , and || denotes the string sequence. It is presented in figure

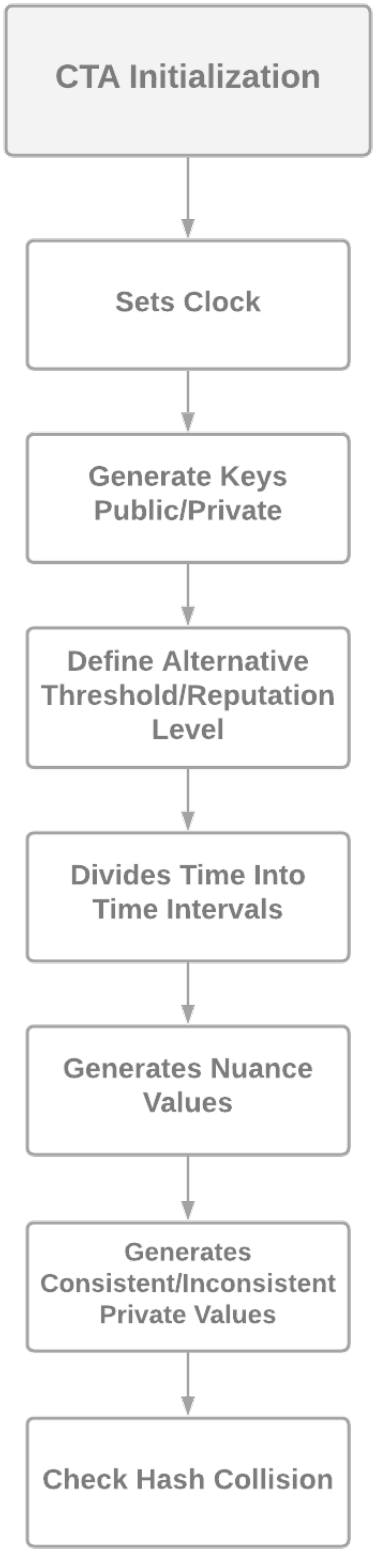

Figure 3.2: Initialization of Cloud Trusted Authority (CTA)

3.3 as:

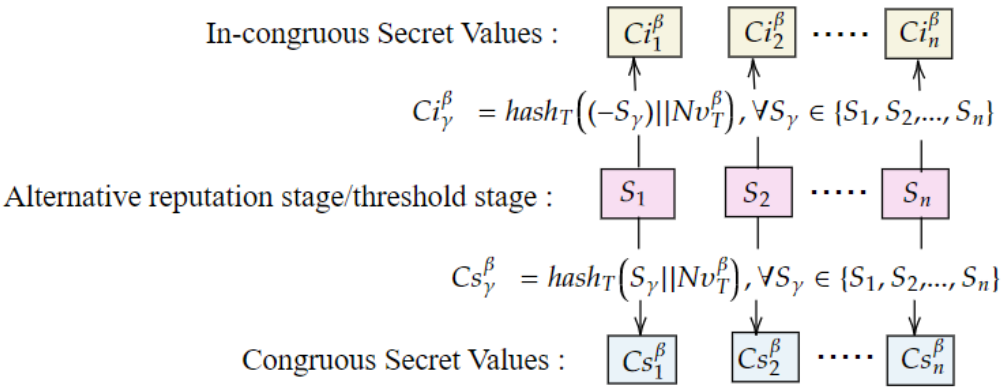

Figure 3.3: Generation of Congruous Secret Values

- As S_1 is defined as 0, the equations $S_1 = -S_1$ and $Cs_1^\beta = Cs_2^\beta$ hold for each T_β, Also if two or more elements in Cs_1^β ,Cs_2^β,... Cs_n^β, Ci_1^β,Ci_2^β,...Ci_n^β are same *i.e.* a minimal rate hash collision occurs, and CTA recalculates equation 3.3.1 with another Nv_T^β (until collision stops) otherwise each records $\langle\ \beta, S_\gamma, Cs_\gamma^\beta, Ci_\gamma^\beta\ \rangle$ in the database, where $\forall\ S_{\gamma_1}, S_{\gamma_2} \in (S_1,S_2,....,S_n)$, $\forall\ S_{\gamma_3} \in (S_2,S_3,....,S_n)$, and $S_{\gamma_1} \neq S_{\gamma_2}$, in equations $Cs_{\gamma_1}^\beta \neq Cs_{\gamma_2}^\beta$,$Ci_{\gamma_1}^\beta \neq Ci_{\gamma_2}^\beta$, $Cs_{\gamma_1}^\beta \neq Cs_{\gamma_3}^\beta$ always stay.

3.3.2 Healthcare Platform Registration

The registration of Healthcare Platform (HP) is carried out in the following steps:

- Whenever a new platform is registered with the CTA in offline manner , the CTA assigns i, a unique identifier to it as HP_i.

- The CTA, then, generates the public and private key Pu_{HP_1}, Pr_{HP_1} for HP_i, and

equips HP_i with a Trusted Module (TM) for storing Pu_{CTA}, Pr_{HP_1}, a clock which is in sync with that of the CTA, Cryptography methods, parameters and secret information, exchanged with the CTA or other healthcare platforms to guarantee integrity, confidentiality and availability.

- Next, the CTA derives the serial number for current time interval *i.e.* β and evaluates HP_i's reputation stage $RV^{\beta}_{HP_i}$ with range from [0,1], based on number of HP_i's on board sensors and the resolution, condition and type of each platform's sensor determining together the quality / condition of the data transmitted by HP_i, in offline manner.

- It then converts $RV^{\beta}_{HP_i}$ to the corresponding reputation value $RV^{\beta}_{HP_i}$ as shown in equation 3.3.2:

$$RV^{\beta}_{HP_i} = \begin{cases} S_1 = 0, \text{ if } RV^{\beta}_{HP_i} \in [0, \frac{1}{n}) \\ S2 = \frac{1}{n}, \text{ if } RV^{\beta}_{HP_i} \in [\frac{1}{n}, \frac{2}{n}) \\ S_n = \frac{n-1}{n}, \text{ if } RV^{\beta}_{HP_i} \in [\frac{n-1}{n}, 1] \end{cases} \tag{3.3.2}$$

- Afterwards, the CTA stores HP_i information *i.e.* Pr_{HP_i}, $RV^{\beta}_{HP_i}$ etc. in the database.

- The CTA updates $RV^{\beta}_{HP_i}$ periodically as the data of HP_i's on board sensors may change time to time.

The detail study of clock synchronization and TM are beyond the scope of this research and explained in [17, 45–47] respectively. The initialization of CTA and health platform is shown in the table 3.1 and presented in figure 3.4.

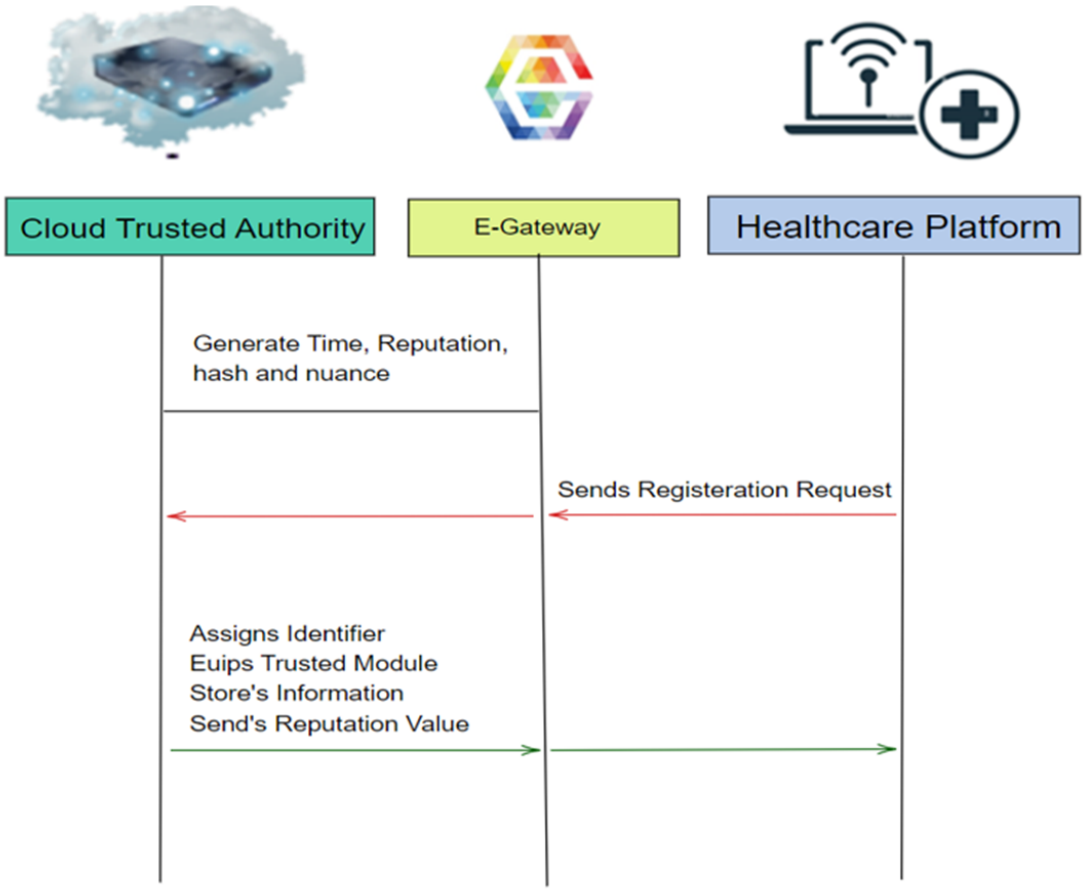

Figure 3.4: Initialization of CTA and Platform Registration Phase(s)

.

The check for trustworthiness of HP / authentication of HP is presented in figure 3.5.

