Conception et développement d'un système d'e-santé intelligent et sécurisé basé sur le raisonnement probabiliste et la technologie blockchain

Hossain Kordestani

HESAM UNIVERSITÉ

le cnam

ÉCOLE DOCTORALE Sciences des Métiers de l'Ingénieur
Le Centre d'études et de Recherche en Informatique et Communications

THÈSE

présentée par : **Hossain KORDESTANI**
soutenue le : **4 juin 2021**

pour obtenir le grade de : **Docteur d'HESAM Université**

préparée au : **Conservatoire national des arts et métiers**

Discipline : **Informatique**

Spécialité : **Informatique**

Design and Development of an Intelligent and Secure E-Health System Based on Probabilistic Reasoning and Blockchain Technology

THÈSE dirigée par :
[**Monsieur BARKAOUI Kamel**] Professeur, Le CNAM

Jury

Mme. Elisabeth METAIS	Professeur, CÉDRIC, Le CNAM	Présidente du jury
M. Mohand-Saïd HACID	Professeur, Département Informatique, University Claude Bernard Lyon 1	Rapporteur
M. François CHAROY	Professeur, LORIA Campus Scientifique, Université de Lorraine	Rapporteur
M. Layth SLIMAN	Enseignant-chercheur, Ecole d'Ingénieurs Informatique - Efrei Paris	Examinateur
M. Abdelghani CHIBANI	Maître de conférences, LISSI, UPEC	Examinateur
M. Wagdy ZAHRAN	Président de l'entreprise Maidis	Invité

THÈSE

Je dédie cette thèse à trois femmes qui ont déterminé ma vie :

À la mémoire de ma mère,

à l'amour de ma femme et

à le sourire de ma fille.

À mon père : mon premier professeur,

et à mes frères : les cinq colonnes de ma vie.

Remerciements

Tout d'abord, je tiens à exprimer ma profonde gratitude à l'entreprise Maidis, qui m'a permis de réaliser cette thèse CIFRE. Je tiens particulièrement à remercier Dr Wagdy ZAHRAN, le président, qui a soutenu ce projet et qui m'a fait confiance. Ces années partagées ont contribué à faire évoluer le sujet au gré de nos échanges, de l'expérience et de l'analyse du terrain, mais aussi de mes motivations.

Je remercie bien sûr mon directeur de thèse et mon directeur de thèse, à commencer par Pr. Kamel BARKAOUI pour sa confiance tout au long du projet. Je remercie également Dr. Abdelghani CHIBANI pour son soutien, ses idées, ses conseils et pour les confrontations très intéressantes entre les deux mondes de la recherche et de l'industrie.

Je remercie également les membres du jury, à commencer par Pr. Elisabeth METAIS pour avoir présidé la séance, je remercie également Pr. Mohand-Saïd HACID et Dr. François CHAROY qui en étaient les rapporteurs. Tous deux ont permis d'apporter des réflexions intéressantes.

Je n'oublie pas tous ceux qui ont eu un rôle moins officiel, mais tout aussi important tout au long de ma vie et sans qui je n'en serais jamais arrivé là. Je tiens également à remercier toutes les personnes qui m'ont soutenu et encouragé durant ces trois années de thèse tant dans le cadre personnel (ma famille, mon ange gardien, mes amis) que dans le cadre professionnel. Sans ce soutien quotidien ou du moins fréquent, je n'aurais jamais réussi à arriver au bout de cette expérience enrichissante mais pleine d'embûches et de doutes.

Résumé

Avec la longévité et un taux croissant de personnes âgées dans de nombreux pays, une préoccupation croissante est de permettre à cette population vieillissante de se produire dans un environnement de convalescence permettant une meilleure qualité de vie avec des coûts réduits, ce qui exige l'utilisation de technologies modernes pour améliorer la qualité de vie des personnes en termes d'autonomie, de sécurité et de bien-être, en particulier dans le contexte de l'e-santé. L'e-santé peut comprendre quatre groupes de services : la télésurveillance, le diagnostic électronique, les soins à distance et la téléconsultation. Ils consistent essentiellement en une saisie de données basée sur l'IdO, une analyse intelligente des données, le stockage et la communication de données entre les utilisateurs dans des lieux éloignés. Le manque de fiabilité des données, l'incertitude des règles médicales et les différentes exigences des patients posent de nombreux défis pour la création d'un système intelligent d'e-santé. En outre, la criticité des données en transit et du service lui-même soulève de multiples préoccupations en matière de sécurité lors du déploiement de l'e-santé. À cette fin, dans cette thèse, nous abordons différents défis dans la conception d'un cadre intelligent et sécurisé pour l'e-santé.

Tout d'abord, nous avons proposé un cadre de télésoins auto-adaptatif. Ce cadre fournit un télémonitorage basé sur l'IdO avec une sélection intelligente d'actions de détection pour fournir une image holistique des patients avec un minimum d'intrusions. Un prototype de ce cadre a été développé et validé par des cliniciens et des ingénieurs de systèmes médicaux collaborant avec la société Maidis dans le cadre du projet ITEA3 Medolution. La performance du diagnostic probabiliste a été évaluée sur la base de deux ensembles de données publiques. Les résultats montrent que la solution de téléassistance que nous proposons est plus performante qu'un classificateur classique, en particulier lorsque plus de 40% des données sont manquantes. Le cadre proposé est également validé à l'aide de quatre scénarios. Les résultats de l'évaluation démontrent la capacité du cadre proposé à aider les patients et les médecins à diagnostiquer et à traiter les conditions et les épisodes médicaux.

Deuxièmement, nous avons proposé un cadre sécurisé pour traiter les questions de robustesse et de respect de la vie privée dans les réseaux IdO. Le cadre proposé est basé sur la technologie de la chaîne de blocs et du stockage distribué pour garantir l'intégrité et la disponibilité dans un réseau décentralisé. Le cadre proposé permet de préserver la confidentialité des données grâce à l'IA, ce qui permet à l'IA de crypter les données sans y donner accès. Le cadre proposé est validé à l'aide d'un cas d'utilisation de la télésurveillance; les résultats de l'évaluation décrivent le maintien de la confidentialité des données dans le cadre proposé.

Troisièmement, nous avons proposé un cadre sûr et robuste pour le partage des données dans le domaine de l'e-santé, ou plus généralement dans les réseaux IdO. Le cadre proposé utilise des algorithmes cryptographiques pour assurer un partage sécurisé des données sans révéler aucune donnée personnelle et en évitant les redondances indésirables. Les exigences de sécurité du partage et du stockage des données dans les réseaux IdO sont définies et leur respect est formellement prouvé dans le cadre proposé. En outre, le cadre proposé dans le cas de l'utilisation de la téléconsultation. Les résultats de l'évaluation montrent la force du cadre proposé pour fournir un cadre sûr et solide pour le partage des données dans le contexte de l'e-santé.

Mots-clés : E-santé, Télémonitorage, Téléconsultation, Sécurité, Chaîne de blocage, Intelligence artificielle, Raisonnement probabiliste, Raisonnement de sens commun, Chiffrement homomorphe, Chiffrement de diffusion, Vérification formelle.

Abstract

With longevity and a growing rate of the elderly in many countries, a rising concern is to enable this population aging to happen within a convalescent environment allowing better quality of lives with reduced costs, which demands using modern technologies to enhance people's quality of lives in terms of their autonomy, safety, and well-being, in the context of AmI, in particular e-health. E-Health can include four groups of services: telemonitoring, e-diagnosis, telecare, and teleconsultation. At their core, they consist of IoT-based data capture, intelligent analysis on data, storage, and communication of data among the users in remote locations. The unreliability of data, the uncertainty of medical rules, and different patient requirements pose many challenges in creating an intelligent e-health system. Moreover, the criticality of the data in transit and the service itself raises multiple security concerns in deploying e-health. To this end, in this thesis, we address the mentioned challenges in designing an intelligent and secure e-health framework.

Firstly, we have proposed a self-adaptive telecare framework. This framework provides IoT-based telemonitoring with a smart selection of sensing action to provide a holistic image of patients with minimal intrusions. A prototype of this framework has been developed and validated by clinicians and medical systems engineers collaborating with the Maidis company in the ITEA3 Medolution project. The probabilistic diagnosis performance has been evaluated based on two public datasets. The results show that our proposed telecare solution outperforms a classical classifier specifically when more than 40% of the data are missing. The proposed framework is also validated using four scenarios. The evaluation results demonstrate the proposed framework's ability to help patients and doctors diagnose and treat medical conditions and episodes. Secondly, we have proposed a secure framework for handling robustness and privacy issues in IoT networks. The proposed framework is based on blockchain technology and distributed storage to ensure integrity and availability in a decentralized network. The proposed framework enables privacy-preserving AI on the data, allowing AI to encrypt data without providing any access to the data. The proposed framework is validated using telemonitoring use case; the evaluation results depict maintaining the privacy of data in the proposed framework. Thirdly, we have proposed a secure and robust framework for data sharing in e-health, or generally in IoT networks. The proposed framework uses cryptographic algorithms to provide secure data sharing without revealing any personal data and while avoiding undesired redundancy. The security requirements of data sharing and storage in IoT networks are defined, and their fulfilments are formally proved in the proposed framework. Moreover, the proposed framework in the teleconsultation use case. The evaluation results show the proposed framework's strength to provide a secure and robust framework for data-sharing in the context of e-health.

Keywords : E-Health, Telemonitoring, Teleconsultation, Security, Blockchain, Artificial Intelligence, Probabilistic Reasoning, Commonsense Reasoning, Homomorphic Encryption, Broadcast Encryption, Formal Verification.

Contents

List of Tables

List of Figures

Chapter 1

Résumé Substantiel

Contents

1.1 Contexte de la recherche

Avec la longévité et l'augmentation du nombre de personnes âgées dans de nombreux pays, une préoccupation croissante est de permettre à cette population de vieillir dans un environnement de convalescence permettant une meilleure qualité de vie et une réduction des coûts [23]. Environ 85 % des personnes âgées souffrent d'au moins une maladie chronique, et environ 72 % de plusieurs maladies. Les populations âgées ont généralement besoin de soins à domicile et de soins médicaux [1] ; en conséquence, le coût des soins à domicile dans le monde a augmenté de 8 % par an, passant de 180 milliards de dollars en 2014 à 300 milliards de dollars en 2020 [2]. L'évolution susmentionnée de la population exige l'utilisation de technologies modernes pour améliorer la qualité de vie des personnes en termes d'autonomie, de sécurité et de bien-être, en particulier dans le contexte de l'intelligence ambiante (AmI). Pieper [79] a introduit le paradigme de l'AmI sur quatre concepts principaux :

- Embarqué : plusieurs dispositifs dédiés sont intégrés de manière invisible dans l'environnement.
- Personnalisé : le système est personnalisé en fonction des besoins de chaque personne.
- Adaptatif : il répond à l'utilisateur et à l'environnement
- Anticipatif : il anticipe les besoins de l'utilisateur sans médiation consciente.

De ce fait, le groupe consultatif sur les technologies des sociétés de l'information a examiné quatre scénarios concernant l'AmI ([31]) et a défini l'AmI comme "la vision de personnes entourées d'une interface intuitive intelligente intégrée à toutes sortes d'objets et d'environnements, capable de reconnaître la présence de différents individus et d'y répondre de manière invisible". Mahmood [61] a défini l'AmI comme des services intelligents qui peuvent reconnaître la présence de l'utilisateur, ses préférences, ajuster l'environnement intelligent pour répondre aux besoins de l'utilisateur, et a introduit des capacités de prise de décision intelligentes dans les réseaux IoT, le facteur critique de ces services.

Compte tenu des défis de santé mentionnés ci-dessus, nous nous concentrons sur les applications de l'AmI liées à la santé, en particulier la e-santé, tout au long de cette thèse. L'e-santé consiste à fournir des soins de santé aux patients à distance, ce qui peut améliorer leur expérience de santé à moindre coût. En raison du premier concept du paradigme AmI, l'embarqué, l'e-santé s'appuie fortement sur l'Internet des objets (IoT) pour fournir une image holistique du patient au système. Les trois autres concepts de l'AmI nécessitent un système intelligent basé sur l'intelligence artificielle (IA) pour traiter les données holistiques et fournir des services de santé personnalisés, adaptatifs et anticipatifs aux patients. Dans cette thèse, nous considérons les groupes de services suivants dans l'e-santé :

- Télésurveillance : consiste à fournir des évaluations fréquentes ou continues de l'état de santé des patients [48]. Maric et al. [62] ont défini la télésurveillance comme l'utilisation des technologies de l'information et de la communication pour surveiller et transférer des données relatives à l'état de santé des patients entre des individus séparés. Dans le contexte de l'AmI, le télémonitoring s'appuie sur les réseaux IoT pour une collecte non intrusive et précise des données.

- E-Diagnostique : fait référence à l'utilisation de la technologie pour analyser les données des patients en vue d'un diagnostic. Il peut s'agir d'une technologie indépendante ou en coopération avec un professionnel de santé. Dans les deux cas, l'e-diagnostic peut être considéré dans la lignée du télémonitoring, avec un service supplémentaire d'analyse des données. Afin d'analyser les données capturées dans le cadre de la télésurveillance, l'e-diagnostic s'appuie sur l'IA pour fournir des soins anticipés et personnalisés.

- Téléassistance : désigne l'utilisation de la technologie pour permettre aux patients de recevoir une assistance sanitaire à leur domicile. Dans le contexte de l'AmI, le téléassistance est un e-diagnostic associé à des soins de santé. Il consiste en une télésurveillance pour la capture des données et en un e-diagnostic pour l'analyse des données ; en outre, le télésoin fournit un traitement réactif basé sur l'e-diagnostic.

- Téléconsultation : fait référence à l'utilisation des technologies de communication pour fournir des consultations d'experts en soins de santé aux patients situés dans des lieux géographiquement différents. Elle peut être considérée comme faisant partie des télésoins, c'est-à-dire qu'elle permet de fournir des soins aux patients lorsque les systèmes intelligents ne sont pas suffisants. Pendant la téléconsultation, les experts en soins de santé peuvent utiliser la télésurveillance ou le télédiagnostic pour saisir les données des patients ou obtenir une aide au diagnostic.

D'autre part, l'e-santé peut être décomposée en quatre couches : (1) la capture des données : collecte des données des patients, par exemple à partir des réseaux IdO ; (2) le calcul : y compris l'analyse des données, l'e-diagnostic, la gestion des soins de santé, les personnalisations ; (3) la communication et le stockage : interaction entre les différentes entités de l'e-santé, par exemple les patients, les professionnels de santé, les modules de calcul et les dispositifs IdO ; (4) l'utilisateur : y compris les patients, les médecins, les soignants et les experts.

La télémédecine est une application des réseaux IdO en plein essor, qui désigne la pratique de la médecine et de la santé publique sur des patients ambulatoires. Dans cette pratique, l'objectif est de mettre en place un système informatique externe souvent sous la forme d'un portail web ou d'une application mobile connectée à un système hébergé dans le cloud, permettant un suivi et une surveillance continue de l'état de santé des patients tout au long de la journée, ce que l'on appelle dans ce contexte la surveillance médicale continue. Ce type de système doit assurer l'acquisition et le traitement des données mobiles du patient, l'utilisation de plusieurs appareils mobiles étant courante dans la vie quotidienne des individus. Le développement de ce type de système fait aujourd'hui l'objet de plusieurs travaux et projets et constitue encore un sérieux défi surtout lorsqu'il s'agit de patients à risque comme les personnes âgées qui souffrent souvent de plusieurs maladies chroniques pour lesquelles une supervision efficace doit être assurée. La majorité des approches de l'état de l'art s'appuient sur des mécanismes à base de règles pour le traitement d'événements complexes sachant que la majorité des sources d'événements sont des capteurs ou des formulaires issus d'applications utilisées par les patients eux-mêmes ou les professionnels de santé et qui servent à rapporter ces événements cliniques importants dans le cadre du suivi du patient.

Plusieurs contraintes sont posées lorsque (i) des applications mobiles sont utilisées, (ii) des capteurs communicants que le patient doit porter en permanence, ou (iii) des capteurs qui sont déployés dans son entourage pour assurer une surveillance continue du patient par un système complexe de traitement des événements. Les contraintes techniques et concernent notamment la fiabilité du système de supervision en raison de la limitation des ressources matérielles et logicielles utilisées, ce qui influence négativement la qualité des données collectées et la robustesse, la fiabilité et l'efficacité du système de supervision. En effet, les plateformes de capteurs ne disposent généralement pas de ressources suffisantes pour garantir une collecte synchronisée et continue des signes vitaux vers des centres médicaux distants. Ces contraintes entraînent donc un risque important de perte de données médicales pendant la collecte. De plus, les données subjectives concernant les signes et symptômes fournies par les patients sont basées sur leur interprétation personnelle qui varie souvent et dépend de l'état mental du patient et du contexte dans lequel le symptôme ou le signe en question est observé.

1.2 Problème de recherche

Compte tenu du contexte ci-dessus, dans ce manuscrit, nous nous concentrons sur le couche de calcul ainsi que sur le couche de communication et de stockage, afin de fournir un cadre d'e-santé doté d'une capture de données basée sur l'IdO. Explicitement, nous abordons le problème de recherche suivant dans ce manuscrit : “Concevoir un système d'e-santé intelligent et sécurisé”. Sur la base des couches discutées de la e-santé, nous pouvons décomposer l'énoncé du problème en problèmes de recherche (PR) détaillés suivants :

- Couche de calcul :

  PR1 Conception d'un cadre de téléassistance auto-adaptatif

  PR2 Concevoir et prouver les exigences de sécurité dans la gestion des calculs

- Couche de communication et de stockage:

  PR3 Conception d'exigences de robustesse en matière de communication et de stockage

  PR4 Conception des exigences de sécurité en matière de communication et de stockage

1.2.1 Conception d'un cadre de téléassistance auto-adaptatif

La téléassistance, composée de la télésurveillance, de l'e-diagnostic et du traitement, est l'essence de l'e-santé dans le contexte de l'AmI. Par conséquent, nous considérons les défis de recherche suivants :

- Quantité énorme de données brutes : si la totalité de la collecte était transférée au médecin et au soignant, le récepteur serait inondé de données pour la plupart inutiles.

- Manque de fiabilité des données : l'ingrédient principal d'un système de télésurveillance est la collecte de données ; cependant, la fiabilité de ces données n'est pas toujours assurée. La collecte des données étant effectuée par les patients ou leurs proches, qui ne sont pas des experts, les risques d'erreurs humaines sont élevés. De plus, les appareils des patients ne sont pas aussi bien entretenus que ceux des médecins, ce qui entraîne des mesures erronées. Dans l'ensemble, la collecte de données dans le cadre de la télésurveillance est sujette à un manque de fiabilité.

- Données incomplètes : pour un e-coaching raisonnable, des informations globales sur le patient sont nécessaires. Cependant, la collecte de ces données en dehors d'un établissement de santé n'est pas réalisable. Elle nécessite différents types d'appareils et la réponse à de nombreux questionnaires, ce qui n'est pas pratique pour un patient moyen à son domicile.

- Règles médicales probabilistes : les règles médicales sont, en général, extraites d'expériences médicales et d'analyses statistiques. Par conséquent, la plupart des règles médicales sont liées à des probabilités, ce qui rend difficile leur modélisation avec un raisonnement traditionnel.

- Personnalisation des produits : chaque patient a des goûts différents et réagit différemment à un accompagnement similaire. Il est donc impossible d'avoir un système d'e-coaching unique pour différents patients.

- Dynamisme des règles : le coaching électronique peut nécessiter un certain réglage avec le temps ; à mesure que le médecin et les experts obtiennent plus d'informations sur le patient, ils peuvent modifier le système de coaching électronique pour ce cas particulier.

- Modélisation des règles médicales : il existe de nombreuses règles médicales disponibles, mais leur modélisation dans notre système est un défi.

- Données hétérogènes : les données recueillies auprès du patient sont de différents types et formats. Il peut s'agir d'une réponse booléenne à un questionnaire, par exemple "Vous sentez-vous secoué ?" ou de connaissances complexes provenant d'une source externe, par exemple le comportement du module de reconnaissance du comportement. Tirer le meilleur parti de ces données hétérogènes est un défi.

1.2.2 Concevoir et prouver les exigences de confidentialité dans la gestion des calculs

Dans l'e-santé, la couche de calcul, en particulier l'IA, fournit les trois concepts du paradigme AmI, à savoir, personnalisé, adaptatif et anticipatif. Les informations médicales des patients utilisées dans le calcul sont privées et doivent rester confidentielles. Cependant, la confidentialité du calcul dans le calcul, en particulier dans une unité de calcul non fiable, est une exigence souvent négligée. Par conséquent, nous considérons les défis de recherche suivants :

- Problèmes de confidentialité : il convient de définir les problèmes de confidentialité pour la gestion des calculs d'un réseau basé sur l'IdO.
- Cadre sécurisé : sur la base des préoccupations définies, il convient de concevoir un cadre sécurisé pour répondre aux préoccupations en matière de respect de la vie privée.
- Preuve : la preuve du respect des préoccupations en matière de confidentialité est nécessaire pour garantir la confidentialité du cadre.

1.2.3 Conception d'exigences de robustesse en matière de communication et de stockage

La robustesse désigne la tolérance aux perturbations de traitement sans affecter le corps fonctionnel d'un système. Dans le domaine de l'e-santé, en particulier de la téléassistance et de la téléconsultation, il est essentiel de maintenir la disponibilité des communications et du stockage. Les solutions actuelles d'e-santé se heurtent à un obstacle de taille : la centralisation, qui augmente la possibilité d'un point de défaillance unique et qui est sujette à des attaques contre la fiabilité et la disponibilité [50]. La décentralisation améliore la robustesse globale des systèmes de santé actuels, garantissant que les données médicales sont protégées contre les attaques malveillantes ou les pertes de données accidentelles [9]. Par ailleurs, nous considérons les défis de recherche suivants :

- Intégrité : les données communiquées et stockées dans le cadre de l'e-santé doivent rester intactes. Toute modification non détectée des données peut perturber les aspects fonctionnels du système.
- Disponibilité : les données et les services d'e-santé doivent rester disponibles pour éviter toute perturbation des services.

1.2.4 Conception des exigences de sécurité en matière de communication et de stockage

Les services d'e-santé s'appuient fortement sur les informations relatives à la santé, notamment les données des capteurs, les signes vitaux, les dossiers médicaux électroniques, les antécédents médicaux, les symptômes et les règles médicales des patients. Les informations de santé doivent impérativement être sécurisées, en particulier dans le cas de la téléassistance et de la téléconsultation, car elles reposent sur l'interaction d'utilisateurs situés dans des lieux différents via les canaux de communication. Par ailleurs, nous considérons les défis de recherche suivants :

- Problèmes de sécurité : les problèmes de sécurité dans le réseau basé sur l'IdO doivent être définis.
- Cadre sécurisé : sur la base des préoccupations définies, il convient de concevoir un cadre sécurisé pour répondre aux préoccupations en matière de sécurité.
- Preuve : la preuve de la satisfaction des préoccupations en matière de sécurité est nécessaire pour garantir la sécurité du cadre.

1.3 Sommaire des contributions

Ce manuscrit comprend trois contributions visant à relever les défis de recherche définis ci-dessus.

1.3.1 Un cadre intelligent de téléassistance auto-adaptatif avec apprentissage automatique et raisonnement logique

Les maladies chroniques ont créé l'un des plus grands défis de la santé publique, car elles sont la principale cause de morbidité et de mortalité [44], en particulier, pendant la pandémie de COVID-19 qui augmente la mortalité des patients atteints de maladies chroniques [75]. Cependant, la qualité de vie et l'espérance de vie des patients atteints de maladies chroniques peuvent être améliorées grâce aux connaissances existantes [105]. Néanmoins, étant donné le nombre élevé de patients atteints de maladies chroniques, leur prise en charge nécessiterait de nombreux efforts médicaux. C'est pourquoi l'utilisation de la téléassistance a été privilégiée. Téléassistance est un terme général qui combine le préfixe grec *tele*, signifiant à distance, et le mot *care*, que [64] ont défini comme l'utilisation des technologies de l'information et des télécommunications pour surveiller les patients et leur fournir des soins de santé à distance. La téléassistance peut être utilisée pour gérer des conditions médicales, y compris les événements aigus importants au cours des conditions médicales, qui sont appelés *épisodes*. [63]. Par exemple, un arrêt cardiaque est un événement de santé aigu qui peut survenir au cours d'un état pathologique cardiovasculaire ; autrement dit, la maladie cardiovasculaire est un état pathologique, tandis que l'arrêt cardiaque est un épisode. En outre, l'épisode hypoglycémique survient couramment chez les patients souffrant de diabète sucré et d'insuffisance rénale chronique.

En réponse à **PR1**, nous avons proposé un cadre de téléassistance auto-adaptatif. Ce cadre fournit une télésurveillance basée sur l'IdO avec une sélection intelligente des actions de détection pour fournir

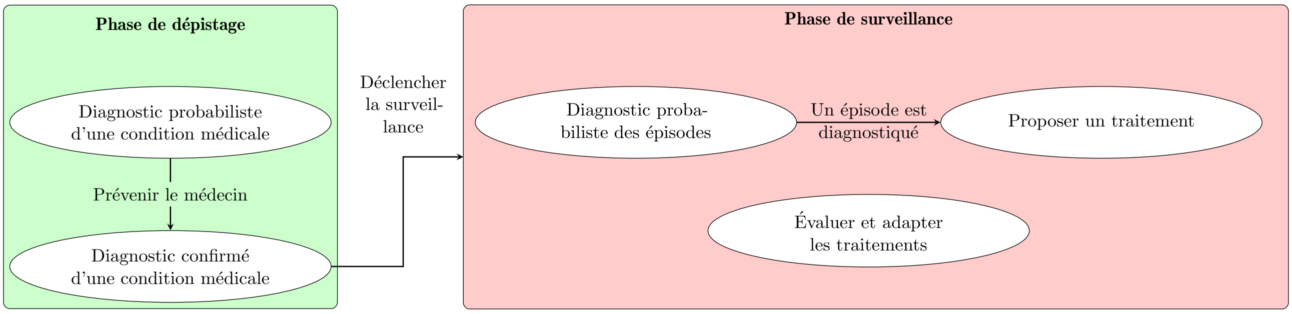

Phase de dépistage
Diagnostic probabiliste d'une condition médicale
Prévenir le médecin
Diagnostic confirmé d'une condition médicale
Déclencher la surveillance
Phase de surveillance
Diagnostic probabiliste des épisodes
Un épisode est diagnostiqué
Proposer un traitement
Évaluer et adapter les traitements

Figure 1.1: Aperçu général du cadre proposé pour la téléassistance auto-adaptative

une image holistique des patients avec un minimum d'intrusions. Il utilise le raisonnement basé sur l'ontologie et le diagnostic probabiliste pour gérer l'hétérogénéité et le manque de fiabilité des données, ainsi que l'incertitude des règles médicales pour fournir un e-diagnostic. Ce cadre intègre la programmation d'ensembles de réponses comme raisonnement de sens commun pour fournir des services de traitement auto-adaptatifs et auto-personnalisés. Les principales contributions du cadre proposé sont les suivantes :

- Un système de télésurveillance basé sur l'IdO qui prend en compte les dossiers médicaux pour un suivi holistique des patients atteints de maladies chroniques.
- Raisonnement basé sur l'ontologie pour obtenir des informations contextuelles.
- Un diagnostic probabiliste : un diagnostic assisté par ordinateur pour gérer les données manquantes, les données incertaines et les règles probabilistes.
- Un traitement auto-adaptatif : Un service de traitement basé sur Answer Set Programming (ASP) avec des règles facilement modifiables et s'adaptant automatiquement à chaque patient.

Comme le montre la figure 1.1, le cadre proposé fonctionne en deux phases pour chaque condition médicale : le dépistage et le surveillance. La condition médicale n'a pas été diagnostiquée pendant la première phase, et le cadre proposé permet de détecter les conditions médicales potentielles. En cas de découverte, le cadre proposé notifie le médecin. Avec ou sans la notification du cadre proposé, le médecin peut établir un diagnostic de l'état pathologique, puis prescrire un suivi pour certains épisodes spécifiques liés à l'état pathologique diagnostiqué. Par conséquent, la phase du cadre proposé passe du dépistage à la surveillance de la maladie diagnostiquée. Le cadre proposé assure la télésurveillance et le traitement auto-adaptatif ; dans cette phase, l'accent est mis sur le diagnostic et la réaction aux épisodes liés à l'état pathologique diagnostiqué.

Étant donné que les conditions médicales ne sont pas exclusives et qu'un patient peut souffrir de plusieurs conditions médicales, le cadre proposé peut être en phase de surveillance pour certaines conditions médicales et en phase de dépistage pour d'autres conditions médicales. Par exemple, dans le cas d'un patient diagnostiqué avec le diabète, le cadre proposé est dans la phase de surveillance du diabète et dans la phase de dépistage d'autres conditions médicales. Dans la première phase, il gère les épisodes liés au diabète, par exemple l'épisode hypoglycémique, tandis que dans la seconde phase, il diagnostique d'autres conditions médicales, par exemple les maladies rénales chroniques.

Pour les deux phases, le cadre proposé utilise des capteurs IdO, des questionnaires et des entrées manuelles pour capturer les données, ainsi qu'un raisonnement basé sur l'ontologie et un raisonnement probabiliste pour permettre un diagnostic probabiliste. Les épisodes sont par définition aigus et temporaires ; il est donc vital de les diagnostiquer en temps réel et de réagir en conséquence. D'autre part, les conditions médicales durent plus longtemps et sont généralement plus complexes à diagnostiquer et à réagir ; par conséquent, dans le cadre proposé, en cas de diagnostic d'une condition médicale, il notifie le médecin pour un traitement ultérieur.

1.3.2 Un cadre sécurisé basé sur la blockchain pour l'IA homomorphique dans les réseaux IdO.

Avec l'émergence des applications IdO, des problèmes de sécurité sont apparus, notamment dans le cas des systèmes centralisés. Les systèmes centralisés sont plus sujets à des problèmes de disponibilité ; dans les applications IdO, une quantité massive de données est produite à grande vitesse, ce qui pourrait dépasser le potentiel des serveurs centralisés. En outre, les données stockées dans les serveurs sont susceptibles d'être modifiées et supprimées ; en d'autres termes, leur intégrité est en danger. En outre, la plupart des applications basées sur l'IdO traitent des données sensibles ; par exemple, dans le domaine de l'e-santé, les données sont considérées comme privées et confidentielles. La confidentialité des données dans les applications IdO est donc une autre préoccupation.

La blockchain, une structure de données décentralisée et inviolable, peut être une solution pour gérer les problèmes de disponibilité et d'intégrité. L'architecture décentralisée de la blockchain évite le point de défaillance unique et étend intrinsèquement la disponibilité du service. En outre, grâce à l'utilisation d'algorithmes cryptographiques dans la blockchain, l'intégrité des données est garantie sur le plan informatique.

Bien que la technologie blockchain améliore la disponibilité et l'intégrité des applications IdO, elle n'est pas entièrement sécurisée en ce qui concerne la vie privée des nœuds [89]. Par conséquent, des solutions hors chaîne pour traiter les problèmes de confidentialité sont nécessaires. Une pratique courante consiste à utiliser des suites de chiffrement asymétrique à cette fin [32]. Le chiffrement asymétrique permet aux consommateurs de données d'accéder aux données, c'est-à-dire que les consommateurs de données sont des entités de confiance. Cependant, dans de nombreux cas, le propriétaire des données peut avoir besoin du calcul d'une entité non fiable, en particulier le fournisseur d'IA. Ce dernier est une nécessité pour une utilisation significative des données IdO ; cependant, comme les données sont précieuses pour les fournisseurs d'IA, on peut choisir de ne pas leur faire confiance. Traditionnellement, ce manque de confiance est traité au moyen d'accords de non-divulgation et de politiques de confidentialité, mais ils se sont avérés faibles dans la pratique. Le chiffrement homomorphe permet de calculer les données chiffrées sans avoir à utiliser les données brutes. Il pourrait donc être une solution pour traiter les problèmes de confidentialité dans le contexte des applications IdO. À notre connaissance, seules quelques études ont envisagé le chiffrement homomorphe pour les données IdO dans la blockchain, mais elles ne tiennent pas compte de l'intelligence artificielle dans le système. Par exemple, BeeKeeper 2.0 [111] propose un schéma de calcul d'externalisation décentralisé basé sur la blockchain Hyperledger Fabric[1].

En réponse à **PR2** et **PR3**, nous avons proposé un cadre sécurisé pour gérer les problèmes de robustesse et de confidentialité dans les réseaux e-santé ou, plus généralement, IdO. Dans le cadre proposé, nous nous concentrons sur l'IA en tant que calcul central des réseaux IdO. Le cadre proposé est basé sur la technologie blockchain et le stockage distribué pour assurer l'intégrité et la disponibilité dans un réseau décentralisé. Le cadre proposé utilise le cryptage homomorphique pour permettre un calcul préservant la confidentialité, en particulier les services d'IA. Les exigences de confidentialité des calculs basés sur l'IdO sont définies, et il est formellement prouvé que le cadre proposé y répond.