3.3.3 Secret Information Query

The query phase of acquiring secret information is carried out in the following steps:

- When a healthcare platform is linked to the e-GW, it may use the infrastructure to request CTA secret information for the current and upcoming time intervals. Particularly, HP_i initially, acquire the current serial number of time interval β by choosing a personalized threshold level $TS^{\beta}_{HP_i}$ from [S_1,S_2,..., S_n] based on it's

Table 3.1: Initialization of CTA and Platform Registration

Cloud Trusted Authority	E-Gateway	Healthcare Platform
Generate Pu_{CTA},Pr_{CTA} defines "n" (where n $\in$ (2,3,..)) substitute threshold and reputation levels S_1. S_2,...,S_n, where $S_1 = 0 < S_2 = \frac{1}{n} < .... < S_n = \frac{n-1}{n}$ Splits time T_1, T_2,..,T_n Generates $N\upsilon_T^\beta$ Generates $\begin{cases} Cs_\gamma^\beta = hash_T(S_\gamma \| N\upsilon_T^\beta) \\ Ci_\gamma^\beta = hash_T(\text{-}S_\gamma \| N\upsilon_T^\beta) \end{cases}$ (3.3.3) Recalculates with another $N\upsilon_T^\beta$ (if hash collision accurs)		Health care Platform Registration
	$\xleftarrow{HP\ requests\ CTA\ for\ Registeration}$	
Assigns "i" to HP Generates Pu_{HP} and Pr_{HP} Equips with "TM" Derives β Evaluates $RV_{HP_i}^\beta = \begin{cases} S_1 = 0, \text{ if } RV_{HP_i}^\beta \in [0, \frac{1}{n}) \\ S2 = \frac{1}{n}, \text{ if } RV_{HP_i}^\beta \in [\frac{1}{n}, \frac{2}{n}) \\ S_n = \frac{n-1}{n}, \text{ if } RV_{HP_i}^\beta \in [\frac{n-1}{n}, 1] \end{cases}$ (3.3.4) Updates $RV_{HP_i}^\beta$ periodically		
	$\xrightarrow{CTA\ euips\ HP_i\ with\ TM\ and\ updates\ reputation\ score\ RV_{HP_i}^\beta}$	

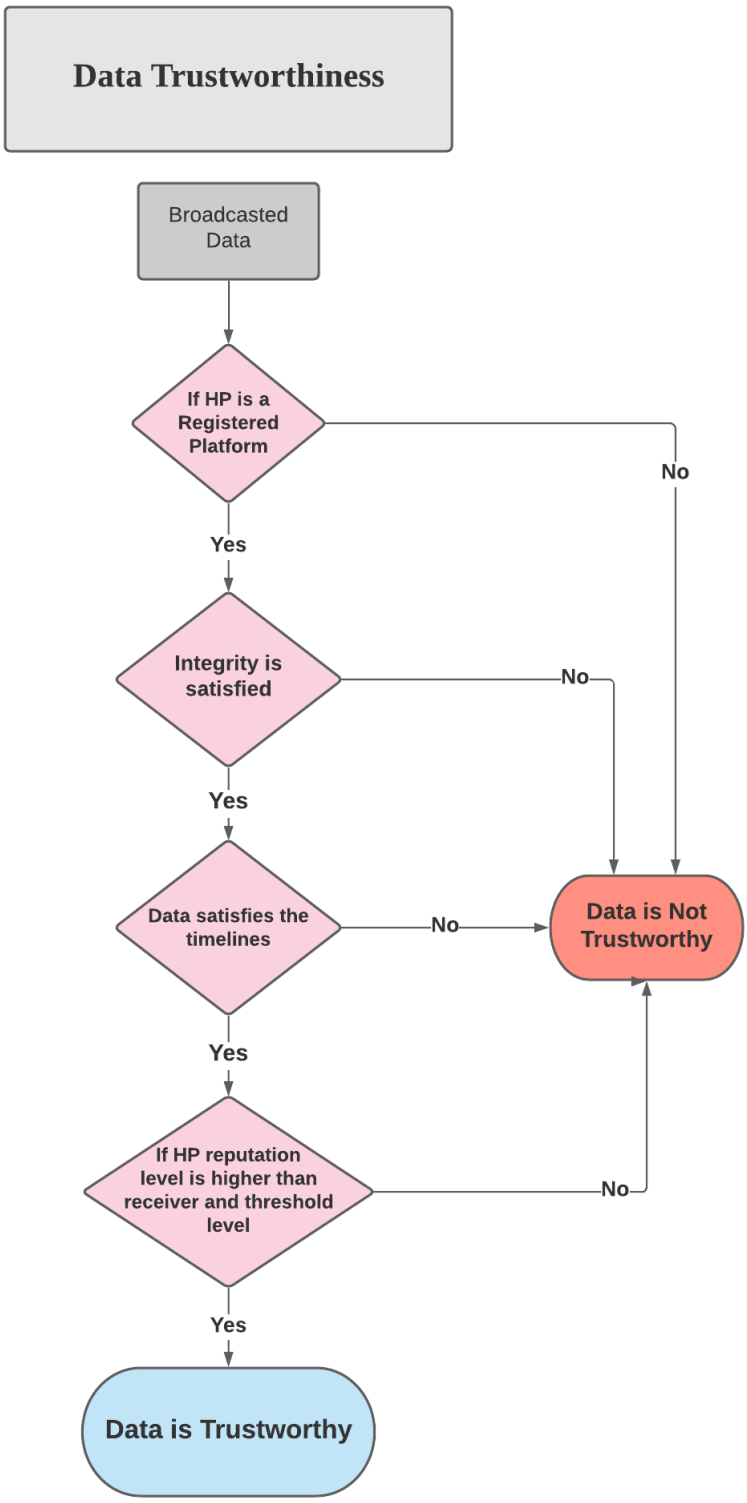

Figure 3.5: Authentication Check of *HP*

clock, and then a query $Q_{HP_i}^{\beta}$ is generated as in the equation 3.3.5:

$$\begin{cases} Q_{HP_i}^{\beta} = Enc_{Pu_{CTA}}(\mathrm{i}||\beta||Dsig_{HP_i}^{\beta}) \end{cases} \tag{3.3.5}$$

where $Dsig_{HP_i}^{\beta}$ is calculated as shown below:

$$\begin{cases} Dsig_{HP_i}^{\beta} = Sign_{Pr_{HP_i}}(\mathrm{i}||\beta||Ts_{HP_i}^{\beta}) \end{cases} \tag{3.3.6}$$

In the the equations 3.3.5 and 3.3.6, $Dsig_{HP_i}^{\beta}$ denotes the digital signature with Pr_{HP_i} on "i||β||$TS_{HP_i}^{\beta}$" and $Enc_{Pu_{CTA}}(*)$ denotes the asymmetric encryption Pr_{CTA} on *.

- Next, HP_i sends $Q_{HP_i}^{\beta}$ to the CTA via e-GW.
- After receiving $Q_{HP_i}^{\beta}$, the CTA first decrypts it with Pu_{CTA} to obtain i, β, $TS_{HP_i}^{\beta}$ and $Dsig_{HP_i}^{\beta}$ and then retrieves HP_i's public key Pu_{HP_i} from the data base.
- The CTA then calculates the current time intervals serial number $\overline{\beta}$ which can vary from β (owing to a replay attack or delay in transmission).
- Next, the CTA verifies the $Dsig_{HP_i}^{\beta}$, β and $TS_{HP_i}^{\beta}$ by:

$$\begin{cases} Dsig_{HP_i}^{\beta} \text{ is in line with equation 3.3.6} \\ \beta = \overline{\beta} \\ TS_{HP_i}^{\beta} \in (S_1, S_2, ..., S_n) \end{cases} \tag{3.3.7}$$

- If any of the above verification fails, the query $Q_{HP_i}^{\beta}$ is discarded by the CTA; Oth-

erwise, the CTA attempts to obtain HPi's secret information from the database, namely the private reputation level set $PR^{\beta}_{HP_i}$ and private threshold level $PL^{\beta}_{HP_i}$.

- If the sent result is not equal to null (which show's that the CTA has calculated these data from HP_i) the CTA adopts the existing $PR^{\beta}_{HP_i}$ and $PL^{\beta}_{HP_i}$ in the collection of results instead of collecting new ones.

- Alternatively, as shown in figure 3.6, the CTA retrieves HP_i reputation stage $RS^{\beta}_{HP_i}$ in T_β from its database. Without the generality loss, it is assumed $RS^{\beta}_{HP_i} = S_{\gamma_i}$ where $S_{\gamma_i} \in [S_1, S_2, ..., S_n]$.

- After that the CTA calculates a private reputation level set $PR^{\beta}_{HP_i}$ for HP_i as:

$$\begin{cases} PR^{\beta}_{HP_i} = \S\ [Cs^{\beta}_1, Cs^{\beta}_2, ...,\ Cs^{\beta}_{\gamma_i},\ Ci^{\beta}_{\gamma_i+1}, ..., Ci^{\beta}_n] \\ \triangle\ [Pr^{\beta}_{HP_i,1}, Pr^{\beta}_{HP_i,2}, .., Pr^{\beta}_{HP_i,n}] \end{cases} \tag{3.3.8}$$

where § signifies the set after the elements are sorted in (*) in lexicographic order and is denoted as "$\triangle$" in equation 3.3.8.

- Moreover, without generality loss, it can be assumed that $\min(TS^{\beta}_{HP_i}, RS^{\beta}_{HP_i}) = S_{\gamma^1_{HP_i}}$, where $S_{\gamma^1_{HP_i}} \in [S_1, S_2, ..., S_{\lceil \gamma_{HP_i} \rceil}]$.

- The CTA sets the HP_i's private threshold limit $PL^{\beta}_{HP_i}$ in T_β as $Ci^{\beta}_{\gamma^1_i}$ (i.e $Pl^{\beta}_{HP_i} = Ci^{\beta}_{\gamma^1_i}$) and stores the record $\langle\ \beta$, i, $PR^{\beta}_{HP_i}$, $Pl^{\beta}_{HP_i}\ \rangle$ on database.

- Afterwords, The HP_1 private information is tried to retrieve by CTA in $T_{\beta+1}$ *i.e.* $PR^{\beta+1}_{HP_i}$, $PL^{\beta+1}_{HP_i}$) from database.

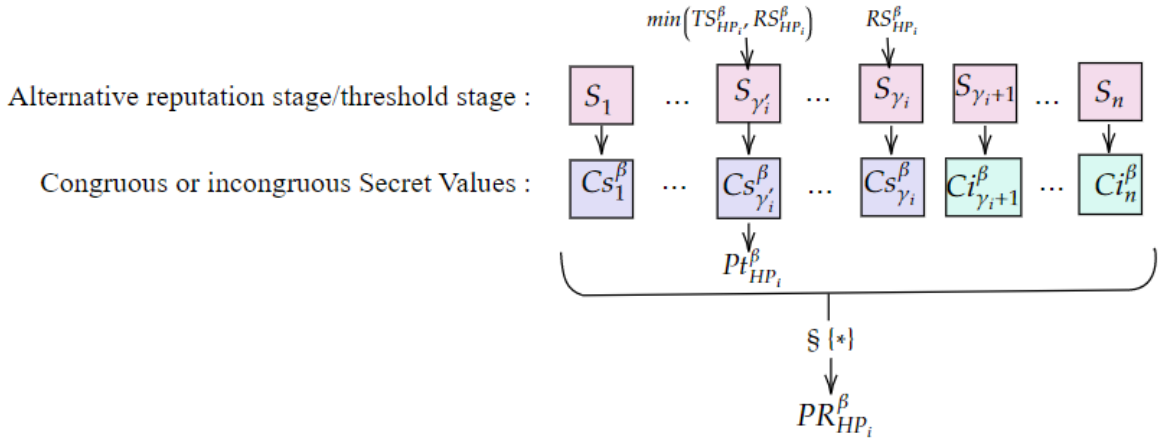

Figure 3.6: Generation of private threshold and reputation levels

- If a set other than zero is the outcome, the CTA embraces the already existant $PR_{HP_i}^{\beta+1}$ and $Pl_{HP_i}^{\beta+1}$ in the result set, instead of calculating new one's.
- Otherwise, the CTA sets $RV_{HP_i}^{\beta+1} = RV_{HP_i}^{\beta}$ and $TS_{HP_i}^{\beta+1} = TS_{HP_i}^{\beta}$.
- As each platform's reputation stage upgrade duration is much greater than each interval and each platform's threshold level does not change frequently, and therefore derives $PR_{HP_i}^{\beta+1}$ and $Pl_{HP_i}^{\beta+1}$ by adopting same methods with those deriving for $PR_{HP_i}^{\beta}$ and $Pl_{HP_i}^{\beta}$.
- Subsequently, record $\langle \beta, \text{i}, PR_{HP_i}^{\beta+1}, Pl_{HP_i}^{\beta+1} \rangle$ is stored in the database by CTA. Not to forget that $PR_{HP_i}^{\beta+1} \neq PR_{HP_i}^{\beta}$ and $Pl_{HP_i}^{\beta+1} \neq Pl_{HP_i}^{\beta}$ almost hold as with different time intervals congruous and incongruous private value's change.

- Eventually, response $Re^{\beta}_{HP_i}$ is generated by the CTA for Hp_i as:

$$\begin{cases} Re^{\beta}_{HP_i} = Enc_{Pu_{Hp_i}}(\langle\ \beta, PR^{\beta}_{HP_i}, Pl^{\beta}_{HP_i}\ \rangle \| \langle\ \beta+1, PR^{\beta+1}_{HP_i}, Pl^{\beta+1}_{HP_i}\ \rangle \| Dsig^{\beta}_{T,HP_i}) \end{cases} \tag{3.3.9}$$

  where $Dsig^{\beta}_{T,HP_i}$ is computed as:

$$\begin{cases} Dsig^{\beta}_{T,HP_i} = Sign_{Pr_CTA}(\langle\ \beta, PR^{\beta}_{HP_i}, Pl^{\beta}_{HP_i}\ \rangle \| \langle\ \beta+1, PR^{\beta+1}_{HP_i}, Pl^{\beta+1}_{HP_i}\ \rangle) \end{cases} \tag{3.3.10}$$

  In equations 3.3.9 and 3.3.10, $Dsig^{\beta}_{T,HP_i}$ denotes the digital signature with Pr_{CTA} on $\langle\ \beta, PR^{\beta}_{HP_i}, Pl^{\beta}_{HP_i}\ \rangle \| \langle\ \beta+1, PR^{\beta+1}_{HP_i}, Pl^{\beta+1}_{HP_i}\ \rangle$ and the asymmetric encryption with Pu_{HP_i} on (*) is denoted by $Enc_{Pu_{HOP_i}}$ (*).

- Next the response $Re^{\beta}_{HP_i}$ is sent by CTA to HP_i via the e-gateway.

- HP_i first decrypts the received response $Re^{\beta}_{HP_i}$ with Pr_{HP_i} to obtain $\langle\ \beta, PR^{\beta}_{HP_i}, Pl^{\beta}_{HP_i}\ \rangle$, $\langle\ \beta+1, PR^{\beta+1}_{HP_i}, Pl^{\beta+1}_{HP_i}\ \rangle$ and $Dsig^{\beta}_{T,HP_i}$ and then verifies $Dsig^{\beta}_{T,HP_i}$ with Pu_CTA. Next HP_i stores $\langle\ \beta, PR^{\beta}_{HP_i}$ and $Pl^{\beta}_{HP_i}\ \rangle$, $\langle$ in its storage if the associated records did not exists previously.

- Additionally, if HP_i does not receive $Re^{\beta}_{HP_i}$ in a timely manner, it may send request again for the TA once it enters another infrastructure's coverage range.

The data fusion query and response exchange is presented in the table 3.2 and shown in figure 3.7.

Table 3.2: Data Fusion Query and Response

Cloud Trusted Authority	E-Gateway	Healthcare Platform
	$\xrightarrow{CTA\ euips\ HP_i\ with\ TM\ and\ updates\ reputation\ Value\ RV^{\beta}_{HP_i}}$	
		Derives β by $TS^{\beta}_{HP_i}$ from [S_1,S_2,..., S_n] Query request $\left\{ Q^{\beta}_{HP_i} = Enc_{Pu_{CTA}}(\text{i}\|\|\beta\|\|Dsig^{\beta}_{HP_i}) \right.$ (3.3.11)
	$\xleftarrow{Query\ request\ Q^{\beta}_{HP_1}\ is\ generated\ to\ CTA}$	
Decrypts with Pu_{CTA} Obtains "i", β, $TS^{\beta}_{HP_i}$ and $Dsig^{\beta}_{HP_i}$ Retrieves Pu_{HP_i} Derives $\overline{\beta}$ Verifies $\begin{cases} Dsig^{\beta}_{HP_i} \text{ is in line with Equation (3.1.4)} \\ \beta = \overline{\beta} \\ TS^{\beta}_{HP_i} \in (S_1,S_2,...,S_n) \end{cases}$ (3.3.12) If Verification fails $Q^{\beta}_{HP_i}$ is discarded Otherwise CTA retrieves $PR^{\beta}_{HP_i}$ and $PL^{\beta}_{HP_i}$ If result is not-empty CTA adopts $PR^{\beta}_{HP_i}$ and $PL^{\beta}_{HP_i}$ Retrieves $RS^{\beta}_{HP_i}$ in T_{β} Without generality loss $RS^{\beta}_{HP_i} = S_{\gamma_i}$ where $S_{\gamma_i} \in [S_1,S_2,...,S_n]$. Generates $\begin{cases} PR^{\beta}_{HP_i} = \S\ [Cs^{\beta}_1,Cs^{\beta}_2,..., Cs^{\beta}_{\gamma_i}, Ci^{\beta}_{\gamma_i+1},...,Ci^{\beta}_n] \\ \triangle\ [Pr^{\beta}_{HP_i,1},Pr^{\beta}_{HP_i,2},..,Pr^{\beta}_{HP_i,n}] \end{cases}$ (3.3.13) We assume min($TS^{\beta}_{HP_i}$, $RS^{\beta}_{HP_i}$) = $S_{\gamma^1_{HP_i}}$, where $S_{\gamma^1_{HP_i}} \in [S_1, S_2,...,S_{\lfloor \gamma_{HP_i} \rfloor}]$. Sets $PL^{\beta}_{HP_i}$ in T_{β} and stores $\langle\ \beta$, i, $PR^{\beta}_{HP_i}$, $Pl^{\beta}_{HP_i}\ \rangle$ After Retrieving $T_{\beta+1}$ (I.e $PR^{\beta+1}_{HP_i}$, $PL^{\beta+1}_{HP_i}$) CTA sets $RV^{\beta+1}_{HP_i} = RV^{\beta}_{HP_i}$ and $TS^{\beta+1}_{HP_i}$ $=TS^{\beta}_{HP_i}$. Derives $PR^{\beta+1}_{HP_i}$ and $Pl^{\beta+1}_{HP_i}$ Stores $\langle\ \beta$, i, $PR^{\beta+1}_{HP_i}$, $Pl^{\beta+1}_{HP_i}\ \rangle$ As $PR^{\beta+1}_{HP_i} \neq PR^{\beta}_{HP_i}$ and $Pl^{\beta+1}_{HP_i} \neq Pl^{\beta}_{HP_i}$ Generates $\left\{ Re^{\beta}_{HP_i} = Enc_{Pu_{Hp_i}}(\langle\ \beta, PR^{\beta}_{HP_i}, Pl^{\beta}_{HP_i}\ \rangle \| \langle\ \beta+1, PR^{\beta+1}_{HP_i}, Pl^{\beta+1}_{HP_i}\ \rangle \| Dsig^{\beta}_{T,HP_i}) \right.$ (3.3.14)		
	$\xrightarrow{After\ retrieving\ HP_i\ private\ information\ PR^{\beta+1}_{HP_i}\ PL^{\beta+1}_{HP_i}\ CTA\ sends\ Response\ Re^{\beta}_{HP_i}}$	
		Decrypts $Re^{\beta}_{HP_i}$ with Pr_{HP_i} to obtain $\langle\ \beta, PR^{\beta}_{HP_i}, Pl^{\beta}_{HP_i}\ \rangle$, $\langle\ \beta+1, PR^{\beta+1}_{HP_i}, Pl^{\beta+1}_{HP_i}\ \rangle$ and $Dsig^{\beta}_{T,HP_i}$ Verifies $Dsig^{\beta}_{T,HP_i}$ with Pu_{CTA}. Stores $\langle\ \beta, PR^{\beta}_{HP_i}$ and $Pl^{\beta}_{HP_i}\ \rangle$, $\langle$ If Hp_i does not receive response on time it could re request.