Nous avons évalué le cadre sécurisé proposé pour l'IA homomorphique en utilisant un cas d'utilisation dans le domaine de la santé. Le télémonitoring peut être une application du cadre proposé. Un service de télésurveillance typique contient les données médicales des patients et des modules d'IA pour les

[1] https://www.hyperledger.org/use/fabric

analyser. Comme les données médicales sont extrêmement confidentielles, le cadre sécurisé proposé pour l'IA homomorphe peut être bénéfique dans ce cas d'utilisation.

Les dispositifs IdO, par exemple les capteurs vestimentaires, les capteurs environnementaux et les téléphones mobiles, produisent des données dans les systèmes de télésurveillance. Ces données sont considérées comme privées, et il est vital de les garder en sécurité. En outre, ces données sont précieuses pour les entreprises d'IA car elles peuvent être utilisées pour améliorer leurs systèmes. Par conséquent, les patients doivent avoir la garantie d'utiliser un service de télésurveillance sans compromettre leurs données.

Dans le cadre sécurisé proposé pour l'IA homomorphique, les données IdO sont cryptées de manière homomorphique. Ces données peuvent être analysées en utilisant l'IA homomorphe sur les données cryptées sans les décrypter. Les résultats de l'IA sont également chiffrés, et seul le détenteur de la clé du schéma HE, c'est-à-dire le propriétaire des données, peut les déchiffrer. Par conséquent, le cadre proposé permet aux patients d'utiliser l'IA pour analyser leurs données médicales dans le cadre des services de télésurveillance, sans fournir leurs données en clair. De plus, les résultats de la télésurveillance restent également privés et ne sont accessibles qu'au patient concerné.

1.3.3 Un cadre sécurisé de partage des données basé sur la blockchain

Bien que la blockchain fournisse une plateforme distribuée pour le stockage et le partage des données, elle ne dispose pas de mesures suffisantes pour garantir la confidentialité des données. Une approche possible consiste à utiliser des blockchains à autorisation, par exemple Hyperledger Fabric, qui intègre un contrôle d'accès. Dans une blockchain à autorisation, l'accès à certains ou à tous les nœuds de la blockchain est limité par un contrôle d'accès. Cette approche est adaptée aux écosystèmes fermés, par exemple les réseaux internes des organisations. Cependant, la gestion du contrôle d'accès nécessite une autorité centralisée qui limite les avantages de l'utilisation de la blockchain. Un autre groupe d'approches consiste à étendre la blockchain publique sans permission avec un système de cryptage comme couche supplémentaire de sécurité. La technologie blockchain elle-même utilise des algorithmes de hachage sécurisés pour garantir l'immuabilité et l'intégrité. En outre, la couche supplémentaire utilise des algorithmes de cryptage pour assurer la protection de la vie privée et la confidentialité.

Certaines des solutions existantes adoptent l'approche bien établie dans les applications centralisées, c'est-à-dire le chiffrement asymétrique, pour l'utiliser dans la blockchain [29, 32]. Dans les applications centralisées, par exemple une application Web en ligne, l'utilisation du cryptage asymétrique est pratique, car chaque canal de communication est dédié à une seule application. En d'autres termes, les données cryptées sont destinées à un seul destinataire. D'autre part, dans un système distribué, plusieurs services peuvent exister sur le canal de communication, nécessitant les mêmes données. Dans ce cas, l'utilisation du cryptage asymétrique crée plusieurs cryptogrammes pour une seule donnée, ce qui entraîne une redondance des canaux de communication.

En réponse à **PR3** et **PR4**, nous avons proposé un cadre sécurisé et robuste pour le partage de données dans l'e-santé, ou plus généralement dans les réseaux IdO. Pour les exigences de robustesse du partage et du stockage des données, la technologie blockchain, le stockage distribué et la signature numérique sont utilisés dans le cadre proposé. En outre, le cadre proposé utilise une combinaison d'algorithmes cryptographiques pour fournir des données sécurisées tout en évitant la redondance indésirable. Il utilise le chiffrement homomorphique pour les requêtes d'informations préservant la vie privée, le chiffrement asymétrique pour la communication sécurisée entre deux personnes et le chiffrement de diffusion pour le partage sécurisé des données entre plusieurs personnes. Les exigences de sécurité du partage et du stockage des données dans les réseaux IdO sont définies, et leur respect est formellement prouvé dans le cadre proposé. En outre, le cadre proposé est évalué dans un cas d'utilisation de téléconsultation.

Afin d'illustrer les avantages du cadre sécurisé proposé pour le partage des données, nous examinons dans cet article un cas d'utilisation dans le domaine de la santé. La téléconsultation est complexe et comprend de nombreux membres qui interagissent et partagent des données entre eux ; c'est pourquoi une application de téléconsultation, appelée HapiChain[56], est présentée comme le cas d'utilisation du cadre proposé.

- Informations sur les médecins : les médecins peuvent décider de publier leurs informations, y compris leurs coordonnées et leurs disponibilités, publiquement ou de les partager uniquement avec leurs patients. Dans ce dernier cas, l'ACL pour les informations de chaque médecin inclut les patients de ce médecin, ce qui permet un contrôle d'accès et une confidentialité à grain fin. De plus, dans les deux cas, comme la disponibilité et l'intégrité font partie des exigences avérées du cadre proposé, il est garanti que les informations des médecins sont disponibles à tout moment et sans risque de modifications non autorisées.

- Informations médicales des patients : Les informations médicales des patients peuvent être partagées avec leurs médecins généralistes et récurrents, leur infirmière, l'hôpital visiteur et Hapicare. Le partage des données peut être total ou partiel, et également permanent ou temporaire. Par exemple, le patient peut préférer ne pas partager ses signes vitaux en permanence avec son médecin, mais avec Hapicare pour lui offrir un service de télésurveillance. En outre, au cours d'une téléconsultation, il peut souhaiter permettre au médecin d'accéder à ses signes vitaux pour les mesurer à distance. Le cadre proposé permet de partager ces informations en toute sécurité, conformément aux exigences susmentionnées. Le contrôle d'accès à grain fin permet tout type d'accès en fonction du type de données et du moment. En outre, les caractéristiques d'intégrité et de disponibilité du cadre proposé garantissent que les données des patients restent intactes et disponibles.

- Informations sur Hapicare : Outre les informations des médecins et des patients, les informations d'Hapicare comprennent les règles médicales et les rapports de suivi. Un médecin fournit les règles médicales pour un (groupe de) patient(s) spécifique(s) ; les règles médicales sont donc envoyées par les médecins à Hapicare, avec une sécurité garantie (confidentialité, intégrité et disponibilité) dans le cadre proposé. En outre, Hapicare génère des rapports de suivi qui sont partagés avec le patient, son infirmière et son médecin, avec une sécurité garantie dans le cadre proposé.

- Transactions : Les événements qui se produisent pendant la téléconsultation sont nécessaires pour les aspects financiers et juridiques. Le cadre proposé permet de disposer d'un journal immuable et disponible des transactions avec une sécurité garantie.

- Information sur les appels : bien qu'il soit possible d'utiliser le cadre proposé pour partager les données relatives aux appels vidéo, le partage direct des données sur les appels permet des performances bien supérieures. En outre, le contenu d'un appel peut devenir énorme à stocker dans le cadre proposé et n'est généralement pas utile. Par conséquent, il est préférable de partager uniquement les événements de l'appel, par exemple, le patient a commencé l'appel, le médecin a rejoint l'appel ou l'appel a été abandonné, en utilisant le cadre proposé.

1.4 Recommandations pour les recherches futures

Plusieurs domaines doivent encore être explorés pour parvenir à une solution complète d'e-santé dans le contexte de l'AmI. Les perspectives résultant de cette thèse peuvent être résumées comme suit :

- Évaluation de la téléassistance dans le monde réel : L'évaluation du cadre de la téléassistance dans l'environnement réel n'est pas simple ; une évaluation complète d'un système de téléassistance nécessite l'accès au suivi de patients du monde réel pour vérifier si la solution proposée est entièrement adaptée à leurs besoins. L'environnement susmentionné n'était pas accessible en raison de restrictions légales et éthiques. Par conséquent, un travail futur intéressant consiste à mettre en œuvre un pilote de téléassistance dans un environnement contrôlé. Ce pilote devrait être réalisé sous la supervision d'experts médicaux afin de modéliser diverses règles concernant les conditions et les épisodes médicaux, puis d'évaluer la surveillance, le diagnostic et le traitement à distance fournis par le cadre de téléassistance proposé.

- Délégation de privilèges dans le contrôle d'accès : La délégation de privilèges peut réduire la complexité et par conséquent améliorer la convivialité, l'évolutivité et la facilité de gestion des contrôles d'accès. À cette fin, c'est un défi inspirant de suivre la délégation des droits dans le cadre sécurisé de partage de données basé sur la blockchain.

- Mise en œuvre applicative du cadre sécurisé proposé pour l'IA homomorphique : La deuxième contribution de cette thèse aborde les problèmes de confidentialité dans l'externalisation des calculs d'IA dans le contexte de l'AmI, et elle est validée à l'aide de cas d'utilisation. Cependant, les avantages de ce cadre peuvent être augmentés après une étude de cas réelle. À cette fin, la mise en œuvre de ce cadre et son évaluation dans des scénarios du monde réel est une perspective applicative intéressante de cette thèse.

- Étude comparative du cadre proposé pour le partage sécurisé des données : La troisième contribution de cette thèse concerne le partage de données entre les participants des réseaux IdO. Une approche existante consiste à utiliser une blockchain à autorisation pour gérer le contrôle d'accès dans un réseau fermé. Bien que le cadre que nous proposons et la blockchain à autorisation soient destinés à des types d'utilisation différents, un travail futur intéressant serait d'étudier et de comparer la sécurité de ces deux approches.

- Attaques de la blockchain : La sécurité de la blockchain, était hors du champ de cette thèse. Il existe de nombreux types d'attaques dans les différentes couches de la blockchain, en particulier dans la couche applicative. Par conséquent, une étude future intéressante est de simuler ces attaques et d'analyser les implémentations du cadre proposé contre de telles attaques.

Chapter 2

Introduction

Contents

2.1 Research Context

With longevity and a growing rate of the elderly in many countries, a rising concern is to enable this population aging to happen within a convalescent environment allowing better quality of lives with reduced costs [23]. About 85% of elderlies suffer from at least one chronic condition, while around 72% from multiple ones. The older populations are commonly in need of home care and medical care [1]; accordingly, the cost of global home health care has grown with the rate of 8% per year, from $180 billion in 2014 to $300 billion in 2020 [2]. The aforementioned evolution of population demands using modern technologies to enhance people's quality of lives in terms of their autonomy, safety, and well-being, in particular in the context of Ambient Intelligence (AmI). Pieper [79] has introduced the paradigm of AmI on four main concepts:

- Embedded: multiple dedicated devices are invisibly built-in the environment
- Personalized: the system is customized for each person's need
- Adaptive: it responds to the user and environment
- Anticipatory: it anticipates the user's need without conscious mediation

Similarly, Information Societies Technology Advisory Group (ISTAG) has discussed four scenarios regarding AmI [31]; and defined AmI as "A vision of people surrounded by intelligent intuitive interface that are embedded in all kinds of objects and environment that is capable of recognizing and responds to the presence of different individuals in an invisible way." Mahmood [61] has defined AmI as intelligent services that can recognize the user's presence, preferences, adjust the smart environment to suit the user's needs, and has introduced intelligence decision-making capacities in IoT networks, the critical factor of such services.

Given the health mentioned above challenges, we focus on health-related applications of AmI, particularly e-health, throughout this thesis. E-health consists of delivering healthcare to patients at a remote distance, which can enhance their health experience at a lower cost. Because of the AmI paradigm's first concept, embedded, e-health strongly relies on the Internet of Things (IoT) to provide a holistic image of the patient to the system. The other three concepts of AmI require a smart system based on Artificial Intelligence (AI) to process the holistic data and provide personalized, adaptive, and anticipatory health services to the patients. In this thesis, we consider the following groups of services in e-health:

- Telemonitoring: refers to providing frequent or continuous assessments of the health situation of patients [48]. Maric et al. [62] have defined telemonitoring as the use of information and communication technologies to monitor and transfer data related to patients' health status between separated individuals. In the context of AmI, telemonitoring relies on IoT networks for non-intrusive and precise collection of data.

- E-Diagnosis: refers to the use of technology to analyze the patients' data for diagnosis. It can be an independent technology or in cooperation with a health professional. In either case, e-diagnosis can be seen in the longitude of telemonitoring, with additional data analysis service. In order to analyze the captured data in telemonitoring, e-diagnosis relies on AI to provide anticipatory and personalized care.

- Telecare: refers to the use of technology to enable patients to receive health assistance at their homes. In the context of AmI, telecare is e-diagnosis empowered with health care. It consists of telemonitoring for capturing data and e-diagnosis for analyzing data; additionally, telecare provides reactive treatment based on the e-diagnosis.

- Teleconsultation: refers to using communication technology to provide healthcare experts' consultations to the patients at geographically different locations. It might be considered across telecare, i.e., providing healthcare to the patients when smart systems are not sufficient. During teleconsultation, the healthcare experts might use telemonitoring or e-diagnosis to capture data from the patients or get diagnosis assistance.

On the other hand, e-health can be decomposed into four layers: (1) data capture: collecting patients' data, e.g., from IoT networks; (2) computation: including data analysis, e-diagnosis, healthcare management, personalizations; (3) communication and storage: interacting among various entities of e-health, e.g., patients, healthcare professionals, computation modules, and IoT devices; (4) user: including patients, doctors, caregivers, and experts.

2.2 Research Problem Statement

Given the context above, in this thesis, we focus on the computations and also communication and storage layers; in order to provide an e-health framework empowered with IoT-based data capture. Explicitly, we address the following research problem in this manuscript: “Designing an intelligent and secure e-health system.” Based on the discussed layers of e-health, we can decompose the problem statement into the following detailed research problems (RP):

- Computation Layer:

  RP1 Designing Self-adaptive Telecare Framework

  RP2 Designing and Proving Security Requirements in Computation Management

- Communication and Storage Layer:

  RP3 Designing Robustness Requirements in Communication and Storage

  RP4 Designing Security Requirements in Communication and Storage

2.2.1 Designing Self-adaptive Telecare Framework

Telecare, composed of telemonitoring, e-diagnosis, and treatment, is the essence of e-health in the context of AmI. Hence, we consider the following research challenges:

- Huge amount of raw data: if the entire collection was transferred to the doctor and the caregiver, the receiver would be flooded with mostly unuseful data.
- Unreliability of data: the primary ingredient of a telemonitoring system is data collection; however, the reliability of these data is not always assured. Since the patients or their relatives, who are not experts, do the data collection, it introduces high risks of human errors. Moreover, the patients’ devices are not well maintained the way the doctor’s devices, which results in erroneous measurements by the devices. Altogether, make the data collection in telemonitoring prone to unreliability.
- Incomplete data: for a reasonable e-coaching, holistic information about the patient is required. However, collecting such data outside a healthcare facility is not feasible. It needs various types

of devices and answers numerous questionnaires, which is not practical for an average patient at his/her home.

- Probabilistic medical rules: the medical rules are, in general, extracted from medical experiments and regarding statistical analysis. Hence, most of the medical rules are intertwined with probabilities, making it challenging to model with traditional reasoning.

- Personalization: each patient has a various taste and reacts differently to similar coaching. Hence, it is impossible to have a unique e-coaching system for various patients.

- Dynamicity of rules: the e-coaching might require some tuning with time; as the doctor and experts attain more information about the patient, they might alter the e-coaching system for that particular case.

- Modeling medical rules: there are many medical rules available, but modeling them into our system is challenging.

- Heterogeneous data: the data collected from the patient comes in different types and formats. The data can be a Boolean response to a questionnaire, e.g., “Do you feel shaky?” or intricate knowledge coming from an external source, e.g., Behavior from the Behavior recognition module. Making the most use of these heterogeneous data is a challenge.

2.2.2 Designing and Proving Privacy Requirements in Computation Management

In e-health, the computation layer, especially AI, provides the three concepts of the AmI paradigm, namely, personalized, adaptive, and anticipatory. The patients’ medical information used in the computation is private and needs to be kept confidential. However, the computation’s privacy in the computation, especially in an untrusted computation unit, is an often neglected requirement. Hence, we consider the following research challenges:

- Privacy concerns: the privacy concerns for the computation management of an IoT-based network should be defined.

- Secure framework: based on the defined concerns, a secure framework to fulfill the privacy concerns should be designed.

- Proof: the proof of meeting privacy concerns is required to guarantee the privacy of the framework.

2.2.3 Designing Robustness Requirements in Communication and Storage

Robustness refers to tolerance of handling perturbations without affecting the functional body of a system. In e-health, particularly in telecare and teleconsultation, it is vital to keep the communication and storage available. Current e-health solutions face a significant impediment in centralization, increasing the possibility of a single point of failure and prone to attacks against reliability and availability [50]. The decentralization improves the overall robustness of current healthcare systems, ensuring that medical data are protected from malicious attacks or accidental data loss [9]. Hence, we consider the following research challenges:

- Integrity: data in the communication and storage of e-health are required to stay intact. Any undetected changes of data might perturb the functional aspects of the system.
- Availability: data and services in e-health should stay available to avoid any perturbation of services.

2.2.4 Designing Security Requirements in Communication and Storage

E-health services strongly rely on health information, including patients' sensor data, vital signs, electronic health records, medical history, symptoms, and medical rules. Health information is critical to be kept secure, particularly in telecare and teleconsultation cases, as they are based on users in different locations interacting via the communication channels. Hence, we consider the following research challenges:

- Security concerns: the security concerns in the IoT-based network should be defined.
- Secure framework: based on the defined concerns, a secure framework to fulfill the security concerns should be designed.
- Proof: the proof of meeting security concerns is required to ensure the security of the framework.

2.3 Contributions Summary

This thesis includes three contributions to address the research challenges defined in Section 2.2.

2.3.1 An Intelligent Self-adaptive Telecare Framework with Machine Learning and Logical Reasoning

In response to **RP1**, discussed in Section 2.2.1, we have proposed a self-adaptive telecare framework. This framework provides IoT-based telemonitoring with a smart selection of sensing action to provide a holistic image of patients with minimal intrusions. It uses ontology-based reasoning and probabilistic diagnosis to handle the data's heterogeneity and unreliability, as well as the medical rules' uncertainty to provide e-diagnosis. This framework embeds answer set programming as common-sense reasoning to provide self-adaptive and auto-personalized treatment services. A prototype of this framework has been developed and validated by clinicians and medical systems engineers collaborating with the Maidis company in the ITEA3 Medolution project. Moreover, e-diagnosis is evaluated using a comparative experiment to illustrate its strength in missing information. Additionally, the proposed framework is validated using four scenarios.

2.3.2 A Blockchain-based Secure Framework for Homomorphic AI in IoT networks

In response to **RP2** and **RP3**, discussed in Sections 2.2.2 and 2.2.3, we have proposed a secure framework for handling robustness and privacy issues in e-health or, generally, IoT networks. In the proposed framework, we focus on AI as the core computation of IoT networks. The proposed framework is based on blockchain technology and distributed storage to ensure integrity and availability in a decentralized network. The proposed framework uses homomorphic encryption to enable privacy-preserving computation, in particular AI services. The privacy requirements of IoT-based computations are defined, and it is formally proved that the proposed framework meets them.

2.3.3 A Secure Data-sharing Framework Based on Blockchain

In response to **RP3** and **RP4**, discussed in Sections 2.2.3 and 2.2.4, we have proposed a secure and robust framework for data sharing in e-health, or generally in IoT networks. For the robustness requirements of data sharing and storage, blockchain technology, distributed storage, and digital signature are used in the proposed framework. Moreover, the proposed framework uses a combination of cryptographic algorithms for providing secure data while avoiding undesired redundancy. It uses homomorphic encryption for privacy-preserving queries of information, asymmetric encryption for secure one-to-one communication, and broadcast encryption for secure one-to-many data sharing. The security requirements of data sharing and storage in IoT networks are defined, and their fulfilments are formally proved in the proposed framework. Additionally, the proposed framework is evaluated in a teleconsultation use case.

2.4 Organization of the Thesis

The remaining of this thesis is organized into five chapters as follows. Chapter 3 provides the background knowledge required throughout the thesis. Chapter 4 gives an overview of existing works related to this thesis's problem statements. Chapter 5 presents the first contribution of this thesis, a self-adaptive telecare framework. First, we discuss the motivations of such a telecare framework and then present the architecture and the building blocks of the proposed framework. Then we provide the evaluation of the proposed framework. Chapter 6 delivers the second contribution of this thesis, a privacy-preserving framework for computation in IoT networks. First, we discuss the motivations and privacy requirements of such a framework and then introduce the architecture and the components of the proposed framework. Afterward, we formally evaluate the proposed framework against the privacy requirements. The chapter ends with a use case of the framework in telemonitoring. Chapter 7 describes the third contribution of this thesis, which is a secure framework for data sharing in IoT networks. First, we discuss the motivations and security requirements of data sharing and storage in IoT networks. Second, we propose a secure framework for data-sharing by giving out its architecture, components, and algorithms. Then, the workflow and dataflow in the proposed framework are detailed. Afterward, the security of the proposed is formally evaluated. Subsequently, a teleconsultation use case is used to validate the proposed framework in the context of e-health. Lastly, this thesis is concluded by recalling the contributions and discussing the possible future research directions in the closing chapter.

Chapter 3

Preliminaries

This chapter introduces the notations and terminologies that we will in this manuscript. We begin with the definitions of frequently used terms throughout this manuscript. Then we present the concept of uncertainty and approaches to handle it. Later we discuss ontology and answer set programming, which are the essence of the first contribution. After this, we introduce blockchain technologies and two encryption schemes, which are used in the second and third contributions.

Contents

3.1 Telecare

Telecare is a general term of combining the Greek prefix *tele*, meaning at a distance, and the word *care*, which Miller and O'Toole [64] have defined as the use of information and telecommunications technologies to monitor patients and deliver health care to them remotely. Telecare can be used for managing medical conditions, including the important acute events during medical conditions, which are called *episodes* [63].

For instance, a cardiac arrest is an acute health event that may occur during a cardiovascular medical condition; i.e., cardiovascular disease is a medical condition, while cardiac arrest is an episode. Moreover, hypoglycemic episode commonly occurs in patients suffering from Diabetes Mellitus (DM) and Chronic Kidney Disease (CKD).

3.2 Uncertainty

In the field of reasoning, uncertainty is classified into two categories: 1) *Aleatory* uncertainty is the intrinsic changing behavior, i.e., the observations differ in each experiment. 2) *Epistemic* uncertainty is rooted in the insufficiency of knowledge, i.e., principally, this type of uncertainty can be avoided with additional knowledge, and hence it is reducible [11, 53]. In the field of telecare, both types of uncertainties are possible. Faulty sensors, system failures, and human errors cause aleatory uncertainty while lacking enough information about patients, their medical files, and family history result in epistemic uncertainty. Since the medical rules are obtained during study and experiments, they are rarely absolute and deterministic. For instance, Table 3.1 depicts the features of hypercortisolism and their probabilities that Friedman [37] has provided. It shows that the medical rule for hypercortisolism diagnosis using its features would consist of probabilistic relationships, and hence, it would be an uncertain rule. Several approaches exist for handling uncertainty, such as BN [22], Dempster-Shafer (DS)[26, 86], and fuzzy logic; [109]; each of them is suitable for a specific purpose. DS is beneficial for gathering uncertain information from different sources and reasoning to a conclusion. However, fuzzy logic is a better fit when the states with low probability (membership values) are vital, e.g., diagnosing the early stage of a disease, since DS and BN mainly focus on the states with high probabilities. Verbert et al. [99] have thoroughly compared the DS and BN; the overall summary of their comparison is shown in Table 3.2.

Table 3.1: Features of hypercortisolism and their according probabilities [37].

Feature	**Percentage of Patients**
Fat redistribution	95
Menstrual irregularities	80
Thin skin and plethora	80
Moon facies	75
Increased appetite	75
Sleep disturbances	75
Nocturnal hyperarousal	75
Hypertension	75
Hypercholesterolemia and hypertriglyceridemia	70
Altered mentation	70
Diabetes mellitus and glucose intolerance	65
Striae	65
Hirsutism	65
Proximal muscle weakness	60
Psychological disturbances	50
Decreased libido and erectile dysfunction	50
Acne	45
Osteoporosis and pathological fractures	40
Easy bruisability	40
Poor wound healing	40
Virilization	20
Edema	20
Increased infections	10
Cataracts	5

3.2.1 Bayesian Network

Bayesian Network (BN) is a probabilistic graphical model via a directed acyclic graph. It embeds the conditional probabilities of various events and can predict their probability based on the evidence. In the graphical representation of a BN, each event is depicted as a vertex, and each edge shows a causal relationship between two events [22]. The relative dependence between two events is modeled into a conditional probability. For instance, if A and B be two binary events, and the event B causes event A to happen in one-fourth of times, this is depicted as $P(A|B) = 0.25$. A BN enables Bayesian inference, which deduces the probability of an event based on its causes and its effects as shown formally in the following equations; where e, c, and s respectively represent *Event*, *Cause*, and *effect* (Symptom) [22]. Equation 3.1 is the formal equation to calculate causal inference. Informally, based on the law of total probability, the probability of an event is the sum of the probability of its conjunction with each of its causes.

$$P(e) = \sum_{c_i \in causes(e)} \Big(P(e|c_i) \times P(c_i) \Big) \tag{3.1}$$

Equation 3.2 is the formal equation to calculate the inverse reference. Informally, based on Bayes' rule, the probability of an event regarding the observed effects is the ratio of the probability of their conjunction on the event; the numerator can be expanded to the product of the inverse conditional probability and the probability of the effect.

$$P(e|s) = \frac{P(s|e) \times P(e)}{P(s)} \tag{3.2}$$

3.2.2 Fuzzy Logic

Zadeh [109] has coined the term *fuzzy logic* to describe a logic where the truth values are not binary but a real number between 0 and 1 both inclusive. *Fuzzy sets* are the core of fuzzy logic, they are defined as an extension of classical sets, such that its members have a grade of membership. The formal definition of fuzzy sets is as follows:

Let X be the universal set, a *fuzzy set*A is characterized by a *membership function* $\mu_A : X \to [0, 1]$, i.e., each element of X is mapped to its truth value, often called *membership value.* The latter quantifies the grade of membership of an element in the fuzzy set A.

For example, if $HER = 80$ depicts a heart rate reading of 80 beats per minute; it can be argued that $\mu_{NormalHeartRate}(HER = 80) = 0.70$, $\mu_{Palpitation}(HER = 80) = 0.25$, and $\mu_{Bradycardia}(HER = 80) = 0.05$.

The fuzzy logic is used in three main steps: (1) fuzzification, (2) rules, and (3) de-fuzzification. In the first step, all the input values are mapped into fuzzy membership functions. Consequently, the fuzzy rules are executed over the fuzzified inputs. For example, the basic operators of conjunction and disjunction are replaced with their equivalent fuzzy logic operators as depicted in Equation 3.3 and Equation 3.4, respectively.

$$\mu_A(x) \wedge \mu_B(x) = min[\mu_A(x), \mu_B(x)], x \in X \tag{3.3}$$

$$\mu_A(x) \vee \mu_B(x) = max[\mu_A(x), \mu_B(x)], x \in X \tag{3.4}$$

Lastly, depending on the application of the fuzzy logic, the results are de-fuzzified. That is, instead of the use of fuzzy values, they are mapped into a crisp one based on their regarding membership values.

The fuzzy sets provide a formal approach to handle nonbinary states; hence fuzzy logic is beneficial for the inference and decision-making based on such data.

3.2.3 Dempster Shafer Theory

The Dempster-Shafer (DS) framework has been developed for handling imperfect information [26, 86]. The formal definition of this framework is as follows:

Let Ω be a set of all possible states of the system, its *power set* is denoted as 2^{Ω} includes all its subsets. A *mass function* is a function $m : 2^{\Omega} \rightarrow [0, 1]$ such that $\sum_{A \subseteq 2^{\Omega}} m(A) = 1$ and $m(\emptyset) = 0$. The mass function $m(A)$ expresses the percentage of all supporting evidence that the current state belongs to A but to no specific subset of A. The upper and lower limits of probability $P(A)$ can be obtained from the mass assignments. It is bounded by two ongoing nonadditive measures called *belief* ($bel(A)$) and *plausibility* ($pl(A)$). The former is defined as the collective mass of all its subsets; and the latter

Table 3.2: Comparison of DS and BN reasoning [99]

Feature	**BN**	**DS**
Fit for causal and diagnostic reasoning	+	-
Fit for information fusion	-	+
Fit for making decision	+	+
Inference coherence	+	-
Adaptable	+	+

as the collective mass of all overlapping subsets. They are formally defined as follows:

$$bel(A) = \sum_{B|B \subseteq A} m(B) \tag{3.5}$$

$$pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B) \tag{3.6}$$

Since in DS, the mass function considers all the evidence, it can conclude from various and even conflicting information. For example, if two different doctors believe their patient has a sickness S by the probability of %99; the calculation of mass function provides complete support of this diagnosis as $m(S) = bel(S) = 1$. All the aforementioned solutions have been widely used to handle uncertainty, and they can often be used interchangeably. However, based on the main focus of the application, one solution might be favored. DS is beneficial for gathering uncertain information from different sources and reasoning to a conclusion; on the other hand, BN is better for reasoning based on causes and effects. Notably, probabilistic reasoning can be considered as a type of fuzzy logic, in which the membership functions are their probabilities. However, fuzzy logic is a better fit when the states with low probability (membership values) are vital, e.g., diagnosing the early stage of a disease, since DS and BN mainly focus on the states with high probabilities. Verbert et al. [99] have thoroughly compared the DS and BN; the overall summary of their comparison is shown in Table 3.2.

3.3 Ontology

Borst [13] has defined ontology as a "formal specification of a shared conceptualization," thus allowing the formal depiction of information and their relationships [33]. In the medical field, ontology is attracting growing interest for formalizing and reasoning medical data. For example, Disease Ontology (DO) is an open-source ontology for biomedical data associated with human disease. Its vocabulary consists of 8,757 terms with unique maximal cross-references with other terminologies like National Cancer Institute Thesaurus and the National Drug File - Reference Terminology (NDF-RT) [52]. Systematized Nomenclature of Medicine (SNOMED) is a well-known general terminology widely used as a medical ontology with over 120,000 terms. *International Health Terminology Standards Development Organization* have freely provided Systematized Nomenclature of Medicine – Clinical Terms (SNOMED–CT). It includes four core components: 1) Concept Codes: numerical codes identifying clinical conditions organized in hierarchies, 2) Descriptions: text describing the concept codes, 3) Relationships between the concept codes, and 4) Reference Sets: limits and ranges for classification [34]. Figure 3.1 shows a sample visualization of concept *fever* in SNOMED–CT.

SNOMED-CT is exploited by a growing number of medical applications, including clinical decision support systems, electronic health records, e-Prescription, and health research. For instance, the National Board of Health and Welfare of Sweden has implemented medical alert information using SNOMED-CT, which involves documentation of patients' information regarding critical conditions, such as allergies and contagious disease [92]. Moreover, *Snow Owl MQ* is a big-data platform that allows grouping the patients with similar characteristics, inspecting their health records for trends and correlation, and statically analyzing them for verification of clinical hypotheses [98].

3.4 Answer Set Programming

Answer Set Programming (ASP) is a form of declarative programming that focuses on severe search problems. ASP represents knowledge using logical phrases, and it derives new knowledge using automated reasoning. The concept of ASP is to represent a particular computational problem through a logic program and find the solutions, called answer sets, for that problem using automated reasoning performed by an ASP solver. ASP syntax is derived from Prolog, and its semantics are described using stable model semantics introduced by Michael Gelfond and Vladimir Lifschitz [35]. An ASP rule consists of two main parts: (i) Head and (ii) Body; if the body of an ASP rule is true, the ASP solver

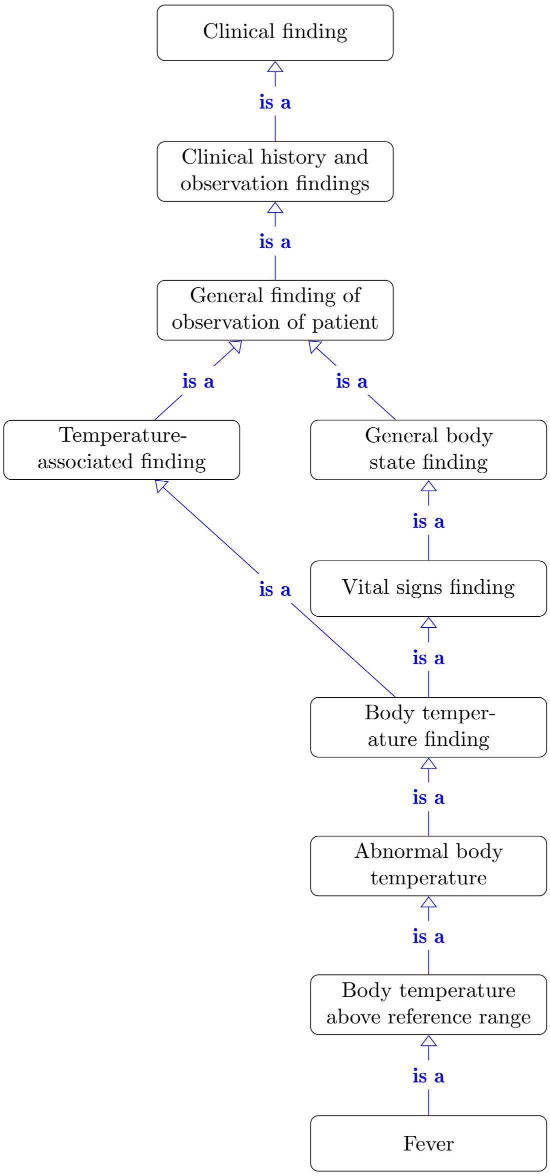

Clinical finding
is a
Clinical history and observation findings
is a
General finding of observation of patient
is a
is a
Temperature-associated finding
General body state finding
is a
Vital signs finding
is a
is a
Body temperature finding
is a
Abnormal body temperature
is a
Body temperature above reference range
is a
Fever

Figure 3.1: Visualization of concept *Fever* in SNOMED-CT [104]

Figure 3.2: ASP example

```
%Facts
condition("hypotension", 50, 5).
suggestion("hypotension", "eating", 50).
suggestion("hypotension", "takingpill", 40).

%Rule
simpletreatment(Episode, Action, T) :— condition(Episode,Pc,T),suggestion(Episode,Action,Ps).
```

concludes that the head of that rule is also true. An ASP rule is formalized as follows:

$$\begin{aligned} &Rule : Head \leftarrow Body. \\ &\qquad l \leftarrow b_1, ..., b_k, not\, b_{k+1}, ..., not\, b_{k+n} \quad (k, n \geq 0). \end{aligned} \tag{3.7}$$

where $b_1, ..., b_k, not\, b_{k+1}, ..., not\, b_{k+n}$ represents the rule's body and l represents its *head*. In ASP, a finite collection of ASP rules constructs an ASP program. Each rule in the ASP program can be seen as a limitation on answer sets of that ASP program. An answer set includes knowledge inferred using the reasoning on the ASP program. For instance, if an ASP program consists of the rule used in Equation 3.7 and its answer set includes all of $b_1, ..., b_k$ atoms and none of $b_{k+1}, ..., b_{k+n}$ atoms, it should also include l. An answer set, also called stable model, is minimal and justified and composed of ground atoms, which are atoms with no variables. The formal definition of answer set is as follows [59]: Suppose the program Π consists of ASP rules. Grounding is performed on the program Π to replace the variables used in the program with all the constants appearing in the program. S is a set of ground atoms obtained using grounding. A Reduct Π^S, with no negated atoms, is obtained using two main steps: (i) for each atom $a \in S$, drop rules with *not a* in their body, (ii) drop literals *not a* from all other rules. The minimal model of the reduct (Π^S) is the answer set S.

Consider the illustrative example depicted in Figure 3.2 with an ASP program Π with three facts and one rule, where a fact is a rule without body and with a single disjunct in the head. The predicate **condition(Episode,**`Pc`,`T`) represents the fact that there is a specific episode **Episode** with the probability `Pc` at the timestamp `T`. The predicate **suggestion(Episode,**`Action`,`Ps`) describes the fact that the suitable treatment for the specific episode **Episode** is the action `Action` with the probability `Ps`. The predicate **simpletreatment(Episode,** `Action`, `T`) depicts the fact that there is one possible treatment using the action `Action` at the timestamp `T` for the episode **Episode**. Where the **Episode** and timestamp are `"hypotension"` and `T`, respectively. Two actions as possible treatements are "eating" and "takingpill". The inferred information, **simpletreatment**("hypotension ", "eating", 5) and **simpletreatment**("hypotension", "takingpill", 5), are obtained using reasoning performed by answer set solvers.

The rich knowledge representation and efficient solvers are the main characteristics of ASP. Moreover, the non-monotonicity of ASP motivates us to use it for the treatment.

3.5 Blockchain

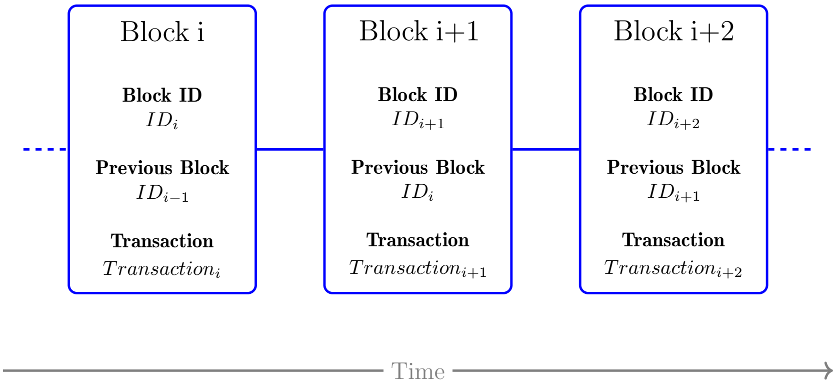


Figure 3.3: A schema of blockchain network

Blockchain belongs to Decentralized Ledger Technology (DLT) [71], which is a consensus of replication, share, and synchronization of data. The distribution can be completely decentralized without central data storage. Fundamentally, blockchain is a data structure formed by linked lists of blocks, chains of blocks. Each block stores a replication of data, as they all are distributed and shared across a peer-to-peer network. The connection between the blocks is via hash values and digital signature, which are one-way cryptographic functions; hence, modification of a block is only possible if all the blocks after that are modified [72]. Fig. 3.3 depicts a schema of blockchain, in which the ID_i is a cryptographic signature based on the contents of ith block; i.e., any change in the contents of this block would invalidate that signature. The most popular application of blockchain is a peer-to-peer anonymous cash system, named BitCoin [**? ?**].

Although blockchain was around for several years, with BitCoin's success, the background technology has also gained more attention; in other words, researchers try to apply blockchain in various domains. However, in BitCoin, the transactions are simple cryptocurrency transfer. To use blockchain for more complex transactions in other domains, structured scripting for application development in blockchain is needed, which promotes the second generation of blockchain. In which, *smart contracts* have been introduced to enable using blockchain for workflows. Smart contracts can be defined as a program in which its execution is triggered once a transaction occurs. This feature is useful for enforcing required actions upon different transactions, e.g., enforcing tax-payment upon the purchase of a product or service.

3.5.1 Second-generation Blockchain

Ethereum [17] is a second-generation blockchain platform; which features smart contracts. In Ethereum, a modified version of the blockchain is used, i.e., it consists of state machines. Each transaction causes a change in the state. In other words, the next state depends on the data of the current state and the current transaction. For instance, in a finance system, the states are balance sheets of members of that system, and transactions are any activities that affect the states of the system. E.g., "Alice has 2 euros, and Bob has 5 euros" is a state, while "Bob transfers 1 euro to Alice" is a transaction which will change the state to the new state of "Alice has 3 euros, and Bob has 4 euros." In Ethereum, each script's execution would cost a fee depending on the complexity of the script referred to *gas*.

3.5.2 Permissioned Blockchain

Although Ethereum improves the blockchain platform to use more complex transactions, however, in Ethereum, similar to blockchain, all the nodes have a copy of the entire states; they are designed to be public; although they can be set up in a private network, any members of that network have access to node contents. They provide anonymity and transparency to the fullest, but the tradeoff is privacy and scalability; which makes them unsuitable for the application that a controlled transparency is required. Although this limitation can be mitigated with the installation of Ethereum in a private manner; it rests another limitation in its throughput, as it is limited to only 7 to 15 transactions per second [3]. To this end, another approach is private blockchain networks, commonly known as permissioned blockchain. The latter features protocols for authentication, authorization, and permission of actions. They often have central identity management and hence are not ideal for a very large number of nodes. However, their throughput can be more than 10,000 transactions per second [101], which is

Table 3.3: A simplified overview of comparison between Hyperledger Fabric versus Ethereum [101]

Criterion	**Ethereum**	**Hyperledger Fabric**
Type of membership	Permissionless and Permissioned	Only Permissioned
User Identifications	Decentralized Anonymous Users	Centralized, the nodes are known
Sybil attack protection	Using huge power required for computing proof-of-work	Using identity management
Latency	Poor-up to 1 hour	Depends of implementation-matter of milliseconds
Throughput	15 transactions per second	More than 10,000 with the existing implementations
Temporary forks	Possible (might lead to double-spending attacks)	Not pssible
Consensus finality	No	Yes
Consensus Protocol	Proof-of-Work[17]	Practical Byzantine Fault Tolerance[101]
Smart Contract Language	Solidity[17]	Go and Java [101]
Scalability[27]	Less scalable	More scalable

incomparable with the throughput of BitCoin and Ethereum. *Hyperledger Fabric* is an implementation of permissioned blockchain for running smart contracts, using familiar technologies [6]. For instance, the smart contracts in Hyperledger Fabric are called *chaincode*, and they can be written in Go[1] or Java languages[2]. Hyperledger Fabric is built on a modular architecture and allows scalable consensus mechanism which enhance the global scalibilty of the application use case. Table 3.3 presents an overview of differences between the Ethereum and Hyperledger Fabric. In which, *Sybil attack* is an attack wherein the attacker creates numerous fake identities to affect the system. Moreover, *consensus finality* is the affirmation that blocks in the blockchain are final and will not be revoked.

The Hyperledge Fabric is a permissioned network, the nodes are all identified and can have three roles, namely *Clients*, *Peers*, and *Ordering Service Nodes*, defined as follows [6]:

- *Clients* submit the proposal of transactions.
- *Peers* execute the proposals of transactions, and validate them. Peers keep blockchain ledgers. The latter contains the immutable records of transactions. Only a subset of peers, namely the

[1] https://golang.org/

[2] https://www.java.com/

endorsing peers (also known as endorsers) can execute the proposals. The committing peers (also known as committers) validate the transactions.

- *Ordering Service Nodes* (also known as orderers) collects the transactions that are approved by endorser and distribute them between the committers.

3.5.3 Security Characteristics of Blockchain

Weber et al. [103] have demonstrated the read availability of blockchain is typically high. On the other hand, blockchain guarantees data integrity leveraging cryptographic techniques, particularly, two following mechanisms [24]:

- A linked list of blocks: this structure enforces that the latest appended block should include the hash value of the proceeding block. Hence, any modification of the previous blocks invalidates all the subsequent blocks.

- Merkle tree structure: in this structure, each block holds a root hash of a Merkle tree of all the transactions. In Merkle-tree, each non-leaf node is the hash value of the concatenated values of its children nodes. Therefore, any modification on the logs of transactions causes a new hash value in the above layer, leading to a falsified root hash. Ensuring any modification is easily detectable.

  Moreover, Park et al. [78] have discussed that incremental construction of Merkle tree is $O(h)$, where h is the height of the tree; however, reconstructing the root of the tree would require up to $O(2^h)$ time or space complexity. The exponential complexity of modification of the Merkle tree shows that blockchain computationally guarantees integrity.

The confidentiality in the blockchain typically depends on the type of the blockchain. The public blockchains do not aim for confidentiality, and they focus on transparency; while, the permissioned blockchains follow confidentiality requirements using access controls.

3.5.4 Distributed File Systems

Blockchain is not a general-purpose technology [21]. One of this technology's limitations is scalability in extensive data storage, as storing them on-chain can grow very expensive; one of the well-established solutions is storing data off-chain and managing it on-chain [45]. Blockchain-based distributed file systems, e.g., InterPlanetary File System (IPFS)[8], can be classified into seven layers [47]:

1. Identity layer: this layer allows each node of the distributed file system to have a unique identification information.
2. Data layer: this layer allows organizing the file structure in the distributed file system.
3. Data-swap layer: this layer allows formulating the file-sharing strategy among the nodes.
4. Network layer: this layer allows discovering, establishing connections, and exchanging files among the nodes.
5. Routing layer: this layer allows each piece of files to be found and accessed by the nodes.
6. Consensus layer: this layer ensures the correctness in the ledger recording transactions and maintains the network's consistency.
7. Incentive layer: this layer allows reward/punishment mechanisms to encourage the nodes to be active and honest.

3.6 Homomorphic Encryption

Homomorphic encryption (HE) is a type of cryptographic scheme to address outsourcing computations' privacy issues. It can be traced back to the 1970s when Rivest et al. [84] have discussed privacy homomorphisms. A typical HE scheme consists of four procedures [89]:

- Key generation: in this procedure, the encryption is set up, and the related keys are generated.
- Encryption: in this procedure, the user encrypts his/her data using the generated key.
- Evaluation: in this procedure, the user submits his/her encrypted data and a function to the server and gets an encrypted result.
- Decryption: in this procedure, the user can decrypt the evaluation function's encrypted result.

HE allows the third party to execute (limited types of) operations on the cryptograms without decrypting them. This definition is presented in Equation 3.8, in which m_1 and m_2 are any two messages and E represents encryption function. If an encryption scheme satisfies this equation's condition, it is defined as homomorphic over some operator $\odot$.

$$\forall m_1, m_2 \quad : \quad Enc(m_1) \odot Enc(m_2) = Enc(m_1 \odot m_2) \tag{3.8}$$

Based on the limitation on the operator, HE schemes can be classified into three categories (generations):

- Partially HE: The first generation of HE handles one type of operation on the ciphertext. For instance, the Rivest et al. [85]'s asymmetric encryption scheme is homomorphic for multiplication. In other words, multiplication of two cryptogram yields to a cryptogram equivalent of the cryptogram of the multiplication of their plain text:

  $$\begin{aligned} c_i &= Enc_{RSA}(m_i) \\ \prod m_i &= Dec_{RSA}(\prod c_i) \end{aligned}$$

  Similarly, Paillier [76]'s encryption scheme is additionally homomorphic.

- Somewhat HE: The HE algorithms allowing both addition and multiplication. However, in somewhat HE, only a limited number of computation is allowed in contrast to partially HE, which allows an unlimited number of computations.

- Fully HE: The idea of Fully HE is to remove any computations' constraints, neither the type nor the times.

Shrestha and Kim [89] have generally described a generic HE's procedures as follows:

- $KeyGen(1^\lambda, \alpha)$: given security parameter λ and an auxiliary input α, $KeyGen$ function yields key triplets of private key, public key, and evaluation key (sk, pk, evk).

- $Encrypt(pk, m)$ given the message m and the public key pk, $Encrypt$ function allows encrypting it and yields to a crypogram $c \in C$.

- $Decrypt(sk, c)$ using the private key sk, the user might decrypt the ciphertext c to yield plaintext message m.

- $Evaluate(evk, C)$ given the set of all cryptograms C and the evaluation key, $Evaluate$ function yields a new cryptogram c^*.

3.7 Broadcast Encryption

Broadcast encryption [36] is a type of encryption scheme aiming to handle delivering encrypted contents, e.g., multimedia contents, over a broadcast channel with multiple receivers, while only authorized users, e.g., those who have paid the subscription fee, can decrypt it. Each receiver has a unique private key in this scheme, and the broadcaster has a dedicated key. Assume $R = r_1, r_2, ..., r_n$ is the set of receivers, if the broadcaster B decides a subset $S \subseteq R$ of receivers with whom share his/her content M. Hence any user in this subset should be able to decrypt it. In contrast, any member outside this subset should not access any information about M. A broadcast encryption (BE) scheme consists of three randomized algorithm [12]:

- *Setup*(t, n): given $t \in \mathbb{Z}$ is a security parameter and n is the number of receivers, *setup* function yields n private keys $d_1, ..., d_n$ and a broadcaster key T.
- *Encrypt*(S, T): this function yields a pair of header h and symmetric encryption key k. The message m is encrypted symmetrically using the key k, yielding a ciphertext C_m. The latter is sent in the broadcasting channel alongside h and S.
- *Decrypt*(S, d_i, h): If the peer is in the list of authorized receivers ($i \in S$), then *decrypt* function yields the message encryption key k. Given the encryption key, the receiver can decrypt the ciphertext C_m and achieve the message m.

Boneh and Silverberg [12] have proposed an efficient solution of BE using n-mulinear maps. The proposed BE scheme consists of no header, and the size of private keys is $O((\log t)^2)$.

Chapter 4

Related Works

In this chapter, we go through the existing studies related to the contributions of this manuscript. We classified the existing studies into four general categories: telecare, blockchain for IoT, homomorphic encryption for IoT, and homomorphic encryption in blockchain.

Contents

4.1 Telecare

Telecare has received massive attention; particularly after the emergence of IoT sensors, researchers tend to apply them for telecare and telemedicine. The telecare systems can be arguably divided into three groups in terms of functionality: (1) telemonitoring: collecting patient's information in real-time, (2) telemonitoring with alerts: in addition to telemonitoring, raising the alarm based on patients' situations, and (3) teletreatment: in addition to telemonitoring with alerts, suggesting treatments to patients based on their conditions. Most of the existing telecare systems are limited to telemonitoring systems with alerts.

Telecare systems are commonly targeted to the public, but some are specialized for patients with a specific type of chronic condition. For instance, Parati et al. [77], Glykas and Chytas [38], and Bucholc et al. [15] have focused on high blood pressure, asthma, and Alzheimer's disease, respectively. Parati et al. [77] have proposed a solution for patients to better control their condition and support the doctor for better follow-up. They point out that blood pressure telemonitoring allows improving the quality of lives of patients, improving the treatments, decreasing the face-to-face consultation sessions, and reducing the costs of healthcare [77]. While Glykas and Chytas [38] have introduced a web-based tool called AsthmaWeb. It accomplishes data gathering and monitoring to manage patients according to their personalized asthma action plan. Similarly, Bucholc et al. [15] have introduced a decision support system to predict the severity of Alzheimer's disease based on biological and clinical measures.

In recent years, many researchers have started working on *teletreatment* systems. For instance, Xu et al. [107] have proposed Cloud-MHMS, a monitoring system to help doctors diagnose patients' conditions better. In this system, patients' information is collected through their mobiles. Moreover, it uses process mining and alpha algorithm to propose a treatment plan based on similar patients' medical files. Even though Cloud-MHMS is useful for doctors' holistic diagnosis, it does not engage patients in their treatment. Likewise, Wong et al. [106] have focused on reducing adverse drug events in intensive care units. This study discusses the effectiveness of decision systems regarding their rate of overridden alerts. Another approach in teletreatment is using the smart environment to help patients. For instance, Loreti et al. [60] have proposed a framework intertwining a rule-based complex event processing with reactive event calculus to suggest a reaction based on patients' states.

Many studies focus on using IoT sensors and mobile for data collection; however, Habib et al. [42]

have designed a decision system based on wireless body sensor networks. As a subset of wireless sensor networks, the latter enables continuous monitoring of patients' vital signs, which is useful for patients in a critical state. In this study, the fuzzy inference system is used to calculate the weight of patients' risk. This study, like the mainstream of studies, overlooked personalization and customizability. On the contrary, Afzal et al. [5] have introduced a mechanism to personalize wellness recommendations, using contextual information, e.g., location and weather, along with the recommendations. In this study, the general health recommendations are personalized to provide the ones that are best suited, based on the requirements, interests, and demands of the user.

Similarly, Rahimi and Wang [81] have designed and implemented a framework to run a patient-specific clinical decision model. It enables collecting the patients' preferences and selecting one of the various options for them. This framework relies on decision trees, which allows a full customizable decision model; however, it requires patients to respond to multiple questionnaires, which is inconvenient for most patients.

4.2 Blockchain for IoT

The characteristics of blockchain, in particular immutability and decentralized, made it a suitable solution for IoT networks. Blockchain allows a secure sharing of resources immune to malicious tampering. Numerous works have studied blockchain for IoT, some of which are discussed in this section.

A group of studies altered blockchain networks to meet the limitations and requirements of IoT applications. IoT networks' main limitation is the computation power, and its most considered requirement is data security. Dwivedi et al. [32] have modified blockchain for the application of IoT devices. They have applied asymmetric encryption suites to design a decentralized platform for secure and efficient medical data transmission. They discuss their proposed architecture to withstand Denial-of-Service attacks, mining attacks, storage attacks, and dropping attacks. In their proposed architecture, the healthcare data is stored in cloud storage servers connected to an overlay network. The latter is a peer-to-peer network of IoT devices. In the overlay network, the nodes are authenticated using valid certificates. In this approach, before sending, the message is signed with the sender's private key and encrypted with the receiver's public key. Upon receipt, the message is decrypted with the receiver's private key and verified with the sender's public key. They have used smart contracts to evaluate whether an IoT reading is normal or abnormal; consequently, to send an alert in the latter case. Similarly, Dorri et al. [29] have proposed Lightweight Scalable Blockchain (LSB) based on overlay network for preserving security in the context of IoT requirements. They have optimized the algorithms for use in the IoT environment; they have opted for a lightweight consensus algorithm. In the overlay network, the nodes are grouped in clusters; and each cluster has an elected cluster head which allows managing the blockchain network. Transactions in LSB are secure using asymmetric encryption, digital signatures, and cryptographic hash functions. The LSB concept is explored in smart home [29] and smart vehicle [28] settings to demonstrate its resistance against common security attacks. Block4Forensic[18] is a lightweight blockchain infrastructure proposed for forensics services in smart vehicle environments. Public key infrastructure is used in the proposed blockchain. Moreover, a fragmented ledger is designed to store detailed information about the vehicle. Block4Forensic allows trustless, traceable, privacy-preserving forensics of smart vehicles for after accidents. Furthermore, Rahulamathavan et al. [82] have proposed a privacy-preserving blockchain architecture based on attribute-based encryption techniques. Attribute-based encryption scheme allows confidentiality and access control in single encryption [39]. This scheme utilizes distributed management nodes to overcome the need for a centralized access control server.

Another group of studies focuses on the data storage of IoT in the blockchain. A well-established architecture for data sharing using blockchain, including IoT-based applications, uses distributed file systems (see Section 3.5.4). Moreover, for data confidentiality, an encryption scheme is used on top of the blockchain architecture. For instance, Sharma et al. [87] have applied such architecture of healthcare, using IPFS, Ethereum, and AES as distributed file systems, blockchain infrastructure, and encryption schemes, respectively. Another approach is using asynchronous encryption schemes. For instance, Li et al. [58] have proposed a scheme for storing and protecting IoT data. In this scheme, certificateless cryptography is used for transaction security. This choice is to reduce the redundancies of traditional public key infrastructure. In certificateless cryptography, users establish their private keys using their secrets and partial private keys; the key generation center provides the latter. In the proposed scheme, the IoT data are stored distributedly using blockchain technology.

Another aspect of IoT network which can be handled using blockchain is access management. Novo [73] has proposed a blockchain-based access control system for arbitrating roles and permissions in IoT. The proposed architecture eliminates the need for a centralized access management system. It simplifies the process and minimizes the communication overheads using a single smart contract. In this proposed approach, the direct integration of blockchain technology into IoT devices is avoided to allow higher usability, given IoT devices' limited capabilities. Furthermore, BCTrust[43] is an authentication mechanism for wireless sensor networks based on blockchain technology. The proposed mechanism is suitable for environments with resource constraints. BCTrust utilizes Ethereum for the blockchain layer. Only a group of trustworthy nodes have access to write on the blockchain of this approach. The proposed approach presents a decentralized authentication system with a global vision as blockchain networks. Hence, the operations are realizable autonomously, transparently, and securely.

Artificial intelligence is a robust analytic tool for IoT networks; it can provide a scalable and accurate data analysis in real-time. Singh et al. [91] have reviewed the existing studies of big data analysis and computation load-balancing in IoT applications and classified them into four main categories: (1) Cloud analysis: big data analysis takes place in a single cloud server. The use of a centralized server limits the accuracy, speed, latency, and computational storage of such systems. (2) Fog analysis: data collection and load balancing occur in a distributed manner to analyze the IoT data. However, a central controller manages the fog intelligence, causing some issues such as resource management and scalability. (3) Edge analysis: the edge nodes complete the training while the cloud server completes the processing task. Feature extraction and data scaling are part of the edge's task to enable the cloud server's data analysis. (4) Device analysis: the nodes are connected to each other in a peer-to-peer network to provide data analysis on the data.

An example of load-balancing in big data analysis is BlockDeepNet[83], which is a blockchain-based platform for deep learning. They propose a collaborative deep learning paradigm, i.e., each IoT device employs deep learning over its data to prepare a local model. The smart contract allows collecting the parameters of local models; and also the global generated collaborative model. BlockDeepNet is based on a private blockchain and uses a lightweight consensus mechanism instead of the resource-intensive proof-of-work. They have pointed out that local deep learning might become too consuming for some IoT devices. In this case, they have suggested offloading the deep learning to the edge server using blockchain transactions.

4.3 Homomorphic Encryption for IoT

Conventional encryption systems might be incompetent for securing the intermediary services against privacy leakage [89]. Homomorphic Encryption (HE), described in Section 3.6, is a privacy-preserving encryption method. In an HE-enabled cloud system, the data are encrypted using a HE algorithm prior to storing on the cloud, and then for data retrieval, the query is also encrypted and sent to the cloud; the cloud server can execute a prediction algorithm to retrieve the query results without knowing the contents of the query nor the data [89].

Song et al. [93] have proposed using HE for providing security and privacy in the context of VANETs. The location and distance information is proposed to be encrypted using HE prior to comparison. The result of the latter does not reveal any information regarding the location data. Moreover, the location information is verified using secure multiparty computation schemes. The proposed approach can withstand the attacks on the privacy of the vehicles. Another work in the context of smart vehicles is [96], in which use of FHE is proposed to secure the location and identity privacy in the vehicle to grid networks. In the latter, electric vehicles transmit their identity, consumption pattern, parking, and charging spots to the power grid network. In this proposed approach, the aforementioned link is secured using FHE. Similarly, Rabieh et al. [80] have proposed a privacy-preserving route reporting scheme in smart vehicles environment using HE. In the proposed approach, each vehicle encrypts its route information using HE scheme and transmits the encrypted message to roadside units. Hence, the vehicles can report their future routes without leaking their private information.

The drone system is another scope of IoT, where HE can be used for security. Cheon et al. [20] have proposed a linearly homomorphic authenticate encryption (LinHAE) scheme for securing the ground control center of drones. LinHAE allows verification of the authenticity of the message as well as its confidentiality. It supports linear operation between the ciphertexts with fast encryption, evaluation, and verification procedures for the real-time controller. Hence, LinHAE provides security against attacks on the confidentiality and integrity of the messages.

Jiang et al. [49] have proposed a privacy-preserving authentication protocol based on HE. The proposed protocol allows the users to arbitrarily generate any number of authenticated identities resulting in full anonymity in an IoT network. The proposed protocol protects its users from being tracked by peer users, service providers, authentication servers, or any other infrastructure. However, in the cases of disputes or malicious activities, the attacker can be traced and identified using authentication servers which execute collaboration protocols. The proposed protocol is lightweight and suitable for IoT requirements.

4.4 Homomorphic Encryption in Blockchain

Blockchain provides secure distributed and decentralized ledgers; however, it is still not entirely secure regarding nodes' security and privacy [89]. HE can be applied in blockchain for additional security; however, a limited number of works have been done to integrate HE in the blockchain in the context of IoT, e.g., BeeKeeper 1.0 [110] and BeeKepper 2.0 [111]. Hence, we also included Engima[112] and Nebula[40] which utilize HE in the blockchain for other contexts, but their idea might be applicable for IoT.

4.4.1 BeeKeeper 2.0

Zhou et al. [110] have extended their work into BeeKeeper 2.0 [111], which is a decentralized outsourcing computation scheme based on blockchain. BeeKeeper 2.0 allows homomorphic computations on the data from IoT devices without obtaining any plaintext data of them. In BeeKeeper 2.0, the publicly verifiable information is stored in the blockchain; such data include verification keys, responses, encrypted numbers, and commitments of core-shares; hence, the validators in the network can verify them, which result in the credibility of the system.

In BeeKeeper 2.0, any peer of the blockchain can become a server of the IoT devices when required by the peer nodes and owner of the IoT devices. The data are stored in the blockchain, and peer nodes manage them distributedly. Devices and servers do not need to maintain large storage since encrypted data are recorded in the blockchain; moreover, the devices do not require high-performance power since servers perform the computation of data and validators perform the verification works.

Since blockchain is tamper-resistance, once data have been stored in the blockchain, it can be considered as immutable. Hence, the verification key, which is used for all verification algorithms of BeeKeeper 2.0, is stored in the blockchain to be immutable. Moreover, the validators verify the aforementioned key prior to storage to ensure its validity.

Hyperledger Fabric blockchain is used as the carrier of the BeeKeeper 2.0 system. Hyperledger

Fabric blockchain is used as the carrier of the BeeKeeper 2.0 system, and their evaluation using this blockchain depicts the servers can process the encrypted data up to 10-degree polynomial with an acceptable computation time.

4.4.2 Enigma

Enigma [112] is a peer-to-peer privacy-preserving decentralized computation platform using cryptographic algorithms, including HE. Enigma uses secure multi-party computation for computing data queries in a distributed way. Data is distributed among various nodes, and they can cooperatively compute without leaking any information to other nodes. Consequently, as no nodes hold the entire data, Enigma is private. Enigma avoids unnecessary replication, in contrast to blockchain networks. Because of HE in Enigma, the nodes can run computations on data without having access to the raw data.

Engima consists of an off-chain network connected to an existing blockchain. In the former network, private and intensive computations are performed, including storage, privacy-enforcing computation, and other heavy processing. Engima uses a distributed hash-table to store data, such that data are stored securely outside of the blockchain. In contrast, access-control protocols and references to the data are stored in the blockchain. This approach allows storing large amounts of encrypted data and performing computations while preserving both privacy and correctness.

Apart from distributed hash-table, two other types of databases are used in Engima: public ledger and multiparty computation. The former holds the history of transactions in an immutable manner which is stored in the blockchain. In the latter, the secret shares are distributed to computing nodes that store their shares locally, while only the owner can request the whole secret. Enigma uses HE for secure multiparty computation, ensuring a verifiable secret sharing scheme.

4.4.3 Nebula

Nebula [40] is a decentralized genomic data generation, sharing, and analysis platform which is based on blockchain and utilizes HE scheme. Since genomic data are beneficial in the recognition of diseases and the development of drugs, they are shared among the researchers. Nebula is proposed to eliminate centralized genomic data generation and intermediaries in order to avoid privacy leakages.

Nebular consists of two core services: Keep and Crunch. The former is a distributed storage system, enabling scalable storage of big data with high throughput data access and efficient data management. The latter is a workflow management engine enabling flexible creation and execution of data analysis pipelines.

Each genomic data is assigned with some survey responses which describe the contents of such data. Data owners encrypt their survey responses and genomic data with the public key and upload them in the Keep. Because the former is need to be queried by the data buyers, it is encrypted using HE. While the latter is encrypted using Advanced Encryption Standard (AES), a well-established symmetric encryption algorithm. Data buyers construct a query and encrypt it with the public key; the encrypted query is executed on the encrypted survey responses, resulting in an encrypted list of data owners and their addresses. The validator nodes decrypt the result and re-encrypt it with the buyer's key. The buyer can now directly communicate with the owner to agree on costs. The owner can grant data access using a smart contract. The validator nodes verify the integrity of the requested data with the comparison of hashes and re-encrypt the data with the buyer's public key. Lastly, the access permissions are registered in the blockchain, and the tokens are sent to the owner's wallet. The use of HE in Nebula enables the data buyers to query the existing information without revealing any information about them.

Chapter 5

An Intelligent Self-adaptive Telecare Framework with Machine Learning and Logical Reasoning

This chapter investigates the use of artificial intelligence alongside knowledge representation and reasoning. The proposed framework is discussed in the context of Telecare application with self-adaptive treatment.

Contents

5.1 Motivation

Chronic diseases have created one of the biggest challenges in public health, as they are the leading cause of morbidity and mortality [44], in particular, during the pandemic of COVID-19 that increases the mortality of patients with chronic diseases [75]. However, the quality of life and life expectancy of patients with chronic diseases can be improved using the existing knowledge [105]. Nevertheless, given a high number of patients with chronic diseases, managing them would require a lot of medical efforts. To this end, the use of telecare has been favored.

Studies have shown that telecare enables patients to feel safe and reassured. It also provides an opportunity for better treatment to the physicians [16, 25]. Predominantly, the technology is used to remove the physical barrier between the medical team and patients to enable treating patients at their homes. For example, Bhatti et al. [10] have focused on treating patients in a remote area. However, the research is trending to facilitate the treatment by providing useful suggestions to patients and comprehensive information to their doctors. Parati et al. [77] have shown that telecare can better control the patient's condition and support doctors to optimize the treatment and, consequently, decrease healthcare expenditure.

Using telecare with different sensors would result in the holistic monitoring of patients. Nevertheless, there are various pitfalls to avoid [16]. The essential principle in designing such systems is customizability; because each patient has a unique requirement, and a simplistic design might not be useful for all patients [94, 100]. Another overlooked problem of such systems is the uncertainty of collected data, which would hugely affect the use of that data [54]. Monitoring vital signs, activities, and any other aspects of human life and health patterns must consider the heterogeneity of the different source types producing observations and the uncertainty of the observations.