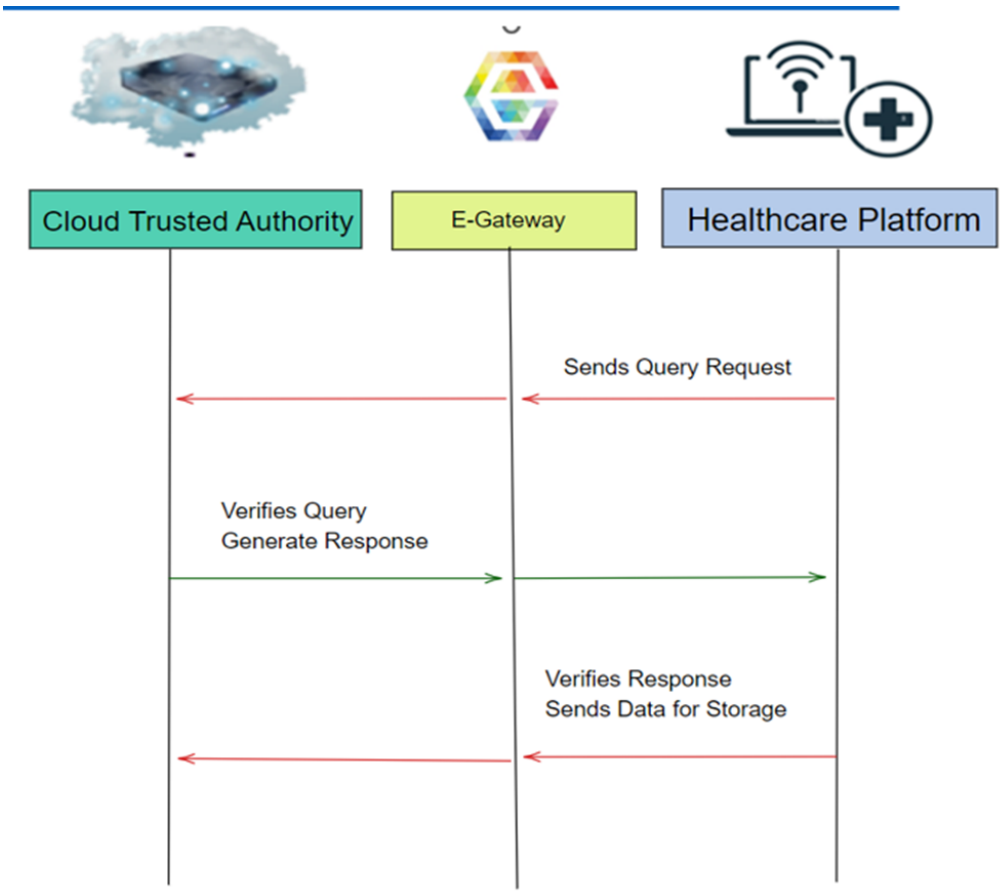

Figure 3.7: Data Fusion Query and Response Exchange

3.3.4 HP-2-HP Communication

The phase of HP-2-HP Communication is carried out in the following steps:

- In the proposed technique *i.e.* PPFHI; it is assumed that each time interval T_β contains κ where $\kappa \in \{1,2,...\}$ is time intervals of equal lengths $T_{\beta,1}$, $T_{\beta,2}$, ...,$T_{\beta,\kappa}$ and indicate the length of each $T_{\beta,\alpha}$ where $\alpha \in [1,2,..,\kappa]$, each platform after generating a fragment of data from the data collected by its on-board sensors, transmits it to surrounding platforms.

- Specifically, each healthcare platform *e.g.* HP_i as data provider, based on its clock, carries out derivation of current time interval β and it's serial number α,

and then information gets fused into the formatted data content $FD_{HP_i}^{\beta,\alpha}$ input from the on-board sensors.

- Next, as shown in figure 3.8, HP_i retrieves its private reputation level set $PR_{HP_i}^{\beta}$ in T_β from it's storage and generates a nuance value $Nv_{HP_i}^{\beta,\alpha}$, and the calculates the hash value $\widehat{Hv}_{HP_i,\gamma}^{\beta,\alpha}$ for each $Pr_{HP_i,\gamma}^{\beta} \in PR_{HP_i}^{\beta}$ as:

$$\begin{cases} \widehat{Hv}_{HP_i,\gamma}^{\beta,\alpha} = hash_{HP}\left(\beta \| \alpha \| FD_{HP_i}^{\beta,\alpha} \| Pr_{HP_i,\gamma}^{\beta} \| Nv_{HP_i}^{\beta,\alpha}\right) \end{cases} \tag{3.3.15}$$

where $hash_{HP}(*)$ indicates the hash function share among all the registered platforms.

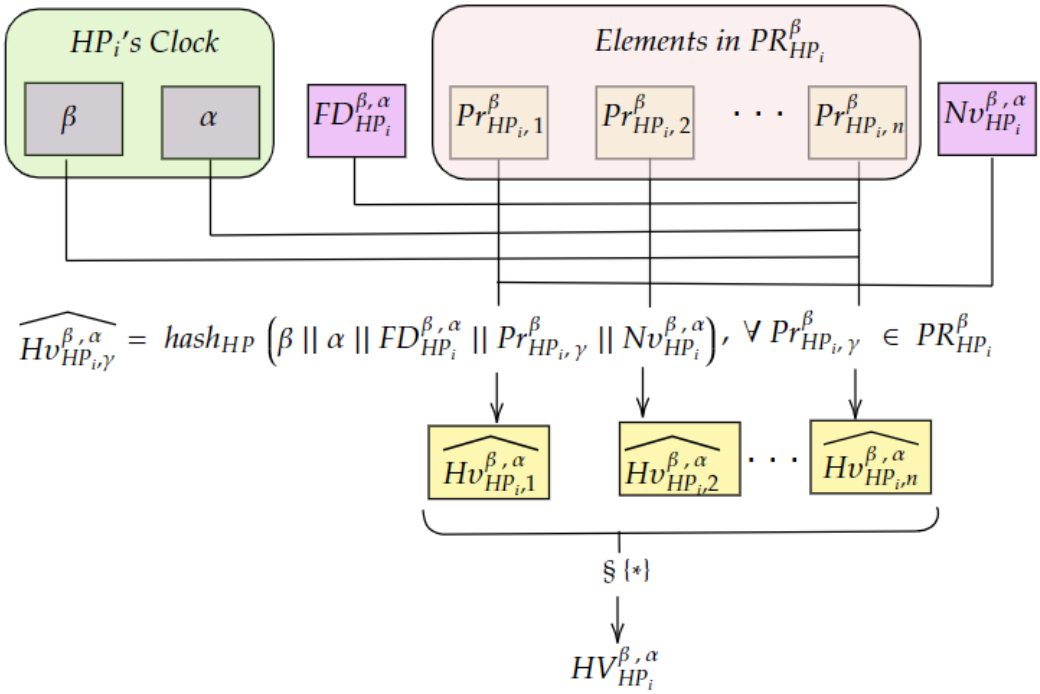

Figure 3.8: Generation of Hash value(s)

- If two or multiple elements in $\widehat{Hv}_{HP_i,1}^{\beta,\alpha}$, $\widehat{Hv}_{HP_i,2}^{\beta,\alpha}$, ..., $\widehat{Hv}_{HP_i,n}^{\beta,\alpha}$ are same *i.e.* a

collision of hash occurs with a minor rate, HP_i re-evaluates the equation 3.3.15 with another $N\upsilon_{HP_i}^{\beta,\alpha}$ until hash collision stops.

- Afterwords HP_i generates a hash value set $HV_{HP_i}^{\beta,\alpha}$ as:

$$\begin{cases} HV_{HP_i}^{\beta,\alpha} = \S[\ \widehat{H\upsilon}_{HP_i,1}^{\beta,\alpha}, \widehat{H\upsilon}_{HP_i,2}^{\beta,\alpha}, ..., \widehat{H\upsilon}_{HP_i,n}^{\beta,\alpha}] \\ \triangle[H\upsilon_{HP_i,1}^{\beta,\alpha}, H\upsilon_{HP_i,2}^{\beta,\alpha}, ..., H\upsilon_{HP_i,n}^{\beta,\alpha}] \end{cases} \tag{3.3.16}$$

and a piece of data generated as in:

$$\begin{cases} GDt_{HP_i}^{\beta,\alpha} = FD_{HP_i}^{\beta,\alpha} || HV_{HP_i}^{\beta,\alpha} || N\upsilon_{HP_i}^{\beta,\alpha} \end{cases} \tag{3.3.17}$$

and then broadcast $GDt_{HP_i}^{\beta,\alpha}$ to nearby healthcare platforms.

- Additionally, each platform continuously receives data from nearby platforms and checks it's reliability.

- Whenever each platform (e.g., HP_j) receives data from nearby platform (e.g., HP_i) HP_j first derives the current time interval's serial number $\overline{\beta}$ and the current time unit's serial number $\overline{\alpha}$ based on the clock (where $\overline{\alpha}$ may be different from α and $\overline{\beta}$ from β due to replay attack or transmission delay), and the each part of $GDt_{HP_i}^{\beta,\alpha}$ is extracted and private threshold limit $PL_{HP_i}^{\beta}$ in $T_{\overline{\beta}}$ is retrieved from it's storage.

- Next HP_j generates $\overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}}$ as:

$$\begin{cases} \overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}} = hash_{HP}(\overline{\beta} || \overline{\alpha} || FD_{HP_i}^{\beta,\alpha} || PL_{HP_i}^{\overline{\beta}} || N\upsilon_{HP_i}^{\beta,\alpha}) \end{cases} \tag{3.3.18}$$

and checks whether $\overline{H\upsilon}^{\beta,\alpha,\overline{\beta},\overline{\alpha}}_{HP_i,HP_j} \in HV^{\beta,\alpha}_{HP_i}$ holds.

- If this is true, the following conclusions are true at a high rate which always equates to value 1, unless a hash collision happens *e.g.*, $\beta \neq \overline{\beta}$, $\alpha \neq \overline{\alpha}$i, or $PL^{\overline{\beta}}_{HP_i} \notin PR^{\beta}_{HP_i}$, but $\overline{H\upsilon}^{\beta,\alpha,\overline{\beta},\overline{\alpha}}_{HP_i,HP_j} \in HV^{\beta,\alpha}_{HP_i}$ with a minute rate c, where c denotes false positive and it's actual value can be determined by both n and $hash_{HP}$(*)):

  1. HP_i(i.e., $GDt^{\beta,\alpha}_{HP_i}$ data provider)is a registered healthcare platform
  2. Integrity is satisfied by $GDt^{\beta,\alpha}_{HP_i}$
  3. $GDt^{\beta,\alpha}_{HP_i}$ satisfies the timeliness *i.e.* $\beta = \overline{\beta}$, $\alpha = \overline{\alpha}$)
  4. HP_i's reputation stage $RS^{\beta}_{HP_i}$ is no not less than the HP_j's threshold limit $TS^{\beta}_{HP_j}$ and reputation level $RS^{\beta}_{HP_j}$ *i.e.* $RS^{\beta}_{HP_i} \geq \min(TS^{\beta}_{HP_j}, RS^{\beta}_{HP_j})$.

  Based on Definition II, HP_j regards $GDt^{\beta,\alpha}_{HP_i}$ as trustworthy and maintains storage for the upcoming data fusion.

- Otherwise, HP_j will be unable to make any of the preceding conclusions 1-4 then, HP_j regards $GDt^{\beta,\alpha}_{HP_i}$ as untrustworthy and discards it directly.

The communication flow between healthcare platforms is presented in table 3.3 and shown in figure 3.9.