5.2 An Intelligent Self-adaptive Telecare Framework with Machine Learning and Logical Reasoning

5.2.1 General Overview

As depicted in Figure 5.1, the proposed framework operates in two phases for each medical condition: screening and monitoring. The medical condition has not been diagnosed during the former, and the

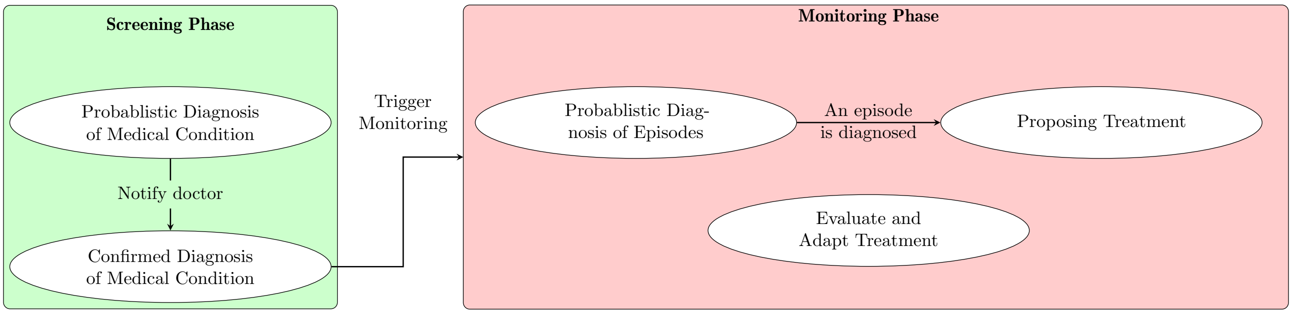

Screening Phase
Probablistic Diagnosis of Medical Condition
Notify doctor
Confirmed Diagnosis of Medical Condition
Trigger Monitoring
Monitoring Phase
Probablistic Diagnosis of Episodes
An episode is diagnosed
Proposing Treatment
Evaluate and Adapt Treatment

Figure 5.1: General overview of the proposed self-adaptive telecare framework

proposed framework allows detecting potential medical conditions. In the case of any finding, the proposed framework will notify the doctor. With or without the notification of the proposed framework, the doctor can establish a diagnosis of the medical condition; and then prescribes monitoring for some specific episodes related to the diagnosed medical condition. Consequently, the phase of the proposed framework is changed from screening to monitoring for the diagnosed medical condition. The proposed framework provides telemonitoring and self-adaptive treatment; in this phase, the focus is to diagnose and react to episodes related to the diagnosed medical condition.

Since medical conditions are not exclusive and a patient might suffer from multiple medical conditions, the proposed framework can be in the monitoring phase for some medical conditions and in the screening phase for other medical conditions. For instance, in the case of a patient diagnosed with DM, the proposed framework is in the monitoring phase of DM and the screening phase of other medical conditions. In the former phase, it manages episodes related to DM, e.g., hypoglycemic episode, while in the latter phase, it diagnoses other medical conditions, e.g., chronic kidney diseases.

For both phases, the proposed framework applies IoT sensors, questionnaires, and manual inputs for capturing data, and ontology-based reasoning and probabilistic reasoning for enabling probabilistic diagnosis. The episodes are by definition acute and temporary; therefore, it is vital to diagnose them in real-time and react accordingly. On the other hand, the medical conditions last longer and usually are more complex to diagnose and react; therefore, in the proposed framework, in the case of diagnosing a medical condition, it notifies the doctor for further treatment.

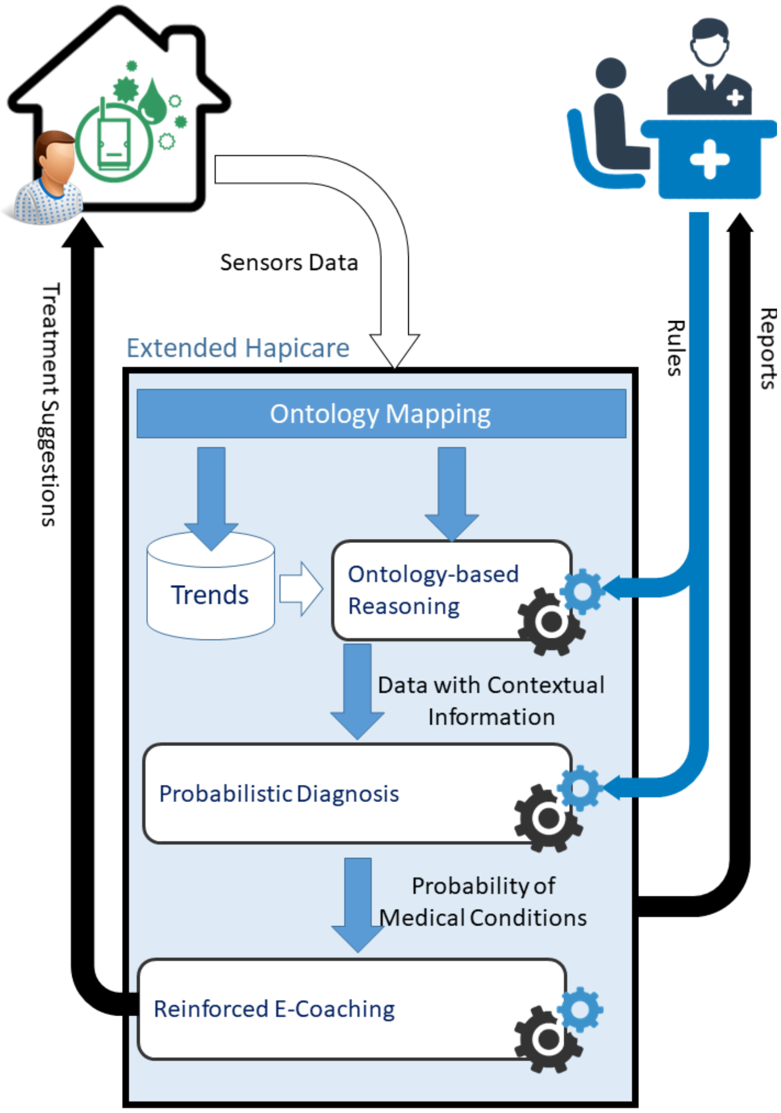


Figure 5.2: Overall architecture of the proposed self-adaptive telecare framework

5.2.2 Architecture

The proposed framework is designed in three primary layers taking advantage of three technologies. The first layer is ontology-based reasoning, which provides contextual information; the second layer is Bayesian reasoning, which provides probabilistic diagnosis; and the third one is the ASP layer, which provides the self-adaptive treatment. Figure 5.2 shows the overall architecture of the proposed framework.

5.2.3 Data

A broad range of data can be beneficial in diagnosing medical conditions/episodes; therefore, in the proposed framework, various types of data are considered. The data used in the proposed framework can be classified into the following four categories:

- Medical information: The medical information provided by the patient, including vital signs, e.g., body temperature.

- Medical file: The medical information provided by an expert; blood test results.

- Non-medical information: The information which is not considered health or medical information, but might be useful, e.g., room temperature.

- Deduced knowledge: The information that is not primarily fed to the proposed framework, but deduced from the analysis and reasoning on the data in the proposed framework.

Various sources might provide the aforementioned data. The source of data used in the proposed framework can be classified into four following categories:

- IoT Sensors: IoT sensors capture the information from the patient either continuously or on-demand, and then the captured information is transferred to Hapicare.

- Manual Input: For capturing the information without sensor, e.g., pain, the data are provided manually by the patient.

- External system: For the medical file, the information are stored in a health information system, which can be transferred to the proposed framework.

- Internal reasoning: The deduced knowledge is produced internally in the proposed framework; hence, the latter is the source of information resulting from reasoning on input data.

Any data, regardless of its type and source, are transmitted to the ontology-based reasoning component to be mapped to an ontology and then processed.

5.2.4 Ontology-based Reasoning

The data within the proposed framework are modeled in ontological terms for uniform depiction, i.e., the collected data are transformed into an ontology for further investigations. To this end, from the ontologies discussed in Section 3.3, we have opted for SNOMED-CT ontology due to its prevalence in research.

The data perceived from sensors are raw and sometimes meaningless on their own; hence, the ontology-based part processes the collected data and provides contextual information. In the proposed framework, trends and thresholds are applied for composing the context of patients. The threshold-based context depicts how the data compare with predefined thresholds, while the trend-based context shows how the collected data compare with the previous readings for the same user. Both types of context are necessary in order to model the health situation of a patient. For instance, a healthy weight is defined using a threshold. However, sudden weight gain or weight loss is a symptom of many medical conditions, which is modeled as a trend context.

In the proposed framework, we have implemented the ontology-based reasoning using JBoss Drools, which is a business rule management system with a rich feature set [7]. The rules used in this reasoning are based on ontological definitions of vital signs, medical conditions, and episodes by experts, i.e., doctors.

5.2.5 Probabilistic Diagnosis

Reliable treatment is not possible except with a reliable diagnosis and reasoning; in other words, telecare can only provide a suggestion for a successfully diagnosed condition. To this end, we have used probabilistic reasoning to diagnose medical conditions and episodes. It analyzes the collected data which have undergone ontology-based reasoning. The challenge for using these data is their unreliability, i.e., the collected data from lay patients in their homes are not reliable because they might mismeasure their vital signs or make a mistake in the manual input. Moreover, some manufacturers produce sensors for personal use only and not for high precision measurement. Albeit a patient might not notice a faulty sensor for some time. Hence, we cannot afford to throw away the unreliable data as they might include valuable information about patients. Another challenge is collecting all the required information about patients at their homes. It would require many sensors and even many questions that are not convenient for patients; hence, missing data should deprive a diagnosis.

Moreover, each medical condition/episode has a set of causes or risk factors that affect the probability of that medical condition/episode. It causes some changes in patients, including symptoms. Because BN is successfully applied for causal inference and prediction and estimation when the data are missing or unreliable [4]; it is selected for probabilistic reasoning for both phases of screening and monitoring in the proposed framework. Such that all undiagnosed medical conditions and the episodes of diagnosed medical conditions are each modeled in an individual BN, where the causes are modeled as parents and the symptoms as descendants of a medical condition/episode, to maintain the existent causal relation of features to the medical rule. Each medical condition/episode is modeled separately in order to avoid the complication of models and undesired interrelations between medical rules.

5.2.5.1 Creation of Bayesian Network

The creation of BNs has two steps: (1) creation of its structure, i.e., the shape of the graph; and (2) creation of the Conditional Probability Table (CPT), i.e., the probabilistic relationships between the nodes of the graph. These two steps can be performed as knowledge-driven, data-driven, or hybrid. In BN's knowledge-driven creation, both steps should be carried out by experts; who model the known probabilistic relationship between events (symptoms, causes, and medical conditions/episodes) and then use them to create a Bayesian network of each medical condition/episode. For example, the last row of the medical rule presented in Table 3.1, is modeled as $P(cataracts|hypercortisolism) = 0.05$. The prevalence of *cataracts* enables a backward inference for diagnosing *hypercortisolism* based on this conditional probability. In data-driven training, the BN is trained using data. In the hybrid training, experts provide the structure of the BN. At the same time, CPTs are extracted using data records of different patients regarding each medical condition/episode. In the proposed framework, data-driven training is not used in order to assure the expected structure that symptoms and causes of a medical condition/episode are respectively represented as descendants and parents of that condition in the BN. Therefore, in the proposed framework, the structure of BN is created by experts with respect to the description given in Section 5.2.5.2. Since the medical conditions are numerous and usually complex to be modeled by experts, the CPTs of BNs used in the screening phase are created using datasets. Moreover, as the accuracy of the diagnosis of episodes is critical, the CPTs of BNs used in the monitoring phase are created by experts (doctors) according to their experiences on the intended patient or other similar patients. In other words, the creation of BNs used in the screening phase and monitoring phase are, respectively, performed hybrid and knowledge-driven.

5.2.5.2 Bayesian Inference for Medical Diagnosis

As established in Section 3.2 diagnostic rules for medical conditions/episodes are a set of cause-effect rules. The modeling of medical conditions and episodes are performed similarly; for the sake of illustration, we depict the modeling of Acute Kidney Injury (AKI), which is one of the possible episodes of CKD [46]. AKI can dramatically increase the chance of mortality and morbidity [51]; however, Yang et al. [108] have discussed that AKI can remain undetected in the majority of cases (74 %). The cause-effect rules can be classified into three main categories:

- Immediate causes: Those are the events that affect the probability of a diagnosis in the short term. For example, hypotension and infection episodes increase the chance of AKI [51]. Hence these medical conditions are considered as immediate causes of AKI.
- Background causes: Those are the underlying events that affect the probability of a diagnosis in an extended period of time. For example, Friedman [37] has discussed that comorbidities are significant risk factors for AKI; i.e., patients with DM and heart diseases are more susceptible to AKI [51]. Hence these medical conditions are considered as background causes of AKI.
- Symptoms: Those are the effects of the medical condition/episode, i.e., the events whose probability is affected by the occurrence of a medical condition/episode. For instance, reduced body weight, irregular heart rate, and swelling are more plausible in AKI [37]

Figure 5.3 depicts simplified modeling of cause-effect relationships for diagnosis of AKI episodes, where the simple red arrow, doubled green arrow, and dotted blue arrow show the cause-effect relationships of immediate causes, background causes, and symptoms, respectively. Each cause-effect relationship can hold a probability, and each event can have a probabilistic value. Hence, Bayesian inference can deduce the probability of this episode with respect to any information on the other events. The second step in modeling an episode is the creation of the CPTs. For modeling medical conditions in BN, the creation of CPTs can be done using the datasets, i.e., training a BN with the given structure using the existing datasets. In the case of modeling episodes in BN, experts provide

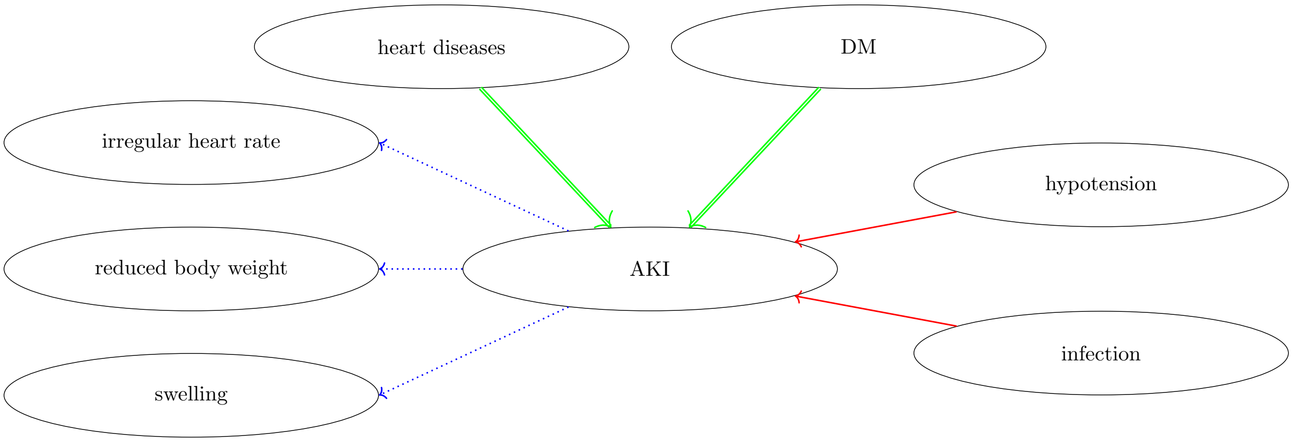


Figure 5.3: BN modeling of cause-effect relationships for diagnosis of AKI

CPTs. However, the expert might benefit from the explicit probabilities provided in the medical studies. For instance, regarding the background causes of AKI, Khadzhynov et al. [51] have reported that $P(heart\ failure \cap AKI) = 6.12\%$, $P(DM \cap AKI) = 8.81\%$, $P(heart\ failure) = 11.49\%$, and $P(DM) = 18.50\%$, which result in the following conditional probabilities:

$$P(AKI|heart\ failure) = 53.26\% \tag{5.1}$$

$$P(AKI|DM) = 47.62\% \tag{5.2}$$

In some cases, the existing medical rules do not include explicit probabilities; e.g., Lehman et al. [57] have reported that for each hour of severe hypotension, the probability of AKI increases by 22%. Both explicit and implicit conditional probabilities can help the experts in creating CPT of the BN of the medical condition/episode.

5.2.6 Self-adaptive Treatment

It is vital to help patients regarding their medical episodes to improve their quality of life. However, a common pitfall is uniforming the treatment services for all patients. Each patient has a different set of requirements, and even for one patient, the treatment might vary through time [100]. Hence, a treatment service should be developed that facilitates manual modification by doctors and automatic customization for patients' needs. To this end, in the proposed framework, we have implemented a self-adaptive treatment service using commonsense reasoning implemented by Answer Set Programming (ASP).

As discussed in Section 3.4, ASP finds the answers to achieve the defined objective, considering the facts and rules. The predefined facts include different episodes and their possible treatments; the input facts are the episodes with their associated probabilities. The experts, e.g., doctors and caregivers, provide the predefined facts, while the probabilistic diagnosis component computes the input facts. The objective is a state with no episodes with a probability above a threshold, which is a safe state in the patients' lives. In this study, self-adaptive treatment is formalized in ASP.

For the automatic customization of treatment, the ASP-solver is called in any diagnosis update to adapt the treatment service for patients. Each solver call results in answer sets that extract the following information: (i) the updated episodes and the most recent action, (ii) the possible actions at this point, and (iii) the obtained awards based on the updated episodes and the most recent action. When the system suggests the next action, and the patient performs this action, the system will observe the new state, i.e., the system monitors the episodes after performing the selected action. The system knowledge is then updated using changing the award of the selected action for that episode.

In ASP, a transition system $D_m = \{S_m, A, f_m\}$ is represented where S_m represents a set of states and always is discrete, A represents a set of actions, and f_m represents transition function, $f_m : S_m \times A \rightarrow S_m$ and always is deterministic. This kind of representation allows fast decision-making for treatment. The set of states is represented by predicates representing episodes that may change their true values at different times, such as **condition**(`"hypotension"`, `50`, `5`), where `5` is the time step of the predicate. Actions, A, are selected actions for the treatment based on the episodes and possible treatments. For instance, the predicate **selectedAct**(`"takingpill"`, `"hypotension"`, `5`) is used to represent that the action `"takingpill"` is selected as action at time step `5` for treating the episode `"hypotension"`. The main ASP rules used in the self-adaptive treatment are shown in Figure 5.4. The award value for a specific action is changed according to the effect of that action on episodes. After D time step of the conduction of the selected action for the treatment, observation is done to obtain updated episodes in order to update knowledge. The changing amount of award value in each iteration, **step**(`S`), is defined as a predefined fact, see line 1 in Figure 5.4, provided by the doctor. Line 3 is used to obtain all possible actions regarding the diagnosed episodes and their possible treatments, represented by **suggestion(Episode,** `Action`, `Ps`). Line 4 is used to illustrate that if the probability of the episode decreases with performing the specific action, the award value of that action should increase, i.e., the selected action has been suitable to mitigate the episode. Therefore, the ASP rules are updated after each iteration. On the contrary, line 5 is used

Figure 5.4: Main ASP rules for self-adaptive treatment

```
1  step(2).
2  timeStep(4).
3  possibleAction(Action, award(Action, Episode,Ps), T) :- condition(Episode,Pc,T), suggestion(Episode,Action,Ps).
4  possibleAction(Action, award(Action, Episode,NewValue), T+ D):-possibleAction(Action, award(Action, Episode, Value), T),
       condition(Episode, Pc, T), condition(Episode, PcNew, T+ D), PcNew<Pc, timeStep(D), NewValue=Value+S, step(S).
5  possibleAction(Action, award(Action, Episode,NewValue), T+ D):-possibleAction(Action, award(Action, Episode, Value), T),
       condition(Episode, Pc, T), condition(Episode, PcNew, T+ D), PcNew>Pc, timeStep(D), NewValue=Value-S, step(S).
6  1{max_sel_weight(X)}1 :- possibleAction(_, award(_,_, X), _), #max {V : possibleAction(Action, award(Action, Episode, V), T)} = X.
7  selectedAct(Action, Episode, T):-max_sel_weight(X), possibleAction(Action, award(Action, Episode, X), T).
8  #show selectedAct/3.
```

to illustrate that if the probability of episode increases with performing that action, the award value of that action should decrease, i.e., the selected action has been ineffective in mitigating the episode. Line 6 and line 7 are used to select an action with the maximum award value for the treatment. The rules show that in the proposed framework, self-adaptive treatment is online with choosing a suitable action, observing the consequences of that action, and changing the award value of that action.

5.2.7 Workflow

5.2.7.1 Workflow in Screening Phase

A typical workflow of screening phase in the proposed framework is presented in Figure 5.5. First, the data are collected from the patient, either directly from the IoT sensors or manually input by the patient or his/her caregiver. Later, the data are processed using ontology-based reasoning to yield contextual information. Afterward, all the data and their context are processed within the probabilistic diagnosis to estimate the probability of different medical conditions. For each medical condition, the *sensing action state* is defined as when the probability of that medical condition is between TH_{wary} and TH_{diag}, where the former is a predefined threshold for suspecting a medical condition, while the latter is a predefined threshold for diagnosing a medical condition. *Sensing action state* signifies that the medical condition is possible, but more evidence is required to confirm or reject this diagnosis. To this end, the proposed framework selects a cause/symptom related to that medical condition, where no recent information exists about that cause/symptom. Then the proposed framework requests the measurement of the selected cause/symptom. In order to minimize sensing actions, in the proposed framework, the most effective cause/symptom of that medical condition is selected, i.e., they are close

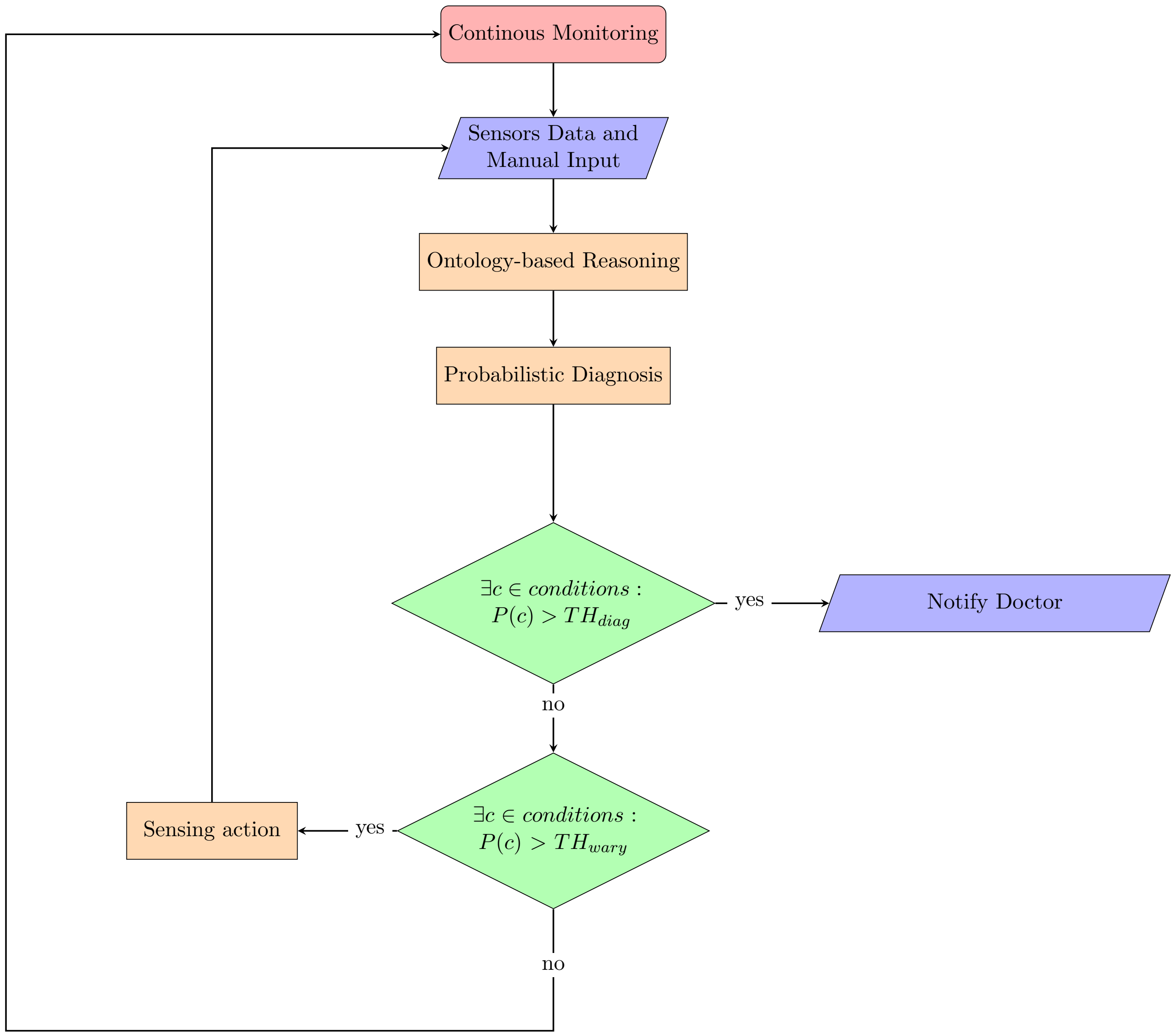

Continous Monitoring
Sensors Data and Manual Input
Ontology-based Reasoning
Probabilistic Diagnosis
$\exists c \in conditions : P(c) > TH_{diag}$
yes
Notify Doctor
no
$\exists c \in conditions : P(c) > TH_{wary}$
yes
Sensing action
no

Figure 5.5: Flowchart of screening phase in the proposed self-adaptive telecare framework

to being considered as *pathognomonic* or *sine qua non* causes/symptoms. The former case is used to confirm the presence of the medical condition, i.e., if the response is positive, it is most likely that the medical condition is present; however, the latter case is exploited to confirm the absence of the medical condition, i.e., if the response is positive, it is most likely that the medical condition is absent. This cycle continues until either the probabilities fall below TH_{wary}, which shows it was a false doubt, or one surpasses TH_{diag}, which signifies detection of a possible medical condition. In the former case, the flow of continuous telemonitoring continues, while in the latter case, the patient's doctor is notified.

If the doctor establishes the diagnosis of a medical condition, which can be as a result of a notification from the proposed framework; the doctor can prescribe telemonitoring for managing that medical condition. To this end, the doctor triggers the monitoring phase for the episodes related to the diagnosed medical condition.

5.2.7.2 Workflow in Monitoring Phase

A typical workflow of the monitoring phase in the proposed framework is presented in Figure 5.6. The data gathering and ontology-based and probabilistic reasoning parts are the same as those in the screening phase. The chance of occurrence of various episodes is low, as they are temporary and acute; on the other hand, the occurrence of multiple medical conditions is not rare, but many medical conditions are more susceptible in the presence of other medical conditions. Hence, in the *sensing action state* of the monitoring phase, differential diagnosis is performed; that is, the selection of sensing action is affected by all the susceptible episodes with their probabilities between TH_{wary} and TH_{diag}. Similar to the screening phase, when no episode is susceptible, the proposed framework pursues continuous telemonitoring. However, if one or multiple episodes are detected, i.e., their probabilities surpass TH_{diag}; these diagnosed episodes alongside their probabilities are forwarded to the self-adaptive treatment component to process them using ASP and propose a customized treatment to the patient. Furthermore, during this treatment, the patient's state is closely monitored to modify the patient's profile based on changes in the probabilities of episodes. If the suggested treatment were effective, the patient would start feeling better, and the probability of the episode would decrease; hence, the probability of the ASP rule related to that suggested treatment should increase. When the recommended treatment is not beneficial for the patient, the ASP reduces the recommended treatment probability.

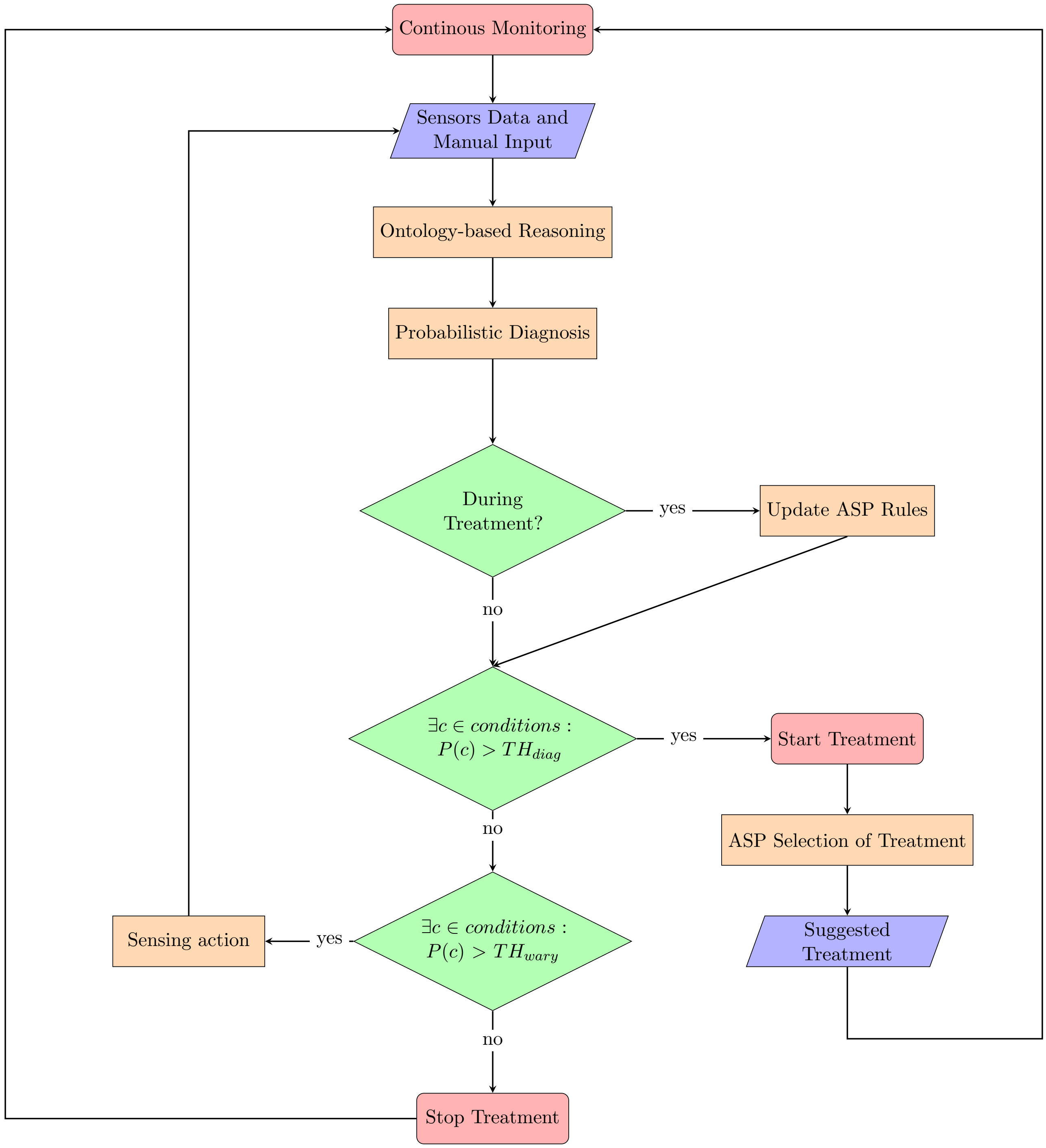
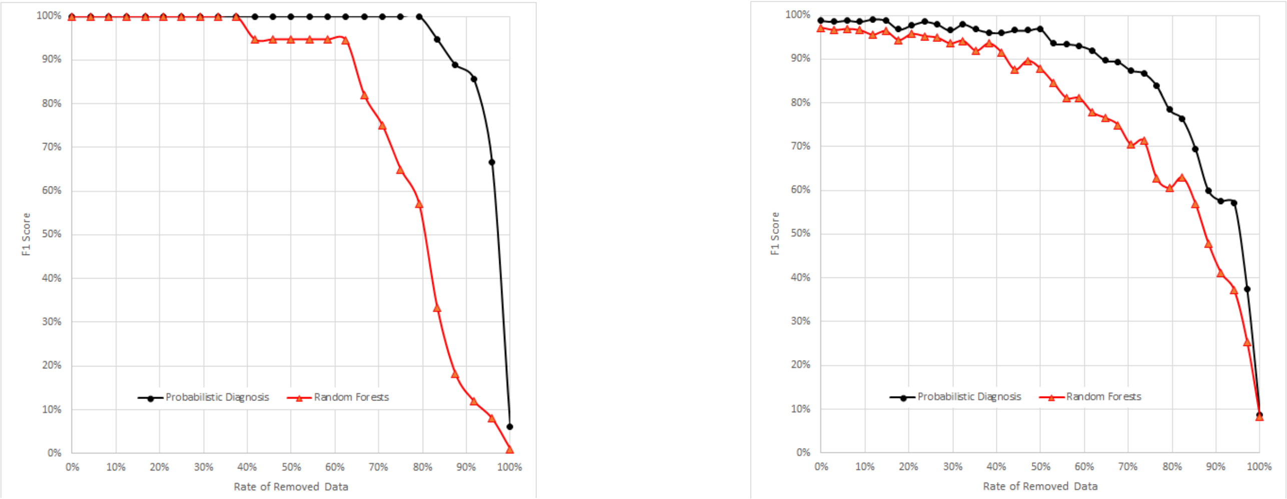


Figure 5.6: Flowchart of telemonitoring phase in the proposed self-adaptive telecare framework



(a) CKD dataset

(b) Dermatology dataset

Figure 5.7: Comparison of F1 score between the proposed self-adaptive telecare framework and random forest

5.3 Evaluation

A comprehensive evaluation of the proposed framework requires access to monitor real-world patients to verify whether the proposed solution is fully adapted to their needs. The aforementioned environment was not accessible due to legal and ethical restrictions; however, the proposed framework is validated by clinicians and medical systems engineers collaborating with the Maidis company in the context of the ITEA3 Medolution project.

Probabilistic diagnosis is an integral component in screening and monitoring phases; in the screening phase, a reliable screening is only possible with a powerful diagnosis component; moreover, in the monitoring phase, as the self-adaptive treatment component relies on the result of the probabilistic diagnosis component, the performance of the latter can depict the expected performance of the whole system. Hence, a comparative study of the diagnosis component is provided to illustrate its strengths, specifically in the case of patients' inadequate information. Moreover, four scenarios are provided to validate the proposed framework.

5.3.1 Comparative Study

Since the use of probabilistic diagnosis is similar in both phases, in the rest of this section, without losing generality, we focus on the use of probabilistic diagnosis in the screening phase, i.e., for detecting plausible medical conditions. The performance of the probabilistic diagnosis component is evaluated based on a classical classifier method, namely random forest, to demonstrate how the loss of data affects their respective performance. First, the numeric features were processed using the ontology-based reasoning component of the proposed framework to produce a nominal context. Then both random forest and BN are trained using the dataset. We have created BN using hybrid creation for a valid comparison (see Section 5.2.5.1). The evaluation is made by comparing the performance of both methods to predict while increasingly removing random data features.

5.3.1.1 Dataset description

In order to demonstrate the performance of the probabilistic diagnosis in the proposed framework, we have implemented the evaluation using two datasets on medical conditions, namely, chronic kidney disease and dermatology datasets.

Chronic kidney disease Dataset: The first dataset is for the prediction of chronic kidney diseases; this dataset includes 11 numeric and 13 nominal features to predict chronic kidney disease, from which 4 are causes and 20 are symptoms. The causes in this model are *age*, *hypertension*, *DM*, and *coronary artery disease*; while the symptoms include the measurements of *blood pressure*, *hemoglobin*, *red blood cell count*, and *white blood cell count.* The dataset is collected from 400 patients in a hospital in India and labeled regarding the presence of chronic kidney disease. This dataset is available on the machine learning repository of the University of California, Irvine [30].

Dermatology Dataset: The second dataset focuses on the diagnosis of different types of erythemato-squamous illnesses in dermatological patients. They all share clinical characteristics of erythema and scaling, and hence it is challenging to distinguish them. Psoriasis, seborrheic dermatitis, lichen planus, rosea pityriasis, chronic dermatitis, and pityriasis rubra pilaris are the illnesses in this group. Unfortunately, the exact diagnosis requires biopsy in most cases. To this end, Güvenir et al. [41] have collected this dataset to decrease the cost of prediction among these illnesses. The dataset has 33 linear and one nominal features and include 366 records, of which 2 are causes, and the rest are symptoms.

This dataset is accessible on the machine learning repository of the University of California Irvine [30].

5.3.1.2 Comparison

Since the random forest model obtains a well-established prediction performance for the aforementioned datasets, it was selected as a rival for comparison. Figure 5.7 depicts how removing the data from the records affects the performance of the proposed framework and random forest. They perform almost similarly when the data are complete. However, as the number of missing data increases, the drop of performance in the random forest model is enormous, which confirms the selection of BN for handling missing data in the core of reasoning in the proposed framework.

5.3.1.3 Robustness Analysis

A BN is robust when the values of its target class slightly depend on the inputs of its causes [97]. In a robust BN, the changes in one of the causes hardly result in dramatic changes in the target class. Chan and Darwiche [19] have implemented a tool integrating multiple approaches for quantification of robustness, also known as sensitivity analysis. We have used their tool to conduct sensitivity analysis based on the algorithm of [88], on the target class, representing the medical condition. The results show that the three causes can raise the probability of the disease to 37.04% in the case of the *chronic kidney diseases dataset*. For the *dermatology dataset*, there are only two causes and they can raise the probability to 68.72% on average for the six illnesses of this dataset. This high sensitivity is due to the unbalanced values of the two causes as they are similar in 66.56% of data records.

5.3.2 Use case

In this use case, the patient is *Frank Smith*, a user of the application of the proposed framework, named *Hapicare*; the summary of his medical file is shown in Table 5.1. In Hapicare, Frank Smith is under screening for various chronic diseases including high blood pressure, chronic kidney disease, and COVID-19. He is also under monitoring for diabetic and heart attack episodes. In this use case, we assume $TH_{wary} = 70\%$ and $TH_{diag} = 90\%$.

Table 5.1: Summary of medical file of patient

Item	Value
Name	Frank Smith
Gender	Male
Year of Birth	1960
BMI	36 kg/m^2 (10-Jan-2021)
Smoking	Yes
Chronic Diseases	DM Type II (5-Dec-2016)
Medical History	Heart Attack (7-Jun-2018)
Family Doctor	Dr. Anna Doe

5.3.2.1 Scenario 1

Once Frank felt shortness of breath, he consulted the Hapicare application. In the latter, the new information, *shortness of breath*, is processed as follows:

- The data are mapped into SNOMED-CT ontology; hence, Dyspnea (Concept Id: 267036007) is obtained and then added to the knowledge base.

- Ontology-based reasoning processes the ontology-mapped data; since no recent activity is present in the knowledge base, Dyspnea at rest (Concept Id: 161941007) is deduced using the ontology-based reasoning and then added to the knowledge base.

- Medical conditions/episodes that are related to the new finding are narrowed down to COVID-19 medical condition.

- The probability of COVID-19 is recalculated based on the knowledge base (including the new finding), which results in P(COVID-19) = 73%.

- Since the obtained probability is between TH_{wary} and TH_{diag}, the application selects Fever (Concept Id: 386661006) as missing information for sensing action.

- The sensing action is mapped to the following patient-friendly statement: "Please measure your body temperature."

- Frank uses an IoT sensor to measure his body temperature, which is $40°C$.

- The obtained temperature is analyzed and then mapped into SNOMED-CT; hence, Fever (Concept Id: 386661006) is obtained and then added to the knowledge base.
- Medical conditions/episodes that are related to the new finding are narrowed down to COVID-19 and hypoglycemia medical conditions.
- The probabilities of COVID-19 and hypoglycemia are recalculated based on the knowledge base (including the new finding), which results in P(COVID-19) = 91% and P(Hypoglycemia) = 64%.
- Since the probability of COVID-19 exceeds TH_{diag}, the diagnosis is added to the knowledge base. Hapicare notifies Frank's doctor about the possible presence of COVID-19 medical condition and encourages Frank to book an appointment with his doctor.

Frank visits his doctor for further diagnosis and possible treatments of his medical condition.

5.3.2.2 Scenario 2

In the second scenario, Dr. Anna Doe, Frank's family doctor, notices an alert from Hapicare reminding a complete blood test for Frank as the previous one is old; hence, she orders a blood test for him. The results of his blood tests are processed in Hapicare similar to the steps discussed in Section 5.3.2.1. Given the medical file and the new information from the blood test, using the model described in Section 5.3.1, the probability of CKD is calculated, and the obtained probability exceeds the diagnosis threshold; hence, Dr. Doe is notified. She examines Frank and confirms CKD diagnosis. Dr. Doe decides to start the monitoring phase for Frank regarding CKD. Frank is currently following peritoneal dialysis treatment at home that uses the lining of the abdomen to filter the blood inside his body; moreover, he is under continuous monitoring for the related episodes, e.g., Urinary Tract Infection (UTI) and Acute Kidney Injury (AKI), and further complications due to comorbidities. The doctor also prescribes him some antibiotics to take in the case of a UTI.

5.3.2.3 Scenario 3

Few months after diagnosis of CKD, Frank displays symptoms of *hypotension* and consequently Hapicare diagnoses a hypotension episode. It follows the treatment discussed in Section 5.2.6. As mentioned in Section 5.2.5.2, episodes of hypotension affect the probability of AKI. Hence, after the diagnosis of hypotension, the probability of AKI is increased, but it is still under TH_{wary}; so no further action is performed.

5.3.2.4 Scenario 4

Frank feels sick and consults Hapicare; the application selects the sensing action *measuring body temperature*. Ontology-based reasoning deduces *fever* and then adds it to the knowledge base. Given the knowledge base, probabilistic reasoning results in increasing the probabilities of COVID-19 and UTI, which are now both surpassing TH_{wary}. For differential diagnosis of the two possible choices, Hapicare selects *cough* as a sensing action. Once Frank responds that he does not cough, the probability of COVID-19 is reduced. Since the probability of UTI is still over TH_{wary}, Hapicare asks Frank for *a burning sensation while urinating* as a sensing action; his positive response increases the probability of UTI. For validation of this diagnosis, Hapicare selects a pathognomonic sensing action and asks Frank to use test strips for UTI detection[1]. As Frank reports the color of the UTI test, Hapicare confirms the presence of UTI, and based on the prescribed antibiotics in the treatment rules, the self-adaptive treatment module selects one of them and recommends Frank to start taking it. After a few days, Hapicare asks Frank to take UTI test strips to see the changes in the state of infection. As Frank's health state regarding UTI is not improving, Hapicare reduces the selected antibiotic's probability and recommends the patient take another antibiotic. After a few days, Hapicare increases the probability of the second antibiotic as Frank recovers from UTI. Dr. Doe is notified at each step, and she can directly take over the medical treatment or change the treatment rules on the fly.

5.4 Discussion

In telecare, most sensors require an action from the patient, e.g., it needs to put the hand inside the cuff of a sphygmomanometer for measuring blood pressure. Moreover, for some of the critical information, there are no sensors; for example, no sensor can measure the pain's location. However, the more sensing actions, the more inconvenient for the user, and patients will eventually opt out if a system asks too many questions. On the other hand, holistic treatment is only possible with all the information. Hence, in the proposed framework, we have applied BN for diagnosis, which resulted in

[1] UTI test strips are diagnosis kits that change color in contact with urine and can be used at home. The presented color shows the presence and type of infection.

semi-holistic treatment with minimal data.

The evaluation results support the selection of the BN; they show the performance of diagnosis with minimum information; for both datasets, the results are hugely in favor of the BN. Albeit having more than half of the data, BN and random forest performed similarly.

On the other hand, one of the common pitfalls is depending too much on specific causes for diagnosis, which results in a biased prediction. In medical diagnosis, the causes only affect the general probability of a medical condition, and a diagnosis is only possible using the effects, also known as symptoms. The robustness analysis shows that the BN's training is efficient and learning about all the causes is not enough for the diagnosis of CKD in the first dataset. However, for the second dataset, there are only two causes, and they have similar values for most of the records; hence the BN training is not robust.

The scenarios presented in Section 5.3.2 demonstrate how an application of the proposed work, Hapicare, can help patients and doctors with screening and monitoring of medical conditions/episodes. In the first two scenarios, the screening of a patient is shown. In the first scenario, the proposed framework captures the patient's information at his home for screening an urgent medical condition. The step-by-step interactions of Hapicare illustrate the workflow of the system. In the second scenario, the medical file and external data are used for screening a medical condition. These two scenarios depict how Hapicare can help a doctor in his/her diagnosis by providing insights into possible medical conditions.

In the last two scenarios, the monitoring phase is shown. In the third scenario, the treatment process and self-adaptive treatment are discussed. Moreover, we have demonstrated that a diagnosis of an episode might affect the probabilities of other medical conditions/episodes. In the last scenario, we have shown that Hapicare can help patients diagnose and treat medical conditions/episodes even when the first treatment is not sufficient. This process might take several weeks with traditional methods of diagnosis and treatment. The result of the self-adaptive treatment can also help doctors treat other medical conditions/episodes; because the doctor can see which treatments were more effective.

5.5 Conclusion

The number of patients struggling with chronic diseases is growing; the traditional treatments are inefficient and massively costly. Efficient and holistic treatment is required to enhance their quality of life. However, comprehensive data collection is intrusive and expensive. In this chapter, we have proposed a framework to achieve the semi-holistic diagnosis with the least possible intrusion. Additionally, we have used ASP to adapt treatment to the specific needs of each patient. The proposed framework collects data from IoT-based sensors as well as self-assessment; these data are processed through ontology-based reasoning for contextual information of collected data. The probabilistic diagnosis is responsible for diagnosing medical conditions/episodes. The diagnosis is based on the patients' contextual information of collected data. Hence, it creates a list of episodes and their associated probabilities and forwards it to the ASP component to suggest the most suitable treatment to each patient. Our experiments have shown the performance of probabilistic diagnosis in comparison with random forest model. Moreover, the validation of the proposed framework using four scenarios demonstrates its effectiveness in diagnosing and treating medical conditions and episodes of patients.

Chapter 6

A Blockchain-based Secure Framework for Homomorphic AI in IoT networks

This chapter investigates the security concerns of data-sharing in distributed IoT networks. These concerns are handled using a blockchain-based framework empowered by various encryption schemes. We use homomorphic encryption for privacy-preserving identification and broadcast encryption for efficient, secure data-sharing. This chapter also includes the formal description and proof of the proposed framework.

Contents

6.1 Motivations

IoT networks are becoming integral parts of the modern-day; there are various IoT applications, e.g., smart homes, smart cars, healthcare. Given the criticality of data, there are arising concerns on security and privacy in the IoT context. The well-established approach regarding the integrity of messages is using the immutable architecture of blockchain. However, in the blockchain, the data are not encrypted, and privacy is achieved using anonymity rather than confidentiality. Hence, the confidentiality and privacy of IoT data in public blockchain networks are still open issues, see Section 3.5.

As we discussed in Section 4.2, many works use asymmetric encryption suits for securing the data in the blockchain network. This process is similar to the existing works before blockchain; the data sender encrypts the messages with the receiver's public key and transmits the encrypted message. It is a suitable approach for the trusted peers because the receiver peers can decrypt the message. However, the idea of blockchain is security without trust. Hence, in this chapter, we introduce a solution using HE schemes to achieve confidentiality in a trustless network.

6.2 Security Requirements

In this proposed framework, we tackle the privacy concern of using third-party services in a distributed network. We focus on IoT networks as they are continually growing in the various fields of applications. We consider the blockchain backend for its well-established characteristics in a secure distributed platform. Singh and Singh [90] have signified the importance of convergence of blockchain, AI, and IoT technologies. They have mentioned that IoT-based applications are connected, flexible, and efficient; blockchain offers additional security and transparency, while AI provides data analysis. Hence, in this chapter, we focus on the emergence of AI into blockchain-based IoT networks in a privacy-preserving manner.

The security concerns that are addressed in the proposed framework can be listed as follows:

- Ownership: In our framework, we ensure that the data owners endure their ownership; and they are the sole who can control their data.
- Transparency: In our framework, all the shared activities should be publicly traceable; so an

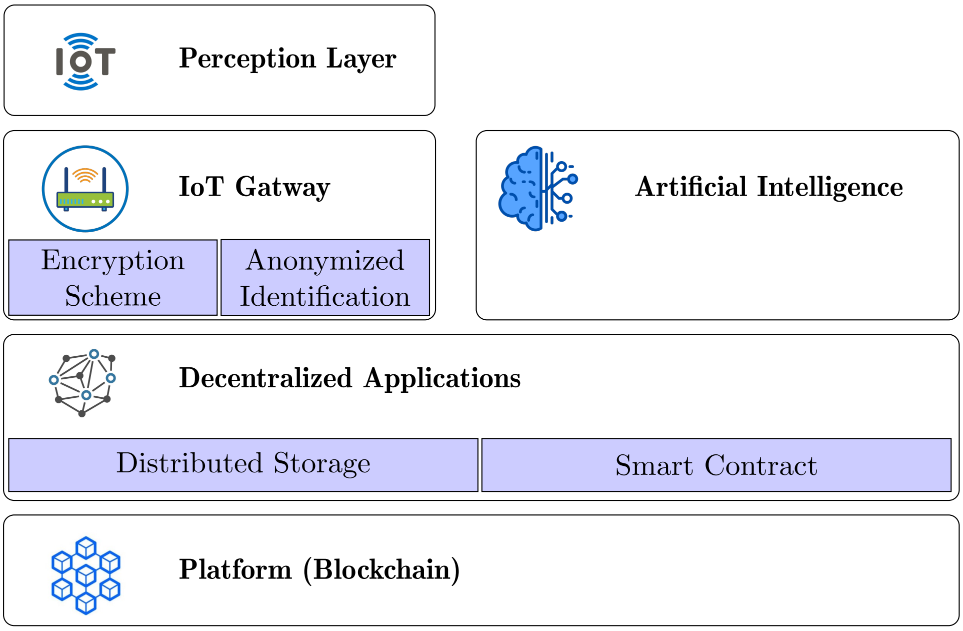


Figure 6.1: General architecture of the proposed blockchain-based secure framework for homomorphic AI in IoT networks

auditor can verify all the procedures.

- Confidentiality: In completion of ownership, only allowed users, set by the data owner, can access the data; and any unauthorized access should be prohibited.
- Integrity: Any data in our framework should have guaranteed integrity, and any changes in the data should be detectable.

6.3 A Blockchain-based Secure Framework for Homomorphic AI in IoT networks

6.3.1 Architecture

The proposed framework is designed in two primary layers to take advantages of two technologies: blockchain and homomorphic encryption. The former provides a distributed immutable platform for ensuring integrity. The latter is for enabling computation without compromising the data itself. The general architecture of the proposed framework is depicted in Figure 6.1; and details of each component are discussed in the following sections.

6.3.2 Platform Layer

In the blockchain, distributed ledger can be considered as an immutable decentralized database. Moreover, smart contracts are robust tools for executing event-driven actions in the blockchain. Hence, the proposed framework is based on blockchain technology. Blockchain plays the role of the proposed framework's backbone as all the other components of the proposed framework communicate through the blockchain network.

6.3.3 Decentralized Application Layer

A Decentralized Application (DApp) is stored and executed via a decentralized network, like blockchain. DApp is a logical layer that orchestrates the proposed framework. The tasks of DApp can be broken into two parts: smart contracts and data storage. The former handles the logical workflow of the system, while the latter handles the distributed data storage. These two sub-components are detailed in the following sections.

6.3.3.1 Distributed Storage Component

Blockchain is not a general-purpose database; i.e., storing large data is not efficient in the blockchain. Hence, in the proposed framework, we propose to store the data off-chain, i.e., outside the blockchain. One of the common technology for this purpose is InterPlanetary File System (IPFS)[8], a decentralized file system that allows access, storage, and security of files on a distributed network, see Section 3.5.4. It is a combination of distributed hash tables, incentivized block-exchange, and self-certifying namespaces. This protocol allows peers to store a file and serve it with its content address. The other peers can find and request the content using a distributed hash table. The combination of IPFS and blockchain allows storing the distributed hash table in the blockchain for easy immutable access.

6.3.3.2 Smart Contract Component

Smart contracts are the event-driven scripts executed in blockchain. Because of the intrinsic features of blockchain, smart contracts can guarantee the execution of some processes in the case of some events. We use smart contracts in our proposed framework for managing the whole system, i.e., to guarantee the various components work as expected.

Upon creating new data in the system, it is encrypted in the encryption scheme, and accompanying the anonymized identification is stored in the distributed storage. The smart contract is triggered upon receiving a piece of new information and arranges its storage. The hash of the stored data with the identification of the data owner is stored in the blockchain. The smart contract also triggers AI's execution on the new data; the resulting information is similarly stored, and the data owner is informed about the result of the AI component. All of these processes are managed using smart contracts.

The use of smart contracts assure that any data in the blockchain is securely saved in a distributed storage, its hash exists in the blockchain, it has been processed by AI, and its analysis results exist in the blockchain.

6.3.4 IoT Gateway Component

IoT gateways traditionally collect and send the data from IoT devices to external platforms. In our proposed framework, IoT gateways connect the IoT network of one peer to the blockchain. IoT gateways are considered to have more computation powers compared to IoT devices, as they are required to perform some preprocessing on the data that are discussed in the following processes:

6.3.4.1 Encryption Process

As we detailed in Section 3.6, homomorphic encryption is an answer for secure outsourcing of computation. The main objective of such a scheme is to enable computation on the data without revealing its content. The outsourced computation of the proposed framework is classification; hence, we propose to encrypt the IoT data using HE schemes.

Since raw IoT data do not contain any identifying information and are usually meaningless on their own. Hence, we assume the raw IoT data is not critical. Therefore, after the data are collected and labeled with the user's identification in the IoT gateway, they are encrypted prior to storage in the distributed storage.

6.3.4.2 Anonymized Identification Process

Blockchain is known for a network of anonymous trustless peers. Hence, for the storage of data, we follow the same approach. In the proposed framework, we use secure hash algorithms to identify the data. Secure hash algorithms are one-way functions whose reverse are not computationally feasible.

The peers can efficiently compute their identifications using a hash algorithm; however, others can not get the peer from their identifications. Moreover, it allows knowing which data belong to the same peer, necessary for AI.

6.3.5 Artificial Intelligence Component

The benefits of AI in IoT data analysis made it an integral part of any smart IoT application, e.g., AI-based diagnosis in smart healthcare applications. AI algorithms usually require high computation power, making them not feasible to be executed locally in IoT networks. Moreover, AI providers, similar to any other services, prefer to provide their services online instead of locally; to avoid reverse engineering and losing their benefits. On the other hand, since the data are invaluable for AI providers to extend their works, they might collect any data at their disposal. Therefore, users are conscious of using AI services for their privacy of data.

In this proposed approach, we propose to use homomorphic AI on the data that allows the users to benefit from AI services without losing the confidentiality of data. An adequate AI service for our proposed framework should be able to work efficiently on encrypted messages.

Various existing works explore the aforementioned requirement, i.e., homomorphic machine learning. For instance, Bost et al. [14] have discussed three classification models for encrypted data: hyperplane decision-based classifier, naïve Bayes classifier, and decision trees. Similarly, Sun et al. [95] have improved HE scheme for similar classification algorithms. Besides traditional classification, Orlandi et al. [74] have addressed the secure data processing for neural networks. They pointed out the confidentiality of data as well as the configuration information.

6.3.6 Perception Layer

The closest layer to the physical layer in our proposed framework is the perception layer. The perception layer consists of the IoT devices that interacts with the outside world. The data are collected by IoT devices, e.g., wearable sensors, mobile phones, and environmental sensors.

6.3.7 Workflow

A typical workflow of the proposed framework is presented using Business Process Model and Notation (BPMN) in Figure 6.2. The actors of the proposed framework are the followings:

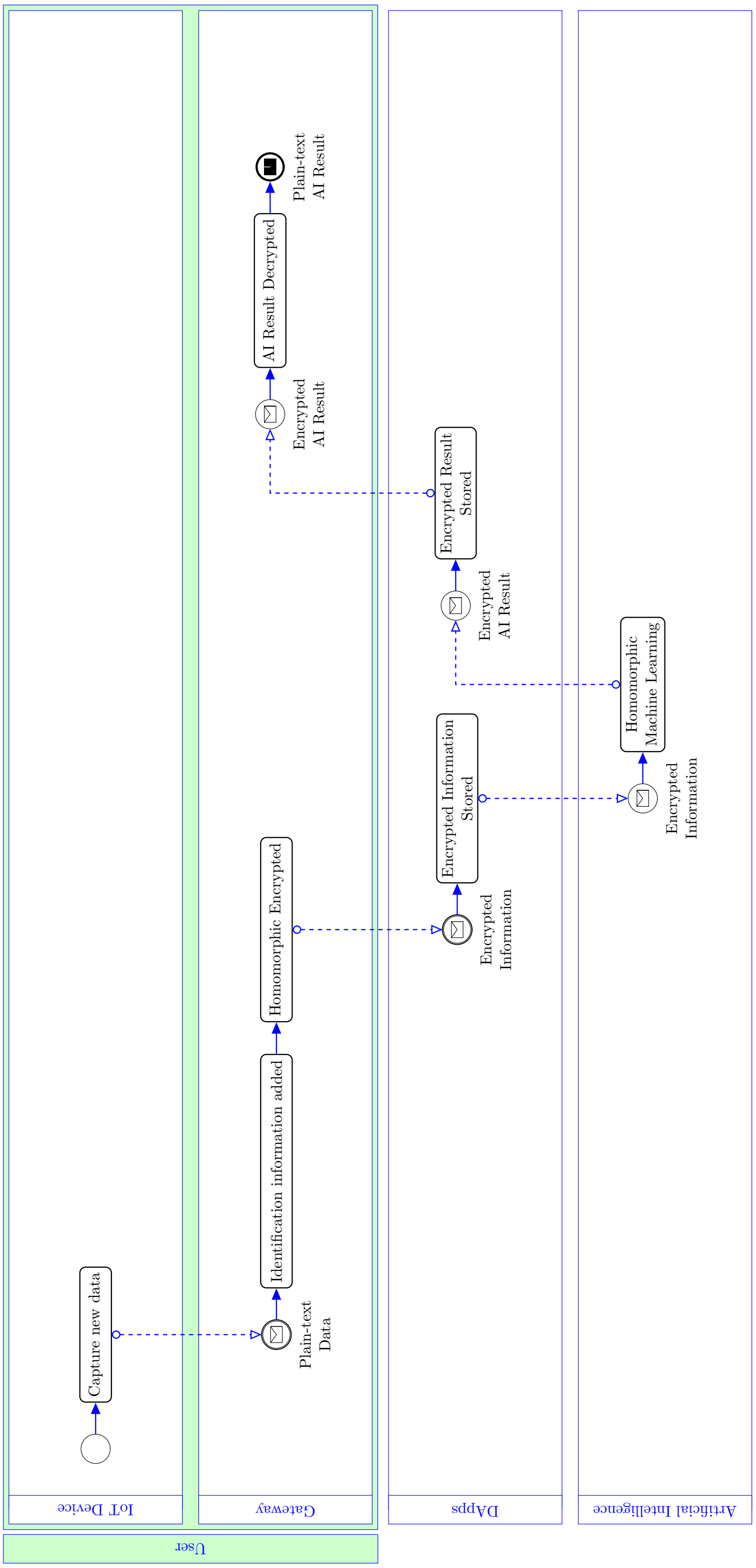


Figure 6.2: General workflow of the proposed blockchain-based secure framework for homomorphic AI in IoT networks

- User
  - IoT device (Section 6.3.6)
  - IoT gateway (Section 6.3.4)
- DApps
  - Distributed storage (Section 6.3.3.1)
  - Smart contracts (Section 6.3.3.2)
- AI (Section 6.3.5)

For simplicity, the interrelation between distributed storage and smart contracts is not depicted in the workflow. However, any data arriving at DApps, either from IoT gateway or AI, are verified using their hash values for integrity. The contents are then sent to the distributed storage for saving the data, which would produce a hash value of the contents and a unique address for accessing the data.

The process starts with capturing data in the IoT devices, e.g., measuring heartrate using an IoT pulsometer. These data are transferred in plain form to the gateway for formatting. The gateway prepares a message based on the anonymized identification and the encrypted data to be sent to DApps. In the latter, the smart contracts verify the message's integrity prior to sending to the distributed storage for saving the cryptogram of new data. When the cryptogram is stored, the AI component can access it using a publicly available address. The AI component may execute homomorphic machine learning algorithms on the cryptogram and send it back to DApps for storage. DApps follow the same approach for storing the encrypted AI results in the distributed storage. Afterward, the user can access the encrypted AI result and decrypt it with his/her key to access the AI results.

6.4 Security Evaluation

In this section, we analyze the security aspects of the proposed framework. The notations used in this section are summarized in Table 6.1.

Table 6.1: Notations used in formal modeling of the proposed secure framework for homomorphic AI

Notation	Description
d	Raw Data
$E(m)$	Encryption on message m
c_m	Cryptogram on message m
$D(c)$	Decryption of message m
$AI(m)$	Classification of message m
r	Result of AI
$H(m)$	Secure hash digest on message m
h_m	Digest value of message m
a_m	Address of message m in the distributed storage
u	User of the system
$Own(u, m)$	Is u the owner of message m
$Acc(u, m)$	Has u read access to message m

6.4.1 Formal Description

6.4.1.1 Proposed Framework

In the proposed framework, an IoT device generates data d. The latter is processed in the IoT gateway by adding the anonymized identification information h_{id} and encrypting the message with the homomorphic encryption scheme ($c_d \leftarrow E(d)$). The collection of these messages creates a message m to be stored in the blockchain.

In the blockchain, after evaluating the integrity of the message, the message is transferred to be stored in the distributed storage. The latter provides the address a_m of the stored message that can be used for further access. Smart contracts store the data's hash, owner's anonymized identification information, and accessible address in the blockchain.

Smart contracts may inform the AI component of the new information. AI accesses the data from the distributed storage. Before other processes, AI verifies the hash value of the message. Afterward, AI executes machine learning algorithms on the encrypted data ($c_r \leftarrow AI(c_d)$). These data are transferred back to smart contracts to be processed and stored with a similar process discussed above for the process and storage of new data.

Smart contracts may inform the user via his/her IoT gateway of the presence of the machine learning results. IoT gateway retrieves the encrypted result h_r using the address a_r. After verification of the hash value, the encrypted result is decrypted $r_d \leftarrow D(c_r)$.

It worth mentioning that decrypted result is, in fact, the result of machine learning algorithms on the data d itself, $r_d = AI(d)$; however, the AI component never has access to the data d, itself. We have summarized the logical flow of the proposed framework in Figure 6.3.

6.4.1.2 Requirements

In any framework, three main broad security requirements are to be addressed: (1) Confidentiality: only authorized entities can read data. (2) Integrity: only authorized entities can modify data. (3) Availability: the access of authorized entities is always available. In all the mentioned security concerns, the authorized entity is to be agreed. Only data owners are authorized to access their data and the derivative of their data in our proposed framework. Moreover, since the IoT data are essential to be kept as history, no entity is authorized to modify the data once entered into the framework. Given the above assumptions, the security requirements discussed in Section 6.2 are mapped to confidentiality requirements formalized as the followings:

$$\text{Ownership} \quad : \quad \forall Message\ m, \forall User\ u : Own_t(u,m) \Longrightarrow Own_{t+1}(u,m) \tag{6.1}$$

$$\text{Transparency} \quad : \quad \forall Transaction\ e, \forall User\ u : Acc(u,e) \tag{6.2}$$

$$\text{Confidentiality} \quad : \quad \forall Message\ m, \forall User\ u : Own(u,m) \Longleftrightarrow Acc(u,m) \tag{6.3}$$

$$\text{Integrity} \quad : \quad \forall Message\ m : m_t \Longrightarrow m_{t+1} \tag{6.4}$$

6.4.1.3 Assumptions

In the proposed framework, we have assumed the IoT gateway is secure. Moreover, we rely on the security of the cryptology algorithms used in this framework. Therefore, we assume the security of hash functions and homomorphic encryption schemes used in this framework. Moreover, we assume the blockchain technology guarantees data integrity and availability, based on the discussion of blockchain security presented in Section 3.5.3. The key assumptions needed for the formal proof are the following:

- Because of the blockchain's intrinsic feature, any data in the blockchain will remain intact over time (Equation 6.5).
- Any data in the blockchain are publicly available for all the users (Equation 6.6).
- Blockchain logs all traces of the transactions in an immutable ledger (Equation 6.6).