Each data receiver is capable of determining reliably the trustworthiness of received data through the aforementioned method in a non-interactive, lightweight and privacy-preserving manner. The aforementioned method ensures that HP_i can only achieve a $PR^{\beta}_{HP_i}$ and a $Pl^{\beta}_{HP_i}$ for each T_β though requesting the CTA more than once, and ensure

Table 3.3: Data Fusion and Communication Between Healthcare Platforms

Healthcare Platform as Data provider	On Device Unit	Healthcare Platform as Data Reciever
Assumes T_β contains κ(where $\kappa \in [1,2,...]$) equal-length time intervals $T_{\beta,1}$, $T_{\beta,2}$, ...,$T_{\beta,\kappa}$ Denotes $T_{\beta,\alpha}$ (where $\alpha \in [1,2,..,\kappa]$) HP_i as data provider derives β and α Fused data $FD_{HP_i}^{\beta,\alpha}$ Retrieves $PR_{HP_i}^{\beta}$ in T_β Generates $N\upsilon_{HP_i}^{\beta,\alpha}$ Calculates $\left\{\widehat{H\upsilon}_{HP_i,\gamma}^{\beta,\alpha} = hash_{HP}\,(\beta\|\alpha\|FD_{HP_i}^{\beta,\alpha}\|Pr_{HP_i,\gamma}^{\beta}\|\,N\upsilon_{HP_i}^{\beta,\alpha})\right.$ (3.3.19) Checks for hash Collision Generates $\begin{cases} HV_{HP_i}^{\beta,\alpha} = \S[\,\widehat{H\upsilon}_{HP_i,1}^{\beta,\alpha}, \widehat{H\upsilon}_{HP_i,2}^{\beta,\alpha}, ..., \widehat{H\upsilon}_{HP_i,n}^{\beta,\alpha}] \\ \triangle[H\upsilon_{HP_i,1}^{\beta,\alpha}, H\upsilon_{HP_i,2}^{\beta,\alpha}, ..., H\upsilon_{HP_i,n}^{\beta,\alpha}] \end{cases}$ (3.3.20) Data generated $\left\{GDt_{HP_i}^{\beta,\alpha} = FD_{HP_i}^{\beta,\alpha}\|HV_{HP_i}^{\beta,\alpha}\|N\upsilon_{HP_i}^{\beta,\alpha}\right.$ (3.3.21)		
	$\xrightarrow{\text{Generated Data } GDt_{HP_i}^{\beta,\alpha} \text{ is Broadcasted}}$	
		HP_j receives data from nearby platform (e.g., HP_i Derives $\overline{\beta}$ and $\overline{\alpha}$ $PL_{HP_i}^{\beta}$ in $T_{\overline{\beta}}$ is retrieved Generates $\left\{\overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}} = hash_{HP}(\overline{\beta}\|\overline{\alpha}\|FD_{HP_i}^{\beta,\alpha}\|PL_{HP_i}^{\overline{\beta}}\|N\upsilon_{HP_i}^{\beta,\alpha})\right.$ (3.3.22) Verifies whether $\overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}} \in HV_{HP_i}^{\beta,\alpha}$ holds (1) If $\beta \neq \overline{\beta}$, $\alpha \neq \overline{\alpha}$i, or $PL_{HP_i}^{\overline{\beta}} \notin PR_{HP_i}^{\beta}$, but $\overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}} \in HV_{HP_i}^{\beta,\alpha}$) HP_i(i.e., $GDt_{HP_i}^{\beta,\alpha}$ If $HP_i$is a registered healthcare platform, (b) Integrity is satisfied by $GDt_{HP_i}^{\beta,\alpha}$ (c) $GDt_{HP_i}^{\beta,\alpha}$ satisfies the timeliness If all of above are true HP_j regards $GDt_{HP_i}^{\beta,\alpha}$ as trustworthy and stores it for the upcoming data fusion Otherwise, regards $GDt_{HP_i}^{\beta,\alpha}$ as untrustworthy and discards it directly.

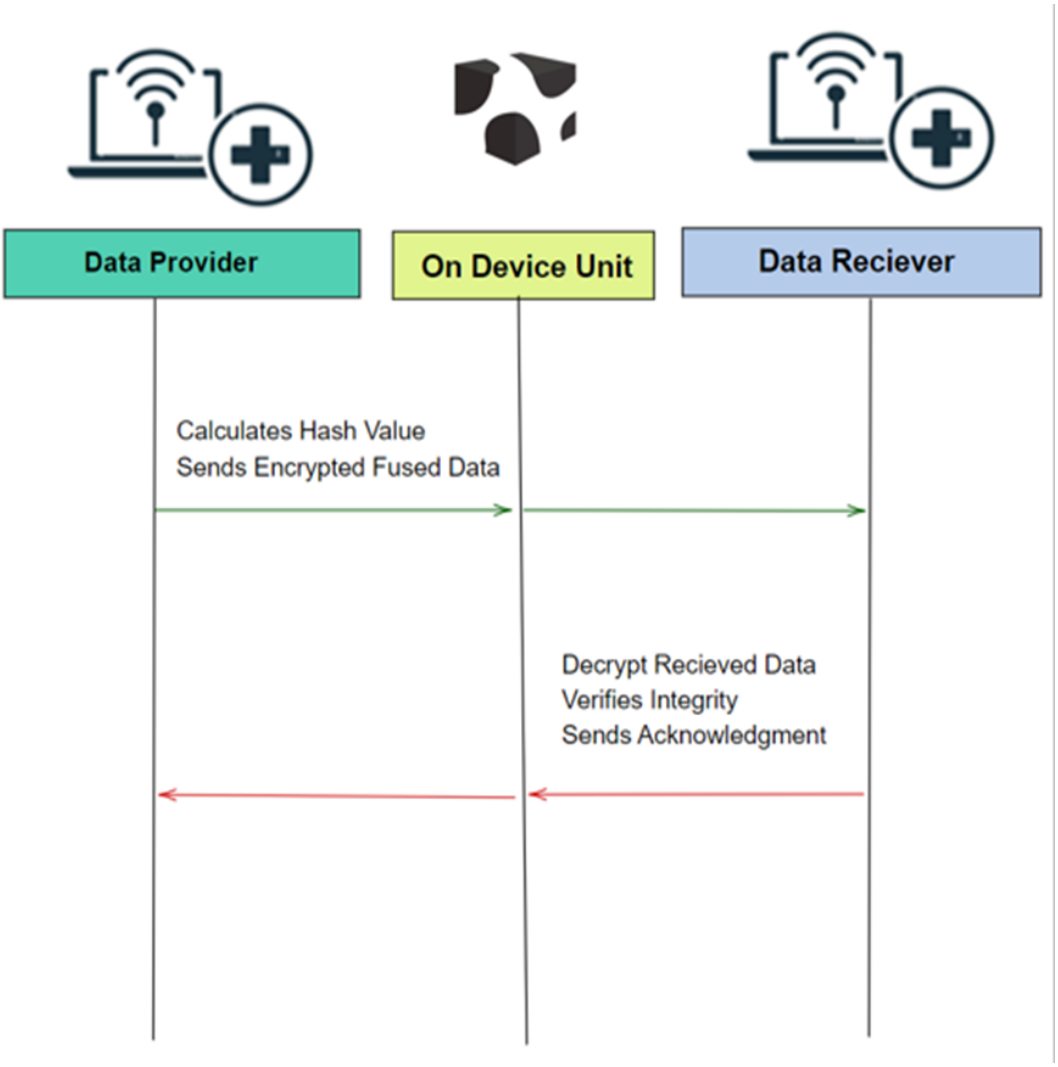

Figure 3.9: Communication Flow Between Healthcare Platforms

that HP_i is not able to obtain a private threshold corresponding to any threshold / reputation level S_{γ_i} greater than HP_i's actual reputation value $RV^{\beta}_{HP_i}$ *i.e.*, $S_{\gamma_i} \in [S_{\gamma_i+1}, S_{\gamma_i+1},..., S_n]$, even though HP_i deliberately sets a threshold level $Pl^{\beta}_{HP_i}$ higher then $RV^{\beta}_{HP_i}$ in it's query $Q^{\beta}_{HP_i}$. The above strategy as a result enhances the capability of privacy preservation to enhance the robustness of PPFHI scheme to a larger extent.

3.4 Summary

In this chapter, the framework for heterogeneous data fusion in IoT devices in healthcare sector was discussed with a detailed description of entities and their communication flow. Security goals and assumptions were presented. A novel proposed mutual authentication protocol with all phases was put forward in detail. Chapter 4 will present the formal analysis of proposed scheme in terms of performance, security and efficiency.

Chapter 4

Performance Analysis

4.1 Overview

The formal analysis of proposed in terms of security, performance and efficiency are discussed in this chapter. The analysis is carried out in three sections. Firstly, the security features and robustness of our enhanced suggested authentication system are scrutinised and analysed. Secondly, a comparative analysis is drawn with the existing schemes and lastly, performance analysis is carried out in terms of computation overhead and comparative analysis.

4.2 Security Analysis

The security features of privacy preservation, accuracy of trust evaluation and soundness of our proposed scheme PPFHI is discussed in this section.

4.2.1 Privacy Preservation

As discussed in security assumptions, it has already been declared that the CTA is trustworthy and securely stores the healthcare platforms data in its storage, and only few of the infrastructures and platforms are curious about the privacy of other platforms and try to disclose their privacy by capturing the information broadcasted by them. Furthermore, HP registration processes are handled in an off-line manner between the CTA and each new HP, ensuring that the platform's privacy is protected during the HP registration stage. Besides, in initial stage where the secret information is requested, CTA's public key Pu_{CTA} is used to encrypt each request (e.g., $Q^{\beta}_{HP_i}$) and HP_i's public key Pu_{HP_i} is used to encrypt each response (e.g., $Re^{\beta}_{HP_i}$). In this way the decrypted contents in $Q^{\beta}_{HP_i}$ or $Re^{\beta}_{HP_i}$ are secured from the curious infrastructures and the other platforms. Moreover, the limited number of $Q^{\beta}_{HP_i}$ or $Re^{\beta}_{HP_i}$ and differing time intervals prevents the linkage of $Q^{\beta}_{HP_i}$ or $Re^{\beta}_{HP_i}$ by the curious infrastructures and the other platforms to reveal HP_i's privacy. Following that, the notion of strong privacy preservation can be explained in detail as follows:

- In the PPFHI scheme, HP_i's data $GDt^{\beta,\alpha}_{HP_i}$ does not contain the $RS^{\beta}_{HP_i}$ of each data provider (*e.g.*, HP_i) and for each T_β every data recipient (*e.g.*, HP_j) can only obtain a $PL^{\beta}_{HP_i}$ (generated randomly by the CTA) even though it requests the CTA for multiple. This is how the trustworthiness of $GDt^{\beta,\alpha}_{HP_i}$ in the view of HP_j is judged but cannot disclose $RS^{\beta}_{HP_i}$.