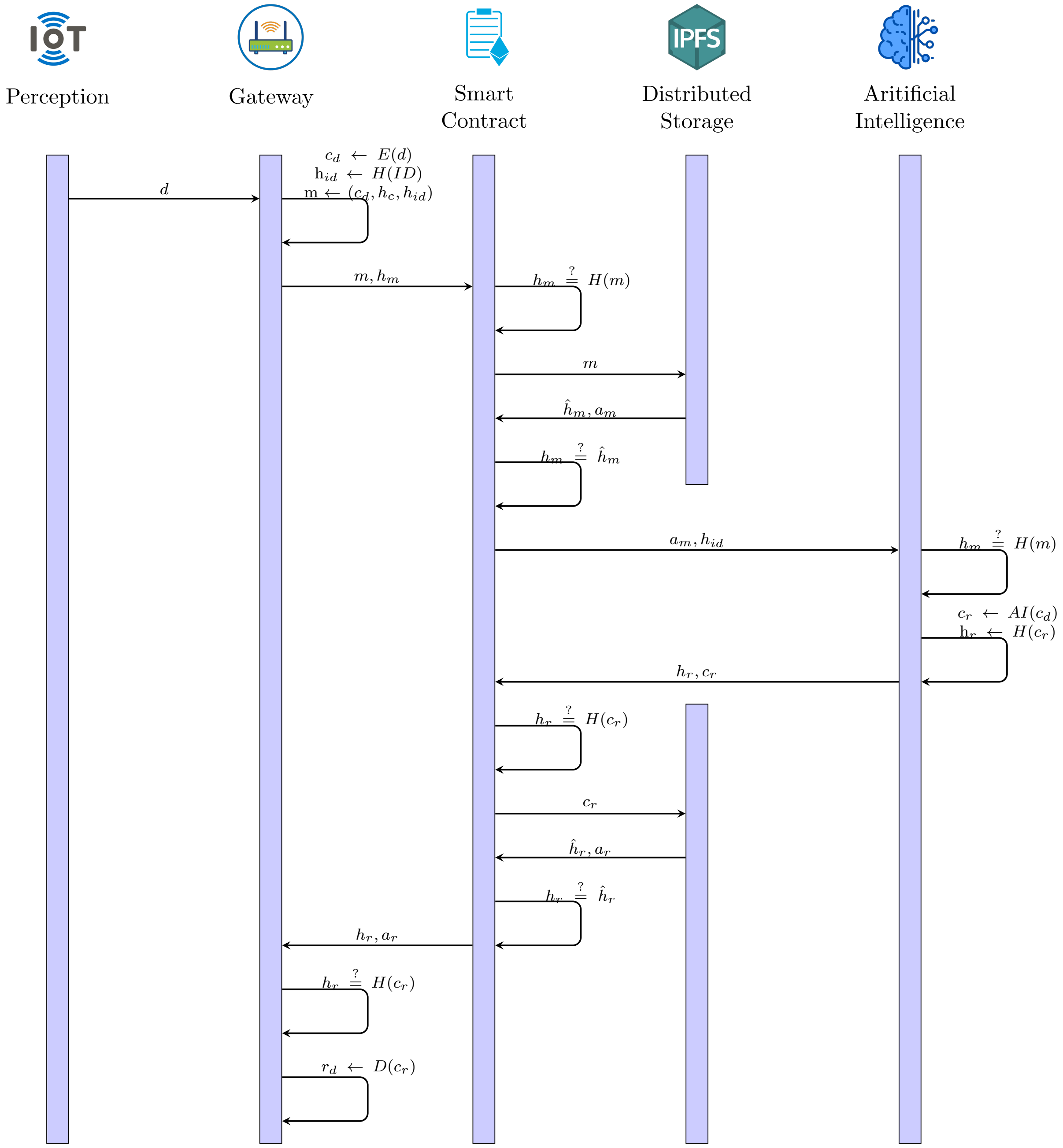

Perception
Gateway
Smart Contract
Distributed Storage
Aritificial Intelligence
d
$c_d \leftarrow E(d)$
$\mathrm{h}_{id} \leftarrow H(ID)$
$\mathrm{m} \leftarrow (c_d, h_c, h_{id})$
m, h_m
$h_m \stackrel{?}{=} H(m)$
m
$\hat{h}_m, a_m$
$h_m \stackrel{?}{=} \hat{h}_m$
a_m, h_{id}
$h_m \stackrel{?}{=} H(m)$
$c_r \leftarrow AI(c_d)$
$\mathrm{h}_r \leftarrow H(c_r)$
h_r, c_r
$h_r \stackrel{?}{=} H(c_r)$
c_r
$\hat{h}_r, a_r$
$h_r \stackrel{?}{=} \hat{h}_r$
h_r, a_r
$h_r \stackrel{?}{=} H(c_r)$
$r_d \leftarrow D(c_r)$

Figure 6.3: Logical flow execution of the proposed framework

- The hash values of two different messages are never identical (Equation 6.8).

- Only with having the encryption key, a user can get plain text data from a cryptogram (Equation 6.9).

- Only the data owner has the encryption key (Equation 6.10).

$$\begin{aligned}
\text{Blockchain characteristics} &: \forall x \in \mathit{blockchain} : x_t \implies x_{t+1} && (6.5)\\
\text{Blockchain characteristics} &: \forall x \in \mathit{blockchain}, \forall User\ u : Acc(u, x) && (6.6)\\
\text{Blockchain characteristics} &: \forall Transaction\ e : e \in \mathit{blockchain} && (6.7)\\
\text{Hash characteristics} &: \forall m_1, m_2 : H(m_1) = H(m_2) \iff m_1 = m_2 && (6.8)\\
\text{Encryption characteristics} &: \forall x : \Big(Acc(u, c_x) \implies Acc(u, x)\Big) \implies Acc(u, key_{c_x}) && (6.9)\\
\text{Encryption characteristics} &: \forall User\ u : Acc(u, key_{c_x}) \implies Own(u, x) && (6.10)
\end{aligned}$$

6.4.2 Formal Verification

6.4.2.1 Integrity

In order to prove integrity in the proposed framework we should prove that $\forall m \in \mathcal{M} : m_t \implies m_{t+1}$. As discussed in Section 6.4.1.1, for each data arriving in the blockchain, its hash value is computed and verified with the provided hash value; and only if the two hash values are matched, the data is stored in the distributed storage, while its hash value and its address are stored in the blockchain (m_0). In the blockchain, all the data are immutable and available; hence we can deduce that $H(m)_t \implies H(m)_{t+1}$. In order to change a message in the distributed storage from m to $H(\hat{m})$, its hash value should be updated to another value in the next timestep $H(\hat{m})_{t+1}$ (Eq. 6.11). Because of the immutability of blockchain and characteristics of hashing algorithm, the integrity of the message in the proposed framework is formally proved as the followings:

$$\begin{aligned}
\text{Equation 6.5} \quad &: \quad \forall Message\ m : H(m)_t \implies H(m)_{t+1} \qquad (6.11)\\
\text{Change in the hash} \quad &: \quad Message\ m, \hat{m} : H(m)_t \implies H(\hat{m})_{t+1}\\
&\implies \quad H(\hat{m})_{t+1} = H(m)_{t+1}\\
\text{Equation 6.8} \quad &: \quad \forall m_1, m_2 : H(m_1) = H(m_2) \iff m_1 = m_2\\
&\implies \quad \hat{m} = m\\
&\implies \quad m_t \implies m_{t+1}\\
&\implies \quad \forall Message\ m : m_t \implies m_{t+1}
\end{aligned}$$

6.4.2.2 Ownership

Each data stored in the proposed framework accompanies a h_{id} which corresponds to the data owner. Since this value is stored in the blockchain, similar to the proof above; it can be proven that:

$$\begin{aligned}
&\quad Own(u, m) \iff h_{id} \qquad (6.12)\\
\text{Integrity} \quad &: \quad h_{idt} \iff h_{idt+1}\\
&\implies \quad \forall Message\ m, \forall User\ u : Own_t(u, m) \implies Own_{t+1}(u, m)
\end{aligned}$$

6.4.2.3 Transparency

For this proof, we rely on the transparency and availability characteristics of blockchain, formalized in Equations 6.5 and 6.6, respectively. Since blockchain logs all the transactions in an immutable manner we can prove the transparency of the framework as follows:

$$\begin{aligned}
\text{Equation 6.6} \quad &: \quad \forall x \in blockchain, \forall User\ u : Acc(u, x) \qquad (6.13)\\
\text{Equation 6.7} \quad &: \quad \forall Transaction\ e : e \in blockchain\\
&\implies \quad \forall Transaction\ e, \forall User\ u : Acc(u, e)
\end{aligned}$$

6.4.2.4 Confidentiality

The proposed framework's data always appear in the encrypted form, except before the IoT gateway, which we assumed secure. According to the encryption scheme's security, no one would have any additional information from the cryptogram, which depicts the confidentiality of the proposed framework.

We use indirect proof for confidentiality: contradictory assumption is that there is an unauthorized user $\hat{u}$ with access to the message m (Equation 6.14). Since the IoT gateway is secure, the user $\hat{u}$ has access to the message through the blockchain (Equation 6.15). However, due to the encryption scheme security characteristics (Equation 6.9), which contradicts the assumptions and proves the confidentiality (Equation 6.16).

$$\begin{aligned}
& \exists \hat{u}, m : Acc(\hat{u}, m) \wedge Own(\hat{u}, m) = false && (6.14)\\
\Longrightarrow\ & \exists \hat{u}, m : Acc(\hat{u}, c_m) \Longrightarrow Acc(\hat{u}, m) \wedge Own(\hat{u}, m) = false && (6.15)\\
\text{Equations 6.9, 6.10}\ :\ & \Big(Acc(\hat{u}, c_m) \Longrightarrow Acc(\hat{u}, m)\Big) \Longrightarrow Own(\hat{u}, m) &&\\
\Longrightarrow\ & \bot && (6.16)\\
\Longrightarrow\ & \forall Message\ m, \forall User\ u : Own(u, m) \Longleftrightarrow Acc(u, m) &&
\end{aligned}$$

6.4.2.5 Availability

In the proposed framework, due to the distributed system of blockchain, storage, and AI, there are computationally enough resources that a single user cannot deprive others. As long as the number of distributed nodes remains large enough, our proposed framework is available.

6.5 Use Case

We have evaluated the proposed secure framework for homomorphic AI using a healthcare use case. Telemonitoring can be an application of the proposed framework; hence, it is discussed as a use case. A typical telemonitoring service contains patients' medical data and AI modules to analyze them. Since the medical data are critically confidential, the proposed secure framework for homomorphic AI can be beneficial in this use case.

IoT devices, e.g., wearable sensors, environmental sensors, and mobile phones, produce data in telemonitoring systems. These data are considered private, and it is vital to keep them secure. Moreover, these data are valuable for AI companies as they can be used for improving their systems. Hence, patients need to be guaranteed to use a telemonitoring service without compromising their data.

In the proposed secure framework for homomorphic AI, the IoT data are homomorphically encrypted. These data can be analyzed using homomorphic AI on the encrypted data without decrypting them. The AI results are also encrypted, and only the holder of the HE scheme's key, i.e., data owner, can decrypt it. Therefore, the proposed framework enables patients to use AI to analyze their medical data for telemonitoring services; without providing their data in plain-form. Moreover, the results of telemonitoring are also kept private for only the intended patient.

6.6 Conclusion

In this chapter, we first discussed the importance of security in IoT networks, particularly in the case of AI outsourcing. Then we proposed a blockchain-based framework using homomorphic encryption schemes to address such a network's security concerns. We have also provided a formal description of the framework and then proved the proposed framework's security requirements.

Chapter 7

A Secure Data-sharing Framework Based on Blockchain: Teleconsultation Use-case

Contents

7.1 Motivations

Many studies work on integrating blockchain technology in IoT applications, given the security, interoperability, scalability, and availability benefits of blockchain. Although blockchain provides a distributed platform for storing and sharing data, it lacks enough measures for guaranteeing the confidentiality of the data. One of the approaches is using permissioned blockchains, e.g., hyperledger fabric, which has embedded access control. In permissioned blockchain, access to some or all the nodes of blockchain are restricted using access control. This approach is suitable for closed ecosystems, e.g., internal organization networks. However, the management of access control requires a centralized authority which limits the benefits of using blockchain. Another group of approaches is extending public permissionless blockchain with cryptosystem as an additional layer of security. Blockchain technology itself uses secure hash algorithms for guaranteeing immutability and integrity. Moreover, the additional layer uses encryption algorithms for privacy and confidentiality. We have discussed the existing works on this subject in Section 4.2. Most of them adopt the well-established approach in centralized applications, i.e., asymmetric encryption, for use in the blockchain.

In centralized applications, e.g., an online web application, using asymmetric encryption is convenient; because each communication channel is dedicated to only one application. In other words, the encrypted data is meant for only one recipient. On the other hand, in a distributed system, various services might exist on the communication channel, requiring the same data. In such cases, using asymmetric encryption creates multiple cryptograms of a single data, causing redundancy of communication channels.

In this chapter, we propose a privacy-preserving framework for the data sharing of IoT-based applications on the blockchain platform. In our framework, we use broadcasting encryption, see Section 3.7, which allows secure and efficient data sharing.

7.2 Privacy Problem

This proposed framework tackles the confidentiality and privacy concerns regarding data sharing in distributed IoT applications. We consider blockchain as the distributed platform of our work, given its security and interoperability features.

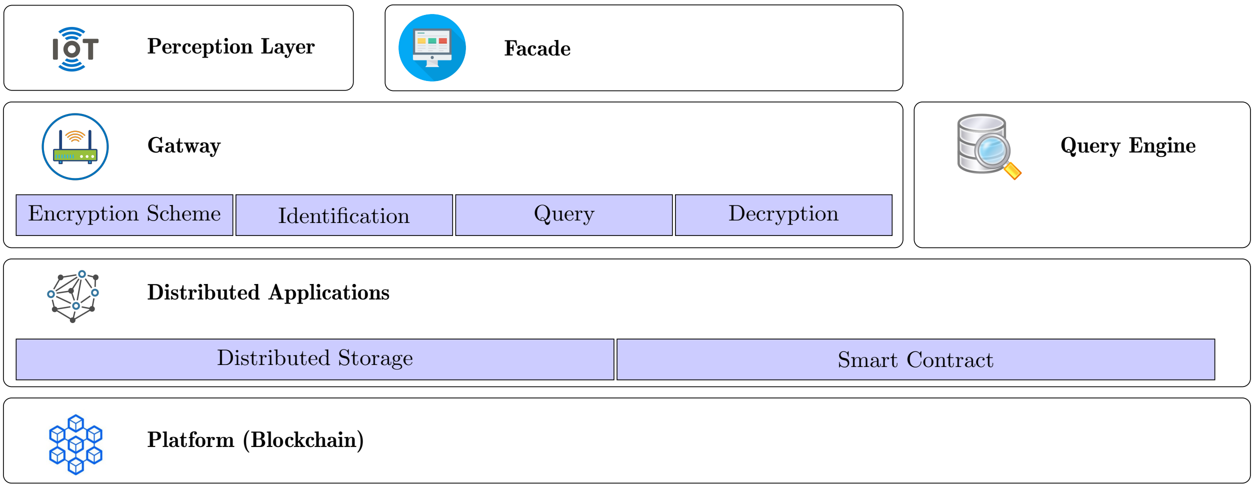


Figure 7.1: General architecture of the proposed framework

The security addressed in the proposed framework is similar to the ones discussed in Section 6.2; with the following additional concerns:

- Fine-grained access control: Owners should be able to adapt their data access in our framework. In our framework, data owners, at any time, may alter the access to their data, i.e., allowing/revoking access to the data.

- Traceable access control: The transactions regarding the access controls should be traceable in our proposed framework. In other words, requests for data access and their response, i.e., granting or denying their requested access, should be available for auditors to verify.

7.3 A Secure Data-sharing Framework Based on Blockchain

7.3.1 Architecture

The proposed framework is designed in two primary layers to take advantage of two technologies: blockchain and broadcast encryption. The former provides a distributed immutable platform for ensuring integrity. The latter is for enabling secure data sharing without unnecessary replications. The general architecture of the proposed framework is depicted in Figure 7.1. This proposed framework's perception and platform layers are similar to the ones discussed in Section 6.3.6 and 6.3.2, respectively. The other components are discussed in the following sections.

7.3.2 Decentralized Applications Layer

The DApps layer of the proposed framework is close to the one discussed in Section 6.3.3. The distributed storage component is similar in the two frameworks. In contrast, the smart contract components of the two DApps are different.

7.3.2.1 Smart Contract Component

The smart contracts provide event-driven execution of scripts that allows having a guaranteed flow of works. The smart contract in the proposed framework manages the underlying transactions, including the following ones:

- Register user: This transaction allows new users to join the network. All users should have been registered prior to any other activities. This transaction allows meaningful traceability.
- Retrieve access control list (ACL): This transaction allows users to get the ACL regarding their data.
- Grant access: This transaction allows data owners to grant access to their data.
- Decline access: This transaction allows data owners to remove existing access to their data.
- Request access: This transaction allows peers to request access to the data; the data owner can use grant/decline transactions to respond to request transactions.

7.3.3 Gateway Component

Gateways in the proposed framework are analogous to IoT devices described in Section 6.3.4. The peers in the proposed framework can be data producers and data consumers. Hence, there are two supplementary processes in the gateway of this proposed framework that enable querying and retrieving the data. Moreover, this proposed framework's encryption scheme and identification are slightly different from the aforementioned gateway.

7.3.3.1 Encryption Process

As we detailed in Section 3.7, broadcast encryption (BE) is an answer for distributing information among multiple recipients. The idea of such an encryption system is to enable multiple parties to decrypt the ciphertext with only one encryption. BE scheme provides a multi-purpose secure channel without the need for redundant encryption of the same message.

In the proposed framework, we utilize BE in the IoT gateways to encrypt the incoming IoT data for the authorized group of recipients. This process exports a single cryptogram which all the peers can decrypt in the authorized group.

7.3.3.2 Identification Process

In Section 6.3.4.2, we have discussed an anonymized identification process to only allow the data owners to find their data. However, in this framework, we need to allow other peers to find their interest data. Hence, in this framework, we encrypt the identification information regarding the data using an HE scheme, see Section 3.6. Using HE allows the data owner to keep the confidentiality of their identification information and allows other peers to query the existing data based on their requirements. Since HE allows computation without revealing the data itself, HE encryption of identification information allows querying based on this information without revealing any additional knowledge from the identification information.

7.3.3.3 Query Process

Since our proposed framework's data are kept private, data consumers can not easily search or select their intended data using traditional approaches. Data consumers can establish some structured query language (SQL)-like queries and encrypt them. The queries are sent to the query execution component to allow its homomorphic execution.

7.3.3.4 Decryption Process

Data consumers are allowed to decrypt any cryptograms that are destined for them. Since the data are encrypted using BE scheme, the data consumers can decrypt the encrypted data using their key without any further required step.

7.3.4 Query Engine Component

A traditional query execution component is designed to allow queries on the stored data in a database and retrieve the query's data. In the proposed framework, the database is stored distributively, and the query-able information is encrypted homomorphically.

The encrypted queries from the data consumers are processed on homomorphically encrypted identification data and yield encrypted results. The result of a query consists of the data owners' address, which the data consumer would use to create a request access transaction.

7.3.5 Facade Layer

In order to facilitate the use of the proposed framework, a facade layer is proposed on top of all the other layers. The facade layer is the abstraction of the underlying complex layers.

7.3.6 Cryptology Algorithms

Cryptology algorithms are designed to provide confidentiality and integrity of the messages. In the proposed framework, we take advantage of various cryptology algorithms. The use of cryptology algorithms in the proposed framework is summarized as follows:

- Hash before Encryption: we assume the integrity of all messages is essential. Hence, all the data in the proposed framework accompany their secure hash signature prior to their encryption. The secure hash signature allows verification of their integration upon receipt.

- BE on IoT data: we assume all the IoT data are confidential and can be destined to multiple recipients. Therefore, IoT data are encrypted using BE scheme respecting the authorized group of access.

- HE on identification information: the identification information allow data consumers to find their intended data; however, they also contain private information that should be kept confidential. Hence, in the proposed framework, the HE scheme is applied to the identification information to allow querying without revealing the contents.

- Asymmetric encryption on queries/query results: The query and their result might contain sensitive information. However, since it is only destined for one recipient, applying BE is not encouraged. Therefore, in the proposed framework, such messages are encrypted with the public key of the recipient.

7.3.7 Workflow

In the proposed framework, two prominent roles, data owner and data consumer, exist. Moreover, smart contracts, distributed storage, gateways, and query engines are intermediate roles that enable the interactions between the two prominent roles. Three main workflows occur in the proposed framework: data owner registration, request access, and data sharing. A recurrent workflow in the proposed framework is data storage. We describe these workflows in the next sections.

7.3.7.1 Storage

The workflow regarding storing any data in the proposed framework is depicted in Figure 7.3. For any data storage: upon arrival of new data, regardless of its type, smart contracts verify the data's integrity and its sender; then the contents of data is transmitted to the distributed storage; when the latter has saved the data, it responds with the address of the stored data; the address of the data in the distributed storage is saved in the blockchain.

7.3.7.2 Data Owner Registration

Figure 7.4 depicts the general workflow of data owner registration. This simple workflow aims to store the data owner's identification information for future queries. The data owner prepares his/her identification information and homomorphically encrypts it. This encrypted information is stored using the storage subprocess, discussed in Section 7.3.7.1. Once the identification information is successfully stored, the registration workflow is terminated.

7.3.7.3 Request Access

The request access workflow involves all the major roles of the proposed framework; this workflow is depicted in Figure 7.4. Data consumer prepares a query based on his/her requirement. The query might include intended identification information such as the name of the data owner. This query is encrypted homomorphically and send to the DApps. Before forwarding the query to the query engine, DApps store the query for future references. The query engine executes the query homomorphically and re-encrypts the result for the requestor. The result might include the encryption of access information, e.g., the intended data owner's address and public key. DApps store and forward the encrypted result

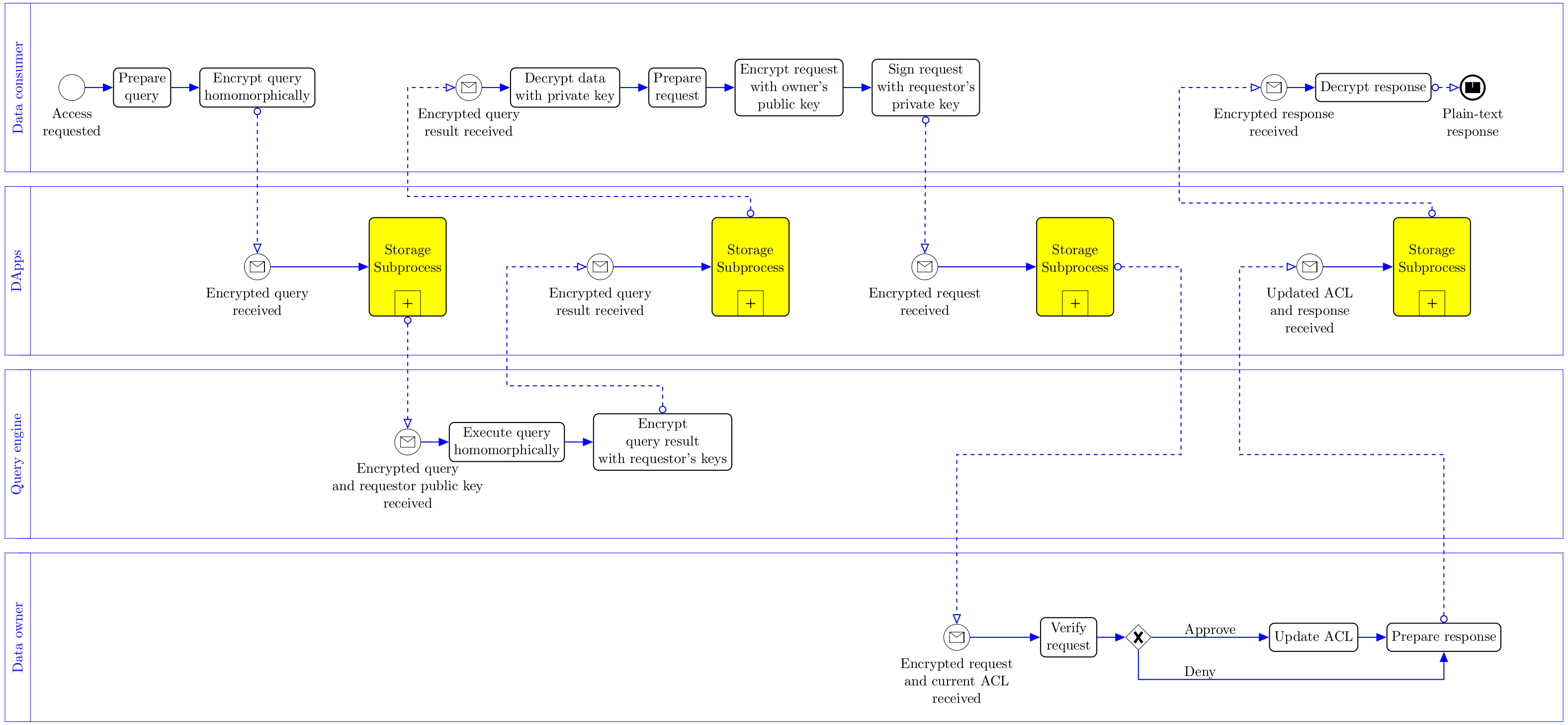

Data consumer
Access requested
Prepare query
Encrypt query homomorphically
Encrypted query result received
Decrypt data with private key
Prepare request
Encrypt request with owner's public key
Sign request with requestor's private key
Encrypted response received
Decrypt response
Plain-text response
DApps
Encrypted query received
Storage Subprocess
Encrypted query result received
Storage Subprocess
Encrypted request received
Storage Subprocess
Updated ACL and response received
Storage Subprocess
Query engine
Encrypted query and requestor public key received
Execute query homomorphically
Encrypt query result with requestor's keys
Data owner
Encrypted request and current ACL received
Verify request
Approve
Deny
Update ACL
Prepare response

Figure 7.2: Workflow of request access in the proposed secure framework based on blockchain

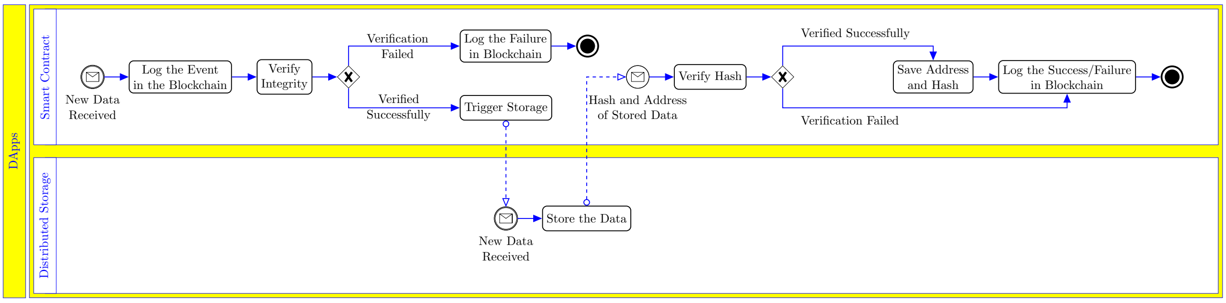


Figure 7.3: Workflow of storage of any data in the proposed secure framework based on blockchain

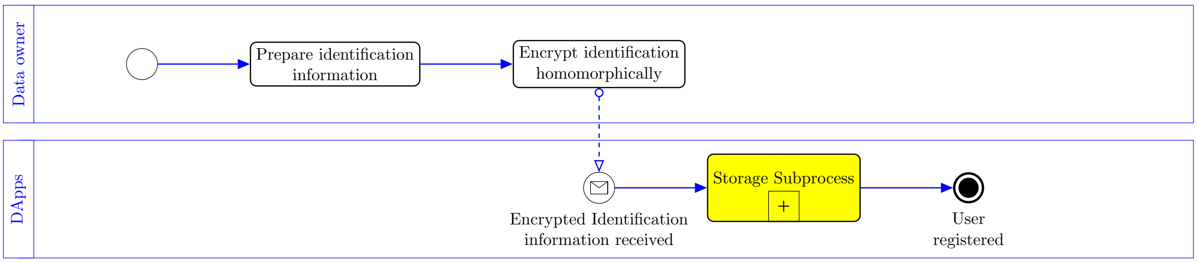

Data owner
DApps
Prepare identification information
Encrypt identification homomorphically
Encrypted Identification information received
Storage Subprocess
+
User registered

Figure 7.4: Workflow of registration in the proposed secure framework based on blockchain

to the data consumer; then, the latter decrypts and prepares an access request based on the query results. The data consumer signs the access request with his/her key and encrypts it with the data owner's public key. DApps store and shares the access request with the data owner. The latter might decrypt the request with his/her key and verify the signature based on the requestor's public key. Data owner might update his/her ACL if he/she decides to grant permission to the requestor. The request access workflow terminates by informing the data consumer about the data owner's response to the access request.

7.3.7.4 Data Sharing

Data sharing workflow is arguably the most frequent workflow compared to registration and request access ones. Figure 7.5 presents an overview of data sharing workflow. The workflow initiates with capturing new data in an IoT device on the data owner's side. Data owner might retrieve their ACL from DApps. ACL includes the keys of authorized recipients, which is used for the broadcast encryption scheme. Upon receiving the ACL, the data owner's gateway can encrypt the new data using the broadcast encryption scheme and transfer it to DApps for storage. Any peers can retrieve the encrypted data from DApps. While only the authorized recipients defined in the ACL can decrypt and access the

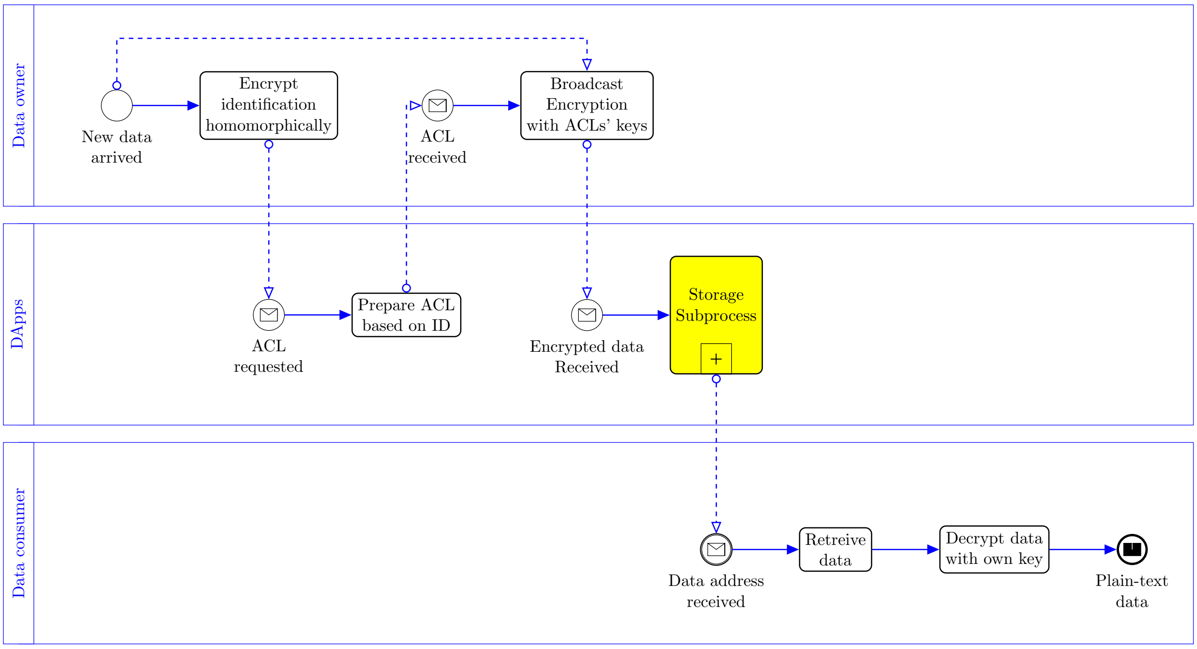


Figure 7.5: Workflow of data sharing in the proposed secure framework based on blockchain

new data in plain form. The latter terminates the data sharing workflow.

The required encryption and storage for data sharing in the proposed framework is $O(N_{messages})$; in which $N_{messages}$ represents the number of messages, respectively. In the existing frameworks in this context that use asymmetric encryption for privacy conserving data sharing, the complexity is $O(N_{messages} \times N_{receivers})$, in which $N_{receivers}$ represents the number of receivers.

7.4 Security Evaluation

In this section, we analyze the security characteristics of our proposed framework. The notations used in this section are summarized in Table 7.1.

7.4.1 Formal Description

7.4.1.1 Proposed Framwork

As discussed in Section 7.3.7, there are three main processes in the proposed framework. The logical dataflow regarding these three processes is depicted in Figure 7.6, where the processes are segregated using a horizontal dotted separator. The inner processes of the data lanes are omitted for simplicity.

Table 7.1: Notations used in formal modeling of the proposed blockchain-based secure data-sharing framework

Notation	Description
ID	Identification information
qry	Query
lst	List of matched items, i.e., the result of a query
c	Data consumer
o	Data owner
d	Raw Data
acl	ACL, i.e., the keys of authorized recipients
P_u	Public key of user u
req	Access request
π	Key used for homomorphic encryption
S_u	Signature of user u.
$E(m)$	Encryption on message m
$C_m^{P_u}$	Cryptogram on message m with the public key of user u
C_m^{π}	Cryptogram on message m using homomorphic encryption
C_m^{acl}	Cryptogram on message m using broadcast encryption
$D(c)$	Decryption of message m
$Query(qry)$	Classification of message m
$H(m)$	Secure hash digest on message m
h_m^S	Digest value of message m with the signature S
a_m	Address of message m in the distributed storage
u	User of the system
$Own(u, m)$	Is u the owner of message m
$Acc(u, m)$	Has u read access to message m
$Send(u_1, u_2, m)$	User u_1 has send message m to user u_2
$Allow(u_1, u_2, m)$	User u_1 has granted access for message m to user u_2

All messages sent to DApps are coupled with their hash digest for verification in smart contracts; however, these hash digests are not shown in the data flow for better readability.

Data owner's registration: For the registration, data owners prepare their identification information internally (id) and encrypt it using homomorphic encryption (C_{id}^{π}). This message, alongside the hash digest value of the identification information (h_{id}) and hash digest of the whole message, is transmitted to DApps. DApps store these data in the blockchain.

Request access: For requesting access, the first step is to find the data owner. To this end, the data consumer prepares a query (qry) and encrypts using homomorphic encryption (C_{qry}^{π}) which are sent to DApps alongside its hash value. After internal verification and storage, DApps allow the query

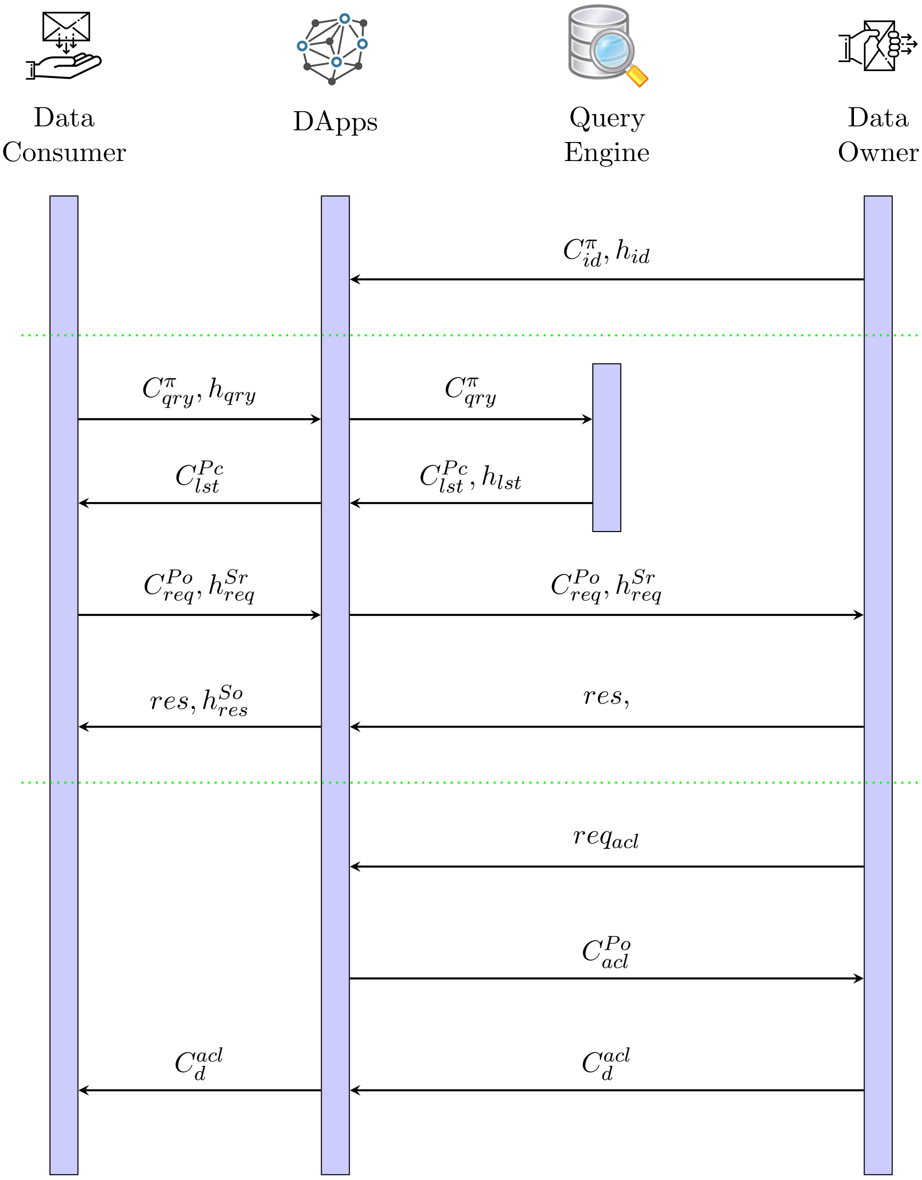


Figure 7.6: Logical flow execution of the proposed framework

engine to read and execute the encrypted query. The latter re-encrypts the results using the requestor's public key (C_{lst}^{Pc}), and sends it alongside the hash digest of the results (h_{lst}) and the hash digest of the whole message to DApps. DApps internally verify and store the encrypted results and allows the data consumer to read the encrypted result. The data consumer, who has issued the query, can decrypt the result of his/her query (lst). He/she may now use this information and form an access request(s) (req); he/she then encrypt using the public key of the intended data owner(s) (C_{req}^{Po}) and sign it using his/her signature (h_{req}^{Sr}). These messages are transmitted to the data owner(s) via DApps. Data owner decides whether to approve or reject the request using a response message res and sign it using his/her signature (h_{res}^{So}). In the case of approval, he/she updates ACL and encrypts it using his/her key (C_{acl}^{Po}); and sends all these messages to DApps for storage. Although for simplicity, the workflow and dataflow did not include message type for the access revocation, the latter is easily addable to the request and response, allowing a fine-grained access control.

Data sharing: For data sharing, the first step is to retrieve the list of authorized recipients. The data owner can retrieve his/her the encrypted ACL from DApps (C_{acl}^{Po}) and decrypt it locally. Then data owner uses the keys of authorized recipients to encrypt the new data (d) using broadcast encryption, resulting in a single cryptogram (C_{d}^{acl}). The latter, alongside its hash value, is verified and stored using DApps. Any authorized recipient can now retrieve and decrypt the cryptogram to access the data (d) in plain form.

7.4.1.2 Requirements

The proposed framework requirements include the requirements explained in Section 6.4.1.2. The confidentiality requirement of this proposed framework is categorized into threes forms:

- confidentiality of identification information and query (Conf. of query),
- confidentiality of metadata, e.g., request, results, ACL (Conf. of metadata),
- confidentiality of data IoT captured data.

The first two forms can be formalized using Equations 7.1 and 7.2; while the last one is, in fact, the access control requirement.

$$\text{Conf. of identification} \quad : \quad \forall ID, \forall u : Acc(u, ID) \iff Own(u, ID) \vee u = q \tag{7.1}$$

$$\text{Conf. of metadata} \quad : \quad \forall m, \forall u : Acc(u, m) \iff Own(u, m) \vee Send(o, u, m) \tag{7.2}$$

Regarding the additional security requirements of this proposed framework, reviewed in Section 7.2, The traceable access control requirement is, in fact, a subset of transparency, Equation 6.13; and the access control requirement can be formalized in Equation 7.3:

$$\begin{aligned} \text{Access control} : \forall Message\ m, \forall User\ u \quad &: \quad Acc(u, m) \\ &\iff \quad User\ o : Own(o, m) \wedge Allow(o, u, m) \end{aligned} \tag{7.3}$$

7.4.1.3 Assumptions

In this proposed framework, we have similar assumptions reviewed in Section 6.4.1.3. However, in this proposed framework, all users have access to the public keys of all nodes, and only the owner of the key has access to the private key.

Moreover, we also assume the security of the broadcast encryption scheme as well as asymmetric encryption one; both can be similarly formalized as Equation 6.9. Moreover, at least one trusted party is required for handling the homomorphic encryption and decrypt the query results; for simplicity, we assumed this is the same user as the query engine; however, this trusted user can be any other node.

7.4.2 Formal Verification

Since the procedure of storage of the data and verification of hash digest is similar, we can use the same proofs presented in Section 6.4.2 for proving the integrity, ownership, availability, and transparency of the proposed framework.

7.4.2.1 Confidentiality of Identification Information and Query

The identification information of users is never stored in the blockchain nor in the distributed storage. Based on the security of the homomorphic encryption scheme used for this encryption scheme, no user can gain any information about them. We use indirect proof for confidentiality: The contradictory assumption is that there is an unauthorized user $\hat{u}$ with access to the identification or query m (Equation 7.4). Since we assumed the security of the owner, the unauthorized user have accessed the message from its cryptogram (Equation 7.5). However, due to the encryption scheme security characteristics formalized in Equation 6.9, it narrows down the unauthorized user to the holder of the encryption key, which contradicts the assumptions of secure party handling the homomorphic encryption and proves the confidentiality.

$$\begin{aligned}
& & & \exists \hat{u}, m : Acc(\hat{u}, m) \wedge \overline{Own(\hat{u}, m)} \wedge u \neq q & (7.4)\\
& & \Longrightarrow\; & \exists \hat{u}, m : Acc(\hat{u}, c^{\pi}_{m}) \Longrightarrow Acc(\hat{u}, m) \wedge \overline{Own(\hat{u}, m)} \wedge u \neq q & (7.5)\\
& \text{Equation 6.9} & : \; & \Big(Acc(\hat{u}, c^{\pi}_{m}) \Longrightarrow Acc(\hat{u}, m)\Big) \Longrightarrow Own(\hat{u}, \pi) & (7.6)\\
& & \Longrightarrow\; & \hat{u} = q \Longrightarrow \bot & (7.7)\\
& & \Longrightarrow\; & \forall ID, \forall u : Acc(u, ID) \Longleftrightarrow Own(u, ID) \vee u = q & (7.8)
\end{aligned}$$

7.4.2.2 Confidentiality of Metadata

The metadata is intended for only one recipient; hence, in the proposed framework, they are encrypted using normal asymmetric encryption schemes. The sender encrypts the metadata using the receiver's public key and stores it in the blockchain and distributed storage, and the receiver can only decrypt it using his/her private key. The confidentiality of metadata can be formally proved the same as the confidentiality of query information, presented in Section 7.4.2.1.

We use indirect proof for the confidentiality of data: The contradictory assumption is that there is an unauthorized user $\hat{u}$ with access to data m (Equation 7.9). Since we assumed the security of the client-sides and the fact the data is always encrypted in the proposed framework; hence, the unauthorized user has accessed the message from its cryptogram (Equation 7.10). However, due to the encryption scheme security characteristics formalized in Equation 6.9, it narrows down the unauthorized user to the holder of the private keys. Since only the receiver has the private key of the encryption (Equation 7.11), it contradicts the security of the client-side. This contradiction proves the initial assumption is impossible and hence proves the confidentiality of data.

$$
\begin{align*}
& & \exists \hat{u}, m : Acc(\hat{u}, m) \wedge \overline{Own(\hat{u}, m)} \wedge \overline{Send(\hat{u}, o, m)} \tag{7.9}\\
& \Longrightarrow & \exists \hat{u}, d : Acc(\hat{u}, c_m^{P_u}) \Longrightarrow Acc(\hat{u}, m) \wedge \overline{Own(\hat{u}, m)} \wedge \tag{7.10}\\
\text{Equation 6.9} & : & \Big(Acc(\hat{u}, c_m^{P_u}) \Longrightarrow Acc(\hat{u}, m)\Big) \Longrightarrow Acc(\hat{u}, Pr_u)\\
& \Longrightarrow & Send(o, \hat{u}, m) \tag{7.11}\\
& \Longrightarrow & \perp\\
& \Longrightarrow & \forall m, u : Acc(u, m) \Longleftrightarrow Own(u, m) \vee Send(o, u, m)
\end{align*}
$$

7.4.2.3 Confidentiality of Data (access control)

The data is protected via the security of the broadcast encryption scheme. Hence, the confidentiality of data depends on the confidentiality of metadata, i.e., ACL, and the confidentiality of the encryption schemes, which are proved and assumed, respectively. We use indirect proof for the confidentiality of data: The contradictory assumption is that there is an unauthorized user $\hat{u}$ with access to data d (Equation 7.12). Since we assumed the security of the client-sides and the fact the data is always encrypted in the proposed framework; hence, the unauthorized user has accessed the message from its cryptogram (Equation 7.13). However, due to the encryption scheme security characteristics formalized in Equation 6.9, it narrows down the unauthorized user to the holder of the encryption keys of broadcast or the receiver of ACL. Since in the proposed framework ACL is only shared with the data owner (Equation 7.14); it contradicts the confidentiality of metadata proved in Section 7.4.2.2. This contradiction proves the initial assumption is impossible and hence proves the confidentiality of data.

$$
\begin{align*}
& & \exists \hat{u}, m : Acc(\hat{u}, d) \wedge \overline{Own(\hat{u}, d)} \wedge \overline{Allow(\hat{u}, o, d)} \tag{7.12}\\
& \Longrightarrow & \exists \hat{u}, d : Acc(\hat{u}, c_d^{acl}) \Longrightarrow Acc(\hat{u}, d) \wedge \overline{Own(\hat{u}, d)} \wedge \tag{7.13}\\
\text{Equation 6.9} & : & \Big(Acc(\hat{u}, c_d^{a}cl) \Longrightarrow Acc(\hat{u}, d)\Big) \Longrightarrow Acc(\hat{u}, acl)\\
\text{Section 7.4.2.2} & : & Acc(\hat{u}, ACL) \Longleftrightarrow Own(\hat{u}, d) \vee Send(o, \hat{u}, d)\\
& \Longrightarrow & Send(o, \hat{u}, m) \tag{7.14}\\
& \Longrightarrow & \hat{u} = o \Longrightarrow \perp\\
& \Longrightarrow & \forall d, u : Acc(u, m) \Longleftrightarrow User\ o : Own(o, m) \wedge Allow(o, u, m)
\end{align*}
$$

7.5 HapiChain: A Teleconsultation Use Case of the Proposed Secure Data-sharing Framework

In order to depict the benefits of the proposed secure framework for data-sharing, we discuss a healthcare use case in this chapter. Teleconsultation is complex and includes numerous members interacting and sharing data with each other; hence, a teleconsultation application, named HapiChain, is discussed as the use case of this proposed framework.

7.5.1 Motivation

The emergence of technology has important impacts on several application domains, including ambient assisted living [65], [66], [68], rehabilitation [67], and healthcare [55], [56], [69], [70]. In the healthcare domain, the necessity of a teleconsultation system is inevitable, especially with the outbreak of coronavirus (SARS-CoV-2) in 2020.

As the governments implement lockdowns to stop the contamination cycles of this lethal virus, access to medical care become more difficult. The need is more critical for the elderly and people with chronic diseases prone to deaths because of coronaviruses [75]; they are also in need of regular visits to their doctors. On normal days, there are 0.6 consultations per year for each aged person [102]. However, the need for consultation arises during a health crisis. On the other hand, access to clinicians is not evenly distributed; people living in suburbs and small towns might require long trips to visit their doctors, which puts them at higher risk of contamination. Therefore, teleconsultation has recently achieved a huge amount of attention to deal with these challenges.

Besides the health benefits of teleconsultations during viral outbreaks, remote treatments tend to be economically and environmentally better solutions as they remove the need to travel for medical treatments.

Although teleconsultation has undeniable benefits over face-to-face consultations, there are various concerns that slow the deployment of such systems. These concerns can be generally categorized into two classes: (1) the security of medical information and (2) the financial aspects of such systems. One of the most tangible examples of the former class of concerns is the confidentiality of the data in transit required for teleconsultation, e.g., medical files, prescription, and video calls. Because of the criticality of medical information, they are used to be stored in an isolated network; however, for

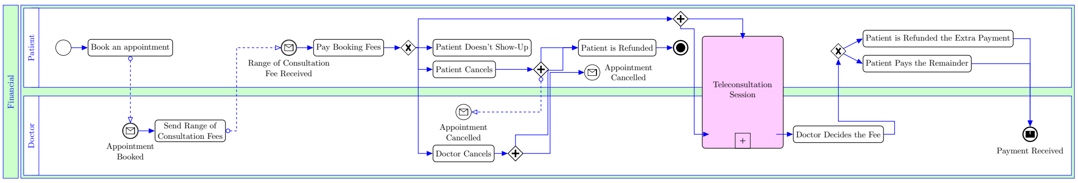


Figure 7.7: The simplified process model of a financial aspects of a teleconsultation session

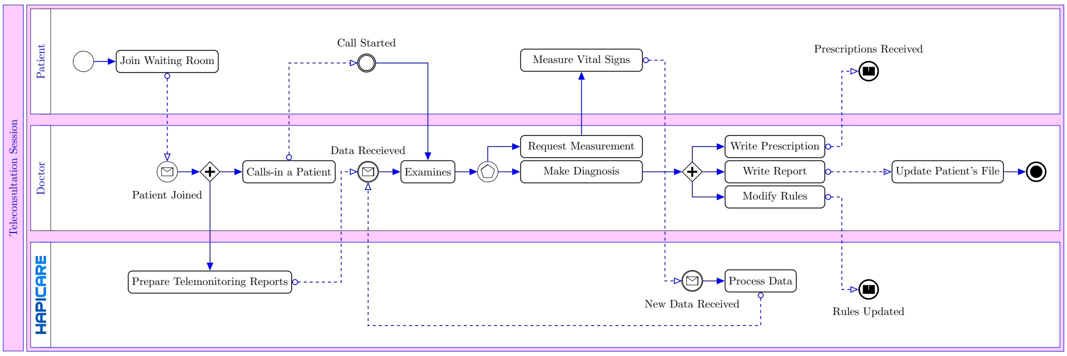


Figure 7.8: The simplified process model of a teleconsultation session

a teleconsultation, this practice is no longer applicable, which raises this concern of confidentiality. The latter class of concerns roughly exists in most online financial systems; however, teleconsultation is arguably more complicated. Common online systems only consist of providers and consumers; contrarily, in teleconsultation, insurance acts as a third actor.

Given the above requirements, we discuss using the proposed secure blockchain-based framework for teleconsultation. The proposed framework provides secure data-sharing of private information and also enables transparent transaction history for the audits.

7.5.2 Teleconsultation

Teleconsultation is defined as medical consultation services provided at a distance. Teleconsultation typically concerns two actors, namely a doctor and a patient, while the shared resources are agenda, waiting room, video channel, prescription, and consultation report. However, in the HapiChain teleconsultation service, we include Hapicare as an additional actor, and consequently, its rules, sensor data, contextual information, and diagnosis reports as additional resources. Multiple workflows form a teleconsultation service. The followings are the main workflows in teleconsultation:

1. Agenda management: The doctor provides a list of his/her available times, and the patient selects

and books one of them.

2. Virtual examination and prescription: The doctor asks the patient for symptoms and exam results for his/her diagnosis and consequently providing a prescription.

3. Finance management: The doctor bills the patient based on the type of consultation, based on the insurance plan of the patient, the sum will be paid to the doctor via the patient and the insurance company.

4. Income share: The amount might be received by a clinic, which would share the amount with the doctor and the infrastructure provider.

An overview of the third workflow is shown in Figure 7.7; in which the second workflow is depicted as a box. Before the teleconsultation session starts, a few prerequisite steps happen: a doctor previously has provided some time-slot for the times at which he/she is available for teleconsultation sessions; then a patient upon his/her will or based on the suggestion of Hapicare books one of them. It is common in medical consultations that the exact fee is not decided until the end of the session, as different types of consultations cost differently. At the time of booking the appointment, the patient would select a reason for consultation; on that basis, a range of consultation fees are provided. One of the doctors' concerns is false bookings; i.e., since the procedure of booking is very simple, malicious users book the time-slots of a doctor without the intention of using them; it blocks real patients from accessing the doctors. To this end, one common practice is prior payment; since the exact consultation fee is not decided in the booking time, the patient would pay a defined fee enough to discourage false booking. Either party can cancel the appointment, resulting in the refunding the prepayment. Upon concluding the teleconsultation session, the doctor decided on a consultation fee; if this fee is lower than the prepayment, the extra amount is returned to the patient; otherwise, the patient pays the difference. Once the doctor has received the payment, it will be stored in his/her logbook for future reference for insurance reimbursements and fiscal aspects.

The core of teleconsultation is the *virtual examination and prescription* workflow; the simplified process model of this core workflow is shown in Figure 7.8. The core workflow, depicted as a box in Figure 7.7 and detailed in 7.8, starts on the time of appointment. The patient logs-in to the system to join the virtual waiting room. Similar to physical consultation, a waiting room is vital to provide flexibility for doctors to spend enough time with each patient. Upon joining each patient, the doctor is notified, and the waiting room is updated. The doctor (or his/her secretary) can always check who is in the waiting room, consultation reasons and scheduled appointments, and how long patients are online in the waiting room. This information allows the doctor to manage better his/her time to visit all the patients in a timely manner.

Moreover, Hapicare starts preparing all the data about the patient who has joined the waiting room. Once the doctor selects a patient to start the video call, this data is transferred to the doctor to help him to have holistic information about the patient. The doctor examines the Hapicare file and asks some questions; he/she might ask for additional measurements. The patient can provide them using his/her connected sensors. After measuring the vital signs, Hapicare will collect the information and transfer them to the doctor. Once the doctor has enough information about the patient to make a diagnosis; he/she would typically write a prescription for the patient; the prescription might include additional diagnosis steps, such as blood exams or referral letters for a specialist. The doctor would also write some reports in the medical file of the patient. If necessary, he/she might also update some of the telemonitoring rules in Hapicare, based on his/her diagnosis; with the conclusion of the teleconsultation session, the rest of general workflow (see Figure 7.7) continues with *the doctor deciding a consultation fee.*

7.5.3 Discussion

Since the proposed framework is dedicated to data-sharing, we explore the data used in teleconsultation and how they can be handled using the proposed framework.

- Doctors' information: the doctors might decide to publish their information, including their contact information and availabilities, publicly or only share it with their patients. In the latter case, the ACL for each doctor's information includes the patients of that doctor, resulting in fine-grained access control and confidentiality. Moreover, in both cases, because availability and integrity are in the proven requirements of the proposed framework, it is guaranteed that the doctors' information is available at all times and without the risk of unauthorized changes.

- Patients' medical information: the patients' medical information might be shared with their generalist, recurrent doctors, nurse, and visiting hospital and Hapicare. The data sharing might be in full or partial, and also permanent or temporary. For instance, the patient might prefer not

Table 7.2: Summary of handling the teleconsultation data in the proposed secure blockchain data-sharing framework

Data	Data Owner	Data Consumer(s)	Security Scheme	Most Critical Concern
Doctors' info.	Doctors	Patients	BE[1]+DS[2]	Integrity
Patients' info.	Patient	Doctors, nurses, hospitals, Hapicare, etc.	BE+DS	Confidentiality
Hapicare Rules	Doctor	Hapicare and Patient	BE+DS	Confidentiality
Hapicare Results	Hapicare	Patient, nurse, and doctor	BE+DS	Confidentiality
Transactions	Doctors, Patients, Hapicare, the framework itself	Everyone	none + DS	Integrity
Call info.	Patients and Doctors	Doctors and Patients	externally	Availability
Call events	Patients and Doctors	Doctors and Patients	none + DS	Integrity

[1] Broadcast Encryption

[2] Digital Signature

to share his/her vital signs continuously with his/her doctor; but with Hapicare to avail him/her a telemonitoring service. Moreover, through the course of teleconsultation, he/she might want to allow the doctor to access their vital signs for remote measurement. This information can be securely shared with the above requirements using the proposed framework. The fine-grained access control enables any type of access based on the data type and time. Moreover, the proposed framework's integrity and availability characteristics guarantee the patients' data stay intact and available.

- Hapicare information: Besides the doctors' and patients' information, Hapicare information includes medical rules and monitoring reports. A doctor provides the former for a specific (group of) patient; hence, the medical rules are sent from doctors to Hapicare, with guaranteed security (confidentiality, integrity, and availability) in the proposed framework. Moreover, Hapicare generates monitoring reports and is shared with the patient, his/her nurse, and doctor, with guaranteed security using the proposed framework.

- Transactions: The events occurring during teleconsultation are required for financial and legal aspects. The proposed framework enables an immutable and available logbook of transactions with guaranteed security.
- Call information: although it is possible to use the proposed framework for sharing the data related to video calls, direct data-sharing of call information enables much higher performance. Moreover, a call's contents can grow huge to be stored in the proposed framework and usually are not useful. Hence, it is better to share the call events only, e.g., the patient started the call, the doctor joined the call, or the call is abandoned, using the proposed framework.

In this use case, we purposely put some information for the public to depict that the proposed framework can be used for public data sharing. However, based on the requirement, public data sharing, e.g., call information, can be transferred via secure data sharing. The summary of the above data and how they are handled is presented in Table 7.2.

7.6 Conclusion

In this chapter, we first discussed the importance of security in IoT networks, particularly in data sharing. Then we proposed a blockchain-based framework using a combination of cryptographic algorithms and blockchain to handle the security concerns in IoT networks. We formally modeled and evaluated our proposed framework regarding the security requirements. Moreover, the proposed framework is validated using a teleconsultation use case, named HapiChain, to depict how our proposed framework can handle security concerns in teleconsultation.

Conclusions and Perspectives

This chapter concludes the thesis. The first section of this chapter summarizes our contributions, while the second section provides an overview of the future research direction and how the contribution of this thesis might be extended.

Contents

7.7 Conclusion

The main objective of this thesis is to propose an intelligent and secure e-health system based on IoT to allow better handling of uncertainty and privacy concerns in the context of AmI. In this thesis, we have proposed three contributions to address the research challenges of e-health that are discussed in Section 2.2; which can be summarized as follows:

- We have proposed a self-adaptive telecare framework. This framework provides IoT-based telemonitoring with a smart selection of sensing action to provide a holistic image of patients with minimal intrusions. It uses ontology-based reasoning and probabilistic diagnosis to handle the data's heterogeneity and unreliability, as well as the medical rules' uncertainty to provide e-diagnosis. This framework embeds answer set programming as common-sense reasoning to provide self-adaptive and auto-personalized treatment services. A prototype of this framework has been developed and validated by clinicians and medical systems engineers collaborating with the Maidis company in the ITEA3 Medolution project. Moreover, e-diagnosis is evaluated using a comparative experiment to illustrate its strength in missing information. Additionally, the proposed framework is validated using four scenarios.

- We have proposed a secure framework for handling robustness and privacy issues in e-health or, generally, IoT networks. In the proposed framework, we focus on AI as the core computation of IoT networks. The proposed framework is based on blockchain technology and distributed storage to ensure integrity and availability in a decentralized network. The proposed framework uses homomorphic encryption to enable privacy-preserving computation, in particular AI services. The privacy requirements of IoT-based computations are defined, and it is formally proved that the proposed framework meets them.

- We have proposed a secure and robust framework for data sharing in e-health, or generally in IoT networks. For the robustness requirements of data sharing and storage, blockchain technology, distributed storage, and digital signature are used in the proposed framework. Moreover, the proposed framework uses a combination of cryptographic algorithms for providing secure data while avoiding undesired redundancy. It uses homomorphic encryption for privacy-preserving queries of information, asymmetric encryption for secure one-to-one communication, and broadcast encryption for secure one-to-many data sharing. The security requirements of data sharing and storage in IoT networks are defined, and their fulfilments are formally proved in the proposed framework. Additionally, the proposed framework is evaluated in a teleconsultation use case.

7.8 Perspectives

Several areas have still to be explored to achieve a complete e-health solution in the context of AmI. The perspectives resulting from this thesis can be summarized as follows:

- Real-world Evaluation of Telecare: Evaluation of telecare framework in the real-world environment is not straightforward; a comprehensive evaluation of a telecare system requires access to monitor real-world patients to verify whether the proposed solution is fully adapted to their needs. The aforementioned environment was not accessible due to legal and ethical restrictions. Hence, one interesting future work is to implement a pilot of telecare in a controlled environment. The pilot run should be pursued with medical experts' supervision to model various rules concerning medical conditions and episodes and then evaluate the monitoring, diagnosis, and remote treatment provided by the proposed telecare framework.

- Delegation in Access Control: Delegation of privileges can reduce the complexity and consequently improve usability, scalability, and manageability of access controls. To this end, it is an inspiring challenge to follow the delegation of rights in the secure blockchain-based data-sharing framework.

- Applicative Implementation of the Proposed Secure Framework for Homomorphic AI: The second contribution of this thesis addresses the privacy concerns in outsourcing AI computation in the context of AmI, and it is validated using use cases. However, the benefits of this framework can be augmented after a real-world case study. To this end, implementing this framework and evaluating it in real-world scenarios is an interesting applicative perspective of this thesis.

- Comparative Study of the Proposed Secure Data-sharing Framework: The third contribution of this thesis handles data-sharing among participants of IoT networks. An existing approach is using a permissioned blockchain for managing access control in a closed network. Although our proposed framework and permissioned blockchain are meant for different types of use, an interesting future work would be to study and compare these two approaches' security.

- Blockchain Attacks: The security of blockchain, was out of the scope of this thesis. There are numerous types of attacks in different blockchain layers, particularly in the application layer. Hence, an interesting future study is to simulate these attacks and analyze the proposed framework's implementations against such attacks.

Bibliography

[1] Global health and aging. https://www.who.int/ageing/publications/global_health.pdf, . [Online; accessed: 2020-09-28].

[2] Senior care industry analysis 2020 - cost trends. https://www.franchisehelp.com/industry-reports/senior-care-industry-analysis-2020-cost-trends/, . [Online; accessed: 2020-09-28].

[3] Sharding-FAQs. URL https://eth.wiki/sharding/Sharding-FAQs. Library Catalog: eth.wiki.

[4] Silvia Acid, Luis M. de Campos, Juan M. Fernández-Luna, Susana Rodrıguez, José Marıa Rodrıguez, and José Luis Salcedo. A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. *Artificial Intelligence in Medicine*, 30(3):215–232, March 2004. ISSN 09333657. doi:10.1016/j.artmed.2003.11.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0933365703001325.

[5] Muhammad Afzal, Syed Imran Ali, Rahman Ali, Maqbool Hussain, Taqdir Ali, Wajahat Ali Khan, Muhammad Bilal Amin, Byeong Ho Kang, and Sungyoung Lee. Personalization of wellness recommendations using contextual interpretation. *Expert Systems with Applications*, 96:506–521, April 2018. ISSN 09574174. doi:10.1016/j.eswa.2017.11.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0957417417307480.

[6] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, Srinivasan Muralidharan, Chet Murthy, Binh Nguyen, Manish Sethi, Gari Singh, Keith Smith, Alessandro Sorniotti, Chrysoula Stathakopoulou, Marko Vukolić, Sharon Weed Cocco, and Jason Yellick. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. *Proceedings*

of the Thirteenth EuroSys Conference, pages 1–15, April 2018. doi:10.1145/3190508.3190538. URL `http://arxiv.org/abs/1801.10228`. arXiv: 1801.10228.

[7] Michal Bali. *Drools JBoss Rules 5.0 developer's guide: develop rules-based business logic using the Drools platform.* From technologies to solutions. Packt Publ, Birmingham, 2009. ISBN 978-1-84719-564-7. OCLC: 845547425.

[8] Juan Benet. IPFS - Content Addressed, Versioned, P2P File System. page 11.

[9] Brennan Bennett. Using Telehealth as a Model for Blockchain HIT Adoption. *Telehealth and Medicine Today*, 2(4), May 2018. ISSN 2471-6960. doi:10.30953/tmt.v2.25. URL `http://telehealthandmedicinetoday.com/index.php/journal/article/view/25`.

[10] Anam Bhatti, Asad Ali Siyal, Adeel Mehdi, Huma Shah, Hinesh Kumar, and Muhammad Ali Bohyo. Development of cost-effective tele-monitoring system for remote area patients. In *2018 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–7, Lahore, Pakistan, February 2018. IEEE. ISBN 978-1-5386-3205-5. doi:10.1109/ICEET1.2018.8338646. URL `http://ieeexplore.ieee.org/document/8338646/`.

[11] Roy Billinton and Dange Huang. Aleatory and Epistemic Uncertainty Considerations in Power System Reliability Evaluation. In *Proceedings of the 10th International Conference on Probablistic Methods Applied to Power Systems*, pages 1–8, Puerto Rico, May 2008. IEEE.

[12] Dan Boneh and Alice Silverberg. Applications of Multilinear Forms to Cryptography. Technical Report 080, 2002. URL `http://eprint.iacr.org/2002/080`.

[13] Pim Borst. *Construction of engineering ontologies for knowledge sharing and reuse.* PhD thesis, Centre for Telematics and Information Technology, Enschede, 1997. OCLC: 68151508.

[14] Raphael Bost, Raluca Ada Popa, Stephen Tu, and Shafi Goldwasser. Machine Learning Classification over Encrypted Data. In *Proceedings 2015 Network and Distributed System Security Symposium*, San Diego, CA, 2015. Internet Society. ISBN 978-1-891562-38-9. doi:10.14722/ndss.2015.23241. URL `https://www.ndss-symposium.org/ndss2015/ndss-2015-programme/machine-learning-classification-over-encrypted-data/`.

[15] Magda Bucholc, Xuemei Ding, Haiying Wang, David H. Glass, Hui Wang, Girijesh Prasad, Liam P. Maguire, Anthony J. Bjourson, Paula L. McClean, Stephen Todd, David P. Finn, and KongFatt Wong-Lin. A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. *Expert Systems with Applications*, 130:157–171, September 2019. ISSN 09574174. doi:10.1016/j.eswa.2019.04.022. URL `https://linkinghub.elsevier.com/retrieve/pii/S0957417419302520`.

[16] Magdalena Bujnowska-Fedak and Urszula Grata-Borkowska. Use of telemedicine-based care for the aging and elderly: promises and pitfalls. *Smart Homecare Technology and TeleHealth*, page 91, May 2015. ISSN 2253-1564. doi:10.2147/SHTT.S59498.

[17] Vitalik Buterin. Ethereum White Paper: A Next-Generation Smart Contract and Decentralized Application Platform, 2014. URL `https://github.com/ethereum/wiki/wiki/White-Paper`.

[18] Mumin Cebe, Enes Erdin, Kemal Akkaya, Hidayet Aksu, and Selcuk Uluagac. Block4Forensic: An Integrated Lightweight Blockchain Framework for Forensics Applications of Connected Vehicles. *IEEE Communications Magazine*, 56(10):50–57, October 2018. ISSN 0163-6804, 1558-1896. doi:10.1109/MCOM.2018.1800137. URL `https://ieeexplore.ieee.org/document/8493118/`.

[19] Hei Chan and Adnan Darwiche. Sensitivity Analysis in Bayesian Networks: From Single to Multiple Parameters. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, pages 67–75, Arlington, Virginia, USA, 2004. AUAI Press. ISBN 0-9749039-0-6. event-place: Banff, Canada.