- Both $PR^{\beta}_{HP_i}$, $PL^{\beta}_{HP_i}$ have the high-entropy feature since $N\upsilon^{\beta}_{T}$ is kept in secret by CTA and generated randomly. Thus, by using the brute force approach to obtain

$PL^{\beta}_{HP_i}$ is still infeasible for HP_j (as well as the curious infrastructures and the other healthcare platforms). Even if it originates from the same data supplier or a group of data providers with similar reputation scores, different data contains distinct hash function and nuance value sets, so HP_j cannot infer $PL^{\beta}_{HP_i}$ using the frequency analysis approach. As a result, $PL^{\beta}_{HP_i}$ cannot be linked or disclosed using HP_i's data. Furthermore, because $PR^{\beta}_{HP_i}$ is not contained in HP_i's data and cannot be exposed through $HV^{\beta,\alpha}_{HP_i}$ (because to the one-way aspect of $HA_{HP}(*)$, $PR^{\beta}_{HP_i}$ cannot be revealed or linked via HP_i's data.

- In this scheme, the $Rl^{\alpha}_{HP_i}$ of each data provider is not contained in HP_i data $Dt^{\beta,\alpha}_{HP_i}$ and each data receiver can get the secret threshold level that the cloud server generates at random for each equal length time unit, even if it requests the cloud server numerous times.

- This implies that a HP is able to identify whether the digital signatures of the HP are trustworthy in its view but won't reveal HP's reputation score in equal length time units. $RV^{\beta}_{HP_i}$ is a secret value which is generated randomly and is kept only by the cloud server.

- Furthermore, unlike in [25], the $TS^{\beta}_{HP_i}$ and $PL^{\beta}_{HP_i}$ of each data receiver (e.g., HP_j) do not need to be broadcasted to the data provider (e.g., HP_i), therefore HP_i (as well as the curious infrastructures and other platforms) cannot access $TS^{\beta}_{HP_i}$ or $PL^{\beta}_{HP_i}$ at the HP2HP data exchange stage.

- The data being broadcasted by the HP does not contain the identifier i and other information related to the sensors are replaced by random values and hash values

which obviously vary with different data ensuring that the unique identifier and sensors information is not revealed.

- Lastly, each data (e.g., $GDt_{HP_i}^{\beta,\alpha}$) broadcasted by HP_i does not contain unique classifier, and its onboard sensor information is replaced with $N\upsilon_{HP_i}^{\beta,\alpha}$ and $HV_{HP_i}^{\beta,\alpha}$ which differ with different data. This helps in prevention of HP_i's unique identifier and information from onboard sensors, being leaked or linked in any way.

4.2.2 Trust Evaluation Accuracy

Trust assessment primarily examine the accuracy of determining the trustworthiness of data $GDt_{HP_i}^{\beta,\alpha}$ broadcasted by a data provider HP_i and analyzed in recipient's view HP_j. This is done by verifying if $\overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}} \in HV_{HP_i}^{\beta,\alpha}$ holds in PPFHI scheme. Without forfeiting generality, we assume $RS_{HP_i}^{\beta} = S_{\gamma_i}$ and $\min(TS_{HP_j}^{\beta}, RS_{HP_j}^{\beta}) = S_{\gamma_j}$ where S_{γ_i},S_{γ_j} $\in \{S_1, S_2,..., S_n\}$. The following is a detailed analysis:

- if and only if (a) HP_i is a registered Platform, it can acquire $PR_{HP_i}^{\beta}$ and construct $HV_{HP_i}^{\beta,\alpha}$. Furthermore, the conclusion $PL_{HP_i}^{\overline{\beta}} = Cs_{\gamma_j}^{\beta} \in \{Cs_1^{\beta}, \{Cs_2^{\beta},\ldots, Cs_{\gamma_j}^{\beta},\ldots,Ci_{\gamma_i}^{\beta},\ldots, Ci_{\gamma_{i+1}}^{\beta},\ldots, Ci_{\gamma_{i+n}}^{\beta}\} = PR_{HP_i}^{\beta}$ holds, if and only if (a) HP_i is a registered platform, (b) $GDt_{HP_i}^{\beta,\alpha}$ fulfills the timeliness (i.e., $\beta = \overline{\beta}$ and $\alpha = \overline{\alpha}$), and (c) in the view of HP_j, HP_i is trustworthy (i.e., $S\gamma_i = RS_{HP_i}^{\beta} \geq \min(TS_{HP_i}^{\beta}, RS_{HP_i}^{\beta}) = S\gamma_j$.

- Moreover, the conclusion $\overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}} \in HV_{HP_i}^{\beta,\alpha}$ holds for sure if

  1. HP_i is a registered platform,

2. $GDt_{HP_i}^{\beta,\alpha}$ fulfills the timeliness,

3. in the view of HP_j, HP_i is trustworthy, and

4. $GDt_{HP_i}^{\beta,\alpha}$ fulfils the integrity.

- Alternatively, the conditions mentioned above holds with a high rate of 1-F if the conclusion $\overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}} \in HV_{HP_i}^{\beta,\alpha}$ holds, where F is an insignificant false positive rate.

- Precisely, it is supposed that n $\in$ {5,10,50,100} and the HA_{HP}(*) is calculated by using non-cryptographic hash function and can get a high-speed hash with a low conflict rate.

- According to definition II, in the view of HP_j, $GD_{HP_i}^{\beta,\alpha}$ is reliable if and only if $GD_{HP_i}\beta,\alpha$ meets the above requirements 1-4. Thus, by checking if $\overline{H\upsilon}_{HP_i,HP_j}^{\beta,\alpha,\overline{\beta},\overline{\alpha}} \in HV_{HP_i}^{\beta,\alpha}$ holds, HP_j an judge whether $GD_{HP_i}^{\beta,\alpha}$ is trustworthy in its view.

4.2.3 Soundness

The soundness of the scheme is analyzed in which the cloud authority is assumed as a semi-trusted entity, Some infrastructures may be down or malevolent, and some health platforms may turn harmful. In PPFHI, both the cloud server and honest infrastructures are not participating in the HP-2-HP stage of data exchange. When a HP_i is in range of an infrastructure, it can generate a request for its confidential data to the cloud server in the current time interval and next time via framework. If the connection is not established, then the request would be regenerated while the health platform is in coverage range of an infrastructure. This implies that if the cloud server is temporarily

unavailable or inaccessible for some duration, the normal procedures of the infrastructure would still not be greatly influenced.

As the malicious infrastructure cannot decrypt $Q_{HP_i}^{\beta}$ or $Re_{HP_i}^{\beta}$ to obtain HP's secret information because both would be encrypted. The eavesdropping and analysis of $GDt_{HP_i}^{\beta,\alpha}$ cannot help any hostile infrastructure because it contains no secret information or information that can be traced to a device. Following are the attack strategies possible against the scheme:

1. Interprets and replays requests
2. Manipulates or forges requests
3. Discards requests

In our scheme, each request contains the digital signatures and each signature contains the hash value set. Therefore, modification to any request can be perceived and becomes infeasible to forge them as well. The serial numbers of time unit and intervals are contained in digital signatures which are validated by the data receiver. As a result, the second attack tactic is ineffective against every malevolent infrastructure. If requests are being discarded by the malicious infrastructure, then the effects would be the same as generated by any unavailable infrastructure. Some of the attack strategies that can be adopted by a compromised health platform include:

1. Manipulation of digital signatures by changing values of alpha and beta.
2. Generation and broadcasting of false data using expired secret reputation level.
3. Interception of digital signatures and replay them in $T^{\beta,\alpha}$.

4. Invading of cloud server's database or TM to modify the health platform reputation information and generate / broadcast false data using the secret information of health platform respectively.

Each data element is checked for relevance using the data receivers in our proposed approach, and any generated data element with an expired secret information level is destroyed and judged invalid. All the data elements would go through an integrity check through the data receivers thus any modification would become impractical. The secret information of a HP is well maintained by its corresponding TM and thus any other platform cannot steal this critical information. We can safely assume that the cloud server would protect its database and would try to eliminate any malicious attempts that any adversary is about to launch. This goes to show that the scheme is quite secure, robust, and provides protection against multiple attacks from the malicious entities. While in the HP-2-HP data exchange stage the compromised infrastructures may carry out attacks, in the HP-2-HP stage of data exchane in PPFHI scheme, the CTA and honest infrastructures are not engaged.

Furthermore, if infrastructure's coverage range covers platform HP_i, through the infrastructure it can request the CTA for its private data in the current and next time intervals. It sends request to the CTA once again it enters the coverage range of another infrastructure if HP_i does not receive a timely response resulting in the acquisition of its private information in advance in the following time interval (*i.e.*, $T_{\beta+1}$) from the CTA at any time (when the CTA is accessible and available, and the platform is within the coverage range of an available infrastructure). As a result, the temporary inaccessibility of the CTA will have little impact on the PPFHI scheme's normal operation if the interval of

the inaccessibility and / or unavailability is less than ζ, and the unavailability of some of the infrastructures will have minimal effect on the PPFHI scheme's normal operation considering each platform can come into another available infrastructure's coverage range within a duration less than ζ. The soundness of our proposed scheme PPFHI is discussed as follows:

- Since $Q_{HP_i}^{\beta}$ and $Re_{HP_i}^{\beta}$ are encrypted with Pr_{CTA} and Pr_{HP_i} respectively, malicious infrastructure cannot decrypt $Q_{HP_i}^{\beta}$ and $Re_{HP_i}^{\beta}$ to obtain HP_i's secret information. Since $GDt_{HP_i}^{\beta,\alpha}$ does not include any private or linkable information, malevolent infrastructure cannot gain from eavesdropping and examining $GDt_{HP_i}^{\beta,\alpha}$. Instead, to compromise the PPFHI scheme, each malevolent infrastructure could use one or more of the attack vectors listed below:

  1. It controls or falsifies $Q_{HP_i}^{\beta}$, $Re_{HP_i}^{\beta}$, or $GDt_{HP_i}^{\beta,\alpha}$.
  2. It captures and replays $Q_{HP_i}^{\beta}$, $Re_{HP_i}^{\beta}$, or $GDt_{HP_i}^{\beta,\alpha}$.
  3. It intentionally discards $Q_{HP_i}^{\beta}$, $Re_{HP_i}^{\beta}$, or $GDt_{HP_i}^{\beta,\alpha}$.

- Any modification to $Q_{HP_i}^{\beta}$, $Re_{HP_i}^{\beta}$, or $GDt_{HP_i}^{\beta,\alpha}$ can be simply observed, and counterfeiting them is also infeasible due to the fact that each $GDt_{HP_i}^{\beta,\alpha}$ contains the hash value set $HV_{HP_i}^{\beta,\alpha}$ and each $Q_{HP_i}^{\beta}$ and $Re_{HP_i}^{\beta}$ contain the digital signatures $Dsig_{HP_i}^{\beta}$, $Dsig_{T,HP_i}^{\beta}$ respectively.

- It guarantee that even though the CTA receives $Q_{HP_i}^{\beta}$ for multiple times or HP_i receives $Re_{HP_i}^{\beta}$ for multiple times, for each T_{β} , healthcare platform (e.g., HP_i) can only attain a $PR_{HP_i}^{\beta}$, $PL_{HP_i}^{\beta}$. In the mean time, each data recipient validates

the time's serial numbers and that of time unit (*i.e.*, β and α, respectively) contained in $GDt_{HP_i}^{\beta,\alpha}$ Thus, the second attack strategy provides no gain to malicious infrastructure.

- If a malicious infrastructure intentionally keeps discarding $Q_{HP_i}^{\beta}$ or $Re_{HP_i}^{\beta}$, it will generate the similar negative impact as that produced by an inaccessible infrastructure. This will not significantly impact the normal running of the PPFHI scheme. Besides, data provider (e.g., HP_i) broadcasts $GDt_{HP_i}^{\beta,\alpha}$ to the nearby platforms, the normal running of the PPFHI scheme intentionally will also not be obviously influenced by rejecting $GDt_{HP_i}^{\beta,\alpha}$ by a malicious infrastructure. Thus, the third attack strategy is also mitigated.

4.3 Informal Security Analysis

In this section, we examine the PPFHI scheme's resistance against a variety of malicious attacks in depth and security features of our proposed scheme. To compromise the PPFHI scheme, one or more of the following attack methods are adopted by HP_k:

1. It employs its own expired private reputation level set to construct and broadcast fake data $GDt_{HP_k}^{\beta,\alpha}$ (which also contains false $FD_{HP_i}^{\beta,\alpha}$, as shown below).

2. It intercepts $GDt_{HP_i}^{\beta,\alpha}$ which is broadcasted by HP_i in $T_{\beta,\alpha}$) and replays it in $T_{\beta,\alpha}$(where $\widehat{\alpha} \neq \alpha$ or $\widehat{\beta} \neq \beta$).

3. It can control $GDt_{HP_i}^{\beta,\alpha}$ by altering one or more parts of $GDt_{HP_i}^{\beta,\alpha}$.

4. By implementing the brute force approach and utilizing $PL^{\beta}_{HP_k}$ or $PR^{\beta}_{HP_k}$, it reconstructs $PL^{\beta}_{HP_k}$ to hold one or multiple coherent secret values which relates to multiple reputation levels that are greater than the HP_k's actual value of reputation level $RS^{\beta}_{HP_k}$ in T_{β}.

5. By utilizing the brute force approach, it recreates $HV^{\beta,\alpha}_{HP_k}$ to have one or more hash values linking to one or more reputation levels greater than HP_k's actual reputation level $RS^{\beta}_{HP_k}$ in T_{β}.

6. It creates and transmits false data by invading HP_i's TM to obtain HP_i's secret information.

7. To directly alter the reputation information of healthcare platforms, it invades the CTA's database.

The security features and resistance against all known attacks, as discussed above, are discussed below:

- Any malevolent platform (e.g., HP_k) cannot decrypt $Q^{\beta}_{HP_i}$ or $Re^{\beta}_{HP_i}$ to get the secret data of another registered healthcare platform as they are encrypted with Pr_{CTA} and Pr_{HP_i}, respectively thus gaining resistance against the 1^{st} attack.

- HP_k cannot gain from eavesdropping and evaluating $GDt^{\beta,\alpha}_{HP_i}$ as it does not hold any private or linkable data. In fact, the data in PPFHI scheme will be checked by receiver for timeliness and regarded as invalid and rejected it is created by using an expired private reputation level set. In this way, HP_k will be unable to gain any advantage from the 1^{st} and 2^{nd} attack method.

- Furthermore, the data receivers will examine each data for integrity and data alteration will be detected so HP_k will not gain from the third attack technique. Furthermore, at the requesting stage of private data, HP_k is unable to get the $PL^{\beta}_{HP_k}$ coherent to any reputation stage greater than $RS^{\beta}_{HP_k}$ in T_β. Each Cs^{β}_{γ} and $H\upsilon^{\beta,\alpha}_{HP_{k,\gamma}}$ (where $\gamma \in \{1,2,..,n\}$) having the high-entropy feature making it infeasible for HP_k to obtain one or more consistent private hash values that are corresponding to one or more reputation stages higher than HP_k actual value of reputation level $RS^{\beta}_{HP_k}$ in T_β by using brute force method, preventing HP_k from benefiting from the 4^{th} and the 5^{th} attacks.

- The HP_i's secret knowledge is presumed to be maintained properly by its TM, and HP_k is unable to acquire this information mitigating 6^{th} attack technique. It is plausible to believe that the CTA's database is well-protected, and that HP_k won't be able to directly edit the platform's reputation information if it invades it preventing the 7^{th} attack. As a consequence, the PPFHI system can provide high resistance to a variety of malicious attacks.

4.4 Performance Analysis

The performance analysis is discussed in this section.

4.4.1 Comparative Analysis

The attributes defined for our comparison are as follows:

- A_1: Mutual Authentication
- A_2: Privacy Preservation
- A_3: Scalability
- A_4: Message Confidentiality / Resistance to Eavesdropping
- A_5: Security against Impersonation Attack
- A_6: Message Integrity / Resistance to Message Modification Attack
- A_7: Security against Replay Attack
- A_8: Security against MITM Attack
- A_9: Session Key Security
- A_{10}: Heterogeneous IOT Data

The comparison is demonstrated in table 4.1 with ✓shows the existence of said attribute.

Table 4.1: Comparative Analysis

Research	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
[44]		✓	✓	✓					✓	✓
[45]		✓		✓					✓	
[50]	✓	✓		✓		✓	✓		✓	
[52]	✓	✓	✓	✓		✓			✓	✓
PPFHI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

4.4.2 Computation Overhead

In the PPFHI policy, the frameworks are only used as interfaces for communication between the healthcare platforms and CTA where and the CTA is supposed to have sufficient processing power. As a result, the calculation overhead has a significantly bigger influence on the actual cooperative healthcare platform's performance of safety applications than it does on the CTA's infrastructure. Meanwhile, the computation overhead sustained by the data (shared between nearby platforms in an HP-2-HP manner) has a substantially bigger effect than requests and responses on the performance of cooperative vehicular safety applications (transmitted between platforms and infrastructures in an HP2I manner). We'll focus on the data computation overhead on the healthcare platform side to reduce space. By counting the number of times signatures are generated, verified, hash functions are executed, and other operations in the data creation and data trustworthiness evaluation processes, we examine the computational overhead of creating a piece of data and concluding a trust evaluation for a piece of data.

When data $GDt_{HP_i}^{\beta,\alpha}$ is generated by HP_i, it has to calculate the $hash_{HP}$(*) for n times to generate $\widehat{H\upsilon}_{HP_i,1}^{\beta,\alpha}$, $\widehat{H\upsilon}_{HP_i,2}^{\beta,\alpha}$, ..., $\widehat{H\upsilon}_{HP_i,n}^{\beta,\alpha}$ and to carry out the sorting of n elements (i.e $\widehat{H\upsilon}_{HP_i,1}^{\beta,\alpha}$, $\widehat{H\upsilon}_{HP_i,2}^{\beta,\alpha}$, ..., $\widehat{H\upsilon}_{HP_i,n}^{\beta,\alpha}$) to obtain $HV_{HP_i}^{\beta,\alpha}$. $\gamma_h ash$ denotes the execution of computation overhead of $hash_{HP}$ (*) once and γ_{sort}(n) denotes sorting n elements, therefore the computation overhead of generation of a data fragment can be approximately evaluated from:

$$\gamma_G \approx n \cdot \gamma_h ash(*) + \gamma_{sort}(n) \tag{4.4.1}$$

4.5 Summary

The formal analysis of proposed in terms of security, performance and efficiency was discussed in this chapter. The analysis was carried out in three sections. Firstly, the security features and robustness of our enhanced suggested authentication system were scrutinised and analysed. Secondly, a comparative analysis was drawn with the existing schemes and lastly, performance analysis was carried out in terms of computation overhead and comparative analysis. Chapter 5 concludes the research and discusses some of the future works.

Chapter 5

Conclusion

5.1 Overview of Research

Data Fusion at edge computing plays an important role in IoT infrastructure and a lot of research has already been carried out in this domain on privacy preservation of homogeneous data fusion. However, there still remains a dire need to design a secure and lightweight privacy preserving scheme for heterogeneous data fusion which should be dynamic and adaptive in nature. To address the issues with authentication of the devices connected to CSP for communication purposes and for fused data communication while providing privacy preservation, a mechanism is required which provides high security as well as performs efficiently. If a device from healthcare is establishing a connection with the cloud server than there must be a proper mechanism for authentication of the device while it attempts to establish a connection with another entity. This research focused to take data from multiple sensors *i.e.* heterogeneous data and then use data fusion techniques to accurately identify the action needed to be taken autonomously by

the underlying machine. The main objective of this study was to put forward a secure and efficient scheme to overcome these prevailing issues. This thesis explored the possibility of providing privacy preservation and authentication mechanism while using data fusion in the field of IoT and presented a novel scheme for heterogeneous IoT devices in e-healthcare domain.

5.2 Summary of Research Contributions

The research work in this thesis builds up its foundation from literature review of the existing techniques being used for heterogeneous IoT devices using data fusion. The literature review in chapter 2 was done from various academic sources. This research then narrowed down to the privacy preservation and authentication of IoT devices connecting to the cloud while listing down the drawbacks of existing schemes and formulating the problem. Then in chapter 3, it discussed the construction of scheme privacy preserving data fusion in e-healthcare IoT devices, in detail and thoroughly covered the literature, design and implementation part of the thesis. A formal analysis was carried out in chapter 4 to show the efficiency in terms of performance and security of the proposed mechanism.

5.3 Conclusion

In this study, we have proposed a novel PPFHI scheme for heterogeneous data fusion in IoT devices that can efficiently balance privacy and trust assessment while requiring

little overhead in terms of computation, communication, or storage to enable distributed data fusion across the e-healthcare sector. Additionally, we have provided in-depth theoretical research, and the findings have shown that the PPFHI scheme is better compared to state-of-the-art schemes in many ways, including the accuracy of fusion outcomes.

5.4 Future Work

In the future, we will continue to assess the effectiveness of our suggested scheme, PPFHI, in other healthcare applications. Additionally, by adding an anonymous component and improved computational robustness, we will further strengthen the PPFHI scheme by adding new security aspects, such as resistance against traceability, Sybil attack etc.

References

[1] M. G. Avram, Advantages and challenges of adopting cloud computing from an enterprise perspective, Procedia Technology 12 (2014) 529:534.

[2] M. Nakayama, C. Chen, C. Taylor, the effects of perceived functionality and usability on privacy and security concerns about cloud application adoptions, Journal of Information Systems Applied Research 10 (2) (2017) 529–534

[3] 3. Dark Reading. 2019. Cloud Customers Faced 681M Cyberattacks in 2018. [online] Available at: <https://www.darkreading.com/attacks-breaches/cloud-customers-faced-681m-cyberattacks-in-2018> [Accessed 28 December 2020].