[20] Jung Hee Cheon, Kyoohyung Han, Seong-Min Hong, Hyoun Jin Kim, Junsoo Kim, Suseong Kim, Hosung Seo, Hyungbo Shim, and Yongsoo Song. Toward a Secure Drone System: Flying With Real-Time Homomorphic Authenticated Encryption. *IEEE Access*, 6:24325–24339, 2018. ISSN 2169-3536. doi:10.1109/ACCESS.2018.2819189. URL `https://ieeexplore.ieee.org/document/8325268/`.

[21] Mohammad Jabed Morshed Chowdhury, Alan Colman, Muhammad Ashad Kabir, Jun Han, and Paul Sarda. Blockchain Versus Database: A Critical Analysis. page 6.

[22] Gregory F. Cooper and Edward Herskovits. A Bayesian method for the induction of probabilistic

networks from data. *Machine Learning*, 9(4):309–347, October 1992. ISSN 0885-6125, 1573-0565. doi:10.1007/BF00994110. URL `http://link.springer.com/10.1007/BF00994110`.

[23] National Research Council et al. Preparing for an aging world: The case for cross-national research. 2001.

[24] Hong-Ning Dai, Zibin Zheng, and Yan Zhang. Blockchain for Internet of Things: A Survey. *IEEE Internet of Things Journal*, 6(5):8076–8094, October 2019. ISSN 2327-4662, 2372-2541. doi:10.1109/JIOT.2019.2920987. URL `https://ieeexplore.ieee.org/document/8731639/`.

[25] Jeroen S. de Bruin, Christian Schuh, Walter Seeling, Eva Luger, Michaela Gall, Elisabeth Hütterer, Gabriela Kornek, Bernhard Ludvik, Friedrich Hoppichler, and Karin Schindler. Assessing the feasibility of a mobile health-supported clinical decision support system for nutritional triage in oncology outpatients using Arden Syntax. *Artificial Intelligence in Medicine*, 92:34–42, November 2018. ISSN 09333657. doi:10.1016/j.artmed.2015.10.001. URL `https://linkinghub.elsevier.com/retrieve/pii/S0933365715001396`.

[26] Arthur P Dempster. Upper and lower probabilities induced by a multivalued mapping. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pages 57–72. Springer, 2008.

[27] Tien Tuan Anh Dinh, Ji Wang, Gang Chen, Rui Liu, Beng Chin Ooi, and Kian-Lee Tan. BLOCKBENCH: A Framework for Analyzing Private Blockchains. *arXiv:1703.04057 [cs]*, March 2017. URL `http://arxiv.org/abs/1703.04057`. arXiv: 1703.04057.

[28] A. Dorri, M. Steger, S. S. Kanhere, and R. Jurdak. BlockChain: A Distributed Solution to Automotive Security and Privacy. *IEEE Communications Magazine*, 55(12):119–125, December 2017. ISSN 1558-1896. doi:10.1109/MCOM.2017.1700879. Conference Name: IEEE Communications Magazine.

[29] Ali Dorri, Salil S. Kanhere, Raja Jurdak, and Praveen Gauravaram. LSB: A Lightweight Scalable BlockChain for IoT Security and Privacy. *Journal of Parallel and Distributed Computing*, 134:180–197, December 2019. ISSN 07437315. doi:10.1016/j.jpdc.2019.08.005. URL `http://arxiv.org/abs/1712.02969`. arXiv: 1712.02969.

[30] Dheeru Dua and Casey Graff. *UCI Machine Learning Repository*. University of California, Irvine, School of Information and Computer Sciences, 2017. URL `http://archive.ics.uci.edu/ml`.

[31] Ken Ducatel, Union européenne. Technologies de la société de l'information, Union européenne. Institut d'études de prospectives technologiques, and Union européenne. Société de l'information conviviale. Scenarios for ambient intelligence in 2010. 2001.

[32] Ashutosh Dwivedi, Gautam Srivastava, Shalini Dhar, and Rajani Singh. A Decentralized Privacy-Preserving Healthcare Blockchain for IoT. *Sensors*, 19(2):326, January 2019. ISSN 1424-8220. doi:10.3390/s19020326. URL `http://www.mdpi.com/1424-8220/19/2/326`.

[33] Natalia Díaz Rodríguez, Manuel P. Cuéllar, Johan Lilius, and Miguel Delgado Calvo-Flores. A fuzzy ontology for semantic modelling and recognition of human behaviour. *Knowledge-Based Systems*, 66:46–60, August 2014. ISSN 09507051. doi:10.1016/j.knosys.2014.04.016. URL `https://linkinghub.elsevier.com/retrieve/pii/S0950705114001385`.

[34] Shaker El-Sappagh, Francesco Franda, Farman Ali, and Kyung-Sup Kwak. SNOMED CT standard ontology based on the ontology for general medical science. *BMC Medical Informatics and Decision Making*, 18(1):76, December 2018. ISSN 1472-6947. doi:10.1186/s12911-018-0651-5. URL `https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-018-0651-5`.

[35] Esra Erdem, Michael Gelfond, and Nicola Leone. Applications of Answer Set Programming. *AI Magazine*, 37(3):53, October 2016. ISSN 0738-4602, 0738-4602. doi:10.1609/aimag.v37i3.2678. URL `https://aaai.org/ojs/index.php/aimagazine/article/view/2678`.

[36] Amos Fiat and Moni Naor. Broadcast Encryption. In Douglas R. Stinson, editor, *Advances in Cryptology — CRYPTO' 93*, pages 480–491, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg. ISBN 978-3-540-48329-8.

[37] Theodore Friedman. Adrenal Gland. In *Andreoli and Carpenter's Cecil Essentials of Medicine*, pages 642–652. Saunders, 9 edition, 2015. ISBN 1-4377-1899-X.

[38] Michael Glykas and Panagiotis Chytas. Technological innovations in asthma patient monitoring and care. *Expert Systems with Applications*, 27(1):121–131, July 2004. ISSN 09574174. doi:10.1016/j.eswa.2003.12.007. URL `https://linkinghub.elsevier.com/retrieve/pii/S0957417403002148`.

[39] Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data. In *Proceedings of the 13th ACM Conference*

on Computer and Communications Security, CCS '06, pages 89–98, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1-59593-518-5. doi:10.1145/1180405.1180418. URL https://doi.org/10.1145/1180405.1180418. event-place: Alexandria, Virginia, USA.

[40] Dennis Grishin, Kamal Obbad, Preston Estep, Kevin Quinn, Sarah Wait Zaranek, Alexander Wait Zaranek, Ward Vandewege, Tom Clegg, Nico César, Mirza Cifric, and George Church. Accelerating Genomic Data Generation and Facilitating Genomic Data Access Using Decentralization, Privacy-Preserving Technologies and Equitable Compensation. *Blockchain in Healthcare Today*, 1:1–23, 2018. ISSN 2573-8240. doi:10.30953/bhty.v1.34. URL https://blockchainhealthcaretoday.com/index.php/journal/article/view/34.

[41] H.Altay Güvenir, Gülşen Demiröz, and Nilsel İlter. Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. *Artificial Intelligence in Medicine*, 13(3):147–165, July 1998. ISSN 09333657. doi:10.1016/S0933-3657(98)00028-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0933365798000281.

[42] Carol Habib, Abdallah Makhoul, Rony Darazi, and Raphaël Couturier. Health risk assessment and decision-making for patient monitoring and decision-support using Wireless Body Sensor Networks. *Information Fusion*, 47:10–22, May 2019. ISSN 15662535. doi:10.1016/j.inffus.2018.06.008. URL https://linkinghub.elsevier.com/retrieve/pii/S156625351730790X.

[43] Mohamed Tahar Hammi, Patrick Bellot, and Ahmed Serhrouchni. BCTrust: A decentralized authentication blockchain-based mechanism. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, Barcelona, April 2018. IEEE. ISBN 978-1-5386-1734-2. doi:10.1109/WCNC.2018.8376948. URL https://ieeexplore.ieee.org/document/8376948/.

[44] Randall E Harris. *Epidemiology of chronic disease: global perspectives*. Jones & Bartlett Learning, 2019.

[45] Haya R Hasan, Khaled Salah, Raja Jayaraman, Junaid Arshad, Ibrar Yaqoob, Mohammed Omar, and Samer Ellahham. Blockchain-based Solution for COVID-19 Digital Medical Passports and Immunity Certificates. 8:222093–222108, December 2020. doi:10.1109/ACCESS.2020.3043350.

[46] Yutaka Hatakeyama, Taro Horino, Hiromi Kataoka, Tatsuki Matsumoto, Kazu Ode, Yoshiko Shimamura, Koji Ogata, Kosuke Inoue, Yoshinori Taniguchi, Yoshio Terada, and Yoshiyasu Okuhara. Incidence of acute kidney injury among patients with chronic kidney disease: a single-center retrospective database analysis. *Clinical and Experimental Nephrology*, 21(1):43–48, February 2017. ISSN 1437-7799. doi:10.1007/s10157-016-1243-2. URL https://doi.org/10.1007/s10157-016-1243-2.

[47] Huawei Huang, Jianru Lin, Baichuan Zheng, Zibin Zheng, and Jing Bian. When Blockchain Meets Distributed File Systems: An Overview, Challenges, and Open Issues. *IEEE Access*, 8:50574–50586, 2020. ISSN 2169-3536. doi:10.1109/ACCESS.2020.2979881. URL https://ieeexplore.ieee.org/document/9031420/.

[48] Sally C Inglis, Robyn A Clark, Finlay A McAlister, Jocasta Ball, Christian Lewinter, Damien Cullington, Simon Stewart, and John GF Cleland. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. *Cochrane database of systematic reviews*, (8), 2010.

[49] Wei Jiang, Dan Lin, Feng Li, and Elisa Bertino. Randomized and Efficient Authentication in Mobile Environments. *Cyber Center Publications*, page 16, February 2014. URL https://docs.lib.purdue.edu/ccpubs/633/.

[50] Zhanpeng Jin and Yu Chen. Telemedicine in the Cloud Era: Prospects and Challenges. *IEEE Pervasive Computing*, 14(1):54–61, January 2015. ISSN 1536-1268. doi:10.1109/MPRV.2015.19. URL http://ieeexplore.ieee.org/document/7030248/.

[51] Dmytro Khadzhynov, Danilo Schmidt, Juliane Hardt, Geraldine Rauch, Peter Gocke, Kai-Uwe Eckardt, and Kai M. Schmidt-Ott. The Incidence of Acute Kidney Injury and Associated Hospital Mortality. *Deutsches Aerzteblatt Online*, May 2019. ISSN 1866-0452. doi:10.3238/arztebl.2019.0397. URL https://www.aerzteblatt.de/10.3238/arztebl.2019.0397.

[52] Warren A. Kibbe, Cesar Arze, Victor Felix, Elvira Mitraka, Evan Bolton, Gang Fu, Christopher J. Mungall, Janos X. Binder, James Malone, Drashtti Vasant, Helen Parkinson, and Lynn M. Schriml. Disease Ontology 2015 update: an expanded and up-

dated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Research*, 43(D1):D1071–D1078, January 2015. ISSN 1362-4962, 0305-1048. doi:10.1093/nar/gku1011. URL `http://academic.oup.com/nar/article/43/D1/D1071/2435381/Disease-Ontology-2015-update-an-expanded-and`.

[53] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112, March 2009. ISSN 01674730. doi:10.1016/j.strusafe.2008.06.020. URL `https://linkinghub.elsevier.com/retrieve/pii/S0167473008000556`.

[54] Bran Knowles, Alison Smith-Renner, Forough Poursabzi-Sangdeh, Di Lu, and Halimat Alabi. Uncertainty in current and future health wearables. *Communications of the ACM*, 61(12):62–67, November 2018. ISSN 00010782. doi:10.1145/3199201. URL `http://dl.acm.org/citation.cfm?doid=3293542.3199201`.

[55] Hossain Kordestani, Roghayeh Mojarad, Abdelghani Chibani, Aomar Osmani, Yacine Amirat, Kamel Barkaoui, and Wagdy Zahran. Hapicare: A Healthcare Monitoring System with Self-Adaptive Coaching using Probabilistic Reasoning. In *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, pages 1–8, Abu Dhabi, United Arab Emirates, November 2019. IEEE. ISBN 978-1-72815-052-9. doi:10.1109/AICCSA47632.2019.9035291. URL `https://ieeexplore.ieee.org/document/9035291/`.

[56] Hossain Kordestani, Kamel Barkaoui, and Wagdy Zahran. HapiChain: A Blockchain-based Framework for Patient-Centric Telemedicine. In *2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH)*, pages 1–6, Vancouver, BC, Canada, August 2020. IEEE. ISBN 978-1-72819-042-6. doi:10.1109/SeGAH49190.2020.9201726. URL `https://ieeexplore.ieee.org/document/9201726/`.

[57] Li-wei Lehman, Mohammed Saeed, George Moody, and Roger Mark. Hypotension as a Risk Factor for Acute Kidney Injury in ICU Patients. *Comput Cardiol*, page 11, 2010.

[58] R. Li, T. Song, B. Mei, H. Li, X. Cheng, and L. Sun. Blockchain for Large-Scale Internet of Things Data Storage and Protection. *IEEE Transactions on Services Computing*, 12(5):762–771, September 2019. ISSN 1939-1374. doi:10.1109/TSC.2018.2853167. Conference Name: IEEE Transactions on Services Computing.

[59] Vladimir Lifschitz. Thirteen Definitions of a Stable Model. In Andreas Blass, Nachum Dershowitz, and Wolfgang Reisig, editors, *Fields of Logic and Computation*, volume 6300, pages 488–503. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-15024-1 978-3-642-15025-8. doi:10.1007/978-3-642-15025-8_24.

[60] Daniela Loreti, Federico Chesani, Paola Mello, Luca Roffia, Francesco Antoniazzi, Tullio Salmon Cinotti, Giacomo Paolini, Diego Masotti, and Alessandra Costanzo. Complex reactive event processing for assisted living: The Habitat project case study. *Expert Systems with Applications*, 126:200–217, July 2019. ISSN 09574174. doi:10.1016/j.eswa.2019.02.025. URL `https://linkinghub.elsevier.com/retrieve/pii/S0957417419301381`.

[61] Zaigham Mahmood. *Guide to Ambient Intelligence in the IoT Environment: Principles, Technologies and Applications*. Springer, 2019.

[62] Biljana Maric, Annemarie Kaan, Andrew Ignaszewski, and Scott A Lear. A systematic review of telemonitoring technologies in heart failure. *European journal of heart failure*, 11(5):506–517, 2009.

[63] Benjamin Miller and Marie O'Toole. episodes, 2003. URL `https://medical-dictionary.thefreedictionary.com/episodes`.

[64] Benjamin Miller and Marie O'Toole. telecare, 2003. URL `https://medical-dictionary.thefreedictionary.com/telecare`.

[65] R. Mojarad, F. Attal, A. Chibani, S. R. Fiorini, and Y. Amirat. Hybrid Approach for Human Activity Recognition by Ubiquitous Robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5660–5665, October 2018. doi:10.1109/IROS.2018.8594173. ISSN: 2153-0866.

[66] R. Mojarad, F. Attal, A. Chibani, and Y. Amirat. Automatic Classification Error Detection and Correction for Robust Human Activity Recognition. *IEEE Robotics and Automation Letters*, 5 (2):2208–2215, April 2020. ISSN 2377-3774. doi:10.1109/LRA.2020.2970667.

[67] R. Mojarad, F. Attal, A. Chibani, and Y. Amirat. Context-aware Adaptive Recommendation System for Personal Well-being Services. In *Proceedings of 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, November 2020.

[68] R. Mojarad, F. Attal, A. Chibani, and Y. Amirat. A Context-aware Hybrid Framework for Human Behavior Analysis. In *Proceedings of 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, November 2020.

[69] R. Mojarad, F. Attal, A. Chibani, and Y. Amirat. A Context-based Approach to Detect Abnormal Human Behaviors in Ambient Intelligent Systems. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD)*, September 2020.

[70] R. Mojarad, F. Attal, A. Chibani, and Y. Amirat. A Hybrid Context-aware Framework To Detect Abnormal Human Daily Living Behavior. In *Proceedings of IEEE World Congress on Computational Intelligence (WCCI)*, July 2020.

[71] Solvej Gradstein Helen Natarajan, Harish Krause. *Distributed Ledger Technology and Blockchain*. World Bank, 2017. doi:10.1596/29053. URL `https://elibrary.worldbank.org/doi/abs/10.1596/29053`. _eprint: https://elibrary.worldbank.org/doi/pdf/10.1596/29053.

[72] Michael Nofer, Peter Gomber, Oliver Hinz, and Dirk Schiereck. Blockchain. *Business & Information Systems Engineering*, 59(3):183–187, June 2017. ISSN 2363-7005, 1867-0202. doi:10.1007/s12599-017-0467-3. URL `http://link.springer.com/10.1007/s12599-017-0467-3`.

[73] Oscar Novo. Blockchain Meets IoT: An Architecture for Scalable Access Management in IoT. *IEEE Internet of Things Journal*, 5(2):1184–1195, April 2018. ISSN 2327-4662. doi:10.1109/JIOT.2018.2812239. URL `https://ieeexplore.ieee.org/document/8306880/`.

[74] C. Orlandi, A. Piva, and M. Barni. Oblivious Neural Network Computing via Homomorphic Encryption. *EURASIP Journal on Information Security*, 2007:1–11, 2007. ISSN 1687-4161, 1687-417X. doi:10.1155/2007/37343. URL `http://jis.eurasipjournals.com/content/2007/1/037343`.

[75] John Paget, Peter Spreeuwenberg, Vivek Charu, Robert J Taylor, A Danielle Iuliano, Joseph Bresee, Lone Simonsen, and Cecile Viboud. Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. *Journal of Global*

Health, 9(2):020421, December 2019. ISSN 2047-2978, 2047-2986. doi:10.7189/jogh.09.020421. URL `http://jogh.org/documents/issue201902/jogh-09-020421.pdf`.

[76] Pascal Paillier. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In Jacques Stern, editor, *Advances in Cryptology — EUROCRYPT '99*, pages 223–238, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg. ISBN 978-3-540-48910-8.

[77] Gianfranco Parati, Eamon Dolan, Richard J. McManus, and Stefano Omboni. Home blood pressure telemonitoring in the 21st century. *The Journal of Clinical Hypertension*, 20(7):1128–1132, July 2018. ISSN 15246175. doi:10.1111/jch.13305. URL `http://doi.wiley.com/10.1111/jch.13305`.

[78] Daejun Park, Yi Zhang, and Grigore Rosu. End-to-End Formal Verification of Ethereum 2.0 Deposit Smart Contract. In Shuvendu K. Lahiri and Chao Wang, editors, *Computer Aided Verification*, pages 151–164, Cham, 2020. Springer International Publishing. ISBN 978-3-030-53288-8.

[79] Roel Pieper. From Devices to "Ambient Intelligence": The Transformation of Consumer Electronics, June 1998. URL `http://epstein.org/wp-content/uploads/DLR-Final-Internal.ppt`.

[80] Khaled Rabieh, Mohamed M. E. A. Mahmoud, and Mohamed Younis. Privacy-Preserving Route Reporting Schemes for Traffic Management Systems. *IEEE Transactions on Vehicular Technology*, 66(3):2703–2713, March 2017. ISSN 0018-9545, 1939-9359. doi:10.1109/TVT.2016.2583466. URL `http://ieeexplore.ieee.org/document/7497514/`.

[81] Seyyed Mohammadreza Rahimi and Xin Wang. Location Recommendation Based on Periodicity of Human Activities and Location Categories. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu, editors, *Advances in Knowledge Discovery and Data Mining*, volume 7819, pages 377–389. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-37455-5 978-3-642-37456-2.

[82] Yogachandran Rahulamathavan, Raphael C.-W Phan, Muttukrishnan Rajarajan, Sudip Misra, and Ahmet Kondoz. Privacy-preserving blockchain based IoT ecosystem using attribute-based encryption. In *2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pages 1–6, Bhubaneswar, December 2017. IEEE. ISBN 978-1-5386-2347-3. doi:10.1109/ANTS.2017.8384164. URL `https://ieeexplore.ieee.org/document/8384164/`.

[83] Shailendra Rathore, Yi Pan, and Jong Hyuk Park. BlockDeepNet: A Blockchain-Based Secure Deep Learning for IoT Network. page 15, 2019.

[84] R L Rivest, L Adleman, and M L Dertouzos. On Data Banks and Privacy Homomorphisms. *Foundations of Secure Computation, Academia Press*, pages 169–179, 1978.

[85] R. L. Rivest, A. Shamir, and L. Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Commun. ACM*, 21(2):120–126, February 1978. ISSN 0001-0782. doi:10.1145/359340.359342. URL `https://doi.org/10.1145/359340.359342`. Place: New York, NY, USA Publisher: Association for Computing Machinery.

[86] Glenn Shafer. Dempster-Shafer Theory. *Encyclopedia of artificial intelligence*, 1:330–331, 1992.

[87] Pratima Sharma, Rajni Jindal, and Malaya Dutta Borah. Healthify: A Blockchain-Based Distributed Application for Health care. In Suyel Namasudra and Ganesh Chandra Deka, editors, *Applications of Blockchain in Healthcare*, volume 83, pages 171–198. Springer Singapore, Singapore, 2021. ISBN 9789811595462 9789811595479. doi:10.1007/978-981-15-9547-9_7. URL `http://link.springer.com/10.1007/978-981-15-9547-9_7`. Series Title: Studies in Big Data.

[88] Prakash P. Shenoy and Glenn Shafer. Propagating Belief Functions with Local Computations. *IEEE Expert*, 1(3):43–52, September 1986. ISSN 0885-9000, 2374-9407. doi:10.1109/MEX.1986.4306979. URL `https://ieeexplore.ieee.org/document/4306979/`.

[89] Rakesh Shrestha and Shiho Kim. Integration of IoT with blockchain and homomorphic encryption: Challenging issues and opportunities. In *Advances in Computers*, volume 115, pages 293–331. Elsevier, 2019. ISBN 978-0-12-817189-9. doi:10.1016/bs.adcom.2019.06.002. URL `https://linkinghub.elsevier.com/retrieve/pii/S0065245819300269`.

[90] Pushpa Singh and Narendra Singh. Blockchain With IoT and AI: A Review of Agriculture and Healthcare. *International Journal of Applied Evolutionary Computation*, 11(4):15, 2020.

[91] Sushil Kumar Singh, Shailendra Rathore, and Jong Hyuk Park. BlockIoTIntelligence: A Blockchain-enabled Intelligent IoT Architecture with Artificial Intelligence. *Future Generation Computer Systems*, 110:721–743, September 2020. ISSN 0167739X. doi:10.1016/j.future.2019.09.002. URL `https://linkinghub.elsevier.com/retrieve/pii/S0167739X19316474`.

[92] Socialstyrelsen. SNOMED CT-Grant of License of the Swedish National Release, July 2015. URL `https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/dokument-webb/ovrigt/snomed-ct-swedish-national-release-affiliate-licence.pdf`.

[93] Jun Song, ChunJiao He, Fan Yang, and HuanGuo Zhang. A privacy-preserving distance-based incentive scheme in opportunistic VANETs. *Security and Communication Networks*, 9(15):2789–2801, 2016. ISSN 1939-0122. doi:https://doi.org/10.1002/sec.1211. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sec.1211`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sec.1211.

[94] Mehwish Sultan, Kerry Kuluski, Warren J McIsaac, Joseph A Cafazzo, and Emily Seto. Turning challenges into design principles: Telemonitoring systems for patients with multiple chronic conditions. *Health Informatics Journal*, page 146045821774988, January 2018. ISSN 1460-4582, 1741-2811. doi:10.1177/1460458217749882. URL `http://journals.sagepub.com/doi/10.1177/1460458217749882`.

[95] Xiaoqiang Sun, Peng Zhang, Joseph K. Liu, Jianping Yu, and Weixin Xie. Private machine learning classification based on fully homomorphic encryption. *IEEE Transactions on Emerging Topics in Computing*, pages 1–13, 2018. ISSN 2168-6750. doi:10.1109/TETC.2018.2794611. URL `http://ieeexplore.ieee.org/document/8260844/`.

[96] Zhong Wei Sun and Wen Xiao Yan. A Privacy Preserving Scheme for Vehicle to Grid Networks Based on Homomorphic Cryptography. *Advanced Materials Research*, 1014:516–519, July 2014. ISSN 1662-8985. doi:10.4028/www.scientific.net/AMR.1014.516. URL `https://www.scientific.net/AMR.1014.516`.

[97] Lambert Surhone, Mariam Tennoe, and Susan Henssonow, editors. *Robust Bayes Analysis*. Betascript Publishing, 2011. ISBN 613-6-02418-7.

[98] Sonja Ulrich. Snow Owl MQ, September 2017. URL `http://b2i.sg/snow-owl-mq/`.

[99] K. Verbert, R. Babuška, and B. De Schutter. Bayesian and Dempster–Shafer reasoning for knowledge-based fault diagnosis–A comparative study. *Engineering Applications of Artificial Intelligence*, 60:136–150, April 2017. ISSN 09521976. doi:10.1016/j.engappai.2017.01.011. URL `https://linkinghub.elsevier.com/retrieve/pii/S0952197617300118`.

[100] Michele Vitacca, Alessandra Montini, and Laura Comini. How will telemedicine change clinical practice in chronic obstructive pulmonary disease? *Therapeutic Advances in Respiratory Disease*, 12:175346581875477, January 2018. ISSN 1753-4666, 1753-4666. doi:10.1177/1753465818754778. URL `http://journals.sagepub.com/doi/10.1177/1753465818754778`.

[101] Marko Vukolić. Hyperledger fabric: towards scalable blockchain for business. page 18, 2015.

[102] Yingying Wang, Kate Hunt, Irwin Nazareth, Nick Freemantle, and Irene Petersen. Do men consult less than women? An analysis of routinely collected UK general practice data. *BMJ Open*, 3(8):e003320, August 2013. ISSN 2044-6055, 2044-6055. doi:10.1136/bmjopen-2013-003320. URL `http://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2013-003320`.

[103] Ingo Weber, Vincent Gramoli, Alex Ponomarev, Mark Staples, Ralph Holz, An Binh Tran, and Paul Rimba. On Availability for Blockchain-Based Systems. In *2017 IEEE 36th Symposium on Reliable Distributed Systems (SRDS)*, pages 64–73, Hong Kong, Hong Kong, September 2017. IEEE. ISBN 978-1-5386-1679-6. doi:10.1109/SRDS.2017.15. URL `http://ieeexplore.ieee.org/document/8069069/`.

[104] Patricia L Whetzel, Natalya F Noy, Nigam H Shah, Paul R Alexander, Csongor Nyulas, Tania Tudorache, and Mark A Musen. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic acids research*, 39(suppl_2):W541–W545, 2011.

[105] WHO, editor. *Preventing chronic diseases: a vital investment.* World Health Organization ; Public Health Agency of Canada, Geneva : [Ottawa], 2005. ISBN 978-92-4-156300-0. OCLC: ocm62288123.

[106] Adrian Wong, Mary G Amato, Diane L Seger, Christine Rehr, Adam Wright, Sarah P Slight, Patrick E Beeler, E. John Orav, and David W Bates. Prospective evaluation of medication-related

clinical decision support over-rides in the intensive care unit. *BMJ Quality & Safety*, 27(9): 718–724, September 2018. ISSN 2044-5415, 2044-5423. doi:10.1136/bmjqs-2017-007531. URL http://qualitysafety.bmj.com/lookup/doi/10.1136/bmjqs-2017-007531.

[107] Boyi Xu, Lida Xu, Hongming Cai, Lihong Jiang, Yang Luo, and Yizhi Gu. The design of an m-Health monitoring system based on a cloud computing platform. *Enterprise Information Systems*, 11(1):17–36, January 2017. ISSN 1751-7575, 1751-7583. doi:10.1080/17517575.2015.1053416. URL https://www.tandfonline.com/doi/full/10.1080/17517575.2015.1053416.

[108] Li Yang, Guolan Xing, Li Wang, Yonggui Wu, Suhua Li, Gang Xu, Qiang He, Jianghua Chen, Menghua Chen, Xiaohua Liu, Zaizhi Zhu, Lin Yang, Xiyan Lian, Feng Ding, Yun Li, Huamin Wang, Jianqin Wang, Rong Wang, Changlin Mei, Jixian Xu, Rongshan Li, Juan Cao, Liang Zhang, Yan Wang, Jinhua Xu, Beiyan Bao, Bicheng Liu, Hongyu Chen, Shaomei Li, Yan Zha, Qiong Luo, Dongcheng Chen, Yulan Shen, Yunhua Liao, Zhengrong Zhang, Xianqiu Wang, Kun Zhang, Luojin Liu, Peiju Mao, Chunxiang Guo, Jiangang Li, Zhenfu Wang, Shoujun Bai, Shuangjie Shi, Yafang Wang, Jinwei Wang, Zhangsuo Liu, Fang Wang, Dandan Huang, Shun Wang, Shuwang Ge, Quanquan Shen, Ping Zhang, Lihua Wu, Miao Pan, Xiting Zou, Ping Zhu, Jintao Zhao, Minjie Zhou, Lin Yang, Wenping Hu, Jing Wang, Bing Liu, Tong Zhang, Jianxin Han, Tao Wen, Minghui Zhao, and Haiyan Wang. Acute kidney injury in China: a cross-sectional survey. *The Lancet*, 386(10002):1465–1471, October 2015. ISSN 01406736. doi:10.1016/S0140-6736(15)00344-X. URL https://linkinghub.elsevier.com/retrieve/pii/S014067361500344X.

[109] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965. ISSN 0019-9958. doi:https://doi.org/10.1016/S0019-9958(65)90241-X. URL http://www.sciencedirect.com/science/article/pii/S001999586590241X.

[110] Lijing Zhou, Licheng Wang, Yiru Sun, and Tianyi Ai. AntNest: Fully Non-interactive Secure Multi-party Computation. Technical Report 735, 2018. URL http://eprint.iacr.org/2018/735.

[111] Lijing Zhou, Licheng Wang, Yiru Sun, and Pin Lv. BeeKeeper: A Blockchain-Based IoT System With Secure Storage and Homomorphic Computation. *IEEE Access*, 6:43472–43488, 2018. ISSN 2169-3536. doi:10.1109/ACCESS.2018.2847632. URL https://ieeexplore.ieee.org/document/8386749/.

[112] Guy Zyskind, Oz Nathan, and Alex Pentland. Enigma: Decentralized Computation Platform with Guaranteed Privacy. *arXiv:1506.03471 [cs]*, June 2015. URL http://arxiv.org/abs/1506.03471. arXiv: 1506.03471.

Hossain KORDESTANI

Design and Development of an Intelligent and Secure E-Health System Based on Probabilistic Reasoning and Blockchain Technology

Résumé : Avec la longévité et le taux croissant de personnes âgées, il est vital de permettre aux personnes âgées d'avoir une meilleure qualité de vie avec des coûts réduits en utilisant l'e-santé. Le manque de fiabilité des données, l'incertitude des règles médicales, les problèmes de sécurité et la personnalisation posent de nombreux défis pour la création d'un système de l'e-santé intelligent et sûr; dans cette thèse, nous abordons les défis mentionnés. Tout d'abord, nous avons proposé un cadre de télésoins auto-adaptatif. Ce cadre fournit un télémonitorage basé sur l'IdO avec une sélection intelligente d'actions de détection pour fournir une image holistique des patients avec un minimum d'intrusions. Deuxièmement, nous avons proposé un cadre sécurisé pour traiter les questions de robustesse et de respect de la vie privée dans les réseaux IdO. Le cadre proposé est basé sur la technologie de la chaîne de blocs et du stockage distribué pour garantir l'intégrité et la disponibilité dans un réseau décentralisé. Le cadre proposé permet de préserver la confidentialité des données grâce à l'IA, ce qui permet à l'IA de crypter les données sans y donner accès. Troisièmement, nous avons proposé un cadre sûr et robuste pour le partage des données dans le domaine de l'e-santé, ou plus généralement dans les réseaux IdO. Le cadre proposé utilise des algorithmes cryptographiques pour assurer un partage sécurisé des données sans révéler aucune donnée personnelle et en évitant les redondances.

Mots clés : E-Santé, Sécurité, Blockchain, IA, Chiffrement homomorphe et diffusion.

Abstract : With longevity and a growing rate of the elderly, it is vital to enable the elderlies to have better quality of lives with reduced costs using e-health. The unreliability of data, the uncertainty of medical rules, security concerns, and personalization pose many challenges in creating an intelligent and secure e-health system; in this thesis, we address the mentioned challenges. Firstly, we have proposed a self-adaptive telecare framework. This framework provides IoT-based telemonitoring with a smart selection of sensing action to provide a holistic image of patients with minimal intrusions. Secondly, we have proposed a secure framework for handling robustness and privacy issues in IoT networks. The proposed framework is based on blockchain technology and distributed storage to ensure integrity and availability in a decentralized network. The proposed framework enables privacy-preserving AI on the data, allowing AI to analyze data without providing any leverage. Thirdly, we have proposed a secure and robust framework for data sharing in e-health, or generally in IoT networks. The proposed framework uses cryptographic algorithms to provide secure data sharing without revealing any personal data and while avoiding redundancy.

Keywords : E-Health, Security, Blockchain, AI, Homomorphic and Broadcast Encryption.