[4] How To WP Blogging. 2019. Top 17 Most Popular File Sharing Websites 2019 update list. [online] Available at: <https://www.howtowpblogging.com/popular-file-sharing-websites/> [Accessed 28 December 2020].

[5] i-SCOOP. 2020. IoT endpoints: the industries and use cases driving growth. [online] Available at: <https://www.i-scoop.eu/internet-of-things-iot/iot-endpoints-2020/> [Accessed 28 December 2020].

[6] Prati, A., Vezzani, R., Fornaciari, M., & Cucchiara, R. (2013). Intelligent

video surveillance as a service. In Intelligent multimedia surveillance (pp. 1-16). Springer, Berlin, Heidelberg.

[7] Gagolewski, Marek. (2015). Data Fusion: Theory, Methods, and Applications.

[8] Botta, Alessio & Donato, Walter & Persico, Valerio & Pescapè, Antonio. (2015). Integration of Cloud computing and Internet of Things: A survey. Future Generation Computer Systems.

[9] Kalamkar, Shrida & A, Geetha. (2020). Heterogeneous Data Fusion for Healthcare Monitoring: A Survey.

[10] Liu, Zhiquan & Ma, Jianfeng & Weng, Jian & Huang, Feiran & Wu, Yongdong & Wei, Linfeng & Li, Yuxian. (2021). LPPTE: A lightweight privacy-preserving trust evaluation scheme for facilitating distributed data fusion in cooperative vehicular safety applications. Information Fusion. 73. 10.1016/j.inffus.2021.03.003.

[11] Hsu, Yu-Liang, Po-Huan Chou, Hsing-Cheng Chang, Shyan-Lung Lin, Shih-Chin Yang, Heng-Yi Su, Chih-Chien Chang, Yuan-Sheng Cheng, and Yu-Chen Kuo. 2017. "Design and Implementation of a Smart Home System Using Multisensor Data Fusion Technology" Sensors 17, no. 7: 1631. https://doi.org/10.3390/s17071631

[12] Federico Castanedo, "A Review of Data Fusion Techniques", The Scientific World Journal, vol. 2013, Article ID 704504, 19 pages, 2013.

[13] H. Boström, S.F. Andler, M. Brohede, R. Johansson, A. Karlsson, J. Van Laere,

L. Niklasson, M. Nilsson, A. Persson, T. Ziemke, On the definition of information fusion as a field of research, Institutionen för Kommunikation och Information, 2007.

[14] H. Lee, B. Lee, K. Park, R. Elmasri, Fusion techniques for reliable information: a survey, Int. J. Digit. Content Technol. Appl. 4 (2010) 74–88.

[15] F. Alam , R. Mehmood , I. Katib , N.N. Albogami , A. Albeshri , Data fusion and IoT for smart ubiquitous environments: a survey, IEEE Access 5 (2017) 9533–9554.

[16] B. Khaleghi, A. Khamis, F.O. Karray, S.N. Razavi, Multisensor data fusion: a review of the state-of-the-art, Inf. Fusion 14 (2013) 28–44.

[17] Y. Zheng, Methodologies for cross-domain data fusion: an overview, IEEE Trans. Big Data 1 (2015) 16–34.

[18] O. Vermesan, P. Friess, Internet of things: Converging Technologies for Smart Environments and Integrated Ecosystems, River Publishers, 2013.

[19] G. Santucci, The internet of things: between the revolution of the internet and the metamorphosis of objects, in: Vision and Challenges for Realising the Internet of Things, 2010, pp. 11–24.

[20] F. Mattern, C. Floerkemeier, From the Internet of Computers to the Internet of Things, in: From Active Data Management to Event-Based Systems and More, Springer, 2010, pp. 242–259.

[21] W. Ding, X. Jing, Z. Yan, L.T. Yang, A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion, Inf. Fusion 51 (2019) 129–144.

[22] F. Qu, Z. Wu, F.Y. Wang, W. Cho, A security and privacy review of VANETs, IEEE Trans. Intell. Transp. 16 (6) (2015) 2985–2996.

[23] J. Kang, R. Yu, X. Huang, Y. Zhang, Privacy-preserved pseudonym scheme for fog computing supported internet of vehicles, IEEE Trans. Intell. Transp. 19 (8) (2017) 2627–2637.

[24] C. Xu, R. Lu, H. Wang, L. Zhu, C. Huang, PAVS: A new privacy-preserving data aggregation scheme for vehicle sensing systems, Sensors 17 (3) (2017) 1–18.

[25] J. Wang, Y. Zhang, Y. Wang, X. Gu, Rprep: A robust and privacy-preserving reputation management scheme for pseudonym-enabled VANETs, Int. J. Distrib. Sens. N. 12 (3) (2016) 1–15.

[26] Q. Wu, J. Domingo-Ferrer, U. González-Nicolás, Balanced trustworthiness, safety, and privacy in vehicle-to-vehicle communications, IEEE Trans. Veh. Technol. 59 (2) (2009) 559–573.

[27] Y.M. Chen, Y.C. Wei, A beacon-based trust management system for enhancing user centric location privacy in VANETs, J. Commun. Netw. 15 (2) (2013) 153–163.

[28] H. Hu, R. Lu, C. Huang, Z. Zhang, Tripsense: A trust-based vehicular platoon crowdsensing scheme with privacy preservation in vanets, Sensors 16 (6) (2016) 1–17.

[29] H. Hu, R. Lu, C. Huang, Z. Zhang, PTRS: A privacy-preserving trust-based relay selection scheme in VANETs, Peer Peer Netw. Appl. 10 (5) (2017) 1204–1218.

[30] Z. Liu, F. Huang, J. Weng, K. Cao, Y. Miao, J. Guo, Y. Wu, BTMPP: Balancing trust management and privacy preservation for emergency message dissemination in vehicular networks, IEEE Internet Things J. 8 (7) (2021) 5386–5407.

[31] K. Wang, C. -M. Chen, Z. Tie, M. Shojafar, S. Kumar and S. Kumari, "Forward Privacy Preservation in IoT-Enabled Healthcare Systems," in IEEE Transactions on Industrial Informatics, vol. 18, no. 3, pp. 1991-1999, March 2022.

[32] D. X. Song, D. Wagner, A. Perrig, Practical techniques for searches on encrypted data, in: Proceedings of IEEE Symposium on Security and Privacy, 2000. S&P 2000., 2000, pp. 44–55.

[33] Hamza, Rafik & Muhammad, Khan & Bellavista, Paolo & Titouna, Faiza. (2019). A privacy-preserving cryptosystem for IoT E-healthcare. Information Sciences. 527.

[34] Wei, Z., Wu, Y., Yang, Y., Yan, Z., Pei, Q., Xie, Y., and Weng, J. (2018). Auto-privacy: Automatic privacy protection and tagging suggestion for mobile social photo. Computers & Security.

[35] Natarajan, Bhalaji & Abilashkumar, P. & Aboorva, S. (2020). A Blockchain Based Approach for Privacy Preservation in Healthcare IoT.

[36] H. F. Durrant-Whyte, “Sensor models and multisensor integration,” International Journal of Robotics Research, vol. 7, no. 6, pp. 97–113, 1988.

[37] R. C. Luo, C.-C. Yih, and K. L. Su, “Multisensor fusion and integration: approaches, applications, and future research directions,” IEEE Sensors Journal, vol. 2, no. 2, pp. 107–119, 2002.

[38] K. K. kumar, E. Ramaraj and D. N. V. S. L. S. Indira, "Data Fusion Method and Internet of Things (IoT) for Smart City Application," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 284-289.

[39] Dautov, Rustem & Distefano, Salvatore & Buyya, Rajkumaar. (2019). Hierarchical data fusion for Smart Healthcare. Journal of Big Data.

[40] Dimitrov DV. Medical Internet of Things and Big Data in healthcare. Healthcare Inform Res. 2016;22(3):156–63.

[41] H. Lin, S. Garg, J. Hu, X. Wang, M. Jalil Piran and M. S. Hossain, "Privacy-Enhanced Data Fusion for COVID-19 Applications in Intelligent Internet of Medical Things," in IEEE Internet of Things Journal, vol. 8, no. 21, pp. 15683-15693, 1 Nov.1, 2021.

[42] W. Tang, J. Ren, K. Deng, and Y. Zhang, “Secure Data Aggregation of Lightweight E-Healthcare IoT Devices with Fair Incentives,” IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8714-8726, 2019.

[43] Z. Guan, Y. Zhang, L. Wu, J. Wu, J. Li, Y. Ma, and J. Hu, “APPA: An anonymous and privacy preserving data aggregation scheme for fog enhanced IoT,” Journal of Network and Computer Applications, vol. 125, pp. 82-92, 2019.

[44] Costel, Bogdan & Pop, Florin & Mihaita, Alexandra & Dobre, Ciprian & Castiglione, Aniello. (2019). Data fusion technique in SPIDER Peer-to-Peer networks in smart cities for security enhancements. Information Sciences. 479. 10.1016/j.ins.2018.06.070.

[45] Yang, Linfu & Liu, Bin. (2019). Temporal Data Fusion at the Edge. 10.1109/IUCC/DSCI/SmartCNS.2019.00031.

[46] El Faouzi, Nour-Eddin & Leung, Henry & Kurian, Ajeesh. (2011). Data fusion in intelligent transportation systems: Progress and challenges – A survey. Information Fusion. 12. 4-10. 10.1016/j.inffus.2010.06.001.

[47] Dautov R, Distefano S, Bruneo D, Longo F, Merlino G, Puliafito A. Pushing intelligence to the edge with a stream processing architecture. In: 2017 IEEE international conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). 2017. p. 792–99.

[48] Dautov R, Distefano S. Distributed Data Fusion for the Internet of Things. In: International conference on parallel computing technologies. Berlin: Springer; 2017a. p. 427–32.

[49] Dautov R, Distefano S. Three-level hierarchical data fusion through the IoT, edge, and cloud computing. In: Proceedings of the 1st international conference on Internet of Things and machine learning. New York: ACM; 2017b. p. 1.

[50] X. Su, K. Fan and W. Shi, "Privacy-Preserving Distributed Data Fusion Based

on Attribute Protection," in IEEE Transactions on Industrial Informatics, vol. 15, no. 10, pp. 5765-5777, Oct. 2019, doi: 10.1109/TII.2019.2912175.

[51] James, A. P., & Dasarathy, B. V. (2014). Medical image fusion: A survey of the state of the art. Information fusion, 19, 4-19.

[52] Wang, P., Yang, L. T., Li, J., Chen, J., Hu, S. (2019). Data fusion in cyber-physical-social systems: State-of-the-art and perspectives. Information Fusion, 51, 42-57.